ESTABLISHING LINKAGE DISEQUILIBRIUM BETWEEN SNPs AND POLYMOPRHIC INVERSIONS

Éric García Hoyos

Bachelor's degree final project

Genetics degree

UAB Universitat Autònoma de Barcelona

INTRODUCTION

Identify and genotype inversions (a specific type of structural variant, SV) is a challenging field in genomics due to the:

- Presence of repetitive sequences (Segmental Duplications, SDs; Inverted Repeats, IRs; Mobile Elements Insertion, MEI).
- Non Allelic Homologous Recombination (NAHR) events between pairs of SDs, that contribute to the inversion recurrence [1].

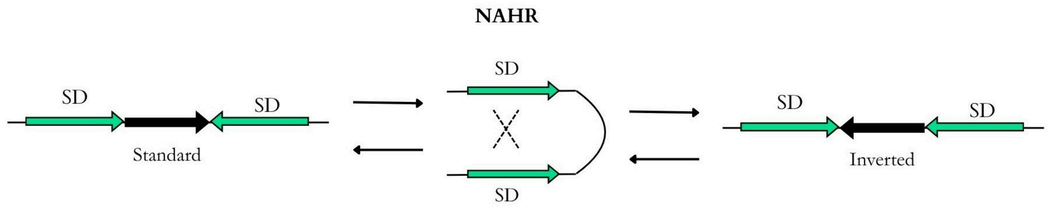

Figure 1: general architecture of an inversion flanked by SDs (marked in green). The inversion status can change depending on NAHR events, from standard to inverted and viceversa.

INVESTIGATION FRAMEWORK

Finding Single Nucleotide Polymorphisms (SNPs) perfectly associated (tagSNPs) to these SVs can serve to validate genotypes [2]. To do so, in this populational study linkage disequilibrium (LD) is calculated using the squared correlation coefficient (r^2) for a total of 44 samples.

With this, it is possible to probe that inversions flanked by SDs are underrepresented among inversions with at least 1 tagSNP and that inversions generated by alternative mechanisms are overrepresented.

Figure 2: locations marked depending on the ancestry of the samples. **EUR** (European Ancestry), **AFR** (African Ancestry), **SAS** (South Asian Ancestry), **EAS** (East Asian Ancestry) and **AMR** (American Ancestry).

OBJECTIVES

Develop a bioinformatic pipeline to provide a single file with merged data containing information for inversions, SNPs and the corresponding LD between alleles.

Evaluate LD depending on the genomic context of inversions and find inversions perfectly tagged by surrounding SNPs.

Determine whether the presence of specific genetic elements, at the ends of inversions, influence the association between alleles of genomic variants.

METHODOLOGY

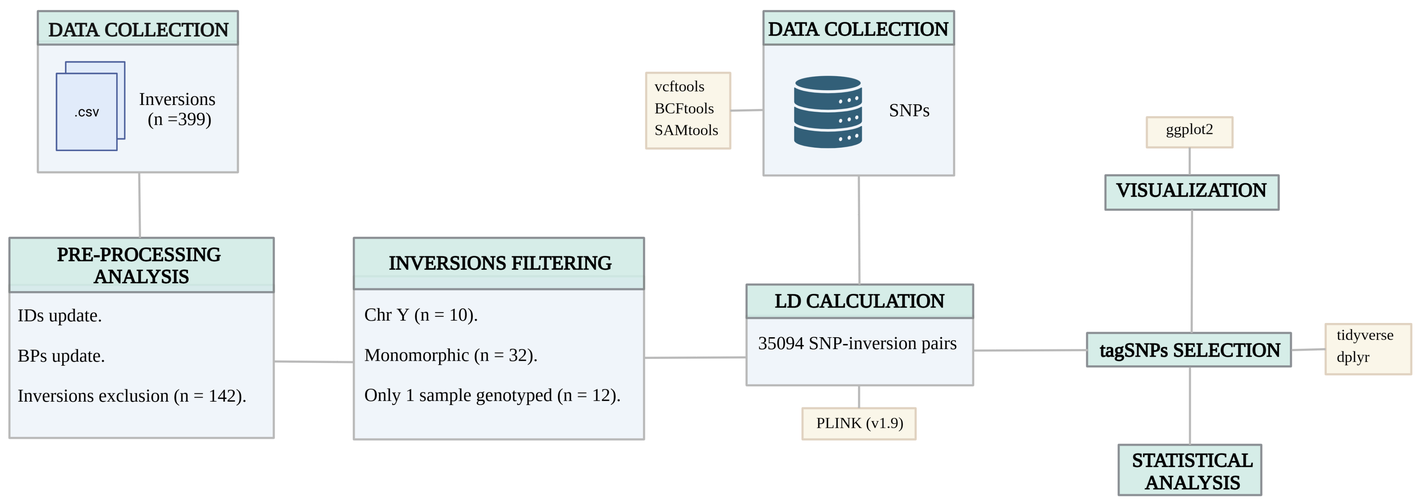

Figure 3: workflow for the different steps in which this research has been divided. Results coming from filtering steps are indicated. Green squares denote the different phases and orange squares mark different programs or packages used. Created using BioRender [3].

RESULTS I

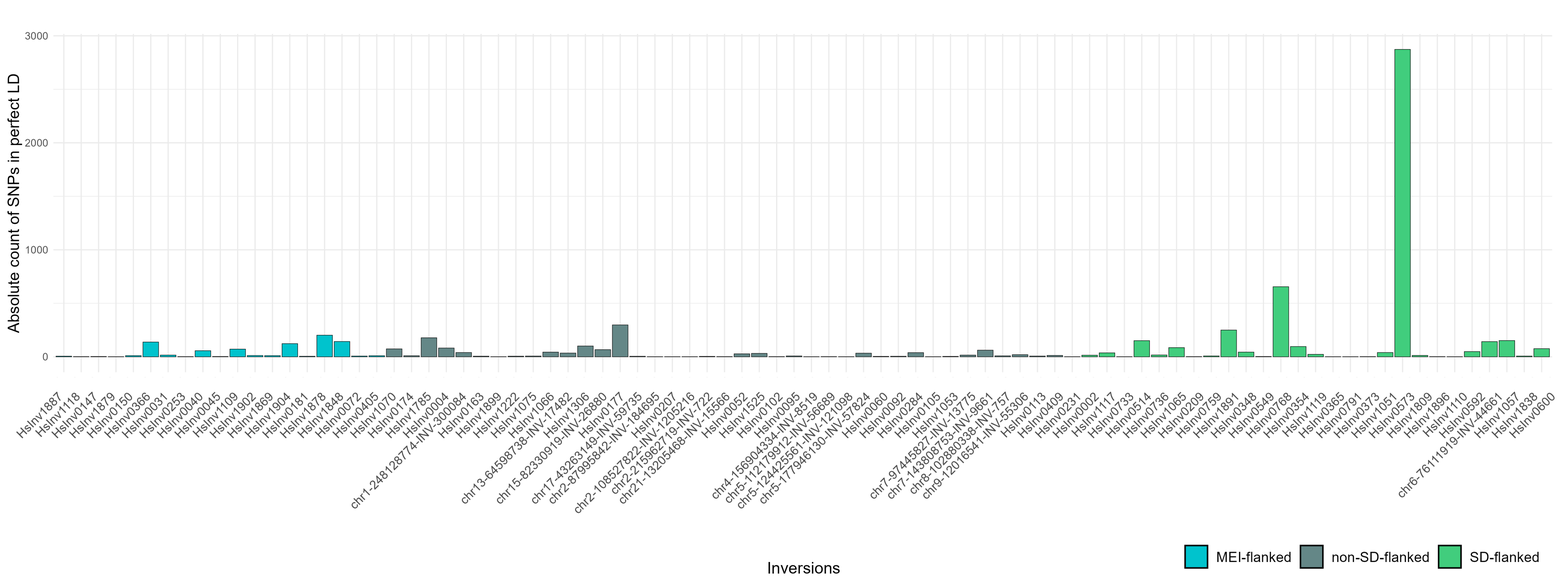

Figure 4: absolute count of tagSNPs (y-axis) divided by inversion (x-axis) and flanking information. Some of these inversions, as they had only 1 tagSNP, show shorter bars.

86 polymorphic inversions contain at least 1 tagSNP (although some of them show a greater number).

- 39 **non-SD-flanked** inversions.
- 28 **SD-flanked** inversions. → HsInv0573 → Tagged by a great number of SNPs, as previously reported [4].
- 19 **MEI-flanked** inversions.

RESULTS II

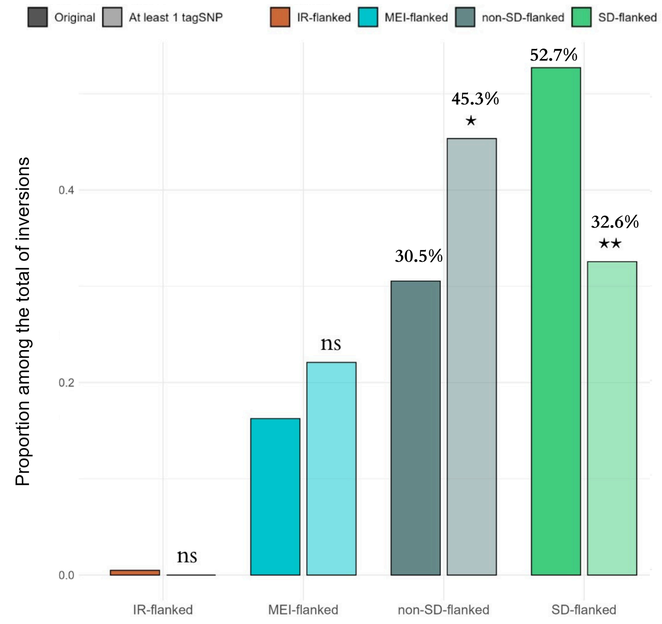

Figure 5: barplot representing proportion (y-axis) among polymorphic inversions (darker colours) and with at least 1 tag SNP (lighter colours). Four categories (x-axis) are divided depending on the genomic context.

Among inversions containing SNPs in perfect LD:

- **SD-flanked** inversions are underrepresented (adj. p-value of 0.0077).
- **non-SD-flanked** inversions are overrepresented (adj. p-value of 0.0429).

CONCLUSIONS

- An online public repository can be accessed via the QR code, where a series of scripts evaluating LD measures, inversions filtering and statistical analyses can be found.
- Haplotype diversification in SD-flanked inversions is coherent with NAHR origin, while inversions that lack SDs at their BPs could be originated in one single inversion event that left the inversion status perfectly associated ($r^2 = 1$) with SNPs' alleles.
- This work could be complemented with studies considering larger sample sizes, population structure biases and alternative LD measures.

REFERENCES

[1] Ebert, P. et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**, eabf7117 (2021).

[2] Giner-Delgado, C. et al. Evolutionary and functional impact of common polymorphic inversions in the human genome. *Nature Communications* **10**, 4222 (2019).

[3] BioRender. *BioRender.com* https://biorender.com. (2024).

[4] Campoy, E., Puig, M., Yakymenko, I., Lerga-Jaso, J. & Cáceres, M. Genomic architecture and functional effects of potential human inversion supergenes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **377**. Epub 2022 Jun 13, 20210209 (2022).