Proteomic profiling of retina and retinal pigment epithelium combined embryonic tissue to facilitate ocular disease gene discovery

Sandeep Aryal
University of Delaware
Deepti Anand
University of Delaware
Hongzhan Huang
University of Delaware
Ashok P Reddy
Oregon Health & Science University
Phillip A Wilmarth
Oregon State University
Larry L David
Oregon State University
Salil A Lachke (salil@udel.edu)
University of Delaware https://orcid.org/0000-0001-8845-010X

Research Article

Keywords: Retina, Proteome, Mass spectrometry, Development, Embryogenesis, Gene discovery, iSyTE

Posted Date: March 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2652395/v1

Proteomic profiling of retina and retinal pigment epithelium combined embryonic tissue to facilitate ocular disease gene discovery

Sandeep Aryal[1], Deepti Anand[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 19713 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

*** Correspondence:**
Salil Lachke
salil@udel.edu

Keywords: Retina, Proteome, Mass spectrometry, Development, Embryogenesis, Gene discovery, iSyTE

Abstract

To expedite gene discovery in eye development and its associated defects, we previously developed a bioinformatics resource-tool iSyTE (integrated Systems Tool for Eye gene discovery). However, iSyTE is presently limited to lens tissue and is predominantly based on transcriptomics datasets. Therefore, to extend iSyTE to other eye tissues on the proteome level, we performed high-throughput tandem mass spectrometry (MS/MS) on mouse embryonic day (E)14.5 retina and retinal pigment epithelium combined tissue and identified an average of 3,300 proteins per sample (n=5). High-throughput expression profiling-based gene discovery approaches–involving either transcriptomics or proteomics–pose a key challenge of prioritizing candidates from thousands of RNA/proteins expressed. To address this, we used MS/MS proteome data from mouse whole embryonic body (WB) as a reference dataset and performed comparative analysis–termed "*in silico* WB-subtraction"–with the retina proteome dataset. *In silico* WB-subtraction identified 90 high-priority proteins with retina-enriched expression at stringency criteria of ≥2.5 average spectral counts, ≥2.0 fold-enrichment, False Discovery Rate <0.01. These top candidates represent a pool of retina-enriched proteins, several of which are associated with retinal biology and/or defects (*e.g.*, Aldh1a1, Ank2, Ank3, Dcn, Dync2h1, Egfr, Ephb2, Fbln5, Fbn2, Hras, Igf2bp1, Msi1, Rbp1, Rlbp1, Tenm3, Yap1, *etc.*), indicating the effectiveness of this approach. Importantly, *in silico* WB-subtraction also identified several new high-priority candidates with potential regulatory function in retina development. Finally, proteins exhibiting expression or enriched-expression in the retina are made accessible in a user-friendly manner at iSyTE (https://research.bioinformatics.udel.edu/iSyTE/), to allow effective visualization of this information and facilitate eye gene discovery.

Introduction

Identification of genes associated with ocular development and its associated defects remains a challenge even though there has been widespread application of transcriptomics and proteomics toward these goals in recent years (Anand and Lachke 2017; Ahmad et al. 2018; Wolf et al. 2022). This is because high-throughput approaches such as RNA-sequencing or mass spectrometry identify thousands of candidates (RNA or protein) and the principal challenge lies in the application of effective downstream analyses for prioritizing candidates that are relevant to the development, homeostasis or pathology of the specific tissue/cell type of interest. Furthermore, prioritization strategies based solely on the extent of expression (*i.e.*, absolute expression) may result in missing important regulatory candidates that are not necessarily among the most highly expressed, but may nevertheless have a key role in development of the tissue/cell type. We developed a user-friendly web-based resource-tool called iSyTE (integrated Systems Tool for Eye gene discovery) to address these challenges in the eye; however, the earlier versions were focused on the lens (Lachke et al. 2012b; Kakrana et al. 2018; Anand et al. 2018; Aryal et al. 2020a). In addition to absolute expression in a tissue/cell type, iSyTE prioritizes candidate genes by comparing the gene expression profiles – in this case, of the lens – to that of a reference dataset, namely, mouse embryonic whole body (WB), a process termed "*in silico* WB-subtraction". This results in prioritization of candidates based on their "enriched-expression" in the lens as opposed to solely based on their absolute expression. Application of the iSyTE approach has effectively identified many new genes (*e.g.*, *Caprin2*, *Celf1*, *Mafg*, *Mafk*, *Rbm24*, *Tdrd7*, etc.) as well as led to the characterization of several regulatory pathways associated with eye/lens development and defects (Lachke et al. 2011, 2012a, b; Kasaikina et al. 2011; Wolf et al. 2013; Manthey et al. 2014; Agrawal et al. 2015; Dash et al. 2015, 2020; Audette et al. 2016; Patel et al. 2017, 2022; Anand and Lachke 2017; Cavalheiro et al. 2017; Siddam et al. 2018; Krall et al. 2018; Padula et al. 2019; Barnum et al. 2020; Aryal et al. 2020b; Anand et al. 2021; Choquet et al. 2021; Lachke 2022).

The present version of iSyTE is restricted to data on the lens, and furthermore, it is predominantly based on transcriptomic data in the form of microarrays or RNA-seq (Lachke et al. 2012b; Kakrana et al. 2018; Anand et al. 2018), with limited data on the proteome (Aryal et al. 2020a). Moreover, it is known that the correlation between RNA profiles and protein profiles is complex and not necessarily linear (Maier et al. 2009). Contributing to this is regulation at the post-transcriptional level – that involves non-coding RNA or RNA-binding protein (RBP)-mediated control over mRNA splicing, transport, stability and translation – all of which determines the cellular proteome (Brinegar and Cooper 2016; Hentze et al. 2018; Gebauer et al. 2021). Indeed, recent findings have demonstrated that RBPs play a conserved role in mediating post-transcriptional control in eye and lens development (Lachke et al. 2011; Lachke and Maas 2011; Dash et al. 2015, 2016, 2020; Siddam et al. 2018; Shao et al. 2020; Barnum et al. 2020; Nakazawa et al. 2020; Sundar et al. 2020; Aryal et al. 2020b; Lachke 2022; Matalkah et al. 2022). Together, these findings suggest that along with the transcriptome, characterization of the proteome is important to determine the factors that are important in a specific cell/tissue of interest. While there are currently many independent transcriptomics studies on the retina (Ratnapriya et al. 2019; Clark et al. 2019; Lukowski et al. 2019; Liang et al. 2019), studies on the retina proteome, especially in embryonic development, are limited (Zhao et al. 2007; Mizukami et al. 2008; Finnegan et al. 2008; Balasubramani et al. 2010). To address these knowledge-gaps in the context of eye development, we report here a tandem mass spectrometry (MS/MS)-based protein profiling of the mouse embryonic day (E) 14.5 retina and retinal pigment epithelium (RPE) combined (termed henceforth as retina) and its comparative analysis with *in silico* WB-subtraction. We demonstrate that while retina protein expression alone (*i.e.,* retina proteome not subjected to *in silico* subtraction) can identify several genes linked to retina biology and defects and is in itself helpful, *in silico* WB-subtraction provides another effective approach in prioritizing key candidates that are not necessarily among the highest expressed proteins in the retina. We generated new expression tracks at the

University of California at Santa Cruz (UCSC) Genome Browser and make this new data accessible through iSyTE.

Methods

Animals

Mice of the background C57BL/6J, obtained from The Jackson Laboratory, were used as wild-type animals in this study, and were bred and maintained at the University of Delaware Center for Animal research facility. The Institutional Animal Care and Use Committee (IACUC) approved the animal protocol (AUP#1226). All the animal experiments described in this study were performed in adherence to the guidelines in the Association of Research in Vision and Ophthalmology (ARVO) statement for the use of animals in ophthalmic and vision research.

Tissue Collection

Mice were bred and pregnant females were euthanized for obtaining embryos for collection of retina tissue. The day on which a vaginal plug was detected was designated as embryonic day (E) 0.5, and tissues was collected at E14.5. Whole retina tissue (retina + retinal pigment epithelium (RPE), henceforth referred to as "retina") were micro-dissected from E14.5 mouse embryos and stored at -80°C until further processing. Whole embryonic tissue minus the eye at E14.5 was considered as "whole embryonic body" (WB). Five biological replicates with each replicate consisting of two retinas isolated from the same embryo were collected. Tissues were processed as previously described (Aryal et al. 2020a). Briefly, tissue samples were suspended in 120 µl of TEAB buffer (167 mM triethyl ammonium bicarbonate buffer) and subjected to probe-sonication in a Fisher Scientific 60 Sonic Dismembrator. To these lysed samples, 40 µl of 20% SDS, 1% DCA and 40 µl of water were added to bring the total volume of each sample to 200 µl (final concentration: 4% SDS, 0.2% DCA, 100 mM TEAB), which were next centrifuged (16000 x *g*, 2 min., room temperature) followed by heating (90°C

for 15 min.). Sample protein quantification was estimated by BCA protein assay kit (Thermo Fisher Cat. No. 23225). For each biological replicate (n = 5 biological replicates), 55 µg of protein/sample was subjected to trypsinization as previously described (Erde et al. 2017). Briefly, a modified enhanced filter-aided digestion protocol (e-FASP) using Amicon 30 kDa ultracentrifugation devices was executed. Samples were subjected to TCEP (Tris Carboxy Ethyl Phosphene) reducing reagent at 90°C for 10 min, followed by transferring to an Amicon filter. Samples were then buffer exchanged into 8 M Urea, 0.2% deoxycholic acid (DCA), 100 mM TEAB. Next, samples were subjected to alkylation with iodoacetamide, exchanged into 0.2% DCA, 50 mM TEAB (pH 8.0) digestion buffer, and subjected to overnight digestion by trypsin (1:20 enzyme:substrate concentration). After overnight trypsin digestion, samples were subjected to centrifugation and the filtrate, which contained the peptides, was subjected to extraction with ethyl acetate, which served to remove DCA. A SpeedVac vacuum concentrator (Thermo Fisher Scientific) was then used to dry the samples which were then resuspended in 100 µl of HPLC-grade water. Next, a Pierce Quantitative Colorimetric Peptide Assay Kit was used to perform a peptide assay on the samples and the average peptide recovery from mouse E14.5 retina samples was estimated to be ~45 µg/sample. Whole embryonic body (WB) tissue (eye removed) sample processing was performed as previously described (Aryal et al. 2020a).

Mass Spectrometry

Mass spectrometry (MS) was performed as previously described (Aryal et al. 2020a). Briefly, protein samples (concentration: 4 µg in 5% Formic acid) were loaded for 5 min on an Acclaim PepMap 0.1 x 20 mm NanoViper C18 peptide trap (Thermo Fisher Scientific)(flow rate: 10 µl/min; mobile phase: in a 2% acetonitrile, 0.1% formic acid). PepMap RSLC C18 2 µm particle, 75 µm x 50 cm EasySpray column (Thermo Fisher Scientific) was used for separating peptides over 205 min on a 7.5–30% acetonitrile gradient (mobile phase: 0.1% formic acid, 300 nl/min flow rate) with Dionex NCS-3500RS UltiMate RSLC nano UPLC system. An Orbitrap Fusion mass spectrometer configured with an

EasySpray NanoSource (Thermo Fisher Scientific) was used to collect tandem MS data, using data dependent analysis (DDA) configuration and a MS/DD-MS/MS instrument menthod (full MS resolutions: 120,000 at m/z 200, mass range 375-1500, charge state 2-7; full MS AGC target: 400,000; intensity threshold: 5,000; max inject time: 50 ms and 10 ppm dynamic exclusion for 60 s; AGC target value for fragment spectra: 5,000; isolation mode: quadrupole; isolation width: 1.6 m/z; isolation offset: off; activation type: CID; collision energy: fixed 35%; maximum injection time: 300 ms; detector type: IonTrap). Centroid mode using positive polarity was used to acquire data. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the Proteomics IDEntifications (PRIDE) database (https://www.ebi.ac.uk/pride/) (Perez-Riverol et al. 2022) partner repository with the dataset identifier PXD039490.

Raw File Conversion and Database Search

MSConvert (Proteowizard toolkit) was used to convert RAW files to MS2 format for the samples as described (Chambers et al. 2012; Aryal et al. 2020a). The retina samples had ~74K MS2 scans per run. A software (available at https://github.com/pwilmart/fasta_utilities.git) was used to download a canonical mouse reference proteome (version 2019.04; 22,287 sequences) from UniProt. To this, a concatenated sequence-reversed decoy database was added along with common contaminants (179 sequences) to obtain 44,932 entries. Peptide sequences were assigned to the MS2 spectra (PSMs) using the search engine Comet (Eng et al. 2013), which was configured as previously described (Aryal et al. 2020a) and were as follows: tryptic cleavage (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 PAW pipeline (https://github.com/pwilmart/PAW_pipeline.git) and the target/decoy method described previously (Elias and Gygi 2007; Wilmarth et al. 2009) was used for post-processing the highest scoring matches for individual PSMs (obtained from Comet) using false discovery rate (FDR) error control. Peptides of different charge states (2+, 3+, and 4+ were considered) and modification state (unmodified or oxidized) were processed to derive accurate delta mass conditional score histograms. FDR values were estimated based on target and decoy score histograms as a function of a Peptide-Prophet-like discriminant score to set thresholds for experiment-wide PSM FDR of 1% as described previously (Keller et al. 2002; Aryal et al. 2020a). A minimum length of 7 amino acids-length were considered for peptide matches. The number of confidently identified (1% FDR) PSMs per sample was 35.4K and the identification rate was 48%.

Protein Inference

The expressed proteins were inferred, using basic parsimony principles, based on the filtered PSM sequences (Nesvizhskii and Aebersold 2005). Homologous protein family members were grouped using an extended parsimony algorithm when evidence to distinguish family members was insufficient. In total, 3,963 proteins were detected after grouping (excluding common contaminant proteins) with 37 decoy matches, for a protein FDR of about 0.9%. The average number of proteins identified per sample was 3,296.

Quantitative Analysis

For the retina and the WB samples, equal amounts of protein were digested and the total spectral counts (SpC, a robust semi-quantitative measure) were measured. Prior to protein inference, the SpC for individual samples were tallied and they independently validated the peptide assay results. Next, the retina and the WB samples were matched by subjecting the individual samples to be scaled to the average total spectral count per sample. Both the retina and the WB samples had about 3,300 protein identifications per sample. Next, the proteins with enriched expression in the retina compared to WB

were determined as follows: For individual proteins, the average SpC for all samples was computed from the scaled data, and only values greater than 2.5 (2,675 proteins) were considered in the differential expressed enrichment analysis between the retina and WB. The bioconductor package, edgeR, was used for differential gene expression analysis (Robinson and Oshlack 2010; Robinson et al. 2010). The default Benjamini-Hochberg multiple testing corrections and the exact test in edgeR were applied in R (version 3.5.3).

Gene Ontology Analysis

For functional annotation by gene ontology (GO) categories, a cluster-based analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6 .8) (Huang et al. 2009) was performed on candidate proteins with retina-expression and retina-enriched expression that were identified by *in silico* WB-subtraction (the cut-offs were: ≥2.5 average spectral counts, ≥2.0 fold-enrichment, FDR <0.01). Benjamini corrected significant *p*-values were considered for prioritization of the pathways and GO categories identified from this analysis as previously described (Aryal et al. 2020a).

Results and Discussion

Embryonic retina proteome generation and quality assessment

We designed an experimental workflow to isolate mouse E14.5 retina, generate its proteome and perform *in silico* WB-subtraction (Fig. 1A). Retina tissue was micro-dissected from mouse E14.5 eyes and processed for protein preparation and proteome analysis. Mouse WB preparation was performed as previously described (Aryal et al. 2020a). Further, proteome downstream analyses were performed according to the outlined workflow (Fig. 1B). From individual retina and WB samples ($n = 5$ biological replicates), 55 µg of protein were subjected to eFASP (enhanced filter-aided sample preparation) and digestion by trypsin. After digestion, equal amounts of peptides were used for high-throughput tandem

mass spectrometry (MS/MS) and spectral count (SpC) data were generated. Application of stringent criteria (≥2 distinct peptides per protein in at least one sample, ≥2.5 average SpC in the retina) to the resulting data led to enrichment analysis of 2,675 proteins in the E14.5 retina (Supplementary Table S1). Across the samples, on average ~35K SpC were detected. Total average SpC was subjected to TMM (trimmed mean of M-values) normalization using edgeR package (Robinson et al. 2010) to account for differences in SpC between retina and WB (Fig. 2A). Next, the quality of data was assessed by boxplots for the normalized SpC datasets that demonstrated that the median expression levels were similar between all the retina and the WB samples (Fig. 2A). Further, multidimensional scaling-based cluster analysis was performed to examine the quality of TMM normalized SpC proteome data. Cluster analysis demonstrated that all five biological replicates of the retina clustered closer to each other and distinctly away from WB samples, which themselves clustered together (Fig. 2B). Sample to sample correlation within the retina and within the WB samples was examined by scatter plot comparisons in all combinations for retina and WB samples, which demonstrated high correlation between samples of the same type (Fig. 2C, D). While the five WB samples correlated with each other at r value >0.98, all the five retina samples correlated with each other at r value >0.97. Finally, comparing the correlation between the average SpC of the retina and that of WB shows that the correlation is much lower ($r = 0.81$) between the retina and WB (Fig. 2E).

Gene Ontology (GO) analysis of proteins expressed in the E14.5 mouse

Before performing other downstream analysis, we characterized the E14.5 proteins that were identified. Many proteins previously linked to retinal development and disease were identified in the proteome analysis, based solely on expression (Supplementary Table S1 for all proteins and Supplementary Table S2 for the top 150). To examine whether specific pathways relevant to retina biology were enriched in this dataset, a cluster-based analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID v6 .8) for functional annotation by gene ontology (GO) categories.

This analysis identified several interesting pathways. These were related to post-transcriptional control of gene expression, *e.g.*, "GO:0003723 RNA-binding", "GO:0030529 intracellular ribonucleoprotein complex", "GO:0006397 mRNA processing" and "GO:0051028 mRNA transport" (Fig. 3)(Supplementary Table S3). Proteins involved in other molecular pathways and processes *e.g.*, "GO:0015031 protein transport", "GO:0055114 oxidation-reduction process" were also identified in the dataset. Finally, pathways in basic cell biological processes were also enriched, *e.g.*, "GO:0007049 cell cycle", "GO:0098641 cadherin binding involved in cell-cell adhesion", and "GO:0003779 actin binding" in the total proteins expressed in E14.5 mouse retina. Together, these represent promising new candidates for future investigations in the retina.

MS/MS *in silico* WB-subtraction identifies proteins exhibiting retina-enriched expression

While GO analysis of total expressed proteins were helpful, to further prioritize the candidates from the E14.5 retina proteome, the "*in silico* WB-subtraction" approach, which has been effectively applied for prioritizing cataract-linked genes in the lens, was applied to this dataset. To do so, we computed the average SpC for all samples and scaled (normalized) data for each protein. Those peptides that passed the filtration criteria of ≥2.5 SpC were considered in the analysis. This approach identified 2675 proteins that could be tested for differential expression between the retina and WB samples. At ≥2.0 fold-enrichment and FDR <0.01 cut-off, 90 proteins had enriched expression in the retina compared to WB (Table 1). These "retina-enriched" proteins identified many proteins linked to retinal defects and revealed several new promising candidates (Fig. 4) demonstrating that the *in silico* WB-subtraction approach can be effectively applied to the retina. Further, compared to absolute expression of proteins, *in silico* WB-subtraction could more effectively prioritize key proteins associated with retina biology and disease. For example, the top 30 proteins ranked on relative abundance in the retina (*i.e.*, not subjected to *in silico* WB-subtraction) did not contain a single protein that has been associated with retina development or defects/disease (Fig. 5A). Indeed, candidates in this list, termed "retina

253 expression" list, were representative of general housekeeping/structural proteins such as
254 Glyceraldehyde-3-phosphate dehydrogenase (Gapdh), Actins (Acta1, Actb), Myosins (Myh3, Myh9,
255 Myh10), Tubulin (Tubb5), Collagen (Col12a1) and several others, not exclusively associated with
256 retina biology. In sharp contrast, the list of the top 30 candidates identified by *in silico* WB-subtraction,
257 termed "retina enriched" list of candidates, contained 1/3rd (10 out of 30) candidates that have been
258 associated with retinal biology and/or defects/disease (Fig. 5B). These are Aldehyde dehydrogenase
259 family 1, subfamily A1 (Aldh1a1), Tyrosinase (Tyr), Keratocan (Kera), Melanocyte protein PMEL
260 (Pmel), Hemicentin-1 (Hmcn1), Retinaldehyde-binding protein 1 (Rlbp1), Harvey rat sarcoma virus
261 oncogene (Hras), Laminin subunit alpha-5 (Lama5), Epidermal growth factor receptor (Egfr),
262 Hephaestin (Heph) and Teneurin-3 (Tenm3). Importantly, the top candidates identified by the *in silico*
263 WB-subtraction approach contained regulatory proteins that are not necessarily among the top highly
264 expressed proteins in the retina. In contrast, no regulatory proteins were detected in the top 30 retina
265 expression list. Finally, the significant differences in SpC levels for different proteins between the
266 retina and WB serves to explain the basis for the effectiveness of the *in silico* WB-subtraction strategy
267 in prioritization of candidates relevant to the retina and its associated defects (Fig. 5B).

268 **Biological and disease relevance of top retina-enriched proteins**

269 Next, we conducted a detailed analysis of all 90 retina-enriched candidates in the context of the
270 published literature to determine their potential relevance to retina biology and defects. Application
271 of this evidence-based curation identified 27 of the 90 (~30%) retina-enriched proteins prioritized by
272 the *in silico* WB-subtraction strategy to be associated with retina biology and/or defects (Table 1). The
273 topmost enriched gene Aldehyde dehydrogenase family 1, subfamily A1 (Aldh1a1; also known as
274 Raldh1 (Retinal dehydrogenase 1)) is shown to regulate dorsal choroidal vasculature development via
275 Sox9 upregulation in retinal pigmented epithelium in mice (Goto et al. 2018). The enriched-expression
276 list also independently identified Tyrosinase (Tyr) protein which is essential for melanin biosynthesis

277 and therefore critical for RPE (retinal pigment cells) and other retinal cells (Jeffery et al. 1994, 1997).
278 Among the candidates is the premelanosome (Pmel) protein, whose deficiency in mice results in cell
279 shape changes, *e.g.*, the normally "oblong" shaped melanosomes turn spherical in RPE cells (Hellström
280 et al. 2011). Further, whole exome sequencing of patients with early-onset age-related macular
281 degeneration (AMD) revealed a single base deletion in another candidate with retina-enriched
282 expression, the hemicentin-1 (HMCN1) gene (Pras et al. 2015).

283 Several proteins involved in signaling pathways were identified among the candidates with retina-
284 enriched expression. Mutations in the candidate, Harvey rat sarcoma virus oncogene (Hras), are
285 associated with retinal dystrophy in two patients with Costello syndrome (Pierpont et al. 2017).
286 Another factor prioritized by *in silico*-WB subtraction is the epidermal growth factor receptor (Egfr),
287 which is associated with retinal cell fate determination (Lillien 1995). Additionally, eph receptor B2
288 (Ephb2) was identified and its deletion in mouse is associated with axonal degeneration in retinal
289 ganglion cell (Fu and Sretavan 2012). The receptor-type tyrosine-protein phosphatase F (Ptprf) is
290 known to be expressed in retinal ganglion cells in mice (Lorber et al. 2005). Further, mutation in
291 Fibulin-5 (FBLN5) is reported in human patients with AMD (Stone et al. 2004; Lotery et al. 2006).
292 Morpholino-based reduction in the retinol-binding protein 1 (Rbp1) has been shown to result in
293 misfolding of outer segments of retina in *Xenopus* (Wang et al. 2010). Recessive mutations in another
294 top candidate, retinaldehyde-binding protein 1 (RLBP1), cause Retinitis punctatta albescens in humans
295 (Morimura et al. 1999).

296 Some RNA-binding proteins (RBPs) involved in post-transcriptional gene expression control that are
297 linked to retina development and differentiation were also among the prioritized proteins. Musashi
298 homolog 1 (Msi1) is an RBP whose deficiency causes photoreceptor morphogenesis defect in mice
299 (Sundar et al. 2020). Another RBP, the insulin-like growth factor 2 mRNA-binding protein 1 (Igf2bp1)
300 is required for retinal ganglion cell axon outgrowth in zebrafish (Gaynes et al. 2015a). Mex3 RNA-

binding family member A (Mex3a) is expressed in the ciliary marginal zone in zebrafish (Naef et al. 2020). The mRNA cap guanine-N7 methyltransferase (Rnmt) is known to be expressed in *Xenopus* retina (Lokapally et al. 2016). Among other proteins with a regulatory function that were identified, deletion in mouse of the transcriptional coactivator Yes-associated protein 1 (Yap1) shows that Yap1 is essential for maintaining retinal pigmented epithelium differentiation (Lu et al. 2020).

Several proteins associated with the cell membrane and/or cytoskeleton were identified. Patients with mutation in the gap junction alpha-1 protein (Gja1) exhibit optic nerve and retinal dysplasia (Gabriel et al. 2011). The ankyrin proteins identified here, ankyrin-2 (Ank2) and ankyrin-3 (Ank3), have been shown to be essential for development of rod photoreceptors in mice (Kizhatil et al. 2009b, a)(Kizhatil et al., 2009a, 2009b)(Kizhatil et al., 2009a, 2009b)(Kizhatil et al., 2009a). Knockout mice for another candidate, the cell adhesion molecule 1 (Cadm1), exhibited impaired response to light stimulation and for structural integrity of rod synapses (Ribic et al. 2014). Deficiency of the transmembrane protein, teneurin-3 (Tenm3) in zebrafish causes abnormal retinal ganglion cell morphology and the lack of Tenm3 in mice lead to defects in binocular visual coordination (Leamey et al. 2007; Antinucci et al. 2013).

Similarly, extracellular proteins as well as other cellular proteins with relevance to retinal biology were identified. Genetic variants in fibrillin-2 C-terminal peptide (FBN2) are reported to contribute to AMD (Ratnapriya et al. 2014). Cytoplasmic dynein 2 heavy chain 1 (Dync2h1) mutations in humans have been associated with non-syndromic inherited retinal disease (Vig et al. 2020). Deficiency in mice of the macrophage migration inhibitory factor (Mif) is associated with reduction in proliferation and inhibition of preretinal angiogenesis (Wang et al. 2017). Further, deficiency in mice of the small leucine-rich proteoglycan family protein Decorin (Dcn) results in structural and microvascular defects in retina (Lim et al. 2018). Finally, deficiency of rootletin (Crocc), recognized as a major component of the ciliary rootlet, is reported to cause retinal degeneration in mice (Yang et al. 2005).

Some genes that function in early retina development were also identified. For example, the transcriptional repressor CTCF (Ctcf) was among these candidates, and its deletion in mouse forebrain at E8.5 causes apoptosis and reduced retinal tissue by E13.5 (Watson et al. 2014). Further, a point mutation identified in the mouse Laminin subunit alpha-5 (Lama5) showed defective retinal cup morphology as early as at E15.5 (Jones et al. 2020). Proteins involved in homeostasis were also identified in this study. For example, hephaestin (Heph) is reported to be essential for iron homeostasis in mice and its deficiency is associated with retinal degeneration (Hahn et al. 2004).

Several other proteins were identified that are independently reported to be expressed in retina, but their function has not been examined in detail. We identified several crystallins (Crybb1, Crybb3, Cryge, and Crym) that are known to be expressed in the retina and have been associated with retinal ganglion cell survival and regeneration (Xi et al. 2003; Piri et al. 2013). Interestingly, knockout mice for another crystallin prioritized in this study, alpha-crystallin A chain (Cryaa), exhibit retinal neovascularization defect (Xu et al. 2015). Similarly, a binding partner (NEDD8 ultimate buster 1 (NUB1)) of the retinal defect-associated protein, Aryl hydrocarbon receptor-interacting protein-like 1 (AIPL1), identified here, is shown to be expressed in developing and adult human retina (Akey et al. 2002). The non-histone chromosomal protein, HMG-14 (Hmgn1), is expressed throughout retina in adult mouse (Lucey et al. 2008). The multi-functional serine and arginine-rich (SR) and desmosome associated protein Pinin (Pnn) is independently reported to be expressed photoreceptors of developing mouse retina (Leu and Ouyang 2006).

A few other proteins identified here have been reported to be associated with eye defects or disease, but their specific function in the retina has not been explored in detail. For example, FRAS1-related extracellular matrix protein 2 (Frem2) is associated with Cryptophthalmos (Yu et al. 2018). Interestingly, keratocan (Kera) deficiency in mice is associated with corneal defects, but its role in the retina has not been examined (Liu et al. 2003). Finally, kinesin-like protein 1A (Kif1a) has been

associated with optic nerve hyperplasia but its mechanistic role is not known in detail (Raffa et al. 2017).

Together, this documented association to retina biology of nearly 1/3rd of the top 90 proteins identified by *in silico*-WB subtraction, renders confidence that other candidates may also have key roles in the retina and may be linked to its defects.

Gene Ontology analysis of retina-enriched proteins

To gain insights into the relevance of the 90 candidates identified by *in silico* WB-subtraction to retina biology, a cluster-based analysis 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 90 retina enriched proteins into several annotation clusters. These are proteins involved in regulatory processes such as chromatin remodeling, *e.g.*, "GO:0006338 chromatin remodeling," "GO:0016569 covalent chromatin modification", "GO:0071564 npBAF complex", "GO:0016514 SWI/SNF complex", "GO:0090544 BAF-type complex", "GO:0071565 nBAF complex", "GO:0006337 nucleosome disassembly", "GO:0043044 ATP-dependent chromatin remodeling", as well as those involved in signaling pathways, *e.g.*, "GO:0043406 positive regulation of MAP kinase activity". Proteins involved in basic cellular processes, *e.g.*, "GO:0008283 cell proliferation", "GO:0007155 cell adhesion" were also identified. Additionally, proteins involved in extracellular matrix were identified, *e.g.*, "GO:0005604 basement membrane", "GO:0005578 proteinaceous extracellular matrix". Finally, proteins with roles in nervous system development were identified (Fig. 6) (Supplementary Table S4). Thus, this analysis identifies key candidates in specific processes relevant to retina biology, which can be functionally characterized in future studies.

Visualization and access of retina-enriched and retina-expressed proteins in iSyTE

Next, we wanted to make this rich proteome information freely available to the research community. Thus, we developed new custom annotation-tracks on the UCSC Genome Browser that provide a heat-map representation of proteins based on their absolute expression or enriched-expression in the E14.5 mouse retina. These tracks are publicly accessible through the web-based resource-tool iSyTE (https://research.bioinformatics.udel.edu/iSyTE/). As examples, the enrichment and expression in the retina, of proteins previously linked to retina defects, *e.g.*, Hras and Tyr, are shown as visualized in iSyTE (Fig. 7A, B). This web-based resource-tool will allow ready and user-friendly visualization of proteins in the E14.5 mouse retina.

Conclusion

Recent studies have highlighted that post-transcriptional regulation of gene expression plays a key role in determining the cellular proteome in eye development. Therefore, it is important to include ocular proteome data to the existing RNA-based profiling datasets to gain new insights into eye development. As a proof-of-principal we previously generated proteomic profiles for the mouse lens and the embryonic whole body (WB) and effectively applied *in silico* WB-subtraction strategy to identify proteins with lens-enriched abundance, which – in addition to consideration of absolute expression scores – allows a prioritized list of proteins for further study. In the present study, we expanded this approach to the mouse embryonic retina. We identified 90 proteins with retina-enriched expression. Nearly 1/3rd of these candidates have been previously reported to be associated with retinal defects. This suggests that *in silico* WB-subtraction was effective in prioritizing select candidates from over 2,600 identified proteins and the other 2/3rd of these identified proteins represent an unexplored pool of candidates for future characterization of their function in the retina. Indeed, there exist independent evidence in the literature for several of these candidates to be expressed in the retina, in agreement with the proteome data reported in the present study. Further, in addition to these "retina-enriched" candidates, nearly 4,000 proteins were found to be present in the mouse E14.5 retina proteome. It

396 should be noted that while many proteins linked to retina biology and pathology were identified in this
397 study, transcription factors (TFs) such as Otx2, Sox2 and Vsx2 with key roles in the retina were not
398 detected. This may be due to the following reasons. While they may be enriched in tissues, TFs are
399 often in lower abundance compared to other expressed proteins (Tacheny et al. 2013). Furthermore,
400 their levels are often spatiotemporally restricted in specific cells within the tissue, information that is
401 compromised when using bulk tissue (as is the case in the present study). In the present study, we
402 measured static protein relative abundances and did not attempt dynamic system measurements (*e.g.*,
403 those informing on protein turnover). Although 2675 quantifiable proteins (from the total 4680
404 proteins detected, which is generally considered a deep proteome) were identified in the present study,
405 since the above mentioned TFs were not among these proteins, this suggests that more sensitive
406 methods would be needed to detect these proteins in future studies. Together, these datasets and their
407 ready accessibility through the web-based ocular gene discovery tool iSyTE represent a rich resource
408 for prioritizing candidates for future hypothesis-driven studies in retina development. Finally, this
409 study serves as a proof of the principle that *in silico* subtraction can also be applied to the retina and
410 RPE to identify promising new candidates in these tissues. In the future, this approach will be expanded
411 to prioritize candidates in other developmental stages of the retina.

412 **Conflict of Interest**

413 *The authors declare that the research was conducted in the absence of any commercial or financial*
414 *relationships that could be construed as a potential conflict of interest*.

415 **Author Contributions**

416 S.A., D.A., H.H., A.P.R., P.A.W., L.L.D. and S.A.L. contributed to the generation and interpretation
417 of the data. S.A., D.A., P.A.W., L.L.D. and S.A.L. analysed the data. H.H. and D.A. contributed to

the development of iSyTE tracks in UCSC Genome Browser. S.A. and S.A.L. wrote the manuscript and all authors contributed to editing the manuscript.

Funding

This work was supported by National Institutes of Health [R01 EY021505 and R01 EY029770] to S.L. S.A. was supported by a Dissertation Fellows Award from the University of Delaware, a Fight For Sight Summer Research Fellowship and a Sigma Xi Grant-in-Aid Research Award. D.A. was supported by a Knights Templar Pediatric Ophthalmology Career Starter Grant Award. Support from the University of Delaware Proteomics and Mass Spectrometry Facility was made possible through funding from the State of Delaware and National Institutes of Health / National Institute of General Medical Sciences INBRE Program Grant [P20 GM103446]. Mass spectrometric analysis was performed by the OHSU Proteomics Shared Resource with partial support from NIH core grants [P30 EY010572 and P30 CA069533] and shared instrument grant [S10 OD012246].

Supplementary Material

Supplementary Material in the form of four tables is submitted along with the manuscript.

References

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. https://doi.org/10.1007/s00439-015-1554-5

Ahmad MT, Zhang P, Dufresne C, et al (2018) The Human Eye Proteome Project: Updates on an Emerging Proteome. Proteomics 18:e1700394. https://doi.org/10.1002/pmic.201700394

Akey DT, Zhu X, Dyer M, et al (2002) The inherited blindness associated protein AIPL1 interacts with the cell cycle regulator protein NUB1. Hum Mol Genet 11:2723–2733. https://doi.org/10.1093/hmg/11.22.2723

Anand D, Al Saai S, Shrestha SK, et al (2021) Genome-Wide Analysis of Differentially Expressed miRNAs and Their Associated Regulatory Networks in Lenses Deficient for the Congenital Cataract-Linked Tudor Domain Containing Protein TDRD7. Front Cell Dev Biol 9:615761. https://doi.org/10.3389/fcell.2021.615761

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

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

450 Antinucci P, Nikolaou N, Meyer MP, Hindges R (2013) Teneurin-3 specifies morphological and
451 functional connectivity of retinal ganglion cells in the vertebrate visual system. Cell Rep
452 5:582–592. https://doi.org/10.1016/j.celrep.2013.09.045

453 Aryal S, Anand D, Hernandez FG, et al (2020a) MS/MS in silico subtraction-based proteomic
454 profiling as an approach to facilitate disease gene discovery: application to lens development
455 and cataract. Hum Genet 139:151–184. https://doi.org/10.1007/s00439-019-02095-5

456 Aryal S, Viet J, Weatherbee BAT, et al (2020b) The cataract-linked RNA-binding protein Celf1 post-
457 transcriptionally controls the spatiotemporal expression of the key homeodomain
458 transcription factors Pax6 and Prox1 in lens development. Hum Genet 139:1541–1554.
459 https://doi.org/10.1007/s00439-020-02195-7

460 Audette DS, Anand D, So T, et al (2016) Prox1 and fibroblast growth factor receptors form a novel
461 regulatory loop controlling lens fiber differentiation and gene expression. Development
462 143:318–328. https://doi.org/10.1242/dev.127860

463 Balasubramani M, Schreiber EM, Candiello J, et al (2010) Molecular interactions in the retinal
464 basement membrane system: a proteomic approach. Matrix Biol 29:471–483.
465 https://doi.org/10.1016/j.matbio.2010.04.002

466 Barnum CE, Al Saai S, Patel SD, et al (2020) The Tudor-domain protein TDRD7, mutated in
467 congenital cataract, controls the heat shock protein HSPB1 (HSP27) and lens fiber cell
468 morphology. Hum Mol Genet 29:2076–2097. https://doi.org/10.1093/hmg/ddaa096

469 Brinegar AE, Cooper TA (2016) Roles for RNA-binding proteins in development and disease. Brain
470 Res 1647:1–8. https://doi.org/10.1016/j.brainres.2016.02.050

471 Cavalheiro GR, Matos-Rodrigues GE, Zhao Y, et al (2017) N-myc regulates growth and fiber cell
472 differentiation in lens development. Dev Biol 429:105–117.
473 https://doi.org/10.1016/j.ydbio.2017.07.002

474 Chambers MC, Maclean B, Burke R, et al (2012) A cross-platform toolkit for mass spectrometry and
475 proteomics. Nat Biotechnol 30:918–920. https://doi.org/10.1038/nbt.2377

476 Choquet H, Melles RB, Anand D, et al (2021) A large multiethnic GWAS meta-analysis of cataract
477 identifies new risk loci and sex-specific effects. Nat Commun 12:3595.
478 https://doi.org/10.1038/s41467-021-23873-8

479 Clark BS, Stein-O'Brien GL, Shiau F, et al (2019) Single-Cell RNA-Seq Analysis of Retinal
480 Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell
481 Specification. Neuron 102:1111-1126.e5. https://doi.org/10.1016/j.neuron.2019.04.010

Dash S, Brastrom LK, Patel SD, et al (2020) The master transcription factor SOX2, mutated in anophthalmia/microphthalmia, is post-transcriptionally regulated by the conserved RNA-binding protein RBM24 in vertebrate eye development. Hum Mol Genet 29:591–604. https://doi.org/10.1093/hmg/ddz278

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. https://doi.org/10.1002/dvdy.24303

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. https://doi.org/10.1002/wrna.1355

Duester G (2022) Towards a Better Vision of Retinoic Acid Signaling during Eye Development. Cells 11:322. https://doi.org/10.3390/cells11030322

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. https://doi.org/10.1038/nmeth1019

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

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. https://doi.org/10.1007/978-1-4939-6747-6_2

Fan X, Molotkov A, Manabe S-I, et al (2003) Targeted Disruption of Aldh1a1 (Raldh1) Provides Evidence for a Complex Mechanism of Retinoic Acid Synthesis in the Developing Retina. Mol Cell Biol 23:4637–4648. https://doi.org/10.1128/MCB.23.13.4637-4648.2003

Finnegan S, Robson JL, Wylie M, et al (2008) Protein expression profiling during chick retinal maturation: a proteomics-based approach. Proteome Sci 6:34. https://doi.org/10.1186/1477-5956-6-34

Fu CT, Sretavan D (2012) Involvement of EphB/Ephrin-B signaling in axonal survival in mouse experimental glaucoma. Invest Ophthalmol Vis Sci 53:76–84. https://doi.org/10.1167/iovs.11-8546

Gabriel LAR, Sachdeva R, Marcotty A, et al (2011) Oculodentodigital dysplasia: new ocular findings and a novel connexin 43 mutation. Arch Ophthalmol 129:781–784. https://doi.org/10.1001/archophthalmol.2011.113

Gaynes JA, Otsuna H, Campbell DS, et al (2015a) The RNA Binding Protein Igf2bp1 Is Required for Zebrafish RGC Axon Outgrowth In Vivo. PLoS ONE 10:e0134751. https://doi.org/10.1371/journal.pone.0134751

Gaynes JA, Otsuna H, Campbell DS, et al (2015b) The RNA Binding Protein Igf2bp1 Is Required for Zebrafish RGC Axon Outgrowth In Vivo. PLoS One 10:e0134751. https://doi.org/10.1371/journal.pone.0134751

520 Gebauer F, Schwarzl T, Valcárcel J, Hentze MW (2021) RNA-binding proteins in human genetic
521 disease. Nat Rev Genet 22:185–198. https://doi.org/10.1038/s41576-020-00302-y

522 Goto S, Onishi A, Misaki K, et al (2018) Neural retina-specific Aldh1a1 controls dorsal choroidal
523 vascular development via Sox9 expression in retinal pigment epithelial cells. Elife 7:.
524 https://doi.org/10.7554/eLife.32358

525 Hahn P, Qian Y, Dentchev T, et al (2004) Disruption of ceruloplasmin and hephaestin in mice causes
526 retinal iron overload and retinal degeneration with features of age-related macular
527 degeneration. Proc Natl Acad Sci U S A 101:13850–13855.
528 https://doi.org/10.1073/pnas.0405146101

529 Hellström AR, Watt B, Fard SS, et al (2011) Inactivation of Pmel alters melanosome shape but has
530 only a subtle effect on visible pigmentation. PLoS Genet 7:e1002285.
531 https://doi.org/10.1371/journal.pgen.1002285

532 Hentze MW, Castello A, Schwarzl T, Preiss T (2018) A brave new world of RNA-binding proteins.
533 Nat Rev Mol Cell Biol. https://doi.org/10.1038/nrm.2017.130

534 Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists
535 using DAVID bioinformatics resources. Nat Protoc 4:44–57.
536 https://doi.org/10.1038/nprot.2008.211

537 Jeffery G, Brem G, Montoliu L (1997) Correction of retinal abnormalities found in albinism by
538 introduction of a functional tyrosinase gene in transgenic mice and rabbits. Brain Res Dev
539 Brain Res 99:95–102. https://doi.org/10.1016/s0165-3806(96)00211-8

540 Jeffery G, Schütz G, Montoliu L (1994) Correction of abnormal retinal pathways found with albinism
541 by introduction of a functional tyrosinase gene in transgenic mice. Dev Biol 166:460–464.
542 https://doi.org/10.1006/dbio.1994.1329

543 Jones LK, Lam R, McKee KK, et al (2020) A mutation affecting laminin alpha 5 polymerisation
544 gives rise to a syndromic developmental disorder. Development 147:.
545 https://doi.org/10.1242/dev.189183

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

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

551 Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the
552 accuracy of peptide identifications made by MS/MS and database search. Anal Chem
553 74:5383–5392. https://doi.org/10.1021/ac025747h

554 Kizhatil K, Baker SA, Arshavsky VY, Bennett V (2009a) Ankyrin-G promotes cyclic nucleotide-
555 gated channel transport to rod photoreceptor sensory cilia. Science 323:1614–1617.
556 https://doi.org/10.1126/science.1169789

557 Kizhatil K, Sandhu NK, Peachey NS, Bennett V (2009b) Ankyrin-B is required for coordinated
558 expression of beta-2-spectrin, the Na/K-ATPase and the Na/Ca exchanger in the inner
559 segment of rod photoreceptors. Exp Eye Res 88:57–64.
560 https://doi.org/10.1016/j.exer.2008.09.022

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

564 Lachke SA (2022) RNA-binding proteins and post-transcriptional regulation in lens biology and
565 cataract: Mediating spatiotemporal expression of key factors that control the cell cycle,
566 transcription, cytoskeleton and transparency. Exp Eye Res 214:108889.
567 https://doi.org/10.1016/j.exer.2021.108889

568 Lachke SA, Alkuraya FS, Kneeland SC, et al (2011) Mutations in the RNA granule component
569 TDRD7 cause cataract and glaucoma. Science 331:1571–1576.
570 https://doi.org/10.1126/science.1195970

571 Lachke SA, Higgins AW, Inagaki M, et al (2012a) The cell adhesion gene PVRL3 is associated with
572 congenital ocular defects. Hum Genet 131:235–250. https://doi.org/10.1007/s00439-011-
573 1064-z

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

576 Lachke SA, Maas RL (2011) RNA Granules and Cataract. Expert Rev Ophthalmol 6:497–500.
577 https://doi.org/10.1586/eop.11.53

578 Leamey CA, Merlin S, Lattouf P, et al (2007) Ten_m3 regulates eye-specific patterning in the
579 mammalian visual pathway and is required for binocular vision. PLoS Biol 5:e241.
580 https://doi.org/10.1371/journal.pbio.0050241

581 Leu S, Ouyang P (2006) Spatial and temporal expression profile of pinin during mouse development.
582 Gene Expr Patterns 6:620–631. https://doi.org/10.1016/j.modgep.2005.11.009

583 Liang Q, Dharmat R, Owen L, et al (2019) Single-nuclei RNA-seq on human retinal tissue provides
584 improved transcriptome profiling. Nat Commun 10:5743. https://doi.org/10.1038/s41467-
585 019-12917-9

586 Lillien L (1995) Changes in retinal cell fate induced by overexpression of EGF receptor. Nature
587 377:158–162. https://doi.org/10.1038/377158a0

588 Lim RR, Gupta S, Grant DG, et al (2018) Retinal Ultrastructural and Microvascular Defects in
589 Decorin Deficient (Dcn −/−) Mice. Microscopy and Microanalysis 24:1264–1265.
590 https://doi.org/10.1017/S1431927618006803

591 Liu C-Y, Birk DE, Hassell JR, et al (2003) Keratocan-deficient mice display alterations in corneal
592 structure. J Biol Chem 278:21672–21677. https://doi.org/10.1074/jbc.M301169200

593 Lokapally A, Metikala S, Hollemann T (2016) Expressional characterization of mRNA (guanine-7)
594 methyltransferase (rnmt) during early development of Xenopus laevis. Int J Dev Biol 60:65–
595 69. https://doi.org/10.1387/ijdb.150409th

596 Lorber B, Hendriks WJAJ, Van der Zee CEEM, et al (2005) Effects of LAR and PTP-BL
597 phosphatase deficiency on adult mouse retinal cells activated by lens injury. Eur J Neurosci
598 21:2375–2383. https://doi.org/10.1111/j.1460-9568.2005.04065.x

599 Lotery AJ, Baas D, Ridley C, et al (2006) Reduced secretion of fibulin 5 in age-related macular
600 degeneration and cutis laxa. Hum Mutat 27:568–574. https://doi.org/10.1002/humu.20344

601 Lu Q, Scott PA, Vukmanic EV, et al (2020) Yap1 is required for maintenance of adult RPE
602 differentiation. FASEB J 34:6757–6768. https://doi.org/10.1096/fj.201903234R

603 Lucey MM, Wang Y, Bustin M, Duncan MK (2008) Differential expression of the HMGN family of
604 chromatin proteins during ocular development. Gene Expr Patterns 8:433–437.
605 https://doi.org/10.1016/j.gep.2008.04.002

606 Lukowski SW, Lo CY, Sharov AA, et al (2019) A single-cell transcriptome atlas of the adult human
607 retina. The EMBO Journal 38:e100811. https://doi.org/10.15252/embj.2018100811

608 Maier T, Güell M, Serrano L (2009) Correlation of mRNA and protein in complex biological
609 samples. FEBS Letters 583:3966–3973. https://doi.org/10.1016/j.febslet.2009.10.036

610 Manthey AL, Lachke SA, FitzGerald PG, et al (2014) Loss of Sip1 leads to migration defects and
611 retention of ectodermal markers during lens development. Mech Dev 131:86–110.
612 https://doi.org/10.1016/j.mod.2013.09.005

613 Matalkah F, Jeong B, Sheridan M, et al (2022) The Musashi proteins direct post-transcriptional
614 control of protein expression and alternate exon splicing in vertebrate photoreceptors.
615 Commun Biol 5:1–15. https://doi.org/10.1038/s42003-022-03990-w

616 Mizukami M, Kanamoto T, Souchelnytskyi N, Kiuchi Y (2008) Proteome profiling of embryo chick
617 retina. Proteome Sci 6:3. https://doi.org/10.1186/1477-5956-6-3

618 Morimura H, Berson EL, Dryja TP (1999) Recessive mutations in the RLBP1 gene encoding cellular
619 retinaldehyde-binding protein in a form of retinitis punctata albescens. Invest Ophthalmol Vis
620 Sci 40:1000–1004

621 Naef V, De Sarlo M, Testa G, et al (2020) The Stemness Gene Mex3A Is a Key Regulator of
622 Neuroblast Proliferation During Neurogenesis. Front Cell Dev Biol 8:549533.
623 https://doi.org/10.3389/fcell.2020.549533

624 Nakazawa K, Shichino Y, Iwasaki S, Shiina N (2020) Implications of RNG140 (caprin2)-mediated
625 translational regulation in eye lens differentiation. J Biol Chem 295:15029–15044.
626 https://doi.org/10.1074/jbc.RA120.012715

627 Nesvizhskii AI, Aebersold R (2005) Interpretation of shotgun proteomic data: the protein inference
628 problem. Mol Cell Proteomics 4:1419–1440. https://doi.org/10.1074/mcp.R500012-MCP200

Padula SL, Anand D, Hoang TV, et al (2019) High-throughput transcriptome analysis reveals that the loss of Pten activates a novel NKX6-1/RASGRP1 regulatory module to rescue microphthalmia caused by Fgfr2-deficient lenses. Hum Genet 138:1391–1407. https://doi.org/10.1007/s00439-019-02084-8

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. https://doi.org/10.1007/s00439-016-1747-6

Patel SD, Anand D, Motohashi H, et al (2022) Deficiency of the bZIP transcription factors Mafg and Mafk causes misexpression of genes in distinct pathways and results in lens embryonic developmental defects. Front Cell Dev Biol 10:981893. https://doi.org/10.3389/fcell.2022.981893

Perez-Riverol Y, Bai J, Bandla C, et al (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res 50:D543–D552. https://doi.org/10.1093/nar/gkab1038

Pierpont ME, Richards M, Engel WK, et al (2017) Retinal dystrophy in two boys with Costello syndrome due to the HRAS p.Gly13Cys mutation. Am J Med Genet A 173:1342–1347. https://doi.org/10.1002/ajmg.a.38110

Piri N, Kwong JMK, Caprioli J (2013) Crystallins in retinal ganglion cell survival and regeneration. Mol Neurobiol 48:819–828. https://doi.org/10.1007/s12035-013-8470-2

Pras E, Kristal D, Shoshany N, et al (2015) Rare genetic variants in Tunisian Jewish patients suffering from age-related macular degeneration. J Med Genet 52:484–492. https://doi.org/10.1136/jmedgenet-2015-103130

Raffa L, Matton M-P, Michaud J, et al (2017) Optic nerve hypoplasia in a patient with a de novo KIF1A heterozygous mutation. Can J Ophthalmol 52:e169–e171. https://doi.org/10.1016/j.jcjo.2017.02.021

Ratnapriya R, Sosina OA, Starostik MR, et al (2019) Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. Nat Genet 51:606–610. https://doi.org/10.1038/s41588-019-0351-9

Ratnapriya R, Zhan X, Fariss RN, et al (2014) Rare and common variants in extracellular matrix gene Fibrillin 2 (FBN2) are associated with macular degeneration. Hum Mol Genet 23:5827–5837. https://doi.org/10.1093/hmg/ddu276

Ribic A, Liu X, Crair MC, Biederer T (2014) Structural organization and function of mouse photoreceptor ribbon synapses involve the immunoglobulin protein synaptic cell adhesion molecule 1. J Comp Neurol 522:900–920. https://doi.org/10.1002/cne.23452

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

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

Shao M, Lu T, Zhang C, et al (2020) Rbm24 controls poly(A) tail length and translation efficiency of crystallin mRNAs in the lens via cytoplasmic polyadenylation. Proc Natl Acad Sci USA 117:7245–7254. https://doi.org/10.1073/pnas.1917922117

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. https://doi.org/10.1371/journal.pgen.1007278

Stone EM, Braun TA, Russell SR, et al (2004) Missense variations in the fibulin 5 gene and age-related macular degeneration. N Engl J Med 351:346–353. https://doi.org/10.1056/NEJMoa040833

Sundar J, Matalkah F, Jeong B, et al (2020) The Musashi proteins MSI1 and MSI2 are required for photoreceptor morphogenesis and vision in mice. J Biol Chem 296:100048. https://doi.org/10.1074/jbc.RA120.015714

Tacheny A, Dieu M, Arnould T, Renard P (2013) Mass spectrometry-based identification of proteins interacting with nucleic acids. J Proteomics 94:89–109. https://doi.org/10.1016/j.jprot.2013.09.011

Vig A, Poulter JA, Ottaviani D, et al (2020) DYNC2H1 hypomorphic or retina-predominant variants cause nonsyndromic retinal degeneration. Genet Med 22:2041–2051. https://doi.org/10.1038/s41436-020-0915-1

Wang J, Lin J, Kaiser U, et al (2017) Absence of macrophage migration inhibitory factor reduces proliferative retinopathy in a mouse model. Acta Diabetol 54:383–392. https://doi.org/10.1007/s00592-016-0956-8

Wang X, Tong Y, Giorgianni F, et al (2010) Cellular retinol binding protein 1 modulates photoreceptor outer segment folding in the isolated eye. Dev Neurobiol 70:623–635. https://doi.org/10.1002/dneu.20798

Watson LA, Wang X, Elbert A, et al (2014) Dual effect of CTCF loss on neuroprogenitor differentiation and survival. J Neurosci 34:2860–2870. https://doi.org/10.1523/JNEUROSCI.3769-13.2014

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. https://doi.org/10.1007/s12177-009-9042-6

Wolf J, Boneva S, Schlecht A, et al (2022) The Human Eye Transcriptome Atlas: A searchable comparative transcriptome database for healthy and diseased human eye tissue. Genomics 114:110286. https://doi.org/10.1016/j.ygeno.2022.110286

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. https://doi.org/10.1093/nar/gkt824

Xi J, Farjo R, Yoshida S, et al (2003) A comprehensive analysis of the expression of crystallins in mouse retina. Mol Vis 9:410–419

Xu Q, Bai Y, Huang L, et al (2015) Knockout of αA-crystallin inhibits ocular neovascularization. Invest Ophthalmol Vis Sci 56:816–826. https://doi.org/10.1167/iovs.14-14734

Yang J, Gao J, Adamian M, et al (2005) The ciliary rootlet maintains long-term stability of sensory cilia. Mol Cell Biol 25:4129–4137. https://doi.org/10.1128/MCB.25.10.4129-4137.2005

Yu Q, Lin B, Xie S, et al (2018) A homozygous mutation p.Arg2167Trp in FREM2 causes isolated cryptophthalmos. Hum Mol Genet 27:2357–2366. https://doi.org/10.1093/hmg/ddy144

Zhang X, Wang D, Dongye M, et al (2019) Loss-of-function mutations in FREM2 disrupt eye morphogenesis. Exp Eye Res 181:302–312. https://doi.org/10.1016/j.exer.2019.02.013

Zhao J, Izumi T, Nunomura K, et al (2007) MARCKS-like protein, a membrane protein identified for its expression in developing neural retina, plays a role in regulating retinal cell proliferation. Biochem J 408:51–59. https://doi.org/10.1042/BJ20070826

718 **Table 1. Top 90 proteins with enriched expression in mouse E14.5 retina and retinal pigment**
719 **epithelium combined tissue as compared to WB**
720

Rank	UniProt Gene Name	Uniprot Accession	Primary Protein Name	Associated Retina or Eye Defects	Reference
1	Aldh1a1	P24549	Retinal dehydrogenase 1 (RALDH 1; RalDH1)	Expressed throughout development in dorsal retina; Choroidal hypoplasia in mice*	(Fan et al. 2003; Goto et al. 2018; Duester 2022)
2	Tyr	P11344	Tyrosinase	Retinal defects*	(Jeffery et al. 1994, 1997)
3	Kera	O35367	Keratocan (KTN)	-	
4	Pmel	Q60696	Premelanosome protein	RPE menalonosme cellular defect*	(Hellström et al. 2011)
5	Cryge	Q03740	Gamma-crystallin E	-	
6	Crybb3	Q9JJU9	Beta-crystallin B3, N-terminally processed	-	
7	Fras1	Q80T14	Extracellular matrix protein FRAS1	-	
8	Hmcn1	D3YXG0	Hemicentin-1	AMD*	(Pras et al. 2015)
9	Crybb1	Q9WVJ5	Beta-crystallin B1B	-	
10	Nfatc4	Q8K120	Nuclear factor of activated T-cells, cytoplasmic 4 (NF-ATc4; NFATc4)	-	
11	Mylk	Q6PDN3	Myosin light chain kinase, smooth muscle, deglutamylated form	-	
12	Rlbp1	Q9Z275	Retinaldehyde-binding protein 1	Retinitis punctatta albescens*	(Morimura et al. 1999)
13	Specc1l	Q2KN98	Cytospin-A	-	
14	Igdcc4	Q9EQS9	Immunoglobulin superfamily DCC subclass member 4	-	

15	Crym	O54983	Ketimine reductase mu-crystallin	-	
16	Epb41l5	Q8BGS1	Band 4.1-like protein 5	-	
17	Lig3	P97386	DNA ligase 3	-	
18	Champ1	Q8K327	Chromosome alignment-maintaining phosphoprotein 1	-	
19	Tes	P47226	Testin	-	
20	Mdc1	Q5PSV9	Mediator of DNA damage checkpoint 1	-	
21	Hras	Q61411	GTPase HRas, N-terminally processed	Retinal dystrophy*	(Pierpont et al. 2017)
22	Lama5	Q61001	Laminin subunit alpha-5	Retinal defects*	(Jones et al. 2020)
23	Frem2	Q6NVD0	FRAS1-related extracellular matrix protein 2	Detected in outer plexiform layer of the retina; mutations associated with Cryptophthalmos	(Zhang et al. 2019)
24	Mfap2	P55002	Microfibrillar-associated protein 2 (MFAP-2)	-	
25	Smarcd1	Q61466	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1	-	
26	Egfr	Q01279	Epidermal growth factor receptor	Retinal cell fate determination*	(Lillien 1995)
27	Heph	Q9Z0Z4	Hephaestin	Retinal defects*	(Hahn et al. 2004)
28	Rai14	Q9EP71	Ankycorbin	-	
29	Tenm3	Q9WTS6	Teneurin-3 (Ten-3)	Retinal defect in mouse and zebrafish*	(Leamey et al. 2007) (Antinucci et al. 2013)
30	Nelfe	P19426	Negative elongation factor E (NELF-E)	-	

31	Znf326	O88291	Zinc finger protein 326	-	
32	Crocc	Q8CJ40	Rootletin	Retinal defects*	(Yang et al. 2005)
33	Cadm1	Q8R5M8	Cell adhesion molecule 1	Retinal defects*	(Ribic et al. 2014)
34	Fbln5	Q9WVH9	Fibulin-5 (FIBL-5)	Retinal defects*	(Lotery et al. 2006) (Stone et al. 2004)
35	Wiz	O88286	Widely-interspaced zinc finger motifs	-	
36	Cryaa	P24622	Alpha-crystallin A chain	Retinal neovascularization defect*	(Xu et al. 2015)
37	Znf219	Q6IQX8	Zinc finger protein 219	-	
38	Gldc	Q91W43	Glycine dehydrogenase (decarboxylating), mitochondrial	-	
39	Mex3a	G3UYU0	Mex3 RNA-binding family member A	-	
40	Nub1	P54729	NEDD8 ultimate buster 1	-	
41	Uncharacterized protein FLJ45252 homolog	Q6PIU9	Uncharacterized protein FLJ45252 homolog	-	
42	Kif1a	P33173	Kinesin-like protein KIF1A	-	
43	Dync2h1	Q45VK7	Cytoplasmic dynein 2 heavy chain 1	Retinal defects*	(Vig et al. 2020)
44	Gja1	P23242	Gap junction alpha-1 protein	Retinal and other eye defects*	(Gabriel et al. 2011)
45	Fbn2	Q61555	Fibrillin-2 C-terminal peptide	AMD*	(Ratnapriya et al. 2014)
46	Pcca	Q91ZA3	Propionyl-CoA carboxylase alpha chain,	-	

			mitochondrial (PCCase subunit alpha)		
47	Zmym4	A2A791	Zinc finger MYM-type protein 4	-	
48	Emilin3	P59900	EMILIN-3	-	
49	Tbl1x	Q9QXE7	Transducin (beta)-like 1 X-linked	-	
50	Bcat1	P24288	Branched-chain-amino-acid aminotransferase, cytosolic (BCAT(c))	-	
51	Niban1	Q3UW53	Protein Niban 1	-	
52	Ehmt2	Q9Z148	Histone-lysine N-methyltransferase EHMT2	-	
53	Brd7	O88665	Bromodomain-containing protein 7	-	
54	Taf15	Q8BQ46	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor	-	
55	Hspa4l	P48722	Heat shock 70 kDa protein 4L	-	
56	Msi1	Q61474	RNA-binding protein Musashi homolog 1 (Musashi-1)	Retinal defects*	(Sundar et al. 2020)
57	Aldh5a1	Q8BWF0	Succinate-semialdehyde dehydrogenase, mitochondrial	-	
58	Strn	O55106	Striatin	-	
59	Dido1	Q8C9B9	Death-inducer obliterator 1 (DIO-1)	-	
60	Ctcf	Q61164	Transcriptional repressor CTCF	Retinal defects*	(Watson et al. 2014)

61	Sptbn2	Q68FG2	Spectrin beta chain	-	
62	Rnmt	Q9D0L8	mRNA cap guanine-N7 methyltransferase	-	
63	Cux1	P53564	Homeobox protein cut-like 1	-	
64	Ephb2	P54763	EphB2/CTF2	Retinal defects*	(Fu and Sretavan 2012)
65	Golga5	Q9QYE6	Golgin subfamily A member 5	-	
66	Golgb1	E9PVZ8	Golgi autoantigen, golgin subfamily b, macrogolgin 1	-	
67	Ank2	Q8C8R3	Ankyrin-2 (ANK-2)	Retinal defects*	(Kizhatil et al. 2009b)
68	Snx6	Q6P8X1	Sorting nexin-6, N-terminally processed	-	
69	Srsf2	Q62093	Serine/arginine-rich splicing factor 2	-	
70	Smarce1	O54941	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1	-	
71	Hmgn1	P18608	Non-histone chromosomal protein HMG-14	-	
72	Ptprf	A2A8L5	Receptor-type tyrosine-protein phosphatase F	-	
73	Dnajc7	Q9QYI3	DnaJ homolog subfamily C member 7	-	
74	Pnn	O35691	Pinin	-	
75	Mccc2	Q3ULD5	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial (MCCase subunit beta)	-	

76	Nipbl	Q6KCD5	Nipped-B-like protein	-	
77	Ank3	G5E8K5	Ankyrin-3 (ANK-3)	Retinal defects*	(Kizhatil et al. 2009a)
78	Rbp1	Q00915	Retinol-binding protein 1	Retinal defects*	(Wang et al. 2010)
79	Yap1	P46938	Transcriptional coactivator YAP1 (Yes-associated protein 1)	Retinal defects*	(Lu et al. 2020)
80	Smarcc1	P97496	SWI/SNF complex subunit SMARCC1	-	
81	Utrn	E9Q6R7	Utrophin	-	
82	Cald1	E9QA15	Caldesmon 1	-	
83	Ccar1	Q8CH18	Cell division cycle and apoptosis regulator protein 1	-	
84	Cttn	Q60598	Src substrate cortactin	-	
85	Prrc2c	Q3TLH4	Protein PRRC2C	-	
86	Arid1a	A2BH40	AT-rich interactive domain-containing protein 1A (ARID domain-containing protein 1A)	-	
87	Cbx1	P83917	Chromobox protein homolog 1	-	
88	Dcn	P28654	Decorin	Retinal defects*	(Lim et al. 2018)
89	Mif	P34884	Macrophage migration inhibitory factor (MIF)	Associated with proliferative retinopathy*	(Wang et al. 2017)
90	Igf2bp1	O88477	Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2 mRNA-binding protein 1; IMP-1)	Retinal defects in Zebrafish*	(Gaynes et al. 2015b, p. 1)

721

Figure Legends

Fig. 1. Workflow of the experimental strategy to generate MS/MS protein profile of the mouse embryonic retina and retinal pigment epithelium combined tissue. (A) Mouse eyes at embryonic day (E)14.5 were isolated, and the retina and retinal pigment epithelium combined tissue (termed retina) was micro-dissected. The whole body (WB) with eye tissue removed was processed similarly and used as reference for differential protein expression analysis. Retina and WB samples (n = 5 for each sample type, 55 µg protein per sample) were subjected to high-throughput tandem mass spectrometry (MS/MS). (B) The workflow for differential protein expression analysis is outlined. The edgeR pipeline was used to determine differential protein expression using normalized spectral counts. Proteins passing stringency criteria of ≥2.5 average spectral counts, ≥2.0 fold-change (in retina, compared to WB), False Discovery Rate <0.01 were considered to have enriched expression in the retina.

Fig. 2. Quality assessment of MS/MS data. (A) TMM (trimmed mean of *M*-values) normalization of spectral counts in WB and retina samples using edgeR to correct for the dramatic compositional differences. The retina and the WB samples showed comparable median SpCs in boxplots (TMM normalized SpC are shown in *y*-axis). (B) Individual biological replicates of the retina and WB samples clustered together while the overall retina and WB samples clustered separately from each other in Multidimensional scaling analysis (leading dimensions 1 and 2 are represented by the axes). (C) Sample-to-sample consistency was examined by generating a scatter matrix for the five WB samples and (D) the five retina samples. (E) A scatter plot with regression analysis shows no correlation *(r* = 0.81) between the average retina and average WB samples.

Fig. 3. Gene ontology (GO) analysis of proteins expressed in the E14.5 retina and retinal pigment epithelium combined tissue. Proteins expressed in the retina 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 identified candidates representing several GO terms that may be relevant to retinal biology, including those involved in various molecular, cellular and physiological processes. The *x*-axis represents the number of protein candidates identified in the specific GO term shown on the *y*-axis.

Fig. 4. *In silico* WB-subtraction identifies candidates with enriched expression in the mouse embryonic retina and retinal pigment epithelium combined tissue. (A) Proteins with the average SpC ≥ 2.5 between retina and WB samples (*n*=2675) were further processed for identifying differentially expressed candidates. This analysis showed that 90 proteins were enriched in retina compared to WB (≥ 2.0 fold-change, FDR <0.01 cut-off). (B) MA plot (M = log ratio of retina to WB, A = average intensity) representation of differential protein expression profiling wherein the "high" (red, circle, FDR < 0.01), "medium" (green, triangle, 0.01 < FDR < 0.05), "low" probability retina-enriched (blue, square, 0.05 < FDR < 0.1) and non-enriched candidates (magenta, cross, 0.1 < FDR) are indicated. Several candidates associated with retinal defects (*) can be identified in this plot.

Fig. 5. Prioritization of candidates associated with retina defects by *in silico* WB-subtraction. (A) Comparison of top 30 retina expressed proteins with the top 30 retina "enriched-expression" proteins shows that 1/3rd of the candidates in the "enriched-expression" category are associated with the retinal defects (highlighted in grey) and are not amongst the top proteins expressed in the retina. (B) Aldh1a1, Tyr, Pmel, Hmcn1, Rlbp1, Hras, Lama5, Egfr, Heph and Tenm3 associated with retinal defects that were not present in the top 30 "expressed" candidates show significant (p<0.001) enrichment in the retina compared to WB, demonstrating the effectiveness of the *in silico* WB-subtraction strategy in identifying these candidates from 2,675 expressed proteins. The average SpC of individual proteins are shown in *y*-axis.

768 **Fig. 6. Gene ontology (GO) analysis of proteins with enriched expression in the E14.5 retina and**
769 **retinal pigment epithelium combined tissue.** The 90 proteins identified to exhibit "enriched
770 expression" in the mouse retina and retinal pigment epithelium combined tissue were analyzed by the
771 Database for Annotation, Visualization and Integrated Discovery (DAVID v6 .8) for functional
772 clustering and annotation based on gene ontology (GO) categories. This analysis identified candidates
773 representing several GO terms that are relevant to retina biology, including "nervous system
774 development", "positive regulation of MAP kinase activity", "chromatin remodeling" and "cell
775 adhesion". The *x*-axis represent the number of protein candidates identified in the specific GO term
776 shown on the *y*-axis.

777 **Fig. 7. iSyTE allows ready visualization of proteins expressed or enriched-expressed in the retina**
778 **and retinal pigment epithelium combined tissue.** The retina protein expression and enriched-
779 expression data can be visualized through the iSyTE web-resource tool at
780 https://research.bioinformatics.udel.edu/iSyTE/. (A) On the iSyTE main webpage, navigate to
781 "Proteome Retina-enrichment", select the Human hg38 or Mouse mm39 assembly on the UCSC
782 Genome Browser using iSyTE tracks, and input the protein candidate of interest to visualize its
783 expression or enriched expression in the retina. The heat-map color key can be used to estimate the
784 retina protein expression or enriched-expression. (B) As example, visualization of the expression and
785 enriched expression of Hras and Tyr proteins in the retina are shown.
786

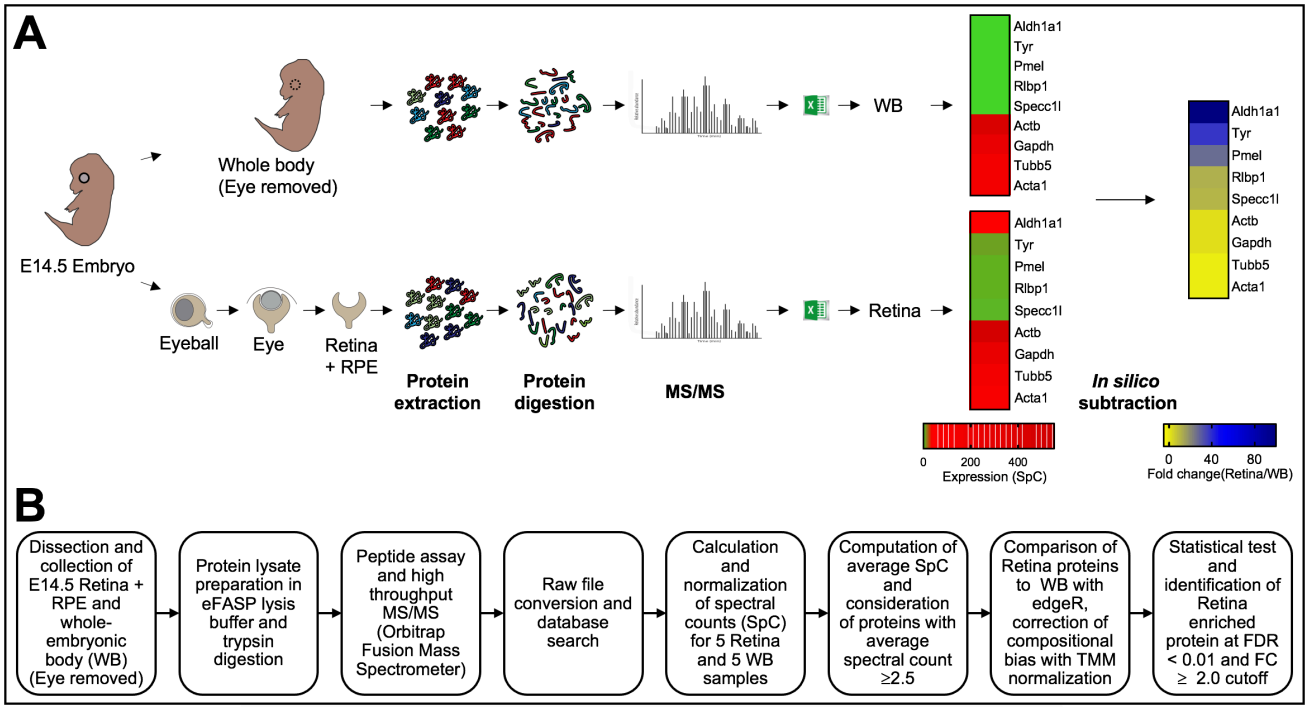
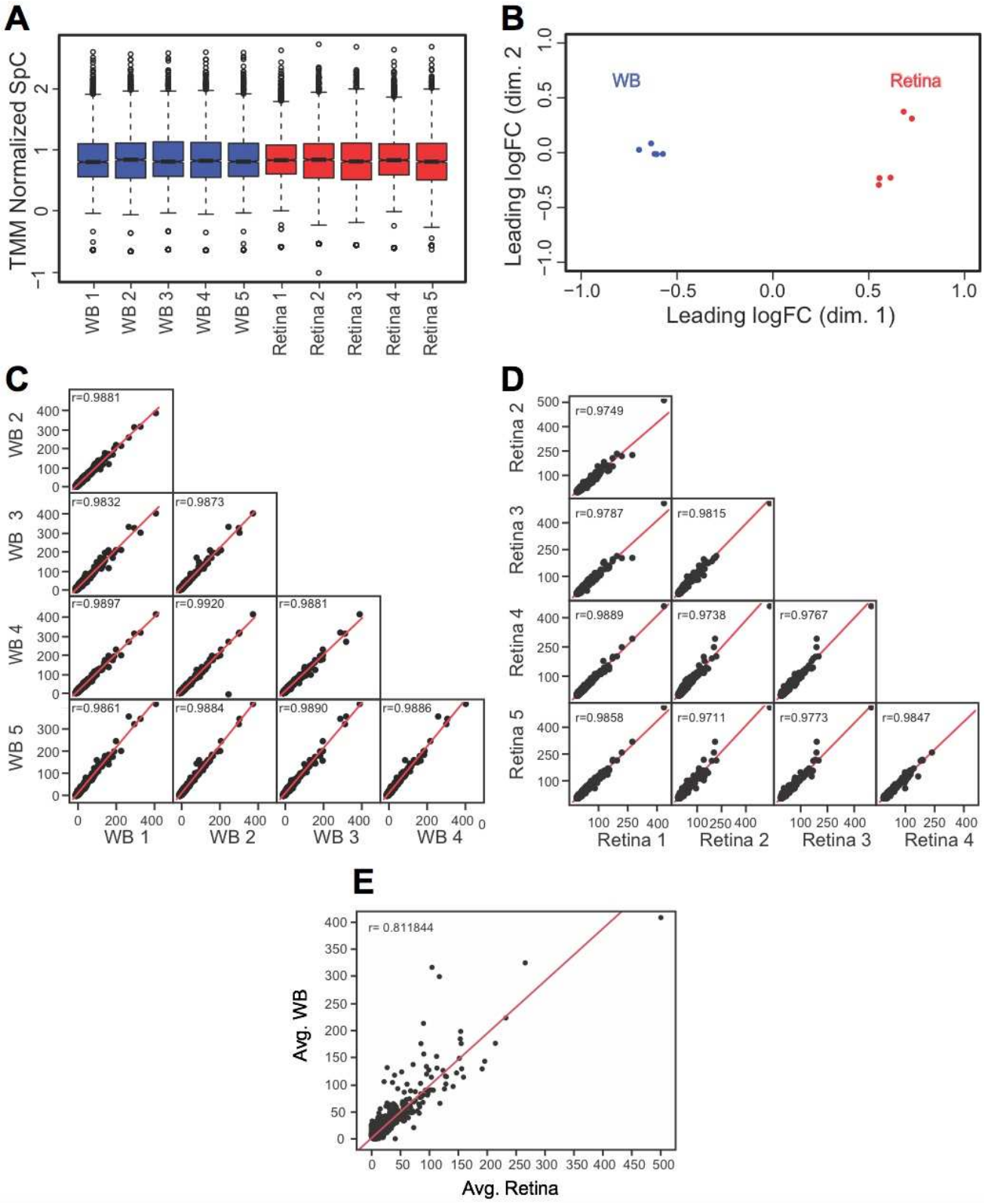


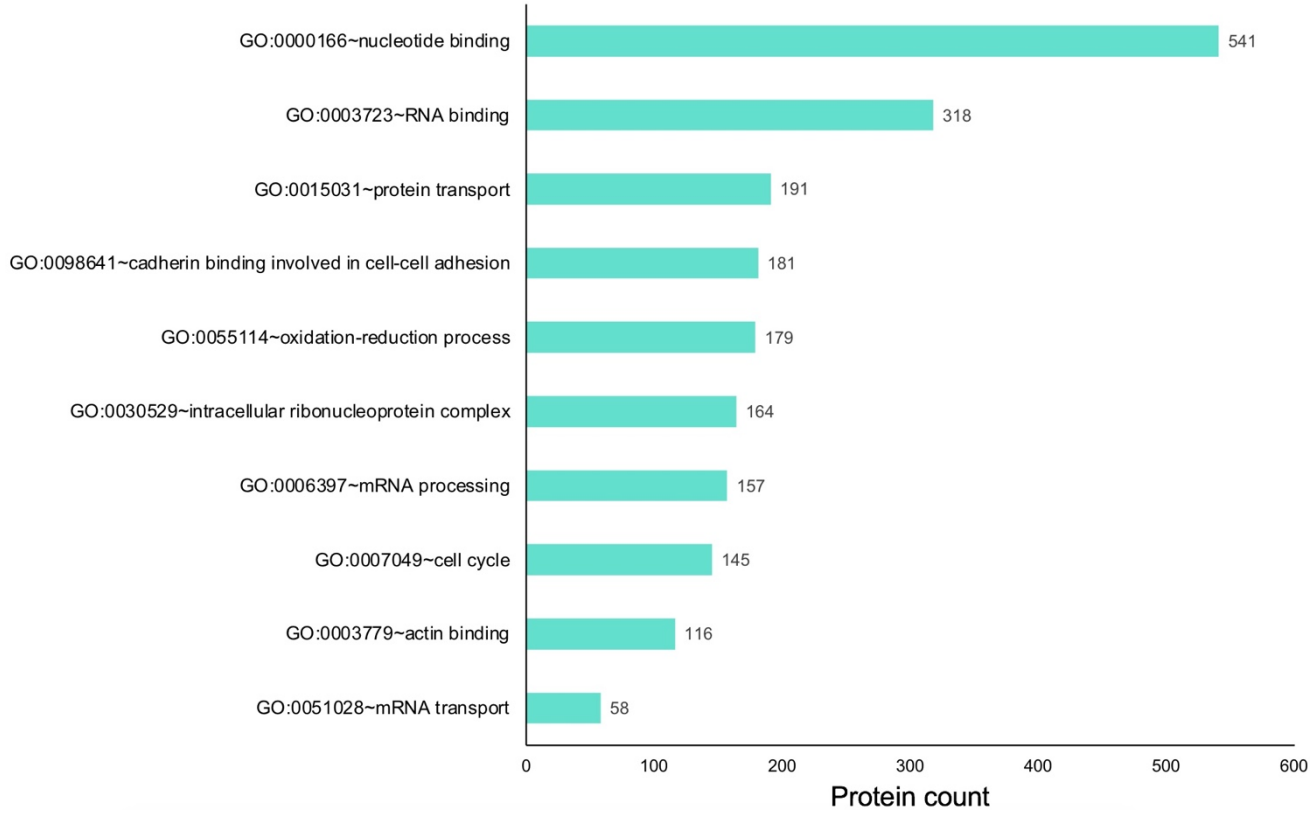
Fig. 1. Workflow of the experimental strategy to generate MS/MS protein profile of the mouse embryonic retina and retinal pigment epithelium combined tissue.



792

793 **Fig. 2. Quality assessment of MS/MS data.**

794



795

796 **Fig. 3. Gene ontology (GO) analysis of proteins expressed in the E14.5 retina and retinal pigment**

797 **epithelium combined tissue.**

798

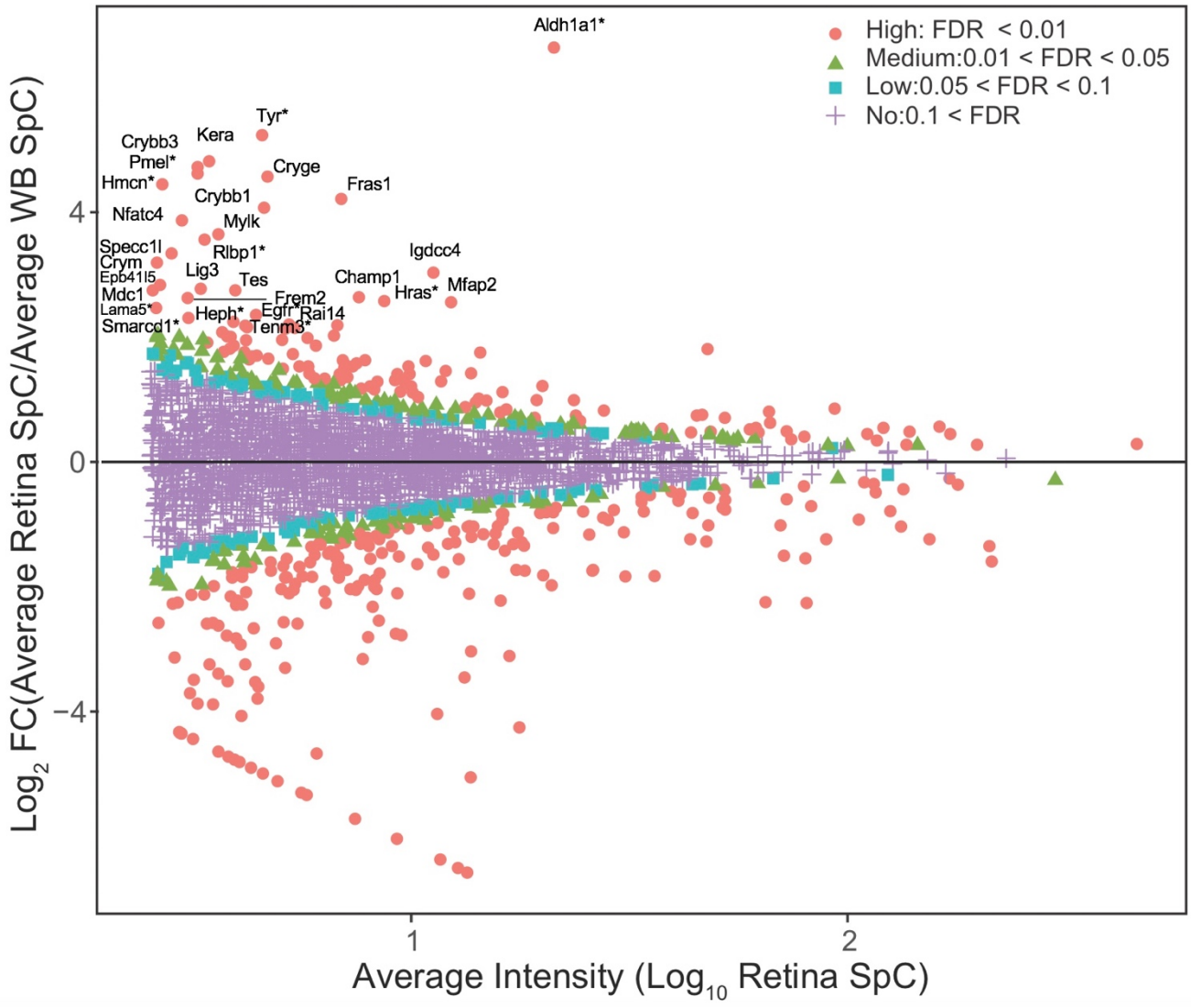


Fig. 4. ***In silico*** **WB-subtraction identifies candidates with enriched expression in the mouse embryonic retina and retinal pigment epithelium combined tissue.**

Top 30 Expressed	Top 30 enriched
Actb	Aldh1a1
Eef1a1	Tyr
Gapdh	Kera
Hsp90ab1	Pmel
Flna	Cryge
Eno1	Crybb3
Myh10	Fras1
Ahnak	Hmcn1
Myh3	Crybb1
Eef2	Nfatc4
Dync1h1*	Mylk
Hspa8	Rlbp1
Col6a3*	Specc1l
Fn1	Igdcc4
Hsp90aa1*	Crym
Sptan1	Epb41l5
Myh9	Lig3
Pkm*	Champ1
Atp5b	Tes
Col12a1	Mdc1
Hbb-bs	Hras
Tln1*	Lama5
Tubb5	Frem2
Vcp	Mfap2
Afp	Smarcd1
Hspd1	Egfr
Vim*	Heph
Hsp90b1	Rai14
Acta1	Tenm3
Flnb	Nelfe

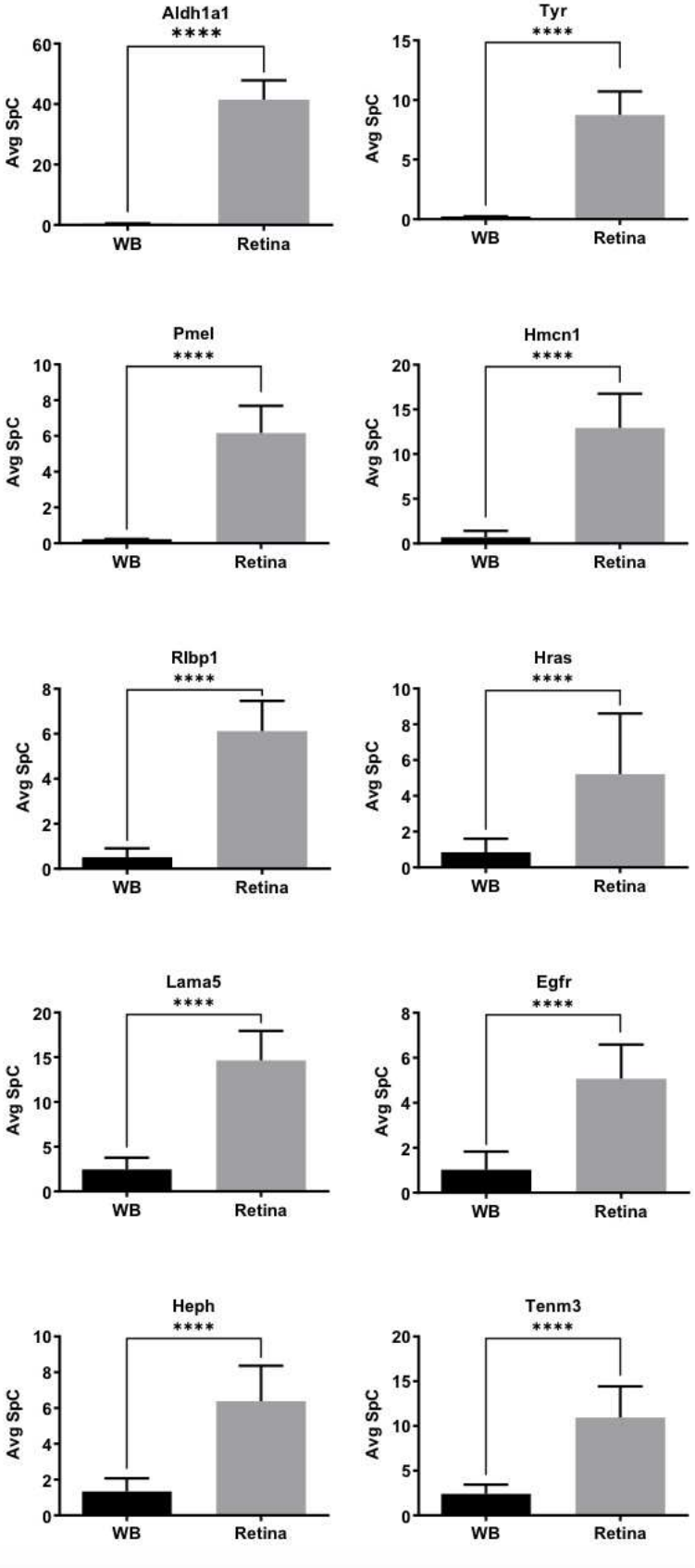


Fig. 5. Prioritization of candidates associated with retina defects by *in silico* WB-subtraction.

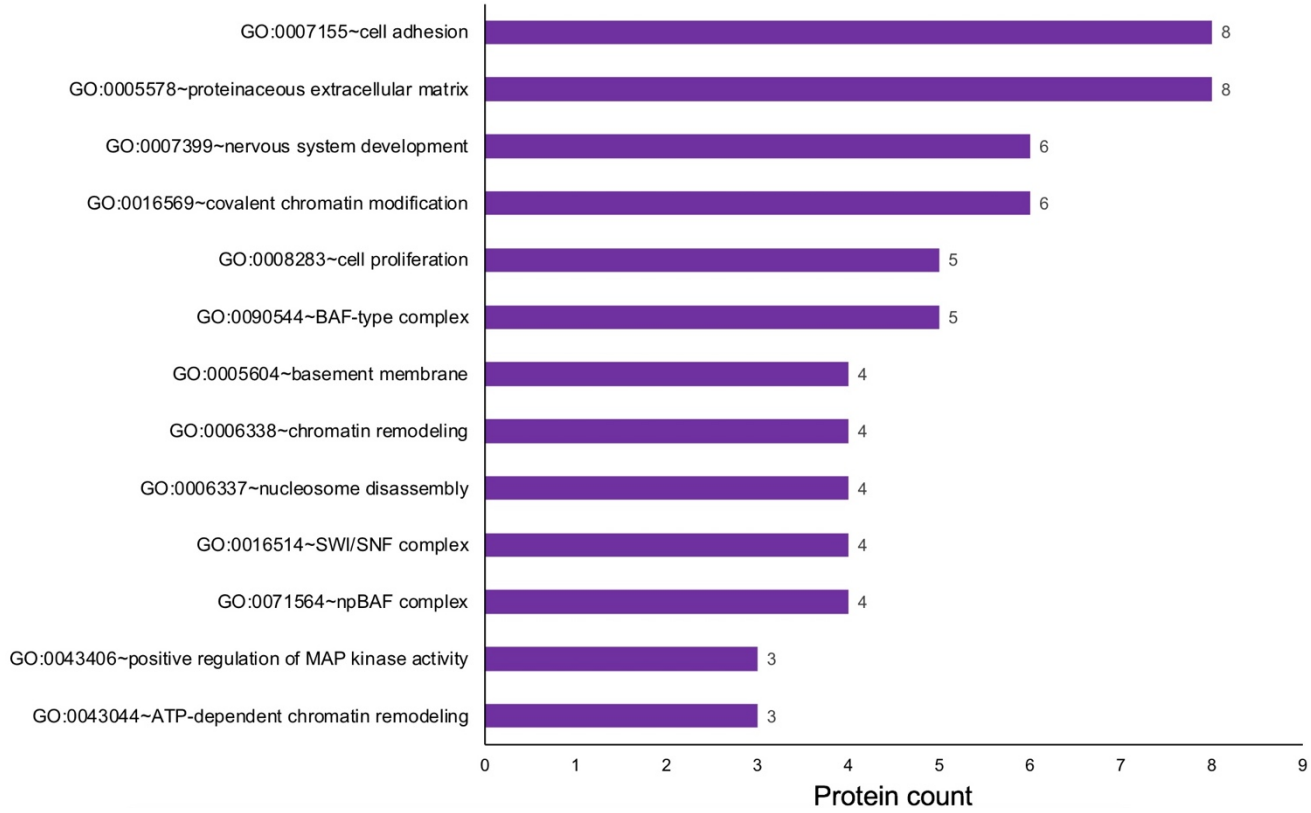


Fig. 6. Gene ontology (GO) analysis of proteins with enriched expression in the E14.5 retina and retinal pigment epithelium combined tissue.

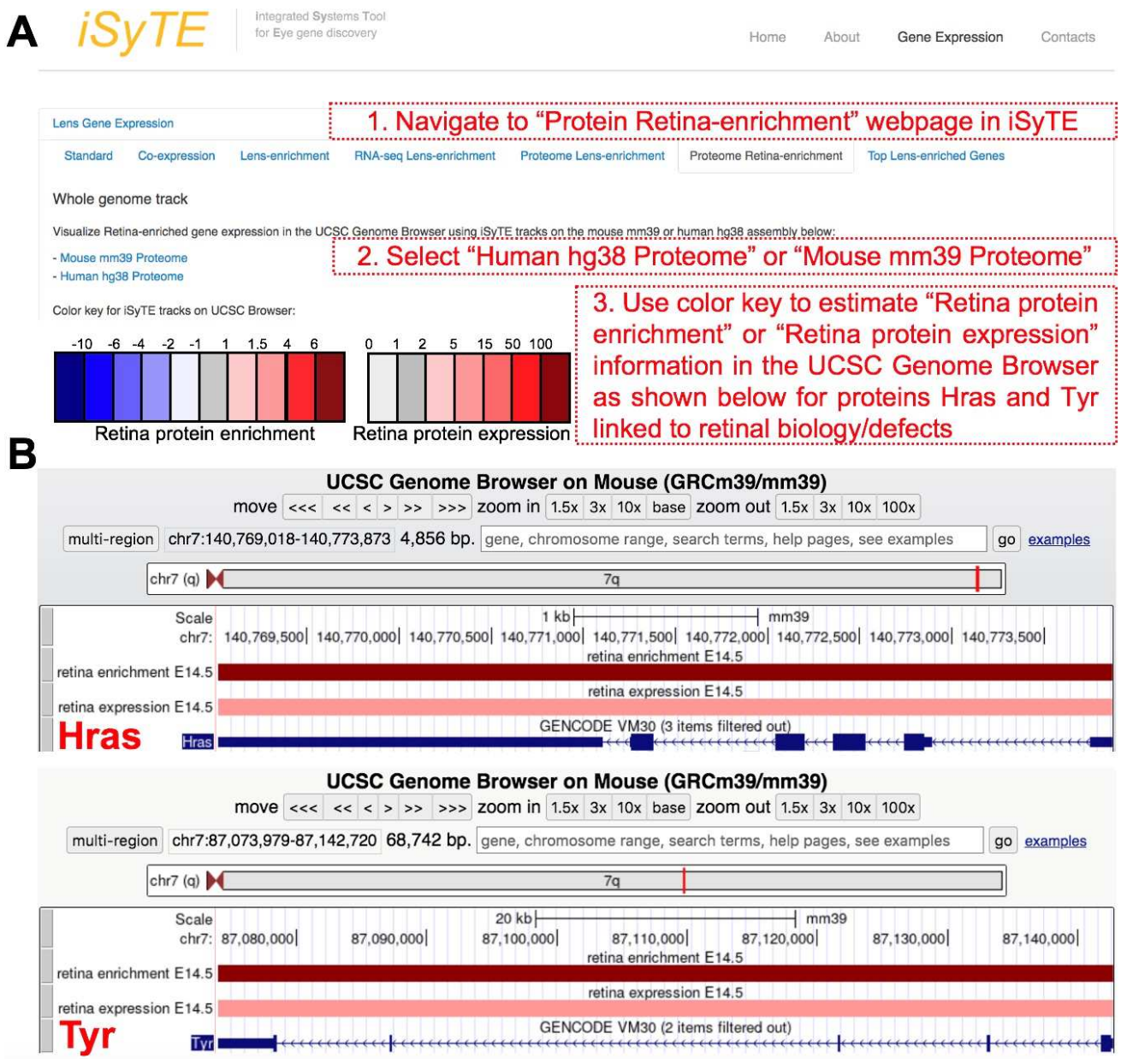


Fig. 7. iSyTE allows ready visualization of proteins expressed or enriched-expressed in the retina and retinal pigment epithelium combined tissue.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Aryal2023SupplementaryTableS1.xlsx
- Aryal2023SupplementaryTableS2.xlsx
- Aryal2023SupplementaryTableS3.xlsx
- Aryal2023SupplementaryTableS4.xlsx