*Nature Reviews Genetics* | AOP, published online
11 November 2014; doi:10.1038/nrg3862


DISEASE GENOMICS

# Triaging risk variants in the non-coding genome

Over the past decade, genome-wide association studies (GWASs) have linked hundreds of loci to diseases, but pinpointing the actual disease-causing variants has been less straightforward. A new algorithm reported in *Nature* promises to aid the identification of causal risk variants in non-coding sequences.

Recent GWASs of autoimmune diseases have suggested an enrichment of risk variants within regulatory elements such as enhancers. "Variants in non-coding sequences pose a particular challenge for establishing pathogenicity," explains Alexander Marson (University of California, San Francisco), one of two lead authors of the study, "as GWASs tend to discern large clusters of single-nucleotide polymorphisms (SNPs) in linkage disequilibrium, making it difficult to distinguish causal SNPs from neutral variants in linkage."

The researchers used high-density Immunochip data from 14,277 individuals with multiple sclerosis and 23,605 healthy controls, which comprises all 1000 Genomes Project common SNPs within 186 loci previously associated with autoimmunity. Based on these data, Broad Institute colleagues Kyle K. Farh and Mark J. Daly developed an algorithm called Probabilistic Identification of Causal SNPs (PICS), which estimates the probability of an individual SNP being a causal variant on the basis of haplotype structure and observed pattern of association at a given locus.

The team applied PICS to 21 autoimmune diseases, using Immunochip data where available or imputation to the 1000 Genomes Project otherwise. Not only was PICS able to identify candidate causal variants based purely on genetic evidence but, importantly, imputation to the 1000 Genomes Project also enabled the prediction of causal variants in other diseases, such as Alzheimer's disease and migraine, based on publicly available GWAS data.

To characterize newly identified potential causal variants, Bradley E. Bernstein, a co-senior author at the Broad Institute, and colleagues generated epigenomic maps of active regulatory elements across several well-defined, primary immune cell types using chromatin immunoprecipitation followed by sequencing (ChIP–seq) and RNA sequencing (RNA-seq). Mapping of the candidate autoimmune disease SNPs determined by PICS to these resources revealed enrichment in immune-cell enhancers.

> "Mapping of the candidate autoimmune disease SNPs determined by PICS … revealed enrichment in immune-cell enhancers"

Further integration of the PICS-SNPs with transcription factor binding maps showed that, despite their close proximity to binding sites of immunity-related transcription factors, the variants most often do not directly disrupt or create recognizable transcription factor binding motifs. "Instead, the genomic sites of disease SNPs strongly suggest the importance of non-canonical sequences with as yet undefined crucial roles in immune-cell gene regulation," says David A. Hafler, a co-senior author from Yale University.

Taken together, the PICS algorithm and epigenomic maps provide a novel interpretative framework to identify and interpret functional effects of non-coding disease variants. Although the study highlights transcription factors, target loci and pathways with disease-specific or general roles in autoimmunity, PICS should also prove useful to the wider research community and help to focus their efforts on likely causal SNPs for other diseases.

*Linda Koch*

**ORIGINAL RESEARCH PAPER** Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* http://dx.doi.org/10.1038/nature13835 (2014)