Project Overview

The main idea of our project is to test if co-expressed genes will exhibit correlated evolutionary rates, suggesting functional similarities between genes and therefore exposure to similar selection pressures.

Project Thesis

Our goal is to use the computational analyses WGCNA (Weighted Gene Co-Expression Network Analysis) and ERC (Evolutionary Rates Covariation) to explore if co-expressed genes in a network have correlated evolutionary rates. We will be exploring our hypothesis across seven different animal species (chicken, mouse, human, rhesus, rat, possum, and rabbit).

Methods

We obtained data consisting of gene expression across different types of tissues, developmental stages, and samples for *Gallus gallus*, *Homo sapien*, *Mus domesticus*, *Macaca mulatta, Mus musculus, Oryctolagus cuniculus*, and *Rattus norvegicus*. Using this data we generated metadata that contained columns specifying tissue type, developmental stage, sample number and species. We used the metadata and individual species gene expression data to practice running intraspecies Weighted Gene Co-expression Network Analysis (WGCNA) for Mmul and Ggal. Of these analyses, gene expression data were categorized by tissue type and developmental stage, allowing us to examine how gene co-expression networks differ within species across various biological conditions. WGCNA was conducted using standard procedures in R, involving data normalization, module identification, and network construction based on gene expression values.

At this point we obtained orthogonal gene data for all seven species that we could use for an interspecies WGCNA. From our ortholog dataset, we found that there were only 338 genes comparable across all respective species. We are currently generating a phylogenetic tree to help determine which species has the least number of comparable orthologs, using Ensembl BioMart to establish the number of shared genes between taxa. We predict that Mmul may need to be removed from our dataset when we perform any further computational analysis, as this species has the fewest orthologs.

Ultimately our interspecies WGCNA will be used to construct co-expression networks for each species and compare their network structures to assess conservation of co-expression modules across evolutionary lineages. After we perform a species-wide WGCNA, we will produce using Evolutionary Rate Covariation (ERC) to examine co-evolution rates between genes we find with similar expression across species. For this analysis, we have generated a gene ID to protein ID table for each species using Biomart. By integrating gene expression networks with ERCs, we aim to identify functional modules that exhibit both conserved co-expression and correlated evolutionary rates.

Results

We are using WGCNA to identify clusters of co-expressed genes in modules and extract key genes from these modules. After completing WGCNA, we will use the data to perform ERC, which will help us assess relationships between different modules by clustering the module eigengenes. By computing the module eigengenes for each cluster, ERC will allow us to identify modules with similar eigengene profiles, which could indicate biologically related processes across different genes or pathways.

Author Contributions

Atiana: Data wrangling, generated Metadata and practiced WGCNA with Ggal and developmental stage.

Connor: Data wrangling, data and project organization, Biomart, practice WGCNA, ERC Exploration

Salvador: Data wrangling, created filtered gene sets, project organization, began intra WGCNA analysis

Triana: Data wrangling, created expression data sets, Bio Mart, began intro to WGCNA analysis