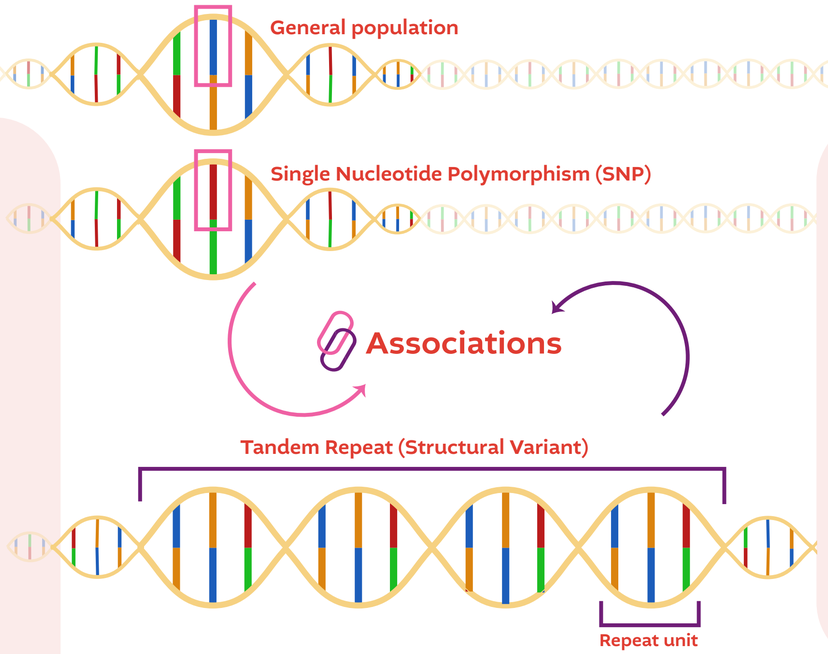
Implications of the Associations Between Structural Variants and Single Nucleotide Polymorphisms for Coronary Artery Disease Risk

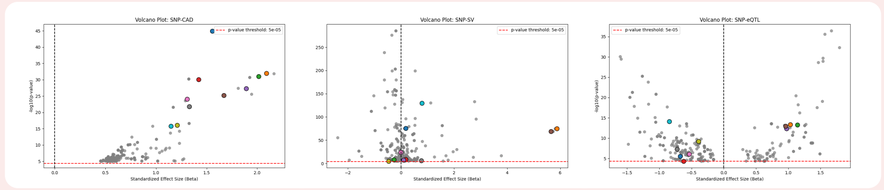
Introduction

- **Coronary artery disease (CAD)** is a condition characterized by the narrowing or blockage of the arteries that supply blood to the heart
- **Genome-wide association studies (GWAS)** identify differences in allele frequency of genetic variants between individuals with shared ancestry who differ phenotypically (exhibit a trait / disease)
- **Single nucleotide polymorphisms (SNPs)** are substitutions of a single nucleotide at a specific position in the genome.
- **Structural variants (SVs)** are larger-scale alterations in the DNA structure, further divisible into subtypes based on the nature of the structural change
- Previous **GWAS** established various SNPs associated with coronary artery disease
- Studying SVs significantly more difficult due to their **larger size and complexity**
- Established association data between SNPs and SVs can be combined with CAD-SNP associations to gain insight into the effect of SNP-SV associations on CAD pathology



Results

- Identified a list of 968 SNP-SV associations highly associated with CAD
- Size of a SV showed no correlation with CAD risk (equal frequency of long/short)
- Identified SNPs highly significant in all 3 datasets:
  - **rs6728861** (orange) associated with SV **chr2_203034349_D_19E0M1** (TE) and increase in expression of gene **CEP63**
  - **rs8046696** (brown) associated with SV **chr16_75395934_D_B8FMF** (TE) and decrease in expression of gene **RBM6**



- Heatmap of effect sizes and genes of the CAD-SNP-SV-eQTLs with all p-values < 5e-08 showed downregulation associated with shorter SVs

Research Question

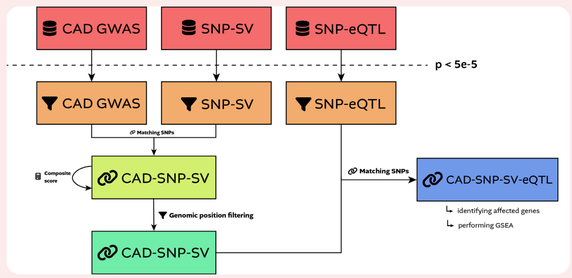
What are the **SVs** associated with **CAD-SNPs** and do the identified **SNP-SV** pairs help explain the **biological association with CAD**?

Methods

- SNP-SV associations are matched with GWAS established CAD-SNPs to obtain CAD-SNP-SV associations
- Associations are then ranked by composite score:

$$\textbf{Composite Score} = |\beta_{std,\,CAD}| + |\beta_{std,\,SNP\text{-}SV}| + \log\left(1/p_{combined}\right)$$

- Associations are filtered by genomic position, retaining the highest composite score SNP per SV within a 500kb window; finally they are matched with CAD-SNP-eQTLs



Discussion

- Previous studies linked SVs to CAD associated genes like **LDLR**, although those genes weren't identified in this study
- Research into the effect of SVs on CAD is **few and far between**
- Highlighted specific SVs which might assist in closing the knowledge gap

- **Future work and improvements:**
  - Composite score overtaken by SNP-SV p-value: introduce a **weighted variant** to balance contribution of terms
  - Sophisticate the **genomic position filtering algorithm**
  - More research in the area is required to reveal the full extent of the role of SVs in the development of CAD

Author: Boris Pavić - bpavic@tudelft.nl

Responsible professor: M. Reinders

Supervisor: N. Tessi

TU Delft