Module 3: Gene Expression and Proteomics

Purpose of the module

Sight-threatening diseases such as diabetic retinopathy, retinopathy of prematurity, glaucoma, retinal ischemic disease, cataract, retinitis pigmentosa, corneal disruption and keratoconus are frequently accompanied by genetic mutations and/or dysfunctional proteins. The mechanisms underlying disruption and/or dysfunction are not necessarily obvious, however deciphering disease mechanisms is a major emphasis of VDI research. In vision research, as in many areas of science/medicine, there has been an explosion in sophisticated technology that yields very large datasets. The onslaught of high-throughput data constitutes a major challenge for bench researchers. This module offers expertise and guidance in gene expression studies and proteomic analysis and facilitates analysis of complex data in a format that is tailored to the needs of vision researchers.

Module director/co-director contact information:

Director: Dr. Yutao Liu (YUTLIU@augusta.edu), phone 706-721-2015
Co-director: Dr. Ashok Sharma (ASSHARMA@augusta.edu), phone 706-721-6335

Consultation on gene expression and proteomic studies

CONSULTATION ON EXPERIMENTAL DESIGN: Dr. Liu is an NEI-R01 funded senior investigator in the Department of Cellular Biology and Anatomy. He has expertise in human genetics/functional genomics related with DNA/RNA sequencing using different platforms. He established strong relationships with many vendors and understands advantages/limitations of different sequencing platforms in biomedical research. He assists with experimental design consultation related with case/control, time series, and multiple groups design as well as experimental platform/sequencing platform/kits choices. Please contact him at the email or phone number listed above.

CONSULTATION ON DATA ANALYSES: Dr. Sharma is an NEI R01-funded investigator with research expertise in development of algorithms for biomarker discovery, using genomic and proteomic data analysis and integration. He has expertise in bioinformatics and big data analysis including computational and biological aspects of biomedical research. He is ideally suited to guide users in data analysis for the gene expression/proteomics studies of this module. Please contact him at the email or phone number listed above.

Assistance with sample preparation and statistical analysis

This module is staffed by two PhD level scientists who can guide users in preparing samples and appropriate methods for analysis of the data.

Jingwen Cai, PhD, (JINCAI@augusta.edu), phone 706-721-4843
Tae Jin Lee, PhD, (TALEE@augusta.edu), phone 706-721-3515

Jingwen Cai, PhD is a research associate with experience in human genetics and genomics. She will work closely with core investigators to extract RNA samples using a variety of different reagents and to evaluate RNA quality/quantity using Tecan and Bioanalyzer RNA chips. She will run expression validation/assays and assist with experimental troubleshooting. She will also guide users in protein extraction as needed. Dr. Cai will participate in gene expression consultation sessions with core directors. Please contact her at the email or phone number listed above.

Tae Jin Lee, PhD is a trained biostatistician with emphasis on mathematical modeling of biological processes. He is an expert in the use of informatics tools such as FASTQC, Bowtie2, Protein Discoverer and Skyline for RNA-Seq and proteomic data analyses. He utilizes R packages such as DESeq2 and edgeR to extract differentially expressed genes from RNA-Seq data. Given his background in statistics, he is poised to guide users in the appropriate statistical methods for analysis of proteomics and gene expression data. Please contact him at the email or phone number listed above.

Description of services:

Gene Expression

EXPERIMENTAL DESIGN CONSULTATION: Investigators first consult with Dr. Yutao Liu to discuss gene expression studies. Biological questions will be ascertained, after which the study design will be planned with considerations of case-control, time series, multiple control groups. This will inform the choice of the appropriate experimental platform for expression profiling. Studies may have varying yields of RNA ranging from a few nanograms to micrograms, necessitating specific quality control procedures and library preparation schemes. A number of considerations may guide the choice of the sequencer, which Dr. Liu will provide.

GUIDANCE FOR SAMPLE PREPARATION/DATA ACQUISITION: Dr. Liu will arrange/participate in a meeting between the vision researcher and the personnel that run samples in the Integrated Genomics Shared Resource core. This meeting will determine: appropriate sequencing technology, suitable sequencing library preparation kit, RNA preparation kit/ technology/RNA amount available for sequencing, estimated timeline of the experiment, and estimated cost of the project. Assistance with RNA isolation will be provided by research associate, Jinwen Cai, PhD. Please note, the services of the Integrated Genomics Core must be borne by the individual investigator.

BIOINFORMATICS: After sequencing, investigators will consult with Dr. Ashok Sharma. The consultation will include: (1) Processing sequencing reads (including alignment), (2) Estimation of individual gene expression levels, (3) Normalization, (4) Identification of differentially expressed (DE) genes, (5) Cluster analysis/Pathway Analysis. To improve our understanding of visual diseases at system level, a network analysis is performed to identify biological processes synergistically contributing to the disease. These analyses facilitate interpretation of data and formulation of novel hypotheses. Bioinformatics tools (Ingenuity Pathway Analysis, Cytoscope, Bioconductor packages) are used to connect the identified genes/proteins with the biological pathways involved during the pathogenesis of ocular disease.

RNA-Seq Data Analysis**:** Raw data will be acquired in FASTQ file format and quality control analysis will be performed using FastQC. Subsequently, raw data reads will be mapped to reference genome using Bowtie2, a well-known short read alignment tool that uses Burrows-Wheeler Transformation. It outputs BAM files for each sample, which will be indexed using SAMtools. The case/control samples will be tested for differential gene expression using DESeq2 implemented in R and finally FDR adjusted p-values will be used to filter the genes with significant changes in expression. Statistical analyses will be performed using the R language and

environment for statistical computing (version 3.4.1; R Foundation for Statistical Computing; www.r-project.org).

Validation/confirmation: Upon identification of DE genes, investigators will be advised on appropriate methodologies to validate/confirm their findings. For gene expression, realtime PCR or droplet digital PCR can be used to validate original findings or to replicate findings. For long coding RNAs and miRNAs, Dr. Liu will consult with investigators on how to select appropriate technologies and design specific assays for validation.

Proteomics

EXPERIMENTAL DESIGN CONSULTATION: Investigators will consult with Dr. Yutao Liu to discuss proteomic analysis. During this meeting the biological question will be established, appropriate sample size and controls determined, and advantages/limitations of the technology discussed.

GUIDANCE FOR SAMPLE PREPARATION/DATA ACQUISITION: Core investigators (or their laboratory personnel) will be advised on best methods to prepare samples for LC/MS-MS analyses under the guidance of Research Programmer Analyst TaeJin Lee, PhD, a staff member of the module. Dr. Sharma will arrange/participate in a meeting between the vision researcher and the personnel that run samples in the Proteomics Core Laboratory at Augusta University. This meeting will determine estimated timeline of the experiment, and estimated cost of the project. Please note, the services of the Proteomics Core must be borne by the individual investigator. Samples will be provided to the proteomics core and standard methods will be used by proteomics core personnel to digest/clean protein samples and then analyze using the Orbitrap Fusion Tribrid mass spectrometer coupled with an Ultimate 3000 nano-UPLC system.

BIOINFORMATICS: Following LC-MS/MS analyses, investigators will consult with Dr. Ashok Sharma for protein identification and quantification analyses. Raw MS data is processed using Proteome Discoverer software (v.2.2) and submitted for SequestHT search against the Uniprot human database. The percolator peptide spectrum matching (PSM) algorithm will be used for validation. SequestHT search parameters will be set as 10 ppm precursor and 0.6 Da product ion tolerance, with static carbidomethylation (+57.021 Da) for cysteine and dynamic oxidation (+15.995 Da) for methionine and dynamic phosphorylation (+79.966 Da) for serine, threonine and tyrosine. Proteins that contain similar peptides, which cannot be differentiated based on MS/MS analysis alone will be grouped to satisfy the principles of parsimony. Proteins sharing significant peptides will be grouped into clusters. A protein report comprising the identities and spectrum counts (number of PSM) for each protein will be exported as a semi-quantitative measure for relative protein levels in all samples.

PATHWAY AND NETWORK ANALYSIS will be performed using the previously-mentioned tools for protein expression. The analyses will provide insights about molecular function/protein networks involved in specific ocular disorders.

VALIDATION/CONFIRMATION: The selected proteins from the discovery step could be evaluated using more targeted medium-low throughput methods such as western blotting, ELISA, customized protein array, and parallel reaction monitoring mass spectrometry (PRM-MS). PRM-MS is an ion monitoring technique based on high resolution and high precision MS to provide absolute quantification of proteins and peptides. PRM-MS experiments will be performed on the same LC-MS/MS platform as in discovery step using the same LC elution conditions. The MS will be running in the targeted MS mode for quantification purpose. PRM assay will generate the full fragment spectra, which can be used to further confirm the identity of the target peptide, meanwhile, one signature fragment (generally the most intense fragment) for each candidate peptide will be selected to

calculate the peak area on the extracted ion chromatograph for that peptide. The peak area for each peptide will then be normalized by spiked internal standard (stable isotopic counterpart of each candidate peptide) across different samples to compensate for possible experimental variations.

Costs associated with module use

The module provides the following service free of charge for all P30 users:

- ✓ Consultation on experimental design
- ✓ Arrangement on sequencing and proteomics platform selection
- ✓ Professional advice on RNA and protein sample preparation
- ✓ Data analysis of RNA-Seq and proteomics
- ✓ Bioinformatics analyses (Pathway Analysis, Network analysis)

The cost of proteomics run using mass spectrometry and RNA-Sequencing including library preparations as well as data storage is of the responsibility of individual investigators.

Citation

Please cite the NEI Center Core Grant for Vision Research - **P30EY031631** in your publications if you use the services offered in this module.