Network-based Functional Prediction Augments Genetic Association to Predict Candidate Genes for Inflammatory Bowel Disorder in Mice

Montana K. Lara[1], Karolyn G. Lahue[2], Alisha A. Linton[1], Brigitte Lavoie[1], Qian Fang[3], Mahalia M. McGill[2], Jessica W. Crothers[3], Cory Teuscher[3], Gary M. Mawe[1], Anna L. Tyler[5], Dimitry N. Krementsov[1,2], and J. Matthew Mahoney[1,4]

Departments of Neurological Sciences[1], Biomedical and Health Sciences[2], Medicine[3], and Computer Science[4], University of Vermont, Burlington, VT, The Jackson Laboratory[5], Bar Harbor, ME

The University of Vermont LARNER COLLEGE OF MEDICINE

NEUROSCIENCE GRADUATE PROGRAM The University of Vermont

Introduction

Human genome wide association studies (GWAS) have identified hundreds of candidate genes for **inflammatory bowel disorder (IBD)** while mouse studies have also identified multiple quantitative trait loci (QTL). Here, we use a novel **network-based machine learning** approach to better predict candidate genes controlling IBD susceptibility under the hypothesis that human risk factors and their mouse orthologs are functionally related genes acting in the same biological process. Using **support vector machine** (SVM) classifiers with **gene expression** signatures in mouse immune cells, we prioritized functionally related genes associated with human IBD susceptibility. We trained SVM classifiers to identify subnetworks enriched with IBD GWAS genes in two tissue-specific functional networks in mice: the **intestinal** and **hemolymphoid networks**. By integrating positional, functional, and expression-based information with network-based machine learning, we can better predict candidate genes with strong functional evidence for association to IBD.

Methods

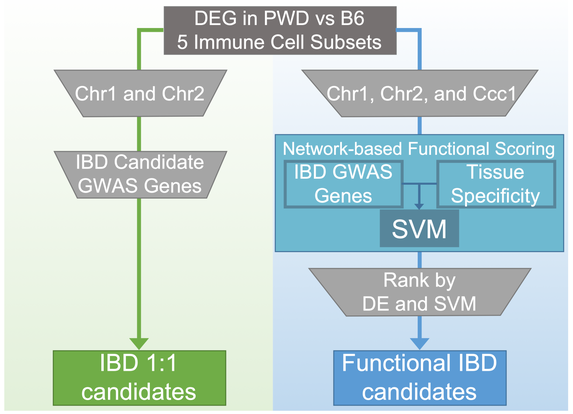

Fig 1. *Workflows* used to identify and prioritize candidate genes. Chromosomes (Chrs) 1, 2, and 12 have been identified as novel loci that profoundly enhance IBD susceptibility. IBD 1:1 candidates were identified by the overlap of highly differentially expressed genes in C57BL/6J (B6) vs. PWD/PhJ (PWD) IBD mouse model immune cells and human IBD GWAS candidate genes. Scores from tissue-specific "Intestine" and "Hemolymphoid" network SVMs that were trained with IBD GWAS genes were integrated with differential expression to identify functional IBD candidates.

Acknowledgements

The Epilepsy, Cognition, and Development group for their guidance and input. The Neuroscience Graduate Program for providing opportunity and support.

Results

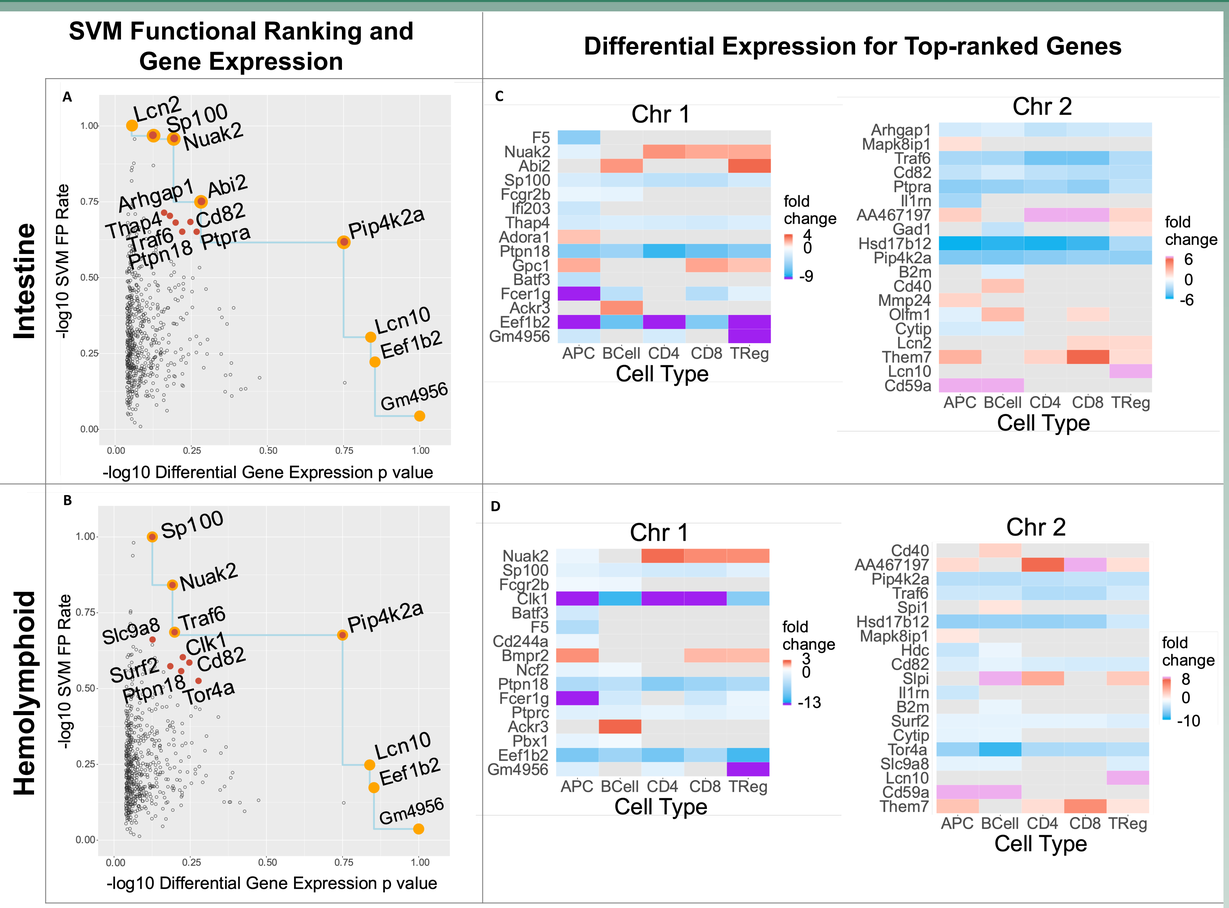

Fig 2. Identification and prioritization of novel gene candidates associated with IBD susceptibility. (**A,B**) Representative plots of SVM scores and differential expression for both networks. SVMs using the intestine and hemolymphoid networks, respectively, were integrated with the differentially expressed gene dataset from B6 and PWD regulatory T cells (Tregs). Significantly differentially expressed genes from chrs 1 and 2 are plotted by normalized -log_{10}(p-value) on the x-axis and normalized SVM -log_{10}(SVM false positive rate) $(FPR)_{SVM}$ on the y-axis. FPR_{SVM} and p-values were normalized by their maximum value. Orange and red indicates top ranked genes that also fall on the Pareto front maximizing high ranking SVM scores and high differential expression. FPR_{SVM} vs. differential expression plots are available for all five immune cell subtypes from the intestine and hemolymphoid network SVMs. (**C,D**) Heatmaps of differential gene expression between PWD and B6 immune cell types ($log2$(fold change PWD/B6)) for the top candidate genes identified by the SVM approach. Purple and pink denote extreme differential expression values, and gray denotes genes whose differential expression was not significant for that cell type (FDR <0.05).

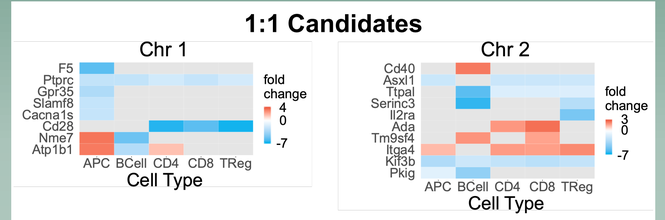

Fig 3. *IBD 1:1 Candidates* Heatmaps of differential gene expression between PWD and B6 immune cell types ($log2$(fold change PWD/B6)) for the top candidates identified by the 1:1 IBD candidate approach. To identify candidates relevant to human IBD from the expression data, genes were filtered using overlapping IBD GWAS candidate loci.

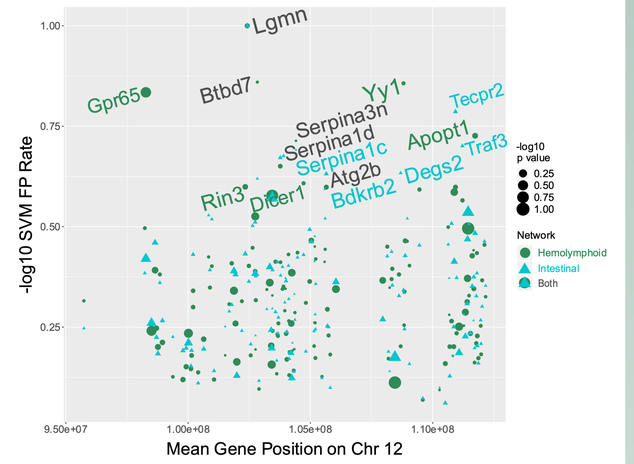

Fig 4. *Ccc1 locus on Chr 12* Given the candidacy of the *Ccc1* locus identified in mice, genes from that interval were ranked using the network-based functional scoring approach. Genes are plotted by genomic location on the x-axis and normalized -log_{10}(FPR_{SVM}) on the y-axis. Each gene has a score from both network SVMs and the size of each point indicates normalized significance of differential expression. Highly ranked genes are located at the top, with *Lgmn* the highest ranked in both networks.

Conclusions

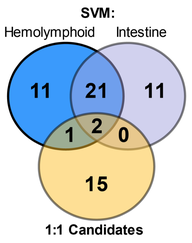

Fig 5. between genes identified using IBD 1:1 candidate approach and the intestine and hemolymphoid network SVMs.

With multiple IBD susceptibility genes identified by human GWAS and QTL mapping studies in mice, gene prioritization is increasingly important. Here, we have used a machine learning approach to augment physical mapping data, providing a manageable list of plausible candidate genes to be pursued. By integrating human IBD GWAS genes, differential expression of immune cells in mice, and tissue-specific functional networks, we ranked candidate genes in relevant genomic locations by their functional association to known to IBD GWAS genes in the mouse intestine and hemolymphoid networks. We identified 46 genes of interest as both highly functionally related to the IBD phenotype and highly differentially expressed in PWD vs. B6 immune cell types. Ideal candidates include *Itga4, Pip4k2a, Traf6, Lgmn,* and *Gpr65*. These genes are known to be involved in a variety of processes, specifically inflammation and immune signaling, including modulating response against pathogen-driven intestinal inflammation and involvement in innate immune signaling *(Traf6)*, regulating lymphocyte entry into inflamed gut and CNS tissue *(Itga4)*, as well as T cell receptor signaling *(Ptpra)* and enzymatic activity involving cellular membrane lipids (*Pip4k2a*). *Lgmn* encodes the cysteine protease legumain, which is highly expressed in dendritic cells and plays a role in antigen processing and presentation. Our results demonstrate the predictive potential of network-based machine learning for candidate gene ranking across species.

Literature Cited

Bearoff, F., del Rio, R., Case, L. *et al.* Natural genetic variation profoundly regulates gene expression in immune cells and dictates susceptibility to CNS autoimmunity. *Genes Immun* **17,** 386–395 (2016).

de Lange, K., Moutsianas, L., Lee, J. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49,** 256–261 (2017).

Tyler AL, Raza A, Krementsov DN, et al. Network-Based Functional Prediction Augments Genetic Association To Predict Candidate Genes for Histamine Hypersensitivity in Mice. *G3 (Bethesda).* 2019;9(12):4223–4233. Published 2019 Dec 3. doi:10.1534/g3.119.400740

Goya, J., et al., *FNTM: a server for predicting functional networks of tissues in mouse.* Nucleic Acids Res, 2015. **43**(W1): p. W182-7.

Rogala, A.R., et al., The Collaborative Cross as a resource for modeling human disease: CC011/Unc, a new mouse model for spontaneous colitis. Mamm Genome, 2014. 25(3-4): p. 95-108.