INVESTIGATION OF TDRD7-BASED REGULATION IN THE OCULAR LENS BY

INTEGRATED MULTIOMICS APPROACH

by

Salma Mohammed Al Saai

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the
Master of Science in Bioinformatics and Computational Biology

Summer 2020

INVESTIGATION OF TDRD7-BASED REGULATAION IN THE OCULAR LENS BY

INTEGRATED MULTIOMICS APPROACH

by

Salma Mohammed Al Saai

Approved: _____

Salil A. Lachke, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Cathy Wu, Ph.D.
Director of the Center for Bioinformatics & Computational Biology

Approved: _____

Levi T. Thompson , Ph.D.
Dean of the College of Engineering

Approved: _____

Douglas J. Doren, Ph.D.
Interim Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

ACKNOWLEDGMENTS

Thanks for everyone who supported and accouraged me through my journey to obtain this degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Clouding of the eye lens, termed cataract, is the leading cause of blindness worldwide. Mutations in *Tdrd7* (Tudor domain containing protein 7) causes congenital cataract in human and mouse. Tdrd7 protein is an RNA granule component that is involved in several aspects of post-transcriptional regulation. Tdrd7 protein is predicted to participate in protein-protein interactions via its Tudor domains and OST-HTH/LOTUS domains. MicroRNAs (miRNAs), a class of small noncoding RNAs, are also known to regulate gene expression post-transcriptionally and have been implicated in the pathogenesis of cataract. However, the impact of Tdrd7 deficiency on miRNAs in the lens is unexplored. In this study, we used *Tdrd7*-targeted germline knockout (*Tdrd7*-/-) mouse, which exhibit fully penetrant cataract, as a model to investigate the impact of Tdrd7 on miRNAs. We performed RNA-seq to analyze small RNA expression in postnatal day 15 *Tdrd7-/-* mouse lens. Differential expression analysis identified significantly mis-expressed miRNAs in *Tdrd7-/-* mouse lens prior to the detection of overt cataract suggesting their potential involvement in gene expression control. Using a multiomics integrated approach, the analysis of these differentially expressed miRNAs in the context of genome-wide mRNA expression profiling data and proteome data on *Tdrd7-/-* mouse lenses informs on the regulatory network underlying lens defects resulting from Tdrd7 deficiency. In sum, this work identifies miRNAs downstream of Tdrd7, and predicts their potential targets, in turn providing new evidence that supports a role for post-transcriptional regulatory factors in lens development and early-onset cataractogenesis.

Chapter 1

INTRODUCTION

1.1 The lens

The eye is a multicomponent organ and the ocular lens is located in its anterior chamber (Figure 1.1). The lens is an avascular tissue which consists of lens epithelial cells positioned at the anterior region and lens fiber cells at the posterior region that make up majority of the lens tissue (Bassnett et al., 2011; Bhat, 2001).

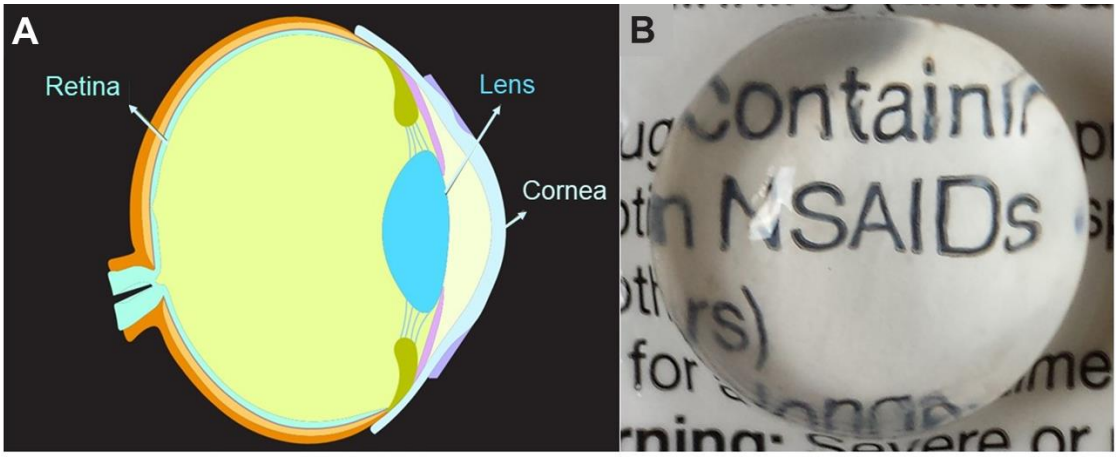


Figure 1.1 **The ocular lens is a transparent component of the eye**
(A) The ocular lens is located in the anterior chamber of the eye (B) The lens is a transparent living tissue (Image in (A) Adapted from Dash et al. 2016; Image in (B) by Salma Al Saai)

Throughout life, epithelial cells proliferate until they reach the "transition zone" at the lens equator, where they exit the cell cycle and begin to terminally differentiate into

fiber cells (Figure 1.2) (Bhat, 2001). Differentiating fiber cells start to elongate while they are being displaced from the lens equator towards the lens center. Fiber cells at the center are considered mature once they undergo degradation of their organelles and nuclei. Organelle degradation eliminates light scattering objects from the lens and creates an organelle free zone around the visual axis (Figure 1.2). This process renders the lens with a high refractive index that is critical for lens transparency (Bassnett, 2009).

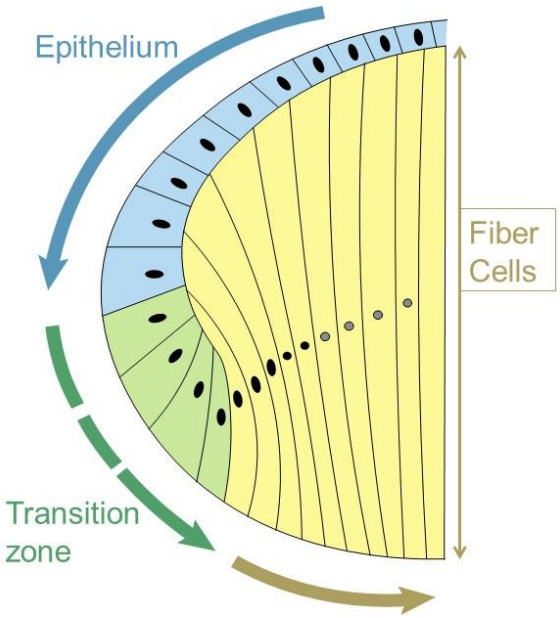


Figure 1.2 **Organization of cells in the lens**
A schematic illustrating the ocular lens tissue structure. Epithelial cells mitotically divide until they reach the "transition zone" at the lens equator, where they exit the cell cycle and starts to differentiate and elongate into fiber cells. This process continues until fiber cells starts to degrade their organelles and nuclei to create a light scatter-free zone around the visual axis of the lens (Adapted and modified from Dash et al. 2016)

1.2 Cataract

Clear and sharp vision relies on light being sharply focused by the transparent cornea and lens onto the retina (Bazan, 1989). According to the National Institute of Health (NIH), clouding of the eye lens, termed cataract, contributes to ~50% of blindness worldwide (Graw, 2009). The NIH-eye projects that 50% of the elderly in the United States (who are above the age of 75) will be affected by age related cataracts. Furthermore, cataracts that are identified at birth or within the first year of life are classified as congenital (Shiels and Hejtmancik, 2013). Currently, cataract surgery is the only available treatment option, which is successful in restoring vision. However, surgery is costly. Additionally, in congenital cataracts well timed surgery is critical to prevent permanent blindness (Kuhli-Hattenbach et al., 2008; Yorston, 2004). It is estimated that between 25% and 50% of congenital cataracts are due to genetic alteration (Shiels and Hejtmancik, 2015). Genetic mutations can be caused by external factors such as rubella infection affecting the fetus in expecting mothers or can be due to inheritance (Bardram and Brændsdrup, 1947; Mets, 2001)

1.3 Tdrd7 is linked to cataract in human and mouse

In human, the *TDRD7* (Tudor domain containing 7) gene (OMIM: 611258) is located on chromosome 9 while in mouse, it is located on chromosome 4. The use of integrated System Tool for Eye gene discovery (iSyTE) lead to the discovery of TDRD7 as new cataract-linked gene (Lachke et al., 2011a, 2012). *Tdrd7* deficiency causes early onset cataract in human, mouse and chicken (Lachke et al., 2011a). After the initial report in 2011, several other reports independently confirmed the importance of *Tdrd7* for lens transparency (Chen et al., 2017; Tan et al., 2017; Zheng et al., 2014).

1.4 Tdrd7 structure and function

Tdrd7 consists of Tudor domains that are predicted to bind to methylated arginines within other proteins, and OST-HTH (Oskar-TDRD5/TDRD7-Helix-Turn-Helix)/LOTUS (Limkain, Oskar, and Tudor containing proteins 5 and 7) domains that are predicted to bind to RNA(Anantharaman et al., 2010; Chen et al., 2011; Côté and Richard, 2005; Cui et al., 2013). Tdrd7 co-staining with Pyronin Y, a dye that binds RNA, indicated the close association of Tdrd7 with RNA in the lens (Lachke et al., 2011a). Additionally, it was recently shown that the OST-HTH/LOTUS domains of Tdrd7 can bind and stimulates the DEAD-box RNA helicase (Jeske et al., 2015, 2017), that function in regulating several aspects of RNA metabolic processes (Linder and Jankowsky, 2011), which further suggests the close association of Tdrd7 to RNA and post-transcriptional control.

Figure 1.3 **A schematic representation of the TDRD7 Protein**
TDRD7 contains three OST-HTH/LOTUS domains that are predicted to bind to RNA. And three Tudor domains that are predicted to bind methylated arginine residues within other proteins

In the ocular lens *Tdrd7* transcripts are highly abundant in fiber cells and is conserved among vertebrates (Figure 1.4) (Lachke et al., 2011a). Tdrd7 protein colocalizes with cytoplasmic RNA in a granular pattern in the ocular lens. Interestingly, the granular pattern described in the lens tissue agrees with Tdrd7's earlier description of specialized RNA granules, known as chromatoid bodies, in differentiating sperm (Kotaja and Sassone-Corsi, 2007; Marcello and Singson, 2011).

During spermiogenesis Tdrd7 dynamically remodels chromatoid bodies and due to its essential role, deficiency of *Tdrd7* causes azoospermia and results in male sterility in human and mice (Lachke et al., 2011a; Tanaka et al., 2011a; Tan et al., 2017).

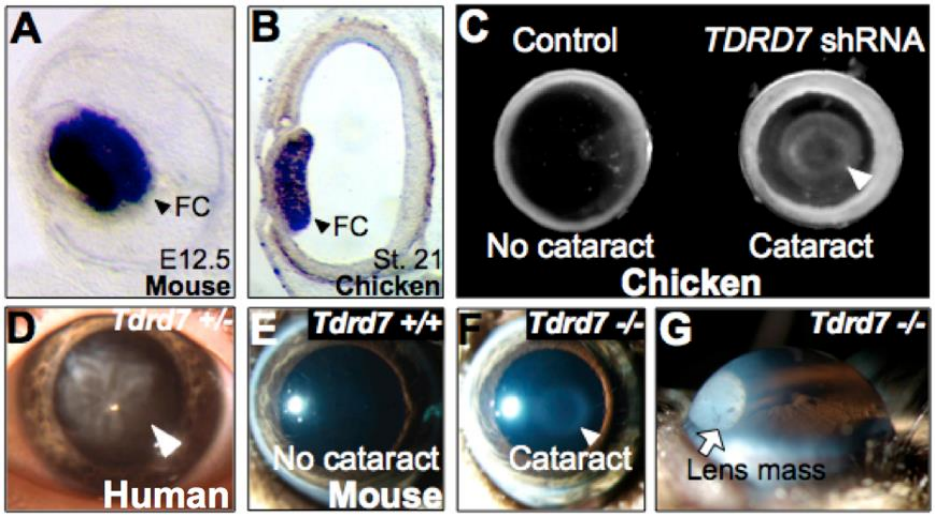


Figure 1.4 **Tdrd7 is expressed in the lens and its deficiency causes cataract**
In situ hybridization for *Tdrd7* mRNA in mouse (A) and chicken (B) lens fiber, both indicates the abundance of Tdrd7 transcripts in lens cells. (C) knockdown of *Tdrd7* in chicken lenses causes cataract. *Tdrd7* Deficiency in human (D) or mouse (E-G) causes cataract and lens mass dislocation (Adapted from: Lachke et al., 2011a)

In the lens, Tdrd7 variably co-localizes with several types of RNA granules, such as transport ribonucleoproteins and processing bodies (Lachke et al., 2011a). Proteins that associate with RNA granules are known to function as post-transcriptional regulators of gene expression (Anderson and Kedersha, 2009; Kishore et al., 2010). Post-transcriptional regulators mediate their control on their target RNA by several ways such as; transporting transcripts to a specific subcellular localization, promoting transcripts stabilization or channeling them to degradation (Figure 1.5)

(Anderson and Kedersha, 2009; Dash et al., 2016; Jiang et al., 2014). However, the mechanism by which Tdrd7 mediates its post-transcriptional control in the lens and how its deficiency affects results in cataract is not well understood.

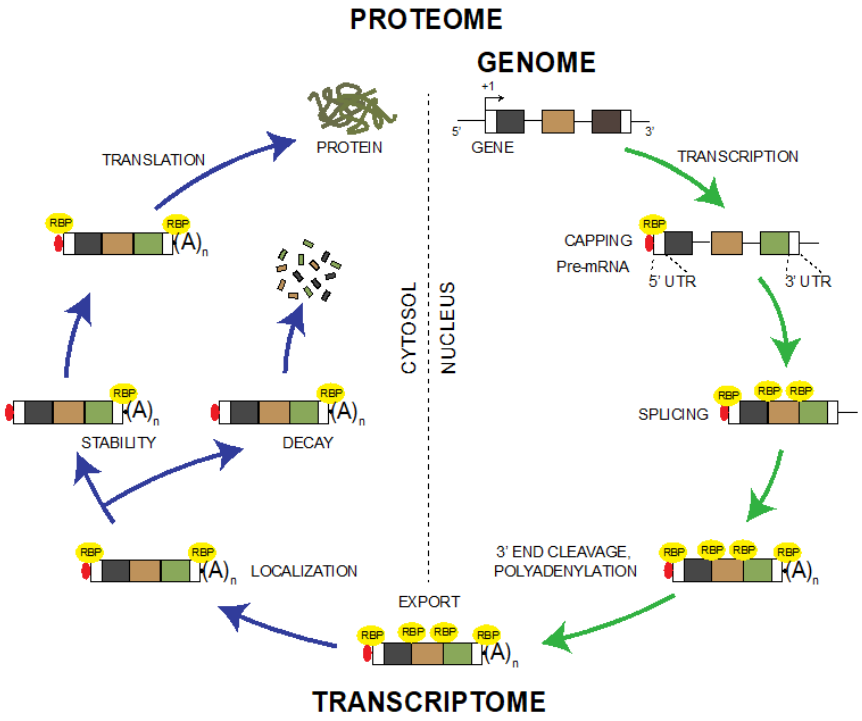


Figure 1.5 Function of **post-transcriptoinal regulators in controlling gene expression**

Post-transcriptoinal regulators mediate a distinct level of gene expression control on their transcript targets. Depending on the cellular context they can: stabilize their transcript targets; transport a transcript target to specific subcellular localization or channel it to degradation (Adapted from Dash et al. 2016)

1.5 *Tdrd7-/-* mice exhibit lens fiber cell defects and misregulated transcriptome prior to overt cataract formation

Previously generated *Tdrd7-/-* mice (Tanaka et al., 2011a) were subjected to detailed characterization recently (Barnum, Al Saai, Patel et al.,2020). At 3 months of

age, all *Tdrd7-/-* developed exhibit ocular and lens that were detected by light microscopy and histological analysis (Figure 1.6-A). However, these defects were not detectable at embryonic and early postnatal stages (E16.5) and (P4). Interestingly, while light microscopy and grid imaging found no discernable lens defects at P18, it identified 100% penetrant cataract only four days later at (P22) in *Tdrd7-/-* mice (Figure 1.6-C) (Barnum, Al Saai, Patel et al.,2020). In addition, abnormal lens fiber cell morphology was identified by scanning electron microscopy (SEM) at stage P18 in *Tdrd7-/-* lenses. And phalloidin staining of *Tdrd7-/-* at stage P15, uncovered fiber cells morphological defects particularly in cells undergoing nuclear degradation. Furthermore, detailed molecular characterization of *Tdrd7-/-* lens,14 days prior to the onset of cataract at stage P4, by RNA-Seq led to the identification of 280 differentially expressed genes at 1.5-fold change cutoff and <0.05 significant *P*-value (Barnum, Al Saai, Patel et al.,2020).

These results suggest that while *Tdrd7*-/- lenses are apparently normal and transparent at stages P18, P15 and P4, their fiber cells manifest signs of severe abnormalities on the cellular and molecular level. These pathological abnormalities are initiated prior to the onset of the phenotype, when not discernible by light microscopy or histology, and their gradual accumulation over time lead to the manifestation of severe opacities by stage P22 (Barnum, Al Saai, Patel et al.,2020).

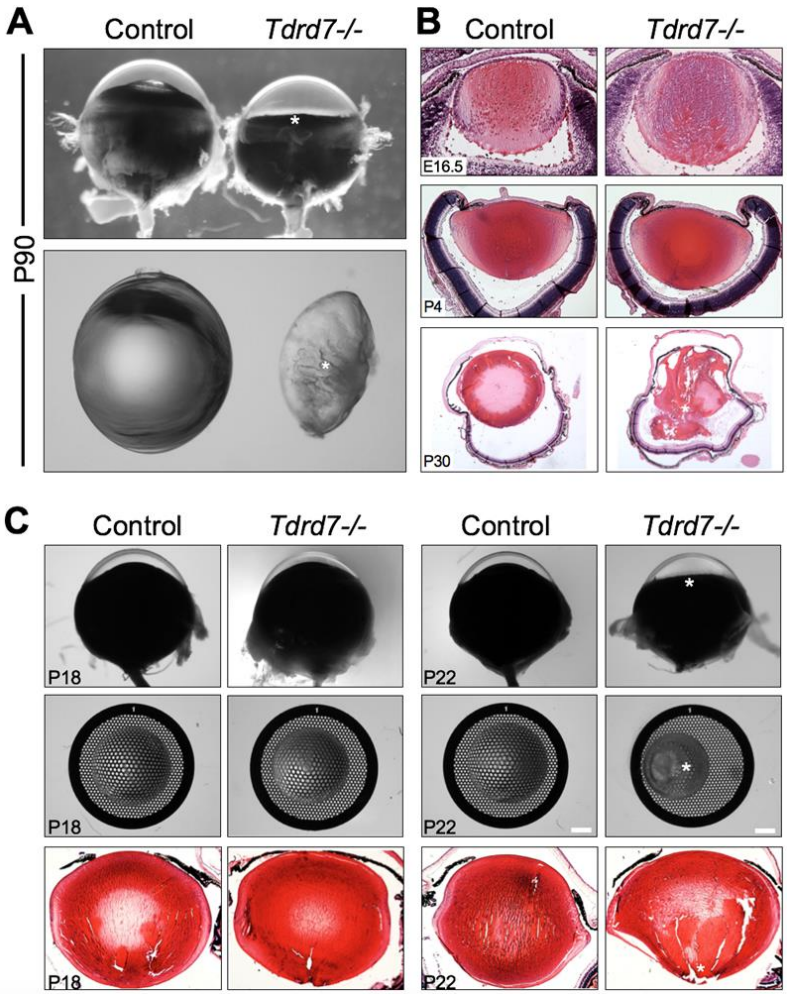


Figure 1.6 **Phenotypic characterization of *Tdrd7-/-* lens and eye defects**. (A) *Tdrd7-/-* mouse lenses exhibit eye defects (left panels, light microscopy). *Tdrd7-/-* eye exhibits a flattened iris and a smaller lens with posterior rupture and cataract is observed in (right panel indicated with asterisks). (B) Histology of *Tdrd7+/-* (control) and *Tdrd7-/-* lenses at various embryonic (E) and postnatal stages. *Tdrd7-/-* lenses show severe lens defects. (C) *Tdrd7-/-* mouse exhibit fully penetrant ocular defects by P22. Light microscopy of *Tdrd7+/-* (control) and *Tdrd7-/-* eye and lens. At P22 *Tdrd7-/-* mice exhibit cataract (asterisk, right image middle panel) and a flattened iris (asterisk, left image middle panel) while control appear normal. Histology shows defects in P22 *Tdrd7-/-* lens (Adapted from Barnum, Al Saai, Patel et al.,2020)

1.6 miRNA in lens biology

MicroRNAs (miRNAs) are small non coding RNAs (~19-25 nucleotide long) (Lu and Rothenberg, 2018; Rolle et al., 2016). miRNAs are important post-transcriptional regulators for gene expression. It is estimated that about 50% of protein coding genes (*i.e.* their mRNAs) are regulated by miRNA in a cellular dependent manner (Stefani and Slack, 2008). miRNAs regulate a wide span of cellular processes such as apoptosis, embryonic stem cell control, cellular metabolic pathways and neuronal differentiation (Dong et al., 2014; Houbaviy et al., 2003; Pasquinelli and Ruvkun, 2002; Sempere et al., 2004).

The biogenesis of miRNAs start in the nucleus and continue in the cytoplasm. miRNAs are initially transcribed by RNA polymerase II in the nucleus which gives rise to a long transcript known as pri-miRNA (Lee et al., 2002). A primary processing step is then performed by Drosha and DGCR8 along with other factors. During primary processing step the long pri-miRNA is trimmed into a ~65 nucleotide pre-miRNA(Wang et al., 2007). Exportins in the nucleus recognizes a short stem of 2–3 nucleotide 3′ overhangs present on the pre-miRNA. This recognition mediates the transportation of pre-miRNA into the cytoplasm (Lund et al., 2004). Once in the cytoplasm, Dicer catalyzes the second processing step. Dicer is a conserved enzyme that is present in nearly all eukaryotic organisms (Bernstein et al., 2003). The second processing step generates miRNA/miRNA* duplex. The strand that remains from the duplex is known as the mature miRNA, while the miRNA* (also known as passenger strand) is usually degraded. Mature miRNAs mediate their function by binding to the 3′ untranslated region of its mRNA target. This binding initiates the assembly of Argonaute (Ago) family proteins into an effector complex known as RNA-induced silencing complex (RISC). This ribonucleoprotein complex controls gene expression

via sequence-specific target gene silencing (Kim et al., 2009; Kobayashi and Tomari, 2016).

Animals that lack miRNAs either die during early embryonic stages or are born with sever developmental defects. This indicates the importance of miRNA function in post-transcriptional gene regulation (Bernstein et al., 2003; Morita et al., 2007; Wang et al., 2007).

Recently, several studies have gathered evidence on the importance of miRNAs regulation in the eye and its implication in the pathogenesis of ocular diseases, such as cataract. In particular, miRNAs are implicated in regulating the Fgf pathway in the lens (Wolf et al., 2013). Several studies have also embarked on miRNA profiling to identify deregulated miRNAs in defective lenses (cataract and myopia) (Kubo et al., 2013; Tanaka et al., 2019). Furthermore, many researchers have investigated the functional roles of misregulated miRNAs in lens development, the pathogenesis of cataract and post capsular opacification (PCO) (Bitel et al., 2012; Dong et al., 2014; Li and Piatigorsky, 2009; Li et al., 2015). However, more research is needed to gain further understanding of the molecular mechanism by which misregulation of miRNAs impact cataract formation.

1.7 The aim of this thesis work

Tdrd7 is a post-transcriptional regulator that is predicted to closely associated with target RNA either directly through its OST-HTH/LOTUS domains or via interactions with other proteins such as the DEAD-box helicase proteins that themselves interact with RNA (Anantharaman et al., 2010; Cui et al., 2013). Further, Tdrd7 was shown to colocalize with RNA and with several classes of ribonucleoprotein complexes in the

lens (Lachke et al., 2011b) However, the mechanism by which Tdrd7 functions in mammalian lens and how its deficiency results in cataract is not fully understood.

Likewise, miRNAs are post-transcriptional regulators that mediate their control of gene expression by initiating the assembly of a ribonucleoprotein complex (RISC). As noted above, miRNAs have been implicated in cataracts (Filipowicz et al., 2008; Kim et al., 2009; Kobayashi and Tomari, 2016). However, the molecular mechanism by which miRNA misregulation underlie cataract formation is poorly understood.

High-throughput total RNA-sequencing for *Tdrd7-/-* lens at postnatal stage P4, fourteen days prior to the onset of cataract, identified the misregulation of several candidate genes that are key for lens biology (Barnum, Al Saai, Patel et al.,2020). However, the impact of *Tdrd7* deficiency on miRNAs in the lens is unexplored.

In this thesis, I will use *Tdrd7*-targeted germline knockout (*Tdrd7*-/-) mouse, which exhibit 100% penetrant cataracts similar to those in humans, as a model to investigate the impact of Tdrd7 on miRNAs. I will do so by performing small RNA-seq to analyze miRNA expression on the lens of postnatal day 15. This stage is selected because at this age, *Tdrd7-/-* lenses are apparently normal and transparent under light microscopy, but exhibit lens fiber cell morphological defects as described in section 1.50 above. Therefore, the analysis of miRNA from small RNA-seq profiles will potentially identify miRNAs that are downstream of Tdrd7 and are important for lens homeostasis and transparency.

These experiments will also me to contribute miRNA expression data generated from the wild type control mouse lens at P15 to the existing (publicly available) set of miRNA normal expression, which represent 6 developmental stages prior to P15. This will extend the existing knowledge of miRNA normal expression in the lens to an additional postnatal stage, which will further our understanding on specific miRNA expression patterns during lens development.

In addition, using a multiomic approach I will integrate data sets obtained from genome-wide mRNA expression profiling data and proteome profiling data on P15 *Tdrd7-/-* mouse lenses to gain a comprehensive understanding on the miRNA regulatory network underlying Tdrd7 deficiency-based cataracts.

Through this thesis work my aim is to improve our understanding on the mechanism of cataract pathogenesis caused by Tdrd7 deficiency, as well as expand our knowledge on normal miRNA expression and the impact of miRNA misregulation on the lens. This will advance our understanding on post-transcriptional regulatory factors in lens development and early-onset cataractogenesis.

Chapter 2

MATERIALS AND METHODS

2.1 Animals studies

Animal studies were conducted in accordance to the Association for Research in Vision and Ophthalmology (ARVO). The University of Delaware Institutional Animal Care and Use Committee (IACUC) approved all experimental protocols that were used for the animal studies. In this thesis work I used *Tdrd7* targeted germline knockout (KO) mouse line that was generated and described earlier (Tanaka et al., 2011a). *Tdrd7* targeted germline KO mice lack Tdrd7 genomic region between exon 8 to exons 12. The University of Delaware Animal Care Facility housed the mice in a 14 hour light to 10 hour dark cycle. Males of *Tdrd7* homozygous null (*Tdrd7-/-*) are sterile (Tanaka et al., 2011b). Thus fertile *Tdrd7-/-* females were bred to *Tdrd7+/-* male mice. However this breeding scheme could not generate a wild-type (*Tdrd7+/+*) to be used as a litter-mate control for the designed experiments. To address this problem, *Tdrd7+/-* males and females were bred to generate *Tdrd7+/+* experimental controls. The resulting male and female offsprings with *Tdrd7+/+* genotype were furthur bred to continualy generate *Tdrd7+/+* for experimental controls.

2.2 DNA isolation from mouse tails

Genomic DNA mouse-tail tissue was used for genotyping. DNA was isolated from tails using Gentra Puregene Mouse Tail Kit (QIAGEN, Cat No. 158267). Mouse-tails were initially digested overnight at 57°C with 300µl cell lysis buffer and 1.5µl of

Protinase K 10 mg/ml (Invitrogen,Grand Island, NewYork). Digested tails were then treated with 1.5µl RNase A solution (QIAGEN, Cat No. 158922) at 37°C for 45 minutes, then brought to room temperature for 15 minutes and were placed on ice for 1 minutes. To precipitate proteins, 100µl of protein precipitation solution (supplied in the kit) were added to the digested tails, followed by vortexing for 20 seconds then centrifugation at 130000g for 5 minutes . The supernatant (~300µl) was transferred to new 1.7 ml Eppendorf tubes containing 300ul of absolute isopropanol for DNA precipitation. Before these Eppendorf tubes being centrifuged at 13.0000g for 5-15 minutes they were gently inverted several times. DNA precipitation was followed by carefully removal of the supernatant and the DNA pellet allowed to air dry to remove residual isopropanol. 300µl of 70% ethanol was added to wash the precipitated DNA pellet and the samples were centrifuged at 13.0000g for 5-15 minutes. After carefully draining the ethanol, the Eppendorf tubes containing the DNA pellet were allowed to air dry completely at room temperature or at 37°C. DNA hydration solution (20µl-50µl) was added to dried DNA pellets and placed on a heating block at 65°C for 1 hour. Isolated DNA Samples were then moved to a rocker and left overnight to completely dissolve the DNA. DNA concentrations were determined using NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo fisher scientific). DNA samples were preserved at -20°C until needed for PCR.

2.3 Mouse genotyping

Isolated DNA described in section 2.2 above from mouse tail tissue was diluted (200ng/µl). In the polymerase chain reaction (PCR) (total reaction volume, 25 µl) in 0.2 ml tubes (VWR: Cat No.490003-692), 1 µl of diluted genomic DNA was added. The following describes the preparation of one PCR reaction mix: 18 µl

nuclease free water; 2.5 µl of 10x CoralLoad PCR Buffer from (QIAGEN: Cat No. 201205); 1.25µl of Dimethyl sulfoxide (Sigma-Aldrich Cat No. D2650); 0.2 µl Taq DNA Polymerase from (QIAGEN: Cat No. 201205); 0.5 µl of prepared 10 mM dNTP set mix (QIAGEN: Cat No.201913); 0.5µl of 100µm of forward primers; 0.5 µl of 100µm of reverse primers The following are the primer sequences used for genotyping: Tdrd7-WT-g-F 5'-GAG TAA CTC TGG GCG CAG TC-3', Tdrd7-WT-g-R 5'-GCC ATA GCA ATC AGT GAG CA-3', expected product size 250 bp; and Tdrd7-KO-g-F 5' GTC TAA CCC ATT CAG GGA TGA AGA 3', Tdrd7-KO-g-R 5' GAA TCC TCA CCA GTT AGC CTC ACC 3'. The expected size product of this PCR is 500 bp, as previously described (Tanaka et al., 2011a). All steps of the PCR preparation were performed on ice. The tubes with the PCR reaction mix were briefly centrifuged before being placed on BioRad® Thermal Cycler PCR machine. The PCR were run under the following conditions for Tdrd7 genotyping PCR: 95°C for 5 minutes, 35 cycles of (95 °C for 30seconds, 56°C for 1 minutes, 42°C for 30 sec.). For short term storage the PCR product was kept at 4°C , while for long term storage the PCR product was kept at -20°C for long term. 1.5% agarose gel (The Lab Depot Cat No. TLDA1705) was used for gel electrophoresis. The agarose gel was mixed with 15µl of 1% Ethidium Bromide Solution (Thermo Fisher Scientific Cat No. BP1302-10). Finally, a 100 base pair ladder (Thermo Fisher Scientific Cat No. SM0242) was loaded on the agarose gel along with the samples to determine facilitate the determination of the amplicon size of the PCR.

2.4 Total RNA sequencing for *Tdrd7-/-* at P15 in mouse lens

Lens tissue were dissected from postnatal stage 15 (P15) of *Tdrd7-/-* and *Tdrd7*+/+. One mouse (2 lenses) were used per biological replicate and a total of three

biological replicates were collected. Dissected lens tissue were immediately flash frozen on dry ice (after being placed in 1.7 ml Eppendorf RNASE free tubes) then were stored at -80°C. mirVana™ miRNA isolation kit (Thermo fisher scientific, Cat No. AM1560) was used for RNA extraction from flash frozen lens tissue using according to the manufacturer's instructions. Extracted total RNA were shipped to DNA LINK were they were assessed for the RNA quality before proceeding to sequencing. All six samples had RIN >7 and total RNA yield >1 µg. All shipped samples passed the quality assessment and were then subjected to library preparation using Truseq Stranded Total RNA H/M/R Prep kit. Illumina Novaseq6000 platform was used to perform Paired-end RNA-seq and sequencing depth was at 50 million reads.

2.5 Analysis of P15 total RNA seq data

Raw FASTQ files were subject to adaptor trimming and removal low quality bases for paired-end reads using Trim Galore version 0.6.5 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). This was followed by mapping of trimmed reads against Mus musculus reference genome (GRCm38) using HISAT2 version 2.2.0 (Kim et al., 2015, 2019). Raw counts of mapped reads were obtained using HTSeq version 0.11.2 (Anders et al., 2015). EdgeR package version 3.28.0 was used for differential expression analysis in R/Bioconductor version. 3.6.2 (Robinson et al., 2010) Benjamini & Hochberg was used for *p*-values and multiple comparisons using FDR method of (1995).

2.6 *Tdrd7-/-* lens of P15 Tandem Mass Tag (TMT) Mass spectrometry

Five P15 mice of each *Tdrd7-/-* and *Tdrd7+/+* lens tissue were dissected (one mouse (2 lenses) was used per biological replicate) followed immediately by flash freezing (lens tissue placed in 1.7 ml Eppendorf RNase free tubes on dry ice) and kept at -80°C until shipping. Collected lens tissue were shipped to Oregon Health & Science University for TMT Mass spectrometry and data analysis. All sent samples were labeled with 10-plex TMT isobaric labeling reagents (Thermo Fisher Scientific). The tryptic digests were then fractionated to obtain sufficient depth for the proteome profiling. Data was acquired by Thermo Orbitrap Fusion instrument to reduce interference and increase reporter ion dynamic range using the SPS MS3. An established pipeline known as PAW (https://github.com/pwilmart/PAW_pipeline) was data processing. Comet was used with mouse reference proteome (2018.01_UP000000589_10090_Mus-musculus_canonical_for. fasta). Differential expression and statistical analysis for the proteome was performed in R/Bioconductor using edgeR package. Jupyter notebook was used to perform exact test on two-sample comparison.

2.7 Small RNA sequencing for Postnatal stage 15 in *Tdrd7-/-* mouse lens

Samples described in section 2.4 were also used for stranded small RNA-seq library preparation using NEBNext® Multiplex Small RNA Library Prep. Single end reads were obtained using Illumina Nextseq500. FASTQ files were downloaded from DNA LINK server. Trimmomatic v0.36 (Bolger et al., 2014) was used for adaptor trimming and removal of low quality bases for single-end reads. Trimmed reads were mapped against Mus musculus reference genome (GRCm38) using Bowtie2 version

2.3.5.1 (Langmead, 2010; Langmead et al., 2009). BAM files were used for studying fragment length distribution for small RNA population in the lens.

2.8 Analysis of miRNA sequencing data

Mature miRNA reference sequence FASTA file was downloaded from microRNA database miRBase (Kozomara and Griffiths-Jones, 2014; Kozomara et al., 2019). FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit) was used to convert miRNA reference sequences to DNA. After conversion, trimmed small RNA reads described above in section 2.7 above were mapped against mature miRNA sequences using Bowtie2 version 2.3.5.1 (Langmead, 2010; Langmead et al., 2009) using default options (end-to-end) which does not allow alignments of mismatches. miRNA sorted alignment BAM files were used to obtain raw counts via samtools idxstat (Li et al., 2009).

2.9 Differential expression analysis of miRNA

Differential expression analysis was performed in R/Bioconductor version 3.6.2. Two packages were used to perform the same analysis: 1) EdgeR package version 3.28.0 (Robinson et al., 2010), 2) DESeq2 version 3.11 (Love et al., 2014). *p*-values were adjusted for multiple comparisons using False Discovery Rate (FDR) method of Benjamini & Hochberg (1995).

2.10 Hierarchical Cluster Heatmap

Hierarchical cluster heatmaps (clustered by rows) were generated in Jupyter notebook using Seaborn library in Python (v2.7). This was used for lens miRNA expression across 6 developmental stages obtained from (Khan et al., 2015) in addition miRNA expression obtained for P15 wild type from this work. RNA-seq

miRNA expression values corresponding to the following developmental stages were used to construct the heatmap: embryonic stage (E15), embryonic stage (E18), Postnatal stage (P0), Postnatal stage (P3), Postnatal stage (P6), Postnatal stage (P9) and Postnatal stage (P15). The miRNA expression were in form of RPM or CPM (Reads per million mapped reads or Counts per million mapped reads) and both are equivalent.

2.11 miRNA target prediction

List of differentially expressed miRNA in *Tdrd7-/-* lens at postnatal stage P15 were used for miRNA gene target prediction. miRDB database (Chen and Wang, 2020; Liu and Wang, 2019; Wang, 2008) was used for miRNA gene target mining. The search was set to include targets with targets score ≥ 60. In miRDB target scores are assigned via a computational target prediction algorithm. The higher the score (highest is 100), the more confidence the prediction is.

2.12 Identification of miRNA targets among differentially expressed genes and proteins in P15 of *Tdrd7-/-* lens

A data set for differentially expressed genes (Figure 2.1) and proteins (Figure 2.2) (described in section 2.5 and section 2.6) for *Tdrd7-/-* lens of P15 was used. The transcript list was composed of 1033 significantly misexpressed genes (FDR $\leq$ 0.05, 1.2 $\leq$ absolute fold change $\leq$ -1.2), while the protein list was composed of 234 significantly missexpressed proteins (P-value $\leq$ 0.05, 1.2 $\leq$ absolute fold change $\leq$ -1.2). The intersection between miRNA targets identified in section 2.11 above and the list of differentially expressed genes as well as proteins in *Tdrd7-/-* were identified in Jupyter notebook with a custom script for Pandas library in Python (v2.7). The intersection list of miRNA targets was used for downstream analysis.

2.13 Identification of lens enriched genes among differentially expressed genes and proteins in *Tdrd7-/-* lens

Lens enrichment data corresponding to gene and protein list described in section 2.12 above was downloaded from iSyTE 2.0 (Kakrana et al., 2018). Lens enrichment values indicate how much (by fold change) a specific gene is expressed in the lens when its expression is compared to the whole body (Kakrana et al., 2018; Lachke et al., 2012). Enrichment values were obtained for nine developmental stages corresponding to each list of genes and proteins described above in section 2.5 and section 2.6 in *Tdrd7-/-* of P15. Genes and proteins that have lens enrichment fold change that is $\geq$1.5 in at least one developmental stage across the nine stages were considered as lens enriched. The nine stages included early to late embryonic development as well as early postnatal to adult stages (E10.5, E11.5, E12.5, E16.5, E17.5, E19.5, P0, P2 and P56).

2.14 Building miRNA gene target network of *Tdrd7*

A Bipartite network for miRNA regulatory network of Tdrd7 was constructed using Cytoscape version 3.6.1. The nodes are miRNA (diamond shape) and miRNA target (ellipse shape). miRNA and miRNA targets (genes and proteins lists that overlapped miRNA targets predicted by miRDB, as described in section 2.12 above and illustrated in Figure 2.1Figure 2.2) used in this network are those that are significantly misexpressed in *Tdrd7-/-* lens of p15. The color of the nodes represents the absolute expression fold change between *Tdrd7-/-* and control in P15, which was obtained from preformed differential expression analysis. Blue nodes indicate reduced expression (downregulation) while red nodes indicate elevated expression (upregulation). The intensity of the color corresponds to the magnitude of the fold change. miRNA gene target nodes that are enriched in the lens tissue according to

iSyTE 2.0 (as described above in section 2.13 above) were distinguished with a font label that is red. The edge wight and color intensity corresponds to the target score obtained from miRDB indicating the prediction confidence level between miRNA node and miRNA target. Target scores in this data set ranges between (100 and 60), increased edge weight and color intensity represents higher target scores.

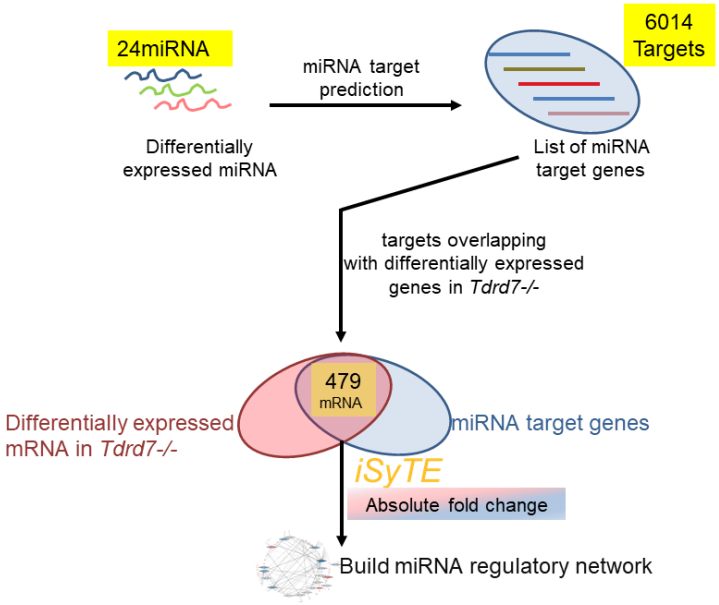


Figure 2.1 **Work flow for building Tdrd7 miRNA regulatory network integrated with RNA seq profiling data from *Tdrd7-/-* of P15**

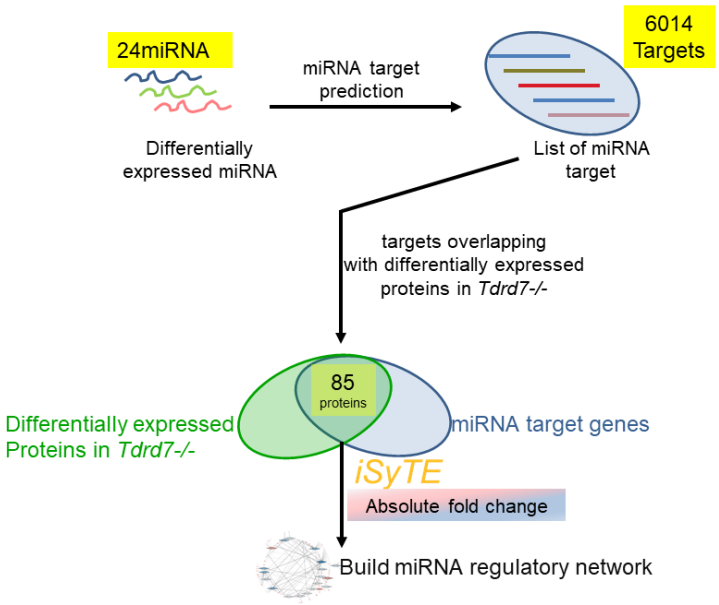


Figure 2.2 **Work flow for building Tdrd7 miRNA regulatory network integrated with proteome profiling data from *Tdrd7-/-* of P15**

2.15 Gene Set Enrichment Analysis (GESA) for miRNA targets

Gene set enrichment analysis for *Tdrd7-/-* lens at p15 (GESA) was performed on differentially expressed genes and differentially expressed proteins that were predicted to be targeted by any of the differentially expressed miRNA. GSEA was performed in R/Bioconductor version. 3.6.2 using the following packages: clusterProfiler, enrichplot, DOSE and org.Mm.eg.db for mouse database (Yu and He, 2016; Yu et al., 2012, 2015). The gene list was ranked based on absolute fold change and analysis for all Gene Ontology (GO) terms were considered significant at P-value=<0.05.

Chapter 3

RESULTS

3.1 Identification of variable fragment lengths from small RNA-seq in the lens tissue

Fragment size length distribution was identified using BAM files from mapped reads against Mus musculus reference genome (GRCm38) using Bowtie2 version 2.3.5.1. On average, 86% of the reads across all samples were mapped to the reference genome. The total mapped reads average across 6 samples was 26,082,246. The fragment size distribution of mapped reads ranged between 18-76 nucleotides. The highest two peaks were for reads that were 22 nucleotides and 76 nucleotides long. There was no significant difference observed in the distribution of the fragment size of mapped reads between *Tdrd7-/-* and its wild type control (*Tdrd7+/+*).

The fragment size distribution profile suggests the presence of the following small noncoding RNA population: fragments ranging between 19–25 nucleotides suggests the presence of miRNA (Rolle et al., 2016) while fragments ranging between 26-31 nucleotides suggests the presence of piRNA (Kirino and Mourelatos, 2007). Further, fragments ranging between 70-100 nucleotides suggests the presence of transfer RNA (tRNA) (Torres et al., 2014). Finally, fragments ranging between 60-76 nucleotides found in the data suggests the presence of small Cajal body-specific RNA which are reported to range in size between 60 to 300 nucleotides (Ronchetti et al., 2012).

For this thesis work, my focus will be on studying miRNA population.

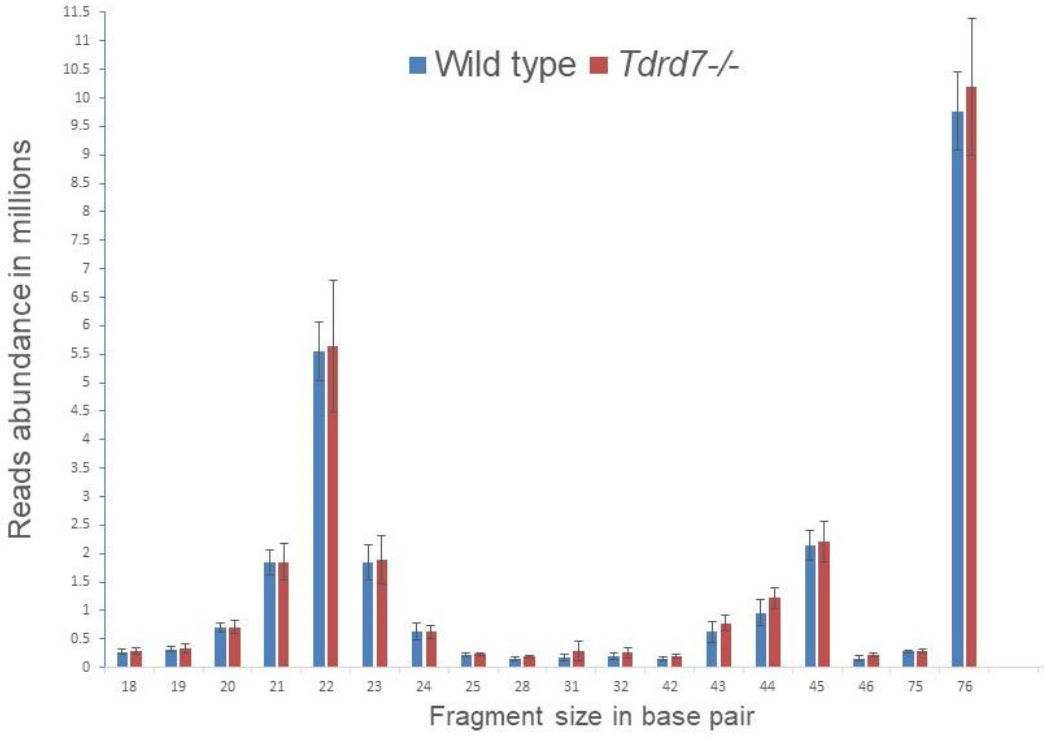


Figure 3.1 **Fragment size distribution of mapped reads in small-RNA seq of Tdrd7-/- at P15**

The fragment distribution of small RNA seq suggests the presence of more than one class of small noncoding RNA. Based on the fragment size the most abundant two peaks could be representing miRNA (22 nucleotides) and tRNA (76 nucleotides)

3.2 miRNA linked to eye defects are found among the misregulated miRNAs in *Tdrd7-/-* lens

miRNA counts were subject to differential expression analysis. EdgeR and Deseq were used to identify miRNA that are differentially expressed in *Tdrd7-/-* lens when compared to its wild type control (*Tdrd7+/+*). Analysis results from Both edgeR and Deseq identified 25 miRNA to be significantly misexpressed in *Tdrd7-/-* (Figure 3.2). However, edgeR identified additional 5 miRNA. Therefore, for all the

downstream analysis I used results that were obtained from edgeR, which consisted of 30 differentially expressed miRNA (Table 3.1).

A miRNA was considered to be differentially expressed if it has FDR $\leq$ 0.5 and $\geq$1.2 or $\leq$ -1.2 absolute fold change. Several miRNA that are linked to eye defects and cataract were among the 30 misregulated miRNAs identified in *Tdrd7-/-* lens, such as miR-124, miR-96 and miR-183 (Wu et al., 2017,Weston et al., 2018).

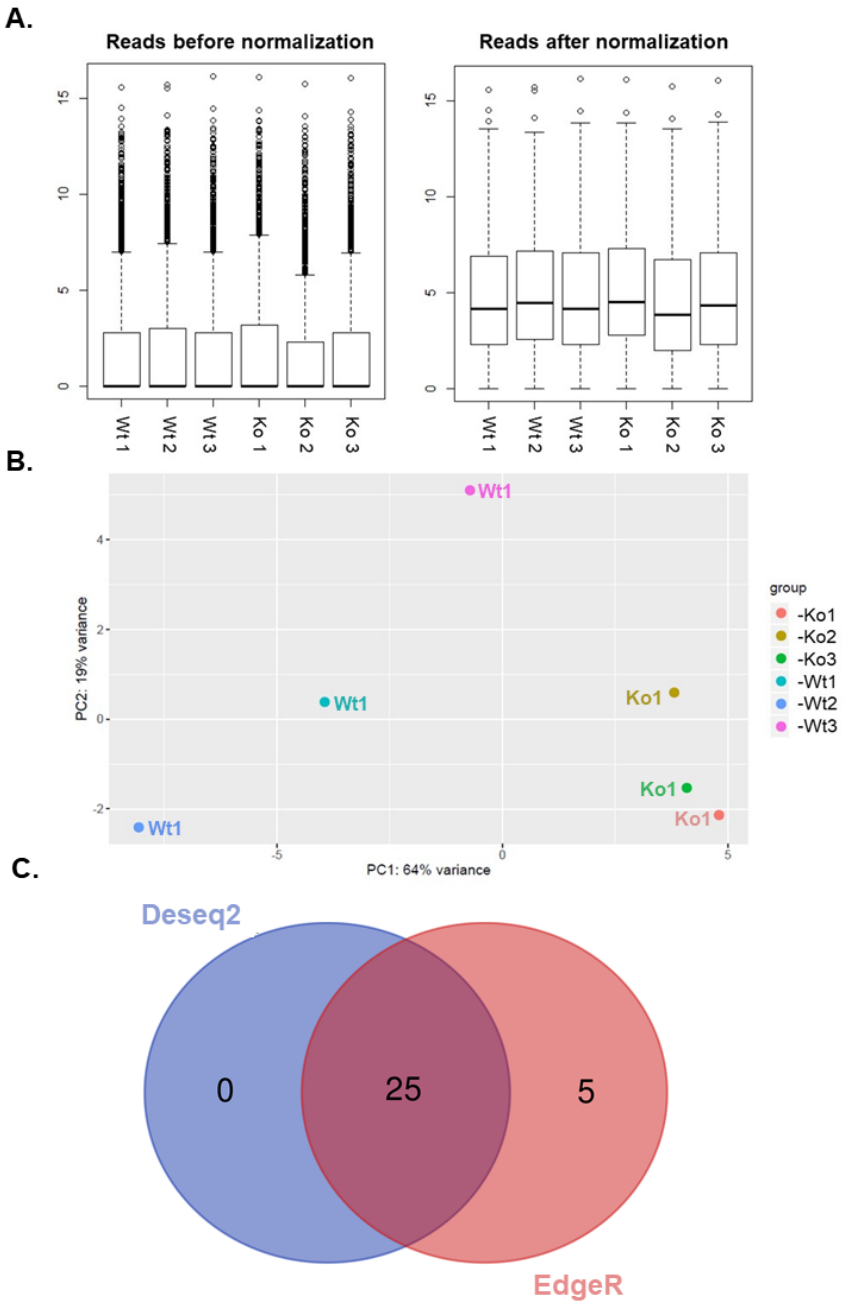


Figure 3.2 **Differential expression analysis for miRNA in edgeR and DESeq** (A) A box plot for the process of counts normalization in edgeR; on the left are the raw counts before normalization and on the right are counts after normalization. (B) a principal component analysis for the samples used in the differential expression analysis; wild type (Wt) control

samples are well separated from the *Tdrd7-/-* (Ko) samples on the right. (C) Venn diagram illustrating that overlap identified in the differential expression analysis results performed by edgeR and Deseq

Table 3.1 **List of differentially expressed miRNA in *Tdrd7-/-* lens**

All miRNAs in this table meet the following criteria (FDR≤ 0.5 and ≥1.2 or ≤ -1.2 absolute fold change). Order is based on absolute fold change. miRNA that are in white font are those that were identified by edgeR only and not DESeq.

miRNA Name	logFC	Absolute fold change	FDR
mmu-miR-124-3p	-5.0	-33.5	3.9E-05
mmu-miR-96-5p	-4.4	-20.6	8.4E-08
mmu-miR-9-5p	-4.3	-19.8	5.8E-09
mmu-miR-182-5p	-4.3	-19.5	3.0E-07
mmu-miR-9b-3p	-4.2	-19.2	1.5E-07
mmu-miR-183-5p	-4.2	-19.1	4.0E-06
mmu-miR-129-5p	-3.9	-16.3	5.0E-05
mmu-miR-9-3p	-3.3	-9.6	3.4E-05
mmu-miR-216b-5p	-2.8	-7.3	3.4E-02
mmu-miR-190b-5p	-2.8	-6.8	8.6E-04
mmu-miR-7b-5p	-2.7	-6.5	3.9E-05
mmu-miR-6540-5p	-2.6	-5.8	1.7E-02
mmu-miR-217-5p	-2.3	-4.6	2.5E-02
mmu-miR-488-3p	-2.2	-4.4	2.4E-05
mmu-miR-135b-5p	-1.8	-3.3	1.7E-02
mmu-miR-690	-1.8	-3.3	1.7E-03
mmu-miR-369-3p	-1.6	-2.9	1.6E-06
mmu-miR-323-3p	-1.4	-2.4	2.7E-02
mmu-miR-412-5p	-1.3	-2.4	6.5E-03
mmu-miR-666-5p	-1.1	-2.1	4.8E-04
mmu-miR-770-3p	-1.1	-2.1	3.8E-05
mmu-miR-433-3p	-1.1	-2.0	9.5E-03
mmu-miR-540-3p	-0.9	-1.7	3.1E-02
mmu-miR-543-3p	-0.7	-1.6	1.8E-02
mmu-miR-434-5p	-0.6	-1.5	2.8E-02
mmu-miR-351-5p	0.9	2.0	3.6E-02
mmu-miR-322-3p	1.0	2.1	1.0E-02

mmu-miR-126a-3p	1.0	2.1	4.0E-02
mmu-miR-126b-5p	1.5	3.1	9.6E-04
mmu-miR-3081-3p	1.7	3.6	2.0E-02

3.3 miRNA are important for lens biology through embryonic and postnatal stages

Next, I sought to examine the normal expression of miRNA in the lens during embryonic (E) and postnatal (P) developmental stages. For this purpose, I used miRNA-seq data described in (Khan et al., 2015). Data from Khan et al., have normal miRNA expression profiles for the following developmental stages: E15, E18, P0, P3, P6 and P9. To those developmental stages, I extended the miRNA profiling in the lens to P15 by adding miRNA expression data of the wild type control that was generated from work in this thesis. To visualize the pattern of miRNA expression through lens development, I generated a hierarchical cluster heatmap (clustered by row). With a miRNA expression cut-off $\geq$ 1 Count Per Million (CPM) the heatmap shows the profiles of 517 miRNA in the lens across 7 developmental stages (Figure 3.2A). The hierarchical cluster heatmap reveals a dynamic pattern for miRNA expression during normal lens development. For example, a cluster of miRNAs exhibits the highest lens expression during embryonic development (E15 and E18), and subsequently the expression of this group of miRNAs gradually decreases as lens development progresses. Another cluster of miRNA exhibits their highest expression at P15 while their expression remains relatively lower at earlier developmental stages.

The hierarchical cluster heatmap in (Figure 3.3A) also contains the normal expression profiles of miRNAs that are misregulated at P15 in *Tdrd7-/-* lens. To examine them closely, I extracted their normal expression profiles into a separate hierarchical cluster heatmap (Figure 3.3-B). Interestingly, the normal miRNA

expression profiles of upregulated miRNA in *Tdrd7-/-* lens clustered together at the bottom, and their normal expression is particularly abundant during embryonic and early postnatal stages. I also observed that the normal expression of miRNA that are downstream of *Tdrd7* are critical for all developmental stages, as the heatmap indicates that the miRNA abundance is variably distributed across all developmental stage. Some miRNAs are highly abundance during early lens development while others are highly abundant during postnatal.

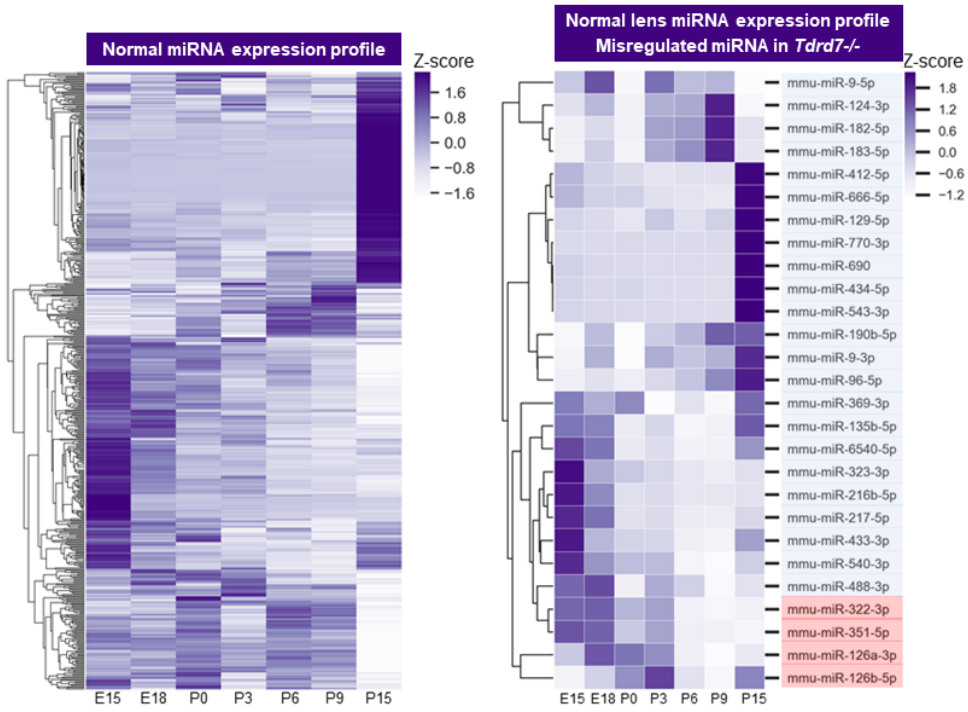


Figure 3.3 **Normal expression of miRNA in lens development**
Cluster heatmap displays a wide span of normal expression pattern across early and late developmental stages for normal and misregulated miRNA in the lens. (A) displays the normal expression profiles of 517. (B) displays the normal expression profiles for differentially regulated miRNA in *Tdrd7-/-*. *Note:* miRNA that are shaded in blue and red corresponds to downregulation and upregulation in in *Tdrd7-/-*, respectively

3.4 Target prediction integrated with multiomics misregulated transcriptome and proteome of *Tdrd7-/-* identifies miRNAs that control key lens genes

miRNAs that are differentially expressed in *Tdrd7-/-* lens at P15 were used for target gene prediction through the miRDB database (Chen and Wang, 2020; Liu and Wang, 2019; Wang, 2008). The following are miRNAs that miRDB did not identify gene targets for: mmu-miR-126b-5p, mmu-miR-182-5p, mmu-miR-3081-3p, mmu-miR-6540-5p, mmu-miR-690, mmu-miR-9-3p and mmu-miR-9b-3p.

A total of 6014 unique gene targets were identified for 24 miRNA. Further, 479 genes and 85 proteins that are differentially expressed in *Tdrd7-/-* lens at P15 overlapped with the miRNA predicted target list. All overlapping genes and proteins had the following cut off (FDR $\leq$0.05, and $\geq$1.2 or $\leq$ -1.2 absolute fold change). These data sets were used to construct a bipartite miRNA-target network for Tdrd7 which will be described in details below in section 3.5

3.5 Building Tdrd7 multiomic miRNA gene/protein target regulatory network

I used cytoscape to build two regulatory miRNA target networks based on differentially expressed miRNAs in *Tdrd7-/-* lens. The first network is based on miRNA targets overlapping with genome-wide mRNA expression profiling data in *Tdrd7-/-* lens (Figure 3.4). The second network is based on miRNA targets overlapping with proteome profiling data in *Tdrd7-/-* lens (Figure 3.8) Details on the data sets used can be found in Chapter 2, in the following sections: 2.4,2.5, 2.6, 2.9, 2.11, 2.12 and 2.13) Details on the network description can be found in section 2.14.

3.5.1 miRNA mRNA target regulatory network of *Tdrd7-/-*

A bipartite miRNA gene target (hairball) network (Figure 3.4) is composed of 503 nodes and 937 edges. The 503 nodes are divided between 24 miRNA nodes

(diamond shape) and 479 gene target (mRNA) nodes (ellipse shape). Among 479 mRNA target nodes, 233 (~48%) nodes (ellipse shape) that are located in the outer most circle are predicted to be regulated by one miRNA (Figure 3.4A). The reminder 246 (~51%) gene target nodes located in the inner most circle are predicted to be regulated by more than one miRNA (miRNA synergism) (Figure 3.4A). A break down for the level of miRNA synergism on their predicted gene targets is illustrated in (Figure 3.4B and Figure 3.4C). mRNA target nodes located in the inner most circle of the network (Figure 3.4A) was further broken into several inner circles based on the increased level of miRNA synergism. In (Figure 3.4B) the network illustrates miRNA synergism breakdown while all the edges are displayed. As for (Figure 3.4C) the edges were hidden for clear visualization of the break down. The inner second circle contains 119 mRNA target nodes (ellipse shape) that are predicted to be synergistically regulated by 2 miRNAs. The third inner circle contains 66 mRNA target nodes (ellipse shape) that are predicted to be synergistically regulated by 3 miRNAs. The inner fourth, fifth and sixth circles contains 47, 7 and 5 mRNA target nodes, respectively, and they are predicted to be synergistically regulated by 4 and 5 miRNAs, respectively. Finally, the predicted mRNA target RAR Related Orphan Receptor A (*Rora*), which is lens enriched, and the predicted mRNA target Pseudopodium Enriched Atypical Kinase 1 (*Peak1*) are synergistically regulated by 7 and 8 miRNAs, respectively.

To closely examine the gene target Peak1,which is synergistically regulated by the highest number of microRNAs in this network, I sub-selected Peak1 along with its first neighbors (Figure 3.5) and along with its second neighbors (Figure 3.6) Interestingly, six of the miRNA regulating Peak1 in the lens of *Tdrd7-/-* are down-

regulated at P15, which could partly explain to the elevated levels of Peak1 in *Tdrd7-/-* lens. To gain more insights into the miRNA predicted mRNA targets network, I have selected the most upregulated and the most down-regulated miRNA from the network along with their first node neighbors (Figure 3.7).

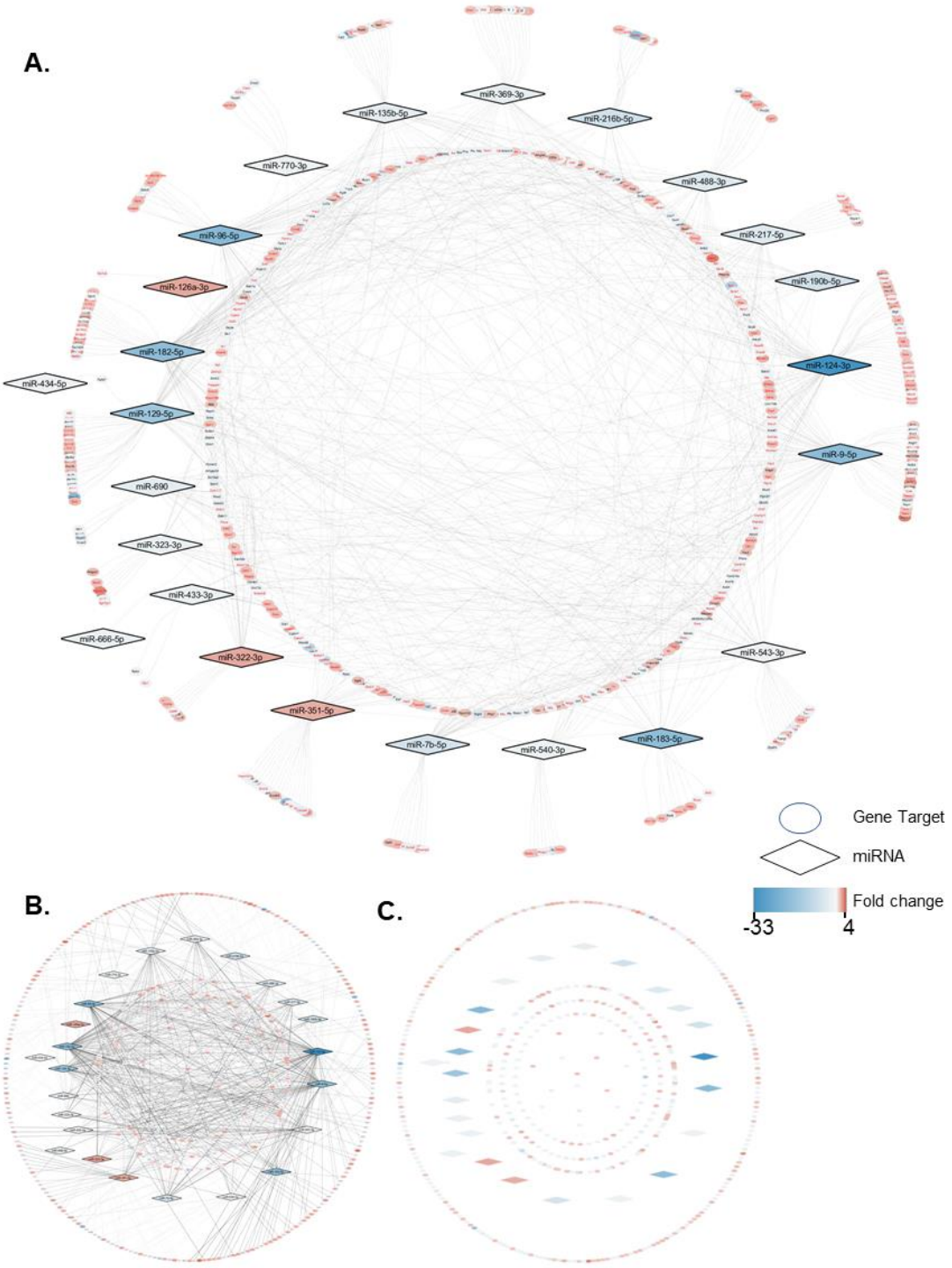


Figure 3.4 **miRNA gene target network in Tdrd7-/- lens**
(A) A hairball bipartite miRNA gene target regulatory network for Tdrd7 which is composed of 503 nodes and 937 edges. 24 miRNA

nodes (diamond shape) are connected to 479 gene target nodes (ellipse shape). 246 gene target nodes are predicted to be regulated by more than one miRNA. These gene target nodes are in the inner most circle. (B) The same network in (A) but rearranged to illustrate miRNA synergism. (C) represents the network in (B) but with hidden edges for better miRNA synergism visualization. The level of miRNA synergism increases as the size of the inner circle decreases. **Network key**: Diamond shape represents miRNA, ellipse shape represents gene target, red font label denotes lens enriched targets, edges wight increases according to the predicted target score from miRDB, node colores reflects fold change in *Tdrd7-/-* lens of P15 as shown in the legend and explain in the section 2.14

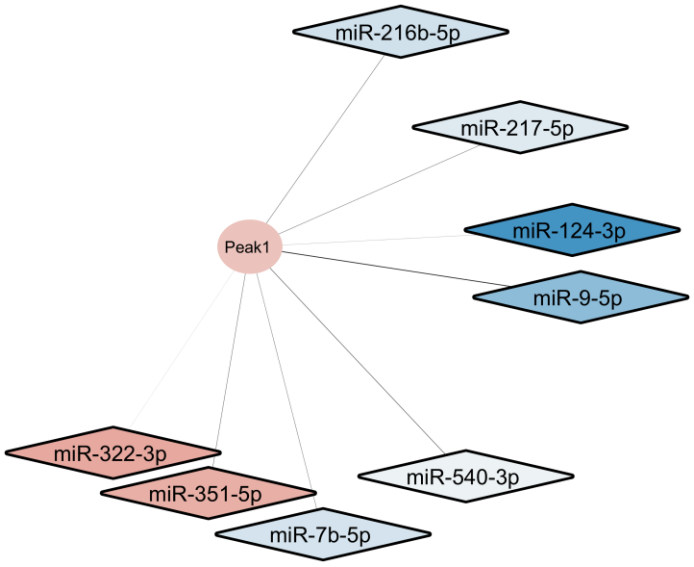


Figure 3.5 **Peak1 is predicted to be synergistically regulated by eight miRNA**
Peak1 gene predicted miRNA target was selected with its first neighbors from the miRNA gene target regulatory network of *Tdrd7-/-* lens at P15. Interestingly, 6 of the miRNA regulating Peak1 in the lens of *Tdrd7-/-* are down regulated at P15, which could partly explain to the elevated levels of Peak1 in *Tdrd7-/-* lens. **Network key**: Diamond shape represents miRNA, ellipse shape represents gene target, red font label denotes lens enriched targets, edges wight increases according to the predicted target score from miRDB, node colores reflects fold change in *Tdrd7-/-* lens of P15 as shown in the legend and explain in the section 2.14

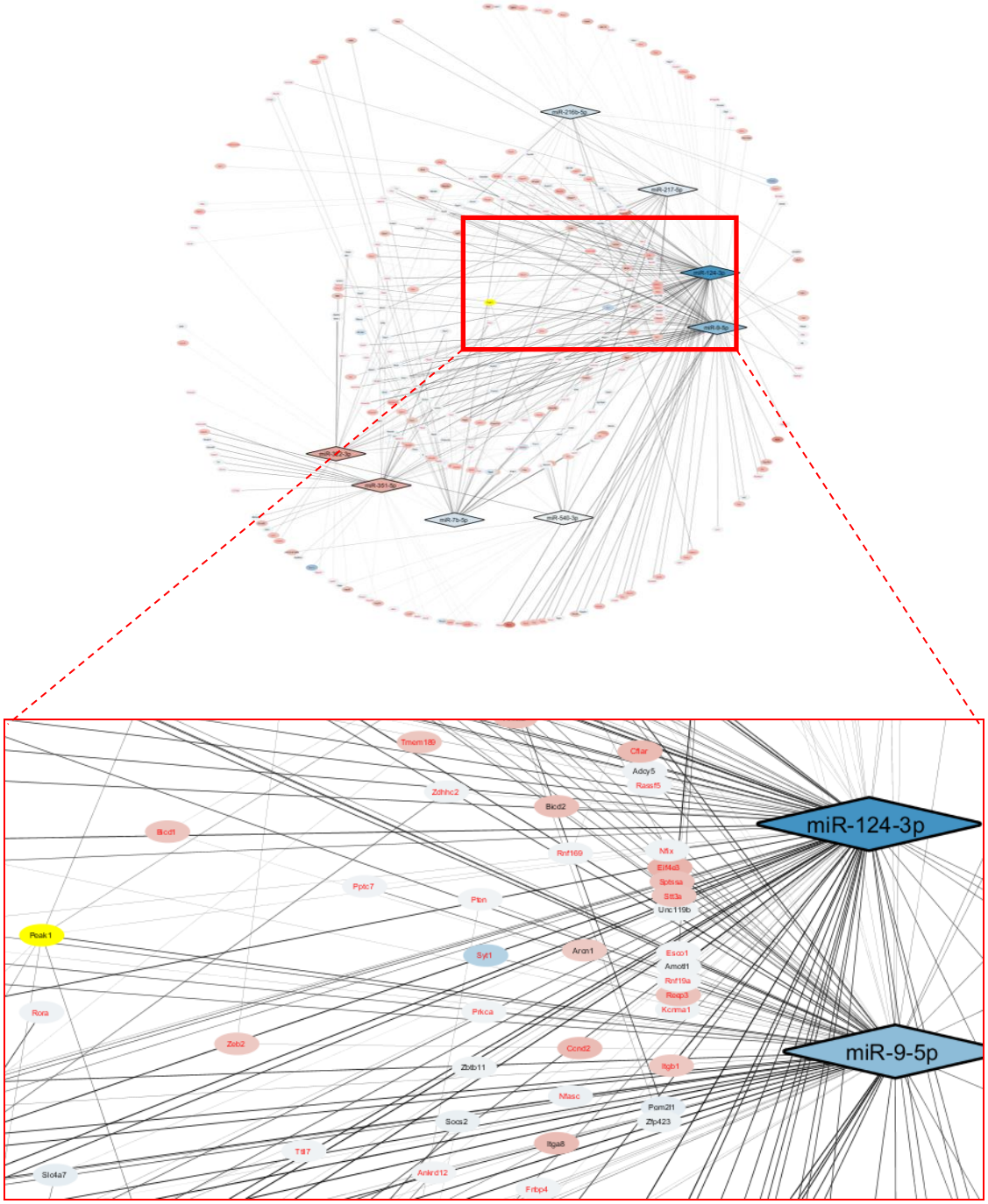
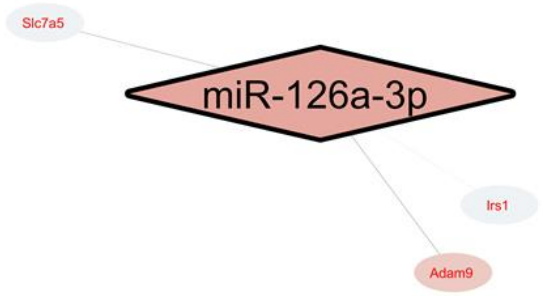


Figure 3.6 **Peak1 alone has majority (60%) of the network nodes when selected to its second neighbors**

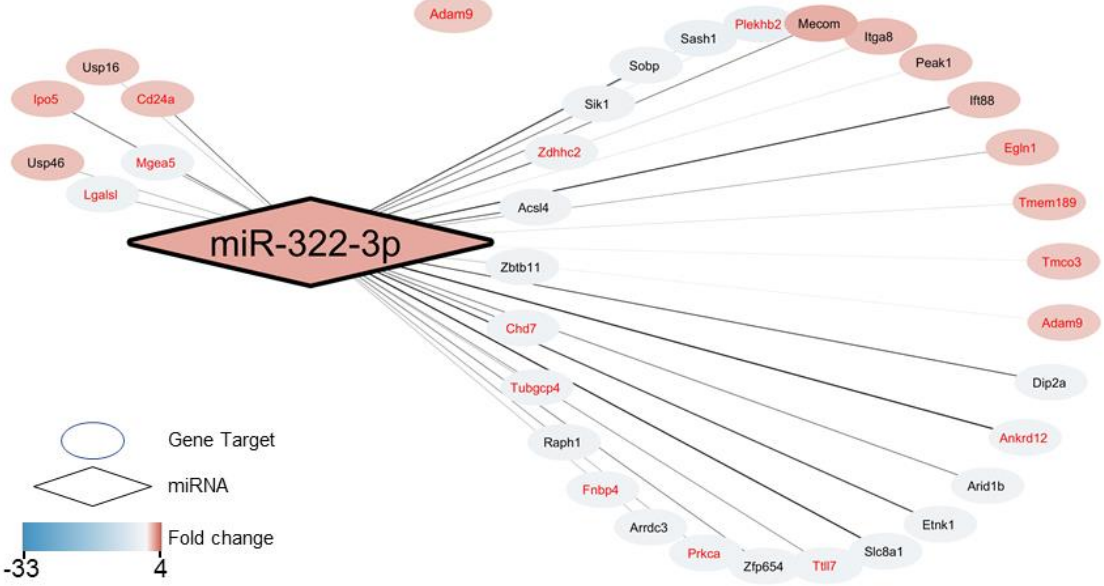
The density of connections to Peak1 gene target ,when selected with its first neighbors (as in Figure 3.), increases immensely when its selected to its second neighbors. Here, Peak1 network has 302 nodes connected

via 384 edges in contrast to 9 nodes connected via 8 edges when it was selected to its first neighbors. Although Peak1 is not lens enriched (node font label is black), many of its second neighbors among other miRNA predicted gene targets are lens enriched (node font label is red). **Network key**: Diamond shape represents miRNA, ellipse shape represents gene target, red font label denotes lens enriched targets, edges wight increases according to the predicted target score from miRDB, node colores reflects fold change in *Tdrd7-/-* lens of P15 as shown in the legend and explain in the section 2.14

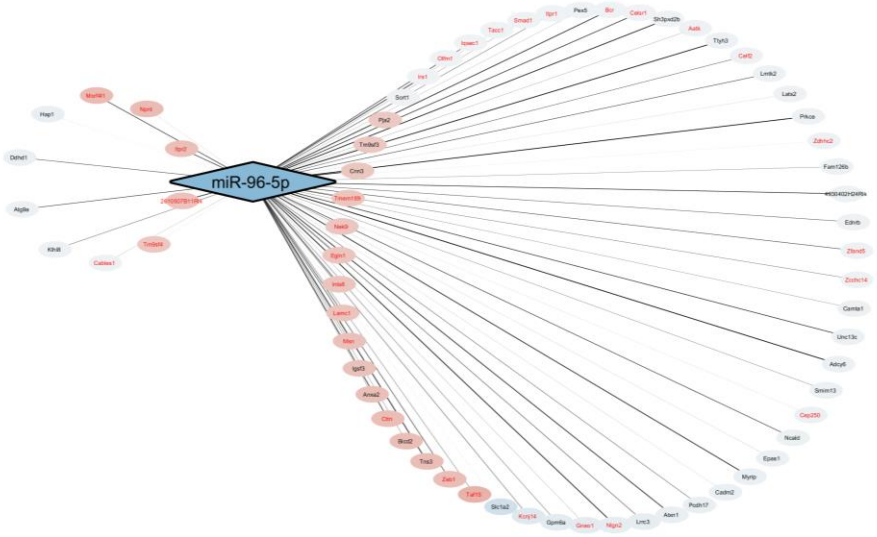
A.


Slc7a5
miR-126a-3p
Irs1
Adam9

B.


Usp16
Ipo5
Cd24a
Usp46
Mgea5
LgalsI
miR-322-3p
Sash1
Plekhb2
Mecom
Itga8
Sobp
Peak1
Sik1
Ift88
Zdhhc2
Egln1
Acsl4
Tmem189
Zbtb11
Tmco3
Adam9
Chd7
Dip2a
Tubgcp4
Ankrd12
Raph1
Arid1b
Fnbp4
Etnk1
Arrdc3
Prkca
Zfp654
Ttll7
Slc8a1
Gene Target
miRNA
Fold change
-33
4

C.


miR-96-5p

D.

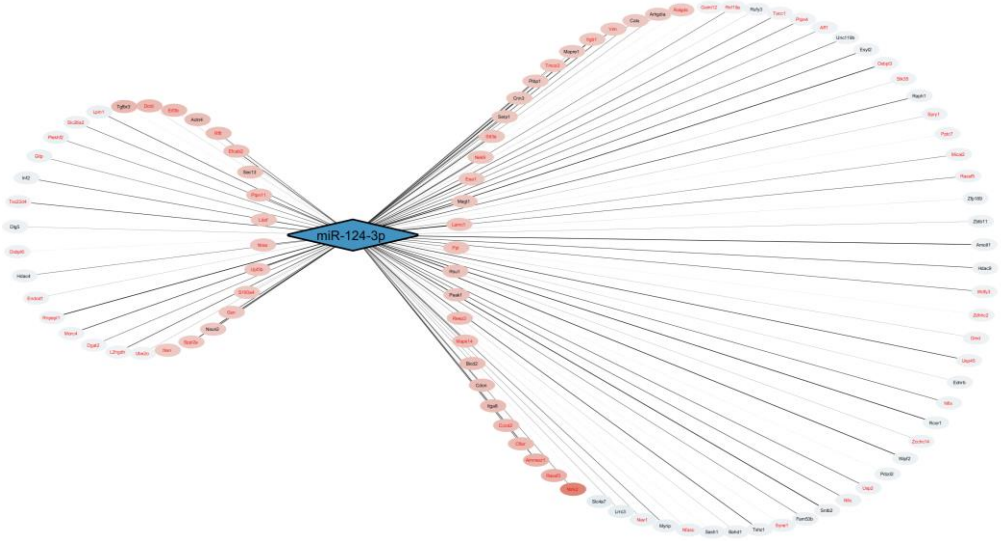


Figure 3.7 **Top misregulated miRNA sub selections from the regulatory miRNA predicted gene target network of Tdrd7-/- lens**
Four sub-selections (A), (B), (C) and (D) corresponding respectively to the two most upregulated and two most downregulated miRNA in Tdrd7 miRNA gene target network miRNA along with their first neighbors. In these networks, all gene target nodes that are located to the left of the miRNA node are unique predicted targets to that miRNA. All gene predicted target nodes located to the right of the miRNA node, are additionally predicted to be regulated by other miRNAs. **Network key**: Diamond shape represents miRNA, ellipse shape represents gene target, red font label denotes lens enriched targets, edges wight increases according to the predicted target score from miRDB, node colors reflects fold change in *Tdrd7-/-* lens of P15 as shown in the legend and explain in the section 2.14

3.5.2 miRNA protein target regulatory network of *Tdrd7-/-*

A bipartite miRNA protein target (hairball) network (Figure 3.8) is composed of 105 nodes and 150 edges. 105 nodes are divided between 20 miRNA nodes (diamond shape) and 85 protein target nodes (ellipse shape). Among 85 protein target nodes, 50 (~59%) nodes that are located in the outer most circle (ellipse shape) are predicted to be regulated by one miRNA (Figure 3.8). The reminder 35 (~41%) protein target nodes located in the inner most circle are predicted to be regulated via

miRNA synergism (Figure 3.8). The following is a break down for the level of miRNA synergism on their predicted protein targets: There are 18 protein target nodes (ellipse shape) predicted to be synergistically regulated by 2 miRNAs, 8 protein target nodes (ellipse shape) predicted to be synergistically regulated by 3 miRNAs, 6 and 2 protein target nodes that are predicted to be synergistically regulated by 4 and 5 miRNAs, respectively. Finally, the predicted protein target Microtubule Associated Protein 1B (Map1b) is synergistically regulated by 6 miRNAs. To closely examine the protein target Map1b, I sub-selected Map1b along with its first neighbors (Figure 3.9) and along with its second neighbors (Figure 3.10). Additionally, to gain more insights into the regulatory miRNA predicted protein targets network in *Tdrd7-/- lens*, I have sub-selected the most upregulated miRNA and the most down regulated miRNA (Figure 3.11) in *Tdrd7-/-* lens along with their first node neighbors.

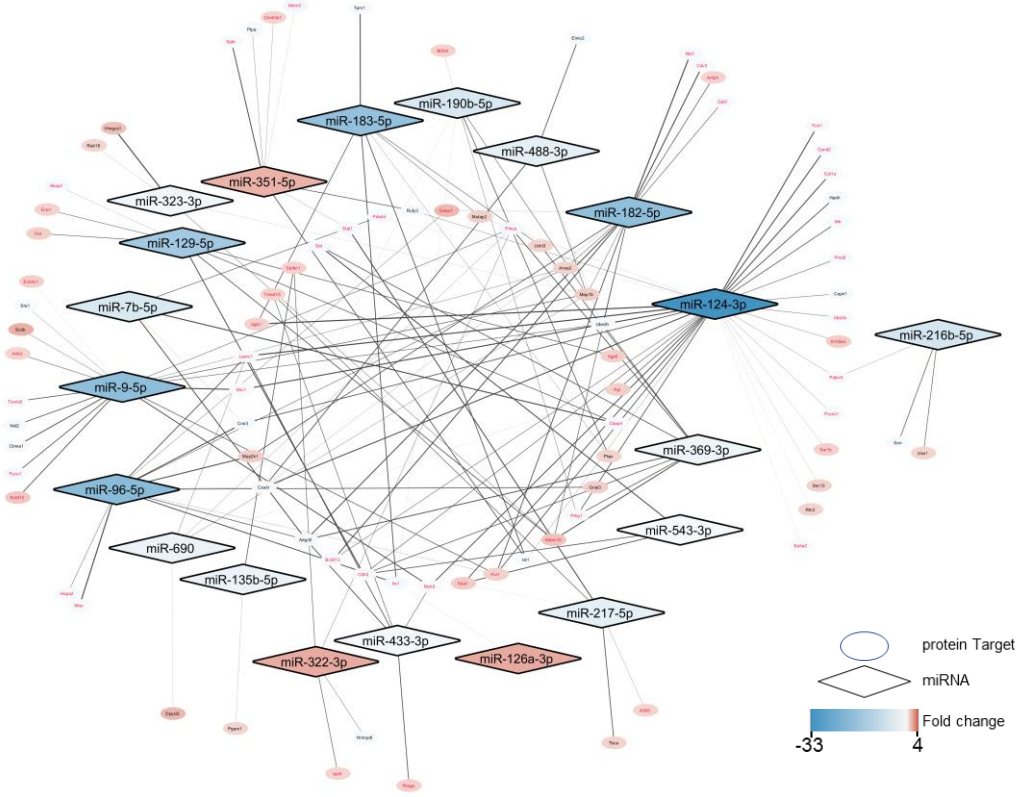


Figure 3.8 **miRNA protein target network in Tdrd7-/- lens**

(A) A hairball bipartite miRNA protein target regulatory network for Tdrd7 which is composed of 105 nodes and 150 edges. 20 miRNA nodes (diamond shape) are connected to 85 protein target nodes (ellipse shape). 35 protein target nodes are predicted to be regulated by more than one miRNA. These protein target nodes are located in the inner most circle. **Network key**: Diamond shape represents miRNA, ellipse shape represents gene target, red font label denotes lens enriched targets, edges wight increases according to the predicted target score from miRDB, node colors reflects fold change in *Tdrd7-/-* lens of P15 as shown in the legend and explain in the section 2.14

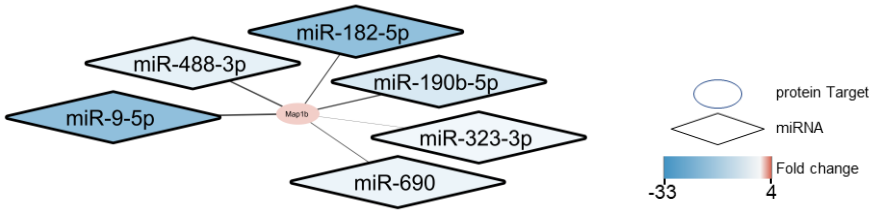


Figure 3.9 **Map1b protein is synergistically regulated by six miRNA in the lens**

Map1b protein predicted miRNA target was selected with its first neighbors from the miRNA protein targe regulatory network of *Tdrd7-/-* lens at P15. Interestingly, all the six miRNA regulating Map1b in the lens of *Tdrd7-/-* of P15 are down regulated. This strongly suggests that the upregulation of Map1b in *Tdrd7-/-* lens is due to the reduction of the six miRNA that are predicted to Map1b

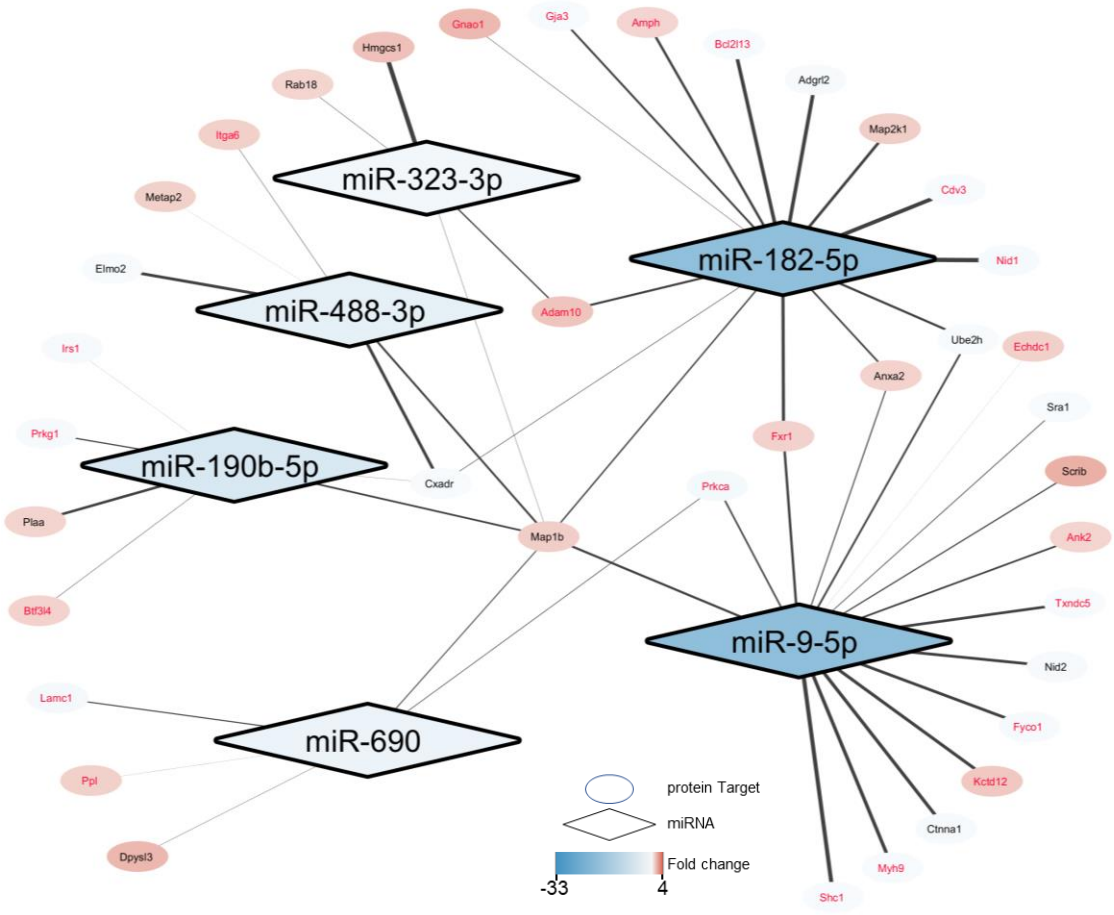
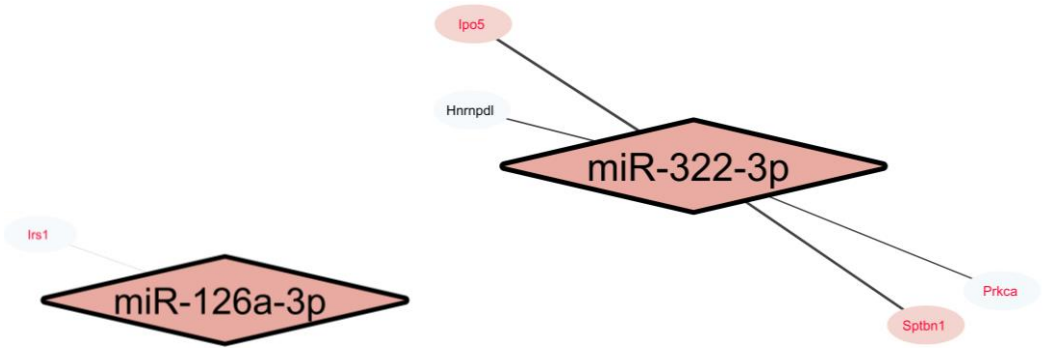


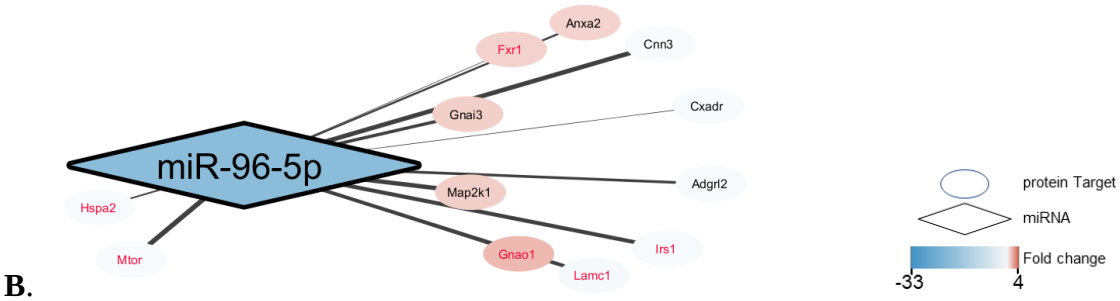
Figure 3.10 **Map1b protein sub-selected along with its second neighbors from miRNA protein target network of Tdrd7-/- lens**

The density of connections to Map1b protein target ,when sub-selected with its first neighbors (as in Figure 3.), increases immensely when its sub-selected with its second neighbors. Here, Map1b network has 44 nodes connected via 50 edges in contrast to 7 nodes connected via 6 edges when Mapb1 was selected with its first neighbors. Although Map1b is not lens enriched (node font label is black), many of its second neighbors among other miRNA predicted protein targets are lens enriched (node font label is red).

A.



B.



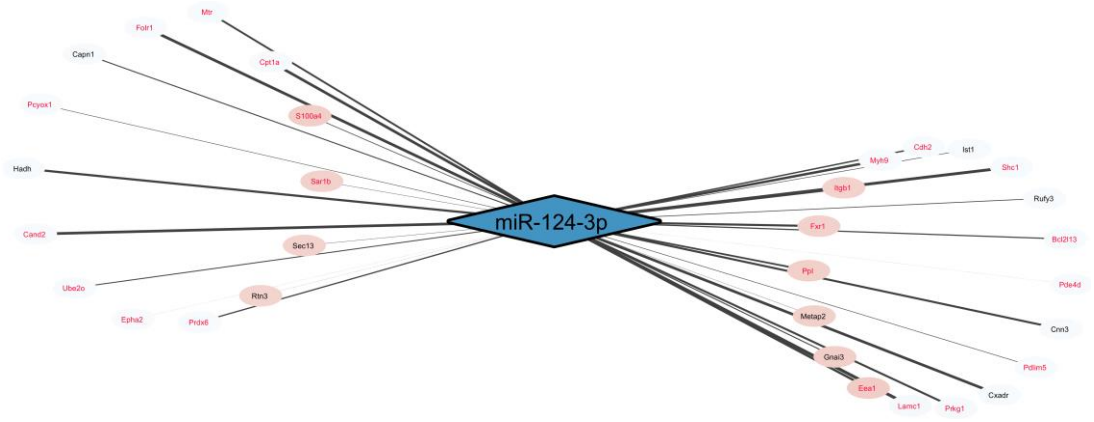


Figure 3.11 **Top misregulated miRNA sub selections from the regulatory miRNA predicted protein target network of Tdrd7-/- lens**
These are four sub selections corresponding to the two most upregulated miRNA (A) and two most downregulated miRNA (B) in Tdrd7 miRNA protein target network. miRNA along with their first protein target neighbors. In these networks, all protein target nodes that

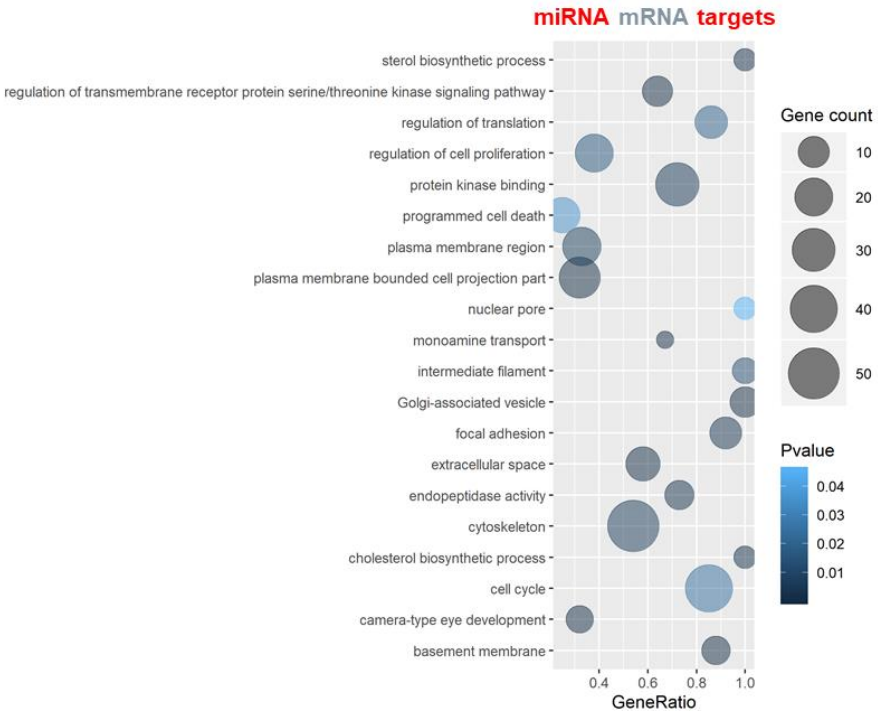
are located to the left of the miRNA node are unique predicted targets of that miRNA. All gene predicted target nodes located to the right of the miRNA node, are additionally predicted to be regulated by other miRNAs. **Network key**: Diamond shape represents miRNA, ellipse shape represents gene target, red font label denotes lens enriched targets, edges wight increases according to the predicted target score from miRDB, node colors reflects fold change in *Tdrd7-/-* lens of P15 as shown in the legend and explain in the section 2.14

3.6 Gene Set Enrichment Analysis (GESA) for miRNA predicted targets identifies processes that are implicated by the absence of *Tdrd7* in the lens

Gene set enrichment analysis is used to identify categories of genes or proteins that are enriched in a large predefined set genes or proteins. The implementation of statistical methods in GSEA enables the identification of significantly enriched sets for a given large profiling data of ranked genes or protein (Subramanian et al., 2005).
In the context of miRNA misregulation in *Tdrd7-/-* lens at P15, I performed GSEA on misregulated miRNA predicted targets which overlapped the transcriptome and the proteome of *Tdrd7-/-* lens. This analysis sheds light on significantly enriched terms that are linked to *Tdrd7-/-* deficiency in the lens, which in turn will add to our understanding on important biological processes that are contributing to cataract pathobiology in *Tdrd7-/-* lens.

GSEA Analysis uncovered terms such as camera-type eye development, cytoskeleton and basement membrane (Figure 3.12A) as well as cellular response to drugs and microtubules to be significantly among the enriched terms for miRNA predicted targets (Figure 3.12B).

A.



B.

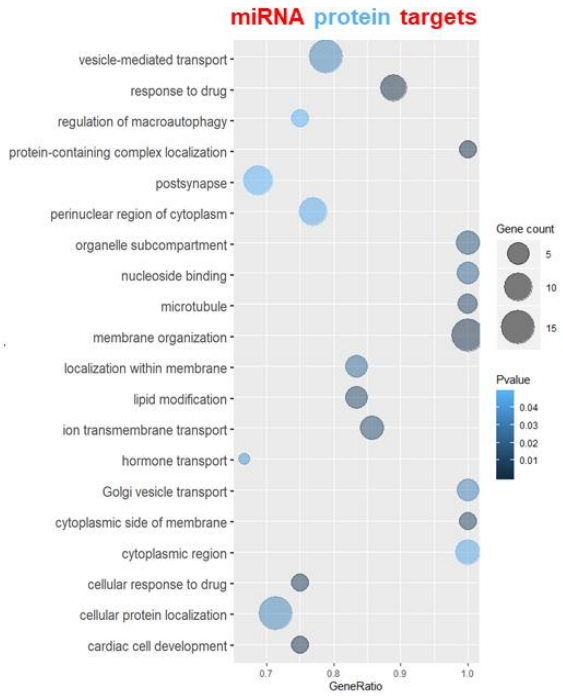


Figure 3.12 **Scatter plots for misregulated miRNA targets in Tdrd7-/- lens**

(A) Scatter plots for enriched GO categories for miRNA targets that overlap with *Tdrd7-/-* transcriptome at P15 (B) Scatter plots for enriched GO categories for miRNA targets that overlapped *Tdrd7-/-* proteome at P15. The y-axis represents the enriched GO categories and the x-axis represents the rich factor of the enriched GO terms. The dot size represents the gene count and the color of the dots represents the P-values. Rich factor is the ratio of differentially expressed gene number enriched in the pathway to the total gene number in a certain pathway

Chapter 4

DICUSSION

Tdrd7, a ribonucleoprotein/RNA granule component, functions in several aspects of post-transcriptional regulation and is linked to congenital cataract in human (Lachke et al., 2011b; Tan et al., 2017; Tanaka et al., 2011b, 2011a; Zheng et al., 2014). miRNAs are also known to regulate gene expression post-transcriptionally and have been implicated in the pathogenesis of cataract (Dong et al., 2014; Li and Piatigorsky, 2009; Tanaka et al., 2019). By investigating miRNAs in *Tdrd7-/-* mouse lens, my work adds a new dimension to our understanding of post-transcriptional regulation in the lens.

In this thesis I investigated the lens of *Tdrd7-/-* lens at P15. At this age, the lens is apparently normal under light microscopy and develops cataract 7 days later. In *Tdrd7-/-* lens I identified 30 significantly misregulated miRNAs. Interestingly, the normal expression pattern of these misregulated miRNAs exhibited variable abundance across embryonic and postnatal lens developmental stages; some miRNAs were highly abundant during embryonic stages such as miR-322-3p, while others were highly abundant during postnatal stages, such as miR-96-5p. This observation suggests that Tdrd7 is upstream of miRNAs that modulate targets that are critical for all lens developmental stages. And the absence of *Tdrd7* disrupts the fined tuned control mediated by these miRNAs on their targets causing accumulation of molecular changes that cause cataracts.

Several miRNAs that are misregulated in *Tdrd7-/-* lens have been implicated in cataracts and eye defects. For example, miR-124 that is downregulated in *Tdrd7-/-* lens, was shown to be important for the process of vesicle proteins transport in squid lenses (Bitel et al., 2012) and it is suggested that it is involved in the pathogenesis of congenital cataract in infants (Wu et al., 2017). Likewise, mouse model of miR-96 and miR-183 misexpression have been linked to cataracts (Weston et al., 2018). Both these miRNAs are reduced in *Tdrd7-/-* lens, further suggesting that *Tdrd7* is important for the key lens miRNAs.

In this thesis, I have taken a systematic multiomics approach for integrating three high throughput profiling data sets for *Tdrd7-/-* lens at P15 coupled with data extracted from two different public databases. This strategy aims to facilitate the identification of the crucial links and interactions among key misregulated miRNAs/ mRNAs/proteins which contributes to the deterioration of lens homeostasis in *Tdrd7* deficiency. To achieve these goals, I started with the analysis of miRNA from small RNA-seq of *Tdrd7-/-* lens. This was followed by identifying predicted targets of those misregulated miRNA from miRDB database. Subsequently, I integrated two omics-levels profiling datasets representing the transcriptome and the proteome of *Tdrd7-/-* and identified their overlap with predicted miRNA targets. I have also overlaid the three profiling datasets described above with their absolute fold change values in *Tdrd7-/-* lens. To enhance the miRNA predicted targets data sets, I added their lens enrichment information that was obtained from iSyTE database. Finally, I used this multiomics integrated data to build a comprehensive Tdrd7 miRNA-target regulatory network in the lens. This network serves as a powerful visualization tool that fulfills the intended goal.

Below, I will illustrate an example on how studying this multiomics network, coupled with other bioinformatics analysis and literature search, can efficiently identifying high priority candidates for further study.

miR-126a is upregulated in *Tdrd7-/-* lens. When miR-126a is extracted from the miRNA-mRNA target network to its first neighbors (Figure 3.7A), it points to the Insulin Receptor Substrate 1 (*Irs1*) gene to be reduced in *Tdrd7-/-* lens. It leads to the speculation that *Irs1* may be negatively regulated by miR-126a. Additionally, the network indicates that *Irs1* is enriched in the lens tissue and is also regulated by other miRNA. Further, it appears that Irs1 protein is also reduced in *Tdrd7-/-* lens (Figure 3.11A). Together, these finding suggests that Irs1 is an important candidate for future follow up studies due to the consistency in its misexpression trend within the Tdrd7-/- lens transcriptome and proteome, and due to its correlation to its miRNA regulator. Furthermore, based on iSyTE microarray expression data for 9 developmental stages (Kakrana et al., 2018), the normal expression of Tdrd7 strongly correlates with the normal expression of *Irs1* in the lens (Pearson correlation = 0.9). Interestingly, literature search indicated that *Irs1* is expressed in the lens of *Xenopous*. The same study showed reduced lens size upon *Irs1* knockdown in *Xenopous* (Bugner et al., 2011). In mouse models, *Irs1* is associated to diabetic-linked eye defects (Lavin et al., 2016; Reddy et al., 2013). All the above suggests that Tdrd7 is upstream of miR-126a which itself potentially regulates *Irs1* and may be important in the lens.

It is important to point that this network is particularly powerful compared to other network construction approaches due to it being solely built on actual profiling data sets in the context of the *Tdrd7-/-* lens of the same age (P15).

Additionally, it is also important to note that the trend of miRNAs expression did not always correlate with their targets. This suggests that those targets could be also controlled by other regulators (Zhang et al., 2019). Thus, the relationship between miRNAs and their predicted targets has to be carefully examined and tested.

Finally, it is important to discuss miRNA synergism. Synergistic mRNA regulation is a way to maximize the efficiency of miRNAs in regulating their targets. Synergy is the observation of miRNA-miRNA "collaboration" for fine-tuning a target output (especially in the same direction of regulation). Synergism helps overcoming the consequences from the dependency (or burden) of a single miRNA to "physically" regulate hundreds of targets and is hence energetically less costly for the cell (Arvey et al., 2010; Liu et al., 2016; Zhou and Yang, 2012). miRNA synergism was identified within both Tdrd7 miRNA target regulatory networks (Figure 3.5, Figure 3.6 and Figure 3.9). This suggest that cells in the lens also uses miRNA synergy as a strategy for economical energy expenditure. This might be of special importance to the lens as differentiated fiber cells face the challenge of limited resources as they produce high levels of crystallin proteins while on a pathway toward organelle degradation. It also suggests that Tdrd7 is upstream of sophisticated regulatory networks consisting of multiple layers of information, and serves to explain why its deficiency disrupts collectively many lens key factors resulting in cataract.

Chapter 5

FUTURE DIRECTIONS

As a follow up of this work, I suggest performing small RNA-seq analysis to identify misregulated small RNA in *Tdrd7-/-* lens from other small RNA populations such as Piwi-interacting RNA (piRNA) and transfer RNA (tRNA). This will expand our knowledge on other as yet unaddressed aspects of post-transcriptional control in the context of *Tdrd7* deficiency and lens biology.

iSyTE database provides lens enrichment information from lens high-throughput profiling total RNA data sets. Total lens RNA is subtracted from a whole-body control without the lens to obtain a list of lens enriched genes. It is hypothesized that genes which are highly expressed in the lens tissue when compared to other tissues in the body are important for lens biology. As an expansion for iSyTE, I suggest performing a small RNA-seq on whole body control without the lens. This high-throughput data will allow the identification of small RNA populations that are enriched in the lens. Several small RNA data sets generated form wild type lenses at different developmental stages, including P15 wild type that was generated from this thesis work, can be then subject to in silico subtraction by comparative analysis with these WB miRNA samples. Further, the datasets generated from this analysis can be added as new small RNA tracks on iSyTE website.

I propose to create public access web tools for the Tdrd7 multiomics miRNA target network that was generated in this thesis. This can be done on the existing iSyTE website that hosts a track for an interactive network visualization.

REFERENCES

Anantharaman, V., Zhang, D., and Aravind, L. (2010). OST-HTH: a novel predicted RNA-binding domain. Biol. Direct *5*, 13.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Anderson, P., and Kedersha, N. (2009). RNA granules: post-transcriptional and epigenetic modulators of gene expression. Nature Reviews Molecular Cell Biology *10*, 430–436.

Arvey, A., Larsson, E., Sander, C., Leslie, C.S., and Marks, D.S. (2010). Target mRNA abundance dilutes microRNA and siRNA activity. Mol. Syst. Biol. *6*, 363.

Bardram, M., and Brændsdrup, P. (1947). MATERNAL RUBELLA DURING PREGNANCY AS CAUSE OF CONGENITAL CATARACT AND OTHER CONGENITAL MALFORMATIONS. Acta Ophthalmologica *25*, 353–367.

Barnum, C.E. (2015). Phenotypic and molecular characterization of cataract in the Tdrd7 knockout mouse mutant. M.S. University of Delaware.

Barnum, C.E., Al Saai, S., Patel, S.D., Cheng, C., Anand, D., Xu, X., Dash, S., Siddam, A.D., Glazewski, L., Paglione, E., et al. (2020). The Tudor-domain protein TDRD7, mutated in congenital cataract, controls the heat shock protein HSPB1 (HSP27) and lens fiber cell morphology. Hum. Mol. Genet.

Bassnett, S. (2009). On the mechanism of organelle degradation in the vertebrate lens. Experimental Eye Research *88*, 133–139.

Bassnett, S., Shi, Y., and Vrensen, G.F.J.M. (2011). Biological glass: structural determinants of eye lens transparency. Philosophical Transactions of the Royal Society of London B: Biological Sciences *366*, 1250–1264.

Bazan, H.E. (1989). The synthesis and effects of eicosanoids in avascular ocular tissues. Prog. Clin. Biol. Res. *312*, 73–84.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. (2003). Dicer is essential for mouse development. Nat. Genet. *35*, 215–217.

Bhat, S.P. (2001). The Ocular Lens Epithelium. Bioscience Reports *21*, 537–563.

Bitel, C.L., Singh, V., and Frederikse, P.H. (2012). miR-124, miR-125b, let-7 and vesicle transport proteins in squid lenses in L. pealei. Curr. Eye Res. *37*, 388–394.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bugner, V., Aurhammer, T., and Kühl, M. (2011). Xenopus laevis insulin receptor substrate IRS-1 is important for eye development. Developmental Dynamics *240*, 1705–1715.

Chen, Y., and Wang, X. (2020). miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res *48*, D127–D131.
Chen, C., Nott, T.J., Jin, J., and Pawson, T. (2011). Deciphering arginine methylation: Tudor tells the tale. Nat. Rev. Mol. Cell Biol. *12*, 629–642.
Chen, J., Wang, Q., Cabrera, P.E., Zhong, Z., Sun, W., Jiao, X., Chen, Y., Govindarajan, G., Naeem, M.A., Khan, S.N., et al. (2017). Molecular Genetic Analysis of Pakistani Families With Autosomal Recessive Congenital Cataracts by Homozygosity Screening. Invest. Ophthalmol. Vis. Sci. *58*, 2207–2217.
Côté, J., and Richard, S. (2005). Tudor Domains Bind Symmetrical Dimethylated Arginines. J. Biol. Chem. *280*, 28476–28483.
Cui, G., Botuyan, M.V., and Mer, G. (2013). (1)H, (15)N and (13)C resonance assignments for the three LOTUS RNA binding domains of Tudor domain-containing protein TDRD7. Biomol NMR Assign *7*, 79–83.
Dash, S., Siddam, A.D., Barnum, C.E., Janga, S.C., and Lachke, S.A. (2016). RNA-binding proteins in eye development and disease: implication of conserved RNA granule components. Wiley Interdiscip Rev RNA *7*, 527–557.
Dong, N., Xu, B., Benya, S.R., and Tang, X. (2014). MiRNA-26b inhibits the proliferation, migration, and epithelial–mesenchymal transition of lens epithelial cells. Mol Cell Biochem *396*, 229–238.
Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature Reviews Genetics *9*, 102–114.
Graw, J. (2009). Genetics of crystallins: cataract and beyond. Exp. Eye Res. *88*, 173–189.
Houbaviy, H.B., Murray, M.F., and Sharp, P.A. (2003). Embryonic stem cell-specific MicroRNAs. Dev. Cell *5*, 351–358.
Jeske, M., Bordi, M., Glatt, S., Müller, S., Rybin, V., Müller, C.W., and Ephrussi, A. (2015). The Crystal Structure of the Drosophila Germline Inducer Oskar Identifies Two Domains with Distinct Vasa Helicase- and RNA-Binding Activities. Cell Rep *12*, 587–598.
Jeske, M., Müller, C.W., and Ephrussi, A. (2017). The LOTUS domain is a conserved DEAD-box RNA helicase regulator essential for the recruitment of Vasa to the germ plasm and nuage. Genes Dev. *31*, 939–952.
Jiang, H., Xu, L., Wang, Z., Keene, J., and Gu, Z. (2014). Coordinating expression of RNA binding proteins with their mRNA targets. Sci Rep *4*, 7175.
Kakrana, A., Yang, A., Anand, D., Djordjevic, D., Ramachandruni, D., Singh, A., Huang, H., Ho, J.W.K., and Lachke, S.A. (2018). iSyTE 2.0: a database for expression-based gene discovery in the eye. Nucleic Acids Res. *46*, D875–D885.
Khan, S.Y., Hackett, S.F., Lee, M.-C.W., Pourmand, N., Talbot, C.C., and Riazuddin, S.A. (2015). Transcriptome Profiling of Developing Murine Lens Through RNA Sequencing. Invest. Ophthalmol. Vis. Sci. *56*, 4919–4926.
Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature Methods *12*, 357–360.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology *37*, 907–915.
Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. *10*, 126–139.
Kirino, Y., and Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2′-O-methylated at their 3′ termini. Nature Structural & Molecular Biology *14*, 347–348.
Kishore, S., Luber, S., and Zavolan, M. (2010). Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. Brief Funct Genomics *9*, 391–404.
Kobayashi, H., and Tomari, Y. (2016). RISC assembly: Coordination between small RNAs and Argonaute proteins. Biochim. Biophys. Acta *1859*, 71–81.
Kotaja, N., and Sassone-Corsi, P. (2007). The chromatoid body: a germ-cell-specific RNA-processing centre. Nat. Rev. Mol. Cell Biol. *8*, 85–90.
Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, D68-73.
Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47*, D155–D162.
Kubo, E., Hasanova, N., Sasaki, H., and Singh, D.P. (2013). Dynamic and differential regulation in the microRNA expression in the developing and mature cataractous rat lens. J. Cell. Mol. Med. *17*, 1146–1159.
Kuhli-Hattenbach, C., Lüchtenberg, M., Kohnen, T., and Hattenbach, L.-O. (2008). Risk factors for complications after congenital cataract surgery without intraocular lens implantation in the first 18 months of life. Am. J. Ophthalmol. *146*, 1–7.
Lachke, S.A., Alkuraya, F.S., Kneeland, S.C., Ohn, T., Aboukhalil, A., Howell, G.R., Saadi, I., Cavallesco, R., Yue, Y., Tsai, A.C.-H., et al. (2011a). Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. Science *331*, 1571–1576.
Lachke, S.A., Alkuraya, F.S., Kneeland, S.C., Ohn, T., Aboukhalil, A., Howell, G.R., Saadi, I., Cavallesco, R., Yue, Y., Tsai, A.C.-H., et al. (2011b). Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. Science *331*, 1571–1576.
Lachke, S.A., Ho, J.W.K., Kryukov, G.V., O'Connell, D.J., Aboukhalil, A., Bulyk, M.L., Park, P.J., and Maas, R.L. (2012). iSyTE: integrated Systems Tool for Eye gene discovery. Invest. Ophthalmol. Vis. Sci. *53*, 1617–1627.
Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics *CHAPTER*, Unit-11.7.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology *10*, R25.
Lavin, D.P., White, M.F., and Brazil, D.P. (2016). IRS proteins and diabetic complications. Diabetologia *59*, 2280–2291.
Lee, Y., Jeon, K., Lee, J.-T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. *21*, 4663–4670.

Li, Y., and Piatigorsky, J. (2009). Targeted deletion of Dicer disrupts lens morphogenesis, corneal epithelium stratification, and whole eye development. Developmental Dynamics *238*, 2388–2400.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
Li, Y., Liu, S., Zhang, F., Jiang, P., Wu, X., and Liang, Y. (2015). Expression of the microRNAs hsa-miR-15a and hsa-miR-16-1 in lens epithelial cells of patients with age-related cataract. Int J Clin Exp Med *8*, 2405–2410.
Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. Nat. Rev. Mol. Cell Biol. *12*, 505–516.
Liu, W., and Wang, X. (2019). Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. Genome Biol. *20*, 18.
Liu, Z., Guo, Y., Pu, X., and Li, M. (2016). Dissecting the regulation rules of cancer-related miRNAs based on network analysis. Sci Rep *6*, 34172.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550.
Lu, T.X., and Rothenberg, M.E. (2018). MicroRNA. J Allergy Clin Immunol *141*, 1202–1207.
Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95–98.
Marcello, M.R., and Singson, A. (2011). Germline Determination: Don't Mind the P Granules. Curr Biol *21*, R155–R157.
Mets, M.B. (2001). Eye manifestations of intrauterine infections. Ophthalmol Clin North Am *14*, 521–531.
Morita, S., Horii, T., Kimura, M., Goto, Y., Ochiya, T., and Hatada, I. (2007). One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. Genomics *89*, 687–696.
Pasquinelli, A.E., and Ruvkun, G. (2002). Control of developmental timing by micrornas and their targets. Annu. Rev. Cell Dev. Biol. *18*, 495–513.
Reddy, V.S., Raghu, G., Reddy, S.S., Pasupulati, A.K., Suryanarayana, P., and Reddy, G.B. (2013). Response of small heat shock proteins in diabetic rat retina. Invest. Ophthalmol. Vis. Sci. *54*, 7674–7682.
Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.
Rolle, K., Piwecka, M., Belter, A., Wawrzyniak, D., Jeleniewicz, J., Barciszewska, M.Z., and Barciszewski, J. (2016). The Sequence and Structure Determine the Function of Mature Human miRNAs. PLOS ONE *11*, e0151246.
Ronchetti, D., Todoerti, K., Tuana, G., Agnelli, L., Mosca, L., Lionetti, M., Fabris, S., Colapietro, P., Miozzo, M., Ferrarini, M., et al. (2012). The expression pattern of

small nucleolar and small Cajal body-specific RNAs characterizes distinct molecular subtypes of multiple myeloma. Blood Cancer Journal *2*, e96–e96.
Sempere, L.F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E., and Ambros, V. (2004). Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol. *5*, R13.
Shiels, A., and Hejtmancik, J. (2013). Genetics of human cataract. Clin Genet *84*, 120–127.
Shiels, A., and Hejtmancik, J.F. (2015). Chapter Twelve - Molecular Genetics of Cataract. In Progress in Molecular Biology and Translational Science, J.F. Hejtmancik, and J.M. Nickerson, eds. (Academic Press), pp. 203–218.
Stefani, G., and Slack, F.J. (2008). Small non-coding RNAs in animal development. Nat. Rev. Mol. Cell Biol. *9*, 219–230.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. *102*, 15545–15550.
Tan, Y.-Q., Tu, C., Meng, L., Yuan, S., Sjaarda, C., Luo, A., Du, J., Li, W., Gong, F., Zhong, C., et al. (2017). Loss-of-function mutations in TDRD7 lead to a rare novel syndrome combining congenital cataract and nonobstructive azoospermia in humans. Genet. Med.
Tanaka, T., Hosokawa, M., Vagin, V.V., Reuter, M., Hayashi, E., Mochizuki, A.L., Kitamura, K., Yamanaka, H., Kondoh, G., Okawa, K., et al. (2011a). Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis. Proc. Natl. Acad. Sci. U.S.A. *108*, 10579–10584.
Tanaka, T., Hosokawa, M., Vagin, V.V., Reuter, M., Hayashi, E., Mochizuki, A.L., Kitamura, K., Yamanaka, H., Kondoh, G., Okawa, K., et al. (2011b). Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis. Proc. Natl. Acad. Sci. U.S.A. *108*, 10579–10584.
Tanaka, Y., Kurihara, T., Hagiwara, Y., Ikeda, S., Mori, K., Jiang, X., Torii, H., and Tsubota, K. (2019). Ocular-Component-Specific miRNA Expression in a Murine Model of Lens-Induced Myopia. International Journal of Molecular Sciences *20*, 3629.
Torres, A.G., Batlle, E., and Ribas de Pouplana, L. (2014). Role of tRNA modifications in human diseases. Trends in Molecular Medicine *20*, 306–314.
Wang, X. (2008). miRDB: A microRNA target prediction and functional annotation database with a wiki interface. RNA *14*, 1012–1017.
Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat. Genet. *39*, 380–385.
Weston, M.D., Tarang, S., Pierce, M.L., Pyakurel, U., Rocha-Sanchez, S.M., McGee, J., Walsh, E.J., and Soukup, G.A. (2018). A mouse model of miR-96, miR-182 and

miR-183 misexpression implicates miRNAs in cochlear cell fate and homeostasis. Sci Rep *8*.
Wolf, L., Gao, C.S., Gueta, K., Xie, Q., Chevallier, T., Podduturi, N.R., Sun, J., Conte, I., Zelenka, P.S., Ashery-Padan, R., et al. (2013). Identification and Characterization of FGF2-Dependent mRNA: microRNA Networks During Lens Fiber Cell Differentiation. G3 (Bethesda) *3*, 2239–2255.
Wu, C.-R., Ye, M., Qin, L., Yin, Y., and Pei, C. (2017). Expression of lens-related microRNAs in transparent infant lenses and congenital cataract. Int J Ophthalmol *10*, 361–365.
Yorston, D. (2004). Surgery for Congenital Cataract. Community Eye Health *17*, 23–25.
Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol. BioSyst. *12*, 477–479.
Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology *16*, 284–287.
Yu, G., Wang, L.-G., Yan, G.-R., and He, Q.-Y. (2015). DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics *31*, 608–609.
Zhang, J., Pham, V.V.H., Liu, L., Xu, T., Truong, B., Li, J., Rao, N., and Le, T.D. (2019). Identifying miRNA synergism using multiple-intervention causal inference. BMC Bioinformatics *20*, 613.
Zheng, C., Wu, M., He, C.-Y., An, X.-J., Sun, M., Chen, C.-L., and Ye, J. (2014). RNA granule component TDRD7 gene polymorphisms in a Han Chinese population with age-related cataract. J. Int. Med. Res. *42*, 153–163.
Zhou, X., and Yang, P.-C. (2012). MicroRNA: a small molecule with a big biological impact. Microrna *1*, 1.

Appendix A

DESEQ DIFFRENTIAL EXPRESSION ANALYSIS RESULTS

Table-A. 1 **Differentially expressed miRNA identified from DESeq**

miRNA Name	log2FoldChange	fold change STE*	padj
mmu-miR-124-3p	-5.09	1.06	1.07E-04
mmu-miR-96-5p	-4.36	0.59	4.61E-11
mmu-miR-9-5p	-4.27	0.55	9.65E-12
mmu-miR-182-5p	-4.27	0.82	2.27E-05
mmu-miR-9b-3p	-4.21	0.60	5.32E-10
mmu-miR-183-5p	-4.20	0.92	3.19E-04
mmu-miR-129-5p	-3.97	0.82	1.07E-04
mmu-miR-9-3p	-3.26	0.62	2.21E-05
mmu-miR-190b-5p	-2.81	0.69	2.39E-03
mmu-miR-7b-5p	-2.71	0.54	3.69E-05
mmu-miR-6540-5p	-2.59	0.79	3.10E-02
mmu-miR-217-5p	-2.27	0.69	3.10E-02
mmu-miR-488-3p	-2.21	0.44	3.69E-05
mmu-miR-690	-1.80	0.44	2.56E-03
mmu-miR-369-3p	-1.57	0.31	3.69E-05
mmu-miR-412-5p	-1.30	0.40	3.19E-02
mmu-miR-433-3p	-1.10	0.32	2.33E-02
mmu-miR-770-3p	-1.09	0.19	2.63E-06
mmu-miR-666-5p	-1.09	0.24	3.89E-04
mmu-miR-543-3p	-0.69	0.17	2.64E-03
mmu-miR-434-5p	-0.64	0.15	1.55E-03
mmu-miR-351-5p	0.87	0.28	4.43E-02
mmu-miR-322-3p	1.00	0.28	1.55E-02
mmu-miR-126b-5p	1.54	0.39	3.41E-03
mmu-miR-3081-3p	1.70	0.52	3.19E-02

*STE: standard error estimate for the log2 fold change

Appendix B

SEQUENCE OF DIFFERENTIALLY EXPRESSED MIRNA SEQUENCE IN TDRD7-/- AT P15 IN THE LENS

Table-B. 1 **Differentially expressed mature miRNA sequence**

Mature miRNA sequence obtained from miRbase data base. The table is merged with edgeR results

miRNA Name	Log2 fold change	Absolute fold change	FDR	mature MiRNA sequence	seq length
mmu-miR-124-3p	-5	-33.5	3.90E-05	UAAGGCACGCGGUGAAUGCC	20
mmu-miR-540-3p	-0.9	-1.7	3.10E-02	AGGUCAGAGGUCGAUCCUGG	20
mmu-miR-9b-3p	-4.2	-19.2	1.50E-07	AUACAGCUAGAUAACCAAAGA	21
mmu-miR-129-5p	-3.9	-16.3	5.00E-05	CUUUUUGCGGUCUGGGCUUGC	21
mmu-miR-488-3p	-2.2	-4.4	2.40E-05	UUGAAAGGCUGUUUCUUGGUC	21
mmu-miR-369-3p	-1.6	-2.9	1.60E-06	AAUAAUACAUGGUUGAUCUUU	21
mmu-miR-323-3p	-1.4	-2.4	2.70E-02	CACAUUACACGGUCGACCUCU	21
mmu-miR-322-3p	1	2.1	1.00E-02	AAACAUGAAGCGCUGCAACAC	21
mmu-miR-183-5p	-4.2	-19.1	4.00E-06	UAUGGCACUGGUAGAAUUCACU	22
mmu-miR-9-3p	-3.3	-9.6	3.40E-05	AUAAAGCUAGAUAACCGAAAGU	22
mmu-miR-216b-5p	-2.8	-7.3	3.40E-02	AAAUCUCUGCAGGCAAAUGUGA	22
mmu-miR-190b-5p	-2.8	-6.8	8.60E-04	UGAUAUGUUUGAUAUUGGGUUG	22
mmu-miR-6540-5p	-2.6	-5.8	1.70E-02	CUAAGGCAGGCAGACUUCAGUG	22
mmu-miR-690	-1.8	-3.3	1.70E-03	AAAGGCUAGGCUCACAACCAAA	22
mmu-miR-666-5p	-1.1	-2.1	4.80E-04	AGCGGGCACAGCUGUGAGAGCC	22
mmu-miR-770-3p	-1.1	-2.1	3.80E-05	CGUGGGCCUGACGUGGAGCUGG	22
mmu-miR-433-3p	-1.1	-2	9.50E-03	AUCAUGAUGGGCUCCUCGGUGU	22
mmu-miR-543-3p	-0.7	-1.6	1.80E-02	AAACAUUCGCGGUGCACUUCUU	22
mmu-miR-434-5p	-0.6	-1.5	2.80E-02	GCUCGACUCAUGGUUUGAACCA	22
mmu-miR-126a-3p	1	2.1	4.00E-02	UCGUACCGUGAGUAAUAAUGCG	22
mmu-miR-126b-5p	1.5	3.1	9.60E-04	AUUAUUACUCACGGUACGAGUU	22
mmu-miR-96-5p	-4.4	-20.6	8.40E-08	UUUGGCACUAGCACAUUUUUGCU	23
mmu-miR-9-5p	-4.3	-19.8	5.80E-09	UCUUUGGUUAUCUAGCUGUAUGA	23
mmu-miR-217-5p	-2.3	-4.6	2.50E-02	UACUGCAUCAGGAACUGACUGGA	23

mmu-miR-135b-5p	-1.8	-3.3	1.70E-02	UAUGGCUUUUCAUUCCUAUGUGA	23
mmu-miR-412-5p	-1.3	-2.4	6.50E-03	UGGUCGACCAGCUGGAAAGUAAU	23
mmu-miR-3081-3p	1.7	3.6	2.00E-02	UUGCGCUCCGAUCUCUGAGCUGG	23
mmu-miR-7b-5p	-2.7	-6.5	3.90E-05	UGGAAGACUUGUGAUUUUGUUGUU	24
mmu-miR-351-5p	0.9	2	3.60E-02	UCCCUGAGGAGCCCUUUGAGCCUG	24
mmu-miR-182-5p	-4.3	-19.5	3.00E-07	UUUGGCAAUGGUAGAACUCACACCG	25

Appendix C

REPRINT PERMISSION

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Aug 08, 2020

This Agreement between S Al Saai ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4884480132025
License date	Aug 08, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	WILEY INTERDISCIPLINARY REVIEWS: RNA
Licensed Content Title	RNA-binding proteins in eye development and disease: implication of conserved RNA granule components
Licensed Content Author	Salil A. Lachke, Sarath Chandra Janga, Carrie E. Barnum, et al
Licensed Content Date	May 1, 2016
Licensed Content Volume	7
Licensed Content Issue	4
Licensed	31

Appendix D

ANIMAL APPROVAL

RECEIVED
AUG 2 0 2019
IACUC

University of Delaware
Institutional Animal Care and Use Committee
Annual Review

Title of Protocol: Investigate the function of genes associated with animal development using mouse	
AUP Number: 1226-2019-2	**← (4 digits only)**
Principal Investigator: Salil A. Lachke	
Common Name: Mouse **Genus Species:** *Mus musculus*	

Pain Category: ***(please mark one)***

USDA PAIN CATEGORY: *(Note change of categories from previous form)*	
Category	**Description**
☐ B	Breeding or holding where NO research is conducted
☒ C	Procedure involving momentary or no pain or distress
☐ D	Procedure where pain or distress is alleviated by appropriate means (analgesics, tranquilizers, euthanasia etc.)
☐ E	Procedure where pain or distress cannot be alleviated, as this would adversely affect the procedures, results or interpretation

Official Use Only

IACUC Approval Signature: Glen Tabin, DVM

Date of Approval: 10.1.19