RNA-Seq: Improving Our Understanding of Retinal Biology and Disease

Michael H. Farkas, Elizabeth D. Au, Maria E. Sousa, and Eric A. Pierce

Ocular Genomics Institute, Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts 02114

Correspondence: eric_pierce@meei.harvard.edu

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org

Over the past several years, rapid technological advances have allowed for a dramatic increase in our knowledge and understanding of the transcriptional landscape, because of the ability to study gene expression in greater depth and with more detail than previously possible. To this end, RNA-Seq has quickly become one of the most widely used methods for studying transcriptomes of tissues and individual cells. Unlike previously favored analysis methods, RNA-Seq is extremely high-throughput, and is not dependent on an annotated transcriptome, laying the foundation for novel genetic discovery. Additionally, RNA-Seq derived transcriptomes provide a basis for widening the scope of research to identify potential targets in the treatment of retinal disease.

The study of whole transcriptome gene expression in disease research is not a novel concept. Early methods of gene expression analysis, such as the use of expressed sequence tags (EST) and serial analysis of gene expression (SAGE), have been in use since the early 1990s. In the late 1990s, microarrays quickly became the method of choice for the study of gene expression, owing to their higher throughput nature. These methods allowed scientists to study transcriptomes in great detail and in less time, giving rise to large amounts of information more quickly than previously thought possible.

Early gene expression studies using EST analysis were first published by Adams et al. (1991), and quickly gained popularity as a means to identify novel alternate splice sites and examine differential expression in data sets. This technique involves sequencing a cloned cDNA and mapping the sequence (100–800 bp) to a genome of interest. In 1999, 83 EST clusters were identified as potential retinal specific genes, with 14 further classified as potential disease genes (Malone et al. 1999). By 2000, the first analysis of the retinal transcriptome was published (Bortoluzzi et al. 2000). Nearly 5000 known retinal genes were studied, levels of expression were estimated, and several genes were noted to be potentially associated with disease. Although EST studies laid the groundwork for analysis of the retinal transcriptome, they are extremely low-throughput, making whole transcriptome analyses time consuming and difficult.

The use of SAGE analysis was first published in 1995 as a means to study differences in gene expression in patients with cancer (Velculescu

Editors: Eric A. Pierce, Richard H. Masland, and Joan W. Miller
Additional Perspectives on Retinal Disorders: Genetic Approaches to Diagnosis and Treatment available at www.perspectivesinmedicine.org

Copyright © 2015 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/cshperspect.a017152
Cite this article as *Cold Spring Harb Perspect Med* 2015;5:a017152

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

et al. 1995). SAGE studies produce a list of short (10–20 bp) sequences, which can then be mapped back to a genome of interest, whereas EST studies are based on the sequencing of one longer sequence. In 2002, SAGE libraries constructed from two eye tissue samples were studied, identifying 26,355 retinal transcripts, and 10,404 RPE (retinal pigment epithelium) transcripts (Sharon et al. 2002). SAGE studies, although an improvement on ESTs, are still low-throughput, and the analysis of an entire transcriptome is both time-consuming and expensive, necessitating the use of more high-throughput analysis methods to comprehensively study the retinal transcriptome (Swaroop and Zack 2002).

DNA microarrays were an answer to the call for higher throughput methods of expression analysis, and were first used to study a "complete" genome in 1997 (Lashkari et al. 1997). Microarrays evolved from the technique of Southern blotting. Complementary sequence that aligns uniquely to each gene of interest are mapped on an array, and used to determine the relative levels of expression of each gene in a given sample. In 2010, microarray analysis identified 154 "signature RPE genes," in which expression in the RPE was at least 10-fold higher than published expression levels in other tissues (Strunnikova et al. 2010). The high-throughput nature of microarrays makes this method preferable to EST and SAGE for whole transcriptome analyses. However, microarrays only allow for relative quantitation of transcripts compared with all other transcripts on the array. Additionally, microarrays are reliant on a properly annotated genome. It is only possible to study previously identified transcripts, without the ability to identify alternate splice sites or novel exons.

RNA-Seq is the newest and most widely used method for transcriptome analyses, combining the qualitative nature of EST and SAGE studies with the high-throughput and quantitative abilities of microarrays. Importantly, RNA-Seq experiments are extremely cost-effective considering the large amount of data produced, and allow for the identification of novel exons, splice sites, and transcripts. Whereas EST and SAGE sequencing studies required the researcher to clone a small section of a genome at a time, RNA-Seq allows for the sequencing of an entire transcriptome (we define the transcriptome as all expressed RNAs, both coding and noncoding). This greatly enhances the possibility of finding novel transcripts, and offers increased insight into the transcriptional dynamics of a particular cell.

Next generation sequencing (NGS) has become a ubiquitous research tool in all fields of science, and has greatly enhanced the field of ophthalmology. The rise of NGS was possible because of the many advantages it offered over previously preferred techniques. Now, with only nanogram levels of starting material, the current NGS technologies can create more than one hundred million paired-end reads, covering a large portion of the genome, or transcriptome, while maintaining sensitivity and specificity (Mardis 2008; Morozova 2008; Shendure 2008; Reis-Filho 2009). The data generated is useful in a variety of applications because of its dual qualitative and quantitative nature (Reis-Filho 2009). This enables individual investigators to pursue projects previously accessible only to larger consortia, particularly because of its cost effectiveness over Sanger sequencing (Schuster 2007; Morozova 2008; Shendure 2008).

One of the most prominent challenges of RNA-Seq has been the bioinformatic analysis of such large data sets. The massive amount of data generated by each sequencing run has created a high demand for pipelines capable of accurately analyzing the data in a timely manner (Shendure 2008). In addition, as RNA-Seq studies continue to produce more and more data, it becomes increasingly difficult to sift through the results to discover pertinent biological insights (Reis-Filho 2009).

NEXT GENERATION SEQUENCING IN OPHTHALMOLOGY

Since its commercial availability in 2004, NGS has been used in a variety of cases within ophthalmology (Mardis 2008). Whole exome sequencing has identified both novel and known mutations in DHDDS, SNRNP200, MAK,

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

NMNAT1, and RPE65, in patients with Leber congenital amuarosa (LCA) and retinitis pigmentosa (RP) (Benaglio 2011; Bowne 2011; Özgül 2011; Tucker 2011; Wang 2012). Similarly, NGS has been used to search for mutations in other ocular conditions. In 2010, mutations were discovered in TSPAN12, in patients who exhibited familial exudative vitreoretinopathy (Nikopoulos 2010). NGS has also been used to investigate variants in ABCA4, which is associated with a number of retinal dystrophies including Stargardt disease, cone–rod dystrophy and RP (Zernant 2011). Furthermore, this technology has aided in the discovery of mutations in GPR179 and ATOH7, which are associated with autosomal-recessive complete congenital stationary night blindness and global eye defects, respectively (Khan 2012).

Increasingly, NGS has been used in a clinical setting for diagnostic purposes within ophthalmology. This method has the ability to aid in the diagnosis of diseases such as RP, which has dozens of causative genes associated with it (Audo 2012; Neveling 2012; Schrader 2011). NGS has the ability to continue to shape both the future of diagnostics and ophthalmology research as a whole.

RNA-SEQUENCING

RNA-Seq is a powerful NGS method that allows for the study of transcriptomes in great detail and with a high degree of accuracy, providing unprecedented resolution of gene expression at the level of the transcript down to a single base (Chepelev et al. 2009; Costa et al. 2010; Marguerat and Bahler 2010). This affords the ability to study RNA editing, alternative splicing, and isoform quantification (Costa et al. 2010; Ozsolak and Milos 2011; Rosenberg et al. 2011; Djebali et al. 2012). Further, the relatively unbiased nature of the technology captures novel transcriptome events including novel isoforms and novel gene expression (Chen et al. 2011; Halvardson et al. 2013; Kim et al. 2012; Mercer et al. 2012). RNA-Seq studies provide a vast wealth of information, and hence require sophisticated methodologies and bioinformatic algorithms for data analysis. As NGS technology continues to improve, the RNA-Seq library preparation techniques and analysis packages are quickly evolving as well. These innovations, along with the development of appropriate models, allow for RNA-Seq to be applied to vision research.

NEXT GENERATION LIBRARY PREPARATION

Initially, RNA-Seq was limited by the relatively large amount (1–10 μg) of total RNA required for library preparation (Farkas et al. 2012; Grant et al. 2011). This amount of RNA was needed because of long protocols with many inefficient purification steps that lead to significant loss of product. Even with such large starting amounts of RNA, the loss of sample at each step required PCR amplification to have enough final product for sequencing. It is important to note that over-amplification has the potential to lead to amplification bias and PCR artifacts (Wang et al. 2009; Costa et al. 2010, 2012). These initial preparations were labor-intensive, relatively low-throughput, multi-day protocols.

Lower input, higher-throughput methods have since been developed, which allow for starting total RNA concentrations as low as 100 ng (Brooks et al. 2011, 2012). An initial concern with using minimal starting material was that samples would be underrepresented when compared with data produced using the earlier, high-input protocols. Brooks et al. (2012), however, showed in mouse retina that the two methods are comparable (Brooks et al. 2011). Although an improvement on the first library preparation protocols, these methods are still subject to PCR bias because of necessary amplification before sequencing, and require multiple days to complete.

A newer innovation involves transposon-mediated fragmentation and adaptor ligation, processes originally designed for DNA sequencing that can be adapted for RNA-Seq library preparation (Syed 2010; Gertz et al. 2012). Unlike previous RNA-Seq preparation methods, in which polyadenylated RNA (polyA RNA) was selected and fragmented, then converted to double-stranded cDNA (ds cDNA), the new method involves a polyA RNA selection step

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

during first-strand cDNA synthesis followed by transposon-based fragmentation of ds cDNA. This offers significant improvements to the process, eliminating purification steps and minimizing loss of product. Further, because there are fewer purification steps in this process, the number of amplification cycles is reduced to only five, and PCR bias is minimized. Finally, this newest method is higher throughput and can be performed in hours rather than days.

BIOINFORMATIC ANALYSIS—ALIGNMENT OF RNA-Seq DATA

One of the most critical steps in any RNA-Seq experiment is the alignment of reads to a reference genome and/or transcriptome. This is also one of the most challenging steps because of splicing of exons to form transcripts. RNA-Seq aligners are continually being developed, and improved, to overcome this challenge. Many of the currently available alignment algorithms have been comprehensively characterized and reviewed (Fonseca et al. 2012; Lindner and Friedel 2012). The most important factors to consider when choosing an RNA-Seq aligner are its ability to: 1) accurately align reads; 2) maximize the number of reads aligned (sensitivity); 3) detect splice junctions (both annotated and novel); and 4) efficiently align reads (i.e., speed of alignment with minimal compute resources).

Currently, no RNA-Seq alignment algorithm is able to outperform the others in all four categories. In choosing which algorithm is most suitable for a particular RNA-Seq experiment, it is important to weigh the benefits of each. For example, of the most popular aligners, RUM (RNA-Seq unified mapper) and GSNAP (Genomic Short-Read Nucleotide Alignment Program) are the most accurate, sensitive, and capable of detecting splice junctions (Grant et al. 2011; Dobin et al. 2013). However, they are only able to achieve this by being compute-intensive and having slower alignment speeds, relative to other programs. The most recently developed algorithm, STAR, is an ultrafast aligner that uses fewer compute resources relative to RUM and GSNAP, but does so at the cost of sensitivity and accuracy (Dobin et al. 2013). Even so, STAR outperforms some of the most popular alignment algorithms such as TopHat and MapSplice.

Although alignment speed is certainly important, with such large data sets being produced by NGS studies, it may be the least important factor, taking into account the amount of data needed to fully cover a transcriptome. We have reported that 100 million reads are necessary to fully cover both the human and mouse neural retina transcriptome (Grant et al. 2011; Farkas et al. 2012, 2013). Because transcriptome complexity varies between tissues and organisms, it is necessary to independently determine the appropriate depth of coverage for each RNA-Seq study (Wang et al. 2008; Barbosa-Morais et al. 2012; Gonzalez-Porta et al. 2012). With the amount of relatively inexpensive resources available on the cloud, if a read depth of 100 million read is considered sufficient, aligner speed becomes less important than accuracy, sensitivity, and junction detection (Karlsson et al. 2012; Kienzler et al. 2012; Morgan et al. 2012; Nekrutenko and Taylor 2012; Zhang et al. 2012).

BIOINFORMATIC ANALYSIS—DIFFERENTIAL EXPRESSION

Although an important part of the analysis process, sequence alignment is only the first step in the bioinformatic processing of RNA-Seq data. Downstream analysis algorithms are necessary to obtain more useful information, and can address a range of issues, from basic tasks such as quantification of differential gene expression, to more challenging projects such as determining alternative splicing and RNA-editing (Peng et al. 2012). Unfortunately, as with the alignment algorithms, development of algorithms to perform such downstream analyses have lagged behind the advances in sequencing technology, and there is limited consensus as to which perform optimally.

Often, one of the first postalignment questions regarding sequencing data is the determination of differential expression (DE) between conditions. This is not a new concept; analysis of transcriptome data for DE features has been per-

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

formed since the first experiments utilizing SAGE (Lorenz and Dean 2002; Baggerly et al. 2003; Lu et al. 2005). Microarrays soon replaced SAGE as the method of choice to study transcriptomes, and a new set of DE algorithms were developed (Cui and Churchill 2003; Smyth 2004). When RNA-Seq was developed, it was assumed that the statistical models developed for microarrays could be adapted for RNA-Seq DE analyses. Indeed, some RNA-Seq DE algorithms borrow statistical assumptions from microarray packages (Oshlack et al. 2010). However, microarray quantification is based on a continuous intensity distribution, whereas RNA-Seq quantification is the result of a discrete measurement (i.e., number of reads mapped to a feature). The differences in data distribution and the biases associated with each method suggest that DE algorithms would not be compatible between the two. Microarray data is modeled using a normal distribution, although a Poisson distribution was initially determined to be the most appropriate for modeling RNA-Seq data (Marioni et al. 2008; Bullard et al. 2010). In fact, analysis of RNA-Seq experiments in which the same sample was run over multiple lanes of a flowcell showed that RNA-Seq data was Poisson distributed (Marioni et al. 2008). Poisson modeling of biological replicates, however, does not account for biological variability, and is prone to high false positive rates (Anders and Huber 2010). Interestingly, the analysis method that most appropriately fits RNA-Seq data was first developed for SAGE analyses. Because these two methods are relatively similar and mainly differ in the amount of data generated from a single experiment, a negative binomial distribution has been considered the most appropriate for DE analysis of RNA-Seq data (Robinson and Smyth 2008; Anders and Huber 2010; Hardcastle 2010; Robinson et al. 2010). Two of the most popular DE algorithms, DESeq and EdgeR, incorporate the negative binomial statistical approach, but differ in their normalization methods (Anders and Huber 2010; Robinson et al. 2010). Multiple studies comparing the two methods have been published with the consensus being that DESeq is more conservative and better limits the false positive rate, although EdgeR is able to call more truly DE genes, but at the expense of a higher false positive rate (Rapaport et al. 2013; Soneson and Delorenzi 2013). Ultimately, the best algorithm will be the one that best fits the research design of the end user.

In addition to quantification of gene expression, RNA-Seq can be used to study RNA splicing. Detecting splice junctions is not a trivial task, as it requires mapping reads to multiple locations in the genome, separated by anywhere from a few bases to as many as a few hundred thousand bases. The ability to perform such mapping functions are built into alignment algorithms, but are performed differently by each. RUM, for example, identifies annotated splice junctions by using a concatenated transcriptome index, which does not include introns or intergenic regions (Grant et al. 2011). This allows RUM to quickly align reads to annotated junctions without the need for gapped alignment. Novel splice junction detection by RUM is more compute intensive because it is performed using the BLAST-Like Alignment Tool (BLAT) (Kent 2002). BLAT uses an "anchor-extension" strategy, in which a portion of the read is aligned to the genome, and is considered to "anchor" the alignment. The remaining read is then aligned to another region in the genome. In contrast, TopHat uses an alternative method for detecting splice junctions, which relies on the aligner Bowtie, assembly algorithm Maq, and canonical splice site information to map reads crossing splice junctions (Li et al. 2008; Trapnell et al. 2009). TopHat splice junction mapping differs from the "anchor-extension" method in that it maps reads to build "islands" that represent exons, using the canonical splice site information to find the exon/intron boundaries.

The "anchor-extension" strategy is acknowledged to be an excellent method for novel splice junction determination. Notably, using RUM, we have identified thousands of novel splicing events in the mouse neural retina, including exon skipping, novel exons, and alternate 3′ and 5′ splicing (Grant et al. 2011). Several new algorithms use statistical modeling methods as an alternative approach. Most recently, Fine-

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

Splice, PASTA, and TrueSight have been developed to use regression modeling to identify splice junctions, and claim to enhance sensitivity and specificity over traditional splice junction detection methods, but have not been independently verified (Li et al. 2013; Tang and Riva 2013; Gatto et al. 2014).

RNA-Seq STUDIES IN THE RETINA

The rapid development of bioinformatic tools has enabled a complete genomic characterization of the retina, which is necessary to properly study the basic biological mechanisms that drive cellular physiology, morphology, and development, as well as disease. To date, only a few RNA-Seq studies have set out to characterize the transcriptome of the normal or diseased retina. Original characterizations of the retina using SAGE and microarray technology, although useful, are antiquated and inadequate for fully sampling the transcriptome (Mortazavi et al. 2008; Wang et al. 2009; Li et al. 2010). Microarrays can comprehensively sample the annotated transcriptome, but are limited to providing information about known features (exons, genes, etc.), and are unable to aid in the discovery of novel features. In contrast, SAGE analyses are able to detect novel features, but are limited by minimal data output (Oshlack et al. 2010). Although RNA-Seq does have its own biases, such as problems caused by PCR amplification, it addresses many issues of the earlier technologies and allows for a much more comprehensive transcriptome analysis (Costa et al. 2012).

Given the importance of mouse models for vision research, and the difficulty in obtaining high quality human retinal samples, it is no surprise that the first RNA-Seq retinal transcriptome articles to be published examine the mouse retina. These studies have found between 16,000–34,000 transcripts expressed in the mouse retina, which is in line with prior SAGE analyses of human retina, where 26,000 transcripts were found to be expressed (Swaroop and Zack 2002; Brooks et al. 2011). Analyses regarding the overlap of transcript expression in the retina between the two species, including isoform abundance, has not yet been performed, but would be particularly useful for studying retinal dystrophies in mouse models.

More comprehensive RNA-Seq studies of the mouse retina have shed light onto the overall expression and alternative splicing of known retinal disease genes. For example, Gamsiz et al. (2012) noted that many of these genes are among the most highly expressed in the retina, with rhodopsin being at the top of the list (Gamsiz et al. 2012). The investigators also point out that these genes are among the largest, and are susceptible to more alternative splicing events. This is interesting, and confirms previous observations that the number of alternative splicing events a transcript undergoes is proportional to the number of exons in the transcript (Pickrell et al. 2010).

Although new RNA-Seq studies are continuously increasing our knowledge of the mouse retina, alternative splicing has not yet been fully characterized, and, in particular, novel alternative splicing events remain to be fully investigated. Grant et al. (2011) set out to identify and validate novel alternative splicing events, including exon skipping, alternate 3′/5′ splice sites, and novel exons, concluding that tens of thousands of novel alternative splicing events occur in the mouse retina (Grant et al. 2011). A majority of the novel features identified were seen in lower abundance than their annotated counterparts, however, there are also novel features with expression levels suggesting they may in fact be the major isoform of a given gene. Further research is required to determine the biological function of these novel features. Additionally, the fact that the highly expressed novel features have not yet been discovered in other tissues raises the question of their specific importance in the retina.

MODELS FOR STUDYING DISEASE

RNA-Seq captures a snapshot of the transcriptome at a given point in time (Wang et al. 2009; Costa et al. 2010; Marguerat and Bahler 2010). This fact complicates studying disease in humans. In the case of inherited retinal dystrophies (IRDs), for example, high quality human retinal tissue is particularly hard to obtain. In

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

addition, transcriptome changes occur because of degeneration of the retina and extended postmortem times to RNA isolation, further complicating the issue (Koppelkamm et al. 2011). To appropriately study the pathogenesis of an IRD, one would need to view the transcriptome before degeneration. In light of these inherent difficulties, mouse models have been studied as an alternative, although these models are still not a perfect solution.

Mouse models are clearly the method of choice for studying visual dystrophies (Buch et al. 2008; Liu et al. 2010; Jaillard et al. 2012). From a disease standpoint, they have played a major role in elucidating the mechanisms of pathogenesis for many types of blindness (Won et al. 2011). Mouse models have helped identify the role of oxidative damage, lipid metabolism, inflammation, and immune response in the development of age-related macular degeneration (AMD) (Hollyfield and Kuttner-Kondo 2010; Pennesi et al. 2012). In this last case, however, there are obvious deficiencies in using mice as a model for AMD. Most notably, mice lack a macula, and hence true AMD has not been recapitulated. Another issue arises in terms of RNA-Seq data itself. Expression levels of gene isoforms in the transcriptomes of mice vary relative to those expressed in humans (Matlin et al. 2005; Meena Kishore et al. 2005; Pan et al. 2005; Konopka et al. 2009; Stahl and Wainszelbaum 2009; Nilsen and Graveley 2010). This can be caused by variation in alternative splicing between the two species, as well as differences in the degree of expression of each isoform in the tissue. It is well known that splicing is not well conserved between species, especially between human and mouse. It is currently not well understood how these variations affect normal tissue function, and it is unknown how they may affect disease models. Despite this issue, mouse and zebrafish models have commonly been considered the best alternative for disease research, since obtaining retinal samples is not always possible. In cases in which disease genes are conserved between species, these models have proven effective. In other instances, such as the study of mutations in the RNA splicing factors, which cause RP, animal models may not be the best choice, because aberrant splicing is hypothesized to underlie disease pathogenesis (Ivings et al. 2008; Graziotto et al. 2011; Farkas et al. 2012; Cvačková et al. 2014).

One possible alternative to animal models is the use of induced pluripotent stem cells (iPS); with iPS, it is possible to derive RPE and photoreceptor-like cells from human, or animal, fibroblast cells or blood (Takahashi et al. 2007; Aoi et al. 2008; Okamoto and Takahashi 2011; Phillips et al. 2012). This technology offers a lot of promise from a research standpoint, as expression profiles closely match those found in human tissue samples, providing a more accurate disease model. Specifically, in cases in which differences in splicing patterns present complications, iPS cells from patients affected with RNA splicing factor RP may provide a better alternative to understanding disease onset. Further, since the discovery, and practicality, of genome editing, affected patient samples are no longer necessary because mutations can be readily introduced into iPS cells. Additionally, the use of iPS cells offers clinical promise, as functional iPS-derived retina cells have the potential for therapeutic use (Carr et al. 2009; Carr 2011; Schmeer et al. 2012). Before clinical use, a thorough characterization is warranted and necessary, because iPS-dervived RPE have been shown to have a short lifespan after transplantation in the rat eye (Carr et al. 2009). Currently, it is unknown why iPS-derived RPE do not survive in vivo, especially since functional characterization suggests the RPE are indeed functional (Kokkinaki et al. 2011; Westenskow et al. 2012).

RNA-Seq STUDIES TO UNDERSTAND DISEASE MECHANISMS

To date, many mouse models of human visual dystrophies have been developed, but only a few have been studied using RNA-Seq. Age-related retinal degeneration (ARD) is the most common cause of vision loss (Mustafi et al. 2012). It is defined as likely being caused by a myriad of factors that are both genetic and environmental (Donoso et al. 2010; Hong et al. 2011; Pennesi et al. 2012; Sui et al. 2012). Little is known about the genetic factors underlying the disease, which

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

makes it an ideal candidate for RNA-Seq studies. In mice, genetic background plays a role in the development of ARD. Mice on the A/J background are more susceptible to retinal degeneration than are BALB/c or B6 mice (Mustafi et al. 2012). In A/J mice, transcriptome analysis of the whole eye, rather than a specific retinal tissue, revealed an increase in proinflammatory factors and a decrease in protective factors such as those involved in the oxidative stress response. These differentially expressed genes were ultimately localized to the RPE, suggesting a role for the RPE in ARD development.

Diabetic retinopathy is another leading cause of vision loss worldwide (Klein et al. 2009; Visser 2010). The pattern of gene expression changes in mouse models of diabetic retinopathy as the disease progresses, are notable (Kandpal et al. 2012). Inflammation has been shown to be critical to the pathogenesis of diabetic retinopathy. RNA-Seq studies have found that the degree of up-regulation of inflammation related genes actually decreases over time as disease severity progresses, while at the same time other genes, such as those involved in the Wnt signaling pathway, crystallins, and various transcription factors are newly up-regulated (Li et al. 2009; Tang and Kern 2011; Kandpal et al. 2012). Further, these RNA-Seq studies note that disease-specific alternative splicing events, which were previously found to occur in the disease state, in fact accumulate as disease progresses. Interestingly, potential inhibitors known to slow the progression of diabetic retinopathy by targeting specific metabolic pathways were shown to regulate the expression of some of the over expressed genes, most notably the crystallins, in the disease models. Additionally, these inhibitors reversed the levels of the disease-specific alternative splicing events implicated in disease progression.

Although RNA-Seq-based transcriptome analyses of the eye are still in their infancy, it is clear the tremendous potential this technology provides the field. RNA-Seq will likely play an important role in uncovering everything from the eye's basic transcriptional landscape to providing unprecedented insight into the mechanisms underlying disease.

CONCLUDING REMARKS

RNA-Seq has proven to be a powerful, albeit underutilized, tool for studying disease in the eye. Studying the transcriptome of patients with retinal disease is difficult and largely dependent on mouse models. Because the transcriptomes of humans and mice differ quite significantly, these models are not always sufficient for RNA-Seq studies. With the advancement of iPS technology, and the ability to derive RPE and photoreceptor progenitor cells, the technology necessary to perform suitable human studies will now be readily available. These studies will be essential for furthering both genetic and clinical studies to help bring therapies to patients.

The increase in demand for a genetic diagnosis by patients with ocular disease is rising because of the recent advances in gene therapy. In addition, the fact that many ocular diseases present with a similar phenotype caused by a heterogeneous group of genes (e.g., RP) necessitates the need to have a genetic diagnosis before treatment is possible. Whole-exome sequencing (WES) provides the quickest and most practical technology available for screening and providing a genetic diagnosis to patients with mutations in protein coding genes. However, it has become evident that we do not have a complete understanding of many diseases, as many patients with an obvious Mendelian disease cannot be diagnosed using current methods. Although this might be attributed to our incomplete understanding of the exome, it is highly likely that noncoding RNAs and other unexplored regions of the genome contribute to disease. RNA-Seq is both more affordable and powerful to explore these regions, compared with other NGS technologies. Future advances in whole exome sequencing technology, RNA-Seq, analysis, bioinformatics, and functional models should provide clinicians with powerful tools to diagnose and treat ocular disease.

REFERENCES

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF. 1991. Complementary DNA sequencing: Expressed

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

sequence tags and human genome project. *Science* **252:** 1651–1656.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11:** R106.

Aoi T, Yae K, Nakagawa M, Ichisaka T, Okita K, Takahashi K, Chiba T, Yamanaka S. 2008. Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321:** 699–702.

Audo I, Bujakowska K, Mohand-Said S, Lancelot M-E, Moskova-Doumanova V, Waseem N, Antonio A, Sahel J-A, Bhattacharya S, Zeitz C. 2010. Prevalence and novelty of PRPF31 mutations in French autosomal dominant rod-cone dystrophy patients and a review of published reports. *BMC Med Genet* **11:** 145.

Baggerly KA, Deng L, Morris JS, Aldaz CM. 2003. Differential expression in SAGE: Accounting for normal between-library variation. *Bioinformatics* **19:** 1477–1483.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Sáolak R. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338:** 1587–1593.

Benaglio P, McGee TL, Capelli LP, Harper S, Berson EL, Rivolta C. 2011. Next generation sequencing of pooled samples reveals new SNRNP200 mutations associated with retinitis pigmentosa. *Hum Mutat* **32:** E2246–E2258.

Bortoluzzi S, d'Alessi F, Danieli GA. 2000. A novel resource for the study of genes expressed in the adult human retina. *Invest Ophthalmol Vis Sci* **41:** 3305–3308.

Bowne SJ, Humphries MM, Sullivan LS, Kenna PF, Tam LCS, Kiang AS, Campbell M, Weinstock GM, Koboldt DC, Ding L. 2011. A dominant mutation in RPE65 identified by whole-exome sequencing causes retinitis pigmentosa with choroidal involvement. *Eur J Hum Genet* **19:** 1074–1081.

Brooks MJ, Rajasimha HK, Roger JE, Swaroop A. 2011. Next-generation sequencing facilitates quantitative analysis of wild-type and $Nrl^{-/-}$ retinal transcriptomes. *Mol Vis* **17:** 3034–3054.

Brooks MJ, Rajasimha HK, Swaroop A. 2012. Retinal transcriptome profiling by directional next-generation sequencing using 100 ng of total RNA. *Methods Mol Biol* **884:** 319–334.

Buch P, Bainbridge J, Ali R. 2008. AAV-mediated gene therapy for retinal disorders: From mouse to man. *Gene Ther* **15:** 849–857.

Bullard JH, Purdom E, Hansen KD, Dudoit S. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11:** 94.

Carr A. 2011. iPS cells—An alternative source of RPE? *Acta Ophthalmol* doi: 10.1111/j.1755-3768.2011.2231.x

Carr AJ, Vugler AA, Hikita ST, Lawrence JM, Gias C, Chen LL, Buchholz DE, Ahmado A, Smart MJK, Hasan S. 2009. Protective effects of human iPS-derived retinal pigment epithelium cell transplantation in the retinal dystrophic rat. *PLoS ONE* **4:** e8152.

Chen G, Li R, Shi L, Qi J, Hu P, Luo J, Liu M, Shi T. 2011. Revealing the missing expressed genes beyond the human reference genome by RNA-Seq. *BMC Genomics* **12:** 590.

Chepelev I, Wei G, Tang Q, Zhao K. 2009. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nucleic Acids Res* **37:** e106.

Costa V, Angelini C, De Feis I, Ciccodicola A. 2010. Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnology* **2010:** 853916.

Costa V, Aprile M, Esposito R, Ciccodicola A. 2012. RNA-Seq and human complex diseases: Recent accomplishments and future perspectives. *Eur J Hum Genet* **21:** 134–142.

Cui X, Churchill GA. 2003. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol* **4:** 210.

Cvačková Z, Matějů D, Staněk D. 2014. Retinitis pigmentosa mutations of *SNRNP200* enhance cryptic splice-site recognition. *Hum Mutat* **35:** 308–317.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. *Nature* **489:** 101–108.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: Ultrafast universal RNA-Seq aligner. *Bioinformatics* **29:** 15–21.

Donoso LA, Vrabec T, Kuivaniemi H. 2010. The role of complement factor H in age-related macular degeneration: A review. *Surv Ophthalmol* **55:** 227–246.

Farkas MH, Grant GR, Pierce EA. 2012. Transcriptome analyses to investigate the pathogenesis of RNA splicing factor retinitis pigmentosa retinal degenerative diseases. (ed. LaVail MM, Ash JD, Anderson RE, Hollyfield JG, Grimm C), pp. 519–525. Springer, New York.

Farkas MH, Grant GR, White JA, Sousa ME, Consugar MB, Pierce EA. 2013. Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics* **14:** 486.

Fonseca NA, Rung J, Brazma A, Marioni JC. 2012. Tools for mapping high-throughput sequencing data. *Bioinformatics* **28:** 3169–3177.

Gamsiz ED, Ouyang Q, Schmidt M, Nagpal S, Morrow EM. 2012. Genome-wide transcriptome analysis in murine neural retina using high-throughput RNA sequencing. *Genomics* **99:** 44–51.

Gatto A, Torroja-Fungairino C, Mazzarotto F, Cook SA, Barton PJ, Sanchez-Cabo F, Lara-Pezzi E. 2014. FineSplice, enhanced splice junction detection and quantification: A novel pipeline based on the assessment of diverse RNA-Seq alignment solutions. *Nucleic Acids Res* **42:** e71.

Gertz J, Varley KE, Davis NS, Baas BJ, Goryshin IY, Vaidyanathan R, Kuersten S, Myers RM. 2012. Transposase mediated construction of RNA-Seq libraries. *Genome Res* **22:** 134–141.

Gonzalez-Porta M, Calvo M, Sammeth M, Guigo R. 2012. Estimation of alternative splicing variability in human populations. *Genome Res* **22:** 528–538.

Grant GR, Farkas MH, Pizarro AD, Lahens NF, Schug J, Brunk BP, Stoeckert CJ, Hogenesch JB, Pierce EA. 2011. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics* **27:** 2518–2528.

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

Graziotto JJ, Farkas MH, Bujakowska K, Deramaudt BM, Zhang Q, Nandrot EF, Inglehearn CF, Bhattacharya SS, Pierce EA. 2011. Three gene-targeted mouse models of RNA splicing factor RP show late-onset RPE and retinal degeneration. *Invest Ophthalmol Vis Sci* **52:** 190–198.

Halvardson J, Zaghlool A, Feuk L. 2013. Exome RNA sequencing reveals rare and novel alternative transcripts. *Nucleic Acids Res* doi: 10.1093/nar/gks816.

Hardcastle TJ. 2010. baySeq: Empirical Bayesian analysis of patterns of differential expression in count data. *BMC Bioinformatics* **11:** 442.

Hollyfield JG, Kuttner-Kondo L. 2010. Animal models for age-related macular degeneration. In Animal models for retinal diseases (ed. Pang IH, Clark AF), pp. 81–98. Springer, New York.

Hong T, Tan AG, Mitchell P, Wang JJ. 2011. A review and meta-analysis of the association between C-reactive protein and age-related macular degeneration. *Surv Ophthalmol* **56:** 184–194.

Ivings L, Towns KV, Matin M, Taylor C, Ponchel F, Grainger RJ, Ramesar RS, Mackey DA, Inglehearn CF. 2008. Evaluation of splicing efficiency in lymphoblastoid cell lines from patients with splicing-factor retinitis pigmentosa. *Mol Vis* **14:** 2357.

Jaillard C, Mouret A, Niepon ML, Clerin E, Yang Y, Lee-Rivera I, Ait-Ali N, Millet-Puel G, Cronin T, Sedmak T. 2012. *Nxnl2* splicing results in dual functions in neuronal cell survival and maintenance of cell integrity. *Hum Mol Genet* **21:** 2298–2311.

Kandpal RP, Rajasimha HK, Brooks MJ, Nellissery J, Wan J, Qian J, Kern TS, Swaroop A. 2012. Transcriptome analysis using next generation sequencing reveals molecular signatures of diabetic retinopathy and efficacy of candidate drugs. *Mol Vis* **18:** 1123–1146.

Karlsson J, Torreno O, Ramet D, Klambauer G, Cano M, Trelles O. 2012. Enabling large-scale bioinformatics data analysis with cloud computing. In *Parallel and Distributed Processing with Applications (ISPA), 2012 IEEE 10th International Symposium*, pp. 640–645. IEEE.

Kent WJ. 2002. BLAT: The BLAST-like alignment tool. *Genome Res* **12:** 656–664.

Khan K, Logan CV, McKibbin M, Sheridan E, Elçioglu NH, Yenice O, Parry DA, Fernandez-Fuentes N, Abdelhamed ZIA, Al-Maskari A. 2012. Next generation sequencing identifies mutations in Atonal homolog 7 (ATOH7) in families with global eye developmental defects. *Hum Mol Genet* **21:** 776–783.

Kienzler R, Bruggmann R, Ranganathan A, Tatbul N. 2012. Large-Scale DNA sequence analysis in the cloud: A stream-based approach. In *Euro-Par 2011: Parallel Processing Workshops*, pp. 467–476. Springer, New York.

Kim J, Zhao K, Jiang P, Lu Z-X, Wang J, Murray J, Xing Y. 2012. Transcriptome landscape of the human placenta. *BMC Genomics* **13:** 115.

Klein R, Klein BEK, Moss SE. 2009. The Wisconsin epidemiological study of diabetic retinopathy: A review. *Diabetes Metab Rev* **5:** 559–570.

Kokkinaki M, Sahibzada N, Golestaneh N. 2011. Human induced pluripotent stem-derived retinal pigment epithelium (RPE) cells exhibit ion transport, membrane potential, polarized vascular endothelial growth factor secretion, and gene expression pattern similar to native RPE. *Stem Cells* **29:** 825–835.

Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, Peng S, Preuss TM, Wohlschlegel JA, Geschwind DH. 2009. Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* **462:** 213–217.

Koppelkamm A, Vennemann B, Lutz-Bonengel S, Fracasso T, Vennemann M. 2011. RNA integrity in post-mortem samples: Influencing parameters and implications on RT-qPCR assays. *Int J Legal Med* **125:** 1–8.

Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci* **94:** 13057–13062.

Li H, Ruan J, Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18:** 1851–1858.

Li J, Wang JJ, Yu Q, Wang M, Zhang SX. 2009. Endoplasmic reticulum stress is implicated in retinal inflammation and diabetic retinopathy. *FEBS Lett* **583:** 1521–1527.

Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26:** 493–500.

Li Y, Li-Byarlay H, Burns P, Borodovsky M, Robinson GE, Ma J. 2013. TrueSight: A new algorithm for splice junction detection using RNA-seq. *Nucleic Acids Res* **41:** e51.

Lindner R, Friedel CC. 2012. A comprehensive evaluation of alignment algorithms in the context of RNA-Seq. *PloS ONE* **7:** e52403.

Liu Q, Zhang Q, Pierce EA. 2010. Photoreceptor sensory cilia and inherited retinal degeneration. *Adv Exp Med Biol* **664:** 223–232.

Lorenz WW, Dean JFD. 2002. SAGE profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). *Tree Physiol* **22:** 301–310.

Lu J, Tomfohr J, Kepler T. 2005. Identifying differential expression in multiple SAGE libraries: An overdispersed log-linear model approach. *BMC Bioinformatics* **6:** 165.

Malone K, Sohocki MM, Sullivan LS, Daiger SP. 1999. Identifying and mapping novel retinal-expressed ESTs from humans. *Mol Vis* **5:** 5.

Mardis ER. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9:** 387–402.

Marguerat S, Bahler J. 2010. RNA-Seq: From technology to biology. *Cell Mol Life Sci* **67:** 569–579.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-Seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18:** 1509–1517.

Matlin AJ, Clark F, Smith CWJ. 2005. Understanding alternative splicing: Towards a cellular code. *Nat Rev Mol Cell Biol* **6:** 386–398.

Meena Kishore S, Bagavathi SP, Yun Ping L, Lee Pern C, Yiting Y, Pandjassarame K. 2005. Alternatively spliced human genes by exon skipping—A database (ASHESdb). *In Silico Biol* **5:** 221–225.

Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, Mattick JS, Rinn JL. 2012. Targeted RNA

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES

sequencing reveals the deep complexity of the human transcriptome. *Nat Biotech* **30:** 99–104.

Morgan JC, Chapman RW, Anderson PE. 2012. A next generation sequence processing and analysis platform with integrated cloud-storage and high performance computing resources. In *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, pp. 594–594. ACM.

Morozova O, Marra MA. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92:** 255–264.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5:** 621–628.

Mustafi D, Maeda T, Kohno H, Nadeau JH, Palczewski K. 2012. Inflammatory priming predisposes mice to age-related retinal degeneration. *J Clin Invest* **122:** 2989–3001.

Nekrutenko A, Taylor J. 2012. Next-generation sequencing data interpretation: Enhancing reproducibility and accessibility. *Nat Rev Genet* **13:** 667–672.

Neveling K, Collin RWJ, Gilissen C, van Huet RAC, Visser L, Kwint MP, Gijsen SJ, Zonneveld MN, Wieskamp N, de Ligt J. 2012. Next-generation genetic testing for retinitis pigmentosa. *Hum Mutat* **33:** 963–972.

Nikopoulos K, Gilissen C, Hoischen A, Erik van Nouhuys C, Boonstra FN, Blokland EAW, Arts P, Wieskamp N, Strom TM, Ayuso C. 2010. Next-generation sequencing of a 40 Mb linkage interval reveals *TSPAN12* mutations in patients with familial exudative vitreoretinopathy. *Am J Hum Genet* **86:** 240–247.

Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463:** 457–463.

Okamoto S, Takahashi M. 2011. Induction of retinal pigment epithelial cells from monkey iPS cells. *Invest Ophthalmol Vis Sci* **52:** 8785–8790.

Oshlack A, Robinson MD, Young MD. 2010. From RNA-Seq reads to differential expression results. *Genome Biol* **11:** 220.

Özgül RK, Siemiatkowska AM, Yücel D, Myers CA, Collin RWJ, Zonneveld MN, Beryozkin A, Banin E, Hoyng CB, van den Born LI. 2011. Exome sequencing and *cis*-regulatory mapping identify mutations in *MAK*, a gene encoding a regulator of ciliary length, as a cause of retinitis pigmentosa. *Am J Hum Genet* **89:** 253–264.

Ozsolak F, Milos PM. 2011. RNA sequencing: Advances, challenges and opportunities. *Nat Rev Genet* **12:** 87–98.

Pan Q, Bakowski MA, Morris Q, Zhang W, Frey BJ, Hughes TR, Blencowe BJ. 2005. Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet* **21:** 73–77.

Peng Z, Cheng Y, Tan BC-M, Kang L, Tian Z, Zhu Y, Zhang W, Liang Y, Hu X, Tan X. 2012. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat Biotechnol* **30:** 253–260.

Pennesi ME, Neuringer M, Courtney RJ. 2012. Animal models of age related macular degeneration. *Mol Aspects Med* **33:** 487–509.

Phillips MJ, Wallace KA, Dickerson SJ, Miller MJ, Verhoeven AD, Martin JM, Wright LS, Shen W, Capowski EE, Percin EF. 2012. Blood-derived human iPS cells generate optic vesicle-like structures with the capacity to form retinal laminae and develop synapses. *Invest Ophthalmol Vis Sci* **53:** 2007–2019.

Pickrell JK, Pai AA, Gilad Y, Pritchard JK. 2010. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet* **6:** e1001236.

Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, Betel D. 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-Seq data. *Genome Biol* **14:** R95.

Reis-Filho JS. 2009. Next-generation sequencing. *Breast Cancer Res* **11:** S12.

Robinson MD, Smyth GK. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9:** 321–332.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139–140.

Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S, Papavasiliou FN. 2011. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. *Nat Struct Mol Biol* **18:** 230–236.

Schmeer CW, Wohl SG, Isenmann S. 2012. Cell-replacement therapy and neural repair in the retina. *Cell Tissue Res* 1–12.

Schrader KA, Heravi-Moussavi A, Waters PJ, Senz J, Whelan J, Ha G, Eydoux P, Nielsen T, Gallagher B, Oloumi A. 2011. Using next-generation sequencing for the diagnosis of rare disorders: A family with retinitis pigmentosa and skeletal abnormalities. *J Pathol* **225:** 12–18.

Schuster SC. 2008. Next-generation sequencing transforms today's biology. *Nat Methods* **5:** 16–18.

Sharon D, Blackshaw S, Cepko CL, Dryja TP. 2002. Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE). *Proc Natl Acad Sci* **99:** 315–320.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nat Biotechnol* **26:** 1135–1145.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3:** 1544–6115.

Soneson C, Delorenzi M. 2013. A comparison of methods for differential expression analysis of RNA-Seq data. *BMC Bioinformatics* **14:** 91.

Stahl PD, Wainszelbaum MJ. 2009. Human-specific genes may offer a unique window into human cell signaling. *Sci Signal* **2:** pe59.

Strunnikova N, Maminishkis A, Barb J, Wang F, Zhi C, Sergeev Y, Chen W, Edwards A, Stambolian D, Abecasis G. 2010. Transcriptome analysis and molecular signature of human retinal pigment epithelium. *Hum Mol Genet* **19:** 2468–2486.

Sui GY, Liu GC, Liu GY, Gao YY, Deng Y, Wang WY, Tong SH, Wang L. 2012. Is sunlight exposure a risk factor for age-related macular degeneration? A systematic review and meta-analysis. *Br J Ophthalmol* **97:** 389–394.

Swaroop A, Zack D. 2002. Transcriptome analysis of the retina. *Genome Biol* **3:** reviews1022.1021–reviews1022.1024.

Syed F. 2010. Application of Nextera technology to RNA-seq library preparation. *Nature Methods* **7:** an2–an3.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131:** 861–872.

Tang J, Kern TS. 2011. Inflammation in diabetic retinopathy. *Prog Retin Eye Res* **30:** 343–358.

Tang S, Riva A. 2013. PASTA: Splice junction identification from RNA-Sequencing data. *BMC Bioinformatics* **14:** 116.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.

Tucker BA, Scheetz TE, Mullins RF, DeLuca AP, Hoffmann JM, Johnston RM, Jacobson SG, Sheffield VC, Stone EM. 2011. Exome sequencing and analysis of induced pluripotent stem cells identify the cilia-related gene male germ cell-associated kinase (MAK) as a cause of retinitis pigmentosa. *Proc Nat Acad Sci* **108:** E569–E576.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. *Science* **270:** 484–487.

Visser L. 2010. What's new in diabetic retinopathy?: Review. *S African J Diabetes Vasc Disease* **7:** 93–98.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* **10:** 57–63.

Wang Y, Guo L, Cai SP, Dai M, Yang Q, Yu W, Yan N, Zhou X, Fu J, Guo X. 2012. Exome sequencing identifies compound heterozygous mutations in CYP4V2 in a pedigree with retinitis pigmentosa. *PLoS ONE* **7:** e33673.

Westenskow PD, Moreno SK, Krohne TU, Kurihara T, Zhu S, Zhang Z, Zhao T, Xu Y, Ding S, Friedlander M. 2012. Using flow cytometry to compare the dynamics of photoreceptor outer segment phagocytosis in iPS-derived RPE cells. *Invest Ophthalmol Vis Sci* **53:** 6282–6290.

Won J, Shi LY, Hicks W, Wang J, Hurd R, Naggert JK, Chang B, Nishina PM. 2011. Mouse model resources for vision research. *J Ophthalmol* doi: 10.1155/2011/391384

Zernant J, Schubert C, Im KM, Burke T, Brown CM, Fishman GA, Tsang SH, Gouras P, Dean M, Allikmets R. 2011. Analysis of the ABCA4 gene by next-generation sequencing. *Invest Ophthalmol Vis Sci* **52:** 8479–8487.

Zhang Y, Erdmann J, Chilton J, Onsongo G, Bower M, Beckman K, Thyagarajan B, Silverstein K, Lamblin AF. 2012. CLIA-certified next-generation sequencing analysis in the cloud. In *BMC Proceedings*, p. P54. BioMed Central Ltd.

Cold Spring Harbor Perspectives in Medicine
www.perspectivesinmedicine.org
CSH PERSPECTIVES