



“
LINSIGHT
infers the
selective
pressure on
non-coding
sites”

Genetic variants that are associated with disease often lie outside of protein-coding genes, and accurately characterizing variation within non-coding regions of the human genome remains challenging. Now, Huang *et al.* present a new computational method, Linear INSIGHT (LINSIGHT), which can predict non-coding disease-causing variants.

The software applies evolutionary methods (using signatures of natural selection over time) to functional genomic and genetic variation data and estimates the chance that non-coding mutations at specific genomic sites will have fitness effects. Specifically, LINSIGHT infers the selective pressure on non-coding sites, and as a result the potential fitness consequences of non-coding mutations, on the basis of genetic variation patterns within a species and patterns of divergence from closely related species. Patterns of genetic variation at a site are contrasted with the patterns at nearby genomic regions that are thought to be free from the influence of selection (‘neutrally evolving regions’). The likelihood of fitness consequences for mutations at each site is assumed to

depend on its specific genomic features, including conservation scores, predicted binding sites and regional annotations. LINSIGHT thus builds on previous approaches by combining the probabilistic graphical model generated by INSIGHT with a generalized linear model, resulting in a faster, more scalable approach with greater genomic resolution and predictive power.

To validate the approach, a data set consisting of human and vertebrate genomic sequences and functional genomic data were analysed. The authors chose 48 genomic features and generated LINSIGHT scores for positions in a human reference genome containing these features. Higher scores correspond to greater evolutionary constraint. Owing to its scalability, LINSIGHT allows for data to be pooled across cell types, enabling scores to reach single-nucleotide resolution. Importantly, the level of evolutionary constraint suggested by LINSIGHT scores generated for non-coding regions of the reference genome were generally consistent with previous observations. In addition, LINSIGHT was tested, alongside several other

methods, for its ability to predict known non-coding genetic variants that are associated with inherited diseases. LINSIGHT outperformed all other methods in all comparisons, confirming its utility for characterizing variation in non-coding genomic regions.

Finally, LINSIGHT was applied to almost 30,000 enhancers that are active in a variety of cell and tissue types. Enhancers that were active across different cell types were found to have a greater degree of evolutionary constraint than those active in fewer cell types. Interestingly, tissue-specific enhancers, from cell types that are associated with rapid evolution (for example, cells from the immune system, the male reproduction tract and olfactory regions), showed lower constraint. Furthermore, the degree of constraint at enhancers correlated with the corresponding promoter, suggesting that the same evolutionary pressures may act at enhancer–promoter pairs.

Taken together, these data suggest that LINSIGHT can accurately predict disease-associated genetic variants that fall outside of protein-coding genes and that it may also be used to characterize regulatory function. The software is publicly available online.

Shimona Starling

ORIGINAL ARTICLE Huang, Y.-F., *et al.* Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. *Nat. Genet.* <http://dx.doi.org/10.1038/ng.3810> (2017)

WEB SITE

LINSIGHT software: <https://github.com/CshSiepelLab/LINSIGHT>