j ocul biol dis inform (2009) 2:161–163
DOI 10.1007/s12177-009-9048-0

EDITORIAL

Bioinformatics and the eye

Justine R. Smith

Received: 9 November 2009 / Accepted: 11 November 2009 / Published online: 22 January 2010

Bioinformatics evolved to become a recognized scientific discipline during the last decade of the twentieth century, due in part to the Human Genome Project and the introduction of microarray expression profiling, both of which involved critical computational activities. By 2000, the National Institutes of Health had established the Biomedical Information Science and Technology Initiative (BISTI), which defined bioinformatics as "Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data [1]." Otherwise stated, the methods of bioinformatics allow the biomedical scientist to more completely interpret his data. Bioinformatics was already an important aspect of ophthalmic and vision research in 2000. For example, following early descriptions of microarray expression profiling in the mid 1990s [2], this methodology was quickly applied to studies of development and aging of the retina [3,4]. The current issue of the *Journal of Ocular Biology, Diseases and Informatics* highlights the role that bioinformatics is playing in various fields of eye-related research today. Scientific reports and reviews contributed from an international group of laboratories demonstrate the use of different computational techniques in studies of normal biology and diseases of the eye.

Supported in part by the Schnitzer-Novack Foundation and Research to Prevent Blindness

J. R. Smith (✉)
Casey Eye Institute, Oregon Health & Science University,
3375 SW Terwilliger Blvd.,
Portland, OR 97239, USA
e-mail: smithjus@ohsu.edu

One could easily argue that the most significant breakthroughs in the field of ocular genomics within the last 5 years pertain to age-related macular degeneration (AMD). In 2005, descriptions of the association between single-nucleotide polymorphisms (SNPs) in the complement factor H (CFH) gene and AMD [5–8] were headline articles in both the scientific and popular literature. Manifestations of AMD differ according to ethnicity, however. In Japanese persons, in comparison to the predominantly Caucasian populations that were studied in these landmark studies, relatively more cases of AMD are of the "wet" phenotype, including the variant known as polypoidal choroidal vasculopathy, and relatively less cases are "dry". In this issue, Goto and Akahori et al. [9] report results of a genome-wide scan in 200 Japanese patients with typical wet or polypoidal choroidal AMD and a similar number of unaffected control subjects. They use microarrays designed to detect approximately 500,000 SNPs, and in addition, consider SNPs not represented on the array, but previously associated with AMD. By applying standard association statistics to data thus obtained, they make the significant observation that associations of AMD with SNPs in the AMRS2/HTRA1 region are considerably stronger than associations with SNPs in the CFH region in the Japanese population.

The issue contains four papers that illustrate the application of microarray-based transcriptomic profiling in different fields of ocular research. Kozulin and Provis [10] demonstrate that gene expression profiling by oligonucleotide array may yield substantial information from rare tissue samples. Using arrays with probes designed to detect over 47,000 transcripts, they characterize gene expression in the human macula during mid-gestation. In addition to standard processing of data generated from macular versus nasal and peripheral retinal tissue (i.e., normalization and

differential expression analysis of data), they perform secondary analyses that include biological process clustering and pathway analysis. These additional analyses provide fascinating observations about macular development, including the high expression of multiple cell adhesion molecules in this region of the retina. As is common in microarray studies, novel observations stimulate new hypotheses for future investigations; the authors suggest that adhesion molecules may participate in the formation of retinal circuitry, in axon guidance, and/or in protection of the macula against the retinal stretch that occurs during development.

Two reports discuss the use of gene expression microarray to elucidate the pathogenesis of retinal vascular diseases. Tea et al. [11] use oligonucleotide arrays representing approximately 27,000 genes, to study retinal gene expression early in the course of the murine model of retinopathy of prematurity that is known as oxygen-induced retinopathy. When comparing transcriptomes, differential gene expression may be analyzed by one of two approaches: according to fold-change (often set at 2-fold or 1.5 fold); or according to statistical significance (often defined as false discovery rate below 5%). In this work, both criteria are applied, and fold-change yields the most informative results. The authors find alterations in levels of multiple transcripts regulated by hypoxia-inducible factor 1-alpha in retinas of neonatal mice exposed to cyclic hyperoxia in comparison to non-exposed control animals. As anticipated, more transcripts are differentially expressed in the retina of the susceptible Sprague-Dawley strain than the resistant Fisher 344 strain. Freeman et al. [12] employ similar microarrays to study gene expression after established hyperglycemia in two different models of diabetic retinopathy in the C57BL/6J mouse. Disease is either induced by exposure to streptozotocin or arising spontaneously in mice heterozygous for the $Ins2^{Akita}$ mutation. The investigators perform secondary network analyses to classify genes that are differentially expressed following exposure to hyperglycemia into functional groups. The importance of considering different models is clear from the finding that different networks predominate in each model: in $Ins2^{Akita}$ heterozygote mice, inflammatory processes are identified; in streptozotocin-exposed mice, genes with known responsiveness to diabetes are highlighted. Apart from providing insights into the development of the disease, the authors justifiably stress the relevance of their work as a method for discovery of new biomarkers of disease.

Ocular toxoplasmosis is one of the most common retinal infections in both Western and developing nations, with the potential to cause irreversible blindness [13]. Many basic aspects of the interaction between the host and the causative protozoan, *Toxoplasma gondii*, remain poorly understood. However, microarray profiling has recently been applied to this subject, and important advances in this field have occurred. Brown and Blader [14] review findings from these expression profiling studies, covering topics that include: determinants of parasite virulence; replication of the parasite within a host cell; conversion between the actively replicating tachyzoite form and the dormant bradyzoite form; and host immune responses to limit the severity of the ocular pathology.

Shotgun proteomics is an emerging profiling technology, offering the possibility of identifying thousands of proteins within a complex biological sample. Wilmarth et al. [15] present a new "proteomic analysis workflow pipeline" for identification of proteins contained in tissue samples that have been first digested with trypsin and subsequently subjected to multi-dimensional chromatographic separation and tandem mass spectrometry. To illustrate their approach, the authors profile lens proteins from human, mouse, cow and chicken. They employ sequence-reversed decoy databases and a discriminant function transformation of SEQUEST software to distinguish real from false peptide and protein identifications, and they consider strategies involving two public protein databases—the International Protein Index and UniProt—and tryptic versus no-enzyme searches. Using this method, the authors identify novel lens crystallins, as well as non-crystallin proteins, that are not detected by the more conventional gel-based 2-dimensional electrophoresis.

Metabolomics is the study of low molecular weight metabolites within a biological sample. Young and Wallace [16] review this relatively under-utilized technology, including an appropriate statistical method for data analysis, known as principal component analysis, which is also useful in gene expression microarray studies. The field of metabolomics has pushed ahead recently as a result of advances in nuclear magnetic resonance and mass spectrometry, and has great potential to identify clinically relevant biomarkers. Most notably the methodology has been used to successfully predict the degree of coronary stenosis in human subjects. [17] The authors also discuss the use of metabolomics for the differential diagnosis of inflammatory eye disease.

In any bioinformatics study, the quality of the output data are limited by the quality of the input data. One aspect of quality assurance in microarrays is addressed by Harrington et al. [18], who present a method for analysis of RNA quality that relies on bioanalyzer electropherograms. They illustrate this method with analyses of RNA extracts that have been obtained from various ocular cells and tissues. Of particular relevance to eye-related studies, the authors highlight the presence of unusual RNA peaks in lacrimal gland and acinar samples. They attribute these peaks to highly abundant transcripts associated with the secretory function of the tissue, identifying one of these as

lipocalin 1. Similar to the situation encountered when working with RNA isolates from whole blood, which contain high levels of hemoglobin transcript, these secretion protein transcripts may reduce the sensitivity of microarray studies of lacrimal cells and tissues. The authors suggest that measures to specifically prevent the labeling of these transcripts may improve the quality of the output data in this setting.

As the eight papers in this issue illustrate, bioinformatics is a powerful tool in diverse studies of ocular biology and eye disease. Although the issue focuses on the application of bioinformatics to molecular data that have been generated in genomic, transcriptomic, proteomic and metabolomic studies, computational approaches also benefit other areas of eye-related research, including, for example, investigations that involve image analysis. Clearly, the power of the tools lies in the software, and as more sophisticated software programs are developed, one can expect to obtain even more information from one's data sets. In order for such development to occur in ophthalmic and vision research, however, it is vital that ocular scientists maintain collaborations with statisticians and computer programmers throughout the course of their investigations.

References

1. http://www.bisti.nih.gov/docs/CompuBioDef.pdf.
2. Ramsay G. DNA chips: state-of-the art. Nat Biotechnol. 1998;16:40–4.
3. Livesey FJ, Furukawa T, Steffen MA, Church GM, Cepko CL. Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx. Curr Biol. 2000;10:301–10.
4. Shelton DN, Chang E, Whittier PS, Choi D, Funk WD. Microarray analysis of replicative senescence. Curr Biol. 1999;9:939–45.
5. Hageman GS, Anderson DH, Johnson LV, et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. Proc Natl Acad Sci USA. 2005;102:7227–32.
6. Edwards AO, Ritter R 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. Science. 2005;308:421–4.
7. Haines JL, Hauser MA, Schmidt S, et al. Complement factor H variant increases the risk of age-related macular degeneration. Science. 2005;308:419–21.
8. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308:385–9.
9. Goto A, Akahori M, Okamoto H, et al. Genetic analysis of typical wet-type age-related macular degeneration and polypoidal choroidal vasculopathy in the Japanese population. J Ocul Biol Dis Inform. doi:10.1007/s12177-009-9047-1.
10. Kozulin P, Provis JM. Differential gene expression in the developing human macula: microarray analysis using rare tissue samples. J Ocul Biol Dis Inform. doi:10.1007/s12177-009-9039-1.
11. Tea M, Fogarty R, Brereton HM, et al. Gene expression microarray analysis of early oxygen-induced retinopathy in the rat. J Ocul Biol Dis Inform. doi:10.1007/s12177-009-9041-7.
12. Freeman WM, Bixler GV, Brucklacher RM, et al. Transcriptomic comparison in the retina of two mouse models of diabetes. J Ocul Biol Dis Inform. doi:10.1007/s12177-009-9045-3.
13. Holland GN. Ocular toxoplasmosis: a global assessment. Part 1: epidemiology and course of disease. Am J Ophthalmol. 2003;136:973–88.
14. Brown KM and Blader IJ. The role of DNA microarrays in Toxoplasma gondii research, the causative agent of ocular toxoplasmosis. J Ocul Biol Dis Inform. doi:10.1007/s12177-009-9040-8.
15. Wilmarth PA, Riviere MA, David LL. Techniques for accurate protein identification in shotgun proteomic studies of human, mouse, bovine and chicken lenses. J Ocul Biol Dis Inform. doi:10.1007/s12177-009-9042-6.
16. Young SP, Wallace GR. Metabolomic analysis of human disease and its application to the eye. J Ocul Biol Dis Inform. doi:10.1007/s12177-009-9038-2.
17. Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HW, et al. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med. 2002;8:1439–44.
18. Harrington C, Winther M, Garred MM. Use of Bioanalyzer electropherograms for quality and target evaluation in microarray expression profiling studies of ocular tissues. J Ocul Biol Dis Inform. doi:10.1007/s12177-009-9046-2.