MS/MS *in silico* subtraction-based proteomic profiling as an approach to facilitate disease gene discovery: application to lens development and cataract

Sandeep Aryal[1], **Deepti Anand**[1], **Francisco G. Hernandez**[1], **Bailey A. T. Weatherbee**[1], **Hongzhan Huang**[2], **Ashok P. Reddy**[3], **Phillip A. Wilmarth**[3], **Larry L. David**[3,4], **Salil A. Lachke**[1,2,*]

[1]Department of Biological Sciences, University of Delaware, Newark, DE 19716 USA

[2]Center for Bioinformatics & Computational Biology, University of Delaware, Newark, DE 19716 USA

[3]Proteomics Shared Resource, Oregon Health & Science University, Portland, OR 97239 USA

[4]Department of Chemical Physiology & Biochemistry, Oregon Health & Science University, Portland, OR 97239 USA

Abstract

While the bioinformatics resource-tool iSyTE (integrated Systems Tool for Eye gene discovery) effectively identifies human cataract-associated genes, it is currently based on just transcriptome data, and thus it is necessary to include protein-level information to gain greater confidence in gene prioritization. Here we expand iSyTE through development of a novel proteome-based resource on the lens and demonstrate its utility in cataract gene discovery. We applied high-throughput tandem mass spectrometry (MS/MS) to generate a global protein expression profile of mouse lens at embryonic day (E)14.5, which identified 2371 lens-expressed proteins. A major challenge of high-throughput expression profiling is identification of high-priority candidates among the thousands of expressed proteins. To address this problem, we generated new MS/MS proteome data on mouse whole embryonic body (WB). WB proteome was then used as a reference dataset for performing "*in silico* WB-subtraction" comparative analysis with the lens proteome, which effectively identified 422 proteins with lens-enriched expression at ≥2.5 average spectral counts, ≥2.0 fold-enrichment (FDR <0.01) cut-off. These top 20% candidates represent a rich pool of high-priority proteins in the lens including known human cataract-linked genes and many new potential regulators of lens development and homeostasis. This rich information is made publicly accessible through iSyTE (https://research.bioinformatics.udel.edu/iSyTE/), which enables user-

*Corresponding author: Salil A. Lachke, Department of Biological Sciences, University of Delaware, 105 The Green, Delaware Avenue, 236 Wolf Hall, Newark, DE, USA, Tel.: (302) 831 3040; Fax: (302) 831 2281; salil@udel.edu.

friendly visualization of promising candidates, thus making iSyTE a comprehensive tool for cataract gene discovery.

Keywords

Lens; iSyTE; Proteome; Embryonic lens development; Protein profiling; Database

Introduction

To predict high-priority candidate genes linked to cataract and lens development, a user-friendly web resource iSyTE (integrated Systems Tool for Eye gene discovery) was recently developed (Lachke et al. 2012b). The present version of iSyTE is based on high-throughput transcriptome data generated by microarrays or RNA-sequencing (RNA-seq) of the lens at different developmental and post-natal stages (Kakrana et al. 2018; Anand et al. 2018). To prioritize lens candidates from these vast transcriptomic data, iSyTE uses a strategy termed "*in silico* whole embryonic body (WB) subtraction". This is based on the principle that comparison of a tissue-specific dataset, such as the lens, with that of a general reference dataset such as the WB, effectively "subtracts" genes with similar levels of expression, in turn leading to the identification of genes that exhibit "enriched" expression in the specific tissue of interest (Anand and Lachke 2017). This "lens-enriched expression" strategy has worked well, and iSyTE has effectively identified several new genes linked to lens defects and cataract (Lachke et al. 2011, 2012a; Kasaikina et al. 2011; Agrawal et al. 2015; Dash et al. 2015; Patel et al. 2017; Siddam et al. 2018) and has impacted the understanding of other pathways in lens development and pathology (Wolf et al. 2013; Manthey et al. 2014; Audette et al. 2016; Wang et al. 2017b; Cavalheiro et al. 2017; Krall et al. 2018).

However, while iSyTE gives rich information on the transcript level of gene expression, its current version does not provide any information on expression at the level of proteins – which are the principle effectors of biological processes. This is an important knowledge-gap because the cellular proteome depends on post-transcriptional control of gene expression that can impact alternative splicing, mRNA stability and translational regulation (Dash et al. 2016). Thus, post-transcriptional control can result in scenarios wherein a specific mRNA is present, but its encoded protein is not (*e.g.* because of mRNA silencing) or a specific protein is present, but its parent mRNA is not (*e.g.* because of differences in mRNA and protein stability). Moreover, alternative splicing can produce differential amounts of distinct protein isoforms in a given cell/tissue. Importantly, iSyTE has identified several post-transcriptional regulatory factors such as *Tdrd7*, *Celf1*, *Rbm24* and *Caprin2* that function in the lens (Lachke et al. 2011; Dash et al. 2015; Siddam et al. 2018). Deficiency of these proteins result in cataract and lens defects in human and/or various animal models. Thus, integrating the rich information of the developing lens proteome in iSyTE is significant as it will serve to further increase confidence in iSyTE's cataract-associated gene predictions in cases when both transcript and protein levels correlate, and importantly, even when the transcript and protein levels do not necessarily correlate. In these cases, it will potentially lead to the identification of new cataract-linked genes that are missed by transcriptomics. While integrating proteome data in iSyTE is essential high-throughput

proteomics poses similar and important challenges to high-throughput transcriptomics, such as parsing through the large amounts of data to prioritize select candidates. Thus, although there are several previous studies on lens protein profiling, these all face the common challenge of identifying high-priority candidates in the lens among the many expressed proteins (Hoehenwarter et al. 2006; Bassnett et al. 2009; Wilmarth et al. 2009; Wang et al. 2013; Khan et al. 2018a, b; Zhao et al. 2019).

To address these challenges, in this work we generated new proteome data in the embryonic lens as well as new proteome data on whole embryonic body tissue that allowed us to perform *in silico* WB-subtraction for the first time on protein datasets, leading to the identification of high-priority lens proteins and new candidates for cataract. We performed high-throughput tandem mass spectrometry (MS/MS) to generate a global protein expression profile of mouse lens and WB at embryonic day (E)14.5. Stage E14.5 was selected for this analysis because it is particularly informative as: (1) lens morphogenesis is completed from the perspective of formation of lens primary fiber cells, (2) the immature anterior lens epithelium is established, and (3) secondary fiber cell differentiation is initiated. Furthermore, at this stage, degradation of subcellular organelles in fiber cells is yet to occur and the lens proteome is poised to initiate the challenging process of fiber cell maturation while committing to highly active synthesis of lens proteins. Indeed, a proteome level analysis of this important stage in lens development has not been described, thus representing a critical knowledge-gap. This approach identified 2118 comparable proteins (out of 2371 identified total proteins) to be expressed in the lens and WB at established cut-off criteria. *In silico* WB-subtraction identified 422 lens-enriched proteins including those previously linked to cataract. We find that while lens protein expression alone (*i.e.* lens proteome not subjected to *in silico* WB-subtraction) could identify several cataract-linked genes, *in silico* WB-subtraction was more effective for prioritization of key cataract-linked candidates, especially those that were not as abundant as crystallins. Moreover, *in silico* WB-subtraction identified many new potential regulators/factors in the lens that were not prioritized by lens expression alone. To make this rich proteome information readily available to the research community, we developed new custom annotation-tracks on the University of California Santa Cruz (UCSC) Genome Browser, a public resource, and made these tracks accessible via iSyTE (https://research.bioinformatics.udel.edu/iSyTE/). Together, these data make iSyTE a comprehensive tool for lens expression analysis and cataract gene discovery.

Materials and Methods

Mouse studies

Wild-type C57BL6/J mice (The Jackson Laboratory) were bred and maintained at the University of Delaware Center for Animal research as per the animal protocol (#1226) that was approved by the Institutional Animal Care and Use Committee (IACUC). Animal experiments were performed following the guidelines in the Association of Research in Vision and Ophthalmology (ARVO) statement for the use of animals in ophthalmic and vision research. Animals were housed in a 14 h light to 10 h dark cycle.

Tissue preparation

For embryonic tissue collection, the day of the detection of the vaginal plug was designated as embryonic day (E) 0.5. Lens tissue from E14.5 mouse embryos (five biological replicates from the same litter; each replicate consists of two lenses from the same embryo) was micro-dissected ensuring that the tunica vasculosa lentis was removed and stored in −80°C until further processing. Mouse E14.5 whole embryonic body (WB) tissue (eye removed) (five biological replicates from the same litter) was isolated and ground in liquid nitrogen with a mortar and pestle. Mouse E14.5 WB samples were transferred to a 2 ml lobind centrifuge tube, suspended in 1.2 ml of 4% sodium dodecyl sulfate (SDS), 0.2% deoxycholic acid (DCA), 100 mM triethyl ammonium bicarbonate (TEAB) (pH 8.0), and heated at 90°C for 30 min. Mouse E14.5 lens samples were suspended in 120 μl of 167 mM triethyl ammonium bicarbonate (TEAB) buffer and probe-sonicated using a Fisher Scientific 60 Sonic Dismembrator. Samples were adjusted to 4% SDS, 0.2% DCA, 100 mM TEAB by addition of 40 μl of 20% SDS, 1% DCA and 40 μl of 4% SDS, 0.2% DCA, 100 mM TEAB to a total volume of 200 μl. Lysed samples were centrifuged for 2 min at 16000 x g at room temp and heated at 90°C for 15 min. Mouse E14.5 WB and lens samples were centrifuged, and protein content was quantified by BCA protein assay kit (ThermoFischer Cat. No. 23225). For both WB and lens, 55 μg of protein/sample (*n*=5 biological replicates) was trypsinized using a modified enhanced filter aided e-FASP digestion protocol using Amicon 30 kDa ultracentrifugation devices (Erde et al. 2017). Briefly, samples were reduced with TCEP by heating at 90°C for 10 min, transferred to the Amicon filter, and buffer exchanged into 8 M Urea, 0.2% deoxycholic acid (DCA), 100 mM TEAB. Samples were then alkylated with iodoacetamide, exchanged into 0.2% DCA, 50 mM TEAB (pH 8.0) digestion buffer and trypsin (1:20 enzyme: substrate) was added for an overnight digestion. The following day, samples were centrifuged and the filtrate containing the peptides extracted with ethyl acetate to remove DCA. Samples were then dried in a SpeedVac vacuum concentrator (Thermo Fisher Scientific), resuspended in 100 μl of HPLC water and a peptide assay done using Pierce Quantitative Colorimetric Peptide Assay Kit. Average peptide recovery from mouse E14.5 WB samples was ~80 μg/sample and from mouse E14.5 lens samples was ~45 μg/sample.

Mass spectrometry

Sample digests (4 μg in 5% Formic acid) were loaded onto an Acclaim PepMap 0.1×20 mm NanoViper C18 peptide trap (Thermo Fisher Scientific) for 5 min at a flow rate of 10 μl/min in a 2% acetonitrile, 0.1% formic acid mobile phase. Peptides were separated using a PepMap RSLC C18, 2 μm particle, 75 μm x 50 cm EasySpray column (Thermo Fisher Scientific) using a 7.5–30% acetonitrile gradient over 205 min in mobile phase containing 0.1% formic acid and a 300 nl/min flow rate using a Dionex NCS-3500RS UltiMate RSLC nano UPLC system. Tandem mass spectrometry (MS) data was collected using a Thermo Orbitrap Fusion mass spectrometer configured with an EasySpray NanoSource (Thermo Fisher Scientific). The instrument was configured for data dependent analysis (DDA) using the MS/DD-MS/MS setup. Full MS resolutions were set to 120,000 at m/z 200, mass range 375–1500, charge state 2–7, full MS AGC target was 400,000, intensity threshold was 5,000, max inject time at 50 ms, and 10 ppm dynamic exclusion for 60 s. AGC target value for fragment spectra was set at 5,000. Isolation mode was quadrupole, isolation width was set at

1.6 m/z, isolation offset was set to off, activation type was CID, collision energy was set to fixed at 35%, maximum injection time set at 300 ms and detector type was IonTrap. All data was acquired in centroid mode using positive polarity.

RAW file conversions

The RAW files were converted to MS2 format files using MSConvert from the open source Proteowizard toolkit for five mouse E14.5 lens samples and five mouse E14.5 WB samples (Chambers et al. 2012). The lens samples had ~50K MS2 scans per run while WB samples had ~88K MS2 scans per run. The peptide assay post-digestion suggested that there were higher numbers of peptides in WB after digestion compared to the lens. There were data from 682,315 scans written to MS2 format files.

Database searching

A canonical mouse reference proteome (version 2019.04; 22,287 sequences) from UniProt was downloaded using software available at https://github.com/pwilmart/fasta_utilities.git. Common contaminants were added (179 sequences) and a concatenated sequence-reversed decoy database was added for a total of 44,932 entries. The open source search engine Comet was used to assign peptide sequences to the MS2 spectra (PSMs) (Eng et al. 2013). Comet was configured for: tryptic enzymatic cleavage (a maximum of two missed cleavages); monoisotopic parent ion mass tolerance of 1.25 Da; monoisotopic fragment ion tolerance of 1.0005 Da; fragment bin offset of 0.4; b-, y-, and neutral loss ions were used in scoring (flanking peaks were not used); variable modification of oxidation (+15.9949 Da) on methionine was specified; static modification of alkylation (+57.0215 Da) of cysteines was specified.

PSM error control

The highest scoring matches (top hits) for each PSM from Comet were post processed for false discovery rate (FDR) error control using the PAW pipeline (https://github.com/pwilmart/PAW_pipeline.git) and the target/decoy method (Elias and Gygi 2007; Wilmarth et al. 2009). Accurate delta mass conditional score histograms were created for peptides of different charge states (2+, 3+, and 4+ were considered) and modification state (unmodified or oxidized). Target and decoy score histograms were used to estimate the FDR as a function of a Peptide-Prophet-like discriminant score and to set score thresholds to achieve an overall experiment-wide PSM FDR of 1% (Keller et al. 2002). Peptide matches had to have a minimum length of 7 amino acids. Of the 682K MS2 scans, 514K met the peptide length and charge state requirements. There were 320,640 scans that passed the score cutoffs with 3,319 decoy matches for an FDR of 1.04%. The overall ID rate (of the 514K spectra) was 62%.

Protein Inference

The sequences of the filtered PSMs were used to infer the proteins present in the samples using basic parsimony principles (Nesvizhskii and Aebersold 2005). An extended parsimony algorithm was used to group homologous protein family members together when evidence to

distinguish family members was insufficient (Madhira 2016). In total, 4,645 proteins were detected (4,561 after grouping) with 73 decoy matches, for a protein FDR of about 1.6%.

Quantitative Analysis

Protein assays were used to estimate protein concentration and an equal amount of protein was digested for both WB and lens samples. Post digest peptide assays indicated that the WB samples had higher signals compared to the lens samples. For each sample, equal amounts of the digests were analyzed for the total spectral counts (SpC, a robust semi-quantitative measure). SpC for each sample were also tallied after protein inference and confirmed the peptide assay results, indicating that the lens samples had lower peptide levels. All samples were scaled to the average total spectral count per sample to match the lens and WB samples. There were about 1,800 proteins detected in the lens samples compared to about 3,500 proteins for the WB samples. Because the central question was to identify proteins with enriched expression in the lens compared to WB, the average SpC for all samples was computed from the scaled data for each protein, and further considered in the analysis only if it was greater than 2.5. This cutoff was chosen so that an average SpC of 5 in one condition (*e.g.* lens) and zero in the other condition (*e.g.* WB) could be still be identified. An average SpC of 5 is above the minimal values of 1 or 2 and is expected to be consistently detected and is therefore suggestive of a protein to be present in the sample. Based on this average SpC cutoff of 2.5, there were 2,118 proteins that could be tested for differential expression between the lens and WB samples. A Bioconductor package for differential gene expression, edgeR was used. edgeR has a built-in normalization method called the trimmed mean of M-values (TMM) that corrects for compositional differences between samples and it was appropriate for this experiment (Robinson and Oshlack 2010; Robinson et al. 2010). The exact test in edgeR was used with default Benjamini-Hochberg multiple testing corrections. Analysis was performed in R (version 3.5.3) using a Jupyter notebook. Numerous data visualizations were used to check the analysis steps. Statistical testing results were added back to the proteomics results in a unified results table for subsequent data exploration.

Immunofluorescence

To examine the expression of select proteins in the lens, mouse embryonic head tissue at stage E14.5 was fixed in 4% PFA for 30 minutes on ice and equilibrated in 30% sucrose overnight at 4°C prior to being mounted in OCT (Tissue-Tech, Doral, FL) and stored at −80°C. The frozen head tissue was subjected to sectioning in a cryostat (Leica CM3050) and sections (12 μm thickness) were blocked in blocking solution containing either 5% chicken serum (Abcam, Cambridge, UK; for the antibodies against Eml2, Nol3, Slc7a5) or 1% Bovine Serum Albumin (Sigma-Aldrich, St.Louis, MO) plus 10% Goat Serum (Jackson ImmunoResearch; for the antibody against Igfbp7) in 0.1% Triton X (Promega) and 1X PBS (phosphate buffer) for one hour at room temperature. After blocking for 1 hr, the sections were incubated with the primary antibody overnight at 4°C. The following primary antibodies were purchased from Abcam and Proteintech and used in the given dilutions in the blocking buffers: Eml2 (13529-1-AP, 1:25 diln.), Igfbp7(13529-1-AP, 1:25 diln.), Nol3(13529-1-AP, 1:25 diln.) and Slc7a5(13752-1-AP, 1:25 diln.).

After overnight incubation at 4°C, slides were washed and incubated with the appropriate secondary antibody conjugated to Alexa Fluor 488 (1:200) (Life Technologies, Carlsbad, CA) and the nuclear stain DAPI (1:1000) (Life Technologies) for 2 hr at room temperature. Slides were washed, mounted using mounting media and imaged using the Zeiss LSM 880 Confocal microscope configured with Diode/Argon laser (405 nm and 488 nm excitation lines) (Carl Zeiss Inc, Oberkochen, Germany). Optimal adjustment of brightness/contrast was performed in Adobe Photoshop (Adobe, San Jose, CA).

Gene ontology analysis for lens enriched proteins

Lens enriched proteins identified by in silico WB-subtraction (2.5 average spectral counts, 2.0 fold-enrichment, FDR <0.01 cut-off) were subjected to cluster based analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6 .8) for functional annotation by gene ontology (GO) categories (Huang et al. 2009). The pathways and GO categories identified were prioritized based on Benjamini corrected significant *p*-values.

Comparison of E14.5 lens proteome and transcriptome

We first identified genes common to mouse E14.5 lens proteome and mouse E14.5 lens RNA-seq data (Anand et al. 2018) with significant expression cutoff of spectral count 2 (for protein data) and 2 counts-per-million (for RNA data). These two datasets were tested by Pearson's correlation coefficient method (Mukaka 2012). Further, correlation between lens-enriched proteins and their corresponding mRNA at E10.5, E12.5, E14.5 and E16.5 (Anand et al. 2018) was also analyzed by Pearson's correlation coefficient method. Analysis was performed under 'R' statistical environment (http://www.r-project.org/) and data was visualized as scatter plots. To identify candidate genes that exhibit extraordinarily high mRNA levels compared to protein and *vice versa* (referred here as "outliers"), the $\log_2$ values of the ratio between RNA (CPM) and protein (SpC) expression for individual genes ($n = 1417$) were calculated. Then, the interquartile range (IQR) for the $\log_2$ values was calculated as third quartile (Q3) minus first quartile (Q1). The lower and upper limit for identification of outliers were defined as $Q1 - (1.5 \times IQR)$ and $Q3 + (1.5 \times IQR)$ respectively, based on a previous approach (Cho and Eo 2016). The outliers with $\log_2$(RNA/protein) $> Q3 + (1.5 \times IQR)$ represent candidates with relatively high RNA expression compared to protein (*i.e.* compared to other candidates), and the outliers with $\log_2$(RNA/protein) $< Q1 + (1.5 \times IQR)$ represent candidates with relatively high protein expression compared to RNA.

iSyTE 2.0 based access for lens proteome data

Web-based publicly accessible custom annotation University of California at Santa Cruz (UCSC) Genome Browser (Mouse GRCm38/mm10 assembly) tracks were developed to visualize protein expression and enrichment scores for E14.5 lens. Lens protein expression and enrichment scores were converted into BED (Browser Extensible Data) format for display as annotation track in the UCSC genome browser. The custom tracks for Human GRCh38/hg38 assembly were also developed to the corresponding mouse genes. These tracks are made accessible through the iSyTE 2.0 webpage via newly developed weblinks under the tab "Mouse lens Proteome" at https://research.bioinformatics.udel.edu/iSyTE/.

Results

Proteome data generation and quality assessment

To generate lens and WB proteomes to allow *in silico* subtraction comparative analysis on the protein level, we followed an established pipeline (Fig. 1A). Mouse E14.5 lens and WB (eye tissue removed) were micro-dissected and subjected to protein analyses steps described in the flow-chart (Fig. 1B). Briefly, 55 μg protein were used for each sample of lens and WB (*n*=5 samples for each of lens and WB) (Table 1). This was followed by trypsin digestion and equal loading of the resulting peptides for high-throughput tandem mass spectrometry (MS/MS) analysis for the generation of spectral counts (SpC). MS/MS detected 2371 proteins in the E14.5 lens based on the following cut-off (2 distinct peptides per protein in at least one sample) (Supplementary Table S1). All the lens samples had an average of 20,670 SpC, while all the WB samples had an average of 40,820 SpC (Table 1). To address these differences in SpC between lens and WB samples, total average SpC was subjected to TMM (trimmed mean of M-values) normalization using edgeR (Fig. 1B).

Next, to assess the quality of the lens and WB TMM normalized SpC proteome data, we performed cluster analysis by multidimensional scaling. This showed that while individual biological replicates of the lens and WB samples clustered together, overall the lens and WB samples clustered separately from each other (Fig. 2A). To further assess data quality, we derived boxplots for the normalized SpC datasets. The median expression levels were similar between all the lens samples and all the WB samples (Fig. 2B).

To assess sample to sample correlation among the lens and WB samples, we performed scatter plot comparisons in all combinations for lens and WB samples (Fig. 2C, D). This analysis shows that all samples of the same type (*i.e.* either lens or WB) were highly correlated. The five lens samples correlated with each other at *r* value >0.97 as did all the five WB samples. Next, we generated a scatter plot to represent the comparison between the average SpC of the lens and WB expressed proteins. This analysis also shows that there is no correlation ($r = 0.4919$) between the lens and WB, in turn confirming the findings of the cluster analysis (Fig. 2E).

MS/MS *in silico* subtraction identifies lens-enriched proteins

To identify high-priority proteins, we sought to take an approach involving "*in silico* WB-subtraction" that has proved to be effective in prioritization of genes from high-throughput microarrays or RNA-seq analysis. *In silico* WB-subtraction identifies genes with enriched expression in the lens compared to WB. To extend an analogous approach on the protein-level, the average SpC for all samples was computed from the scaled (normalized) data for each protein, and those 2.5 SpC were considered in the analysis. This filter identified 2,118 proteins that could be tested for differential expression between the lens and WB samples. At 2.0 fold-enrichment and FDR <0.01 cut-off, 422 proteins were found to have enriched expression in the lens compared to WB (Fig. 3A) (Supplementary Table S2). The *in silico* WB-subtraction approach worked effectively as demonstrated by the following downstream analyses that together show that many proteins linked to lens development and cataract are found among the top lens enriched candidates (Fig. 3B). Importantly, several proteins that

were not detected in the top candidates based on only "expression", were now detected by *in silico* WB-subtraction.

The utility of *in silico* WB-subtraction was explored by first comparing the top 30 proteins in the "lens expression" (not subjected to *in silico* WB subtraction) and the "lens enriched" list of candidates (Fig. 4A). The "lens expression" list contained several crystallins such as Crybb1 (Crystallin, beta B1), Cryaa (Crystallin, alpha A), Crygf (Crystallin, gamma F), Crybb3 (Crystallin, beta B3), Cryba1 (Crystallin, beta A1), Cryga (Crystallin, gamma A), Crygd (Crystallin, gamma D), Crygb (Crystallin, gamma B), Crygc (Crystallin, gamma C) and Cryba2 (Crystallin, beta A2), which is not surprising, given the high expression of crystallin proteins in the lens (Fig. 4A). In addition, the top 30 lens expression list only contained two non-crystallin proteins, namely Vim (Vimentin) and Myh9 (Myosin, heavy polypeptide 9, non-muscle), which are linked to cataract (Heath et al. 2001; Müller et al. 2009). However, the crystallins Cryab (Crystallin, alpha B), Crygn (Crystallin, gamma N) and Crygs (Crystallin, gamma S) were not among the top 30 proteins in the "lens expression" alone list (Fig. 4A). Moreover, majority of these top candidate proteins in the lens expression list were ubiquitously expressed factors, common to the general functioning of cells, and not necessarily specific to the lens. On the other hand, the "lens-enriched" list contained all the crystallins identified by the "lens expression" list and further also identified Cryab, Crygn and Crygs (Fig. 4A). Importantly, the lens enriched list identified several non-crystallin proteins linked to cataract that were not present in the top 30 proteins in the "lens expression" alone list. For example, in *silico* WB subtraction identified the proteins Aldh1a1 (Aldehyde dehydrogenase family 1, subfamily A1), Bfsp1(Beaded filament structural protein 1, in lens-CP94), Bfsp2 (Beaded filament structural protein 2, phakinin), Caprin2 (Caprin family member 2), Cryab, Crygs, Gja8 (Gap junction protein, alpha 8), Lama1 (Laminin, alpha 1), Mip (Major intrinsic protein of lens fiber), Prox1 (Prospero homeobox 1) and Tdrd7 (Tudor domain containing 7), which are all linked to cataract, among the top 30 lens enriched candidates (Fig. 4A). Even though they were not necessarily among the top highly expressed proteins in the lens, all of these candidates exhibited higher expression in the lens compared to WB (Fig. 4B). This explains why the *in silico* WB-subtraction strategy was effective in identifying these important lens proteins.

Detailed analysis of lens-enriched proteins

Comparison of the top 30 lens enriched proteins versus lens expressed showed that the *in silico* WB-subtraction strategy can be applied effectively for predicting important proteins for lens biology and cataract. Furthermore, it showed that the lens enriched list identified many candidates that were missed by analysis of lens expression alone. To gain further insights from these datasets, we extended this analysis and compared the top 150 lens enriched proteins with the top 150 lens expressed proteins. Interestingly, we find that while 60 of 150 proteins (40%) were commonly identified by both lens expression and lens enrichment, majority (90 of 150 proteins; 60%) were unique to each group.

To gain detailed insights into their significance to biology, the proteins identified by *in silico* WB-subtraction were subjected to evidence-based curation in the published literature. This analysis showed that from the 150 lens enriched candidates, 48 proteins were found to be

associated with lens and/or eye defects (Table 2). Importantly, 19 of these known cataract-linked candidates were found only in the top lens protein enrichment list but not in the top lens protein expression list. These are: Arvcf (Armadillo repeat gene deleted in velocardiofacial syndrome), Atp5d (ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit), Cap2 (CAP, adenylate cyclase-associated protein, 2), Cdh2 (Cadherin 2), Celf1 (CUGBP, Elav-like family member 1), Col4a2 (Collagen, type IV, alpha 2), Cryab, Crygs, Eml2 (Echinoderm microtubule associated protein like 2), Lama1, Naa10 (N(alpha)-acetyltransferase 10, NatA catalytic subunit), Pepd (Peptidase D), Pon2 (Paraoxonase 2), Prox1, Sarnp (SAP domain containing ribonucleoprotein), Sipa1l3 (Signal-induced proliferation-associated 1 like 3), Sod1 (Superoxide dismutase 1, soluble), Stk39 (Serine/threonine kinase 39) and Synm (Synemin, intermediate filament protein). Further, proteins linked to other eye defects were also identified among this list (Table 2). In addition to the above lens enriched candidates, the list of the top 150 lens expressed proteins also contains promising candidates (Table 3).

Next, we examined whether gene-specific knockout (KO) mouse models were available for the top 150 lens enriched proteins (uncharacterized), preferably with initial evidence suggesting lens defects/cataract. Therefore, we analyzed mouse KO phenotypes for the top 150 lens enriched proteins in the International Mouse Phenotyping Consortium (IMPC) database. We found KO mouse models with documented preliminary evidence for a lens or eye related phenotype for several new candidates such as *Eml2*, *Samp*, *Cotl1* and *Pepd*, which, importantly, have not been examined in detail or characterized by the lens research community (Table 4). Further, although IMPC KO mouse models with lens defects have been reported for the lens enriched candidates *Lama1*, *Cap2* and *Arvcf*, these have not been characterized in detail, and the cellular, molecular and pathological basis of these phenotypes remains to be examined. Further, *Synm*, a highly lens enriched protein (ranked 21 of 422; among the top 5%) is known to be associated with cataract in human cases of Meckel syndrome, and a KO mouse model for this gene is available at Knockout Mouse Project (KOMP) at University of California, Davis. Similar to the candidates described above, the *Synm* KO mouse has not been characterized in detail and thus represents a novel resource for understanding the pathological basis of cataract, as suggested by our new proteome data.

In addition to these candidates, the top 150 lens enriched proteins include several candidates that are associated with other eye related defects and therefore their expression in the lens may be reflective of their indirect impact on these tissues. For example, the lens enriched protein Gss is linked to rod-cone dystrophy that presents with maculopathy (Burstedt et al. 2009). Other candidates are as follows: Nap1l4 (Nucleosome assembly protein 1-like 4) is associated with refractive errors (Chen et al. 2016), Gsn (Gelsolin) is associated with lattice corneal dystrophy type II (Huerva et al. 2007), Atp6v1g1 (ATPase, H+ transporting, lysosomal V1 subunit G1) is associated with regulation of eye pressure (Nelson and Harvey 1999), Slc7a5 (Solute carrier family 7 (cationic amino acid transporter, y+ system), member 5) is associated with central serous chorioretinopathy (CSC) (Miki et al. 2018), Bcl2l13 (BCL2-like 13) is associated with rough eye phenotypes in *Drosophila* (Nakazawa et al. 2016) and Cbx3 is associated with abnormally patterned eyes/reduced numbers of ommatidia in *Drosophila* (Kato et al. 2007).

Additionally, there are several candidates in the top 150 lens enriched proteins for which there is experimental evidence for lens expression in the published literature/databases, but these have not been functionally characterized in detail, thus making them promising candidates for future studies. These candidates are Ass1 (Argininosuccinate synthetase 1) (Audette et al. 2016; Wang et al. 2017a), Cttn (Cortactin) (Cheng et al. 2013), Cxadr (Coxsackie virus and adenovirus receptor) (Bassnett et al. 2009), Dkk3 (Dickkopf WNT signaling pathway inhibitor 3) (Ang et al. 2004; Forsdahl et al. 2014; Ji et al. 2016), Eml2 (Medvedovic et al. 2006), Hmga2 (High mobility group AT-hook 2) (Lord-Grignon et al. 2006), Hmgn1 (High mobility group nucleosomal binding domain 1) (Lucey et al. 2008), Igfbp7 (Insulin-like growth factor binding protein 7) (Abu-Safieh et al. 2011), Pgam2 (Phosphoglycerate mutase 2) (Hoang et al. 2014), Ppp1cc (Protein phosphatase 1 catalytic subunit gamma) (Srivastava et al. 2017), Rpl13 (Ribosomal protein L13) (Zhao et al. 2019), Rps27 (Ribosomal protein S27) (Zhao et al. 2019), Rps27a (Ribosomal protein S27A) (Srivastava et al. 2017; Zhao et al. 2019) and Sorbs1 (Geisert et al. 2009). Further, there are several candidates in the top 150 lens enriched proteins that are mis-expressed in the lens in animal models with genetic perturbation for known factors linked to lens biology and/or cataract. For example, Ass1, Bpnt1 (Bisphosphate 3'-nucleotidase 1), Cpt2 (Carnitine palmitoyltransferase 2), Dbnl (Drebrin-like), Kif1a (Kinesin family member 1A), Metap1 (Methionyl aminopeptidase 1) and Rilpl1 (Rab interacting lysosomal protein-like 1) are reduced in *Prox1* cKO lens that exhibits fiber cell defects (Audette et al. 2016), Aldh1a7 (Aldehyde dehydrogenase, cytosolic 1) is reduced in *Klf4* cKO lens (Gupta et al. 2013), Ggct (Gamma-glutamyl cyclotransferase) is elevated in *Mip*-mutant (Lop/+) that exhibits cataract (Zhou et al. 2016) and Dst (Dystonin) is reduced in *Ilk* (integrin linked kinase) cKO lens (Teo et al. 2014). Moreover, Pygm (Muscle glycogen phosphorylase) and Bpnt1 (Bisphosphate 3'-nucleotidase 1) are elevated and reduced, respectively, in *Hsf4* (*Heat shock transcription factor 4*) KO lens, which exhibit cataract (He et al. 2010; Tian et al. 2018), while Ube2v1 (Ubiquitin-conjugating enzyme E2 variant 1) and Rpl36a (Ribosomal protein L36A) are reduced in *Sip1* (*Zeb2, Zinc finger E-box binding homeobox 2*) cKO lens that exhibit lens defects (Manthey et al. 2014). Further, within the top 150 lens enriched candidates, there are factors whose expression was altered in response to oxidative stress (a key factor impacting cataract pathology), and therefore are relevant to lens biology. For example, the top lens enriched protein Pfdn4 is elevated due to H_2O_2-induced oxidative stress in human lens epithelial (HLE) cells (Goswami et al. 2003), while Gls is elevated in Glutathione-deficient LEGSKO mouse lens (Whitson et al. 2017). Finally, the lens enriched protein Amph was found to be elevated during trans differentiation from cornea to lens in *Xenopus* (Day and Beck 2011), indicating that genes associated with lens formation are prioritized in the pool of lens enriched proteins. Finally, immunostaining was used to validate the expression of select high-priority proteins, namely Eml2, Igfbp7, Nol3 and Slc7a5, in the lens (Fig. 5). Together, these findings indicate the effectiveness of the *in silico* WB-subtraction based lens enrichment approach toward identifying new promising candidates associated with lens development and cataract.

Proteome-based *in silico* subtraction identifies high-priority lens membrane proteins

Because several lens membrane proteins have been previously linked to cataract (Shiels et al. 1998; Mackay et al. 1999; Berry et al. 2000; Kloeckener-Gruissem et al. 2008; Lin et al.

2013; Swarup et al. 2018), we next sought to identify high-priority candidates in this class of proteins that are expressed/enriched in the lens. We first compared our lens expressed proteins to previously reported lens membrane proteins analysis performed in mouse strain C57BL/6 (Bassnett et al. 2009). Our data identified 92 lens membrane proteins that were also independently identified by the previous study (Supplementary Table S3). Interestingly, of these 92 lens membrane proteins, 33 are found to be lens-enriched based on *in silico* WB-subtraction, identifying these as high-priority candidates (Table 5). Importantly, of the 33 high-priority lens membrane proteins, we identified Mip and Gja8 that are known to be associated with cataract, indicating that other members in the list may also be important to lens biology. Within the lens membrane proteins, members of the solute carrier (Slc) family have been linked to cataract (Kloeckener-Gruissem et al. 2008; Swarup et al. 2018). Therefore, we focused on identifying the other members of this protein family that are expressed/enriched in the lens. We identified several Slc proteins such as Slc7a5, Slc2a1 (GLUT1), Slc3a2, Slc25a4, Slc25a5, Slc25a3, and Slc25a11, which are also identified in a previous study (Bassnett et al. 2009). In addition, we identified several previously unreported new Slc family proteins such as Slc16a1, Slc25a13 and Slc25a12 (Supplementary Table S3) that are expressed in the lens. Further, Slc7a5, Slc2a1 (GLUT1) and Slc3a2 were identified as highly lens enriched (Table 5). Importantly, Slc2a1 (GLUT1) has is already been shown to be linked to cataract (Swarup et al. 2018), thus indicating the effectiveness of the *in silico* WB-subtraction strategy in identifying lens membrane proteins potentially associated with cataract.

Next, we compared our data on lens expressed proteins with previously reported adult human lens membrane proteins (Wang et al. 2013). This analysis commonly identified 24 lens membrane proteins, among which 15 were highly enriched in mouse lens (Table 6). Comparison of these 15 human lens enriched membrane proteins with 33 mouse lens enriched membrane proteins led to identification of 11 common proteins, namely Bcl2l13, Cadm1, Cdh2, Col4a2, Cxadr, Gja8, Itgb1 (Integrin beta 1), Mip, Nrcam (Neuronal cell adhesion molecule), Slc2a1 (GLUT1) and Slc3a2. This list contains several established cataract-linked proteins (Gja8, Mip, Nrcam, Slc2a1 (GLUT1)) as well as several other uncharacterized proteins that represent high-priority candidates for future studies aimed at membrane protein research in lens biology.

Gene ontology analysis of lens enriched proteins

Next to further examine the relevance of lens enriched proteins to lens biology, cluster-based analysis on these candidates was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6 .8) for functional annotation by gene ontology (GO) categories (Fig. 6) (Supplementary Table S4). This analysis assigned 406 (out of 422) lens enriched proteins into 68 annotation clusters. The top clusters included GO categories that are relevant to lens biology and cataract. These were "eye lens protein", "protein folding", "Ribonucleoprotein", "Protein biosynthesis", and "cell-cell adherens junction", among others (Fig. 6) (Supplementary Table S4). Within the cluster for "eye lens protein", other lens-relevant sub-categories were identified such as "structural constituent of eye lens", "Beta/Gamma crystallin", "lens development in camera-type eye", "eye development" and "visual perception". In addition to the established lens proteins, other

potentially important regulatory factors in these identified clusters were RNA-binding proteins, initiation and elongation factors for protein synthesis, DNA-binding factors, chaperones/heat shock proteins, actin-binding proteins and methyl transferases (Fig. 6) (Supplementary Table S4). This analysis shows that the high-priority lens enriched proteins identified by *in silico* WB-subtraction represent an important set of candidates associated with lens biology.

Lens enriched proteins are also enriched in RNA-based iSyTE

Next, we examined whether the top candidates identified by *in silico* WB-subtraction were also independently identified on the RNA level by microarray analysis. Notably, all 30 candidates show lens enrichment at E14.5 at both RNA and protein level. Further, we find that all the top 30 lens enriched proteins (Table 5) are also enriched in the lens on the RNA level at additional mouse embryonic stages ranging from E10.5 through P56 according to Affymetrix microarray analysis (Fig. 7). These include the known lens enriched proteins such as Crystallins as well as other key factors in the lens such as Aldh1a1, Caprin2, Mip, Prox1 and Tdrd7 that are linked to cataract. Three proteins, namely, Synm, Nol3 and Snx18 are not enriched at E10.5 but they are enriched in the later stages. In addition, Caprin2 and Bfsp1 show low enrichment at the RNA level at E10.5 but are sharply elevated at later stages. These findings suggest that the top lens enriched proteins are similarly detected by both the RNA-based and the protein-based iSyTE.

Comparison of lens proteome to lens transcriptome

We recently published RNA-seq data on mouse lenses at E10.5, E12.5, E14.5 and E16.5 (Anand et al. 2018), which offers the opportunity to compare lens gene expression on the protein and RNA levels. We first considered proteins that were expressed at ≥ 2.0 SpC (n =1685) in E14.5 lens for comparative analysis with mRNAs that were expressed at ≥ 2.0 counts per million (CPM) in E14.5 lens. This analysis identified 1417 genes that were commonly expressed in the RNA-seq and the proteome datasets (Supplementary Table S5). This data will direct researchers to compare the RNA and protein levels of lens expressed genes.

Next, we examined the mRNA-protein correlation for these 1417 commonly identified genes in E14.5 lens. This analysis indicated an overall positive correlation between the transcriptome and the proteome ($r = 0.6$) (Fig. 8A). We were next interested in examining if the correlation was higher for candidates that were recognized as "lens-enriched" by *in silico* WB-subtraction in the protein dataset. Furthermore, we were interested to evaluate if this correlation increased with developmental progression. Therefore, we performed correlation analysis of the lens enriched proteins ($n = 422$) with RNA-seq data on E10.5, E12.5 and E14.5 (Fig. 8B–D). The mRNA-protein correlation was generally higher in lens enriched proteins compared to lens expressed proteins (Fig. 8B–D). Furthermore, the mRNA-protein correlation show an increasing trend with progressive development of the lens, E10.5 ($r = 0.63$), E12.5 ($r = 0.80$) and E14.5 ($r = 0.82$). These data indicate that both RNA-seq and protein profiling identify lens enriched genes that exhibit high correlation.

Next, we sought to further examine the RNA and protein datasets to identify and evaluate candidate genes that exhibit extraordinarily high mRNA levels compared to protein and *vice versa*. To identify such candidates, the $\log_2$ values of the ratio between RNA (CPM) and protein (SpC) expression for individual genes ($n =1417$) was calculated. Then, the interquartile range (IQR) for the $\log_2$ values was calculated and candidates lying outside ± 1.5 times IQR based on a previous approach (Cho and Eo 2016), were designated as "outliers" (extraordinarily high mRNA levels compared to protein and *vice versa*). These include genes with $\log_2$(RNA/protein) >1.5 times IQR that represent candidates with high RNA expression compared to proteins, and the genes with $\log_2$(RNA/protein) <1.5 times IQR that represent candidates with high protein expression compared to RNA. These candidates were considered for further analysis (Supplementary Table S6).

Examination of RNA expression trend of these candidates across the various lens developmental stages (E10.5, E12.5, E14.5 and E16.5) showed that the outliers with high $\log_2$(RNA/protein ratio, *i.e.* relatively high RNA compared to protein) showed sharp increase in mRNA level at stage E14.5 compared to the earlier embryonic stages E10.5 and E12.5. There were 10 such candidates which include Cryba1, Cryga, Dkk3, Gja8 and Mip among others (Fig. 9; Supplementary Table S7). For majority of the 31 candidates with low $\log_2$(RNA/protein ratio; *i.e.* relatively low RNA compared to protein), a dynamic change at the RNA level was observed at stage E14.5, which manifested as a sharp decrease or a sharp increase in mRNA expression compared to preceding stages (Fig. 9; Supplementary Table S7).

Finally, we were interested in those proteins that were expressed in the E14.5 lens but for which the corresponding mRNAs were not detected in the E14.5 lens RNA-seq dataset. Interestingly, for 22 such candidates, the corresponding mRNAs were expressed at the earlier developmental stages, E10.5 and E12.5 (Supplementary Table S8).

Together, these data suggest that post-transcriptional and post-translational mechanisms of gene expression control are at play in lens development, leading to stoichiometric differences in RNA versus protein levels of lens-enriched genes..

Public access to high-priority lens proteins via iSyTE and UCSC Genome Browser

Next, we aimed to provide public access to the lens proteome data in a format that would be user friendly and applicable to human genetics-based studies. Therefore, we developed new custom tracks on the University of California at Santa Cruz (UCSC) genome browser mouse (GRCm38/mm10) assembly for the 2118 proteins. These tracks represent either the lens expression or lens enrichment of these proteins. This mouse lens protein enrichment/ expression information was also used to make tracks for the UCSC genome browser human assembly. The access to these tracks are available via the iSyTE website at the following url: https://research.bioinformatics.udel.edu/iSyTE under "Lens Gene Expression" > "Protein Lens Enrichment". These tracks allow effective visualization of lens protein expression or enriched expression in the context of a specific mapped interval and other genome-level databases. Alternately, specific candidates of interest from patient's exome-seq data can be analyzed for their expression/enriched expression in the lens via these tracks. Lens enrichment and lens expression of a candidate protein is indicated by a heat-map, which can

be used for evaluating their relevance to lens biology. Examples of several cataract-linked factors (*e.g.* TDRD7) that can be effectively visualized by this representation are shown (Fig. 10).

Discussion

Application of high-throughput approaches holds promise to define lens biology on the systems level (Anand and Lachke 2017). Indeed, researchers are increasingly applying genome-level transcriptomics and proteomics to characterize wild-type and KO/mutant lens tissue to gain insight into lens development and the pathology of lens defects, including cataract. For example, on the proteome level, Bassnett and colleagues (Bassnett et al. 2009) have characterized the mouse lens membrane protein profile using MudPIT (Multidimensional protein identification technology) while Wang and colleagues (Wang et al. 2013) have characterized the profile of human lens fiber cell insoluble membrane proteins, along with phosphoproteomic analysis. Further, Khan and colleagues (Khan et al. 2018a) reported the protein profile of the developing mouse lens (mouse strain C57BL/6) using tandem mass tag (TMT) based proteomic approaches. More recently, Zhao and colleagues (Zhao et al. 2019) have performed protein profiling using 2D-LC/MS (tandem mass spectrometry) for isolated lens epithelium and fiber cells from newborn mouse lens (mouse strain CD-1).

While these studies have greatly extended our knowledge of lens expressed proteins, they all face the common challenge of effective prioritization of candidates. This is because of the high-throughput nature of the approach that identifies thousands of expressed candidates. Furthermore, use of the proteome approach has been limited for embryonic lens development. Finally, these datasets are deposited in online databases and are not available to the public in ready to use/analysis format. Therefore, in the present study, we had three goals: (1) to generate a new protein profile for mouse embryonic lens at the key stage E14.5, (2) extend the *in silico*-WB subtraction strategy–which has been successful in transcriptomics studies, to proteome-level analysis of the lens–in order to identify high-priority candidates from thousands of expressed proteins, and (3) make this rich information on lens protein expression/enriched expression widely available through iSyTE and UCSC Genome Browser in a user-friendly manner. Furthermore, extending iSyTE by including proteome-level information is significant and necessary for gene discovery as our own studies on various RNA-binding proteins indicate that post-transcriptional control of gene expression is essential for proper development of the lens, and that defects in these regulatory processes result in cataract.

Here, we show that *in silico* WB-subtraction strategy can be successfully applied to proteome datasets for identification of high-priority candidates in lens development and cataract. Indeed, this approach identifies previously known cataract/lens defects-associated proteins as well as many new potential regulators/factors in the lens. Further, the top lens-enriched genes are commonly identified by both RNA-based iSyTE datasets and protein-based data (this study). In addition to lens protein enrichment analysis, comparison of our lens protein expression data with previously reported lens protein profiles commonly identifies hundreds of proteins in the lens, thus giving confidence to the conclusion that

these proteins are present in the lens. This is even more remarkable when one considers that these proteome analyses have performed on lenses at different developmental/post-natal stages by different research groups using different approaches. For example, comparison of the top 1000 expressed proteins in this study to the top 1000 proteins in the E15.5 mouse proteome published by Khan and colleagues (Khan et al. 2018a) shows that 78% of proteins are commonly identified (Supplementary Table S9). Similarly, comparison of the top 1000 expressed proteins in this study to the top 1000 proteins in the P0 mouse fiber and epithelial cell proteome published by Zhao and colleagues (Zhao et al. 2019) shows that 64% of proteins are commonly identified (Supplementary Table S10). Further, comparison of top 1000 proteins in all three studies commonly identify 565 proteins in the lens (Supplementary Table S11).

Although our experiment was not designed to exclusively identify the lens membrane proteins, we analyzed lens expressed membrane proteins in our list to identify high-priority lens membrane protein candidates. Comparisons of our data with previously reported mouse and human lens membrane proteome data commonly identify 33 and 15 proteins, respectively, to be expressed in the E14.5 mouse lens. These analyses identified several promising solute carrier family proteins (*e.g.* Slc7a5 and Slc3a2) that are excellent candidates for future investigations in the lens. Further, the Slc family proteins identified in this study, namely, Slc25a4, Slc25a5, Slc25a3, Slc25a11, Slc25a13 and Slc25a12 are also reported to be bound to inner mitochondrial membrane. In light of the importance of mitochondrial solute transport in other cell types (Haitina et al. 2006; Gutiérrez-Aguilar and Baines 2013), these proteins present as excellent candidates for future investigations in the lens.

A distinct advantage of using unbiased omics-level approaches is the opportunity to identify uncharacterized proteins. Analysis of our data using a criterion of presence of two or more distinct peptides per protein in at least one lens sample, we identify several uncharacterized proteins. These proteins (as denoted by their uniport accession) include Q8C3W1, Q5EBG8, Q9D727, Q91V76, Q9D7E4 and Q9D1K7, all of which are also identified independently by Zhao and colleagues (Zhao et al. 2019). Thus, these lens-expressed proteins represent novel candidates for further studies in the lens.

In addition to identifying many promising candidates for future analyses in the lens, comparison of lens transcriptome and proteome datasets allows us an opportunity to gain new insights into the correlation between RNA and protein level information in the lens. Our findings show that the top lens-enriched proteins are well correlated with the top lens-enriched RNAs. This analysis also reveals a subset of candidates for which the RNA and protein levels do not correspond. This may be due to biological mechanisms or due to technical limitations between the two different types of global analytical methods. In support of the biological basis of these observations, potential explanations may lie in the differences in the rates of transcription and translation that can be further impacted by post-transcriptional (*e.g.* mRNA stability) and/or post-translational (*e.g.* protein stability) control mechanisms. Indeed, evidence in support of these mechanisms in the lens has been suggested as early as in 1981 by Drs. David Beebe and Joram Piatigorsky who demonstrated that while δ-crystallin mRNA levels were similar in early and late chicken lenses, its

translation efficiency was substantially reduced in the late stage (Beebe and Piatigorsky 1981). Further, it was shown that ectopic increase of δ-crystallin mRNA levels could not lead to increase in its translation into protein in older lenses (Beebe and Piatigorsky 1981). Conversely, the protein levels of the cholesterol biosynthesis enzyme HMGRA (3-hydroxy-3-methylglutaryl coenzyme A reductase) can be elevated in the lens without a similar increase in its mRNA levels (Cenedella 1995). Additionally, it was observed that gamma-crystallin mRNAs are present at birth in mouse epithelial cells but are translated only at a later stage (Wang et al. 2004). Indeed, in this study, we find *Cryga* to have relatively high mRNA to protein ratio in the E14.5 lens.

In our analysis, the 10 candidates with high $\log_2$(RNA/protein ratio; *i.e.* relatively high RNA compared to protein) showed a trend of elevated mRNA expression from stages E10.5 through E16.5. This suggests that the overall extent of protein synthesis or protein stability may be relatively low compared to RNA synthesis or RNA stability for these candidates. On the other hand, candidates with low $\log_2$(RNA/protein ratio; *i.e.* relatively low RNA compared to protein) suggest that the overall extent of RNA synthesis or RNA stability may be relatively low compared to protein synthesis or protein stability. Indeed, protein synthesis and translation of specific mRNAs can be affected by specific factors in the translational machinery, as shown for control of gamma crystallin expression by eukaryotic initiation factor eIF3ha in zebrafish polysomes (Choudhuri et al. 2013; Riba et al. 2019). Examination of the 31 candidates in this category showed two broad trends of mRNA expression in stages E10.5 through E16.5. Some mRNAs showed precipitous reduction while others showed precipitous elevation in their levels between these stages. The overall trend of low mRNA to high protein ratio may reflect the dynamics of this process as well as complex post-transcriptional control mechanisms. However, it should be noted that technical limitations may also contribute to these discrepancies (*e.g.* differences in buffer, conditions of protein digest, and inherent protein properties).

Importantly, we have made the rich new lens protein expression information publicly available in a user-friendly format as custom annotation-tracks for both mouse and human genomes at UCSC Genome Browser, which is freely accessible via iSyTE (https:// research.bioinformatics.udel.edu/iSyTE/). This allows for ready visualization of promising candidate proteins in the context of various other rich information on the ICSC Genome Browser, in turn making iSyTE a comprehensive tool for cataract gene discovery and expression analysis at both transcriptome and proteome level.

However, it is important to note that protein profiling is impacted by variables such as the use of different buffers, enzymatic digestion conditions and analytical approaches. This can lead to variations in coverage/depth of proteins identified by any particular method. Further, use of stringent criteria for prioritizing the best candidates is important but can result in false negatives. For example, even though Cdkn1b (p27^Kip1), Crybb2, Dnajb2, Epha2, Gja3, Lim2, Pax6 and Rbm24, are expressed (with average lens SpC >1) and enriched in the lens (2FC) they are among the genes that did not pass the cut-off of average 2.5 average spectral count that we used in this study (Supplementary Table S12). Even with these limitations, the present study has demonstrated the efficacy of *in silico* WB-subtraction to

identify many proteins as promising new candidates for detailed investigation in lens biology and cataract.

Conclusion

In sum, this study reports for the first time the application of *in silico* WB-subtraction strategy on proteome datasets for identifying high-priority proteins linked to human defects and disease, namely cataract. Further, this work provides free public access to this rich lens proteome data via new custom tracks on the UCSC genome browser available through the eye gene discovery tool iSyTE. Importantly, this report can be taken as proof-of-principle that *in silico* WB-subtraction strategy can be effectively applied to other organs/tissues at proteome level to expedite human defects/disease gene discovery.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Abu-Safieh L, Abboud EB, Alkuraya H, et al. (2011) Mutation of IGFBP7 causes upregulation of BRAF/MEK/ERK pathway and familial retinal arterial macroaneurysms. Am J Hum Genet 89:313–319. 10.1016/j.ajhg.2011.07.010 [PubMed: 21835307]

Agrawal SA, Anand D, Siddam AD, et al. (2015) Compound mouse mutants of bZIP transcription factors Mafg and Mafk reveal a regulatory network of non-crystallin genes associated with cataract. Hum Genet 134:717–735. 10.1007/s00439-015-1554-5 [PubMed: 25896808]

AlFadhli S, Abdelmoaty S, Al-Hajeri A, et al. (2012) Novel crystallin gamma B mutations in a Kuwaiti family with autosomal dominant congenital cataracts reveal genetic and clinical heterogeneity. Mol Vis 18:2931–2936 [PubMed: 23288985]

Anand D, Kakrana A, Siddam AD, et al. (2018) RNA sequencing-based transcriptomic profiles of embryonic lens development for cataract gene discovery. Hum Genet 137:941–954. 10.1007/s00439-018-1958-0 [PubMed: 30417254]

Anand D, Lachke SA (2017) Systems biology of lens development: A paradigm for disease gene discovery in the eye. Exp Eye Res 156:22–33. 10.1016/j.exer.2016.03.010 [PubMed: 26992779]

Ang SJ, Stump RJW, Lovicu FJ, McAvoy JW (2004) Spatial and temporal expression of Wnt and Dickkopf genes during murine lens development. Gene Expr Patterns 4:289–295. 10.1016/j.modgep.2003.11.002 [PubMed: 15053977]

Arora AI, Johar K, Gajjar DU, et al. (2012) Cx43, ZO-1, alpha-catenin and beta-catenin in cataractous lens epithelial cells. J Biosci 37:979–987. 10.1007/s12038-012-9264-9 [PubMed: 23151788]

Audette DS, Anand D, So T, et al. (2016) Prox1 and fibroblast growth factor receptors form a novel regulatory loop controlling lens fiber differentiation and gene expression. Development 143:318–328. 10.1242/dev.127860 [PubMed: 26657765]

Bassnett S, Wilmarth PA, David LL (2009) The membrane proteome of the mouse lens fiber cell. Mol Vis 15:2448–2463 [PubMed: 19956408]

Beebe DC, Piatigorsky J (1981) Translational regulation of delta-crystallin synthesis during lens development in the chicken embryo. Dev Biol 84:96–101 [PubMed: 7250504]

Berry V, Francis P, Kaushal S, et al. (2000) Missense mutations in MIP underlie autosomal dominant "polymorphic" and lamellar cataracts linked to 12q. Nat Genet 25:15–17. 10.1038/75538 [PubMed: 10802646]

Berry V, Francis P, Reddy MA, et al. (2001) Alpha-B crystallin gene (CRYAB) mutation causes dominant congenital posterior polar cataract in humans. Am J Hum Genet 69:1141–1145. 10.1086/324158 [PubMed: 11577372]

Bharathidevi SR, Babu KA, Jain N, et al. (2017) Ocular distribution of antioxidant enzyme paraoxonase & its alteration in cataractous lens & diabetic retina. Indian J Med Res 145:513–520. 10.4103/ijmr.IJMR_1284_14 [PubMed: 28862184]

Billingsley G, Santhiya ST, Paterson AD, et al. (2006) CRYBA4, a novel human cataract gene, is also involved in microphthalmia. Am J Hum Genet 79:702–709. 10.1086/507712 [PubMed: 16960806]

Burstedt MSI, Ristoff E, Larsson A, Wachtmeister L (2009) Rod-cone dystrophy with maculopathy in genetic glutathione synthetase deficiency: a morphologic and electrophysiologic study. Ophthalmology 116:324–331. 10.1016/j.ophtha.2008.09.007 [PubMed: 19111905]

Castellini M, Wolf LV, Chauhan BK, et al. (2005) Palm is expressed in both developing and adult mouse lens and retina. BMC Ophthalmol 5:14 10.1186/1471-2415-5-14 [PubMed: 15969763]

Cavalheiro GR, Matos-Rodrigues GE, Zhao Y, et al. (2017) N-myc regulates growth and fiber cell differentiation in lens development. Dev Biol 429:105–117. 10.1016/j.ydbio.2017.07.002 [PubMed: 28716713]

Cenedella RJ (1995) Role of transcription, translation, and protein turnover in controlling the distribution of 3-hydroxy-3-methylglutaryl coenzyme A reductase in the lens. Invest Ophthalmol Vis Sci 36:2133–2141 [PubMed: 7544784]

Chambers MC, Maclean B, Burke R, et al. (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 30:918–920. 10.1038/nbt.2377 [PubMed: 23051804]

Chen F, Duggal P, Klein BEK, et al. (2016) Variation in PTCHD2, CRISP3, NAP1L4, FSCB, and AP3B2 associated with spherical equivalent. Mol Vis 22:783–796 [PubMed: 27440996]

Chen J, Zhou J, Wu J, et al. (2017) Aberrant Epigenetic Alterations of Glutathione-S-Transferase P1 in Age-Related Nuclear Cataract. Curr Eye Res 42:402–410. 10.1080/02713683.2016.1185129 [PubMed: 27348130]

Cheng C, Ansari MM, Cooper JA, Gong X (2013) EphA2 and Src regulate equatorial cell morphogenesis during lens development. Development 140:4237–4245. 10.1242/dev.100727 [PubMed: 24026120]

Cho H, Eo S-H (2016) Outlier Detection for Mass Spectrometric Data. Methods Mol Biol 1362:91–102. 10.1007/978-1-4939-3106-4_5 [PubMed: 26519171]

Cho K, Lee M, Gu D, et al. (2011) Kazrin, and its binding partners ARVCF- and delta-catenin, are required for Xenopus laevis craniofacial development. Dev Dyn 240:2601–2612. 10.1002/dvdy.22721 [PubMed: 22028074]

Choudhuri A, Maitra U, Evans T (2013) Translation initiation factor eIF3h targets specific transcripts to polysomes during embryogenesis. Proc Natl Acad Sci USA 110:9818–9823. 10.1073/pnas.1302934110 [PubMed: 23716667]

Dash S, Dang CA, Beebe DC, Lachke SA (2015) Deficiency of the RNA binding protein caprin2 causes lens defects and features of peters anomaly. Dev Dyn 244:1313–1327. 10.1002/dvdy.24303 [PubMed: 26177727]

Dash S, Siddam AD, Barnum CE, et al. (2016) RNA-binding proteins in eye development and disease: implication of conserved RNA granule components. Wiley Interdiscip Rev RNA 7:527–557. 10.1002/wrna.1355 [PubMed: 27133484]

Day RC, Beck CW (2011) Transdifferentiation from cornea to lens in Xenopus laevis depends on BMP signalling and involves upregulation of Wnt signalling. BMC Dev Biol 11:54 10.1186/1471-213X-11-54 [PubMed: 21896182]

Dong LJ, Chung AE (1991) The expression of the genes for entactin, laminin A, laminin B1 and laminin B2 in murine lens morphogenesis and eye development. Differentiation 48:157–172. 10.1111/j.1432-0436.1991.tb00254.x [PubMed: 1725162]

Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4:207–214. 10.1038/nmeth1019 [PubMed: 17327847]

Eng JK, Jahan TA, Hoopmann MR (2013) Comet: an open-source MS/MS sequence database search tool. Proteomics 13:22–24. 10.1002/pmic.201200439 [PubMed: 23148064]

Erde J, Loo RRO, Loo JA (2017) Improving Proteome Coverage and Sample Recovery with Enhanced FASP (eFASP) for Quantitative Proteomic Experiments. Methods Mol Biol 1550:11–18. 10.1007/978-1-4939-6747-6_2 [PubMed: 28188519]

Field J, Ye DZ, Shinde M, et al. (2015) CAP2 in cardiac conduction, sudden cardiac death and eye development. Sci Rep 5:17256 10.1038/srep17256 [PubMed: 26616005]

Forsdahl S, Kiselev Y, Hogseth R, et al. (2014) Pax6 regulates the expression of Dkk3 in murine and human cell lines, and altered responses to Wnt signaling are shown in FlpIn-3T3 cells stably expressing either the Pax6 or the Pax6(5a) isoform. PLoS ONE 9:e102559 10.1371/journal.pone.0102559 [PubMed: 25029272]

Geisert EE, Lu L, Freeman-Anderson NE, et al. (2009) Gene expression in the mouse eye: an online resource for genetics using 103 strains of mice. Mol Vis 15:1730–1763 [PubMed: 19727342]

Gonzalez-Huerta LM, Messina-Baas OM, Cuevas-Covarrubias SA (2007) A family with autosomal dominant primary congenital cataract associated with a CRYGC mutation: evidence of clinical heterogeneity. Mol Vis 13:1333–1338 [PubMed: 17679936]

Goswami S, Sheets NL, Zavadil J, et al. (2003) Spectrum and range of oxidative stress responses of human lens epithelial cells to H2O2 insult. Invest Ophthalmol Vis Sci 44:2084–2093. 10.1167/iovs.02-0882 [PubMed: 12714647]

Graw J, Klopp N, Neuhäuser-Klaus A, et al. (2002) Crygf(Rop): the first mutation in the Crygf gene causing a unique radial lens opacity. Invest Ophthalmol Vis Sci 43:2998–3002 [PubMed: 12202521]

Greenlees R, Mihelec M, Yousoof S, et al. (2015) Mutations in SIPA1L3 cause eye defects through disruption of cell polarity and cytoskeleton organization. Hum Mol Genet 24:5789–5804. 10.1093/hmg/ddv298 [PubMed: 26231217]

Gupta D, Harvey SAK, Kenchegowda D, et al. (2013) Regulation of mouse lens maturation and gene expression by Krüppel-like factor 4. Exp Eye Res 116:205–218. 10.1016/j.exer.2013.09.010 [PubMed: 24076321]

Gutiérrez-Aguilar M, Baines CP (2013) Physiological and pathological roles of mitochondrial SLC25 carriers. Biochem J 454:371–386. 10.1042/BJ20121753 [PubMed: 23988125]

Ha TT, Sadleir LG, Mandelstam SA, et al. (2016) A mutation in COL4A2 causes autosomal dominant porencephaly with cataracts. Am J Med Genet A 170A:1059–1063. 10.1002/ajmg.a.37527 [PubMed: 26708157]

Haitina T, Lindblom J, Renström T, Fredriksson R (2006) Fourteen novel human members of mitochondrial solute carrier family 25 (SLC25) widely expressed in the central nervous system. Genomics 88:779–790. 10.1016/j.ygeno.2006.06.016 [PubMed: 16949250]

He S, Pirity MK, Wang W-L, et al. (2010) Chromatin remodeling enzyme Brg1 is required for mouse lens fiber cell terminal differentiation and its denucleation. Epigenetics Chromatin 3:21 10.1186/1756-8935-3-21 [PubMed: 21118511]

Heath KE, Campos-Barros A, Toren A, et al. (2001) Nonmuscle myosin heavy chain IIA mutations define a spectrum of autosomal dominant macrothrombocytopenias: May-Hegglin anomaly and Fechtner, Sebastian, Epstein, and Alport-like syndromes. Am J Hum Genet 69:1033–1045. 10.1086/324267 [PubMed: 11590545]

Hoang TV, Kumar PKR, Sutharzan S, et al. (2014) Comparative transcriptome analysis of epithelial and fiber cells in newborn mouse lenses with RNA sequencing. Mol Vis 20:1491–1517 [PubMed: 25489224]

Hoehenwarter W, Klose J, Jungblut PR (2006) Eye lens proteomics. Amino Acids 30:369–389. 10.1007/s00726-005-0283-9 [PubMed: 16583312]

Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57. 10.1038/nprot.2008.211 [PubMed: 19131956]

Huerva V, Velasco A, Sánchez MC, et al. (2007) Lattice corneal dystrophy type II: clinical, pathologic, and molecular study in a Spanish family. Eur J Ophthalmol 17:424–429. 10.1177/112067210701700326 [PubMed: 17534828]

Jakobs PM, Hess JF, FitzGerald PG, et al. (2000) Autosomal-dominant congenital cataract associated with a deletion mutation in the human beaded filament protein gene BFSP2. Am J Hum Genet 66:1432–1436. 10.1086/302872 [PubMed: 10739768]

Ji B, Lim D, Kim J, et al. (2016) Increased Levels of Dickkopf 3 in the Aqueous Humor of Patients With Diabetic Macular Edema. Invest Ophthalmol Vis Sci 57:2296–2304. 10.1167/iovs.15-18771 [PubMed: 27127928]

Kakrana A, Yang A, Anand D, et al. (2018) iSyTE 2.0: a database for expression-based gene discovery in the eye. Nucleic Acids Res 46:D875–D885. 10.1093/nar/gkx837 [PubMed: 29036527]

Kasaikina MV, Fomenko DE, Labunskyy VM, et al. (2011) Roles of the 15-kDa selenoprotein (Sep15) in redox homeostasis and cataract development revealed by the analysis of Sep 15 knockout mice. J Biol Chem 286:33203–33212. 10.1074/jbc.M111.259218 [PubMed: 21768092]

Kato M, Kato Y, Nishida M, et al. (2007) Functional domain analysis of human HP1 isoforms in Drosophila. Cell Struct Funct 32:57–67. 10.1247/csf.06032 [PubMed: 17575412]

Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74:5383–5392. 10.1021/ac025747h [PubMed: 12403597]

Khan SY, Ali M, Kabir F, et al. (2018a) Proteome Profiling of Developing Murine Lens Through Mass Spectrometry. Invest Ophthalmol Vis Sci 59:100–107. 10.1167/iovs.17-21601 [PubMed: 29332127]

Khan SY, Ali M, Kabir F, et al. (2018b) Identification of novel transcripts and peptides in developing murine lens. Sci Rep 8:. 10.1038/s41598-018-28727-w

Kloeckener-Gruissem B, Vandekerckhove K, Nürnberg G, et al. (2008) Mutation of solute carrier SLC16A12 associates with a syndrome combining juvenile cataract with microcornea and renal glucosuria. Am J Hum Genet 82:772–779. 10.1016/j.ajhg.2007.12.013 [PubMed: 18304496]

Krall M, Htun S, Anand D, et al. (2018) A zebrafish model of foxe3 deficiency demonstrates lens and eye defects with dysregulation of key genes involved in cataract formation in humans. Hum Genet 137:315–328. 10.1007/s00439-018-1884-1 [PubMed: 29713869]

Lachke SA, Alkuraya FS, Kneeland SC, et al. (2011) Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. Science 331:1571–1576. 10.1126/science.1195970 [PubMed: 21436445]

Lachke SA, Higgins AW, Inagaki M, et al. (2012a) The cell adhesion gene PVRL3 is associated with congenital ocular defects. Hum Genet 131:235–250. 10.1007/s00439-011-1064-z [PubMed: 21769484]

Lachke SA, Ho JWK, Kryukov GV, et al. (2012b) iSyTE: integrated Systems Tool for Eye gene discovery. Invest Ophthalmol Vis Sci 53:1617–1627. 10.1167/iovs.11-8839 [PubMed: 22323457]

Lassen N, Bateman JB, Estey T, et al. (2007) Multiple and additive functions of ALDH3A1 and ALDH1A1: cataract phenotype and ocular oxidative damage in Aldh3a1(−/−)/Aldh1a1(−/−) knock-out mice. J Biol Chem 282:25668–25676. 10.1074/jbc.M702076200 [PubMed: 17567582]

Lin Q, Zhou N, Zhang N, et al. (2013) Genetic variations and polymorphisms in the ezrin gene are associated with age-related cataract. Mol Vis 19:1572–1579 [PubMed: 23882136]

Litt M, Kramer P, LaMorticella DM, et al. (1998) Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. Hum Mol Genet 7:471–474 [PubMed: 9467006]

Lord-Grignon J, Abdouh M, Bernier G (2006) Identification of genes expressed in retinal progenitor/ stem cell colonies isolated from the ocular ciliary body of adult mice. Gene Expr Patterns 6:992– 999. 10.1016/j.modgep.2006.04.003 [PubMed: 16765103]

Lucey MM, Wang Y, Bustin M, Duncan MK (2008) Differential expression of the HMGN family of chromatin proteins during ocular development. Gene Expr Patterns 8:433–437. 10.1016/j.gep. 2008.04.002 [PubMed: 18502697]

Lyu J, Kim J-A, Chung SK, et al. (2003) Alteration of cadherin in dexamethasone-induced cataract organ-cultured rat lens. Invest Ophthalmol Vis Sci 44:2034–2040. 10.1167/iovs.02-0602 [PubMed: 12714641]

Mackay D, Ionides A, Kibar Z, et al. (1999) Connexin46 mutations in autosomal dominant congenital cataract. Am J Hum Genet 64:1357–1364. 10.1086/302383 [PubMed: 10205266]

Mackay DS, Boskovska OB, Knopf HLS, et al. (2002) A nonsense mutation in CRYBB1 associated with autosomal dominant cataract linked to human chromosome 22q. Am J Hum Genet 71:1216– 1221. 10.1086/344212 [PubMed: 12360425]

Madhira R The Effects of Parsimony Logic and Extended Parsimony Clustering on Protein Identification and Quantification in Shotgun Proteomics. 68

Manthey AL, Lachke SA, FitzGerald PG, et al. (2014) Loss of Sip1 leads to migration defects and retention of ectodermal markers during lens development. Mech Dev 131:86–110. 10.1016/j.mod. 2013.09.005 [PubMed: 24161570]

Medvedovic M, Tomlinson CR, Call MK, et al. (2006) Gene expression and discovery during lens regeneration in mouse: regulation of epithelial to mesenchymal transition and lens differentiation. Mol Vis 12:422–440 [PubMed: 16710166]

Miki A, Sakurada Y, Tanaka K, et al. (2018) Genome-Wide Association Study to Identify a New Susceptibility Locus for Central Serous Chorioretinopathy in the Japanese Population. Invest Ophthalmol Vis Sci 59:5542–5547. 10.1167/iovs.18-25497 [PubMed: 30480742]

Moré MI, Kirsch FP, Rathjen FG (2001) Targeted ablation of NrCAM or ankyrin-B results in disorganized lens fibers leading to cataract formation. J Cell Biol 154:187–196. 10.1083/jcb. 200104038 [PubMed: 11449000]

Mukaka MM (2012) Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J 24:69–71 [PubMed: 23638278]

Müller M, Bhattacharya SS, Moore T, et al. (2009) Dominant cataract formation in association with a vimentin assembly disrupting mutation. Hum Mol Genet 18:1052–1057. 10.1093/hmg/ddn440 [PubMed: 19126778]

Nakazawa M, Matsubara H, Matsushita Y, et al. (2016) The Human Bcl-2 Family Member Bcl-rambo Localizes to Mitochondria and Induces Apoptosis and Morphological Aberrations in Drosophila. PLoS ONE 11:e0157823 10.1371/journal.pone.0157823 [PubMed: 27348811]

Nelson N, Harvey WR (1999) Vacuolar and plasma membrane proton-adenosinetriphosphatases. Physiol Rev 79:361–385. 10.1152/physrev.1999.79.2.361 [PubMed: 10221984]

Nesvizhskii AI, Aebersold R (2005) Interpretation of shotgun proteomic data: the protein inference problem. Mol Cell Proteomics 4:1419–1440. 10.1074/mcp.R500012-MCP200 [PubMed: 16009968]

Ng D (1993) Lenz Microphthalmia Syndrome In: Adam MP, Ardinger HH, Pagon RA, et al. (eds) GeneReviews®. University of Washington, Seattle, Seattle (WA)

Oláhová M, Yoon WH, Thompson K, et al. (2018) Biallelic Mutations in ATP5F1D, which Encodes a Subunit of ATP Synthase, Cause a Metabolic Disorder. Am J Hum Genet 102:494–504. 10.1016/ j.ajhg.2018.01.020 [PubMed: 29478781]

Padma T, Ayyagari R, Murty JS, et al. (1995) Autosomal dominant zonular cataract with sutural opacities localized to chromosome 17q11–12. Am J Hum Genet 57:840–845 [PubMed: 7573044]

Patel N, Anand D, Monies D, et al. (2017) Novel phenotypes and loci identified through clinical genomics approaches to pediatric cataract. Hum Genet 136:205–225. 10.1007/s00439-016-1747-6 [PubMed: 27878435]

Prieto D, Zolessi FR (2017) Functional Diversification of the Four MARCKS Family Members in Zebrafish Neural Development. J Exp Zool B Mol Dev Evol 328:119–138. 10.1002/jez.b.22691 [PubMed: 27554589]

Puk O, Ahmad N, Wagner S, et al. (2011) First mutation in the βA2-crystallin encoding gene is associated with small lenses and age-related cataracts. Invest Ophthalmol Vis Sci 52:2571–2576. 10.1167/iovs.10-6443 [PubMed: 21212184]

Ramachandran RD, Perumalsamy V, Hejtmancik JF (2007) Autosomal recessive juvenile onset cataract associated with mutation in BFSP1. Hum Genet 121:475–482. 10.1007/s00439-006-0319-6 [PubMed: 17225135]

Riazuddin SA, Yasmeen A, Yao W, et al. (2005) Mutations in betaB3-crystallin associated with autosomal recessive cataract in two Pakistani families. Invest Ophthalmol Vis Sci 46:2100–2106. 10.1167/iovs.04-1481 [PubMed: 15914629]

Riba A, Di Nanni N, Mittal N, et al. (2019) Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. Proc Natl Acad Sci USA 116:15023–15032. 10.1073/pnas.1817299116 [PubMed: 31292258]

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. 10.1093/bioinformatics/btp616 [PubMed: 19910308]

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25 10.1186/gb-2010-11-3-r25 [PubMed: 20196867]

Rong X, Qiu X, Jiang Y, et al. (2016) Effects of histone acetylation on superoxide dismutase 1 gene expression in the pathogenesis of senile cataract. Sci Rep 6:34704 10.1038/srep34704 [PubMed: 27703255]

Santhiya ST, Shyam Manohar M, Rawlley D, et al. (2002) Novel mutations in the gamma-crystallin genes cause autosomal dominant congenital cataracts. J Med Genet 39:352–358. 10.1136/jmg.39.5.352 [PubMed: 12011157]

Shiels A, Mackay D, Ionides A, et al. (1998) A missense mutation in the human connexin50 gene (GJA8) underlies autosomal dominant "zonular pulverulent" cataract, on chromosome 1q. Am J Hum Genet 62:526–532. 10.1086/301762 [PubMed: 9497259]

Siddam AD, Gautier-Courteille C, Perez-Campos L, et al. (2018) The RNA-binding protein Celf1 post-transcriptionally regulates p27Kip1 and Dnase2b to control fiber cell nuclear degradation in lens development. PLoS Genet 14:e1007278 10.1371/journal.pgen.1007278 [PubMed: 29565969]

Srivastava R, Budak G, Dash S, et al. (2017) Transcriptome analysis of developing lens reveals abundance of novel transcripts and extensive splicing alterations. Sci Rep 7:11572 10.1038/s41598-017-10615-4 [PubMed: 28912564]

Stephan DA, Gillanders E, Vanderveen D, et al. (1999) Progressive juvenile-onset punctate cataracts caused by mutation of the gammaD-crystallin gene. Proc Natl Acad Sci USA 96:1008–1012. 10.1073/pnas.96.3.1008 [PubMed: 9927684]

Sun H, Ma Z, Li Y, et al. (2005) Gamma-S crystallin gene (CRYGS) mutation causes dominant progressive cortical cataract in humans. J Med Genet 42:706–710. 10.1136/jmg.2004.028274 [PubMed: 16141006]

Swarup A, Bell BA, Du J, et al. (2018) Deletion of GLUT1 in mouse lens epithelium leads to cataract formation. Exp Eye Res 172:45–53. 10.1016/j.exer.2018.03.021 [PubMed: 29604281]

Tawk M, Titeux M, Fallet C, et al. (2003) Synemin expression in developing normal and pathological human retina and lens. Exp Neurol 183:499–507. 10.1016/s0014-4886(03)00240-1 [PubMed: 14552890]

Teo ZL, McQueen-Miscamble L, Turner K, et al. (2014) Integrin linked kinase (ILK) is required for lens epithelial cell survival, proliferation and differentiation. Exp Eye Res 121:130–142. 10.1016/j.exer.2014.01.013 [PubMed: 24472646]

Tian R, Xu Y, Dou W-W, Zhang H (2018) Bioinformatics analysis of microarray data to explore the key genes involved in HSF4 mutation-induced cataract. Int J Ophthalmol 11:910–917. 10.18240/ijo.2018.06.03 [PubMed: 29977800]

Vorontsova I, Lam L, Delpire E, et al. (2014) Identification of the WNK-SPAK/OSR1 signaling pathway in rodent and human lenses. Invest Ophthalmol Vis Sci 56:310–321. 10.1167/iovs.14-15911 [PubMed: 25515571]

Wang B, Hom G, Zhou S, et al. (2017a) The oxidized thiol proteome in aging and cataractous mouse and human lens revealed by ICAT labeling. Aging Cell 16:244–261. 10.1111/acel.12548 [PubMed: 28177569]

Wang X, Garcia CM, Shui Y-B, Beebe DC (2004) Expression and regulation of alpha-, beta-, and gamma-crystallins in mammalian lens epithelial cells. Invest Ophthalmol Vis Sci 45:3608–3619. 10.1167/iovs.04-0423 [PubMed: 15452068]

Wang Y, Terrell AM, Riggio BA, et al. (2017b) β1-Integrin Deletion From the Lens Activates Cellular Stress Responses Leading to Apoptosis and Fibrosis. Invest Ophthalmol Vis Sci 58:3896–3922. 10.1167/iovs.17-21721 [PubMed: 28763805]

Wang Z, Han J, David LL, Schey KL (2013) Proteomics and phosphoproteomics analysis of human lens fiber cell membranes. Invest Ophthalmol Vis Sci 54:1135–1143. 10.1167/iovs.12-11168 [PubMed: 23349431]

Whitson JA, Zhang X, Medvedovic M, et al. (2017) Transcriptome of the GSH-Depleted Lens Reveals Changes in Detoxification and EMT Signaling Genes, Transport Systems, and Lipid Homeostasis. Invest Ophthalmol Vis Sci 58:2666–2684. 10.1167/iovs.16-21398 [PubMed: 28525556]

Wigle JT, Chowdhury K, Gruss P, Oliver G (1999) Prox1 function is crucial for mouse lens-fibre elongation. Nat Genet 21:318–322. 10.1038/6844 [PubMed: 10080188]

Wilmarth PA, Riviere MA, David LL (2009) Techniques for accurate protein identification in shotgun proteomic studies of human, mouse, bovine, and chicken lenses. J Ocul Biol Dis Infor 2:223–234. 10.1007/s12177-009-9042-6 [PubMed: 20157357]

Wolf L, Harrison W, Huang J, et al. (2013) Histone posttranslational modifications and cell fate determination: lens induction requires the lysine acetyltransferases CBP and p300. Nucleic Acids Res 41:10199–10214. 10.1093/nar/gkt824 [PubMed: 24038357]

Zhao Y, Wilmarth PA, Cheng C, et al. (2019) Proteome-transcriptome analysis and proteome remodeling in mouse lens epithelium and fibers. Exp Eye Res 179:32–46. 10.1016/j.exer.2018.10.011 [PubMed: 30359574]

Zhou Y, Bennett TM, Shiels A (2016) Lens ER-stress response during cataract development in Mip-mutant mice. Biochim Biophys Acta 1862:1433–1442. 10.1016/j.bbadis.2016.05.003 [PubMed: 27155571]
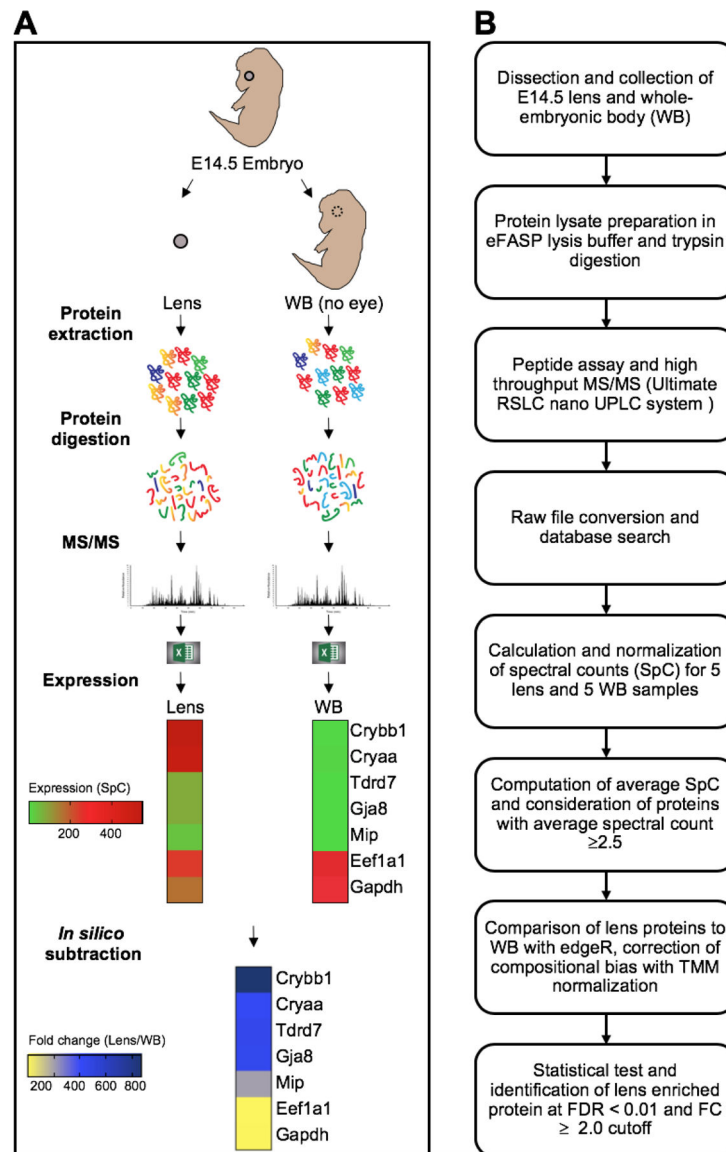
Fig. 1. Generation of mouse E14.5 lens and whole embryonic body (WB) proteome to identify lens enriched genes.

(A) Mouse E14.5 lens and WB (eye tissue removed) were micro-dissected and 55 μg protein of each sample of lens and WB (*n*=5 samples for each of lens and WB) was processed using high-throughput mass spectrometry. Differential protein expression was analyzed by comparing with whole embryonic body (WB) reference proteome. (B) Flow chart showing the pipeline that was followed for differential protein expression analysis. Normalized spectral counts were subjected to differential protein expression using edgeR pipeline. Proteins with false discovery rate (FDR) < 0.01 and fold change (FC) 2 were considered lens enriched.
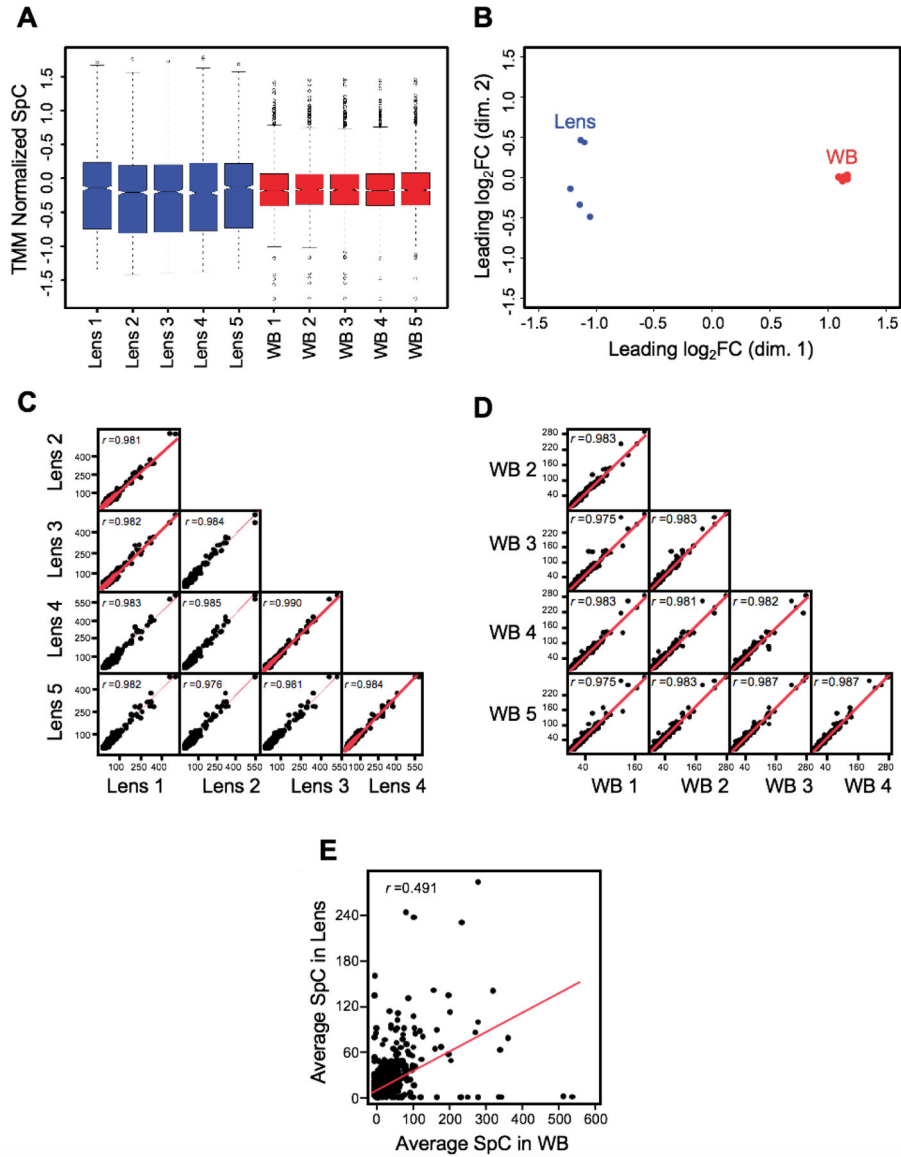
Fig. 2. Quality assessment of MS/MS data of lens and WB.

(A) Multidimensional scaling analysis showed that individual biological replicates of the lens and WB samples clustered together while the overall lens and WB samples clustered separately from each other. The axes show the leading dimensions 1 and 2. (B) Spectral counts in WB and lens samples were subjected to TMM (trimmed mean of M-values) normalization using edgeR to correct for the dramatic compositional differences. The boxplots for the normalized SpC datasets showed comparable median SpCs between the lens and the WB samples. The y-axis represents the TMM normalized SpC. (C) A scatter matrix was generated for five lens samples and correlation was examined for the sample to sample consistency. The lens samples showed a high correlation. (D) A scatter matrix was generated for five WB samples and correlation was examined for the sample to sample consistency. The WB samples showed a high correlation. (E) A scatter plot with regression analysis shows no correlation $(r = 0.4919)$ between the average lens and average WB samples.
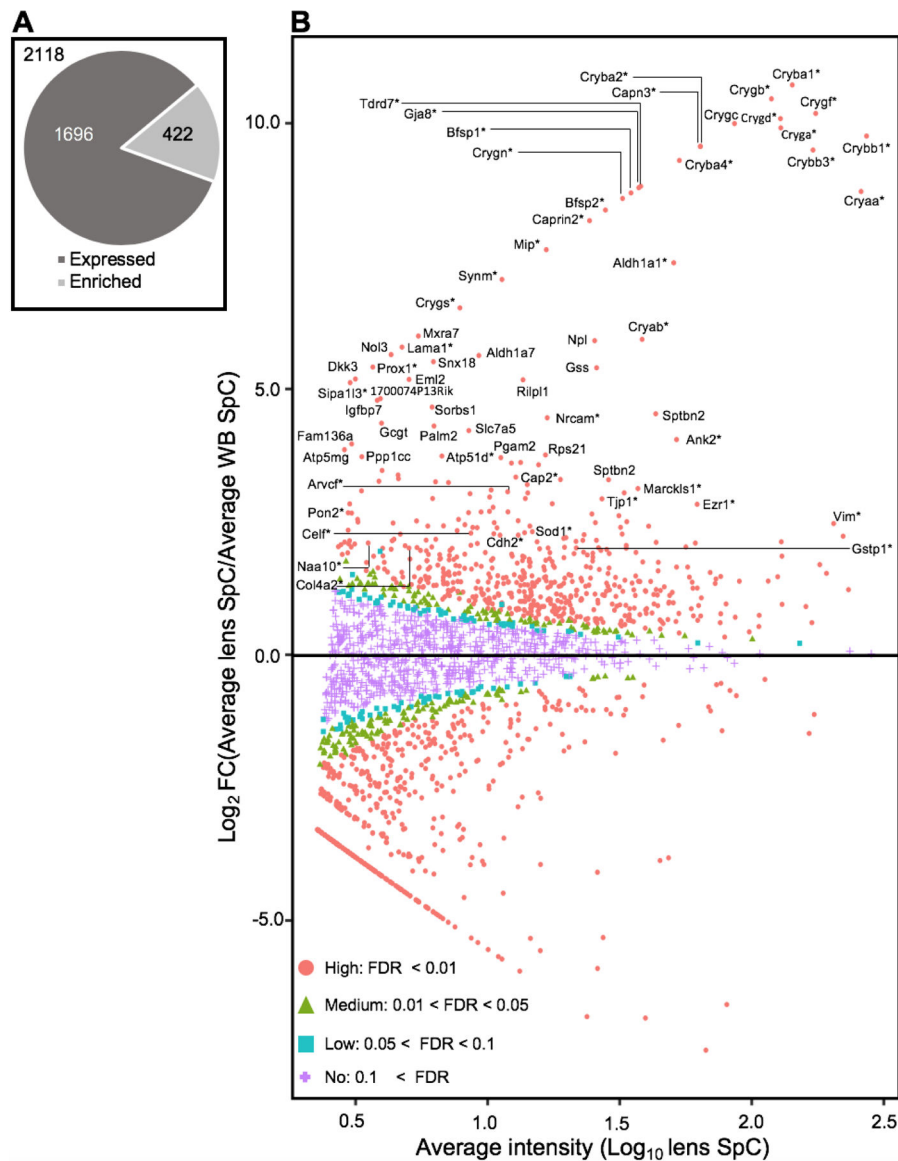
Fig. 3. *In silico* subtraction based identification of lens enriched proteins.
(A) Proteins that passed the average SpC 2.5 between lens and WB samples (*n*=2118)
were considered for differential expression analysis. At 2.0-fold-enrichment and FDR
<0.01 cut-off, 422 proteins were found to have enriched expression in the lens compared to
WB. (B) Differential protein expression profiling of 2118 proteins shown as MA plot (M
represents log ratio of lens to whole body and A represents the average intensity) identified
several lens enriched genes. Among the 422 lens enriched genes include many candidates
that are associated with cataract (denoted with *). Comparison of lens proteome with WB
shows high- (FDR < 0.01, coded red, circle), medium- (0.01 < FDR < 0.05, coded green,
triangle), low-probability lens enriched (0.05 < FDR < 0.1, coded blue, square) and non-
enriched genes (0.1 < FDR, coded magenta, cross). The top 150 lens enriched genes and
cataract associated candidates are indicated.

A

Rank	Top 30 Lens Expressed	Top 30 Lens Enriched
1	Crybb1	Cryba1
2	Cryaa	Crygb
3	Sptan1	Crygf
4	Crygf	Crygd
5	Vim	Crygc
6	Crybb3	Cryga
7	Hsp90ab1	Crybb1
8	Hspa8	Cryba2
9	Cryba1	Capn3
10	Eno1	Crybb3
11	Cryga	Cryba4
12	Crygd	Tdrd7
13	Crygb	Gja8
14	Plec	Cryaa
15	Flna	Bfsp1
16	Sptbn1	Crygn
17	Eef2	Bfsp2
18	Pkm	Caprin2
19	Crygc	Mip
20	Hspd1	Aldh1a1
21	Vcp	Synm
22	Hsp90aa1	Crygs
23	Hspa5	Mxra7
24	Cryba2	Cryab
25	Capn3	Npl
26	Fasn	Lama1
27	Myh9	Nol3
28	Ywhae	Aldh1a7
29	Ezr	Snx18
30	Hbb-bs	Prox1

B

Fig. 4. *In silico* **subtraction effectively prioritizes cataract associated genes.**
(A) Comparison of top 30 lens enriched proteins with the top 30 lens "expressed" proteins shows that *in silico* WB-subtraction is effective in identifying the Crystallins as well as several other cataract associated factors such as Aldh1a1, Bfsp1, Bfsp2, Caprin2, Cryab, Crygs, Gja8, Lama1, Mip, Prox1 and Tdrd7 (grey). (B) Top cataract associated proteins that were not present in the top 30 "expressed" candidates show significant (p<0.001) enrichment in the lens compared to WB. This shows that the *in silico* WB-subtraction strategy is effective in identifying these important non-crystallin lens proteins. The y-axis represents the average SpC individual proteins.
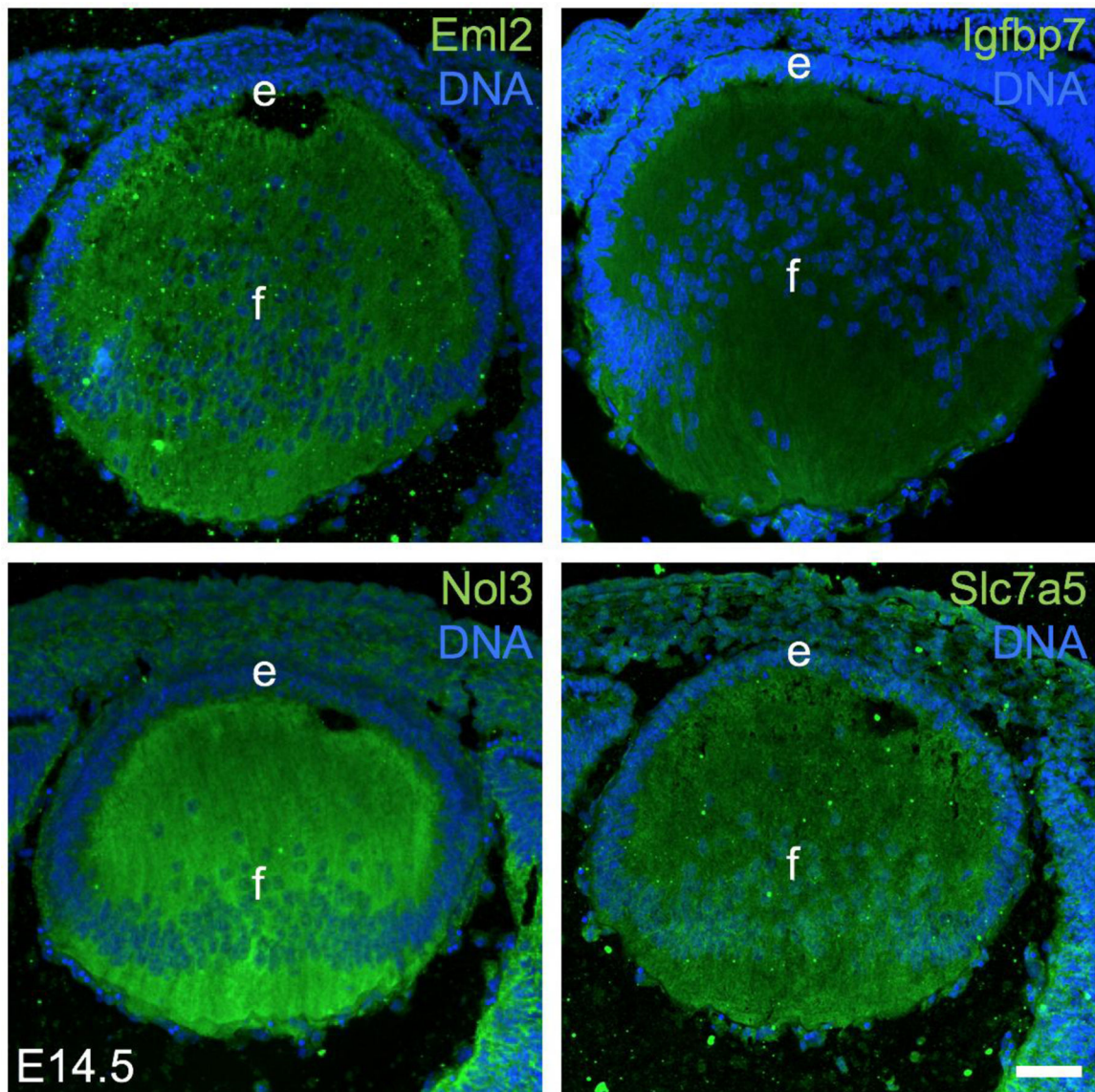
Fig. 5. Immunostaining-based validation of candidate proteins expressed in the lens.
Immunofluorescence using rabbit primary antibodies demonstrates robust expression of select MS/MS predicted proteins in the mouse embryonic day E14.5 lens. The proteins confirmed to be expressed (green) in the lens were Eml2 (Echinoderm microtubule associated protein like 2), Igfbp7 (Insulin-like growth factor binding protein 7), Nol3 (Nucleolar protein 3) and Slc7a5 (Solute carrier family 7 member 5). DNA was stained by DAPI. Lens epithelium (e) and fiber cells (f) are indicated. Scale bar represents 50 μm.
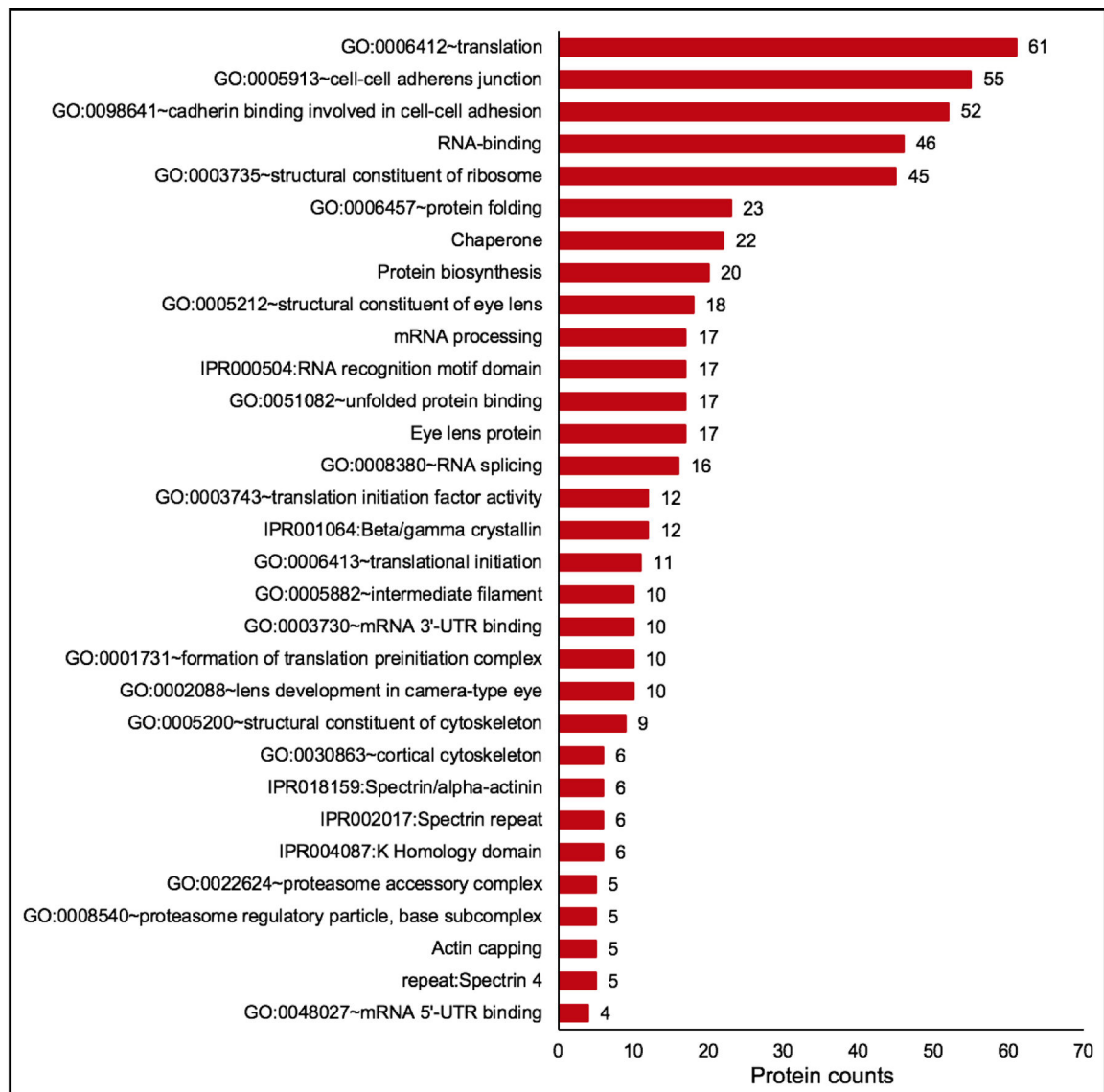
Fig. 6. Gene ontology (GO) analysis of lens enriched proteins identifies candidates relevant to function in the lens.

Lens enriched proteins were subjected to cluster-based analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6 .8) for functional annotation by gene ontology (GO) categories. This analysis assigned 406 lens enriched proteins into 68 annotation clusters. The top clusters included GO categories that are relevant to lens biology and cataract. These were "eye lens protein", "protein folding", "Ribonucleoprotein", "Protein biosynthesis", and "cell-cell adherens junction", among others. Within the cluster for "eye lens protein", other lens-relevant sub-categories were identified such as "structural constituent of eye lens", "Beta/Gamma crystallin", "lens development in camera-type eye", "eye development" and "visual perception". The number on the top of the bar graph shows the number of identified proteins in the category.
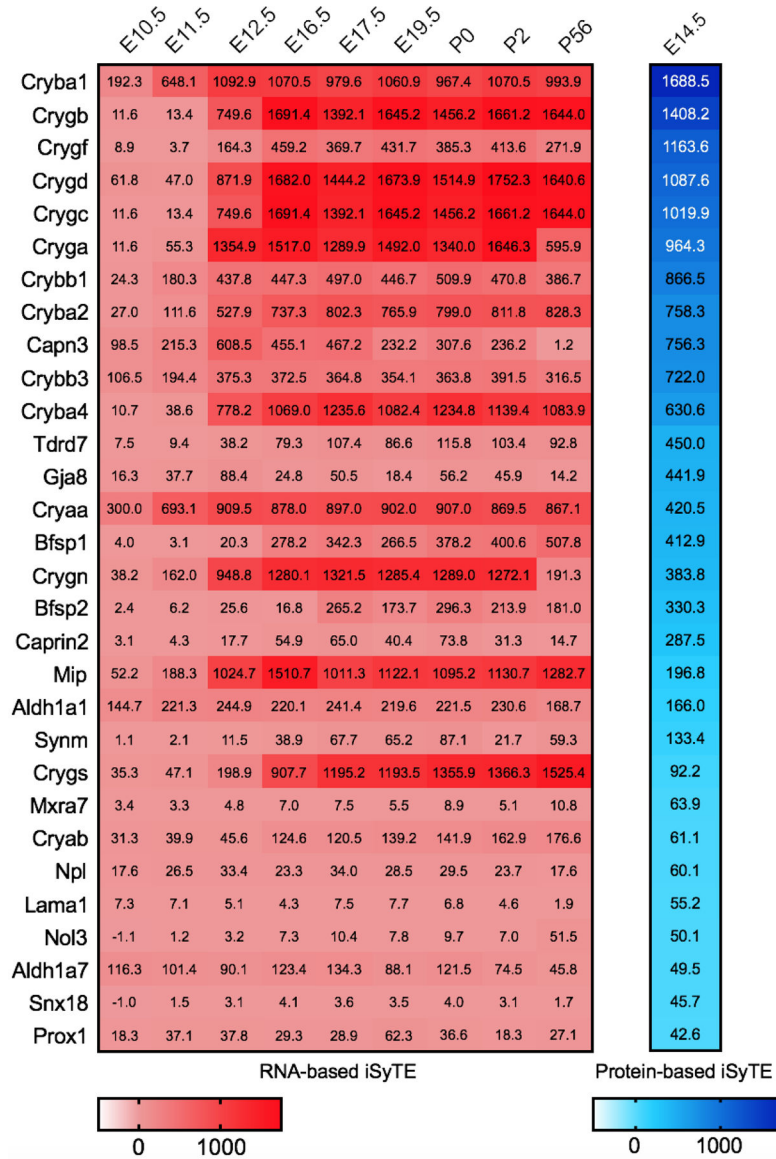
Fig. 7. Comparative analysis of top lens enriched protein candidates in iSyTE microarray datasets.

Comparison of the top 30 candidates identified by *in silico* WB-subtraction to microarray-based mRNA expression data in iSyTE. All the top 30 lens enriched proteins are also enriched in the lens on the RNA level at E14.5 and majority were enriched in one or more of the mouse embryonic stages ranging from E10.5 through P56. The numbers in the table represent the enrichment in fold-change compared to WB and the extent of enrichment is indicated by heat-map for RNA (pink) and protein (blue).
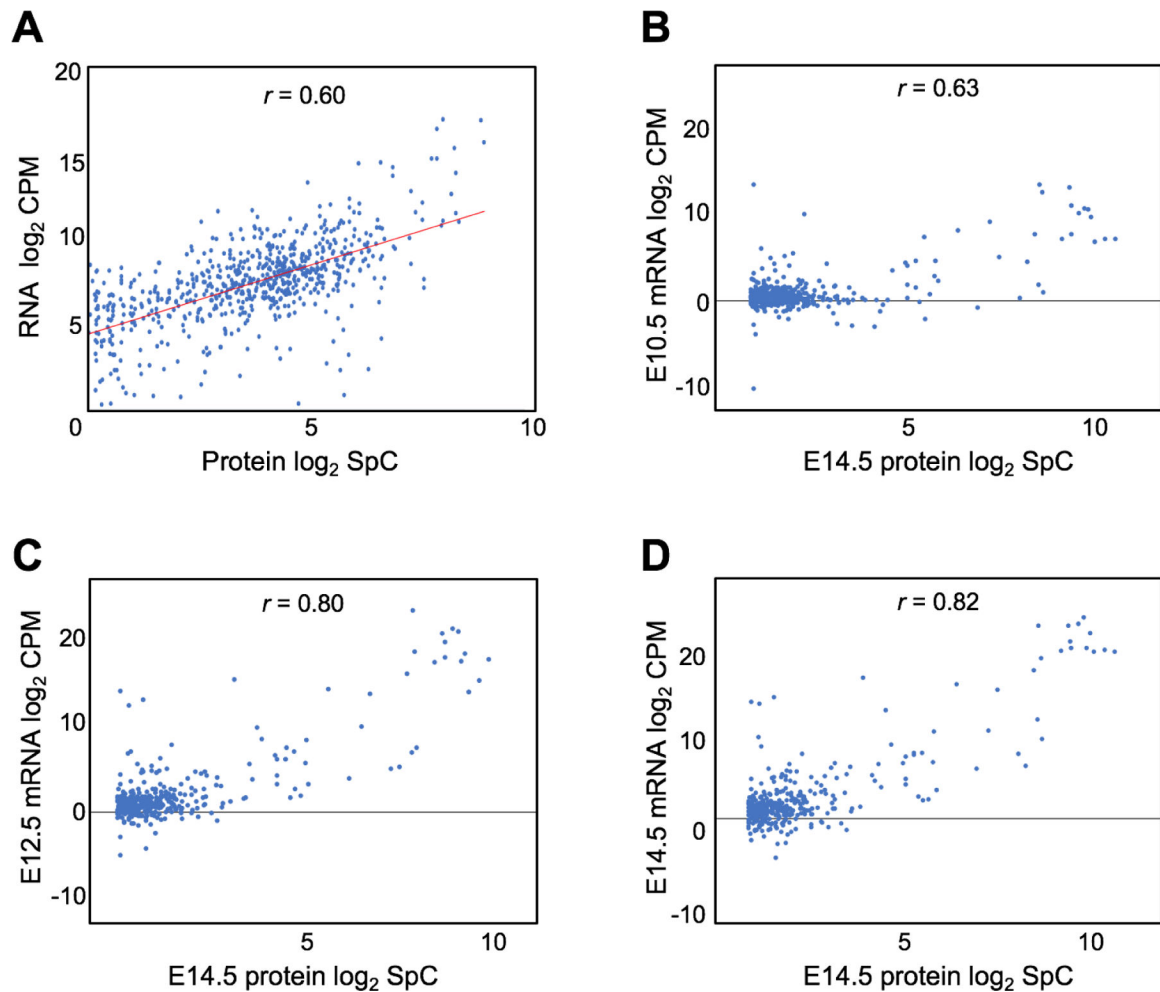
Fig. 8. Comparison of mouse lens proteome with the transcriptome.
(A) Correlation analysis of 1417 genes common to mouse E14.5 lens proteome and mouse E14.5 lens RNA-seq data with significant expression cutoff of spectral count 2.0 (for protein data) and 2.0 counts-per-million (for RNA data). The mRNA-protein correlation was analyzed with Pearson's correlation coefficient (r) which indicated an overall positive correlation between the transcriptome and the proteome ($r = 0.6$). The x-axis represents log_2 protein SpC and the y-axis represents the log_2 mRNA counts. (B-D) Correlation between 422 lens-enriched proteins and their corresponding mRNA at E10.5, E12.5 and E14.5 as analyzed by Pearson's correlation coefficient method. The mRNA-protein correlation is higher in lens enriched proteins compared to lens expressed proteins and increased with progressive development of the lens through stages E10.5 ($r = 0.63$), E12.5 ($r = 0.80$) and E14.5 ($r = 0.82$).
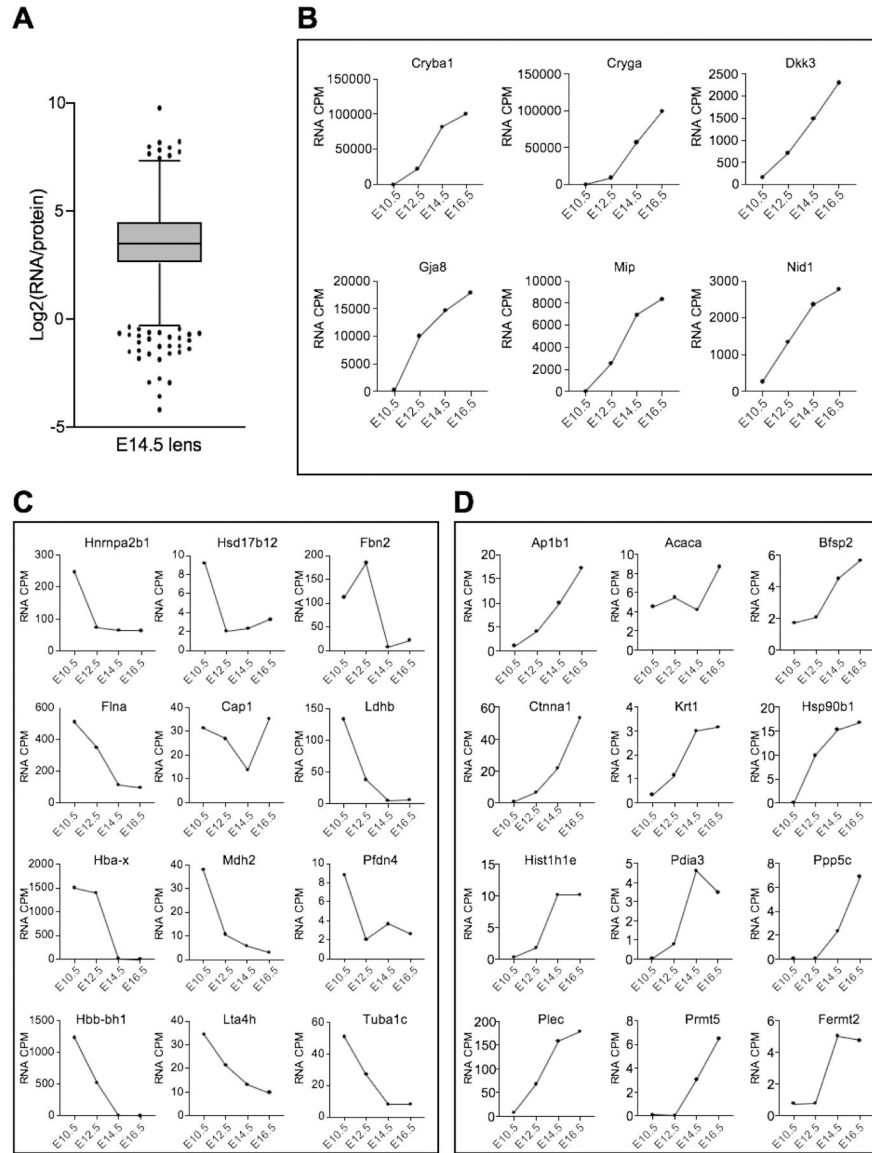
Fig. 9. Candidate genes that exhibit dynamic relationship between mRNA and protein in the lens.

(A) Box plot of the $\log_2$ values of the ratio between RNA (CPM) and protein (SpC) expression for individual genes (n =1417) (y-axis) shows their distribution. The box indicates the interquartile range (IQR) for the $\log_2$ values and the whiskers represent ±1.5 times IQR. The values outside the upper and lower limit of the whiskers are designated as "outliers" (relatively high mRNA levels compared to protein and *vice versa*). (B) Examination of RNA expression trend of the outliers with high $\log_2$(RNA/protein ratio, *i.e.* relatively high RNA compared to protein) showed a sharp increase in mRNA level at stage E14.5 compared to the earlier embryonic stages E10.5 and E12.5. Examination of candidates with low $\log_2$(RNA/protein ratio; *i.e.* relatively low RNA compared to protein) showed a dynamic change at the RNA level at stage E14.5, which manifested as a sharp decrease (C) or increase (D) in mRNA expression compared to preceding stages. The x-axis represents the developmental stages and the y-axis represents the mRNA counts.
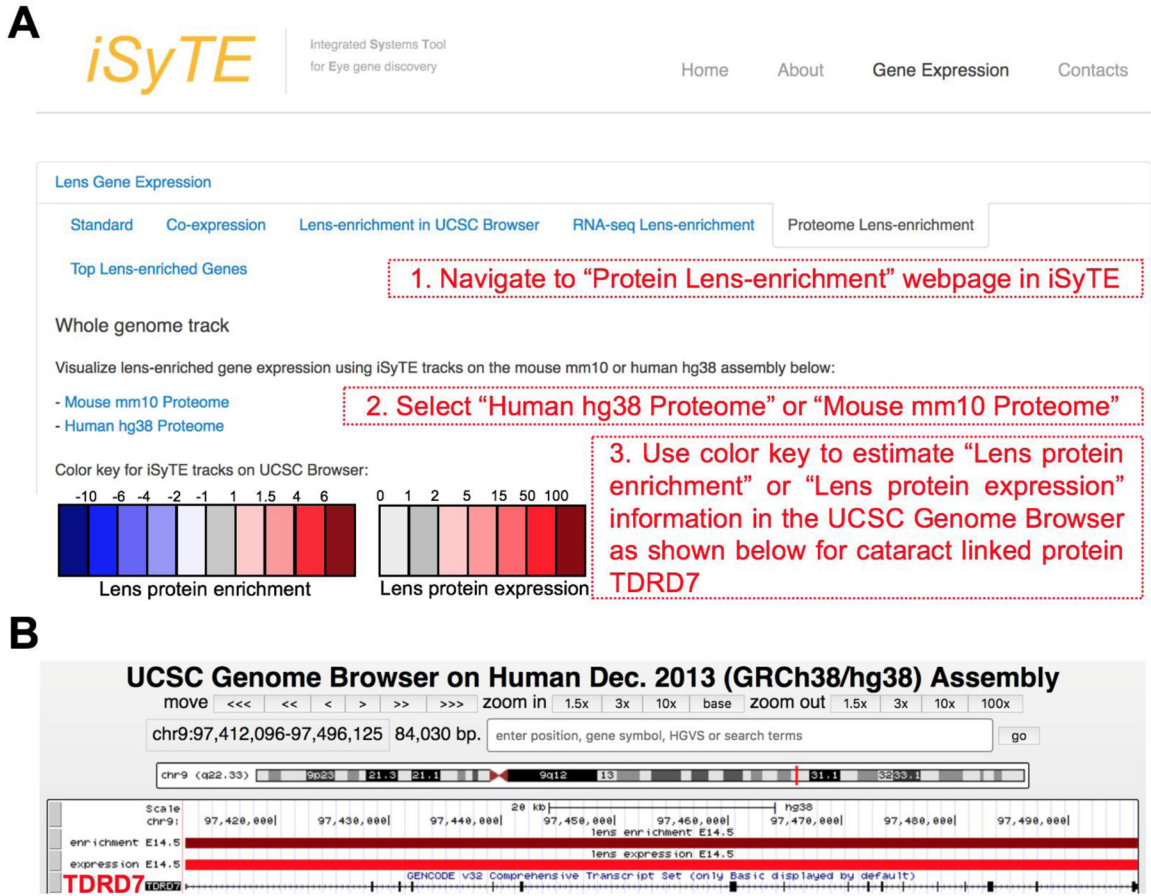
Fig 10. iSyTE based access to UCSC browser custom tracks for visualizing lens proteome data. (A) The lens proteome data is made freely available to the public through custom tracks that can be accessed via the iSyTE 2.0 web-based resource https:// research.bioinformatics.udel.edu/iSyTE/. Under the tab "Mouse lens Proteome", specific links termed "Hg38 proteome" and "mm10 proteome" allow access to custom annotation tracks on the University of California at Santa Cruz (UCSC) Genome Browser for human and mouse, respectively. These tracks allow visualization of expression and enrichment of proteins in the E14.5 lens. The heat maps indicate the extent of lens protein expression and enrichment. (B) As an example of the utility of this resource, visualization of the lens enrichment and expression values for the human cataract-linked protein TDRD7 are indicated.

Table 1.

Total spectral counts for the lens and WB samples.

Sample	Protein (µg)	Total SpC
Lens1	55	19.3K
Lens2	55	22.4K
Lens3	55	21.7K
Lens4	55	21.1K
Lens5	55	18.8K
WB1	55	41.2K
WB2	55	40.3K
WB3	55	40.6K
WB4	55	40.9K
WB5	55	41.1K

Table 2.

Literature analysis of top 150 lens-enriched proteins with regards to eye/lens defects

Rank	UniProt Gene Name	Uniprot Accession	Primary Protein Name	Associated lens or Eye defect	Reference
1	Cryba1	P02525	Beta-crystallin A1	Cataract[*]	(Padma et al. 1995)
2	Crygb	P04344	Gamma-crystallin B	Cataract[*]	(AlFadhli et al. 2012)
3	Crygf	Q9CXV3	Gamma-crystallin F	Cataract[*]	(Graw et al. 2002)
4	Crygd	P04342	Gamma-crystallin D	Cataract[*]	(Stephan et al. 1999)
5	Crygc	Q61597	Gamma-crystallin C	Cataract[*]	(Gonzalez-Huerta et al. 2007)
6	Cryga	P04345	Gamma-crystallin A	Cataract[*]	(Santhiya et al. 2002)
7	Crybb1	Q9WVJ5	Beta-crystallin B1B	Cataract, microcornea[*]	(Mackay et al. 2002)
8	Cryba2	Q9JJV1	Beta-crystallin A2	Cataract[*]	(Puk et al. 2011)
9	Capn3	Q64691	Calpain-3	None found	
10	Crybb3	Q9JJU9	Beta-crystallin B3, N-terminally processed	Cataract[*]	(Riazuddin et al. 2005)
11	Cryba4	Q9JJV0	Beta-crystallin A4	Cataract and microcornea[*]	(Billingsley et al. 2006)
12	Tdrd7	Q8K1H1	Tudor domain-containing protein 7	Cataract[*]	(Lachke et al. 2011)
13	Gja8	P28236	Gap junction alpha-8 protein	Cataract[*]	(Shiels et al. 1998)
14	Cryaa	P24622	Alpha-crystallin A chain	Cataract and micropthalmia[*]	(Litt et al. 1998)
15	Bfsp1	A2AMT1	Filensin	Cataract[*]	(Ramachandran et al. 2007)
16	Crygn	Q8VHL5	Gamma-crystallin N	None found	
17	Bfsp2	Q6NVD9	Phakinin	Cataract[*]	(Jakobs et al. 2000)
18	Caprin2	Q05A80	Caprin-2	Peters anomaly[*]	(Dash et al. 2015)
19	Mip	P51180	Lens fiber major intrinsic protein	Cataract[*]	(Berry et al. 2000)
20	Aldh1a1	P24549	Retinal dehydrogenase 1 (RALDH 1; RalDH1)	Cataract[*]	(Lassen et al. 2007)
21	Synm	Q70IV5	Synemin	Cataract in association with Meckel syndrome in human[*]	(Tawk et al. 2003)
22	Crygs	O35486	Gamma-crystallin S	Cataract[*]	(Sun et al. 2005)
23	Mxra7	Q9CZH7	Matrix-remodeling-associated protein 7	None found	
24	Cryab	P23927	Alpha-crystallin B chain	Cataract[*]	(Berry et al. 2001)
25	Npl	Q9DCJ9	N-acetylneuraminate lyase (NALase)	Not found	
26	Lama1	P19137	Laminin subunit alpha 1	lens morphogenesis and eye development[*]	(Dong and Chung 1991)
27	Nol3	Q9D1X0	Nucleolar protein 3	None found	
28	Aldh1a7	O35945	Aldehyde dehydrogenase, cytosolic 1	None found	

Rank	UniProt Gene Name	Uniprot Accession	Primary Protein Name	Associated lens or Eye defect	Reference
29	Snx18	Q91ZR2	Sorting nexin-18	None found	
30	Prox1	P48437	Prospero homeobox protein 1	lens fiber elongation[*]	(Wigle et al. 1999)
31	Gss	P51855	Glutathione synthetase (GSH synthetase; GSH-S)	None found	
32	Dkk3	Q9QUN9	Dickkopf-related protein 3 (Dickkopf-3; Dkk-3; mDkk-3)	None found	
33	Eml2	Q7TNG5	Echinoderm microtubule-associated protein-like 2 (EMAP-2)	IMPC[*]	
34	Rilpl1	Q9JJC6	RILP-like protein 1	None found	
35	Sipa1l3	G3X9J0	Signal-induced proliferation-associated 1-like protein 3	Cataract[*]	(Greenlees et al. 2015)
36	1700074P13Rik	Q9D9G7	1700074P13Rik protein	None found	
37	Igfbp7	Q61581	Insulin-like growth factor-binding protein 7	None found	
38	Sorbs1	Q62417	Sorbin and SH3 domain-containing protein 1	None found	
39	Sptbn2	Q68FG2	Spectrin beta chain	None found	
40	Nrcam	Q810U4	Neuronal cell adhesion molecule (Nr-CAM)	Cataract[*]	(Moré et al. 2001)
41	Ggct	Q9D7X8	Gamma-glutamylcyclotransferase	None found	
42	Palm2	Q8BR92	Paralemmin-2	Expression study in lens	(Castellini et al. 2005)
43	Slc7a5	Q9Z127	Large neutral amino acids transporter small subunit 1	None found	
44	Ank2	Q8C8R3	Ankyrin-2 (ANK-2)	Cataract[*]	(Moré et al. 2001)
45	Fam136a	Q9CR98	Protein FAM136A	None found	
46	Rps21	Q9CQR2	40S ribosomal protein S21	None found	
47	Atp5f1d	Q9D3D9	ATP synthase subunit delta, mitochondrial	Eye development defect[*]	(Oláhová et al. 2018)
48	Krt76	Q3UV17	Keratin, type II cytoskeletal 2 oral	None found	
49	Pgam2	O70250	Phosphoglycerate mutase 2	None found	
50	Dst	Q91ZU6	Dystonin	None found	
51	Hmga2	P52927	High mobility group protein HMGI-C	None found	
52	Cadm1	Q8R5M8	Cell adhesion molecule 1	None found	
53	Ppp1cc	P63087	Serine/threonine-protein phosphatase PP1-gamma catalytic subunit	None found	
54	Lnpk Uln	Q7TQ95	Endoplasmic reticulum junction formation protein lunapark	None found	
55	Cap2	Q9CYT6	Adenylyl cyclase-associated protein 2 (CAP 2)	Microphthalmia[*]	(Field et al. 2015)
56	Wbp2	P97765	WW domain-binding protein 2 (WBP-2)	None found	
57	Cdv3	Q4VAA2	Protein CDV3	None found	
58	Pygm	Q9WUB3	Glycogen phosphorylase, muscle form	None found	
59	Nedd8	P29595	NEDD8	None found	

Rank	UniProt Gene Name	Uniprot Accession	Primary Protein Name	Associated lens or Eye defect	Reference
60	Slc2a1	P17809	Solute carrier family 2, facilitated glucose transporter member 1	None found	
61	Sf3b5	Q923D4	Splicing factor 3B subunit 5 (SF3b5)	None found	
62	Kif1a	P33173	Kinesin-like protein KIF1A	None found	
63	Marcksl1	P28667	MARCKS-related protein	Small eye [*]	(Prieto and Zolessi 2017)
64	Pgrmc2	Q80UU9	Membrane-associated progesterone receptor component 2	None found	
65	Ccdc115	Q8VE99	Coiled-coil domain-containing protein 115	None found	
66	Arvcf	P98203	Armadillo repeat protein deleted in velo-cardio-facial syndrome homolog	Small eye [*]	(Cho et al. 2011)
67	Nap1l4	Q78ZA7	Nucleosome assembly protein 1-like 4	None found	
68	Rpl13	P47963	60S ribosomal protein L13	None found	
69	Ass1	P16460	Argininosuccinate synthase	None found	
70	Tpm3	P21107	Tropomyosin alpha-3 chain	None found	
71	Bpnt1	Q9Z0S1	3'(2'),5'-bisphosphate nucleotidase 1	None found	
72	Sumo2	P61957	Small ubiquitin-related modifier 2 (SUMO-2)	None found	
73	Tjp1	P39447	Tight junction protein ZO-1	Cataract [*]	(Arora et al. 2012)
74	Ube2v1	Q9CZY3	Ubiquitin-conjugating enzyme E2 variant 1 (UEV-1)	None found	
75	Bola1	Q9D8S9	BolA-like protein 1	None found	
76	Ezr	P26040	Ezrin	Cataract [*]	(Lin et al. 2013)
77	Stk39	Q9Z1W9	STE20/SPS1-related proline-alanine-rich protein kinase	Cataract [*]	(Vorontsova et al. 2014)
78	Rps27	Q6ZWU9	40S ribosomal protein S27	None found	
79	Sarnp	Q9D1J3	SAP domain-containing ribonucleoprotein	IMPC [*]	
80	Srsf2	Q62093	Serine/arginine-rich splicing factor 2	None found	
81	Vamp3	P63024	Vesicle-associated membrane protein 3 (VAMP-3)	None found	
82	Tsc22d3	Q9Z2S7	TSC22 domain family protein 3	None found	
83	Pon2	Q62086	Serum paraoxonase/arylesterase 2 (PON 2)	Cataract [*]	(Bharathidevi et al. 2017)
84	Dbnl	Q62418	Drebrin-like protein	None found	
85	Ybx1	P62960	Nuclease-sensitive element-binding protein 1	None found	
86	Cox5b	P19536	Cytochrome c oxidase subunit 5B, mitochondrial	None found	
87	Ube2m	P61082	NEDD8-conjugating enzyme Ubc12	None found	
88	Jpt2	Q6PGH2	Jupiter microtubule associated homolog 2	None found	
89	Chchd3	Q9CRB9	MICOS complex subunit Mic19	None found	
90	Ywhah	P68510	14-3-3 protein eta	None found	

Rank	UniProt Gene Name	Uniprot Accession	Primary Protein Name	Associated lens or Eye defect	Reference
91	Cfap36	Q8C6E0	Cilia- and flagella-associated protein 36	None found	
92	Rps18	P62270	40S ribosomal protein S18	None found	
93	Sub1	P11031	Activated RNA polymerase II transcriptional coactivator p15	None found	
94	Tomm20	Q9DCC8	Mitochondrial import receptor subunit TOM20 homolog	None found	
95	Vim	P20152	Vimentin	Cataract[*]	(Müller et al. 2009)
96	Rps28	P62858	40S ribosomal protein S28	None found	
97	Pfdn4	Q3UWL8	Prefoldin subunit 4	None found	
98	Cpt2	P52825	Carnitine O-palmitoyltransferase 2, mitochondrial	None found	
99	Ndufa4	Q62425	Cytochrome c oxidase subunit NDUFA4	None found	
100	Basp1	Q91XV3	Brain acid soluble protein 1	None found	
101	Eif2s1	Q6ZWX6	Eukaryotic translation initiation factor 2 subunit 1	None found	
102	Gls	D3Z7P3	Glutaminase kidney isoform, mitochondrial (GLS)	None found	
103	Ube2v2	Q9D2M8	Ubiquitin-conjugating enzyme E2 variant 2	None found	
104	Rpl35a	O55142	60S ribosomal protein L35a	None found	
105	Chmp2a	Q9DB34	Charged multivesicular body protein 2a	None found	
106	Sssca1	P56873	Sjoegren syndrome/scleroderma autoantigen 1 homolog	None found	
107	Sod1	P08228	Superoxide dismutase [Cu-Zn]	Cataract[*]	(Rong et al. 2016)
108	Cttn	Q60598	Src substrate cortactin	None found	
109	Celf1	P28659	CUGBP Elav-like family member 1 (CELF-1)	Cataract[*]	(Siddam et al. 2018)
110	Bcl2l13	P59017	Bcl-2-like protein 13 (Bcl2-L-13)	None found	
111	Nsfl1c	Q9CZ44	NSFL1 cofactor p47	None found	
112	Cxadr	P97792	Coxsackievirus and adenovirus receptor homolog (CAR; mCAR)	None found	
113	Igf2bp1	O88477	Insulin-like growth factor 2 mRNA-binding protein 1	None found	
114	Rpl26	P61255	60S ribosomal protein L26	None found	
115	Cdh2	P15116	Cadherin-2	Cataract[*]	(Lyu et al. 2003)
116	Sptan1	P16546	Spectrin alpha chain, non-erythrocytic 1	None found	
117	Rps20	P60867	40S ribosomal protein S20	None found	
118	Fam49b	Q921M7	Protein FAM49B	None found	
119	Snrpf	P62307	Small nuclear ribonucleoprotein F (snRNP-F)	None found	
120	Krt72	Q6IME9	Keratin, type II cytoskeletal 72	None found	
121	Hmgn1	P18608	Non-histone chromosomal protein HMG-14	None found	

Rank	UniProt Gene Name	Uniprot Accession	Primary Protein Name	Associated lens or Eye defect	Reference
122	Rps14	P62264	40S ribosomal protein S14	None found	
123	Cotl1	Q9CQI6	Coactosin-like protein	IMPC[*]	
124	Adrm1	Q9JKV1	Proteasomal ubiquitin receptor ADRM1	None found	
125	Amph	Q7TQF7	Amphiphysin	None found	
126	Rpl22	P67984	60S ribosomal protein L22	None found	
127	Plec	Q9QXS1	Plectin (PCN; PLTN)	None found	
128	Ewsr1	Q61545	RNA-binding protein EWS	None found	
129	Mri1	Q9CQT1	Methylthioribose-1-phosphate isomerase (M1Pi; MTR-1-P isomerase)	None found	
130	Rps3a	P97351	40S ribosomal protein S3a	None found	
131	Naa10	Q9QY36	N-alpha-acetyltransferase 10	Lenz microphthalmia syndrome[*]	(Ng 1993)
132	Gsn	P13020	Gelsolin	None found	
133	Sgta	Q8BJU0	Small glutamine-rich tetratricopeptide repeat-containing protein alpha	None found	
134	Tfam	P40630	Transcription factor A, mitochondrial (mtTFA)	None found	
135	na	Q6PIU9	Uncharacterized protein FLJ45252 homolog	None found	
136	Atp6v1g1	Q9CR51	V-type proton ATPase subunit G 1 (V-ATPase subunit G 1)	None found	
137	Metap1	Q8BP48	Methionine aminopeptidase 1 (MAP 1; MetAP 1)	None found	
138	Fubp1	Q91WJ8	Far upstream element-binding protein 1 (FBP; FUSE-binding protein 1)	None found	
139	Nudt4	Q8R2U6	Diphosphoinositol polyphosphate phosphohydrolase 2 (DIPP-2)	None found	
140	Pepd	Q11136	Xaa-Pro dipeptidase (X-Pro dipeptidase)	IMPC[*]	
141	Rpl36a	P83882	60S ribosomal protein L36a	None found	
142	Rps3	P62908	40S ribosomal protein S3	None found	
143	Anxa1	P10107	Annexin A1	None found	
144	Timm13	P62075	Mitochondrial import inner membrane translocase subunit Tim13	None found	
145	Rps27a	P62983	40S ribosomal protein S27a	None found	
146	Gstp1	P19157	Glutathione S-transferase P 1 (Gst P1)	Cataract[*]	(Chen et al. 2017)
147	Cbx3	P23198	Chromobox protein homolog 3	None found	
148	Col4a2	P08122	Canstatin	Cataract[*]	(Ha et al. 2016)
149	Marcks	P26645	Myristoylated alanine-rich C-kinase substrate (MARCKS)	None found	
150	Mettl26	Q9DCS2	Methyltransferase-like 26	None found	

Candidates shaded in grey are exclusively detected in the top 150 lens-enriched list of proteins but not in the top 150 lens-expressed list of proteins.

[*] Asterisk denotes connection with eye expression, defect or availability of resource

Table 3.

Top 150 expressed proteins in the E14.5 mouse lens

SN	UniProt Gene Name	Primary Protein Name	Avg. Lens
1	Crybb1	Beta-crystallin B1B	543.4
2	Cryaa	Alpha-crystallin A chain	517.7
3	Sptan1	Spectrin alpha chain, non-erythrocytic 1	365.8
4	Crygf	Gamma-crystallin F	348.9
5	Vim	Vimentin	346.1
6	Crybb3	Beta-crystallin B3, N-terminally processed	340.8
7	Hsp90ab1	Heat shock protein HSP 90-beta	325.5
8	Hspa8	Heat shock cognate 71 kDa protein	285.2
9	Cryba1	Beta-crystallin A1	284.6
10	Eno1	Alpha-enolase	276.5
11	Cryga	Gamma-crystallin A	257.4
12	Crygd	Gamma-crystallin D	256.7
13	Crygb	Gamma-crystallin B	237.3
14	Plec	Plectin (PCN; PLTN)	210.5
15	Flna	Filamin-A (FLN-A)	209.0
16	Sptbn1	Spectrin beta chain, non-erythrocytic 1	203.8
17	Eef2	Elongation factor 2 (EF-2)	203.8
18	Pkm	Pyruvate kinase PKM	184.1
19	Crygc	Gamma-crystallin C	171.9
20	Hspd1	60 kDa heat shock protein, mitochondrial	171.6
21	Vcp	Transitional endoplasmic reticulum ATPase (TER ATPase)	167.0
22	Hsp90aa1	Heat shock protein HSP 90-alpha	133.5
23	Hspa5	Endoplasmic reticulum chaperone BiP	129.8
24	Cryba2	Beta-crystallin A2	127.8
25	Capn3	Calpain-3	127.5
26	Fasn	Oleoyl-[acyl-carrier-protein] hydrolase	125.8
27	Myh9	Myosin-9	112.3
28	Ywhae	14-3-3 protein epsilon (14-3-3E)	112.1
29	Ezr	Ezrin	108.9
30	Hbb-bs	Beta-globin	108.5
31	Hsp90b1	Endoplasmin	107.4
32	Lmnb1	Lamin-B1	106.7
33	Cryba4	Beta-crystallin A4	106.3
34	Tkt	Transketolase (TK)	105.5
35	Ipo5	Importin-5 (Imp5)	104.3
36	Hnrnpa2b1	Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/B1)	103.7
37	Aldh1a1	Retinal dehydrogenase 1 (RALDH 1; RalDH1)	100.7

SN	UniProt Gene Name	Primary Protein Name	Avg. Lens
38	Gsn	Gelsolin	99.2
39	Ank2	Ankyrin-2 (ANK-2)	97.9
40	Hnrnpk	Heterogeneous nuclear ribonucleoprotein K (hnRNPK)	96.2
41	Dync1h1	Cytoplasmic dynein 1 heavy chain 1	94.3
42	Atp5f1a	ATP synthase subunit alpha, mitochondrial	91.6
43	Rps3	40S ribosomal protein S3	90.4
44	Uba1	Ubiquitin-like modifier-activating enzyme 1	88.0
45	Pgk1	Phosphoglycerate kinase 1	87.6
46	Pdia3	Protein disulfide-isomerase A3	87.3
47	Afp	Alpha-fetoprotein	87.0
48	Ywhaz	14-3-3 protein zeta/delta	86.7
49	Aldoa	Fructose-bisphosphate aldolase A	85.7
50	Hspa4	Heat shock 70 kDa protein 4	85.2
51	Ncl	Nucleolin	84.3
52	Sptbn2	Spectrin beta chain	83.0
53	Tcp1	T-complex protein 1 subunit alpha (TCP-1-alpha)	82.4
54	P4hb	Protein disulfide-isomerase (PDI)	79.9
55	Atp2a2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 (SERCA2; SR Ca(2+)-ATPase 2)	78.7
56	Calr	Calreticulin	76.2
57	Hnrnpu	Heterogeneous nuclear ribonucleoprotein U (hnRNPU)	76.2
58	Tdrd7	Tudor domain-containing protein 7	75.8
59	Cryab	Alpha-crystallin B chain	75.6
60	Cct5	T-complex protein 1 subunit epsilon (TCP-1-epsilon)	75.0
61	Gja8	Gap junction alpha-8 protein	74.5
62	Rps3a	40S ribosomal protein S3a	70.6
63	Bfsp1	Filensin	69.6
64	Naca	Nascent polypeptide-associated complex subunit alpha, muscle-specific form	69.3
65	Basp1	Brain acid soluble protein 1	68.6
66	Khsrp	Far upstream element-binding protein 2 (FUSE-binding protein 2)	68.2
67	Cct3	T-complex protein 1 subunit gamma (TCP-1-gamma)	66.9
68	Vdac1	Voltage-dependent anion-selective channel protein 1 (VDAC-1; mVDAC1)	66.9
69	Cct6a	T-complex protein 1 subunit zeta (TCP-1-zeta)	66.9
70	Marcksl1	MARCKS-related protein	66.5
71	Pcbp1	Poly(rC)-binding protein 1	66.2
72	Cct8	T-complex protein 1 subunit theta (TCP-1-theta)	66.0
73	Rack1 Gnb2l1	Receptor of activated protein C kinase 1, N-terminally processed	64.7
74	Crygn	Gamma-crystallin N	64.7
75	Tuba1a	Detyrosinated tubulin alpha-1A chain	64.6
76	Ppia	Peptidyl-prolyl cis-trans isomerase A, N-terminally processed	64.2

SN	UniProt Gene Name	Primary Protein Name	Avg. Lens
77	Tuba1c	Detyrosinated tubulin alpha-1C chain	63.8
78	Hspa9	Stress-70 protein, mitochondrial	63.0
79	Cct7	T-complex protein 1 subunit eta (TCP-1-eta)	62.7
80	Hist1h4a	Histone H4	62.4
81	Ywhaq	14-3-3 protein theta	60.8
82	Gdi2	Rab GDP dissociation inhibitor beta (Rab GDI beta)	60.6
83	Pgam1	Phosphoglycerate mutase 1	60.2
84	Rplp2	60S acidic ribosomal protein P2	59.9
85	Pa2g4	Proliferation-associated protein 2G4	59.6
86	Tln1	Talin-1	59.4
87	Trim28	Transcription intermediary factor 1-beta (TIF1-beta)	59.0
88	Nap1l4	Nucleosome assembly protein 1-like 4	58.7
89	Ctnna1	Catenin alpha-1	58.4
90	Snd1	Staphylococcal nuclease domain-containing protein 1	57.7
91	Prdx1	Peroxiredoxin-1	57.7
92	Ywhah	14-3-3 protein eta	57.2
93	Pdia6	Protein disulfide-isomerase A6	56.3
94	Bfsp2	Phakinin	55.7
95	Ckb	Creatine kinase B-type	55.4
96	Kpnb1	Importin subunit beta-1	55.3
97	Eef1g	Elongation factor 1-gamma (EF-1-gamma)	55.2
98	Pdia4	Protein disulfide-isomerase A4	55.1
99	Cct4	T-complex protein 1 subunit delta (TCP-1-delta)	54.9
100	Ckap4	Cytoskeleton-associated protein 4	54.5
101	Ybx1	Nuclease-sensitive element-binding protein 1	54.0
102	Fubp1	Far upstream element-binding protein 1 (FBP; FUSE-binding protein 1)	54.0
103	Prdx2	Peroxiredoxin-2	53.8
104	Marcks	Myristoylated alanine-rich C-kinase substrate (MARCKS)	53.6
105	Eif2s1	Eukaryotic translation initiation factor 2 subunit 1	53.5
106	Nsfl1c	NSFL1 cofactor p47	53.5
107	Hmgb1	High mobility group protein B1	53.5
108	Npm1	Nucleophosmin (NPM)	53.0
109	Ptbp1	Polypyrimidine tract-binding protein 1 (PTB)	52.6

SN	UniProt Gene Name	Primary Protein Name	Avg. Lens
110	Pygm	Glycogen phosphorylase, muscle form	52.1
111	Fkbp4	Peptidyl-prolyl cis-trans isomerase FKBP4, N-terminally processed	51.7
112	Hnrnpm	Heterogeneous nuclear ribonucleoprotein M (hnRNPM)	50.9
113	Sptb	Spectrin beta chain, erythrocytic	50.6
114	Gss	Glutathione synthetase (GSH synthetase; GSH-S)	50.5
115	Sfpq	Splicing factor, proline- and glutamine-rich	50.3
116	Cap1	Adenylyl cyclase-associated protein 1 (CAP 1)	50.0
117	Npl	N-acetylneuraminate lyase (NALase)	50.0
118	Tpi1	Triosephosphate isomerase (TIM)	49.6
119	Caprin2	Caprin-2	48.5
120	Rpsa	40S ribosomal protein SA	48.0
121	Tjp1	Tight junction protein ZO-1	47.9
122	Pfn1	Profilin-1	47.4
123	Nono	Non-POU domain-containing octamer-binding protein (NonO protein)	46.5
124	Eprs	Proline--tRNA ligase	46.3
125	Hba	Hemoglobin subunit alpha	46.2
126	Rpl12	60S ribosomal protein L12	45.5
127	Acta1	Actin, alpha skeletal muscle, intermediate form	45.4
128	Mdh2	Malate dehydrogenase, mitochondrial	45.2
129	Epb4l2	Band 4.1-like protein 2	45.0
130	Rps4x	40S ribosomal protein S4, X isoform	44.7
131	Phgdh	D-3-phosphoglycerate dehydrogenase (3-PGDH)	44.7
132	Nedd4	E3 ubiquitin-protein ligase NEDD4	44.6
133	Pabpc1	Polyadenylate-binding protein 1 (PABP-1; Poly(A)-binding protein 1)	43.8
134	Rps8	40S ribosomal protein S8	43.6

SN	UniProt Gene Name	Primary Protein Name	Avg. Lens
135	Rps18	40S ribosomal protein S18	43.4
136	Psmd1	26S proteasome non-ATPase regulatory subunit 1	43.3
137	Atp1a1	Sodium/potassium-transporting ATPase subunit alpha-1 (Na(+)/K(+) ATPase alpha-1 subunit)	43.2
138	Dars	Aspartate--tRNA ligase, cytoplasmic	43.1
139	Dbnl	Drebrin-like protein	42.8
140	Ran	GTP-binding nuclear protein Ran	42.7
141	Vars	Valine--tRNA ligase	42.5
142	Hnrnpa1	Heterogeneous nuclear ribonucleoprotein A1, N-terminally processed	42.5
143	Serbp1	Plasminogen activator inhibitor 1 RNA-binding protein	42.4
144	Rps14	40S ribosomal protein S14	42.1
145	Rps27a	40S ribosomal protein S27a	42.1
146	Hbb-y	Hemoglobin subunit epsilon-Y2	41.8
147	Ywhag	14-3-3 protein gamma, N-terminally processed	41.7
148	Hnrnpa3	Heterogeneous nuclear ribonucleoprotein A3 (hnRNPA3)	41.6
149	Rpl3	60S ribosomal protein L3	41.6
150	Idh2	Isocitrate dehydrogenase [NADP], mitochondrial (IDH)	41.4

Table 4.

Lens-enriched proteins with mouse mutants displaying ocular/lens defects

SN	UniProt Gene Name	Mouse phenotype in IMPC	MGI ID
1	Lama1	Abnormal lens morphology and persistence of hyaloid vascular system	MGI:99892
2	Eml2	Abnormal eye morphology	MGI:1919455
3	Cap2	Cataract	MGI:1914502
4	Arvcf	Abnormal eye morphology and cataract	MGI:109620
5	Sarnp	Defects in lens morphology	MGI:1913368
6	Cotl1	Defects in lens morphology	MGI:1919292
7	Pepd	Abnormal optic disc morphology	MGI:97542

Table 5.

Identification of lens-enriched membrane proteins in mouse

UniProt Gene Name	Accession	Log$_2$FC	FC	p-value	FDR	Avg. Lens	Avg. WB
Gja8	P28236	8.0	441.9	1.0E-106	0.00000	74.5	0.2
Mip	P51180	6.8	196.8	6.0E-48	0.00000	33.2	0.2
Nrcam	Q810U4	4.3	22.0	4.0E-38	0.00000	32.2	1.5
Palm2	Q8BR92	4.0	19.8	4.0E-12	0.00000	11.9	0.6
Slc7a5	Q9Z127	4.0	18.6	8.0E-20	0.00000	16.1	0.9
Ank2	Q8C8R3	4.0	16.5	3.0E-93	0.00000	97.9	5.9
Cadm1	Q8R5M8	3.5	11.9	5.0E-31	0.00000	28.8	2.4
Slc2a1	P17809	3.1	9.5	1.0E-10	0.00000	11.5	1.2
Arvcf	P98203	3.0	8.4	4.0E-19	0.00000	21.3	2.5
Tjp1	P39447	2.9	7.6	1.0E-41	0.00000	47.9	6.3
Ezr	P26040	2.8	7.1	3.0E-70	0.00000	108.9	15.3
Pon2	Q62086	2.5	6.3	6.0E-05	0.00018	5.3	0.8
Chchd3	Q9CRB9	2.4	5.7	1.0E-07	0.00000	9.2	1.6
Basp1	Q91XV3	2.4	5.3	5.0E-42	0.00000	68.6	12.8
Bcl2l13	P59017	2.2	4.8	2.0E-12	0.00000	17.5	3.6
Cxadr	P97792	2.2	4.8	3.0E-06	0.00001	7.7	1.6
Cdh2	P15116	2.2	4.8	2.0E-12	0.00000	18.4	3.9
Fam49b	Q921M7	2.2	4.6	3.0E-20	0.00000	31.8	7.0
Col4a2	P08122	1.9	4.0	3.0E-05	0.00010	8.1	2.0
Slc3a2	P10852	1.9	3.8	8.0E-15	0.00000	27.2	7.2
Rac1	P63001	1.8	3.6	1.0E-11	0.00000	21.2	5.9
Itga6	Q61739	1.7	3.2	5.0E-06	0.00002	11.5	3.6
Atp2a2	O55143	1.6	3.1	2.0E-33	0.00000	78.7	25.1
Vdac2	Q60930	1.6	3.1	2.0E-16	0.00000	37.1	11.9
Rala	P63321	1.3	2.5	4.0E-04	0.00090	10.1	4.0
Add2	Q9QYB8	1.3	2.5	9.0E-04	0.00211	9.1	3.7
Palm	Q9Z0P4	1.2	2.4	2.0E-03	0.00402	8.1	3.3
Hsp90ab1	P11499	1.2	2.3	1.0E-83	0.00000	325.5	139.3
Mdh2	P08249	1.2	2.3	1.0E-10	0.00000	45.2	19.8
Sept2	P42208	1.1	2.2	4.0E-07	0.00000	25.1	11.4
Ctnna1	P26231	1.1	2.2	1.0E-14	0.00000	58.4	26.6
Itgb1	P09055	1.0	2.1	4.0E-05	0.00013	19.2	9.4
Rab5c	P35278	1.0	2.0	2.0E-03	0.00400	12.0	5.9

Table 6.

Identification of lens-expressed membrane proteins in human

UniProt Gene Name	Accession	Log$_2$FC	FC	p-value	FDR	Avg. Lens	Avg. WB
Gja8	P28236	8.0	441.9	1.0E-106	0.00000	74.5	0.2
Mip	P51180	6.8	196.8	6.2E-48	0.00000	33.2	0.2
Mxra7	Q9CZH7	5.2	63.9	4.5E-16	0.00000	10.8	0.2
Eml2	Q7TNG5	4.7	36.3	3.5E-13	0.00000	9.8	0.3
Nrcam	Q810U4	4.3	22.0	4.0E-38	0.00000	32.2	1.5
Cadm1	Q8R5M8	3.5	11.9	4.8E-31	0.00000	28.8	2.4
Slc2a1	P17809	3.1	9.5	1.3E-10	0.00000	11.5	1.2
Bcl2l13	P59017	2.2	4.8	1.6E-12	0.00000	17.5	3.6
Cxadr	P97792	2.2	4.8	3.1E-06	0.00001	7.7	1.6
Cdh2	P15116	2.2	4.8	1.8E-12	0.00000	18.4	3.9
Col4a2	P08122	1.9	4.0	3.2E-05	0.00010	8.1	2.0
Slc3a2	P10852	1.9	3.8	8.2E-15	0.00000	27.2	7.2
Cox4i1	P19783	1.5	2.8	5.1E-08	0.00000	18.6	6.6
Ppib	P24369	1.3	2.5	1.4E-09	0.00000	27.9	11.0
Itgb1	P09055	1.0	2.1	4.0E-05	0.00013	19.2	9.4
Slc25a4	P48962	1.0	2.0	7.7E-05	0.00023	26.2	13.3
Atp1a1	Q8VDN2	0.8	1.7	2.8E-06	0.00001	43.2	25.6
Canx	P35564	0.7	1.7	7.3E-04	0.00175	23.7	14.3
Rpn1	Q91YQ5	0.7	1.6	3.2E-04	0.00080	31.4	19.6
Gdi2	Q61598	0.4	1.3	1.5E-03	0.00332	60.6	45.6
Ganab	Q8BHN3	−0.7	−1.6	7.9E-04	0.00187	17.6	28.2
Ncam1	P13595	−1.0	−2.1	3.4E-06	0.00001	12.4	25.4
Por	P37040	−1.3	−2.4	2.7E-04	0.00069	4.5	11.0
Tfrc	Q62351	−1.6	−3.1	3.8E-08	0.00000	5.7	17.8