

Exploring RNA-Seq Expression Data With GxSeq

Nicholas Thrower,^{1,2*} (throwern@msu.edu), Matthew Larson,^{1,2} and Curtis Wilkerson,^{1,2}

¹Michigan State University, East Lansing; ²Great Lakes Bioenergy Research Center, East Lansing, Michigan

<https://gxseq.glbrc.org>

Project Goals: The Great Lakes Bioenergy Research Center is performing basic research to identify new technologies for the conversion of cellulosic biomass to biofuel. One goal is to develop biomass crops with desirable traits leading to more efficient fuel production. Understanding the genetic background for these desirable traits is possible due to advances in high throughput sequencing platforms. RNA-Seq or whole transcriptome sequencing is one of these advances and is becoming a standard for gene expression studies. To assist with storage and analysis of this sequence data, we created a custom web based application called GxSeq. The project began as a tool for investigating gene expression profiles within de-novo assemblies of novel organisms. Over time it has grown to include analyzing and interacting with fully annotated and sequenced genomes.

We have developed a web application (GxSeq) designed to facilitate RNA-Seq expression analysis and its variants. This software provides an array of interactive tables and visualizations customized to the needs of our researchers. The software also provides these researchers with a standard tool to access their data independent of the choice of upstream analysis tools. This reduces the overall complexity of data analysis and so reduces the workload for individual research groups allowing them to focus on experimentation. This software is developed using agile software methodologies, which allows us to quickly adapt the software to the specific needs of individual researchers. By having a close collaboration with researchers and integrating new requests in an expedient manner we are able to accommodate the rapidly changing nature of genomic data analysis. Additionally, this tool allows for center-wide access to large-scale datasets and information derived from analysis of these datasets.

Exploring RNA-Seq data often involves comparison of expression levels between genes or groups of genes. GxSeq has several tools designed to assist in this exploratory analysis including a tabular expression view, dynamic expression charts, k-means clustering and dynamic gene correlation tables. The tabular view can generate a matrix of counts or a ratio from user-selected samples. Interactive clustering allows exploration of gene expression patterns and dynamic gene correlation tables enable deeper exploration of the gene network. Tabular results are displayed alongside functional gene annotations in searchable, sortable tables that can be downloaded for further investigation.

It is also common to explore genes in context of the chromosomal sequence. This genomic view displays annotations and experimental data in a browsable interface. Interactive elements are

intuitive to use and allow dragging along the genome, selecting sequence regions and viewing annotation details. The URL is updated to reflect any changes allowing users to share or save the current view.

The combination of genomic sequence and gene expression data in GxSeq enables researchers to more effectively study the genetic background of desirable biomass traits.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494)