Introduction to GWAS (part I)

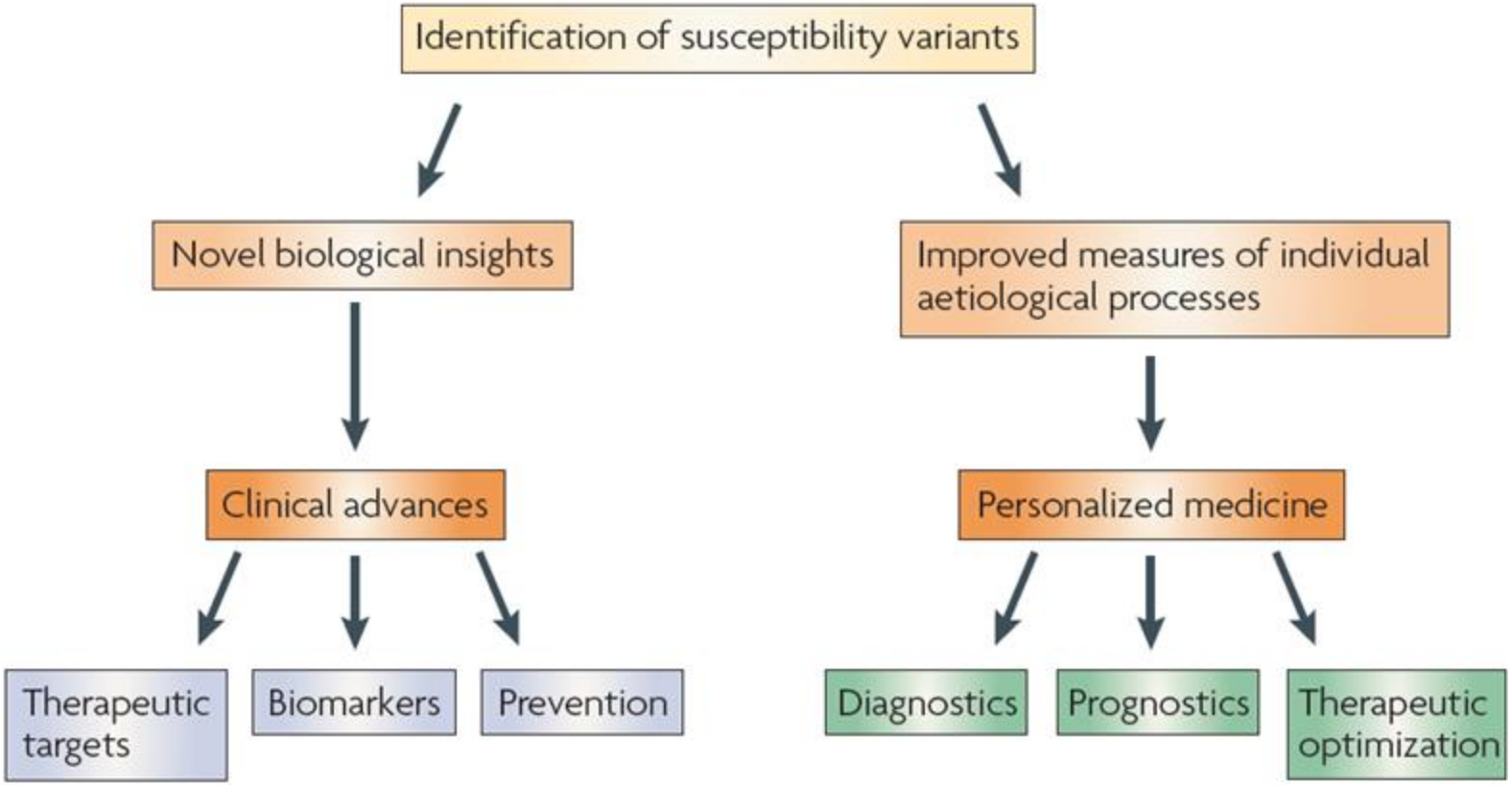
Katrina Grasby and Lucia Colodro Conde

What is it?

- A hypothesis free study of genetic variation across the entire human genome
- Tests for genetic associations with continuous traits or with the presence / absence of disease
- With a focus on low penetrance & high frequency loci
- Tests indirect association

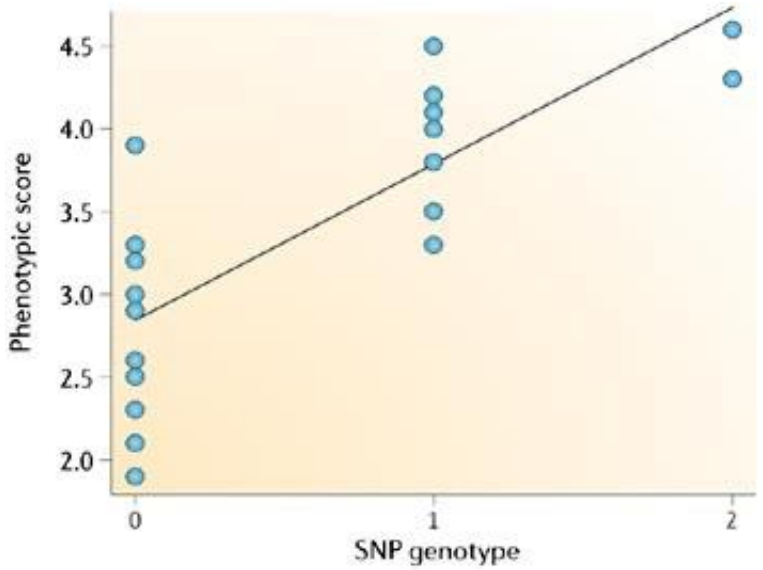
Hirschhorn & Daly. *Nat Rev Genet* (2014)

Why do it?



McCarthy et al. *Nat Rev Genet* (2008)

Quantitative Trait



Linear Regression

$$\hat{Y} = \alpha + \beta X + \varepsilon$$

$\hat{Y}$ = score on phenotype

X = 0, 1 or 2 copies of allele ("G")

$\beta = 0$ no association

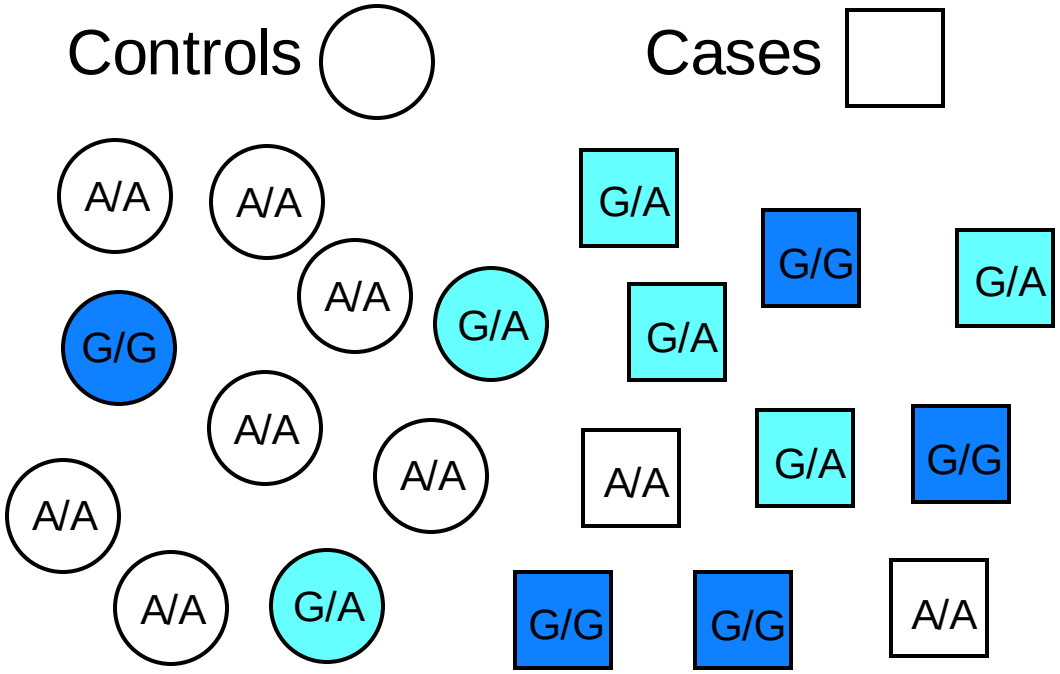
$\beta > 0$ G allele associated with higher score on trait

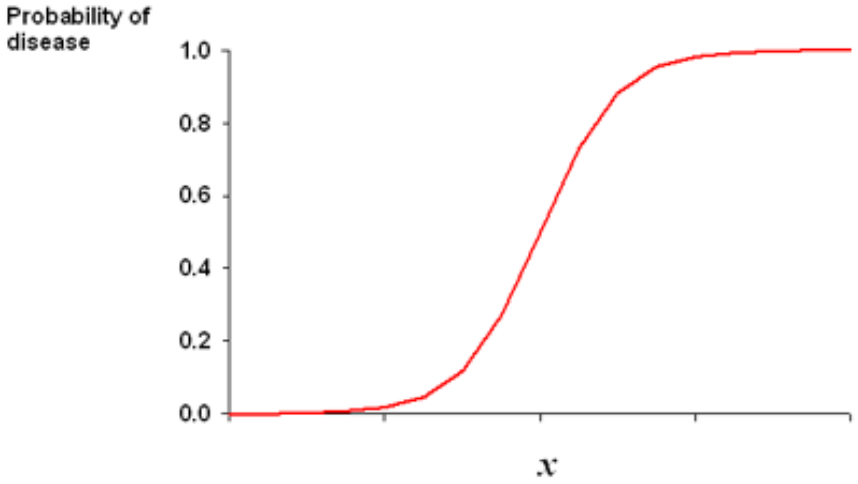
$\beta < 0$ G allele associated with lower score on trait

Balding. *Nat Rev Genet* (2006)

Case-Control

Logistic Regression





$\ln(P/1-P) = \alpha + \beta X + \varepsilon$

β = difference in log odds for cases vs. controls

$e^{(\beta)}$ = difference in odds
= Odd Ratio (OR)

Allelic effect is an OR:
OR > 1 increased risk
OR < 1 decreased risk

The G allele is associated with disease

Relatedness

- Only a few in the total sample = drop

- Random Effects Model

$$\hat{Y} = \alpha + \beta X + G + \varepsilon$$

β = fixed effect of the allele

G = genetic relationship random effect

- Genetic Relationship Matrix (GRM)
  - Sub-sample of SNPs
  - Leave One Chromosome Out (LOCO)