EPOCH EPIGENETICS STUDY PROPOSAL:

Weighted Gene Co-Expression Network Analysis (WGCNA) of adolescents in the EPOCH Study cohort

INVESTIGATOR

Jessica R. Shaw

PROJECT SUMMARY

DNA samples obtained from adolescent participants in the EPOCH Study will be analyzed to identify gene co-expression networks that differ by in utero exposure status to maternal gestational diabetes (GDM). A method of Weighted Gene Co-Expression Network Analysis (WGCNA) will be implemented using the WGCNA R package [1]. The proposed analysis was designed to support the specific aims of the funded R01, "Epigenetic marks of in utero exposure to diabetes" (5R01DK100340-03), which seeks to characterize the epigenetic pathways through which exposure to maternal GDM increases offspring risk for obesity, Type 2 Diabetes, and other adiposity-related conditions.

BACKGROUND

Genome-wide methylation analysis was conducted on samples obtained from a subset of participants at the EPOCH T1 study visit. Differentially methylated regions (DMRs) were identified in association with in utero exposure to maternal GDM. Significant DMRs identified in genome-wide analysis were prioritized for further study based on a review of relevant literature and the presence of related DMRs in the Healthy Start Study cohort. Nine CpG sites within six genes were selected for validation by pyrosequencing. Paired samples from T1 and T2 were pyrosequenced simultaneously to minimize the potential for batch effects. Pyrosequencing produced quantitative estimates (β estimates) of methylation at each CpG site.

Genes residing in close proximity to previously identified DMRs were selected for gene expression profiling. Expression profiling was conducted using samples obtained at the EPOCH T2 study visit from the same subset of participants used in methylation analyses. EPOCH T2 expression profiles have the potential to provide insight into the downstream effects of DNA methylation on gene expression.

Gene ID	No. Prioritized CpGs
PTPRN2	1
SH3PXD2A	1
RNF39	1
ST5	2
DAPL1	3
NPHP4	1

SIGNIFICANCE

The application of WGCNA to expression profiles from the EPOCH Study will facilitate a better understanding of how in utero exposure to maternal GDM impacts long-term gene expression in children. Despite being one of the best performing methods for gene co-expression network construction [2], applications of WGCNA to GDM and associated outcomes in exposed children have not yet been published. Application of WGCNA to EPOCH expression data constitutes a significant scientific opportunity to better understand the biological pathways impacted by GDM exposure.

METHODOLOGY & STATISTICAL CONSIDERATIONS

The WGCNA R package will be used to identify network modules (i.e., clusters) of co-expressed genes present in DNA samples of exposed versus unexposed EPOCH participants at the T2 study visit. WGCNA employs a method of data reduction that takes into account the correlation between gene expression levels and allows for each inter-gene connection to be assigned a weight [3, 4, 11]. WGCNA allows for the manipulation of a number of parameters through which network construction can be informed by biology [4]. WGCNA also improves statistical power by considering clusters of genes rather than individual loci, thereby reducing the multiple testing burden of gene expression analysis [5].

RESEARCH STRATEGY

WGCNA requires the specification of multiple parameters to direct module identification. The choice of these parameters can greatly impact the network modules identified [4]. A thorough literature review of published WGCNA analyses will be conducted to help inform the selection of appropriate parameters. Published analyses on phenotypes related to GDM will be prioritized. Prioritized publications will include:

Medina IR, Lubovac-Pilav Z. Gene Co-Expression Network Analysis for Identifying Modules and Functionally Enriched Pathways in Type 1 Diabetes. *PLOS ONE*. 2016;11(6):e0156006. doi:10.1371/journal.pone.0156006. [6]

Tang W, Gao Y, Li Y, Shi G. Gene networks implicated in diabetic kidney disease. *Eur Rev Med Pharmacol Sci*. 2012;16(14):1967-1973. [7]

Kogelman et al., 2014 "Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. *BMC Medical Genomics*. 2014;7:57. doi:10.1186/1755-8794-7-57. [8]

SPECIFIC AIMS

The proposed analysis will achieve the following three specific aims.

1. Construct a network of gene co-expression present in EPOCH T2 DNA samples.
2. Implement a pathway analysis to identify modules, or clusters of co-expressed genes.
3. Incorporate participant exposure status and clinically relevant covariates to facilitate the identification of biologically plausible modules associated with exposure to GDM.

FUTURE DIRECTIONS

Following the preliminary application of WGCNA to EPOCH expression profiles, future extended analyses may include:

- An analysis of co-expression networks, stratified by genetic risk scores for obesity, anthropomorphic measurements, metabolic measurements, or nutrient levels [7].
- Pathway enrichment analysis of network modules identified by WGCNA [8, 9].
- Use of gene co-expression network modules as predictors, covariates, or outcomes in further analyses.
- Comparison of gene co-expression networks in the EPOCH and Healthy Start Study cohorts [3].

References

1. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. doi:10.1186/1471-2105-9-559.
2. Allen JD, Xie Y, Chen M, Girard L, Xiao G. Comparing Statistical Methods for Constructing Large Scale Gene Networks. *PLOS ONE*. 2012;7(1):e29348. doi:10.1371/journal.pone.0029348.
3. Workshop on weighted gene co-expression network analysis. https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/WORKSHOP/. Accessed November 11, 2016.
4. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*. 2005;4(1). doi:10.2202/1544-6115.1128.
5. Lin X, Barton S, Holbrook JD. How to make DNA methylome wide association studies more powerful. *Epigenomics*. 2016;8(8):1117-1129. doi:10.2217/epi-2016-0017.
6. Medina IR, Lubovac-Pilav Z. Gene Co-Expression Network Analysis for Identifying Modules and Functionally Enriched Pathways in Type 1 Diabetes. *PLOS ONE*. 2016;11(6):e0156006. doi:10.1371/journal.pone.0156006.
7. Tang W, Gao Y, Li Y, Shi G. Gene networks implicated in diabetic kidney disease. *Eur Rev Med Pharmacol Sci*. 2012;16(14):1967-1973.
8. Kogelman et al., 2014 "Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. *BMC Medical Genomics*. 2014;7:57. doi:10.1186/1755-8794-7-57.
9. EnrichNet - Network-based gene and protein set enrichment analysis. http://www.enrichnet.org/. Accessed November 11, 2016.
10. ConsensusPathDB. http://cpdb.molgen.mpg.de/. Accessed November 11, 2016.
11. Tutorials for WGCNA R package. https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/. Accessed November 13, 2016.