Identifying pathogenic gene targets of noncoding GWAS risk loci using gene regulatory circuits

Parker A. Hall[1], Anna M. Barbeau[1], Sagar Nisraiyya[1], Olivia Corradin[1]

1) Whitehead Institute, Cambridge, MA, USA

Identifying the biological consequence of noncoding genetic variants is a substantial challenge which has impeded the translation of GWAS results to clinically actionable findings. The first step in overcoming this challenge is to identify the gene target that mediates the variant's effect on disease risk and determine cell type or context of this effect. Many of these variants lie within transcriptional enhancers and are hypothesized to contribute to disease pathogenesis by disrupting gene regulation. As enhancers frequently regulate distal genes, the current approach is to utilize chromatin interaction studies, such as Hi-C, to identify the gene target of noncoding DNA variants. These results however, often identify physical interactions with many gene targets, which may be active in multiple cell types with potential relevance to disease. It is likely that only a fraction of these enhancer-SNP gene interactions are relevant to understanding disease pathogenesis.

Genes are frequently regulated by multiple enhancers via physical interaction with the promoter. *We have previously demonstrated that SNPs, termed 'outside-variants,' that physically interact with the same target gene as GWAS SNPs can cooperate to alter gene expression and disease risk.* Here, we describe a methodology which utilizes outside-variants to predict the gene and cell type through which risk variants contribute to disease. To evaluate this approach, we first applied it to multiple sclerosis. The majority of MS risk loci were linked to >4 genes via Hi-C. We identified significant outside-variants for 77% of MS loci. We assessed each gene's regulatory circuit i.e. all of the enhancers with which it physically interacts, for enrichment of outside-variants. This approach resulted in a 3-fold reduction in the number of predicted disease-relevant gene targets. These results revealed novel predictions such dysregulation of transcriptional pausing in the brain as contributing to MS pathology and suggest broad applicability to additional complex traits.