Submitter Name: Timothy Reynolds
Submitted email: timothy.reynolds@jax.org
PI Name (if different): Erich Baker, Elissa Chesler
PI email (if different): erich_baker@baylor.edu, elissa.chesler@jax.org

Integrative cross-species analysis uncovers convergent genomic regulatory features associated with complex disease

Timothy Reynolds[1,2], Jason Bubier[1], Arpana Agrawal[3], Erich Baker[2], Elissa Chesler[1]

[1]The Jackson Laboratory; [2]Institute of Biomedical Studies, Baylor University; [3]Department of Psychiatry, Washington University School of Medicine

Complex diseases, including substance use disorders (SUD) and alcohol use disorder (AUD), contribute substantially to the global disease burden, but their mechanisms remain poorly understood. High-throughput functional genomics and genome-wide association studies (GWAS) seek to identify molecular and genetic mechanisms underlying these diseases, leading to the discovery of protein-coding and non-coding variants associated with disease risk. Many of these non-coding variants reside in regions known to exhibit regulatory influences on gene expression. Since these regulatory variants are poorly conserved, they are difficult to characterize, validate, and model in preclinical studies. Although identification of trait regulatory variants is cost-effective in model organisms, the translational relevance of these findings is sometimes uncertain. In the present study, we investigate how non-coding variants associated with AUD risk can be prioritized in GeneWeaver using multi-species epigenomic features, gene networks, and functional genomics experiments. By integrating robust community resources such as NCBI's GEO, the Genotype-Tissue Expression (GTEx) project, and the Encylopedia of DNA Elements (ENCODE) to explore functional genomic and gene regulation, and integrative resources such as Lynx and The Monarch Initiative, we can leverage the individual strength of each resource. Combined with GeneWeaver's integrative data stores and combinatorial toolset, we identify conserved gene regulatory mechanisms across species and pinpoint their effects using expression-based studies. We show that non-coding variants and epigenomic regulators in model organisms can recapitulate human disease non-coding variant effects. Altogether, these results indicate that model organism studies can be used to aid the interpretation of non- coding variants identified in human GWAS.