Submitter Name: Shuo Chen
Submitted email: shuochen@som.umaryland.edu

Genetic Fine-mapping with Dense Linkage Disequilibrium Blocks: genetics of nicotine dependence

Chen Mo[1*], Zhenyao Ye[1*], Kathryn Hatch[1], Yuan Zhang[5], Qing Lu[6], Song Liu[4], Braxton Mitchell[1], Elliot Hong[1], Peter Kochunov[1], Tianzhou Ma[3‡], Shuo Chen[1,2‡]

[1] Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland, Baltimore, Maryland, United States of America
[2] Division of Biostatistics and Bioinformatics, Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, Maryland, United States of America
[3] Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, Maryland, United States of America
[4] Department of Computer Science and Technology, Qilu University of Technology, Jinan, Shandong, China
[5] Department of Statistics, Ohio State University, Columbus, Ohio, United States of America
[6] Department of Biostatistics, University of Florida, Gainesville, Florida, United States of America

[*]These authors contributed equally to this work
‡ Corresponding authors: tma0929@umd.edu (TM); shuochen@som.umaryland.edu (SC)

Fine-mapping is an analytical step for causal prioritization of the polymorphic variants in a trait-associated genomic region observed in genome-wide association studies (GWAS). Prioritization of causal variants can be challenging due to linkage disequilibrium (LD) patterns among hundreds to thousands of polymorphisms associated with a trait. Hence, we propose to disentangle LD patterns by a dense LD block structure. We developed an l_0 graph norm shrinkage algorithm to identify LD blocks consisting of highly correlated single nucleotide polymorphisms (SNPs). Then, we performed dense LD block guided regression shrinkage for selecting a parsimonious set of causal variants. Our approach is computationally efficient and allows fine-mapping of thousands of polymorphisms. We demonstrated its application using a large UK Biobank (UKBB) sample related to nicotine addiction. Our results suggested that polymorphic variances in both neighboring and distant variants can be consolidated into dense blocks of highly correlated loci. Using simulations to evaluate and compare the performance of this new algorithm with existing fine-mapping algorithms, we demonstrated that this method outperformed comparable fine-mapping methods with increased sensitivity and reduced false-positive error rate for causal variant selection. The application of this method to the smoking severity trait in the UKBB sample replicated previously reported loci and suggested the causal prioritization of genetic effects on nicotine dependency.