May 4th, 11:00 AM - 1:00 PM

Using WGCNA to Examine RNASeq Expression Data in the Annual Killifish Austrofundulus Limnaeus

Patrick Clouser
Portland State University

Jason Podrabsky
Portland State University

Clouser, Patrick and Podrabsky, Jason, "Using WGCNA to Examine RNASeq Expression Data in the Annual Killifish Austrofundulus Limnaeus" (2022). *Student Research Symposium*. 12.
https://pdxscholar.library.pdx.edu/studentsymposium/2022/posters/12

Using WGCNA to examine RNASeq Expression data in the annual killifish *Austrofundulus limnaeus*

Patrick R Clouser, and Jason E. Podrabsky
The Center for Life in Extreme Environments
Portland State University, Portland OR 97201
Contact: **Pat6@pdx.edu**

Abstract

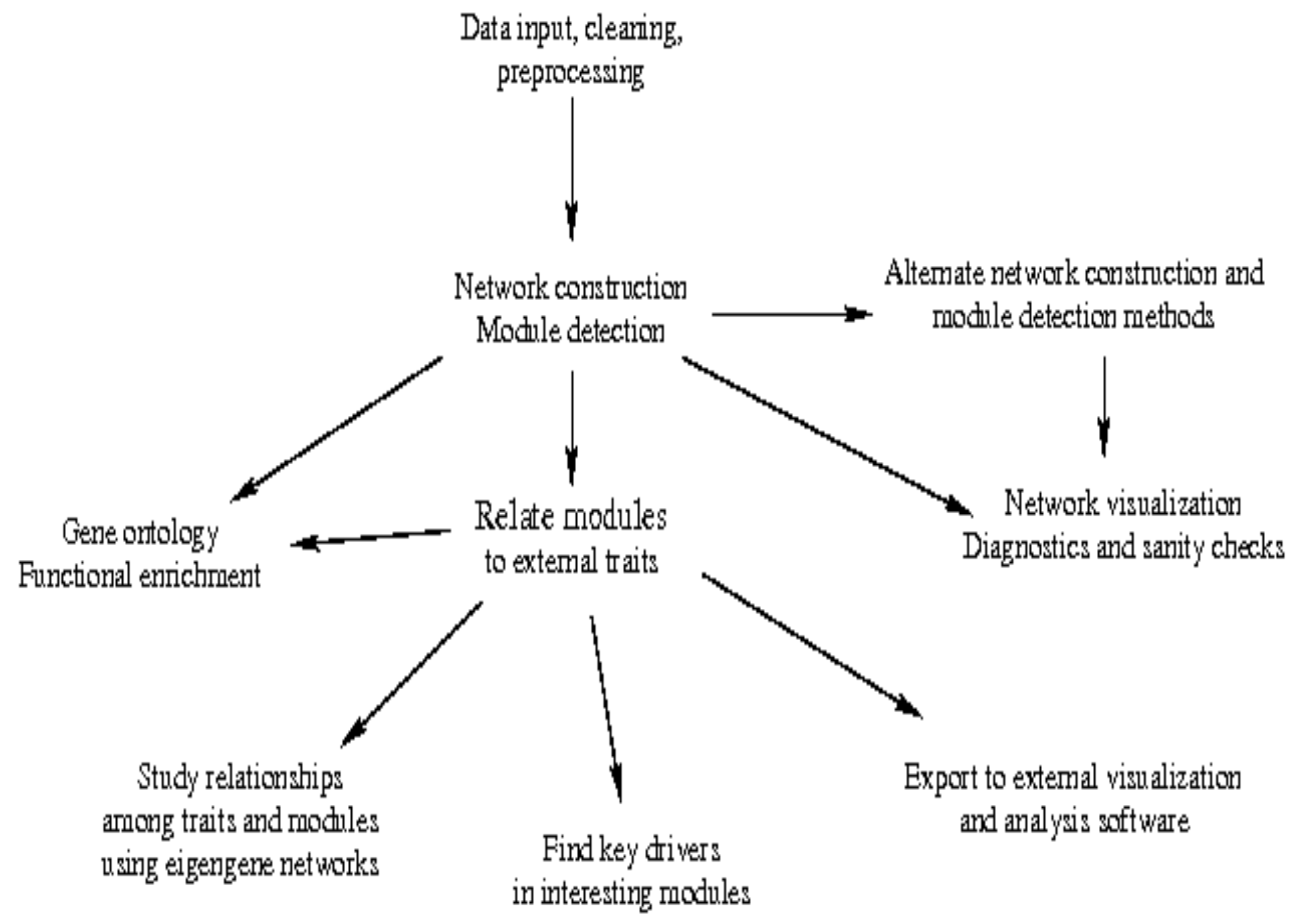
Annual killifish (*Austrofundulus limnaeus*) live in ephemeral ponds of Venezuela. They have evolved a unique life history that includes experience embryonic diapause; a period of developmental dormancy and metabolic arrest that supports survival during the dry season. Developing and diapausing embryos are extremely resistant to environmental stresses; for example, they can survive for months without oxygen (anoxia). *A. limnaeus*' tolerance to anoxia is an important survival mechanism as they can spend months buried in anoxic soil during development. I hypothesize that this extreme anoxia tolerance is supported by gene expression networks that are different from typical vertebrates. To evaluate this hypothesis, I will use weighted gene correlation network analysis (WGCNA) to evaluate RNAseq data from embryos exposed to anoxia, and recovery from anoxia. WGCNA describes patterns of correlation among genes across the experimental treatments, and creates modules of statistically correlated genes. These modules can be compared to one another and across different developmental stages that have differing tolerances of anoxia. I hope to discover unique gene expression modules involved with anoxia tolerance. Understanding how cells can survive without oxygen could lead to critical breakthroughs in medicine and emerging biotechnologies, such as treatments to mitigate damage from heart attacks or strokes.

Background: RNASeq Data From Annual Killifish Embryos

	Control (normoxia)	Short Anoxia (4 hours)	Long Anoxia (24 hours)	Short recovery (2 hours Normoxia)	Long recovery (24 hours Normoxia)
0 days post diapause (dpd)	4	4	4	4	4
4 dpd	4	4	4	4	4
12 dpd	6	6	6	6	6
20 dpd	4	4	4	4	4

Figure 1: The number of transcriptomes (RNA library) made at each of the 5 conditions and 4 stages of development. Each RNA library had approximately 27,000 Loci.

Weighted Gene Co-Expression Network Analysis



Data input, cleaning, preprocessing

Network construction
Module detection

Alternate network construction and module detection methods

Gene ontology
Functional enrichment

Relate modules to external traits

Network visualization
Diagnostics and sanity checks

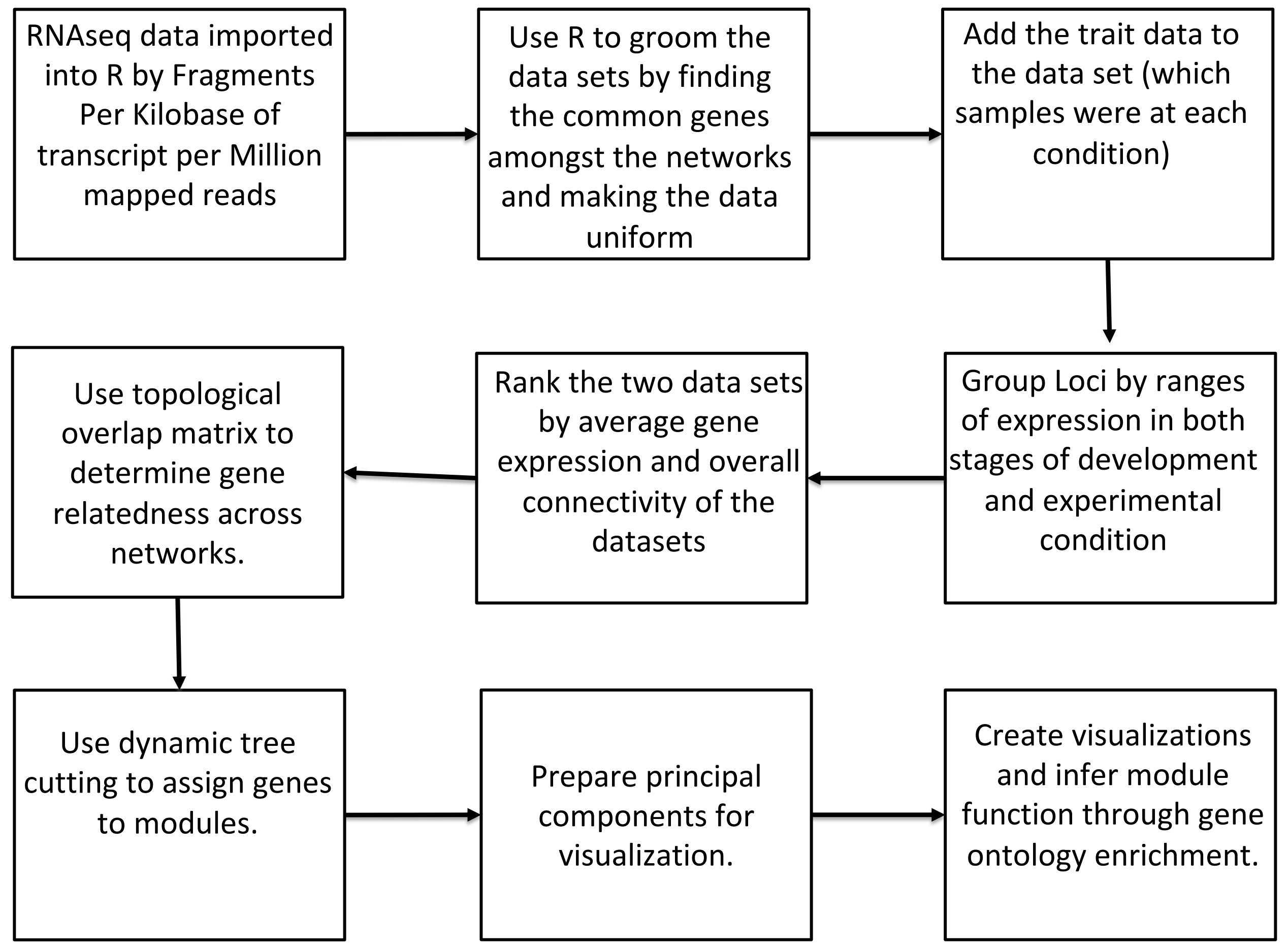
Study relationships among traits and modules using eigengene networks

Find key drivers in interesting modules

Export to external visualization and analysis software

Figure 2: A flow chart of WGCNA shows the workflow and outcomes that can be derived from sequencing data. A powerful tool that utilizes both experimental conditions and high-power computing.

Methods: The General Steps of WGCNA



RNAseq data imported into R by Fragments Per Kilobase of transcript per Million mapped reads

Use R to groom the data sets by finding the common genes amongst the networks and making the data uniform

Add the trait data to the data set (which samples were at each condition)

Group Loci by ranges of expression in both stages of development and experimental condition

Rank the two data sets by average gene expression and overall connectivity of the datasets

Use topological overlap matrix to determine gene relatedness across networks.

Use dynamic tree cutting to assign genes to modules.

Prepare principal components for visualization.

Create visualizations and infer module function through gene ontology enrichment.

Results: 4dpd to 12dpd Network Analysis

- WGCNA pipeline analysis was run to compare the 4 dpd network to the 12 dpd network.

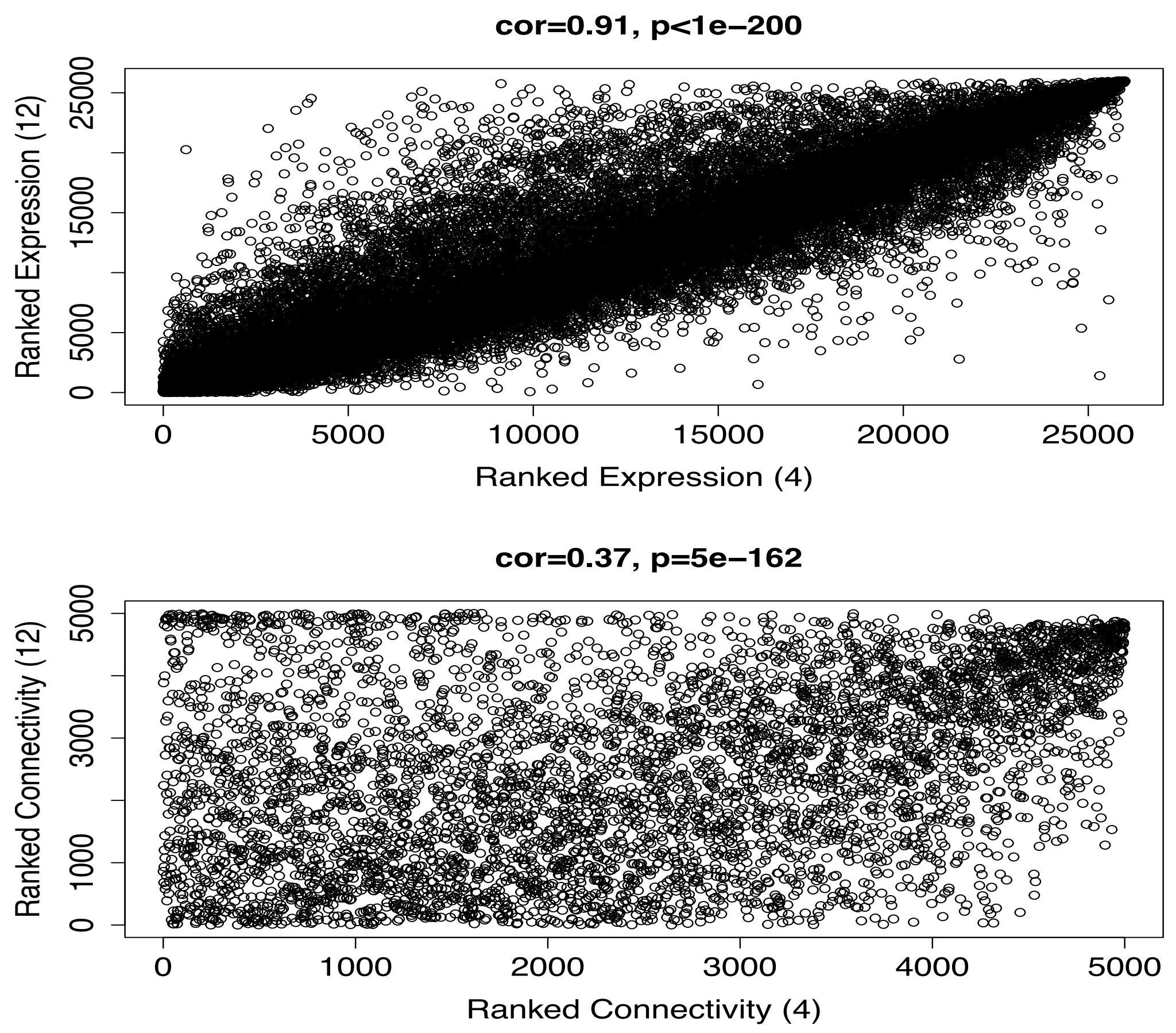


Figure 3: The 4dpd and 12 dpd networks are compared to show ranked expression and ranked connectivity across all Loci.

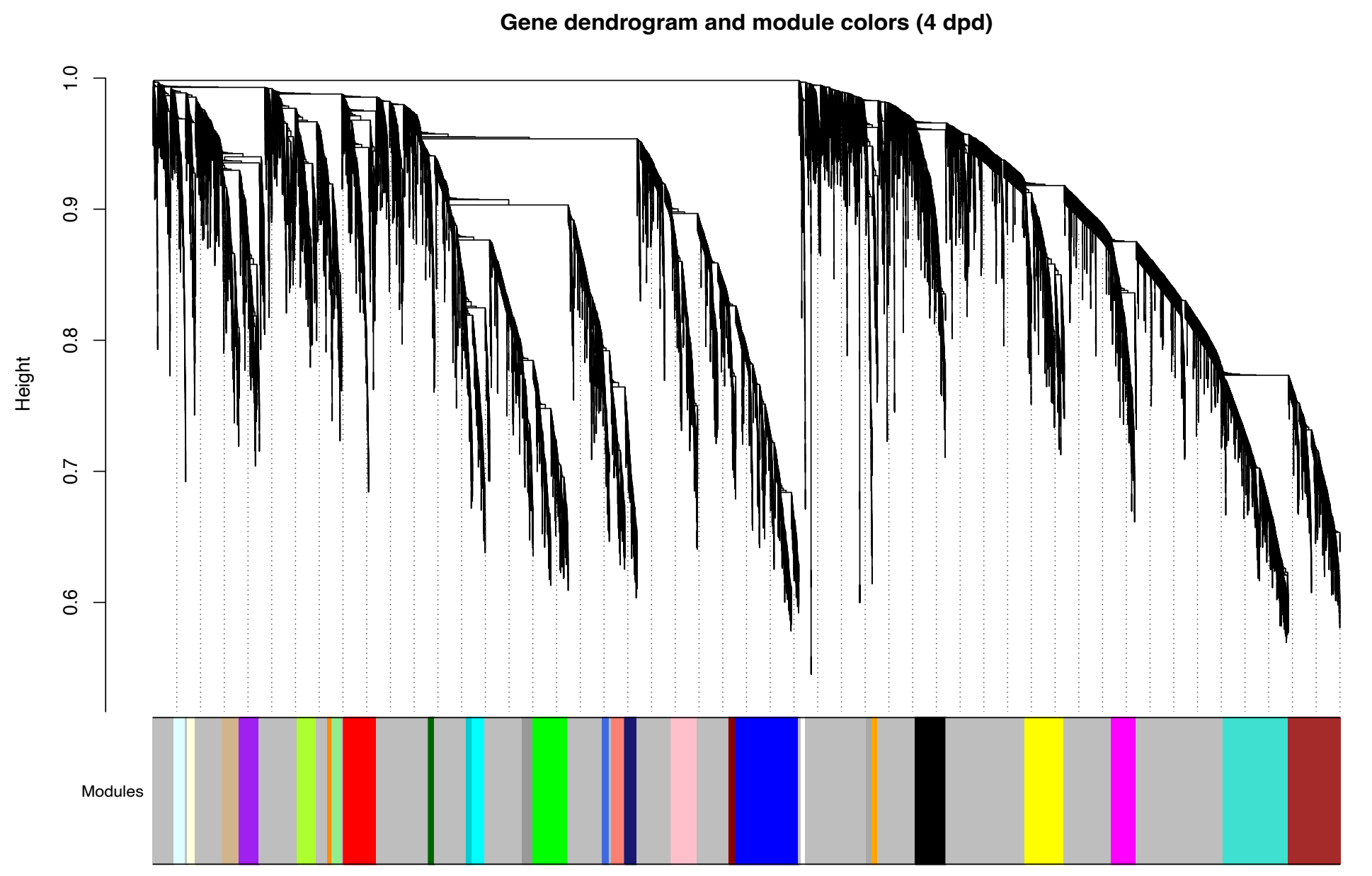


Figure 4: The module membership is determined using dynamic tree cutting. The 4dpd network was used as the control network to determine modules. Shown here the modules are given names based on color. Each color is a module of like genes.

Results: 4 dpd to 12 dpd

Top 10 Preserved Hub Genes between 4dpd and 12 dpd

pink	purple	red	royalblue	salmon	tan
LOC106531234	LOC106533582	LOC106513016	LOC106527387	LOC106523122	LOC106526332
LOC106529154	LOC106517524	LOC106514273	LOC106522982	LOC106511003	LOC106512315
LOC106511341	LOC106517760	LOC106518426	LOC106516498	LOC106514601	LOC106514242
LOC106527412	LOC106518710	LOC106528011	LOC106510909	LOC106515138	LOC106522380
LOC106531789	LOC106523747	LOC106531188	LOC106535257	LOC106519669	LOC106529093
LOC106532408	LOC106528824	LOC106531409	LOC106535828	LOC106522268	LOC106531113
LOC106527175	LOC106529727	LOC106532005	LOC106536019	LOC106535532	LOC106536345
LOC106519900	LOC106534700	LOC106536530	LOC106536756	LOC106536291	LOC106532924
LOC106534587	LOC106517980	LOC106456675	LOC106534202	LOC106537302	LOC106519336
LOC106520678	LOC106516189	LOC106456660	LOC106517575	LOC106537354	LOC106528434

Figure 5: Shown above is the top 10 preserved hub Genes (LOC ID) of 6 of the 28 modules from the comparison of the 4dpd and 12 dpd networks. The hub genes are identified by by comparing the gene membership of the modules in both networks (4dpd and 12 dpd).

purple	gene symbol	description
LOC106533582	egln3	egl-9 family hypoxia-inducible factor 3 [Austrofundulus limnaeus]
LOC106517524	si:dkeyp-72g9.4	uncharacterized
LOC106517760	LOC106517760	sorting nexin-10B-like [Austrofundulus limnaeus]
LOC106518710	LOC106518710	Protein FAM195A-like [Austrofundulus limnaeus]
LOC106523747	LOC106523747	apolipoprotein A-I-like [*Austrofundulus limnaeus*]
LOC106528824	gadd45ab	rowth arrest and DNA-damage-inducible, alpha, b [Austrofundulus limnaeus]
LOC106529727	LOC106529727	diablo homolog, mitochondrial-like [Austrofundulus limnaeus]
LOC106534700	LOC106534700	G0/G1 switch protein 2-like [Austrofundulus limnaeus]
LOC106517980	LOC106517980	rostaglandin I2 synthase [Austrofundulus limnaeus]
LOC106516189	tsc22d3	C22 domain family, member 3 [Austrofundulus limnaeus]

Figure 6: Above is an example of gene ontology for the top 10 preserved hub genes in the purple module between the 4 dpd and 12 dpd networks. Understanding the functions and/ or types of genes in a module can lend insight to what processes that module is a part of.

Discussion

- The information provided is preliminary data and has been used to display the benefits of using WGCNA as a tool for treating and interpreting genomic data sets.
- Genomic sequencing data has become more prevalent in the sciences, and the need for analysis of large data sets is of great importance for the interpretation of this genomic data.
- WGCNA is a powerful tool that allows the user to compare complex data and extrapolate the genetic network controls behind various traits.

Future Directions

- Continue the analysis of the annual killifish anoxia experiment. This includes further enrichment through gene ontology and visualizations of the data.
- Use the module membership data to look for networks not preserved across the different stages of development .
- Find gene targets believed to be vital for the annual killifish's anoxia tolerance to drive future experiments.
- Find abnormal expressions in hypoxic to anoxic environments that may be why *A. limnaeus* can develop normally under extreme environmental conditions other vertebrates cannot survive.
- Compare these data to other vertebrate models of anoxia tolerance.

Acknowledgements

- Funding: NSF Grant IOS-2025832
- The Podrabsky Lab for training and supplying the embryos
- The Center for Life in Extreme Environments