A phenome-wide association study of tandem repeat variation in 168,554 individuals in the UK Biobank

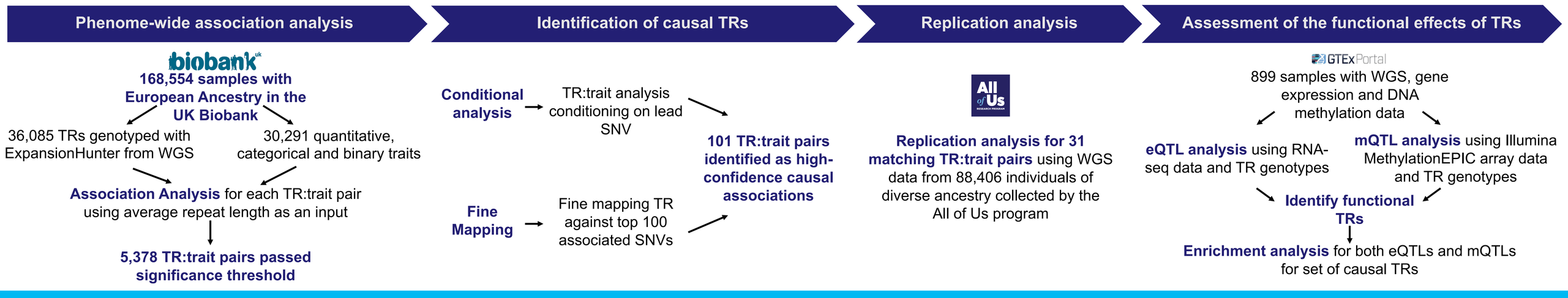
Celine Manigbas[1], Bharati Jadhav[1], Paras Garg[1], Mariya Shradrina[1], William Lee[1], Gabrielle Altman[1], Alejandro Martin-Trujillo[1,+], Andrew Sharp[1,+]

[1] Department of Genetics and Genomic Sciences and Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029. + These authors made equal contributions to this work.

Background

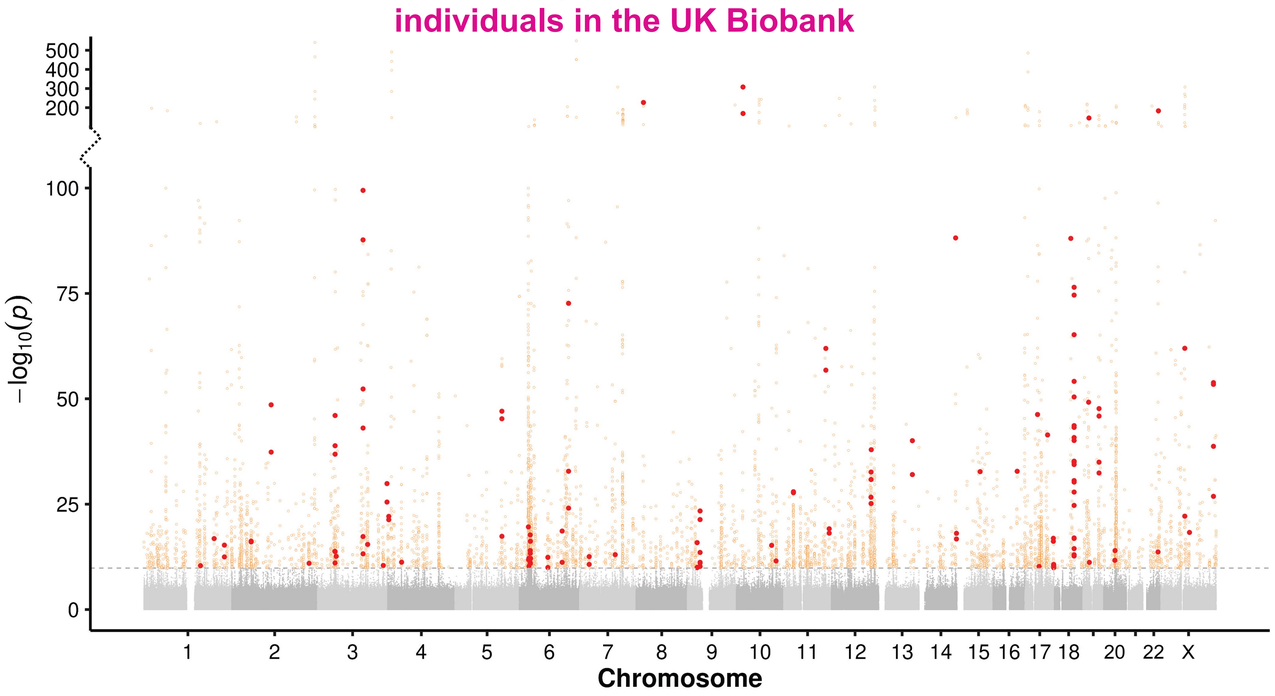
Common tandem repeat (TR) variations have been demonstrated to affect molecular phenotypes by modifying gene expression and DNA methylation[1,2]. However, their influence on human traits has not been systematically investigated. SNV-based GWAS are primarily focused on binary variants, meaning other variant types that may contribute to the 'missing heritability' of the genome, such as multi-allelic TRs, are overlooked in standard genetic studies, creating a knowledge gap. To address this gap, we investigated the role of common length polymorphisms in TRs as modifiers of phenotypic variation for thousands of human traits using phenome-wide association analysis.

Methods



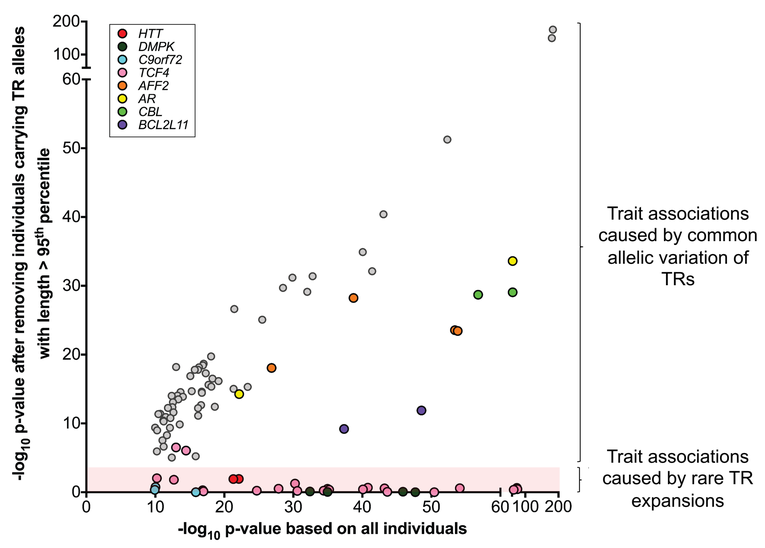
Results

1. Phenome-wide association study of 36,085 TRs and 30,291 traits in 168,554 individuals in the UK Biobank



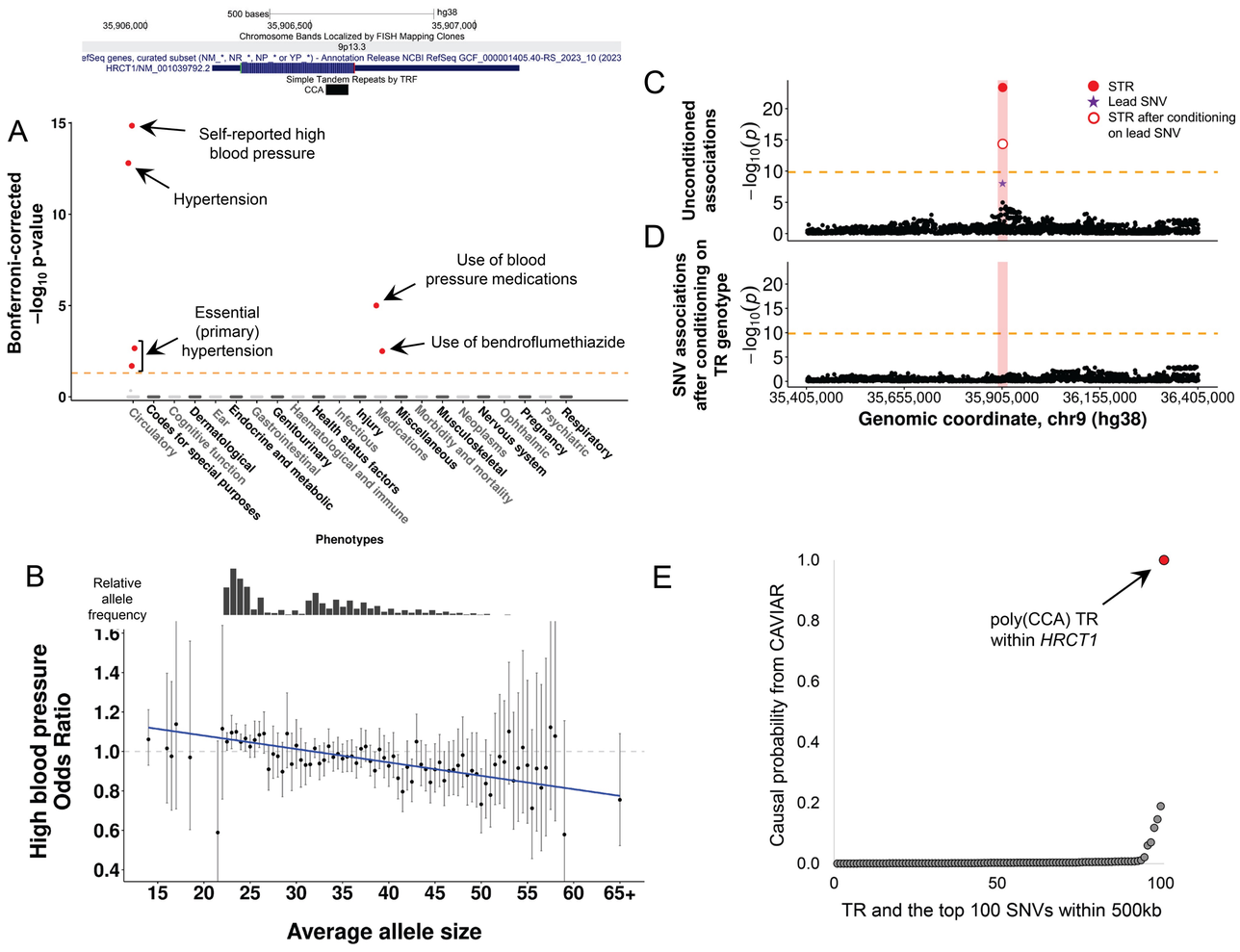
After adjusting p-values for multi-testing, we identified 5,378 significant TR:trait associations (orange). Of these, 101 TR:trait pairs are considered causal after fine mapping and conditional analysis (red).

2. Novel causal associations are driven by common polymorphic variation of TRs rather than extreme expansions

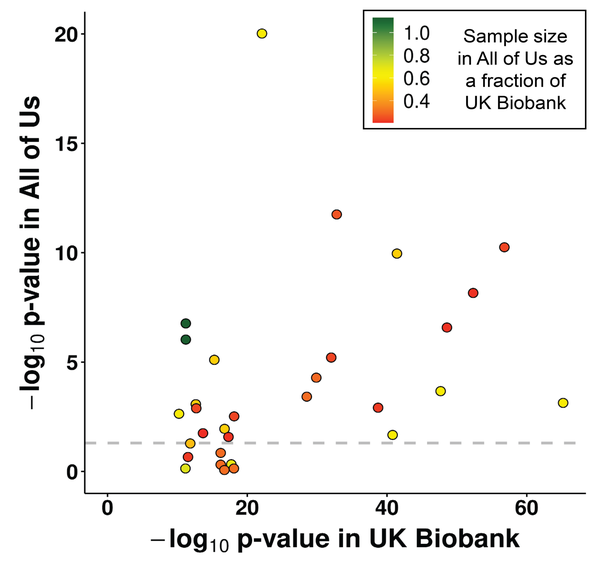


In contrast to known pathogenic repeat expansions, nearly all novel causal associations we found maintain significance upon removal of outlier alleles, indicating that they are caused by common TR variation, not rare expansions.

3. A coding poly(CCA) motif within exon 1 of *HRCT1*, a gene of unknown function, is causally associated with risk of hypertension

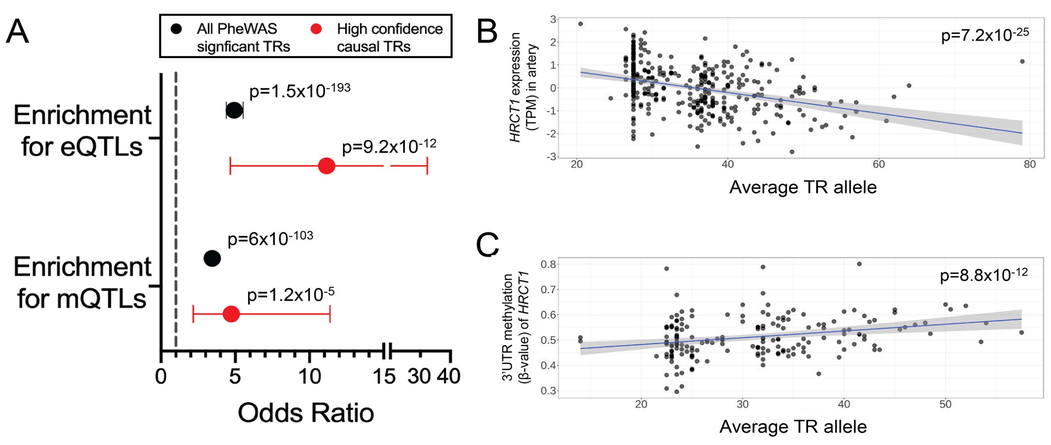


4. Replication analysis in the All of Us cohort



23 out of 31 (74%) causal associations replicated with p<0.05 and consistent effect direction observed in UK Biobank.

5. Causal TRs are strongly enriched for functional effects



(A) The coding CCA repeat in *HRCT1* is strongly associated with blood pressure related traits, (B) with carriers of the shortest 5% of alleles showing, on average, an 11% higher risk of hypertension than those carrying the longest 5% of alleles. (C,E) Causal variant analysis identified the *HRCT1* TR as the causative variant at this locus for increased risk of hypertension, (D) further supported when conditioning SNVs on TR genotypes.

(A) Causal TRs showed an 11.2 and 4.7-fold enrichment for eQTLs and mQTLs, respectively. The coding repeat in *HRCT1* shows both (B) a negative correlation with *HRCT1* expression and (C) a positive correlation with DNA methylation levels of a CpG site located in the 3'UTR of *HRCT1*.

Conclusions

- Our phenome-wide analysis provides a comprehensive evaluation of the impact of TR variation on human traits based on direct TR genotypes.
- In addition to rare TR expansions, our analysis also identified many TRs where common length polymorphism influences human traits.
- Causal variant analysis reveals that a subset of TRs represents the causal variant responsible for the phenotypic variation.
- Causal TRs are strongly enriched for loci involved in the expression and methylation levels of nearby genes, providing insights into the molecular mechanism by which TRs regulate the associated trait.
- Overall, our study highlights the contribution of multi-allelic TRs as a potential contributor to the "missing heritability" of SNV-based GWAS, emphasizing the importance of considering TRs in genetic studies.

Acknowledgements

This work was supported by NIH grants AG075051, NS105781 and HD103782 to A.J.S. and NHLBI BioData Catalyst Fellowship #5120339 to A.M.T.

This work utilized UK Biobank data through application number 82094. We gratefully acknowledge All of Us participants for their contributions, without whom this research would not have been possible

1. Fotsing SF, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S, Goren A, Gymrek M. The impact of short tandem repeat variation on gene expression. Nat Genet. 2019 Nov;51(11):1652-1659. PMID: 31676866

2. Martin-Trujillo A, Garg P, Patel N, Jadhav B, Sharp AJ. Genome-wide evaluation of the effect of short tandem repeat variation on local DNA methylation. Genome Res. 2023 Feb;33(2):184-196. PMID: 36577521

To learn more: https://doi.org/10.1101/2024.01.22.24301630
Contact me: **celineaurora.manigbas@mssm.edu**