**In-silico prediction of the 3D genome of body height-associated haplotypes**

**Wanjun Gu, Erin Gilbertson, Rany Salem, Tony Capra**

**Background**

The majority of disease associated variants identified via GWAS are non-protein-coding. The mechanism by which non-coding variants affect disease is unknown but are thought to play a regulatory function. Consideration of three-dimensional genome architecture may reveal the mechanism by which non-coding variants influence gene expression of target genes. For example, phenotypes like body height, SNP heritability shows significant clustering within topologically associated domains (TADs), and these domains provide a framework for understanding the genetic architecture of body height through extensive GWAS summary statistics.

**Methods**

In our study, we leverage a recently computational method that predicts via using whole genome-sequencing (WGS) data We applied this data to GWAS summary statistics for human adult height, focusing on suggested genome-wide significant regions (top p-value ≤ 5e-5). Utilizing the NHLBI Trans-Omics for Precision Medicine (TopMed) WGS data for ~50,000 subjects, we imputed haplotypes for these significant signals across diverse populations, including Europeans, Africans, East Asians, South Asians, and Admixed/non-admixed Americans. Haplotypes were then edited into the 200-kb reference sequences and fed into the machine-learning model to predict alterations in the 3D genome structure. We quantified 3D genomic structure changes using the Spearman correlation between distance matrices of reference and edited sequences, defining divergence scores as one minus the correlation coefficient.

**Results**

We evaluated 9917 genomic regions, divergence-scoring each for common haplotypes (haplotype count ≥ 30) found in the TopMed dataset. A total of 107 regions (1.08% of all regions) exhibited divergence scores greater than 0.001, and 17 regions (0.17% of all regions) had divergence scores exceeding 0.01. The most notable divergence, with a score of 0.113, was near the *LCOR* gene on chr 10. A specific variant at this locus, rs7477274, likely disrupts 3D genome folding by altering the DNA-binding affinity of the CTCF transcription factor. In addition to *LCOR*, other significant disruptions were observed near the *SLC41A2* and *FGF2* loci.

**Conclusion**

Our findings suggest that certain haplotypes associated with body height significantly disrupt 3D genome folding. In-silico mutagenesis coupled with 3D genome predictions provides a powerful approach to fine-map GWAS signals and identify potentially functional variants. Further work is need to confirm the functional implications of these findings.