Name: Megan Hagenauer, Ph.D.

PI Name: Huda Akil, Ph.D.

Email: hagenaue@umich.edu

PI email: akil@med.umich.edu

The Brain Data Alchemy Project: Teaching Research Reproducibility and Discovery Science While Mining Gold from Archived Genomics Data

Megan Hagenauer, Ph.D.[1]; Cosette Rhoads[2]; Jinglin Xiong[2]; Erin Hernandez[2];
Duy Nguyen[2]; Annaka Saffron[2]; Elizabeth Flandreau, Ph.D.[3];
Stanley Watson Jr., M.D., Ph.D.[1]; Huda Akil, Ph.D.[1]

[1]Michigan Neuroscience Institute, University of Michigan, Ann Arbor, MI USA; [2]Grinnell College, Grinnell, IA USA; [3]Psychology Department, Grand Valley State University, Allendale, MI USA

During the past decade, the landscape of neuroscience research has undergone two major transformations in the way that data are collected, analyzed, and interpreted. First, there has been an intensive push to reform scientific practices to improve research reproducibility. Second, accelerated growth in computing power and omics knowledge has led to a blossoming of "discovery science". In this new landscape, trainees need to acquire skills that are not included in traditional curriculum.

We have addressed this need by creating an intensive summer program that provides direct, hands- on experience with experimental design and statistical issues related to research reproducibility and discovery science. Within the program, trainees conduct a systematic meta-analysis focused on a chosen neuroscience topic using the burgeoning trove (>10,000) of publicly available transcriptional profiling datasets (microarray, RNA-Seq).

We successfully piloted the program in 2022 (n=6 trainees). We found that over the course of a single summer (10 weeks), trainees were able to learn to code in R, survey literature, and run a full genomics meta-analysis that could serve as a small publication or preliminary data for a grant. The topics chosen by the trainees included chronic stress, sleep deprivation, antidepressant usage, viral and bacterial inflammation. Next year, we plan to focus on substance use. Each of the meta-analyses revealed an extensive set of differentially expressed genes that can shed light on neuropsychiatric disorders. These gene sets will be released within reference databases compiled by the PI (Brain.gmt) and broader curation efforts (Geneweaver, MSigDB, Enrichr, PhenoCarta) to advance genomic science.