Name: Andreas Pfenning Email: apfenning@cmu.edu

Fine-mapping candidate neuropsychiatric regulatory variants using cell type-aware comparative genomics

BaDoi N. Phan[1], Alyssa J Lawler[1], Jing He[2], Ashley R. Brown[1], Irene M. Kaplow[1], Amanda Kowalczyk[1], Chaitanya Srinivasan[1], Elinor Karlsson[3,4], Kerstin Lindblad-Toh[5,6], William R. Stauffer[2], Andreas R. Pfenning[1]

[1]Computational Biology Department, School of Computer Science, Carnegie Mellon University; Pittsburgh, PA, USA; [2]Neurobiology Department, School of Medicine, University of Pittsburgh; Pittsburgh, PA, USA; [3]Program in Bioinformatics and Integrative Biology, University of Massachusetts Chan Medical School, Worcester, MA, United States; [4]Broad Institute, Cambridge, MA, United States; [5]Medical Biochemistry and Molecular Biology, Uppsala University, Uppsala, Sweden

Measure of nucleotide sequence conservation across species are useful for identifying functional genomic loci, but can fail when regulatory function is maintained, often in a cell type-specific manner, even when sequence is not. We introduce Cell-TACIT, the Cell Type-Aware Conservation Inference Toolkit, to identify trait-associated regulatory variants. Cell-TACIT integrates sequence conservation scores with cell type-specific open chromatin data collected from a few mammalian species to impute function for hundreds more. Applying Cell-TACIT to neuropsychiatric trait loci identifies higher heritability enrichments and more fine-mapped variants than nucleotide conservation or human chromatin data alone. Our in vivo reporter assays validate predictions for enhancers with risk variants near the DRD2 tobacco use risk locus. By integrating genome conservation and multi-species open chromatin data, Cell-TACIT prioritizes variants within regions of conserved regulatory function for in vivo characterization and addresses a major challenge in translating disease associations to mechanistic understanding.