One-minute responses

- The pedigree content seemed really slow, but I guess there are people without genetics backgrounds

Pedigree analysis: additional topics

- Twin studies: measuring penetrance
- Sib pair analysis
- Ascertainment issues

Twin studies

- Monozygotic (MZ or identical) twins share both genotype and early environment
- Dizygotic (DZ or fraternal) twins share early environment
- Greater disease phenotype similarity among MZ as opposed to DZ twins suggests a genetic component
- MZ concordance rate is an upper limit for genetic contribution
- Many important genetic diseases have MZ concordance around 50% or less
- Usefulness limited by availability of twins

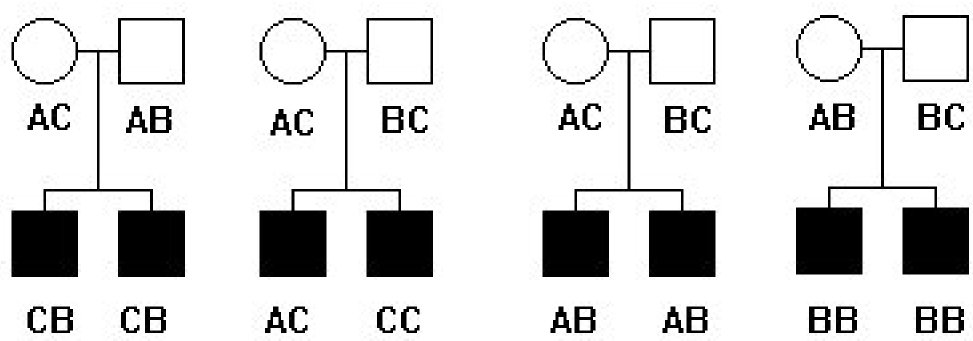
Sib pair analysis

- Pairs of affected siblings offer some information about a disease
- Easier to collect than parents plus children for some diseases
- Measure haplotype sharing (for a particular genetic region) between affected sibs:
  - Random–disease locus probably not in that region
  - Sibs tend to share both haplotypes–recessive disease
  - Sibs tend to share one haplotype–dominant disease

Sib pairs with complete linkage and penetrance

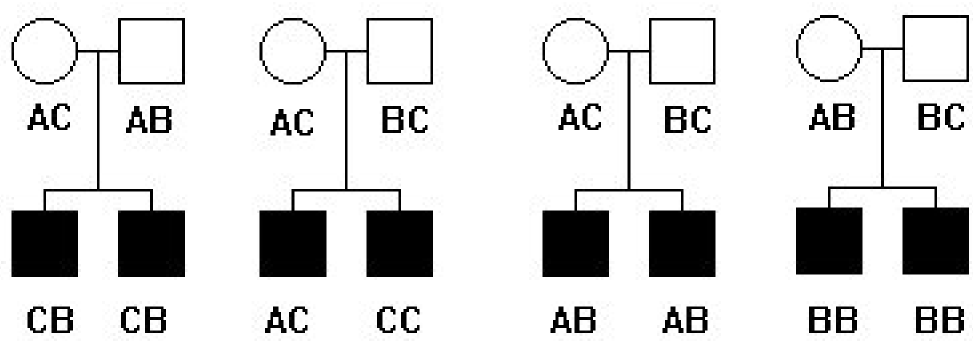
Condition	Share 2	Share 1	Share 0
No linkage	1/4	1/2	1/4
Dominant	1/2	1/2	0
Recessive	1	0	0

Four sib pairs



- Three pairs share both alleles
- One pair shares one allele
- This patterns suggests a nearby recessive gene

Four sib pairs



- Explanations for the second sib-pair:
  - Recombination
  - Gene is not completely recessive (heterozygote sometimes affected)
  - One parent has two copies of the recessive gene
- Four pairs aren't enough
- Childhood diabetes originally mapped with 17 sib pairs

Sib pair extensions

- Some information can also be found in affected/unaffected pairs
- Sib trios are better than pairs, but rarer
- More complex mathematics to handle missing parent data, ambiguous sharing

Missing parent information

- IBD–identical by (recent) descent
- IBS–identical by state
- With parent information, we can often see how many haplotypes siblings share IBD
- Without parents, we can often only score IBS

Missing parent information: example

- In some pairs, siblings share no haplotypes IBS
- These clearly share none IBD
- In others, they share 1 IBS: one child AB, the other AC
- These could share 0 or 1 IBD
  - If one parent was AA, the children's two A's might be unrelated
  - If one parent was AB and the other was AC, again the children's two A's might be unrelated

Missing parent information: example

- If A is very common, the chance of sharing 0 is higher
- How to estimate frequency of A?
- Cannot use all members of families, as they are related
- Two approaches:
  - Use population frequencies (requires sib pairs all to come from the same well-defined population)
  - Use frequencies taken from one randomly chosen sib in each family (requires a lot of sib pairs to get good estimates)
- Once an estimate of frequencies is available, a pair which shares 1 IBS can be weighted appropriately for its chance of sharing 0 or 1 IBD
- Similar calculations apply to a pair which shares 2 IBS

Ascertainment

- Many statistical methods assume that the families were chosen at random
- This is nearly impossible
- Non-random choices can lead to bias in the statistics if not taken into account
- Example:
  - We have a list of all patients in Seattle with a particular genetic disease
  - We call each one and ask them to enroll their family in our study
  - Do we have a random sample of families with at least one patient?

Ascertainment

- Do we have a random sample of families with at least one patient? No!
- Families with 4 patients were called 4 times
- This probably increases the chance they will respond
- Also, families with more patients tend to be more interested in studies
- If we use these data to measure the tendency of our disease to cluster in families, we will over-estimate it

Ascertainment bias

- This is an example of ascertainment bias
- It can be a severe problem for some studies
- Example:
  - In families with a recessive disease, on average 1/4 of children are affected
  - In families with a dominant disease, on average 1/2 of children are affected
  - Over-sampling families with more affected children can lead to mistaking a recessive disease for a dominant one

Example

- Original MZ twin studies for insulin-dependent diabetes suggested concordance of 50%
- Later, more careful studies found 20-30%
- Concordant pairs probably overrepresented in original study
- Two diabetics are more likely to become involved than one
- Families where both twins are affected are more likely to think the disease is genetic

Three-pronged strategy against ascertainment bias

- Try to reduce it
  - In our example, we might call only one patient per family, no matter how many there are
- Document it thoroughly
  - Keep records of how many patients in each family
  - Keep records of which families agreed to be in the study
- Compensate for it statistically
  - We can weight each family by its chance of being ascertained
  - If a 4-patient family is twice as likely to be ascertained as a 1-patient family, we can weight it 1/4 as heavily
  - These adjustments can become very sophisticated

Summary

- Particular family configurations (sib pairs, twins) can be mined for mapping
- Essential to know how the data were collected
- Data collected in a biased fashion lead to biased estimates
- If the data collection method is fully understood, a correction may be possible
- Calling colleagues and saying "Do you have any interesting families with disease X?" can produce a data set unsuitable for statistical analysis