Mapping complex traits in populations: Genome-wide association studies (GWAS)

QTL mapping in a cross/ pedigree

- Localize gene(s) affecting phenotypic difference between original strains/ people
  - Let's say you map a gene causing diabetes in one family to 18p22 (strong effect next to a marker)
  - You map in ***another*** family having diabetes, and you see no effect of 18p22 region
    - What happened?

Another approach: mapping in populations by association studies (e.g., GWAS)

- We discussed this last week for simple traits
  - **If a marker is very close to the disease-causing gene, individuals having one allele will be more likely to have the disease than individuals having the other allele**
- Same principle true even if many genes involved

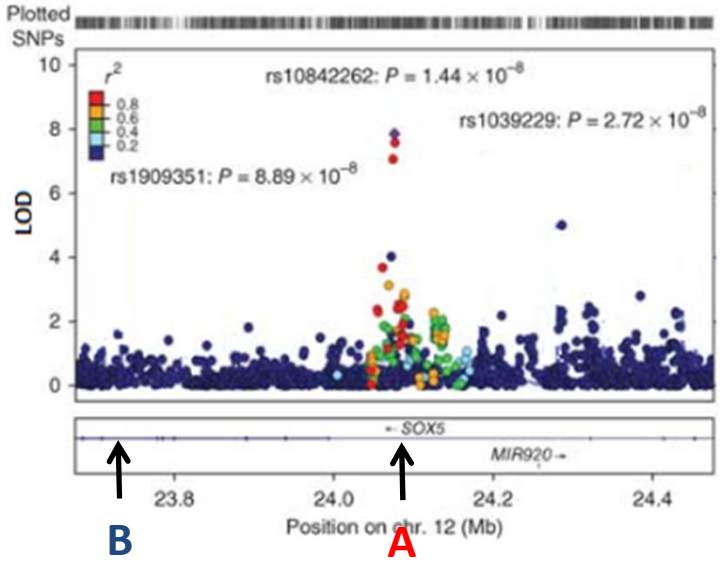
Test many, many, many markers

- Some have an allele associated with disease
  - AA: prob disease: 10%
  - aa: prob disease: 1%
  - Chi-square or other statistic says unlikely difference by chance
- Many have no association
  - BB: prob disease: 1.5%
  - bb: prob disease: 1.6%

Visualize with something like LOD plot

Test many, many, many markers

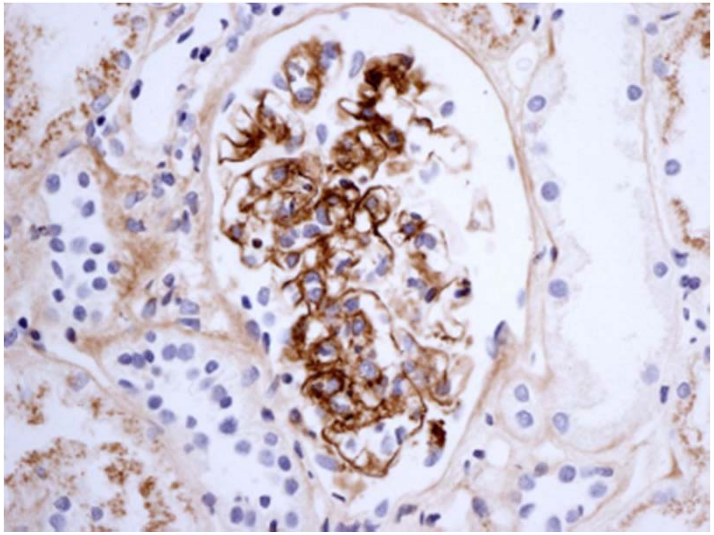
- Some have an allele associated with disease
  - AA: prob disease: 10%
  - aa: prob disease: 1%
  - Chi-square or other statistic says unlikely difference by chance
- Many have no association
  - BB: prob disease: 1.5%
  - bb: prob disease: 1.6%



From Feb 2012 *Nature Genetics*
Azoospermia in Chinese men

Often find not one but many possible factors

- From Feb 2012 issue of *Nature Genetics*:
  - **A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy (8p23, 17p13, 22q12).**



Problem: multiple comparisons

- From lab, chi square test gave probability of getting that association "by chance"
  - If you roll two dice and get snake-eyes twice in a row, are they loaded?
    - ~0.1% chance that would happen, but not ZERO

Problem: multiple comparisons

- From lab, chi square test gave probability of getting that association "by chance"
  - If you roll two dice and get snake-eyes twice in a row, are they loaded?
    - ~0.1% chance that would happen, but not ZERO

  - What if rolled dice 1,000,000 times- what is chance you'd get snake-eyes twice in a row SOMETIME in there?

Problem: multiple comparisons

- From lab, chi square test gave probability of getting that association "by chance"
  - If you roll two dice and get snake-eyes twice in a row, are they loaded?
    - ~0.1% chance that would happen, but not ZERO
  - What if rolled dice 1,000,000 times- what is chance you'd get snake-eyes twice in a row SOMETIME in there? >99.99%!

Problem: multiple comparisons

- Same problem when test 1,000,000 SNPs for association with disease of interest
  - Very likely to get an association that looks strong by chance but no real biology

- As a result, instead of using 5% probability as a cut-off, often use ~0.00001% probability

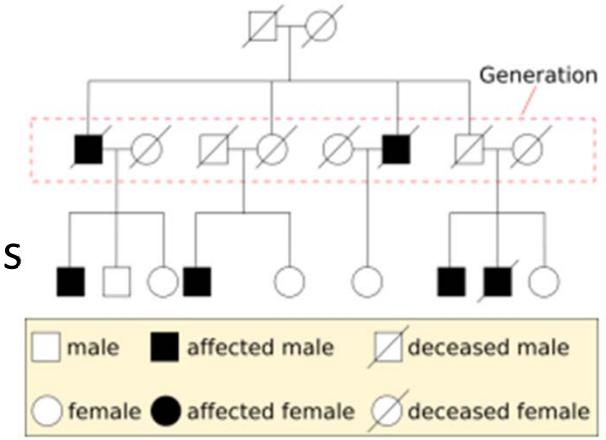
Problem: multiple comparisons

- Same problem when test 1,000,000 SNPs for association with disease of interest
  - Very likely to get an association that looks strong by chance but no real biology

- As a result, instead of using 5% probability as a cut-off, often use ~0.00001% probability
  - BUT this requires a VERY strong association…

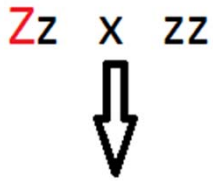
Pedigree/cross vs. Population

- Pedigree/ cross analyses find genetic factors causing differences ***in that family***.
  - Typically, high power for small number of genes
- Population association studies try to find genetic factors causing variation ***across the population***.
  - Typically, lower power but can find more genes and localizes more precisely



Why???

- In cross/ pedigree, limited to mapping genes with effects in that cross, BUT the alleles you're mapping are abundant
  - In simple cross, initially 25% or higher



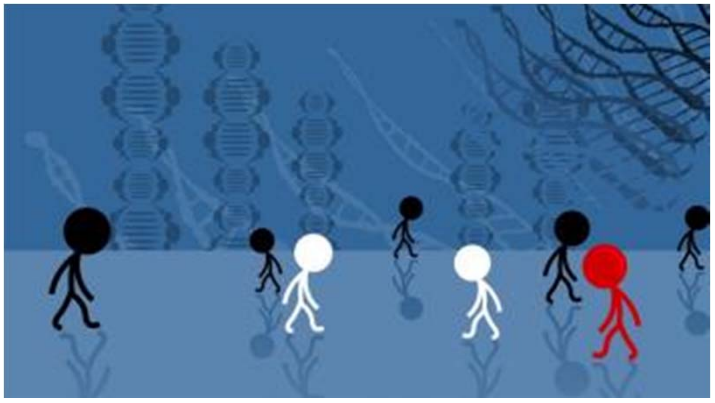
- In population, you can theoretically map alleles at any gene causing effect, BUT you may have *different* very rare variants causing the same effect/ disease

aazz aazz aazz aazz Aazz

aaZz aazz aazz aazz aazz

Genome-wide association studies in populations work for mapping "common disease variants"

- … but are most disease variants common?
- … how often do *different rare* mutations cause the same disease?



Also, can see differences among ethnic groups for associations

- Sometimes see a marker allele associated with disease in one ethnic group but not another

Type II Diabetes

- Four SNPs identified associated with Type II diabetes in East Asian subjects

- One of these four SNPs (on chr 11) had no detectable effect in studies of European populations

Image Credits, Unit 7-3

- Microarray, © 2006 Schutz, CC by-SA 3.0, en.wikipedia.org.
- Afghan child receiving vaccination, © 2010 UNICEF Sverige, CC by 2.0, www.fotopedia.com.
- LOD scatterplot, © 2012 Hu Z. et al and Nature Genetics, all rights reserved, http://www.nature.com/ng/journal/v44/n2/full/ng.1040.html, Nature Genetics 44, 183–186 (2012)
- Nephropathy kidney cells, © 2008 Karamadoukis et al, CC by 2.0, www.jmedicalcasereports.com
- Stick figures and DNA, © vishnukumar, all rights reserved, www.photoxpress.com
- Multicolor handprints, © Willee Cole, all rights reserved, www.photoxpress.com