IBS Methods for Affected Pairs Linkage

Biostatistics 666

Lecture 14

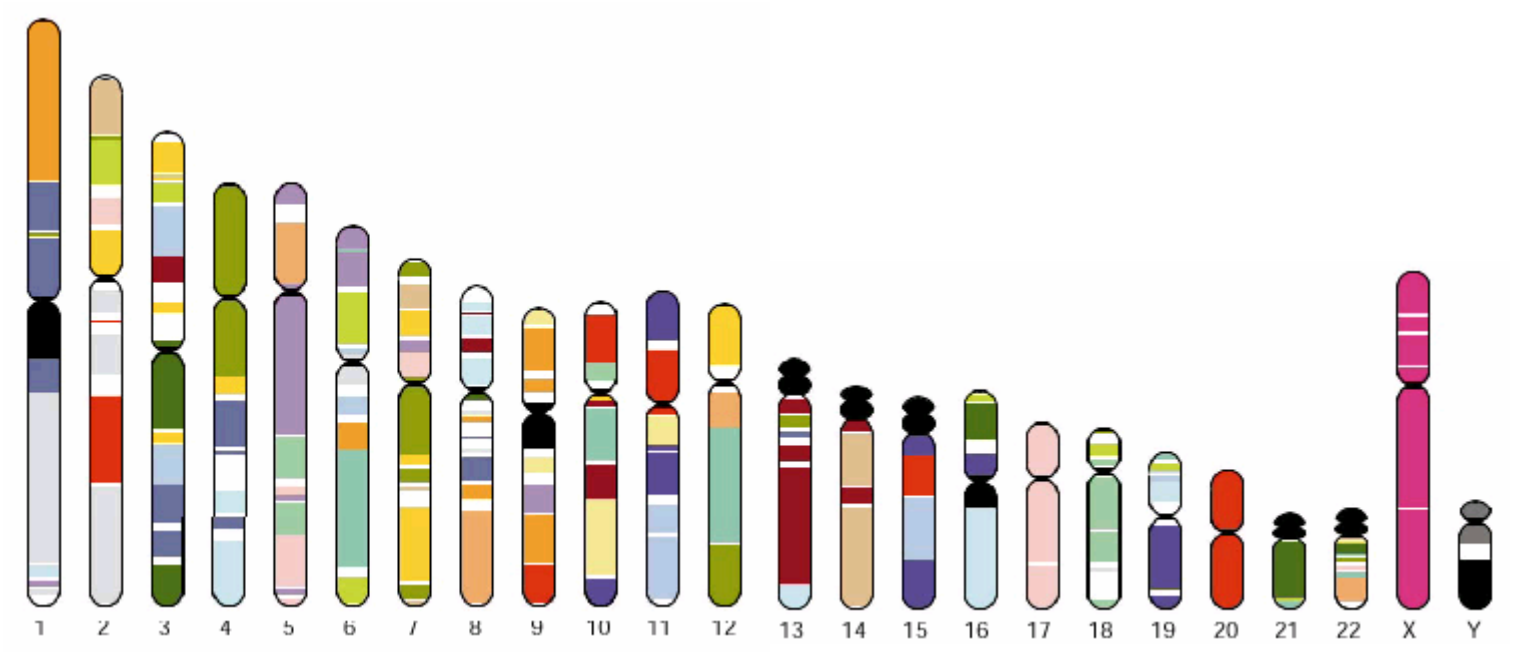
Genetic Mapping

"Compares the inheritance pattern of a trait with the inheritance pattern of chromosomal regions"

Positional Cloning

"Allows one to find where a gene is, without knowing what it is."

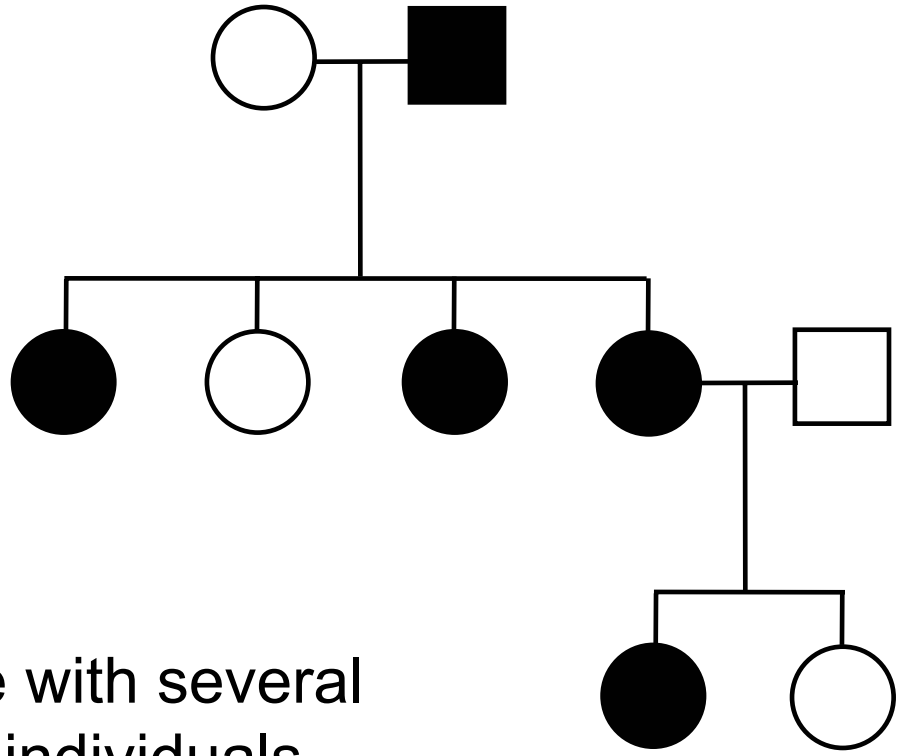
Where are the genes influencing a particular trait?



Intuition for Linkage Analysis

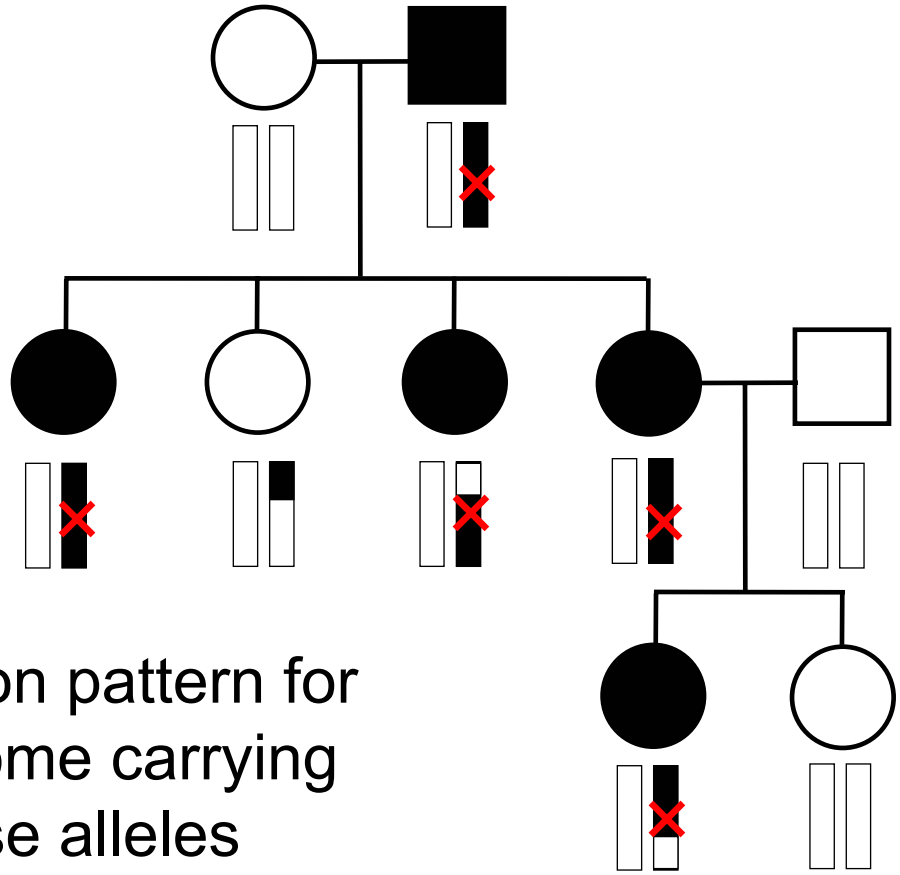
- Millions of variations could potentially be involved
  - Costly to investigate each individually
- Within families, variation is organized into a limited number of haplotypes
  - Sample modest number of markers to determine whether each stretch of chromosome is shared

Tracing Chromosomes



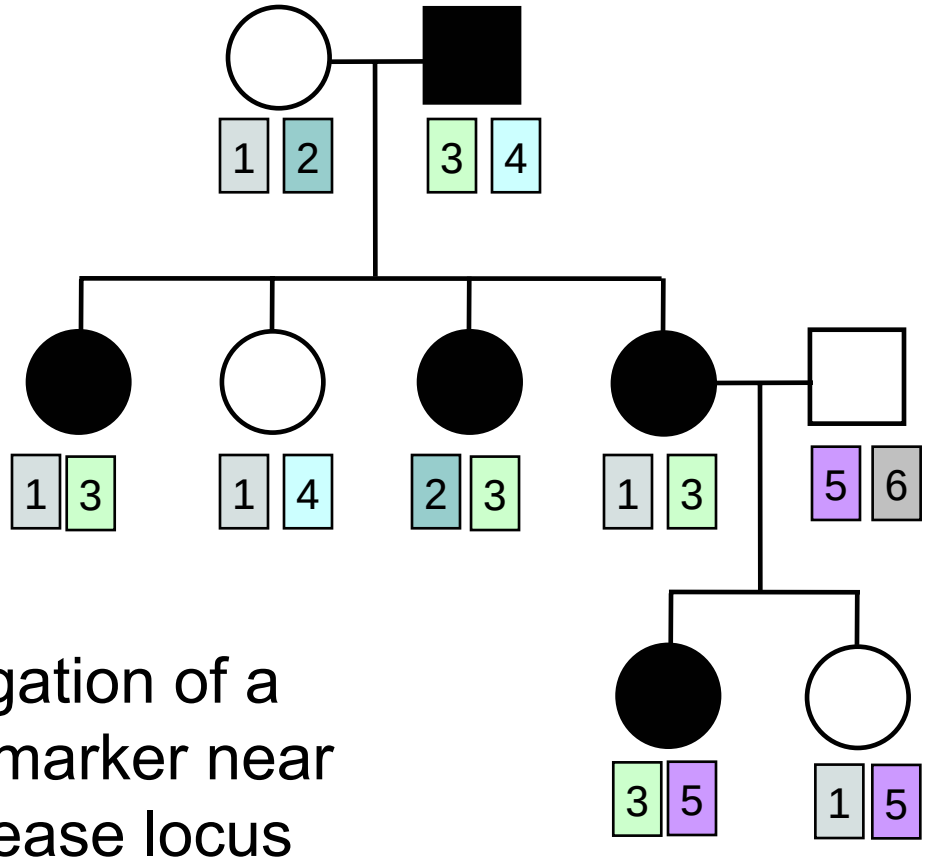
A pedigree with several affected individuals

Tracing Chromosomes



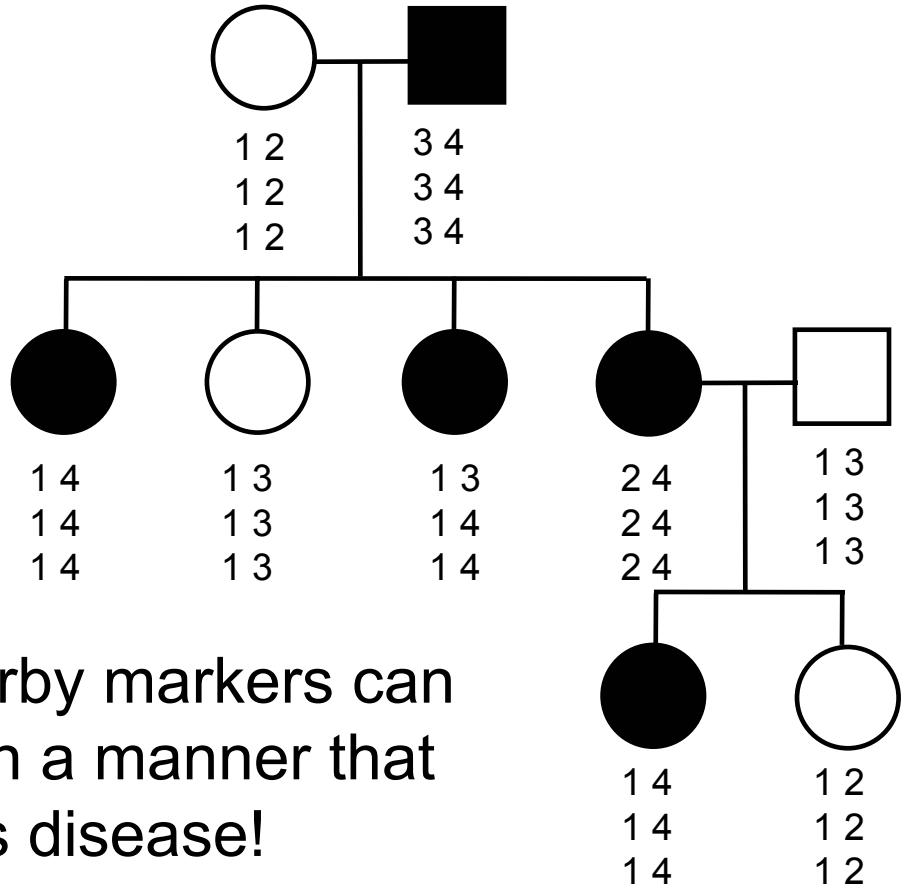
Segregation pattern for chromosome carrying disease alleles

Tracing Chromosomes



Segregation of a specific marker near the disease locus

Tracing Chromosomes



Multiple nearby markers can segregate in a manner that tracks disease!

Today …

- Linkage analysis with sibling pairs
  - Identity-by-State (IBS) based method
- Find markers that are near disease locus
  - Near means recombination fraction $\theta < \frac{1}{2}$
- Minimalist approach …

Reference for Today...

- Power of IBS Methods for Linkage Analysis

- Bishop DT and Williamson JA (1990) *Am J Hum Genet* **46:**254-265

- Recommended Reading

Bishop and Williamson (1990) Opening Line

"The availability of a large number of DNA markers has made possible mapping projects with the certainty that if:

(a) a major gene exists for a trait;

(b) the trait is reasonably homogeneous;

(c) there is sufficient family material available;

then a linked marker can be found."

Data for a Linkage Study

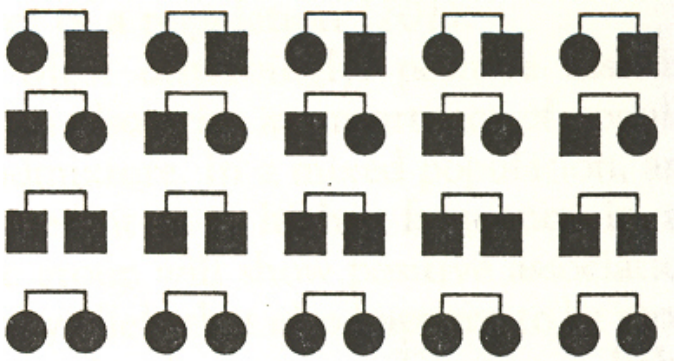
- Pedigree
  - Set of individuals of known relationship
- Observed marker genotypes
  - SNPs, VNTRs, microsatellites
- Phenotype data for individuals

Minimalist Approach

- Pedigree
  - Two individuals of known relationship
- Observed Marker Genotypes
  - A single marker
- Phenotypes
  - Both individuals are affected

Allele Sharing Analysis

- Reject random sharing at a particular region
- Less powerful than classic methods
  - When disease model is known
- More robust than classic methods
  - When disease model is unknown



Consider Autosomal Recessive Locus …

- For a collection of sibling pairs…
- What patterns of sharing do you expect at the disease locus?
- What patterns of sharing to you expect as you move away from the disease locus?

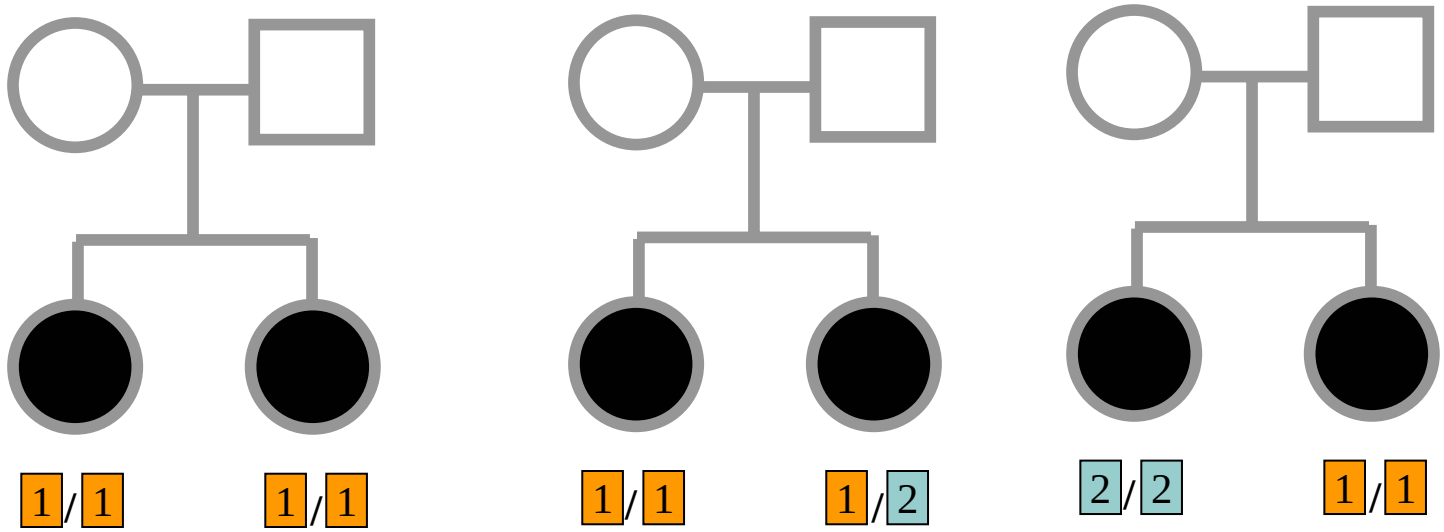
Historical References

- Penrose (1953) suggested comparing IBD distributions for affected siblings.
  - Possible for highly informative markers (eg. HLA)
- Thomson (1986) suggested discarding partially informative families.
- Lange (1986) proposed using IBS information instead of IBD.

IBS Based Methods

- Sample of affected relative pairs
- Examine a marker of interest
- Count alleles shared for each pair
  - This includes both …
  - Chromosomes that are identical-by-descent
  - Chromosomes that simply carry identical alleles

Examples of IBS States

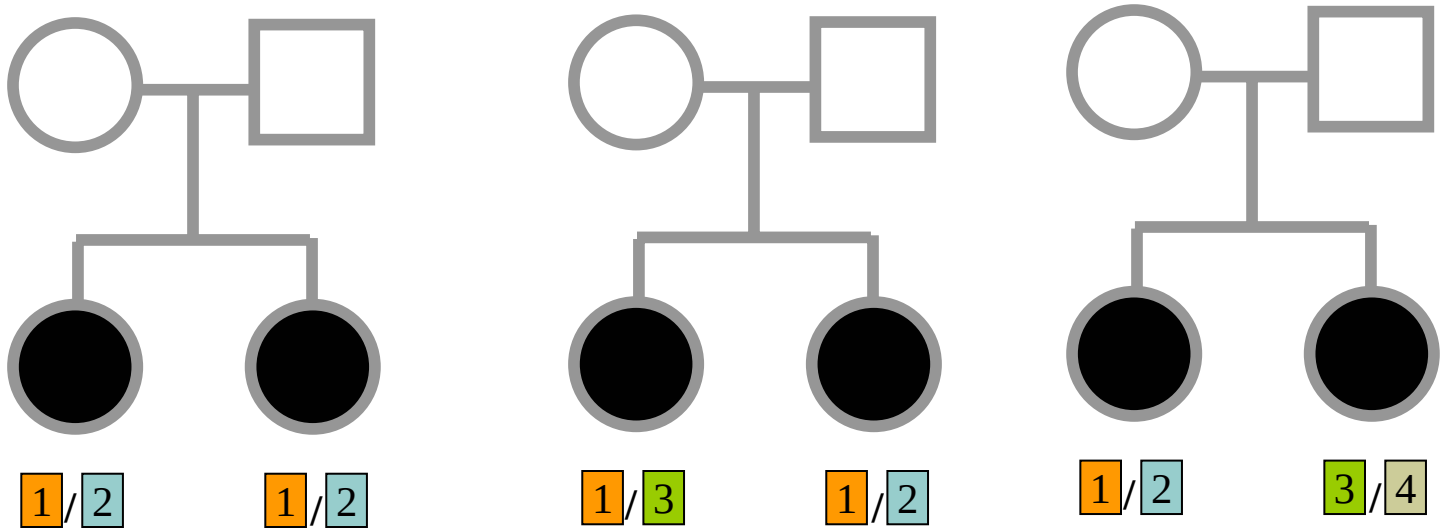


IBS = 2

IBS = 1

IBS = 0

Examples of IBS States



IBS = 2 IBS = 1 IBS = 0

Evidence for Linkage

- Increased similarity in affected pairs

- Compared to:
  - Unselected pairs
  - Unaffected pairs
  - Discordant pairs
  - Expectations derived from allele frequencies

Test for Independence

$$\chi^2_{2df} = \sum_i \frac{[N_{IBS=i} - E(N_{IBS=i})]^2}{E(N_{IBS=i})}$$

(general test, for sibling pairs)

$$\chi^2_{1df} = \frac{[N_{IBS=0} - E(N_{IBS=0})]^2}{E(N_{IBS=0})} + \frac{[N_{IBS>0} - E(N_{IBS>0})]^2}{E(N_{IBS>0})}$$

(grouping often preferable for other relatives)

- Assuming all counts are relatively large
- If counts are small, use binomial or trinomial distribution

Modeling IBS Sharing

- For any relative pair, calculate:
  - Probability of IBD sharing
    - 0, 1 or 2 alleles
  - Conditional probability of IBS sharing
    - 0, 1, 2 alleles
  - IBS sharing >= IBD sharing
    - Why?

IBD

- The underlying sharing of chromosomes segregating within a family
- Siblings share 0, 1 or 2 alleles
  - Probabilities ¼, ½ and ¼
- Unilineal relatives share 0 or 1 alleles
  - Probability of sharing is kinship coefficient ϕ * 4

P(Marker Genotype|IBD State)

Relative		IBD		
I	**II**	0	1	2
(a,b)	(c,d)	$4p_ap_bp_cp_d$	0	0
(a,a)	(b,c)	$2p_a^2p_bp_c$	0	0
(a,a)	(b,b)	$p_a^2p_b^2$	0	0
(a,b)	(a,c)	$4p_a^2p_bp_c$	$p_ap_bp_c$	0
(a,a)	(a,b)	$2p_a^3p_b$	$p_a^2p_b$	0
(a,b)	(a,b)	$4p_a^2p_b^2$	$(p_ap_b^2+p_a^2p_b)$	$2p_ap_b$
(a,a)	(a,a)	p_a^4	p_a^3	p_a^2
Prior Probability		¼	½	¼

Note: Assuming alleles unordered within genotypes

P(IBS = i | IBD = j)

$$P(IBS = 2 \mid IBD = 0) = 2\sum_{i \neq j} p_i^2 p_j^2 + \sum_i p_i^4$$

$$P(IBS = 1 \mid IBD = 0) = 4\sum_{i \neq j} p_i^2 p_j (1 - p_i - p_j) + 4\sum_i p_i^3 (1 - p_i)$$

$$P(IBS = 0 \mid IBD = 0) = \sum_{i \neq j} p_i p_j (1 - p_i - p_j)^2 + \sum_i p_i^2 (1 - p_i)^2$$

P(IBS = i | IBD = j)

$$P(IBS = 2 \mid IBD = 2) = 1$$

$$P(IBS = 1 \mid IBD = 2) = 0$$

$$P(IBS = 0 \mid IBD = 2) = 0$$

$$P(IBS = 2 \mid IBD = 1) = \sum_i p_i^2$$

$$P(IBS = 1 \mid IBD = 1) = \sum_{i \neq j} p_i p_j$$

$$P(IBS = 0 \mid IBD = 1) = 0$$

Example, Assuming Equal Allele Frequencies

	P(IBS=0)	P(IBS=1)	P(IBS=2)
2 alleles, IBD=0	.125	.500	.375
2 alleles, IBD=1	.000	.500	.500
3 alleles, IBD=0	.222	.592	.185
3 alleles, IBD=1	.000	.666	.333

Inference from Example

- IBS approaches IBD as number of alleles increases
- If linkage is being tested with chi-square test, how does the number of alleles (and marker informativeness) affect these two tests:
  - A test of whether $N_{IBS >= 1}$ increases?
  - A test of whether $N_{IBS > 1}$ increases?

IBS Probabilities

No. of Alleles	P(IBS=0)	P(IBS=1)	P(IBS=2)
2	.03	.37	.60
3	.05	.48	.47
4	.08	.51	.40
20	.21	.52	.27
∞	.25	.50	.25

Sibling IBS as a function of allele count, for marker with equally frequent alleles

Results of Bishop and Williamson (1990)

- Effect size, P(IBS | Affected pair)
- Number of alleles at marker
- Different relationships
- Recombination fraction

More Alleles Increase Power

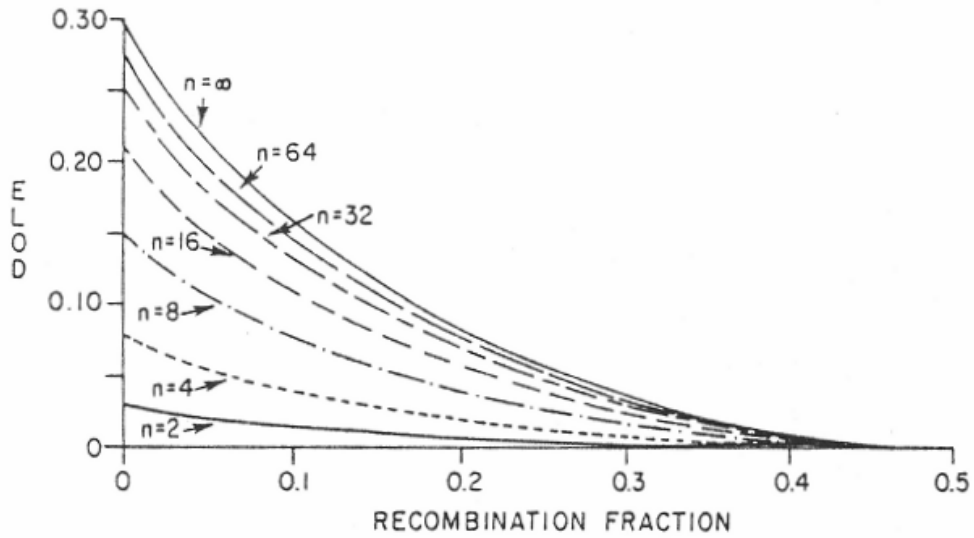


Figure 3 Variation in ELOD as a function of n, the number of alleles at the marker locus. All alleles are assumed to have frequency $1/n$. This calculation is performed for the grandparent-grandchild relationship with a rare trait allele frequency.

Effect of Recombination Varies According to Relationship

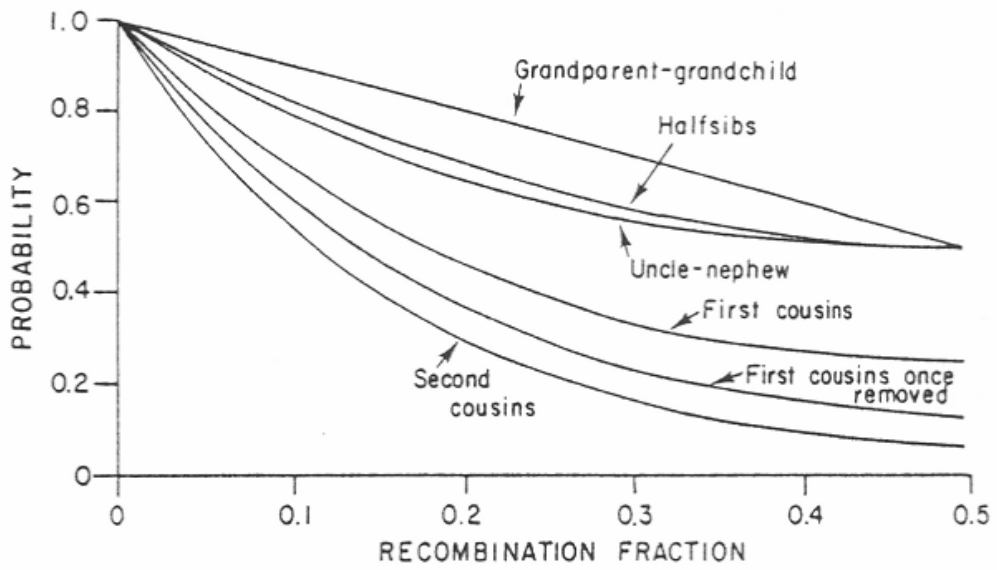


Figure 2 Probability of i.b.d. at a second linked locus conditional on i.b.d. at an index locus, as a function of the recombination fraction r between the loci, for specific genetic relationships. This function is $d_{11}(r)$ in the notation of table 1.

Power vs. P(IBS | Affected Pair)

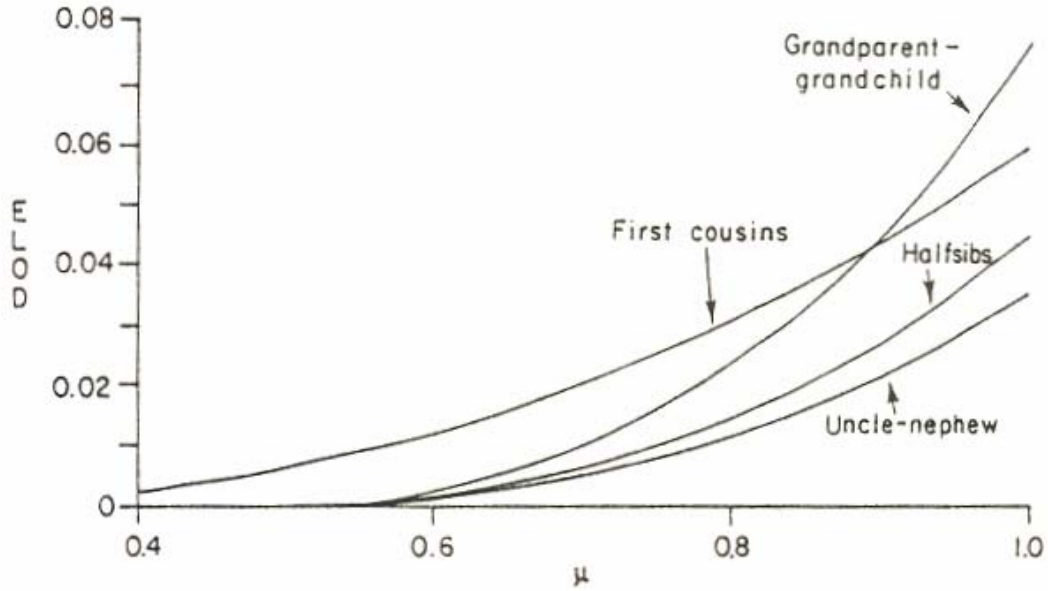


Figure 4 ELOD as a function of μ for an eight-allele marker system for a range of relationships. The recombination fraction between the trait and marker locus is .1.

With no phenocopies, rare alleles are easier to map

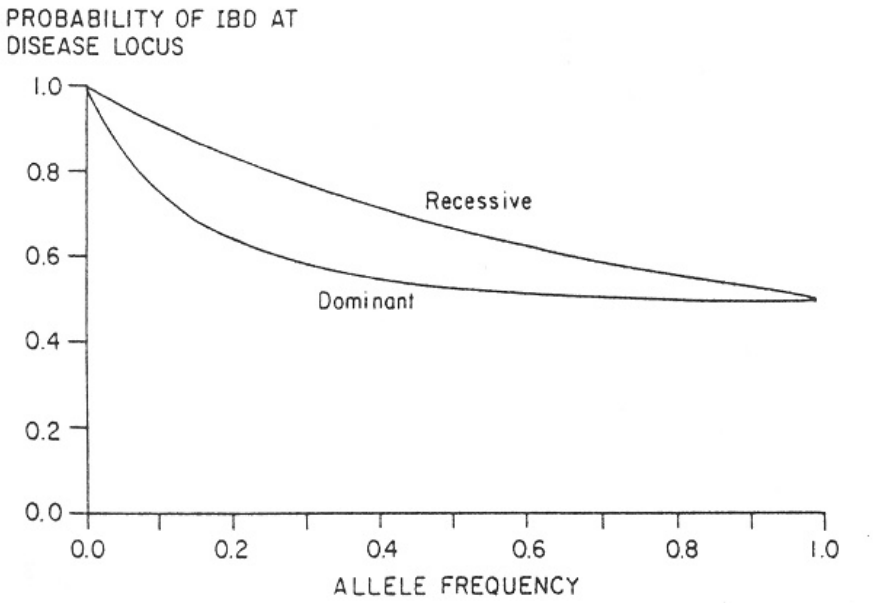


Figure 1 Probability of i.b.d. at a trait locus for two affected related individuals, as a function of the mode of inheritance of the trait. This figure is computed for the relationships with $\phi = .125$.

In general, phenocopies decrease power

Table 2

Average Informativeness for Mapping a Partially Penetrant Dominant Trait with Phenocopies

p and x	μ[a]	Phenocopy Rate	Relative Information Content (%)
.01:			
.000	.96	.00	100
.001	.96	.05	98
.01	.92	.33	81
.02	.88	.50	61
.05	.74	.71	23
.10	.61	.83	5
.10:			
.000	.75	.00	100
.001	.75	.00	99
.01	.74	.04	89
.02	.73	.08	80
.05	.69	.18	56
.10	.64	.30	31

NOTE.—The recombination fraction is .1, and the marker system has eight equally frequent alleles.

[a] For a grandparent-grandchild affected pair.

Shortcomings of IBS Method

- All sharing is weighted equally
  - Sharing a rare allele
  - Sharing a common allele
  - Sharing homozygous genotype
  - Sharing heterozygous genotype
- Inefficient.

Recommended Reading

- Bishop DT and Williamson JA (1990)
  Am J Hum Genet **46:**254-265

- Good introduction to linkage analysis in affected relative pairs, discusses
  - Marker choice
  - Recombination fraction
  - Disease model
  - Type of relative pair