

Gene Expression Analysis of Brook Trout Exposed to Fracking In The Marcellus Shale Region

Will Conrad, Dr. Vincent Buonaccorsi, Dr. Chris Grant

Abstract

The introduction of fracking as a method of oil and natural gas extraction has made it possible for oil companies to remove massive amounts of crude oil and natural gas from areas that would previously have been prohibitively expensive to extract. The risks of environmental damage and bioaccumulation of the heavy metals frequently found in fracking chemical mixtures have led to some concern, and a push to examine the effects of fracking exposure on wildlife. Over the course of this project, we utilized transcriptomics data to analyze the correlation between gene expression and brook trout exposure to various environmental factors associated with stream fracking exposure. The transcriptome data were generated from fish taken from a series of streams in July of 2015, along with several years of water quality data measures taken for those same streams for several years prior to the fish collection. The data were analyzed in R via the Weighted Gene Coexpression Network Analysis method, which works by clustering the transcript reads into a series of modules and comparing the overall expression of the modules against a set of traits or environmental factors. In our preliminary results we found that there was a significant correlation between the gene expression of some modules and exposure to mercury in 2014, and an even stronger set of correlations between some modules and a set of water quality measures associated with fracking in 2015.

Methodology

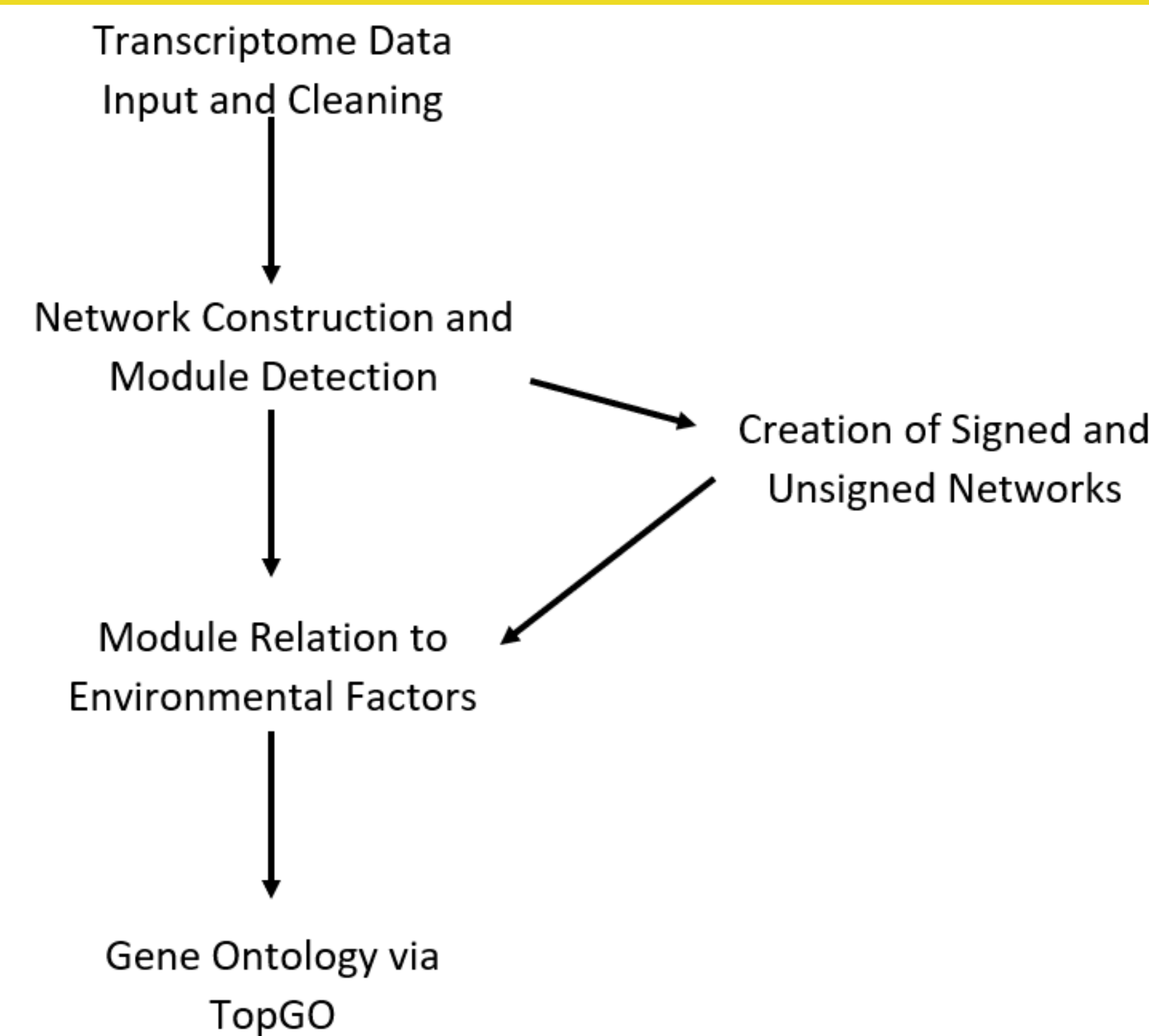


Fig. 1 This figure shows the process that was followed in R for analyzing the data via Weighted Gene Coexpression Network Analysis(WGCNA). The gene ontology was carried out using TopGO, a separate R package that allowed us to compare the transcripts contained in each of the modules generated by the WGCNA process to external information about the function of each of the transcripts.

References

- [1] Grant, C. J., Weimer, A. B., Marks, N. K., Perow, E. S., Oster, J. M., Brubaker, K. M., ... & Lamendella, R. (2015). Marcellus and mercury: assessing potential impacts of unconventional natural gas extraction on aquatic ecosystems in northwestern Pennsylvania. *Journal of Environmental Science and Health, Part A*, 50(5), 482-500.
- [2] Grant, C. J., Lutz, A. K., Kullig, A. D., & Stanton, M. R. (2016). Fracked ecology: Response of aquatic trophic structure and mercury biomagnification dynamics in the Marcellus Shale Formation. *Ecotoxicology*, 25(10), 1739-1750.
- [3] Horvath, S. [DGIT CTRL]. (2013, Aug. 20). *Weighted gene co-expression network analysis* [Video file]. Retrieved from https://www.youtube.com/watch?v=rRIRMMW_RRS4&t=4101s
- [4] Horvath, S. [DGIT CTRL]. (2013, Aug. 21). *WGCNA R Software Tutorial* [Video file]. Retrieved from https://www.youtube.com/watch?v=4h0_iZp6ab0
- [5] Samuels, Myra L., et al. Statistics for the Life Sciences: Global Edition. Pearson, 2016.
- [6] Peterson, Mark. (June, 2017) RNAseq Analysis, GCAT-SEEK Workshop [PDF file], accessed on Juniata College p-drive
- [7] File:Brook trout freshwater fish.jpg. (2017, November 18). *Wikimedia Commons, the free media repository*. Retrieved 19:47, February 3, 2020 from https://commons.wikimedia.org/w/index.php?title=File:Brook_trout_freshwater_fish.jpg&oldid=268090242.
- [8] Dong W, Liu J, Wei L, Jingfeng Y, Chernick M, Hinton DE. 2016. Developmental toxicity from exposure to various forms of mercury compounds in medaka fish (*Oryzias latipes*) embryos. *PeerJ* 4:e2282 <https://doi.org/10.7717/peerj.2282>

Results

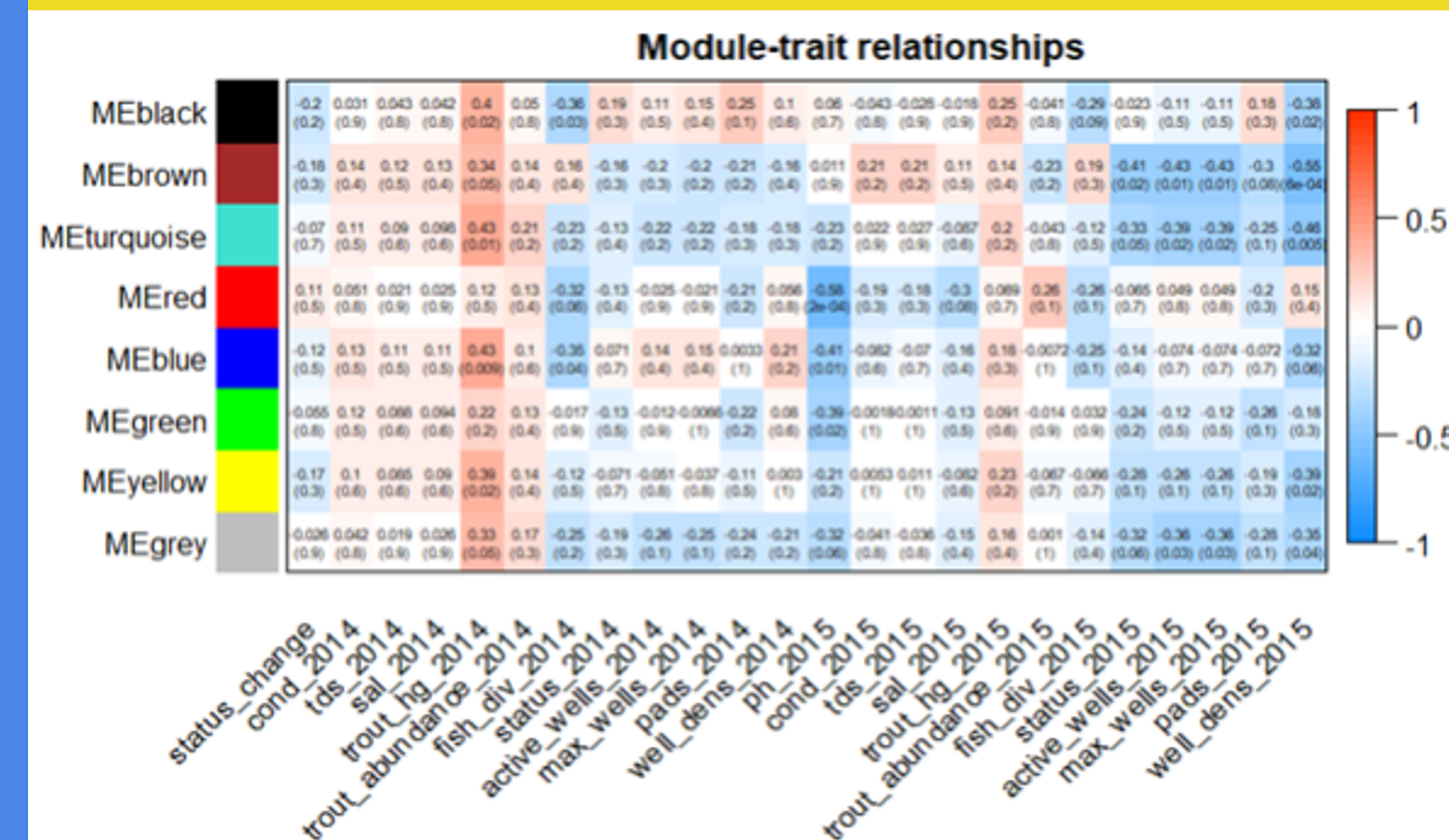


Fig. 2 This figure shows the results of the relation of the signed modules to the environmental characteristics. Significant correlations were found between the red module and water pH in 2015 ($r = -0.58$, $p\text{-value} = 2 \times 10^{-4}$), and the brown, turquoise, and grey modules were found to correlate significantly with the fracking environmental factors after being analyzed for group significance.

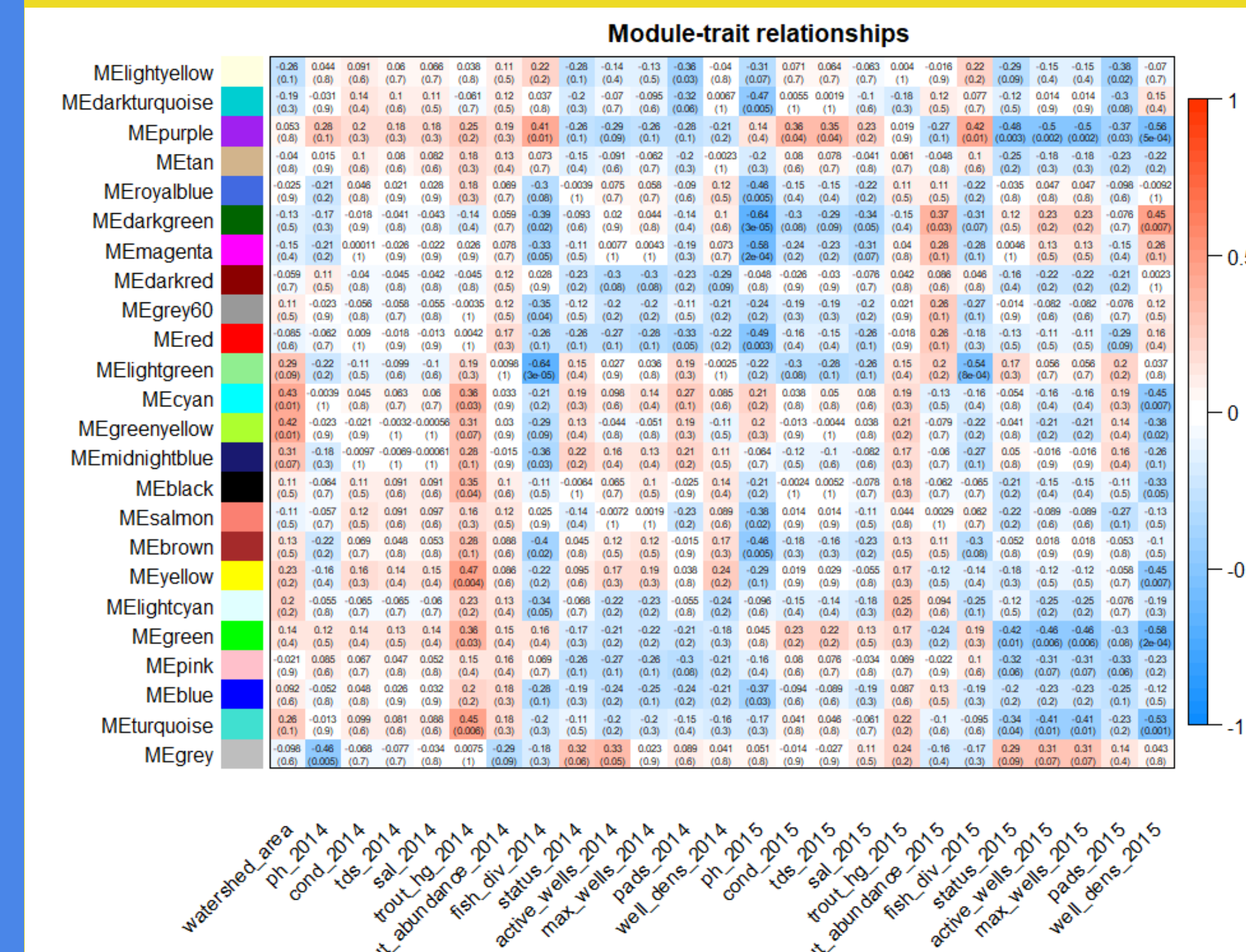


Fig. 3 This figure shows the correlation between the environmental factors, and the unsigned modules. There were no outright significant correlations due to the number of comparisons, but additional Chi-square analysis showed some significant broader trends in mercury levels in 2014, pH in 2015, and in the 2015 fracking factors.

Conclusions and Further Study

It was found that in the modules constructed from both signed and unsigned networks, there was a significant positive correlation between the mercury found in trout in 2014, and gene expression in the trout used in this study. Fracking factors for 2015 all showed negative correlations with gene expression. The pH of the water in 2015 showed a negative correlation with gene expression for some modules. The positive correlation between mercury found in fish taken in 2014 and gene expression should be examined more closely, especially given prior literature findings that showed a connection between the presence of fracking, and the levels of mercury in the water².

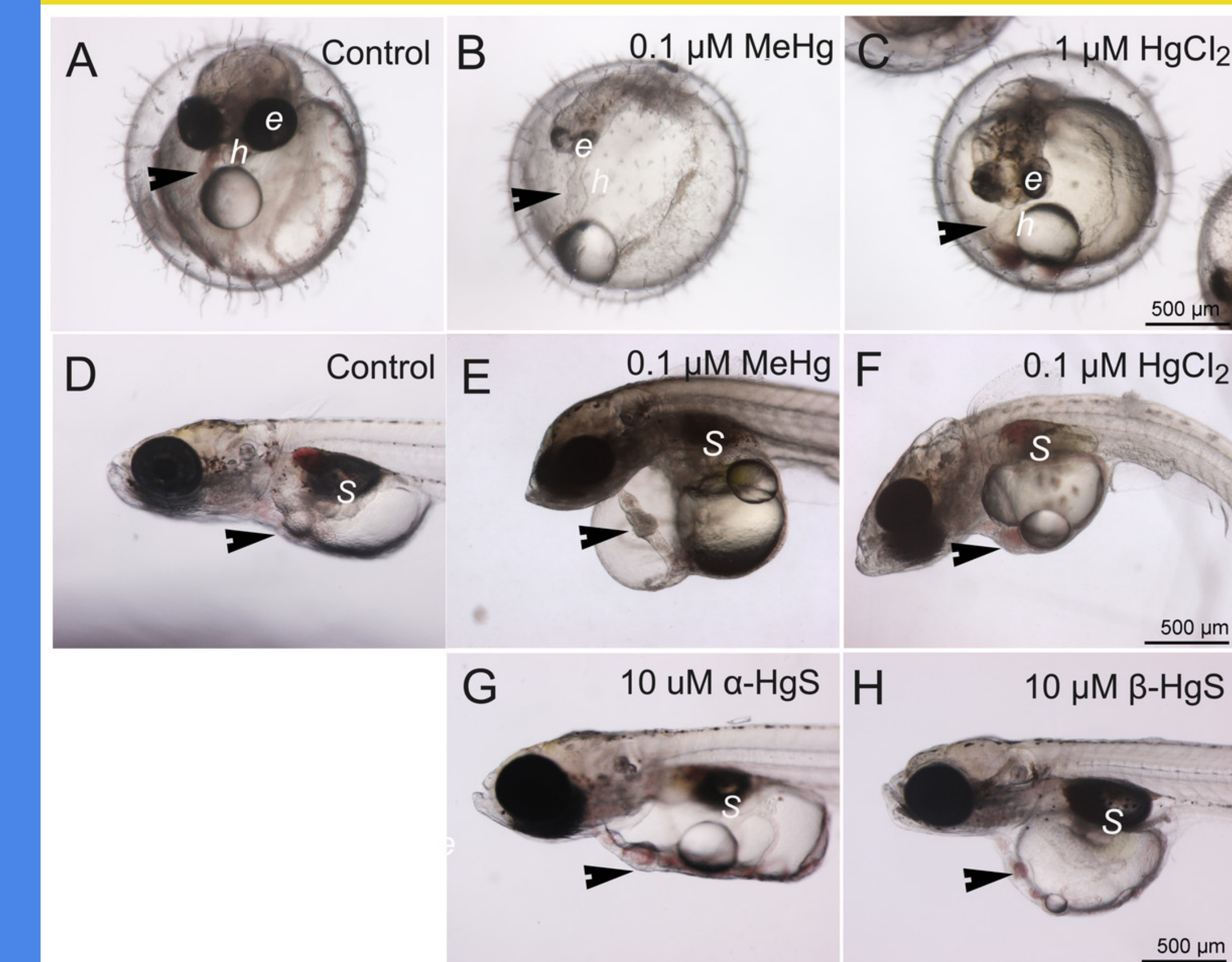


Fig.4 [8] This figure shows the effects of different types of mercury exposure on fish embryos. The mercury used in this study by Dong, et. al. was delivered directly to the fish embryos, and the results they produced indicate that there may be some benefit to further study of the connection between gene expression and mercury exposure.



Introduction

RNA transcriptomics has become an increasingly powerful tool for analyzing the activity of an organism's cells. The transcriptome can give information on which genes are most active⁴, but it has the drawback of containing so much information that it can be difficult to sort through in a timely manner. Weighted Gene Co-Expression Network Analysis (WGCNA) attempts to solve this issue with the use of module eigengenes that stand in for groups of expressed transcripts³.

Module eigengenes are mathematical constructs that WGCNA constructs from groups of transcripts with similar connectivity, which allows them to be clustered together into modules⁴. The module eigengene is then used to represent the entire module, which allows the researcher to compare a handful of the module eigengenes³ to a set of organism traits or environmental factors in place of running thousands of comparisons for each individual gene.

The WGCNA package in R allows a researcher to run these analyses with a high degree of customizability⁴, permitting the analysis to be run on vastly different datasets, and with very different sets of factors for the organism traits or environmental factors. In this instance, the WGCNA has been used to analyze the relationship between a set of environmental factors measured in a series of Pennsylvania streams, and the transcriptomics information for the fish removed from those streams, with special attention paid to environmental factors relating to fracking in the affected areas.

The presence of fracking byproducts in streams has been shown to affect brook trout living in those streams², and this report provides an initial analysis of the genes that have seen a change in expression level in correlation with the presence of certain environmental characteristics, and the presence of some types of fracking infrastructure.