Cornerstones of medicine

Prevention

Diagnosis

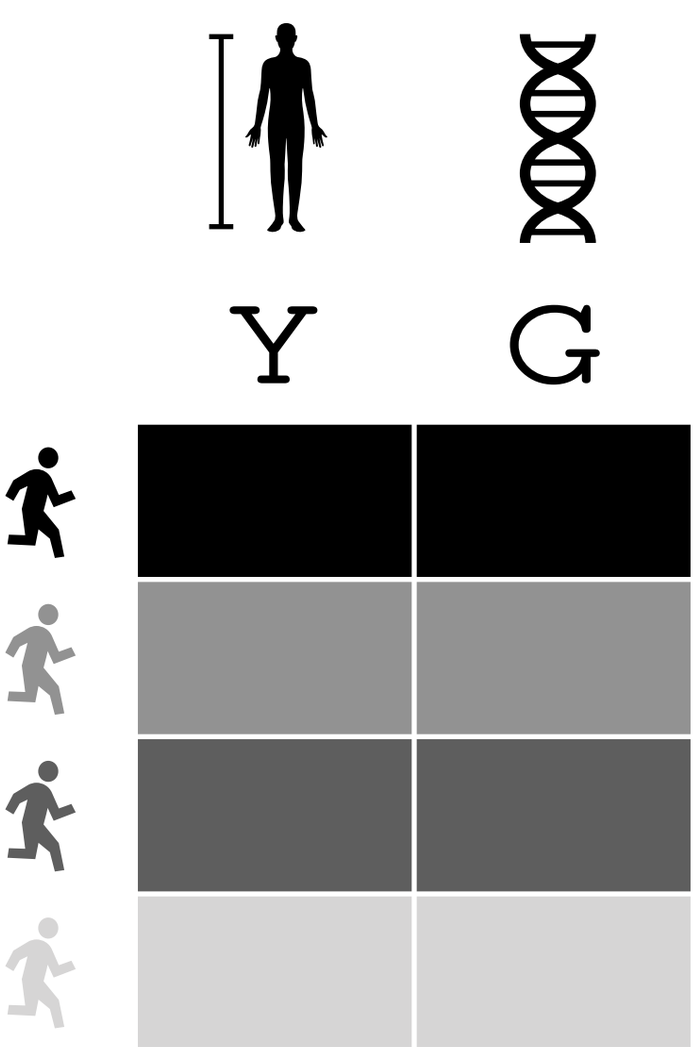
Therapy

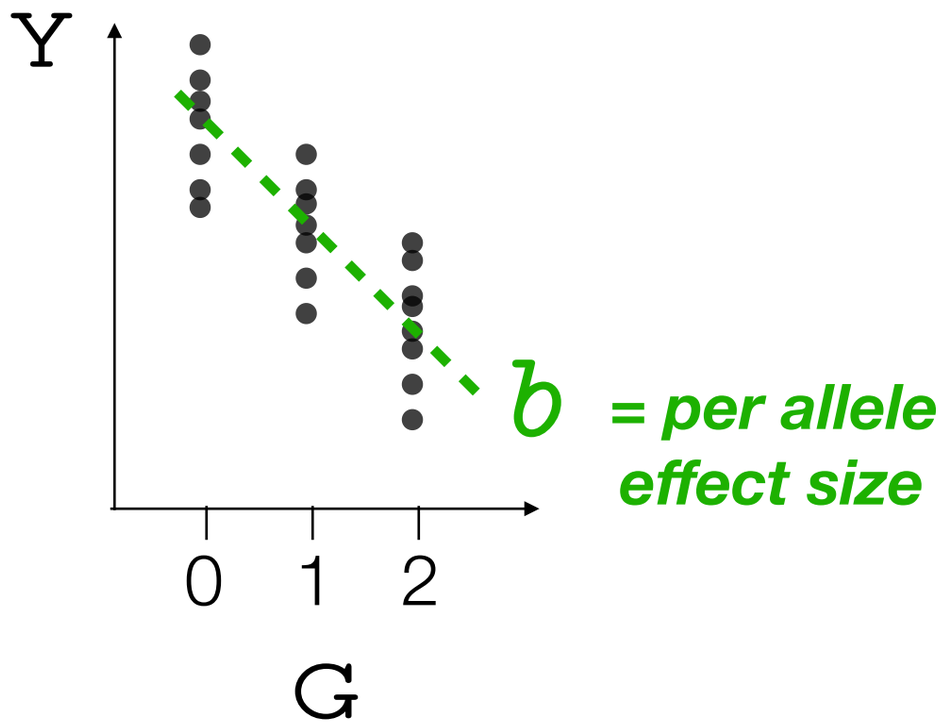
Prognosis

GWASs

$$Y = Gb + E \qquad E \sim N(0, s^2)$$

estimate b the effect of a risk allele of genotype G o

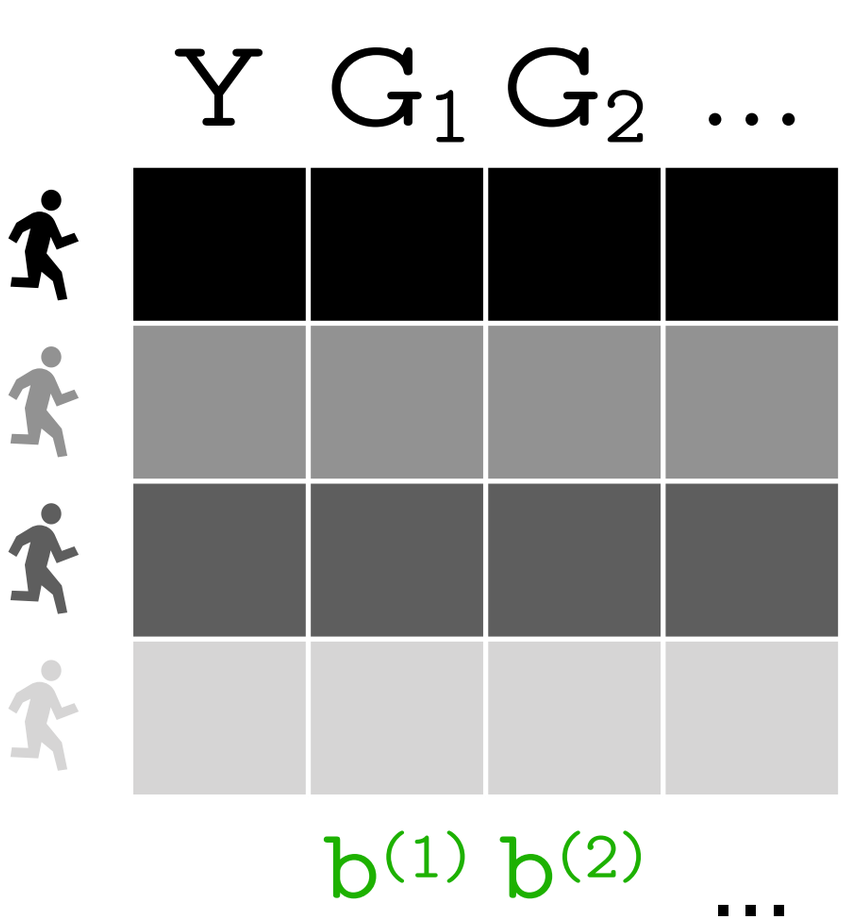


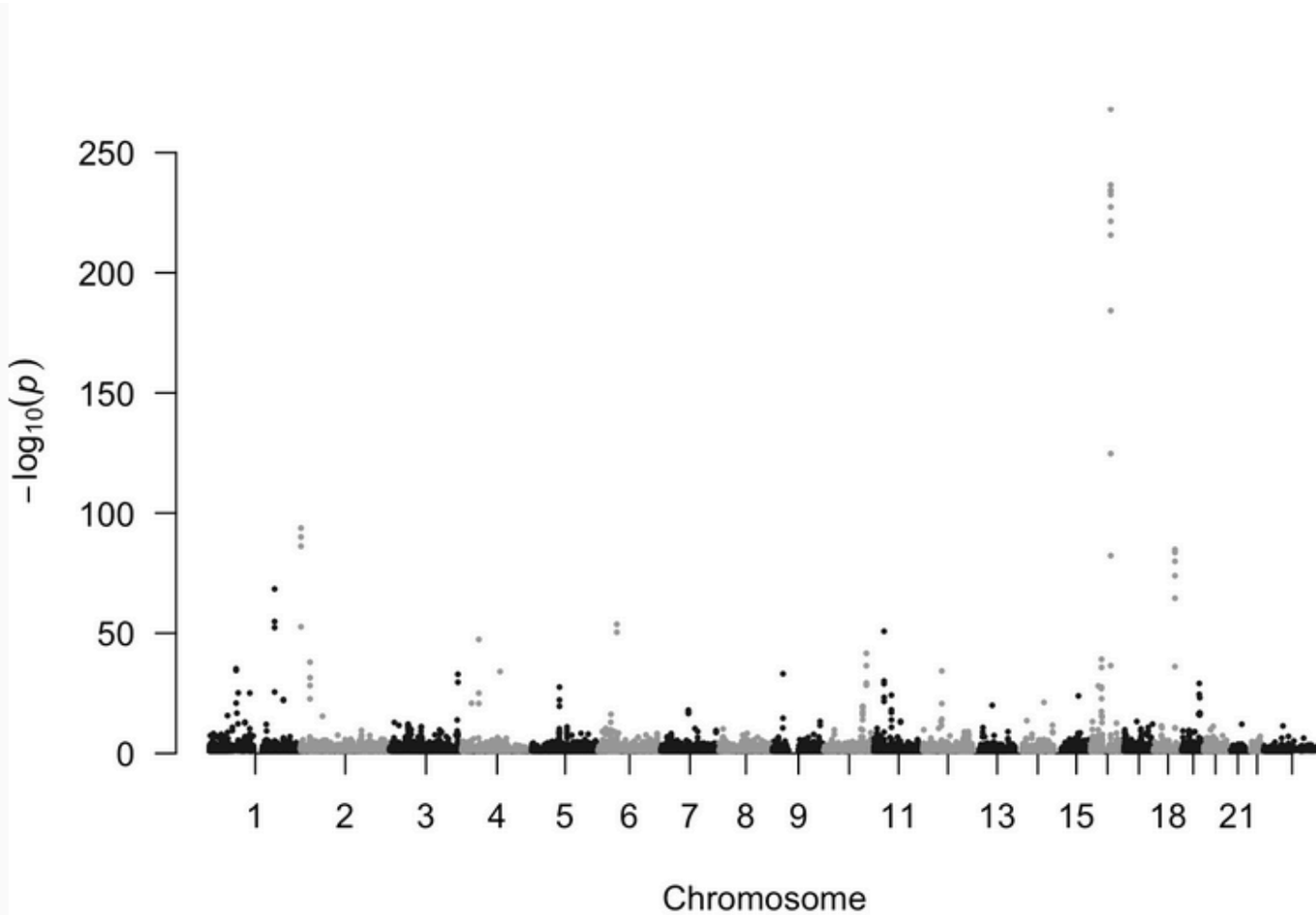


GWASs

$$Y = G_i b^{(i)} + E \qquad E \sim N(0, s^2)$$
$$i = 1, \ldots, \#\text{genotypes}$$

estimate $b^{(i)}$ *the effect of a risk allele of genotype i on Y.*



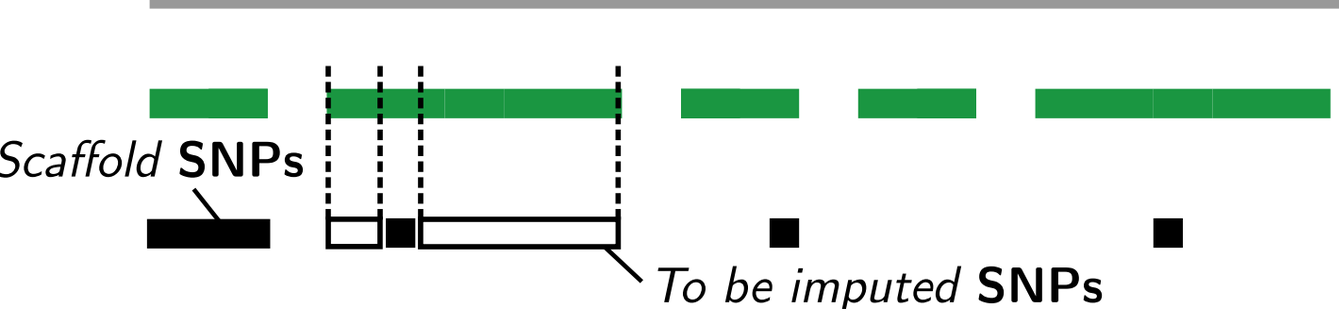


Data

Genome ~ 3 bio bp

1000 Genomes Project ~ 15 mio SNVs

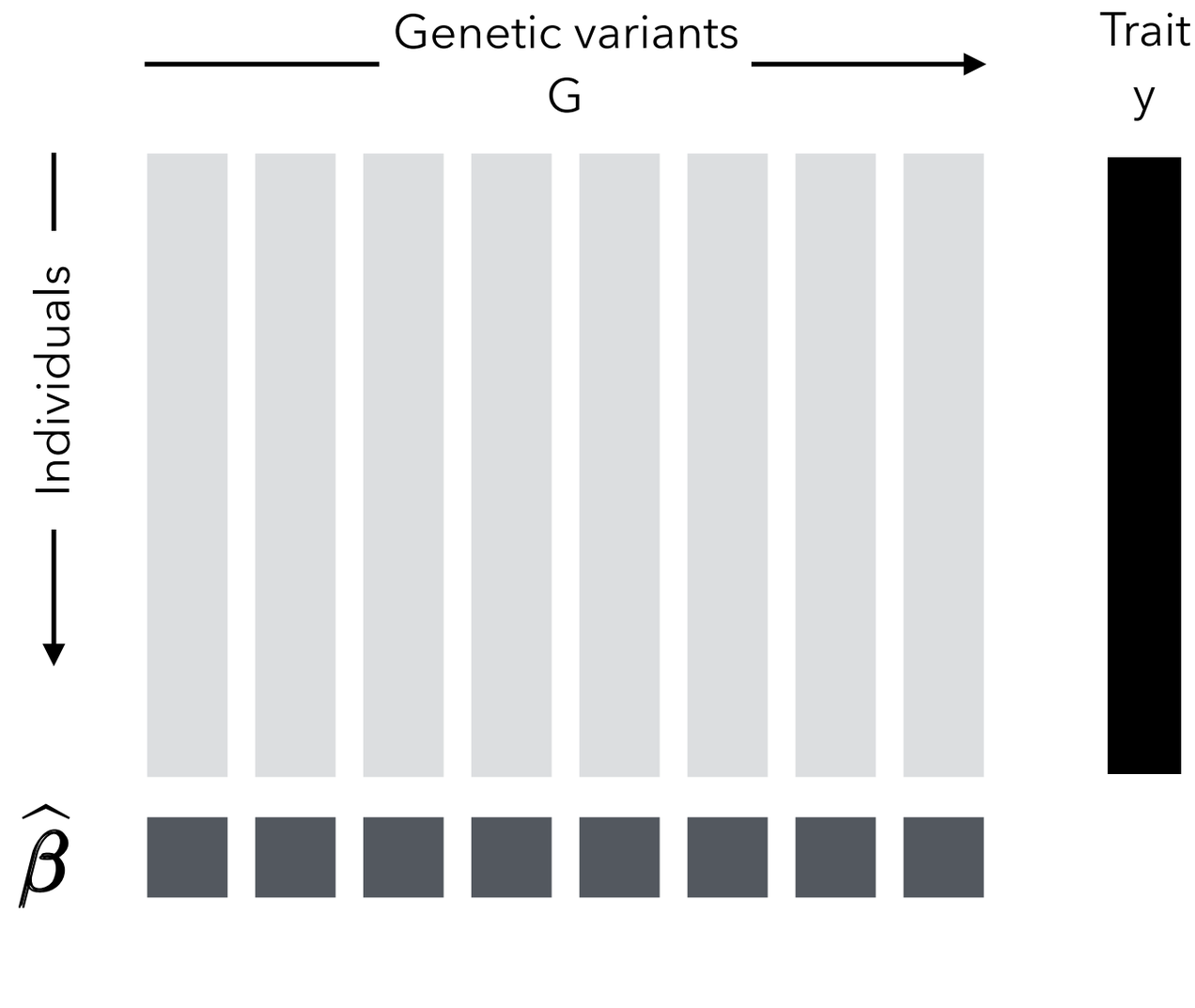
SNP array data



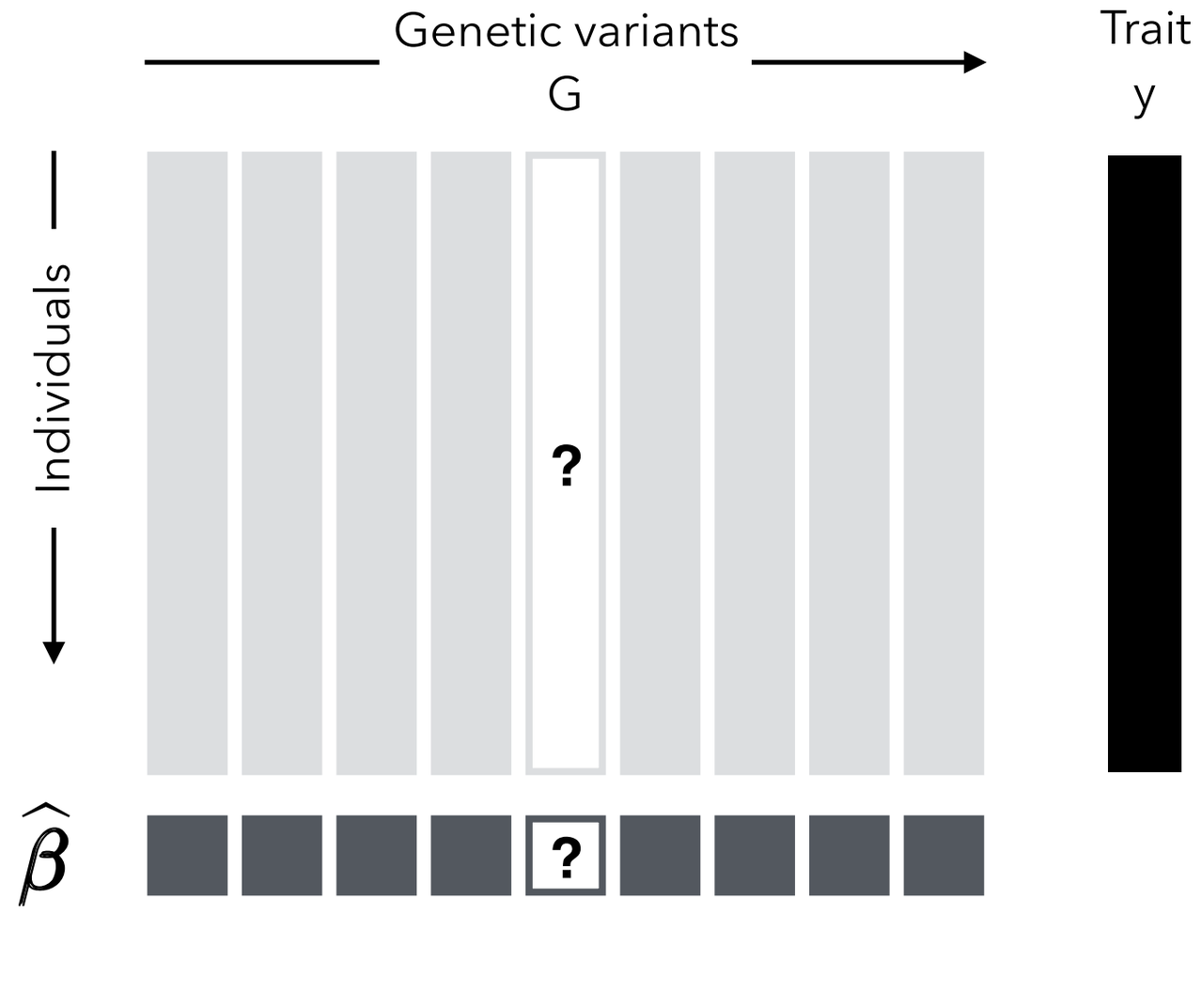
GWAS

Summary statistic

SNP	effect size	standard error	sample size	allele frequency
G_1	$b^{(1)}$			
G_2	$b^{(2)}$			
G_3	$b^{(3)}$			
...				



Summary statistics



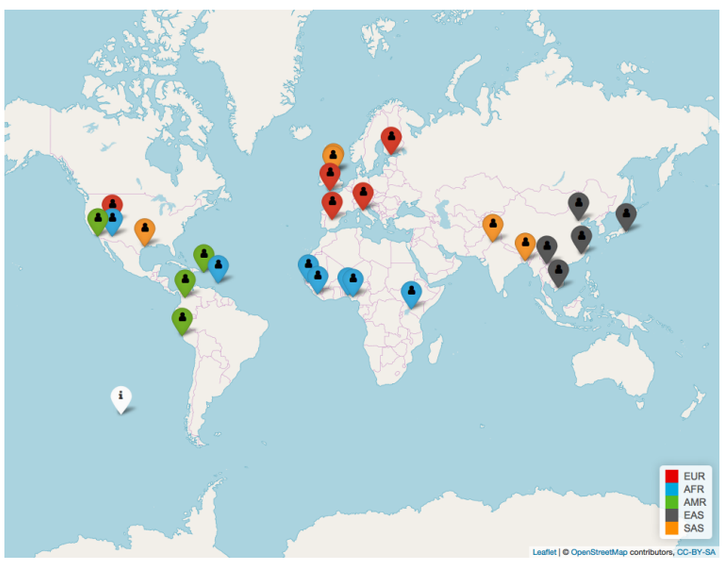
Summary statistics
of unobserved genotype:

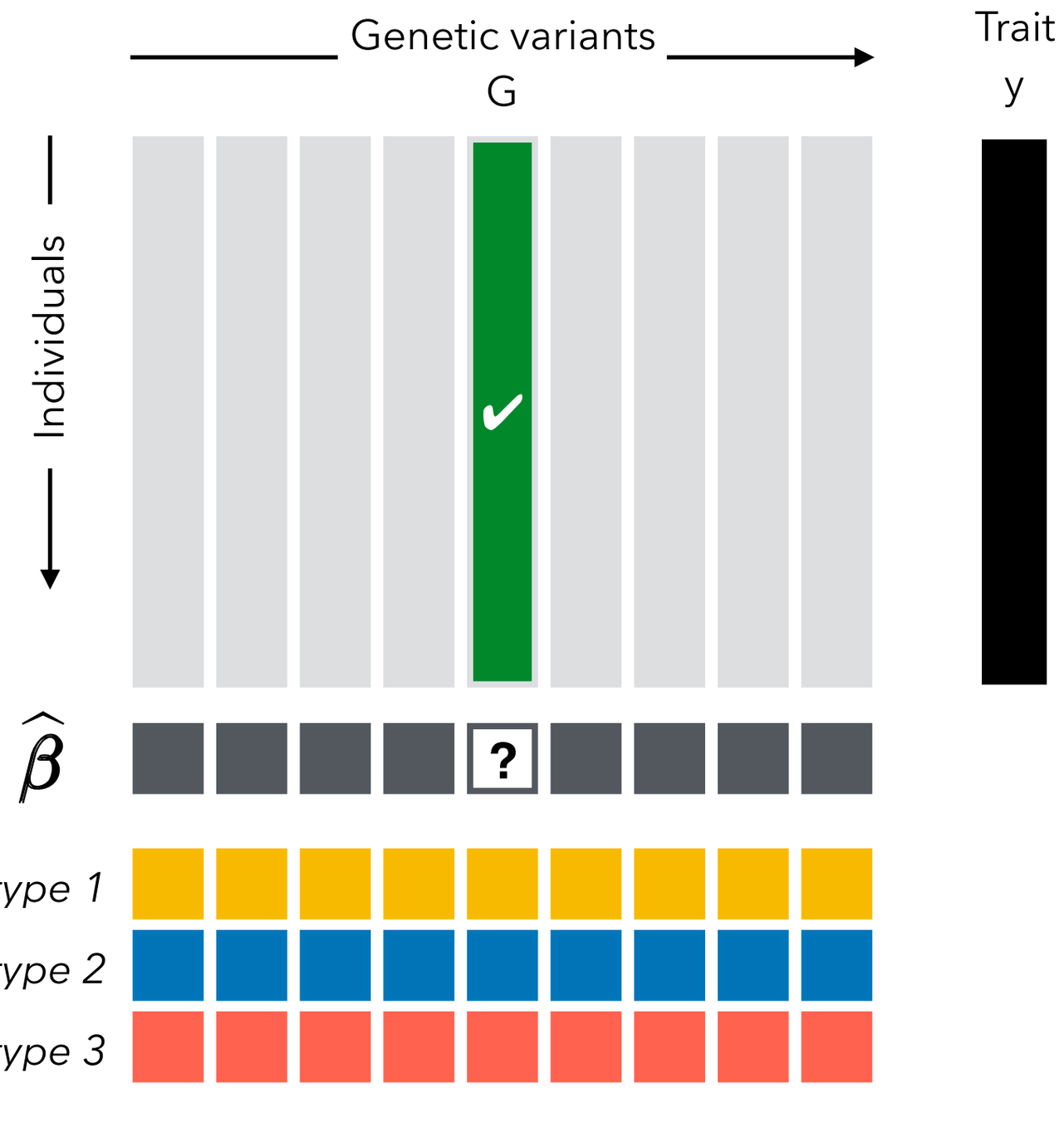
$\widehat{\beta}_i =?$

Genotype imputation

Step 1

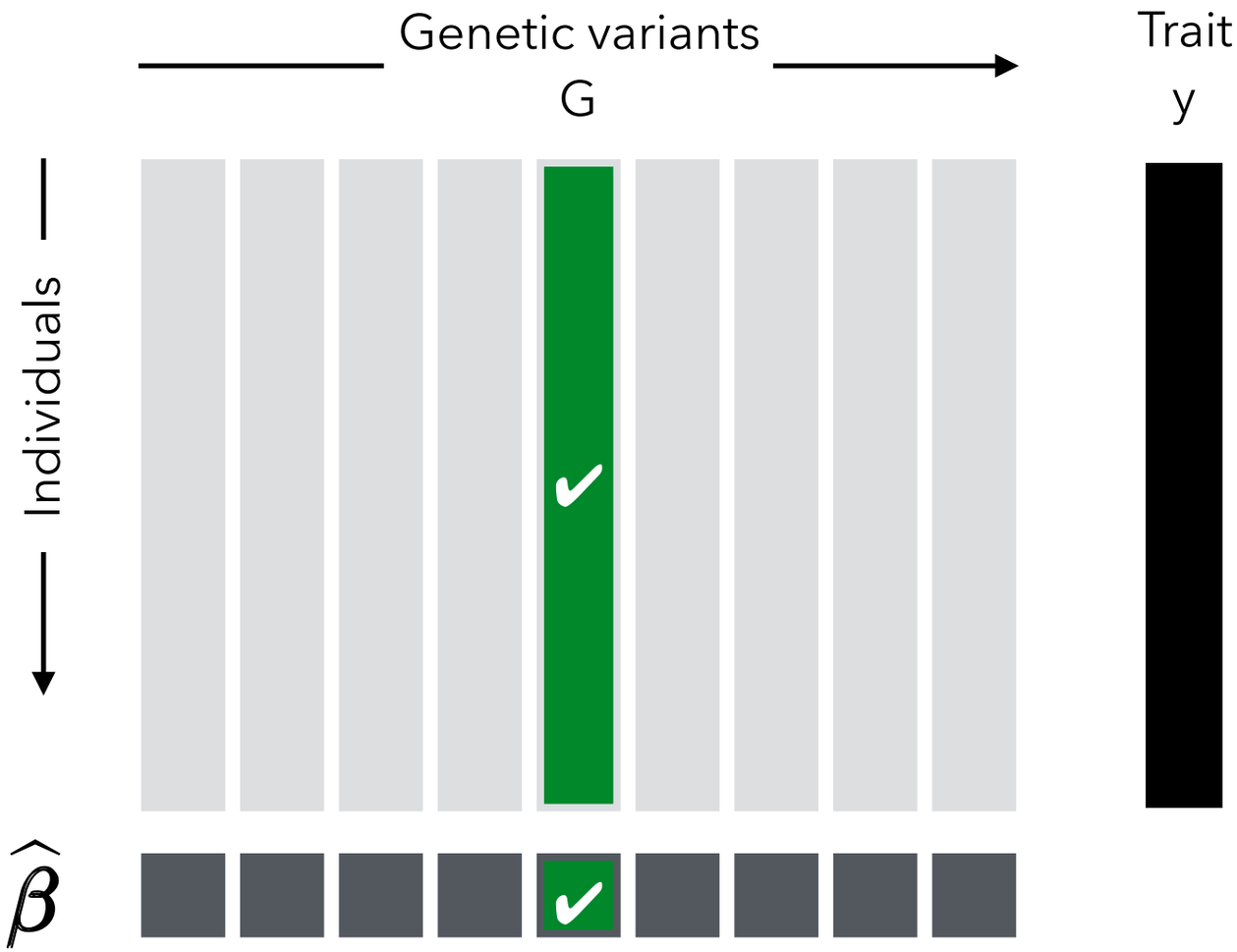
1000 Genomes



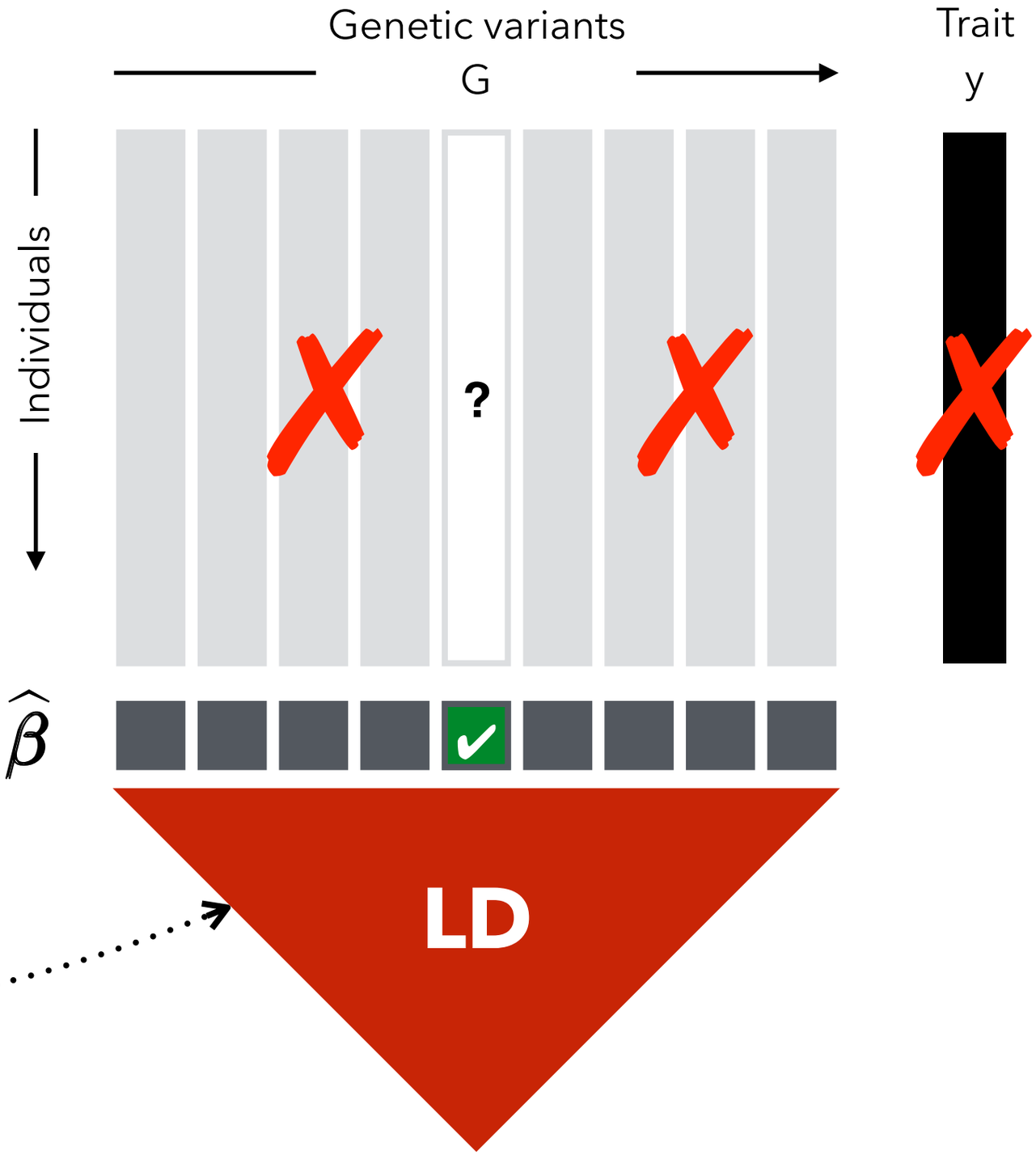


Genotype imputation

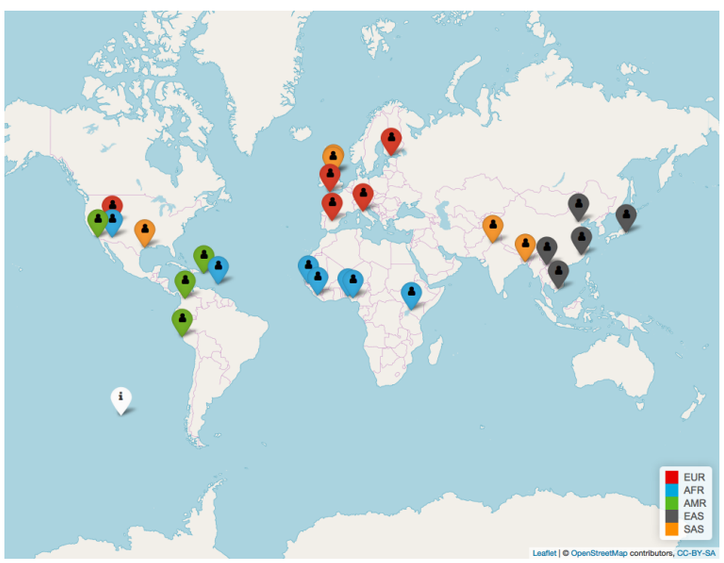
Step 2



Summary statistic imputation



1000 Genomes

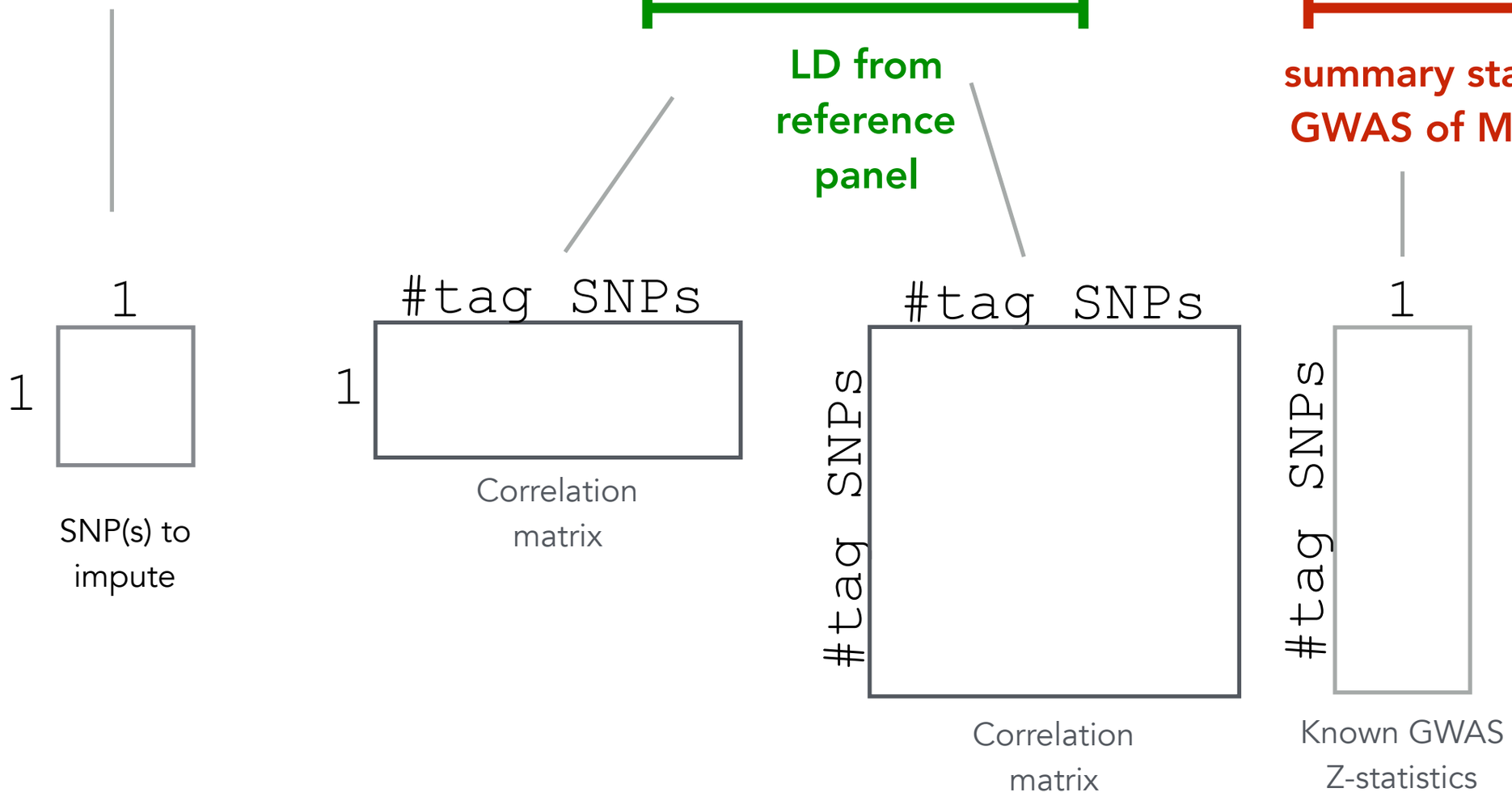


__c__ and __C__ estimate the correlation between the Z-statistics

$$\widehat{Z}_u = \boldsymbol{c}' C^{-1} \boldsymbol{Z}$$

LD from reference panel

summary statistics GWAS of M SNPs



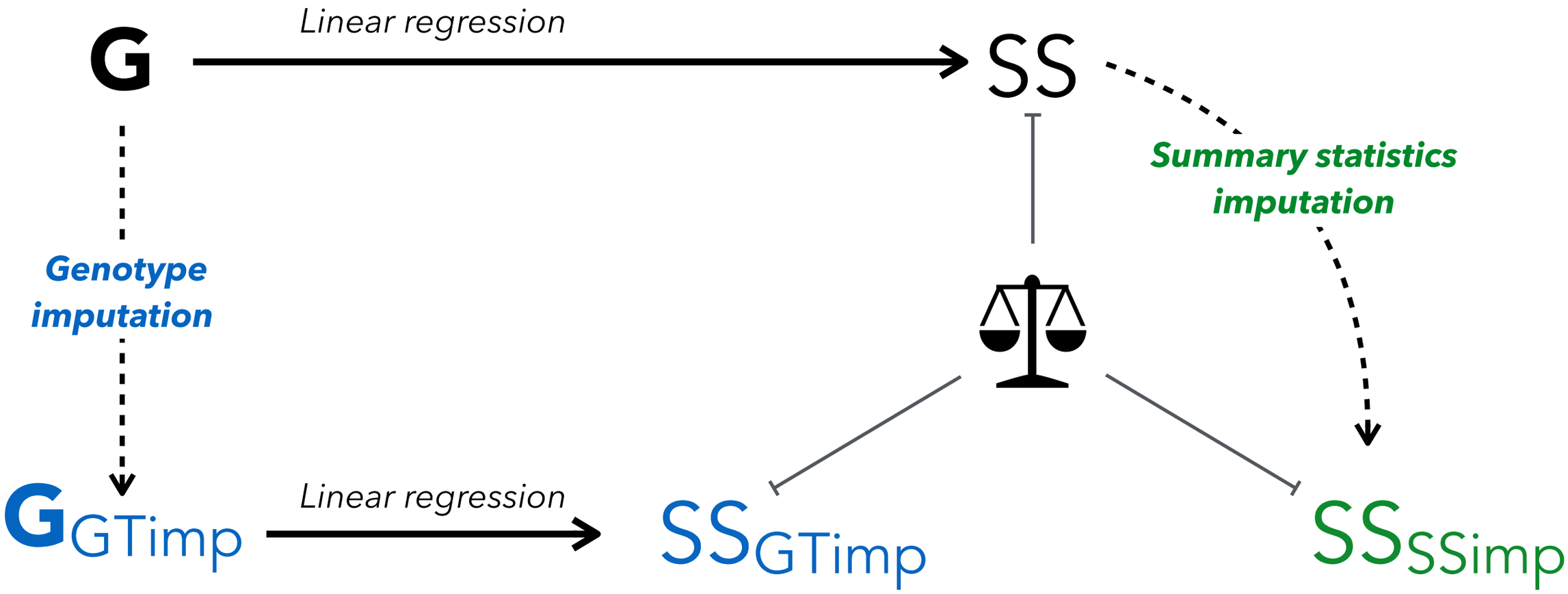
SSIMP

Methodology

- First described by Wen and Stephens (2010)
- Summary statistic imputation needs:
  1. Summary statistics for a set of genotyped markers
  2. LD structure
- Advantages compared to genotype imputation:
  - not bounded to privacy restrictions related to the use of individual-level data
  - lower computation time
  - less storage space
  - updating to newer reference panels easier

SSIMP
My contribution

- Improved the **imputation quality** towards a more accurate, yet fast to compute measure.
- Incorporated **population (ad)mixture** using the weighted LD structure of subpopulations in the reference panel.
- **Investigated shrinkage parameter** λ.
- **Compared** summary statistic imputation to **genotype imputation**,
- Tested the utility of summary statistic imputation on a **real case study** on human height.


G
Linear regression
SS
Summary statistics imputation
Genotype imputation
GGTimp
Linear regression
SSGTimp
SSSSimp