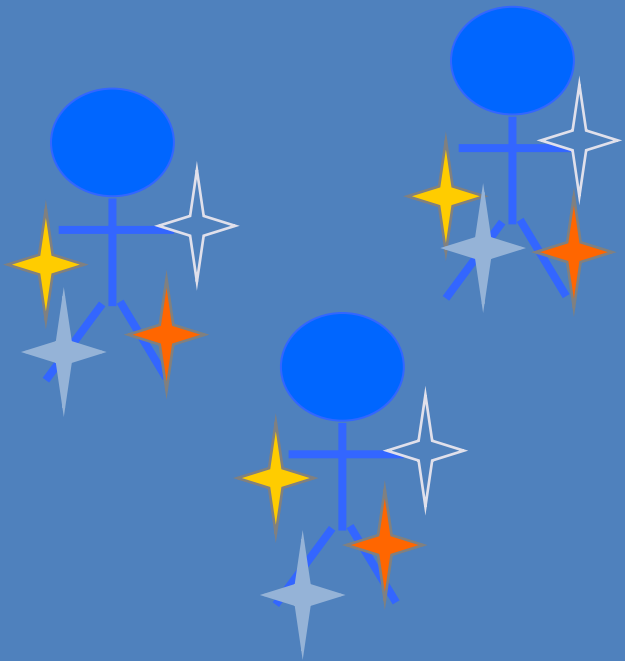


MODULE 2: Genome-wide association study methods

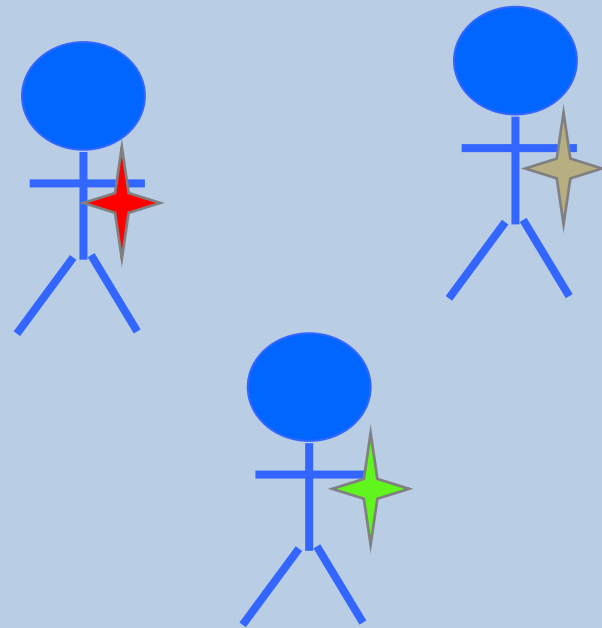
Theory behind GWAS strategy

Common disease – common variant



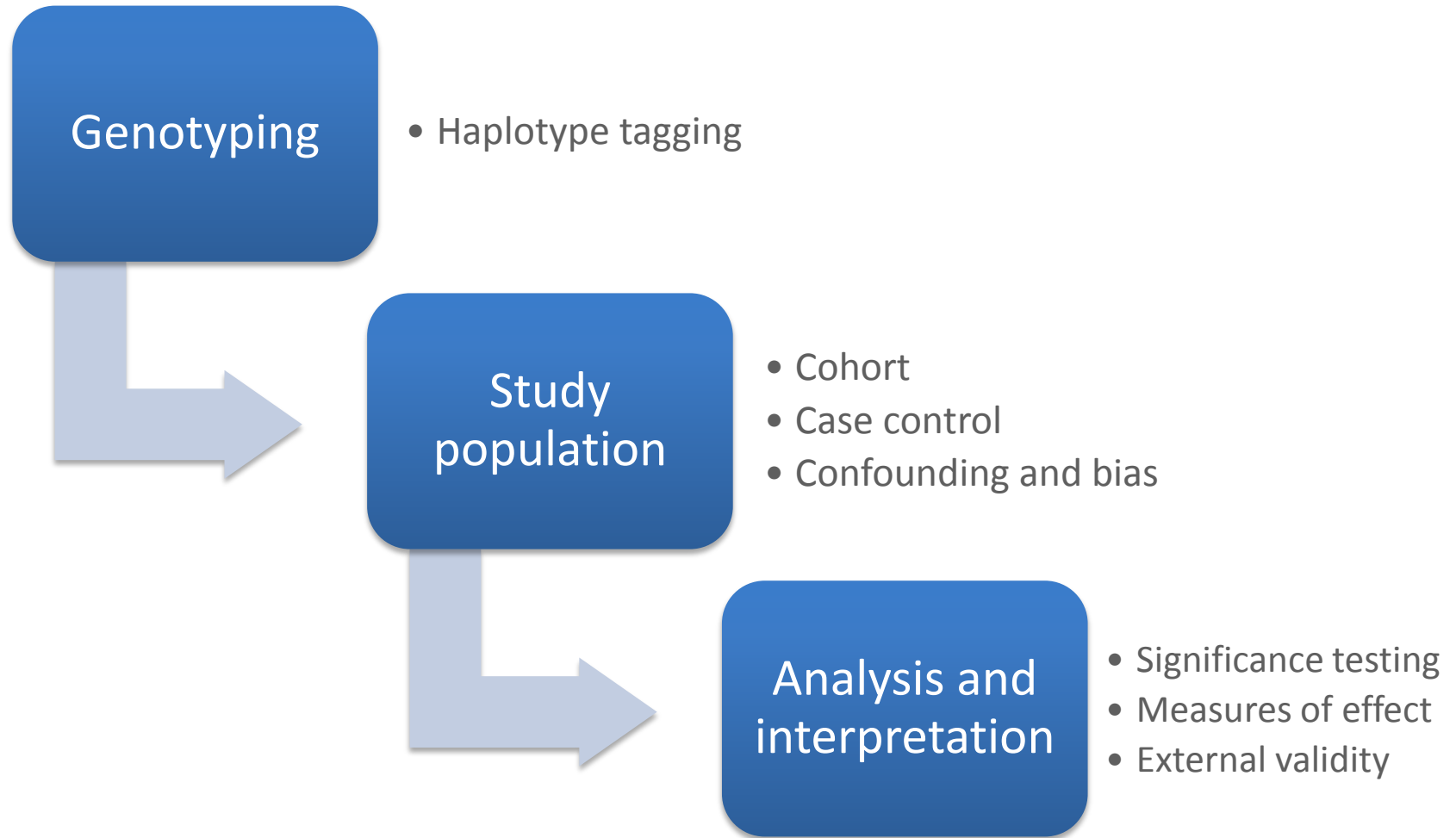
Cumulative effect of many common, low penetrance variants

Common disease – rare variant



Different single, rare, high penetrance variants

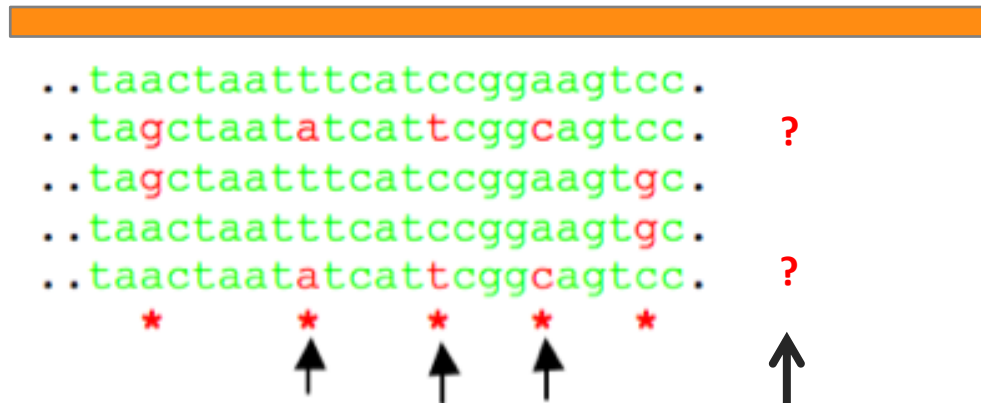
GWAS approach



Genotyping Platform

Genotyping arrays/SNP chips

- 1,000,000 SNPs in one experiment
- Direct and indirect capture of ‘all’ common variants by using ‘tag’ SNPs



Genotyping any ONE of these four captures all

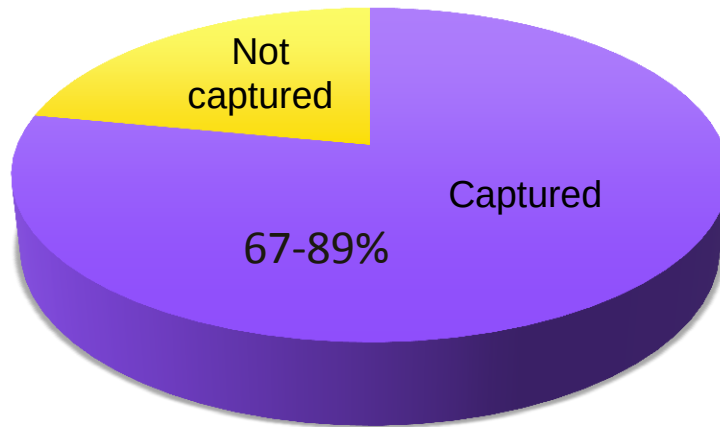


An association with a tag SNP helps define the region (block) harboring the causal variant

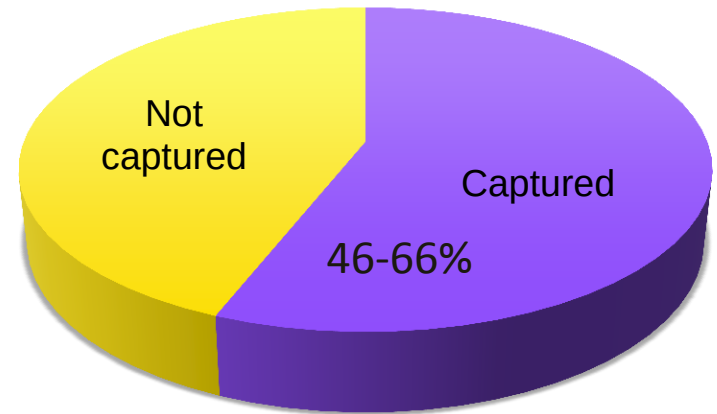
Genomic coverage of SNP chips

- How well do these chips capture common variants?

% of common variants captured by 1M
SNP chip in Europeans/Asians



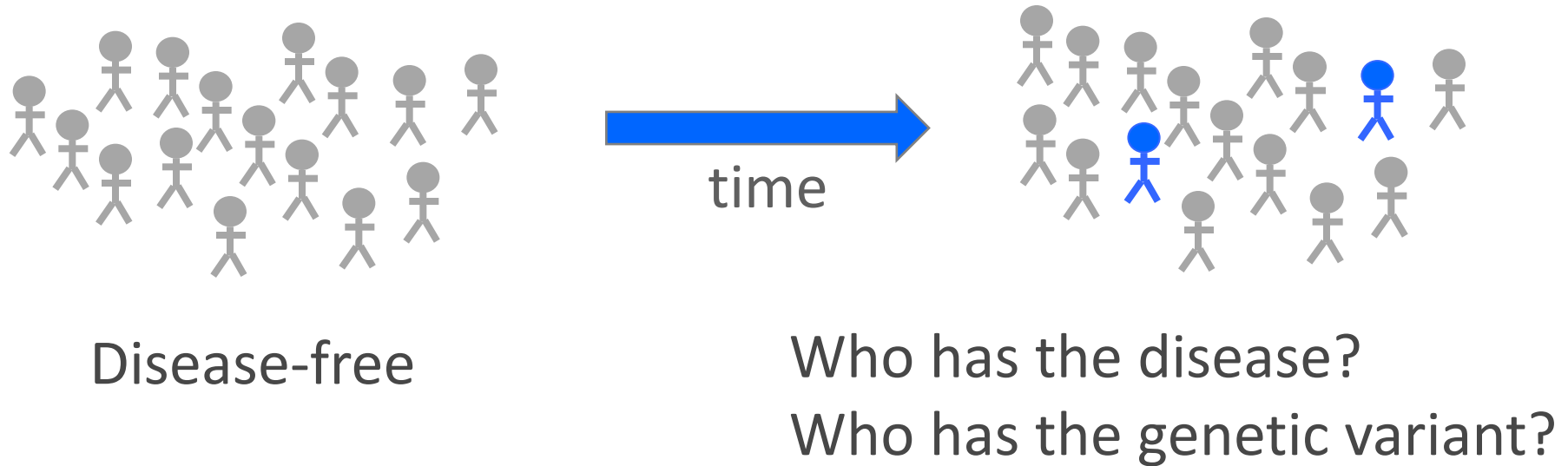
% of common variants captured by 1M
SNP chip in African ancestries



Study Designs

- Common observational studies
 - Cohort
 - Case-control
- Common biases
 - Confounding
 - Misclassification bias

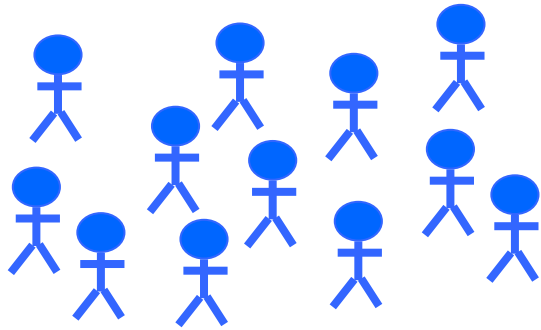
Cohort studies



Drawbacks

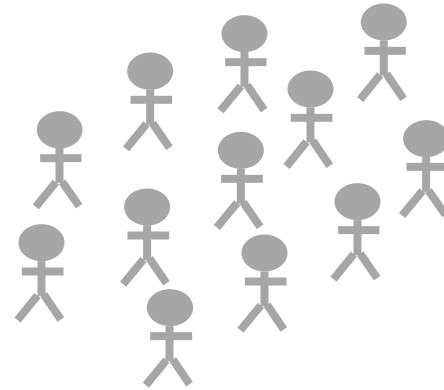
- Need to be large for rare diseases
- Need to follow a long time for diseases with long latency

Case-control studies



With disease

How many have gene variant?



Disease-free

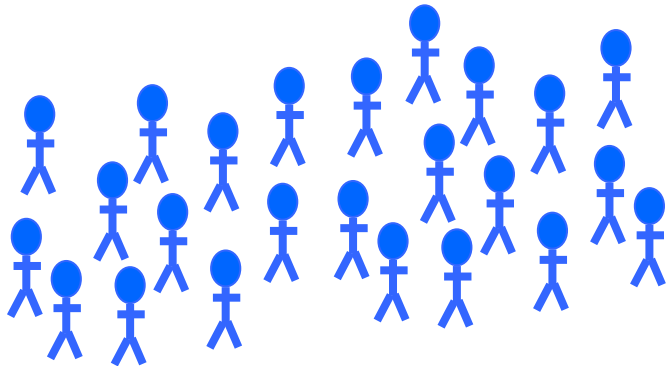
How many have gene variant?

Drawbacks

- Prone to confounding and other biases

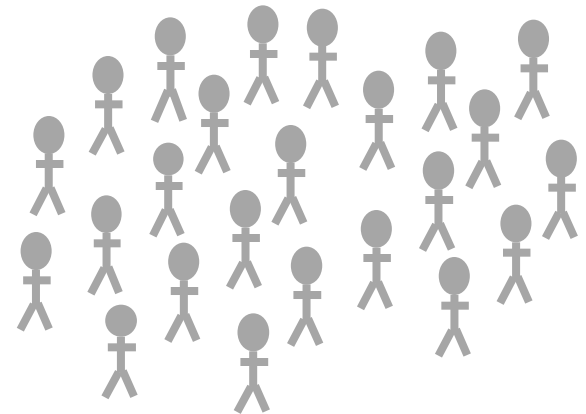
Confounding

Diabetic cases



✓ Genetic variant more prevalent

Non-diabetic controls



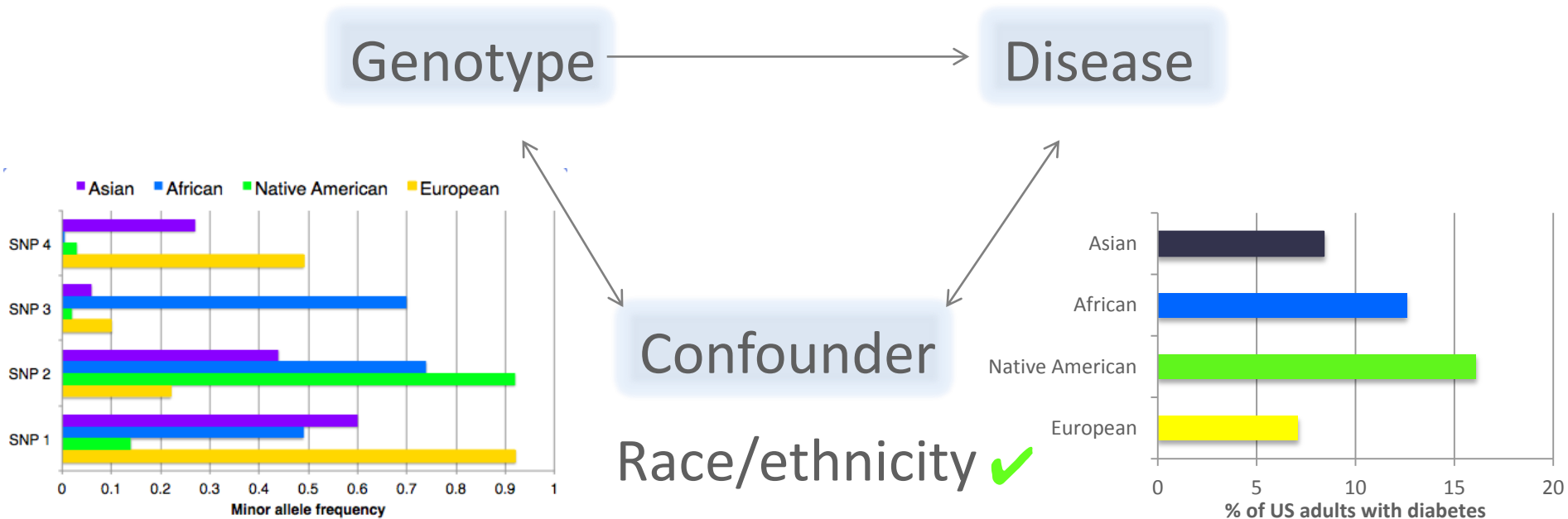
✓ Genetic variant less prevalent

How else might these two groups systematically differ?

Latino
Smoker
Obese

Non-Latino
Non-smoker
Non-obese

Race is a common confounder in GWAS (aka population stratification)

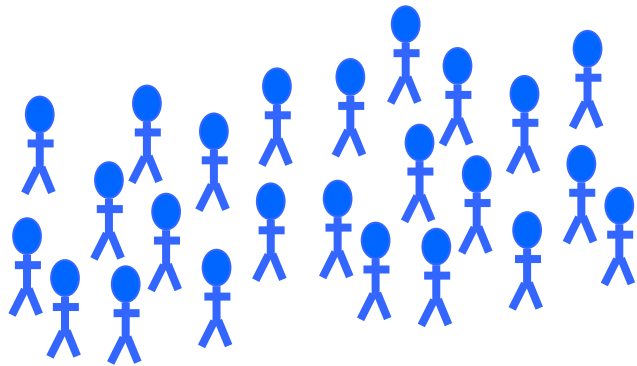


- *Can lead to false positive or false negative associations*
- *Must be controlled in design or analysis*

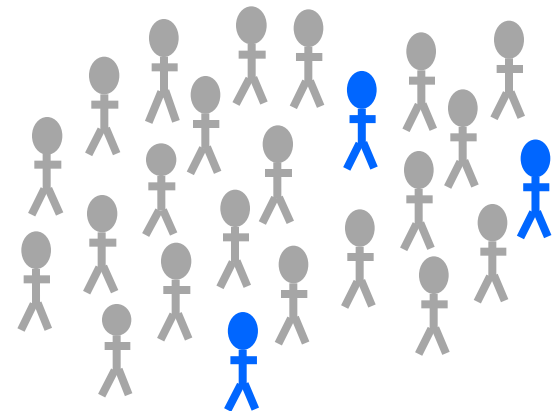
Misclassification bias

- Some cases erroneously classified as controls

Cases



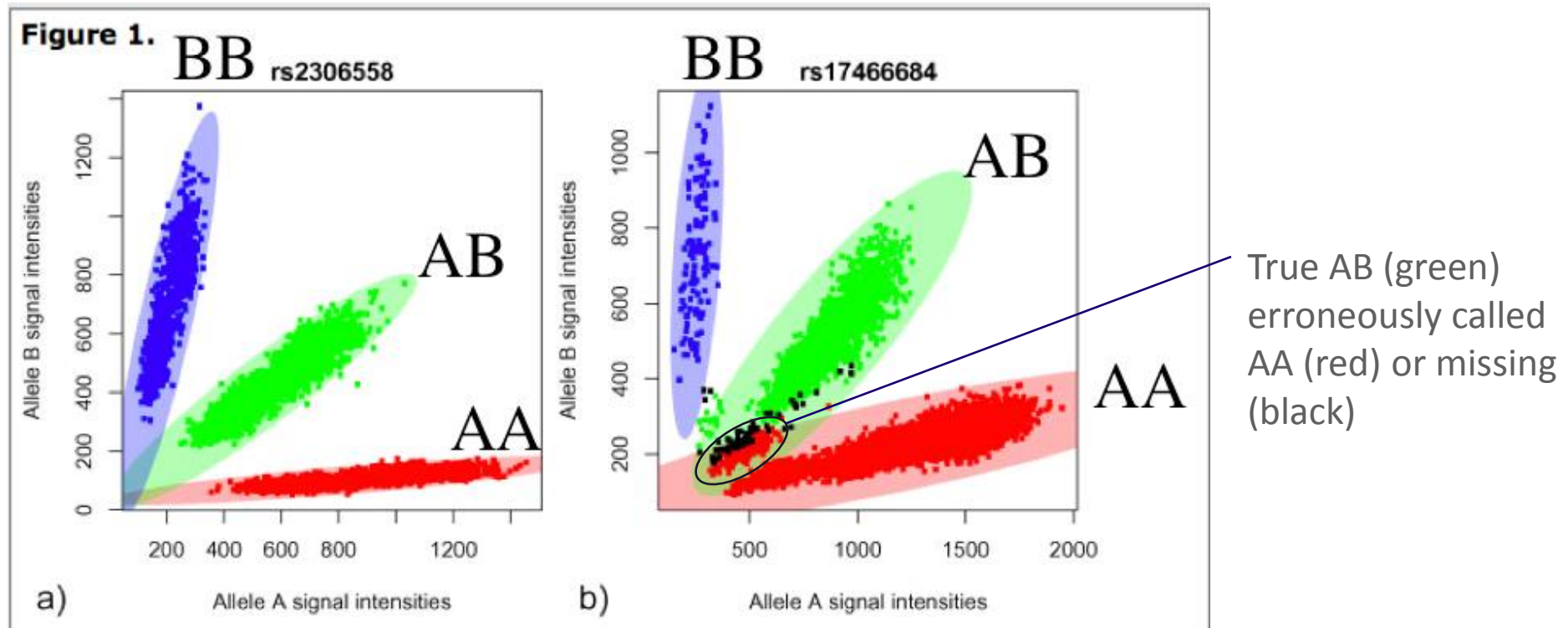
Controls



How does this happen?

Misclassification bias (cont'd)

Genotypes assigned incorrectly



Effect of misclassification bias

Randomly distributed

- E.g. misclassification of disease irrespective of genotype
- E.g. genotyping error equally as likely in cases and controls
- False negative (bias toward the null)

Differentially distributed

- E.g. misclassification of disease in one genotype vs another
- E.g. genotyping error occurs in controls but not cases
- False negative or false positive

Question

Which study design is more prone to confounding and bias?

- A. Case-control
- B. Cohort

Answer

A. Case-control

Case-control studies are more prone to confounding and bias than cohort studies because cases and controls are often difficult to match on important variables