

Introduction to linkage and family based designs to study the genetic epidemiology of complex traits

Harold Snieder



Overview of presentation

- Designs: population vs. family based
- Mendelian vs. complex diseases/traits
- Linkage analysis: model based vs. model free
 - Model based: recombination fraction & LOD score
 - Model free: Identity by descent (IBD) & Haseman-Elston regression
- Genome scans for complex traits
 - Lipoprotein(a) example
- Candidate gene approach
 - Population based or within family association tests
- Linkage versus Association

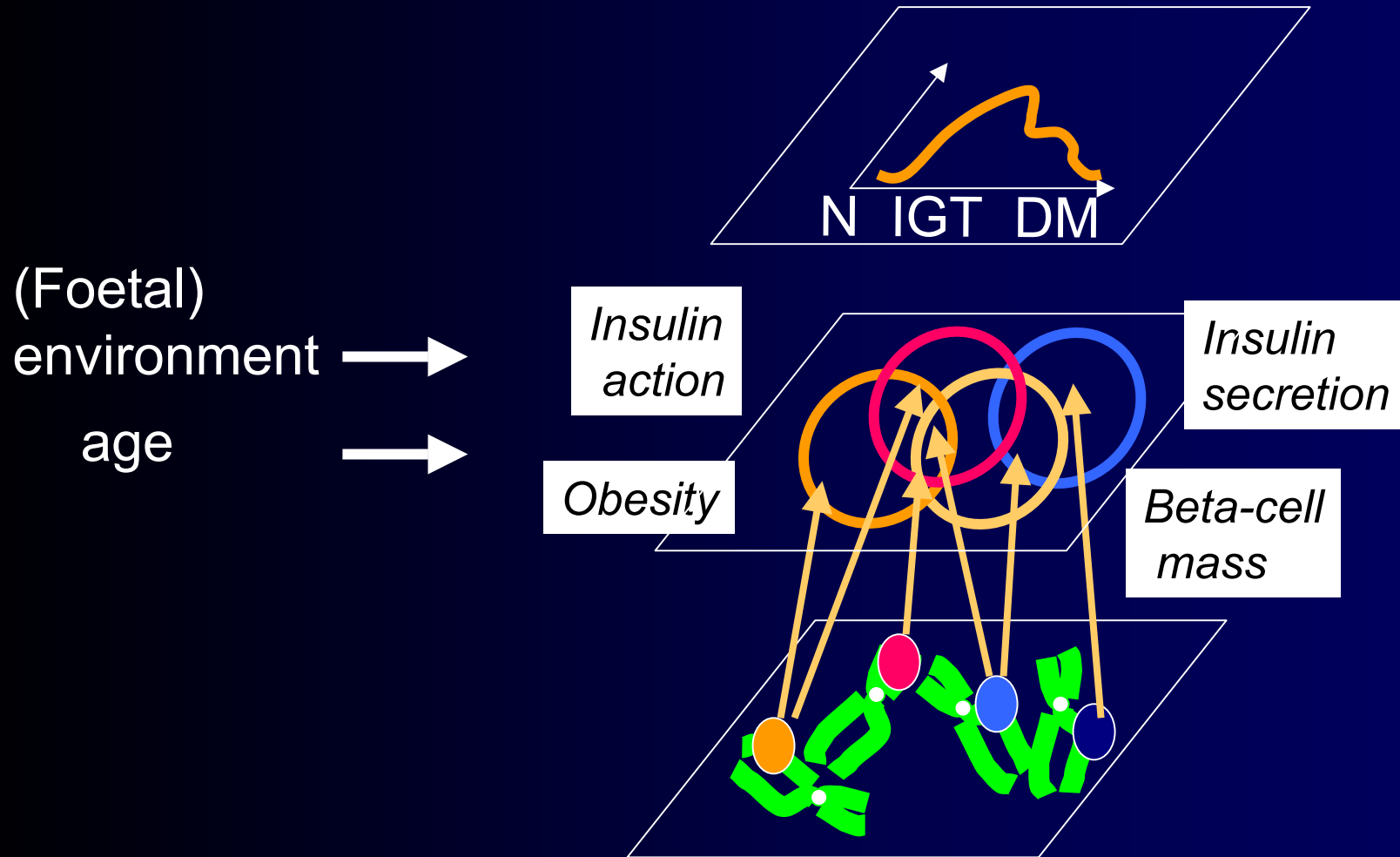
Genetic epidemiological study designs

	Population-based	Family-based
Unmeasured effects	Inbreeding studies Admixture studies	Twin/family studies Segregation analysis Complex segregation
Measured genes	Mutations Allelic association	Linkage Association with family controls (TDT) Combined linkage and association

Localizing genes for Mendelian (single-gene) disease

- Familial aggregation?
 - Family studies: Large pedigrees
- How is the disease transmitted in families (mode of inheritance)?
 - Segregation Analysis
- Do genetic markers co-segregate with disease in pedigrees? Where is the gene located?
 - Classic linkage analysis:
 - Model based: mode of inheritance, penetrance & gene frequencies

Type 2 Diabetes: a complex disease



Challenges in genetics of complex disease

- Identifying genes of small relative effect against a background of substantial variation
- Definition and measurement of phenotype
- Age-related disease/trait expression
- Gene-environment interaction
- Heterogeneity (eg Hypertension)

Localizing genes for complex traits & diseases

- Familial aggregation? Genetic or environmental factors?
 - Twin study
 - Adoption Study
- Is a major gene involved?
 - Complex segregation analysis in pedigrees (*transmission, penetrance, gene frequencies*)
- Where is the gene located?
 - Linkage - eg Genome scan
 - Association - candidate gene approach

Linkage analysis

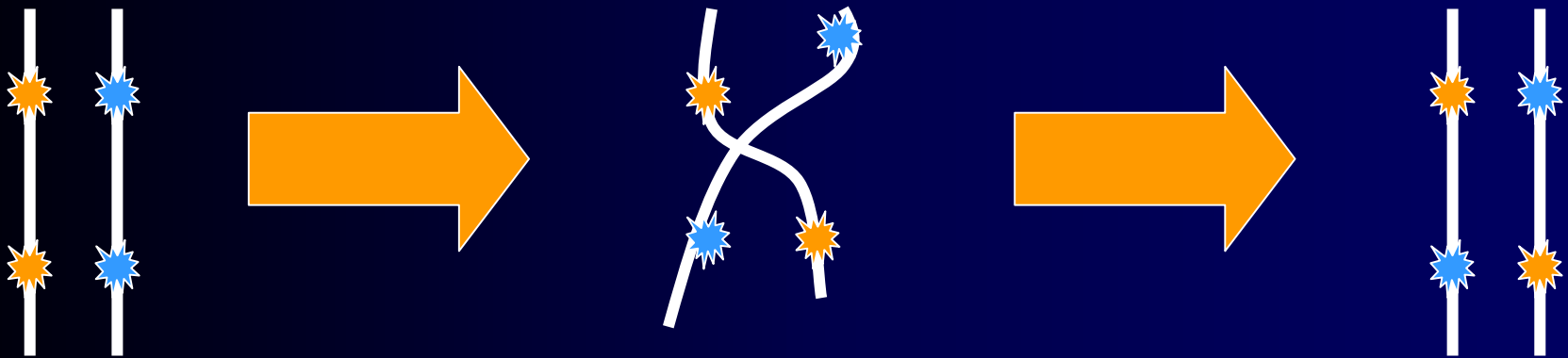
- Tracks co-segregation of genetic markers with diseases/traits in families
- Model based linkage
 - based on information provided by complex segregation analysis
 - limited utility in complex traits (inheritance model unknown)
- Model free linkage
 - family or sib pair design (affected or quantitative traits)
 - Sharing of alleles identical by descent (IBD)
 - No inheritance model needed

Model based linkage: recombination

- Mendel thought that any two different genes were always inherited independently
- This is not true when two genetic loci are physically close together on the genome
- Alleles at loci on different chromosomes are unlinked and have a 50:50 chance of being inherited together (assuming independence).
- The probability of recombination or recombination fraction (denoted θ [theta]) is said to be 0.5 in this case (no linkage)

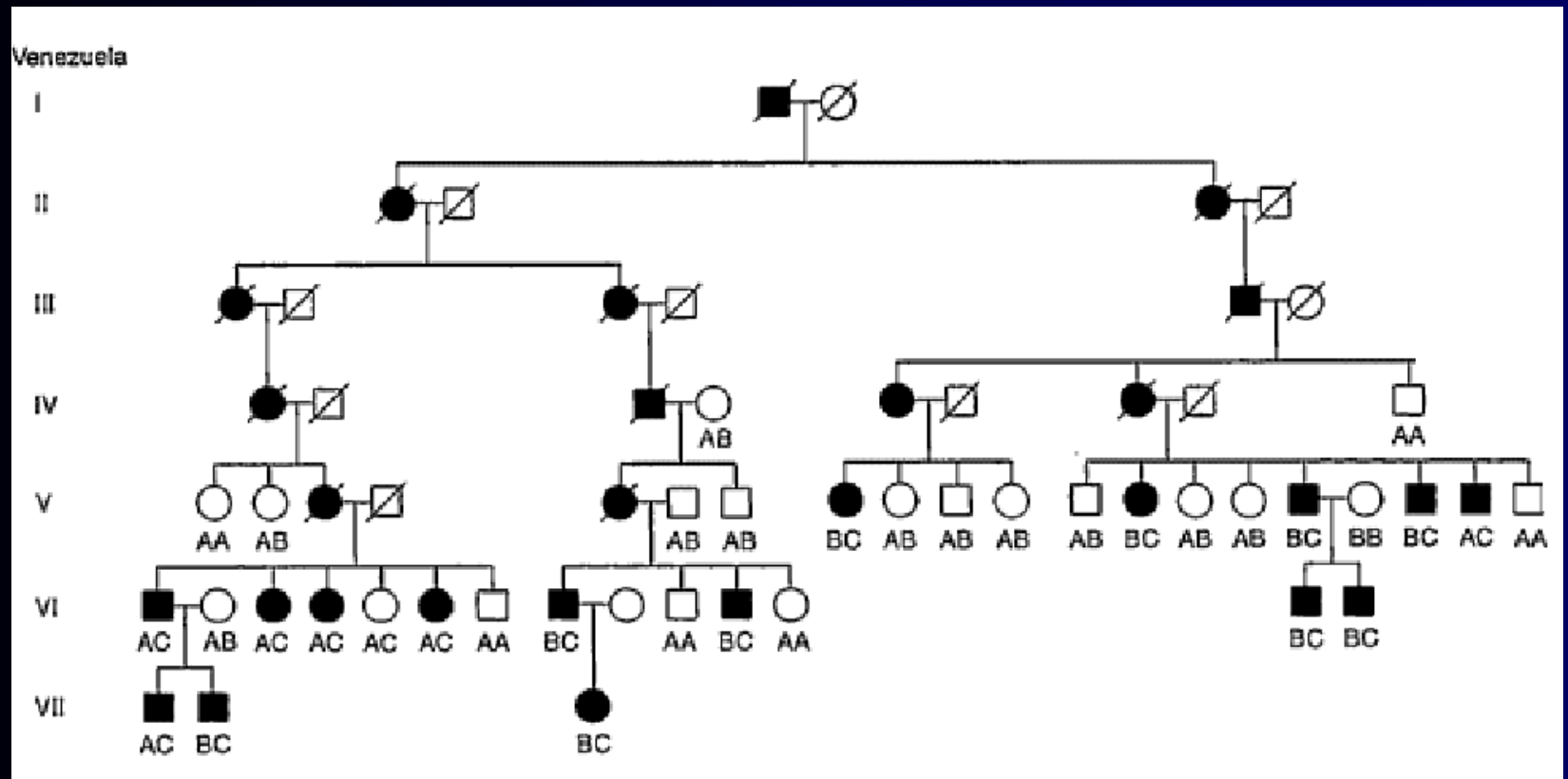
Recombination and linkage

- Recombinations occur due to crossovers at meioses:



- If few patients in family pedigrees show recombination between genetic marker and disease gene, the recombination fraction θ is low (< 0.5) and linkage tight.
- Recombination fraction is also a measure of distance between marker and disease gene (gene mapping)
- Two loci on the same chromosomes are linked if: $\theta < 0.5$, unlinked if $\theta = 0.5$.

Pedigree of Huntington Disease



LOD Score

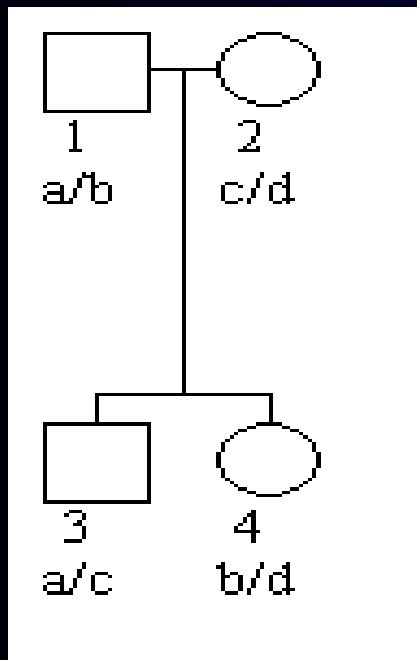
- Usually, we wish to compare the evidence for linkage against the null hypothesis that two loci are unlinked.
- This can be expressed in a Lod Score:
 - $LOD = \log_{10}[(P(\text{data}|\theta)/P(\text{data}|0.50))]$
 - X^2 distributed (1 df): $LOD=3$ corresponds to $X^2 = 13.8$ ($p\text{-value}_{\text{one-sided}}=0.0001$).

Model free linkage

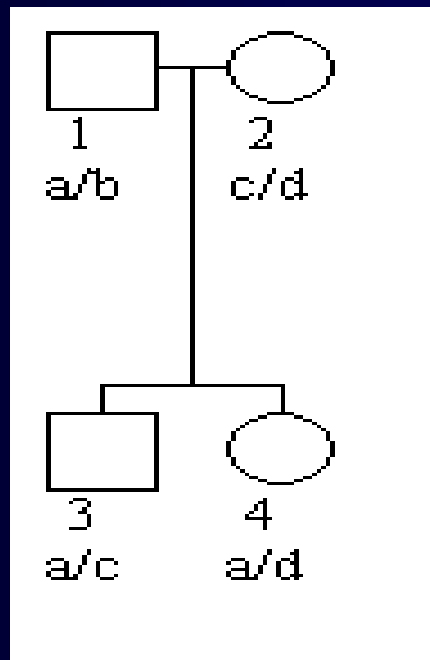
- Affected sib-pairs
 - mapping of disease genes (eg hypertension)
- Quantitative traits in families or sib-pairs
 - mapping of Quantitative Trait Loci or QTLs (eg blood pressure)
- Does similarity in phenotypic variation or disease status correlate with sharing of genetic material?
 - Based on sharing of alleles identical by descent (IBD) at different marker locations along the chromosome

Identity by descent

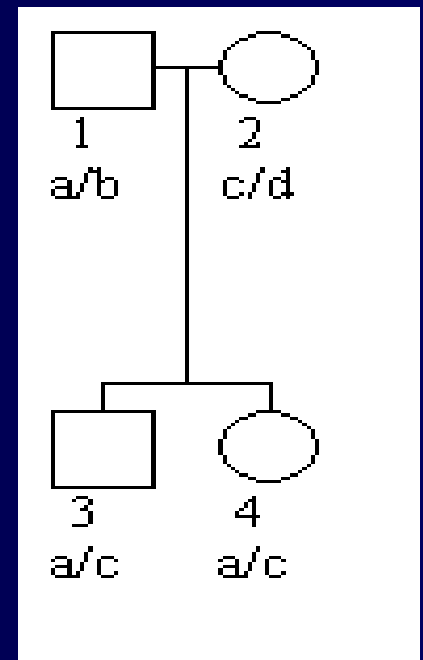
- Two alleles that are copies of the same ancestral gene are said to be identical by descent (IBD) as opposed to just Identical by State (IBS).
- Siblings (3 and 4) may share 0, 1 or 2 alleles IBD:



0 IBD



1 IBD



2 IBD

Identity by descent

- Siblings are expected to share IBD in defined patterns:

IBD status	Genetic Correlation (π)	Expected %
0	0	25%
1	0.5	50%
2	1	25%

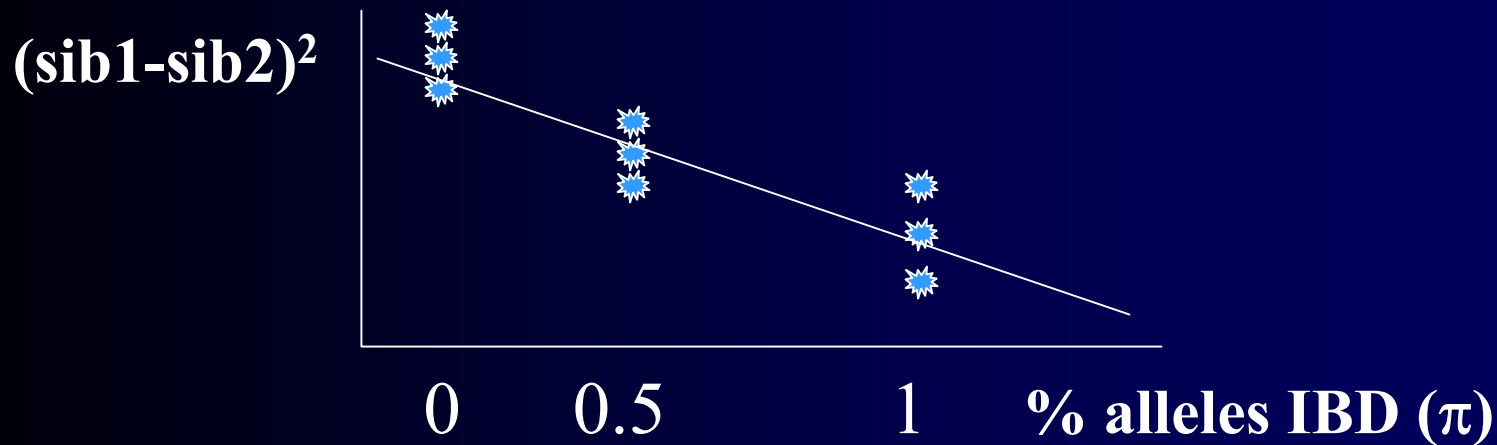
- Linkage exists if eg affected siblings show greater allele sharing IBD at a closely linked marker locus than the expected 50% (ie greater than expected genetic similarity)

QTL linkage in sib pairs

- The first proposed method of quantitative-trait linkage analysis for sib-pair data was the **Haseman-Elston** method.
- This method is based on regression of the squared trait difference of a sib pair, on the estimated proportion of alleles shared identical-by-descent (IBD) at a putative marker locus

Haseman – Elston regression

- For example for blood pressure:
 - $[BP_{sib1} - BP_{sib2}]^2 = \text{intercept} + \beta * \text{IBD sharing}$
- In case of linkage $\beta < 0$ ($\beta = -2 V_{QTL}(1 - \theta)^2$)



- This is what you expect under linkage. The higher the probability of IBD among pairs, the smaller the difference in phenotype.

Genome scan for complex traits

- Perform genome scan for complex quantitative trait (QTL mapping)
- Some significant linkage peaks (eg LOD > 3) are found on some chromosomes
- The **Problem**:
 - Even if these peaks are real they cover a large area (20cM) with sometimes hundreds of candidate genes
 - Fine mapping is **VERY** difficult
 - Only about 5 genes (!!!) for complex traits & diseases have been found this way
- Conclusion: Don't even think about it!

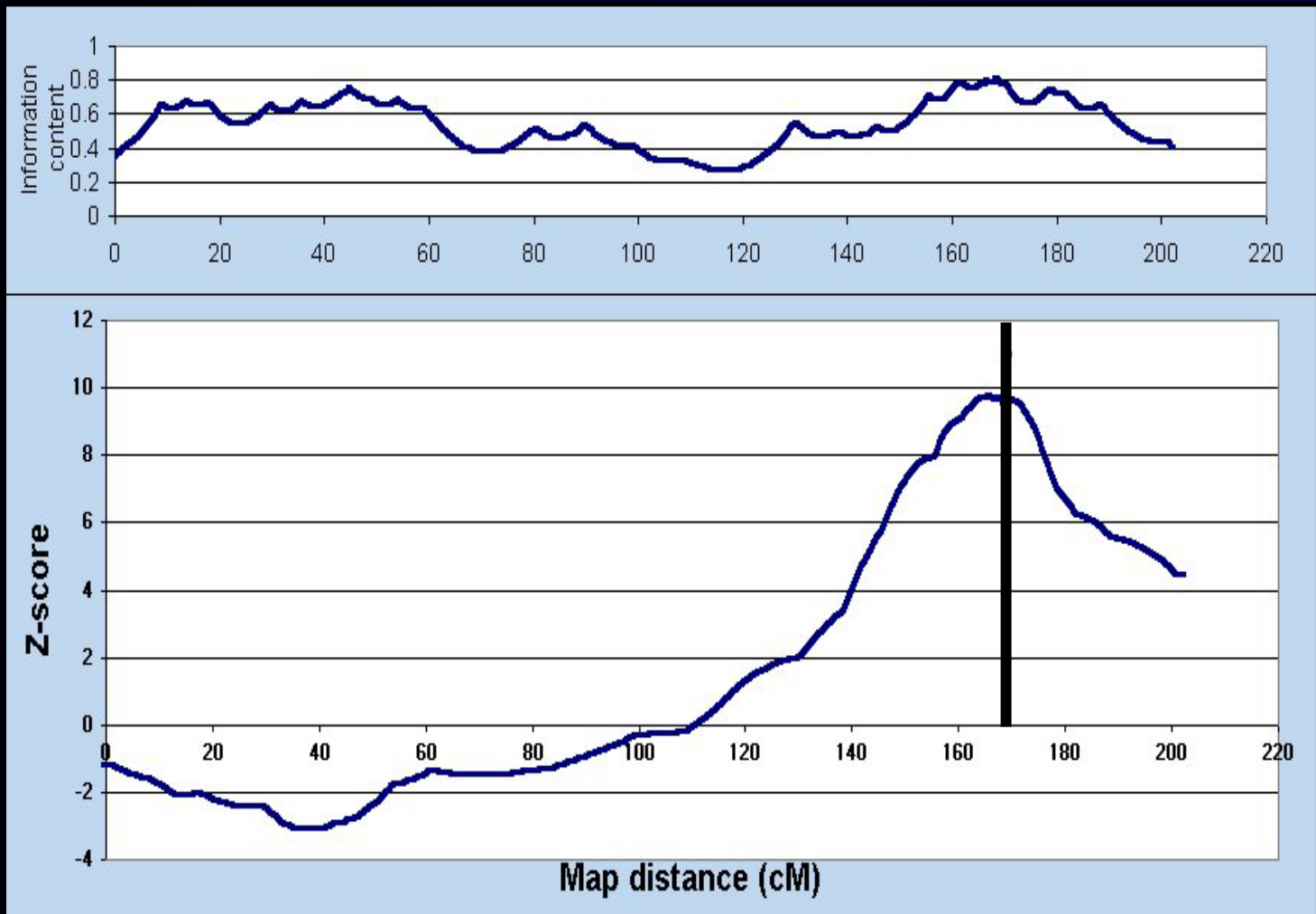
Genome scan for lipoprotein(a)

- 413 MZ & 806 DZ female twin pairs ($h^2 = .88$)
- genome scan of 806 DZ pairs
- non parametric (Z score) multipoint analysis using MAPMAKER/SIBS
- Genome scan of quantitative trait in unselected sib- (or DZ twin) pair population: Does it work?

Lipoprotein(a)

- Lipoprotein(a) [Lp(a)] ideal trait for proof of principle:
 - quantitative risk factor for CHD
 - no confounders
 - heritability $\approx 90\%$
 - the gene [apo(a)] on Chr.6 is known and explains 90%

MapMaker plot: Chr.6 Lp(a)



Candidate gene approach:

	Population-based	Family-based
Unmeasured effects	Inbreeding studies Admixture studies	Twin/family studies Segregation analysis Complex segregation
Measured genes	Mutations Allelic association	Linkage Association with family controls (TDT) Combined linkage and association

Association studies: family based?

Four possible explanations for a significant association between a gene and disease or quantitative trait:

- population stratification/admixture
 - ethnic confounding, use **within family controls**:
Transmission Disequilibrium Test (TDT)
- causal
 - mutation is functional
- linkage disequilibrium
 - physically close to functional locus on same chromosome
- chance

Linkage versus Association

- **Linkage Analysis**

- Follows meiotic events through families for co-segregation of disease and particular genetic variants
 - Large Families
 - Sibling Pairs (or other family pairs)
 - Works **VERY** well for ‘Mendelian’ diseases

- **Association Studies**

- Detect association between genetic variants and disease across families: exploits linkage disequilibrium
 - Case-Control designs
 - Cohort designs
- **May be more appropriate for complex traits**