

# Inversion filtering analysis

*Ruth Gómez Graciani*

We started with the following list of inversions. When (n = m) inversions is stated, it means there are n inversions in the list, of which m are unique inversions and n-m are repeated because they had DEL genotypes.

- INVERIONS STEP01: autosomal inversions (106 = 104)
- INVERIONS STEP02: inversions with imputability info (111 = 109)
- INVERIONS IMPUTE2 = STEP01 int STEP02: autosomal invs with imputability info (94 = 92)
- INVERIONS STEP03: is imputable or has tag SNPs or is non polymorphic in at least one population (93 = 92)
- INVERIONS STEP04 = STEP03 int. STEP01: imputable autosomal inversions (83 = 82)

After the imputation, the genotypes were filtered differently depending on the individual's population. The amount of inversions analyzed with IMPUTE2 is 94 = 92 (autosomal inversions with imputability info), so the table has 1880 rows (20 individuals x 94 inversions).

Genotypes filtering steps:

- GENOTYPES STEP01: Inversions marked as imputable, tagged or non polymorphic were accepted. SAS, ALL and EAS individuals were filtered according to the GLB imputability column. EUR and AFR were filtered according to their corresponding columns. Inversions HsInv0191 and HsInv1075 were not accepted for SAS and ALL individuals because it was observed in AFR and EUR individuals that good quality (all results with probability >0.8) not always implied good uniformity (all 'global' results equal and equal to 'same population' control).
- GENOTYPES STEP02: Quality filter. Individuals without a same population control result (mostly ALL and SAS, but some EAS as well), were required to have probabilities > 0.8 in all the global results. Otherwise, the same population control probability was required to be >0.8.
- GENOTYPES STEP03: Uniformity filter. All genotypes for EUR and AFR individuals were accepted regardless their uniformity, because we trust the same population control provided that the corresponding probability is >0.8. Other populations' genotypes were accepted when all their available measurements predicted the same genotype (all 4 measurements in most EAS individuals and the global results only in most SAS and ALL individuals).
- GENOTYPES STEP04: for a genotype to be accepted, it must pass all three filters. The accepted genotype is the same population control when available, and the global results when the same population control is missing.

Once the genotypes are filtered, they are aggregated and counted to filter inversions by the amount of different haplotypes available.

- INVERIONS STEP05: those inversions with more than 3 individuals genotyped and with both heterozygous and homozygous individuals available (62).
- INVERIONS STEP06 = STEP04 int. STEP05: autosomal imputable good quality genotyped inversions (62)