Integrated Comparison of GWAS, Transcriptome, and Proteomics Studies Highlights Similarities in the Biological Basis of Animal and Human Myopia

Nina Riddell and Sheila G. Crewther

School of Psychology and Public Health, La Trobe University, Melbourne, Victoria, Australia

Correspondence: Sheila G. Crewther, School of Psychology and Public Health, La Trobe University, Melbourne, Victoria 3086, Australia; s.crewther@latrobe.edu.au.

Submitted: August 26, 2016
Accepted: December 21, 2016
Citation: Riddell N, Crewther SG. Integrated comparison of GWAS, transcriptome, and proteomics studies highlights similarities in the biological basis of animal and human myopia. Invest Ophthalmol Vis Sci. 2017;58:660–669. DOI:10.1167/iovs.16-20618

PURPOSE. To identify commonalities between the genes in close proximity to genome-wide association study (GWAS) refractive error and axial length loci, and the genes and proteins differentially expressed in animal models of optically induced refractive error.

METHODS. The GWAS catalog was searched for loci significantly ($P \leq 5 \times 10^{-8}$) associated with refractive error or axial length. PubMed was searched for exploratory animal transcriptome and proteomics studies of optically induced refractive error. A total of 15 GWAS, 7 transcriptome, and 9 proteomics studies met inclusion criteria. Ensembl's BioMart was used to identify human orthologs for the differentially expressed genes and proteins from animal studies. These orthologs were then compared to the protein-coding genes within 1 megabase (Mb), 500 kilobases (kb), and 250 kb of human GWAS loci by using the GeneOverlap R package, and Benjamini-Hochberg-adjusted $P$ values and odds ratios (ORs) were calculated for each intersection.

RESULTS. The genes near human GWAS loci overlapped significantly with the genes downregulated during early myopia induction in animals ($1Mb$: OR = 1.56, $P = 0.025$; 500 kb: OR = 1.92, $P = 0.010$; 250 kb: OR = 2.33, $P = 0.010$). There was also significant overlap between the genes and proteins differentially expressed in late myopia (OR = 4.12, $P = 0.018$). When animal study results were segregated by methodologic parameters, GWAS candidate genes overlapped significantly with the genes differentially expressed at early (OR = 1.50, $P = 0.010$) but not late (OR = 1.04, $P = 0.684$) induction time-points. Gene and protein expression responses also appeared well conserved across model species, and there was no evidence of greater GWAS-transcriptome concordance in similar species to humans (e.g., primates or mammals).

CONCLUSIONS. These findings suggest that genetic and environmental factors control ocular growth via similar biological pathways across species, and support the continued use of animal models for investigating the biological mechanisms underlying human myopia development.

Keywords: GWAS, microarray, proteomics, myopia, animal models

Myopia (short-sightedness) occurs when the eye grows too long for its refractive power, such that distant visual scenes are focused in front of the neural retina. Currently ~1.45 billion people worldwide are myopic, and prevalence and severity are increasing in many urban communities. Individuals showing moderate to high levels (4–6 diopters [D]) of myopic error are at significantly increased risk of vision loss from a range of secondary pathologies, making the development of treatments to limit excessive ocular growth an important global health, quality of life, and socioeconomic priority.

Ocular growth in humans and other animals is controlled by complex gene–environment interactions. Genetic susceptibility has primarily been investigated in humans, initially using candidate gene and family-based linkage approaches and more recently, using hypothesis-free genome-wide association studies (GWAS) with greater power to dissect the genetic underpinnings of common complex traits. These genetic studies have identified a range of loci associated with high myopia, or the entire spectrum of refractive error. Concurrently, the role of visual experience in adapting ocular growth to suit the environment has been extensively investigated in animals where myopia can be induced by rearing with form deprivation occluders or negatively powered defocusing lenses. Transcriptome and proteomics studies of ocular expression profiles in these animal models have associated thousands of genes and hundreds of proteins with environmentally driven growth. Enrichment analyses have aided interpretation of the wealth of data generated by exploratory human and animal studies; however, the key pathways by which genes and environment control ocular growth remain largely unclear.

Thus, further cross-methodology input from human and animal studies will be needed to improve understanding of the biological basis of myopia and identify therapeutic targets. This approach assumes that the biological mechanisms controlling eye growth are similar across species. To date, assessments of cross-species conservation have been hindered by the hetero-
genic methodologies used to investigate myopia etiology in humans and animals (i.e., human GWAS have implicated genes containing genetic variants that influence myopia risk, while animal transcriptome and proteomics studies have implicated genes and proteins whose expression levels change during environmentally driven myopia induction). Despite these differences, several studies have identified commonalities between the genes near human refractive error loci and the genes differentially expressed in environmental animal models. The only systematic investigation of cross-species conservation to date has been conducted by Stambolian, who compared the genes within 1 megabase (Mb) of peaks from the CREAM refractive error meta-analysis with the genes implicated in chick lens-induced hyperopia, primate occlusion-induced myopia, and yeast growth. The GWAS candidate genes share more commonalities with yeast growth than the refractive error models, leading Stambolian to conclude that the biological pathways involved in human and animal myopia must be different.

Opportunities now exist to expand and improve on this initial cross-species comparison. Most notably, Stambolian did not assess the statistical significance of gene overlaps; this has important implications, as the likelihood of commonalities occurring by chance between large lists of genes is far higher than that for small lists. Moreover, as discussed above, Stambolian focused on the results of one GWAS meta-analysis and two animal transcriptome studies. The proliferation of exploratory human and animal studies in recent years (e.g., see Supplementary Table S1) offers the opportunity to expand the comparison to include a further 14 GWAS, 5 transcriptome, and 9 proteomics studies. The transcriptome and proteomics studies now available also encompass a wider range of animal models including six species, five optical manipulations, and multiple tissue types and time-points. With this diversity, it is possible to assess whether some animal paradigms show greater concordance with the genes near human GWAS loci. Thus, here we aimed to systematically assess whether the same genes are associated with refractive error loci in humans and optically induced refractive change in animal models. We compared all genes within 1 Mb of GWAS loci for refractive error and axial length with the genes and proteins differentially expressed in all exploratory animal transcriptome and proteomics studies of refractive error published by July 2016. In our initial comparison, we assessed the statistical significance of overlaps between GWAS candidate genes and the genes and proteins differentially expressed in early and late animal models of refractive error. One day was chosen as the cutoff for “early” expression changes, as this time-point provided the best split between the acute induction period when refractive errors were likely to be small to moderate and the chronic induction period when refractive errors were likely to be moderate to severe (see Gene List Generation in the Methods section). We also assessed the effects of different animal model paradigms by comparing GWAS candidate genes with animal study results segregated on the basis of four methodologic parameters: time-point, species, tissue, and optical manipulation.

Methods

Inclusion Criteria and Gene List Generation

Human Genome-Wide Association Studies. The NHGRI-EBI catalog of published GWAS is a quality-controlled, manually curated, literature-derived collection of all published GWAS provided jointly by the National Human Genome Research Institute (NHGRI) and the European Bioinformatics Institute (EMBL-EBI). The infrastructure is maintained by EMBL-EBI, and a detailed description of the catalog curation process is available on their Web site (www.ebi.ac.uk/gwas/docs/methods, in the public domain). Single-nucleotide polymorphisms (SNPs) significantly associated with refractive error or axial length were identified by searching the NHGRI-EBI catalog on July 26, 2016 (GRCh38.p7; dbSNP Build147). rsID positions from the 23andMe refractive error study (which is not indexed in the catalog) were updated to the GRCh38.p7 assembly by using Ensembl (release 86) and added to the search output. In total, this process identified 93 SNPs reaching genome-wide significance from 15 GWAS (Supplementary Table S1). The rsIDs were then collated into a single list (Supplementary Table S2), and protein coding genes within 250 kb, 500 kb, and 1 Mb of the loci identified with Ensembl's BioMart (Ensembl release 86; Homo sapiens GRCh38.p7). This latter step replicated that used by Stambolian, who considered all genes within 1 Mb of refractive error loci to be candidates from the assumption that SNPs within enhancer elements could affect gene expression at this distance. As SNPs show a strong bias to be close to the genes that they regulate, we also included analyses for 500 kb and 250 kb distances with the aim of identifying sets of higher confidence candidates.

Animal Transcriptome Studies. Transcriptome studies were included if they measured expression changes in posterior ocular tissues on an essentially transcriptome-wide scale (microarray or RNA sequencing), and compared optically induced ocular growth changes to normal growth controls. A PubMed (National Center for Biotechnology Information [NCBI] at the US National Library of Medicine [NLM], and the National Institutes of Health [NIH], Bethesda, MD, USA). search (July 29, 2016) identified seven peer-reviewed transcriptome studies meeting these criteria (Supplementary Table S1). Differentially expressed genes from each study were identified by using the original author's statistical criteria (listed in Supplementary Table S1) and ultimately converted to human ortholog Ensembl Gene IDs to enable cross-species comparisons. Full details of this identifier conversion process are provided in Supplementary Figure S1. Briefly, probe identifiers from microarray transcriptome studies were first updated (where possible) with the latest Affymetrix annotation (Affymetrix release 36). Identifiers were matched to Ensembl Gene IDs and then converted to human orthologs IDs (Homo sapiens GRCh38.p7) using BioMart (Ensembl release 86). For one-to-many ortholog relationships, the ortholog with the highest confidence and match percentage score was chosen. When two orthologs with identical scores were identified, both orthologs were retained in the analysis. The original results from each transcriptome study and subsequent ID conversions are provided in Supplementary Tables S3 through S9.

Animal Proteomics Studies. Proteomics studies of refractive error were included if they measured protein expression changes in ocular tissues by using an exploratory approach, in addition to meeting the experimental design inclusion criteria described for transcriptome studies. A PubMed search (July 28, 2016) identified nine peer-reviewed proteomics studies meeting these criteria (Supplementary Table S1). Differentially expressed proteins from each study were identified by using the original author's statistical criteria (listed in Supplementary Table S1) and ultimately converted to human ortholog Gene IDs. Full details of this identifier conversion process are provided in Supplementary Figure S2. Briefly, protein identifiers were matched to Ensembl Protein and Gene IDs in BioMart (Ensembl release 86). When protein identifiers could not be matched to a single isoform, all possible isoforms were mapped to Ensembl gene IDs and retained in the analysis. In the final conversion step, Ensembl...
Gene IDs were matched to human orthologs IDs (*Homo sapiens* GRCh38.p7) retrieved from BioMart (Ensembl release 86) as described above for transcriptome studies. The original results from each proteomics study and subsequent ID conversions are provided in Supplementary Tables S11 through S18.

**Gene List Generation.** After human ortholog identification, differentially expressed genes and proteins were split into metalists for comparison with human GWAS genes. These metalists were created by separating differentially expressed genes and proteins by the time-point (≤ 1 day = early, > 1 day = late), growth condition (myopia or growth slowing), and fold-change direction (up- or downregulated). As noted in the introduction, 1 day was chosen as the cutoff for early expression changes, as this time-point provided the best split between the acute induction period when refractive errors were likely to be small to moderate and the chronic induction period when refractive errors were likely to be moderate to severe. This was justified as follows. Twelve datasets were collected within 1 day (Supplementary Table S1); nine of these early datasets were collected from chick where fast research suggests that small but measurable changes to refraction and axial dimensions are likely to have occurred.48 The remaining three datasets originated from mouse53 where refractive changes are likely to be minimal.49 Eighteen datasets were collected at >1 day. The animals in 10 of these datasets (Bertrand 2006,29 McGlinn 2007,21 Lam 2007,50 Jostrup 2009,53 Stone 2011,20 Barathí 2014,57 and Wu 201456) are likely to have undergone large refractive shifts (i.e., ~3–4 D or more) at the time of gene/protein profiling.48,50,51 Animals in the remaining eight late time-point datasets (Tkatchenko 2006,22 Frost 2007,31 Zhou 2010,54 Frost 2012,52 and Li 201253) are likely to have achieved moderate refractive shifts (i.e., ~3–4 D) at the time of gene/protein profiling.52,53 The delineation between early and late expression changes was based on induction time rather than refractive state, as most studies did not provide biometric measures for the animals profiled and thus refractive errors could only be roughly estimated from past literature. 

Note that duplicate identifiers resulting from genes or proteins implicated in multiple datasets were only removed once the final metalists were generated. Consequently, identifiers could be included in multiple lists. For example, a gene identified as upregulated in early myopia in one study, and downregulated in early myopia in another study, would be included in both applicable metalists. Such genes showing seemingly inconsistent fold changes were retained in the analysis because their implication in multiple datasets provided stronger support for their involvement in refractive error. Moreover, we suspected that genes showing dynamic expression shifts over time could be particularly important for controlling the progression of refractive errors. The eventual results supported this methodologic choice, as genes showing inconsistent fold changes were well represented in the GWAS-transcriptome overlaps despite being relatively uncommon in the original data (see Supplementary Table S19).

To further assess the impact of methodologic parameters, we also created metalists of the genes and proteins differentially expressed in each species, tissue (retina/RPE/choroid or sclera), time-point (early or late), and optical manipulation (lens-induced myopia, LIM; lens-induced hyperopia, LIH; form deprivation myopia, FDM; lens-induced myopia recovery, LIMR; and form deprivation myopia recovery, FDMR).

**Cross-Method Comparisons**

**Overlap Statistics.** The processes described above resulted in lists of differentially expressed genes and proteins from individual transcriptome and proteomics studies (segmented by the various methodologic parameters), and lists of genes within 1 Mb, 500 kb, and 250 kb of GWAS refractive error and axial length loci. We then compared these GWAS, transcriptome, and proteomics lists by using the GeneOverlap R package (version 1.10.0). This package uses Fisher’s exact test to calculate the significance of each pair of gene lists in comparison to the genomic background (in this case, 22,109 human protein-coding genes). The function returns the number of intersecting genes between the two lists, the P value, and the estimated odds ratio. The null hypothesis is that the odds ratio is no larger than 1; values larger than this indicate a positive association between lists. Heatmap contingency tables were created to display the pairwise overlaps between lists, including the number of intersecting genes, odds ratios, and significant (P < 0.05) Benjamini-Hochberg-adjusted P values (where applicable). Note that Benjamini-Hochberg P value adjustments were made for each family of comparisons within a contingency table (i.e., on a per-contingency table basis).52–55

**Gene Ontology Characterizations.** To characterize the genes implicated in both GWAS-transcriptome (Supplementary Table S19) and transcriptome–proteomics (Supplementary Table S20) studies, we tested them for enrichment of gene ontology (GO) biological process, cellular component, and molecular function terms by using DAVIDs Functional Annotation Chart tool (version 6.7).54 The analyses were conducted by using the *Homo sapiens* population background, a gene count threshold of 3, and an EASE score cutoff of ≤0.05 (the EASE score is a modified Fisher’s exact P value). A number of enriched gene sets contained highly similar genes. To reduce this redundancy, gene sets that clustered together in the Enrichment Map app (version 2.1.0)55 with an overlap coefficient of 0.7 (indicating similar gene contributions) were collapsed into a single annotation cluster. Chord diagrams were then created to display the relationship between single genes and enriched GO clusters by using the GOplot R package (version 1.0.2). The original DAVID enrichment results (i.e., before redundant gene sets were collapsed) are provided in Supplementary Tables S22 (GWAS-transcriptome) and S23 (transcriptome–proteomics). There were too few commonalities between GWAS and proteomics findings (i.e., only eight genes; Supplementary Table S21) to test their enrichment.

**Results**

**Common Findings Across GWAS, Transcriptome, and Proteomics Methodologies**

The results from seven transcriptome and nine proteomics studies of optically induced refractive error in animals were compared with the genes near human genetic loci for refractive error and axial length (collated from 15 GWAS studies). Nine hundred fifty genes were located within 1 Mb of a human refractive error and/or axial length loci. Of these 950 genes, 103 and 8 were also differentially expressed in animal transcriptome (Supplementary Table S19) and proteomics studies (Supplementary Table S21), respectively (Fig. 1A). In total, 66 of the 93 GWAS loci were located within 1 Mb of at least one differentially expressed gene or protein (Fig. 1B). There was also some overlap in the findings of transcriptome and proteomics studies, with 20 genes differentially expressed at both the transcript and protein level (Fig 1A; Supplementary Table S20).

We first assessed the statistical significance of the overlap between the genes near GWAS loci, and the genes differentially expressed in early and late transcriptome and proteomics...
methodologic conditions showed significant overlap with GWAS candidate genes: genes differentially expressed within the first day of lens wear (Fig. 3A), genes differentially expressed in the retina/RPE/choroid (Fig. 3B), and genes differentially expressed during LIM (Fig. 3D). The genes and proteins differentially expressed in four species (chick, mouse, tree shrew, and primate) showed some overlap with the genes near human GWAS loci (Fig. 3C); however, these overlaps were not statistically significant and all had roughly equivalent odds ratios (ranging from 1.3–2.1), indicating a similar degree of concordance.

**Gene Ontology Characterization of the Genes Implicated Across Multiple Methods**

Two genes within 1 Mb of GWAS loci were differentially expressed in both transcriptome and proteomics studies of refractive error. Glia maturation factor, β (GMFB), a neurotrophic factor that mediates brain cell development and growth, is located within 1 Mb of a refractive error locus on chromosome 14. In animal studies of myopia induction, GMFB was downregulated transcriptionally in the retina/RPE, while the encoded protein was downregulated in the sclera. Collagen, type XIV, α1 (COL14A1), which plays an adhesive role by integrating collagen bundles, is located within 250 kb of a refractive error locus on chromosome 8. COL14A1 was upregulated transcriptionally in the retina/RPE during myopia induction and the encoded protein was downregulated in the sclera during myopia induction. These two genes accounted for one-quarter of the commonalities between GWAS and proteomics studies. The remaining six genes implicated in both methodologies are listed in Figure 4A.

Gene ontology enrichment analyses were used to characterize the longer lists of genes implicated across GWAS–transcriptome and transcriptome–proteomics studies. Twelve GO clusters were enriched in the list of common genes from human GWAS and animal transcriptome studies (Fig. 4C), with the largest cluster related to ion binding. Most of the genes contributing to cluster enrichment were downregulated during myopia induction in animal studies (consistent with the overlap results reported in Fig. 2A). The list of genes differentially expressed in animal transcriptome and proteomics studies was enriched for cytoskeleton and nucleotide binding GO clusters (Fig. 4B). Consistent with the single gene overlap findings reported previously (Fig. 2C), many of the genes contributing these enrichments displayed inconsistent fold-change patterns across transcriptome and proteomics studies.

**DISCUSSION**

A previous comparison demonstrated that the genes within 1 Mb of CREAM refractive error loci share only four commonalities with the genes differentially expressed in two animal transcriptome studies of refractive error. In the present study, however, we demonstrated that expanding this initial comparison to include all available exploratory human and animal studies implicates a further 99 genes (i.e., 103 in total) and 8 proteins that are both differentially expressed in animal models and located within 1 Mb of human GWAS loci. We further show that, although a very large number of genes and proteins have been implicated across these studies, the overlap between human GWAS candidate genes and the genes downregulated during early myopia induction in animal models is statistically significant (i.e., more than expected by chance). These results extend the findings of previous targeted studies to provide systematic evidence that many
similar genes are associated with refractive error loci in humans and optically induced refractive change in animal models. Further research is needed to elucidate the biological mechanisms underlying these cross-species commonalities.

Our results are consistent with a model in which genetic and environmental factors control ocular growth via similar biological pathways, and/or a model in which genetic variants alter susceptibility to environmental factors rather than being causative in and of themselves (as previously suggested).

FIGURE 2. Heatmaps showing overlap of the genes within 1 Mb, 500 kb, and 250 kb of GWAS loci with the genes differentially expressed in animal transcriptome studies of (A) myopia induction and (B) growth slowing (GS). Heatmaps showing overlap of the genes within 1 Mb, 500 kb, and 250 kb of GWAS loci with the genes differentially expressed in animal proteomics studies of (C) myopia induction and (D) GS. Heatmaps showing overlap of the genes differentially expressed in animal transcriptome and proteomics studies of (E) myopia induction and (F) GS. The color of each heatmap square represents the odds ratio (larger odds ratios indicate a stronger association between lists), and the number of overlapping genes and significant Benjamini-Hochberg-adjusted P values (in parentheses where applicable) are superimposed on the grids. Arrows indicate the fold-change direction of differential expression in transcriptome and proteomics studies, and the number of genes implicated in each list is shown in parentheses after the list name. Note that the length of each list is a key factor determining overlap significance.
Notably, there was greater concordance between GWAS candidate genes and the genes and proteins that were downregulated (rather than upregulated) in animal studies. This finding is difficult to explain given the heterogeneous data being compared (i.e., genetic variation versus environmentally driven expression changes), but could plausibly reflect a bias for genetic variants that decrease gene expression (e.g., by disrupting transcription factor binding). 57

Although proteomics studies showed some overlap with both transcriptome studies and GWAS candidate genes, only the proteomic–transcriptome overlap was statistically significant. The lower concordance between GWAS and proteomics studies could reflect several factors. Firstly, some degree of discordance is expected given that the ratio of genes to proteins is not one-to-one, and there are an array of intermediate transcriptional and posttranscriptional regulatory mechanisms.58,59 Indeed, correlation is generally low even at the mRNA and protein level60 (an observation concordant with the present results where fold changes across transcriptome and proteomics studies were often inconsistent). Secondly, most of the proteomic studies used two-dimensional gel electrophoresis (2D-E) combined with large fold-change cutoffs (ranging from $\geq 1.3$–3). Two-dimensional gel electrophoresis has difficulty separating hydrophobic membrane proteins, 61 which were enriched in the GWAS–transcriptome commonalities (see Fig. 4C in the Results section). Because the 2D-E approach is limited to the detection of proteins showing large fold changes, it may also decrease concordance with GWAS that profile the entire genome (particularly if selection pressures create a bias for genetic variants linked to modest changes in gene and/or protein abundance).62,63 Indeed, all of the GWAS–proteomic overlaps originated from the two

![Figure 3. Heatmaps assessing the effects of animal study methodological parameters (A) time-point, (B) tissue, (C) species, and (D) optical manipulation) on the pattern of overlapping results. The color of each heatmap square represents the odds ratio (larger odds ratios indicate a stronger association between lists), and the number of overlapping genes and significant Benjamini-Hochberg–adjusted $P$ values (in parentheses where applicable) are superimposed on the grids. The total number of genes in each list is shown in parentheses after the list name. Note that all transcriptome studies sampled R/RPE/choroid, while proteomics studies were split into those sampling R/RPE/choroid and those sampling sclera. Prot, proteomics; R/RPE/choroid, retina, RPE and/or choroid tissue; Trans, transcriptome.](https://tvst.arvojournals.org/article-pdf/58/1/665/5024672/665.pdf)
proteomics studies\(^{32,37}\) with arguably higher sensitivity (based on their statistical parameters and the number of proteins implicated). Finally, it seems likely that some of the environmentally driven expression responses involved in human myopia are not mediated by underlying genetic variation.\(^{7,64}\) The GWAS data do not necessarily capture these "environment-only" responses, presumably lowering the overall concordance with animal studies of environmentally induced refractive error.

In addition to assessing the overall significance of cross-methodology overlaps, we considered the impact of several animal study methodologic parameters (species, tissue, time-point, and optical manipulation) on the pattern of results. These latter comparisons demonstrated that phylogenetically distant model species show similar gene (e.g., chick and primate) and protein (e.g., tilapia and tree shrew) expression changes, suggesting that expression responses are well conserved across animal models of environmentally driven refractive error. Moreover, although most genes associated with each optical manipulation were unique, most models of myopia induction and growth slowing showed significant overlap at the gene and/or protein level, suggesting that a subset of similar biological mechanisms is activated across growth paradigms. Lastly, we observed commonalities between the genes and proteins differentially expressed in the retina/RPE/choroid and the proteins differentially expressed in the sclera.\(^{19,65–67}\) as any such signaling mechanism would presumably elicit related expression responses across multiple tissue layers. The cross-tissue transcriptome–proteome and proteome–proteome commonalities were primarily composed of genes involved in mediating cellular and extracellular structure including cytoskeletal proteins (\textit{MYH8}, \textit{TPM3}, \textit{GMFB}, \textit{ACTB}, \textit{GSN}, \textit{VIM}), collagen (\textit{COL14A1}), and transcripts that stimulate cell growth and proliferation (\textit{ANGPTL7}, \textit{CDCK4}). Molecular chaperones (\textit{CRYBA1}, \textit{CCT8}), albumin (\textit{ALB}), and apolipoprotein A-I (\textit{APOA1}) were also implicated. Notably, these cross-tissue commonalities included the two genes (\textit{GMFB}, \textit{COL14A1}) implicated by all methods (GWAS, transcriptome, and proteomics), providing strong support for their role in

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4}
\caption{(A) Table showing all genes within 1 Mb of GWAS loci that were also differentially expressed in animal proteomics studies (see Supplementary Table S21 for full details). (B) Chord diagram showing enriched GO clusters for the genes differentially expressed in animal transcriptome and proteomics studies (i.e., all overlapping genes from Supplementary Table S20). (C) Chord diagram showing enriched GO clusters for the genes implicated in both human GWAS and animal transcriptome studies (i.e., all overlapping genes from Supplementary Table S19). In each chord diagram, enriched GO clusters are shown on the right, and genes contributing to this enrichment are shown on the left. Squares on the left indicate whether a gene was differentially expressed in animal transcriptome or proteomics studies of myopia induction and/or growth slowing (see the key for gene and protein differential expression [DE]).}
\end{figure}
integrating structural change across multiple posterior ocular layers.

Our assessment of methodologic parameters also identified several conditions that provided increased concordance with human GWAS candidate genes. Although these findings contribute to the existing debate around the relative importance of different animal models for understanding human myopia, they should be broadly interpreted with caution given the heterogeneous designs of the studies included in each methodologic list. Here, we found that GWAS candidates showed greater concordance with the genes differentially expressed at early (relative to late) induction time-points. A similar pattern was not seen for the proteomic data where only one study (with two differentially expressed proteins) met inclusion criteria for early expression changes. Notably, however, five of the eight GWAS–proteomic overlaps originated from the study by Frost and Norton of LIM in the slower tree shrew model. Although this dataset was included in the “late” induction category, the animals had only achieved moderate refractive shifts (~3.4 D). The greater concordance with animal studies at early time-points and/or when refractive shifts were small to moderate is particularly interesting in light of recent findings that many of the CREAM GWAS loci show early onset effects in children that remain stable or progress further with age. In this context, our findings support the biological validity of the proposal by Stone and Khurana, and others (e.g., Guo et al. and He et al.) that different genes are involved in the onset versus the persistence and progression of refractive change.

Genome-wide association study candidate genes also showed greater concordance with the genes differentially expressed in LIM (relative to other optical manipulations). This finding is less convincing, as the effect size was weak and other methodologic factors varied considerably across the different optical manipulation groups (e.g., induction time, species, and the number of studies). Similarly, our analysis demonstrated that the genes differentially expressed in the retina/RPE/choroid (i.e., all of the genes implicated in transcriptome studies) overlapped significantly with the genes near human GWAS loci, while the proteins differentially expressed in the retina/RPE/choroid and sclera do not. In the absence of transcriptome studies profiling sclera, it seems likely that this finding reflects the overall greater concordance of the transcriptome results with GWAS candidate genes.

It is notable that, although four model species (chick, mouse, tree shrew, and primate) showed some commonalities with GWAS candidate genes, there was no evidence of more significant overlap in similar species to humans (e.g., primates or mammals). In combination with the significant cross-species overlap at the gene and protein level discussed above, these findings suggest that model species is not an important factor for determining the degree of overlap between environmentally mediated expression changes in animals and the genes near human refractive error loci. However, it should be noted that data in higher-order animal species such as primates are currently limited and, as such, future studies may well uncover evidence for a species effect.

Heterogeneous experimental designs and incomplete data availability necessitated a few important limitations in our analysis. We identified candidate genes on the basis only of their proximity to refractive error peaks, and thus they are not necessarily those by which a given SNP affects refractive development. Moreover, our analysis was limited to protein-coding genes and may have missed more complex effects mediated by transcriptional and posttranscriptional regulators such as microRNAs (Tedja MS, et al. IOVS 2016;57:ARVO E Abstract 4791). Many of the animal studies included in our analysis used expression profiling techniques with limited gene or protein coverage, and most also imposed fold-change cutoffs (see Supplementary Table S1). These factors presumably decrease concordance with GWAS that profile the entire genome. In addition to using varied statistical criteria, the included animal studies encompassed a wide range of experimental designs (i.e., animal age and circadian timing, measurement platform, tissue, induction time-point, optical manipulation, and species). As mentioned previously, these methodologic variations made it difficult to compare equivalent conditions across transcriptome and proteome methodologies, and to segregate out balanced datasets representing different methodologic parameters. Moreover, these methodologic variations precluded finer evaluation of variables such as tissue type and induction time (which were instead assessed in crude groupings). As such, although our primary finding regarding significant overlap between GWAS candidate genes and the genes differentially expressed in animal models is robust, our remaining findings regarding the effects of different animal study methodologic parameters should be interpreted with caution.

To improve the systems-level understanding of ocular growth control, future studies need to integrate multiple levels of omics data collected under comparable experimental conditions. In particular, expression quantitative trait locus (eQTLs) studies combining genotype and gene expression data from posterior ocular tissues are needed to improve candidate gene predictions for GWAS loci. Studies collecting concurrent transcriptome and proteome measurements in identical animal models may also help to elucidate whether the fold-change discordance we identified across these study types was due to methodologic factors (such as the timing of expression profiling) or posttranslational events. Given the recent emergence of next-generation transcriptome and proteomic technologies, such future studies are also likely to generate better-quality datasets in terms of reproducibility, sensitivity, and coverage. It is also notable that large-scale exploratory transcriptome studies of gene expression in the sclera during refractive error induction are currently unavailable (although some preliminary results have been reported in mouse).

As noted above, our analysis identified significant overlap of proteins differentially expressed in the retina/RPE/choroid and sclera. Given that transcriptome and proteomics results also showed greater within- rather than across-method concordance, it is reasonable to expect that future scleral transcriptome studies would identify additional commonalities with retinal transcriptome responses that may help to elucidate the signaling mechanisms propagating across the posterior eye.

In summary, we conducted the first large-scale comparison of genetic, transcriptome, and proteomic studies of refractive error. We showed that gene and protein expression changes are well conserved across animal models of environmentally driven refractive error, and that the genes implicated in these animal models overlap significantly with the genes near human GWAS refractive error loci. These findings provide strong support for the continued use of animals for investigating the biological basis of human myopia and developing novel therapeutic approaches to control eye growth.

Acknowledgments
The authors would like to thank Loretta Giummarra, Melanie Murphy, and David P. Crewther for their helpful comments on the manuscript.

Supported by Australia Research Council Grant DP110103784 to David P. Crewther and SGC and by Australian Government Research Training Program Scholarship to NR.

Disclosure: N. Riddell, None; S.G. Crewther, None
Integrated Comparison of Myopia Omics Studies

References


