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

**In-silico mutagenesis identifies body height associated haplotypes that modify 3D genome contacts**

**Body height associated haplotypes influence 3D genome contacts**

**Machine learning reveals novel 3D regulatory mechanisms for height-associated haplotypes**

**Wanjun Gu, Erin Gilbertson, Rany Salem, John A. Capra**

**Background**

Variants associated with phenotypes in genome-wide association studies (GWAS) are predominantly non-protein-coding and regulatory in nature. Such variants may influence the three-dimensional (3D) genome architecture, thereby affecting the gene expression of target genes. Some phenotypes, like body height, have enriched SNP heritability within topologically associated domains (TADs). The 3D genome structure can be predicted in silico using machine learning models based solely on DNA sequence information. This provides an opportunity to evaluate 3D genome disruption as a mechanism underlying height-associated loci genome-wide.

and these domains provide a framework for understanding the genetic architecture of body height through extensive GWAS summary statistics.

**Methods**

We analyzed genome-wide significant regions associated with body height (top p-value ≤ 5e-5) from the largest available GWAS. To enable haplotype-aware analyses, we used the NHLBI Trans-Omics for Precision Medicine (TopMed) sequencing dataset to impute haplotypes for these significant loci across diverse populations, including Europeans, Africans, East Asians, South Asians, and Admixed/non-admixed Americans. We then predicted alterations in the 3D genome contacts caused by each common haplotype (count ≥ 30 in TopMed).

**Results**

We evaluated 9917 height-associated regions, and 107 regions (1%) exhibited divergence scores greater than 0.001, and 17 (0.17%) had divergence scores exceeding 0.01. The most notable divergence for a height-associated haplotype was near the *LCOR* gene on chromosome 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**

We identify several haplotypes that likely influence variation in body height by modifying 3D genome folding. However, this functional mechanism is relatively rare among height GWAS hits. Our results demonstrate how in-silico mutagenesis based on powerful sequence-based machine learning models provides an efficient approach to fine-map GWAS signals and identify potentially functional variants and mechanisms.

Further experimental validation is required to confirm the functional implications of these findings.