

Submitter Name: Melissa Gymrek
Submitter email: mgymrek@ucsd.edu

The impact of short tandem repeats on complex traits

Stephanie Feupe Fotsing^{1,2}, Jonathan Margoliash³, Catherine Wang⁴, Shubham Saini³, Richard Yanicky⁵, Sharona Shleizer-Burko⁵, Alon Goren⁵ & Melissa Gymrek^{3,5}

¹Biomedical Informatics and Systems Biology, University of California, San Diego, La Jolla, CA, USA; ²Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA USA; ³Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA USA; ⁴Department of Bioengineering, University of California San Diego, La Jolla, CA USA; ⁵Department of Medicine, University of California San Diego, La Jolla, CA USA

Short tandem repeats (STRs) have been implicated in a variety of complex traits in humans. However, genome-wide studies of the effects of STRs on gene expression thus far have had limited power to detect associations and provide insights into putative mechanisms. Here, we leverage whole genome sequencing and expression data for 17 tissues from the Genotype-Tissue Expression Project (GTEx) to identify STRs for which repeat number is associated with expression of nearby genes (eSTRs). Our analysis reveals more than 28,000 eSTRs. We employ fine-mapping to quantify the probability that each eSTR is causal and characterize a group of the top 1,400 fine-mapped eSTRs. We identify hundreds of eSTRs linked with published GWAS signals and implicate specific eSTRs in complex traits including height and schizophrenia, inflammatory bowel disease, and intelligence. Overall, our results support the hypothesis that eSTRs contribute to a range of human phenotypes and will serve as a valuable resource for future studies of complex traits.