Genetic Linkage

genes that are close together on a chromosome are likely to be inherited together because recombination is less likely to happen

genes that are far away on the same chromosome are likely to be separated during homologous recombination so are not linked

genes on separate chromosomes are never linked

Linkage Analysis

determine loci of genes on a chromosome influencing a specific trait or disease

look for evidence of coinheritance of the trait with other genes or markers whose locations are known

recombinant = odd number of crossovers between 2 loci

non-recombinant = even number of crossovers between 2 loci

Haplotype

combination of alleles that are inherited together on the same chromosome

humans have two haplotypes, one from each parent

phasing separates maternally and paternally inherited copies of each chromosome into haplotypes to determine which alleles are on the same chromosome

e.g. for a two-loci genotype, $Aa\ and\ Bb$, there are two possible haplotypes, $AB\ and\ ab$ or $Ab\ and\ aB$

Recombination

non-recombinant haplotype = haplotype whose alleles are from the same grandparent

recombinant haplotype = haplotype whose alleles are from both grandparents

recombination fraction (θ) = probability of recombination between two loci, related to how closely linked they are

$\underline{\theta = 0}$

genetic marker is the polymorphism causing the disease or the marker is so close to the disease mutation that recombination can never occur

$\underline{\theta = 0.5}$

50% chance the alleles are inherited together

50% chance recombination occurs between the alleles and are not inherited together

two loci are very apart on the same chromosome or located on two different chromosomes

Linkage Analysis

if you know the location of the marker and the distance between the marker and the disease locus, then can approximate the location of disease locus

H_0: $\theta = \frac{1}{2}$ The disease locus and marker segregate independently and are not linked.

H_A: $\theta < \frac{1}{2}$ The disease locus and marker are close together on the same chromosome so are linked.

LOD score = logarithm base 10 of the likelihood ratio

$$LOD(\theta_1) = \log_{10}\left(\frac{L(\theta = \theta_1)}{L(\theta = 0.5)}\right) = \log_{10}\left(\frac{{\theta_1}^R(1-\theta_1)^N}{0.5^{R+N}}\right)$$

$$\hat{\theta} = \frac{R}{R+N}$$

R = recombinants
N = nonrecombinants

LOD Score Method
$LOD\ score > 3$ implies linkage
$LOD\ score < -2$ implies no linkage
LOD scores between -2 and 3 are inconclusive
will need additional individuals or families or utilize additional markers to increase power of test

Multipoint Linkage Analysis
incorporates multiple markers and computes likelihood that a disease is located at a certain position on a chromosome
null hypothesis is that the disease locus is not on the chromosome

Factors Affecting Linkage Analysis

Phenocopy
observed phenotype not due to the genetic factor of interest
phenocopy rate can be introduced into the likelihood for linkage analysis by assigning a probability of observing the disease without the disease genotype

Penetrance
probability of expressing the disease given a specific phenotype
reduced penetrance = not expressing the trait even with the disease allele
e.g. $P(disease|DD\ or\ Dd) < 1$ for autosomal dominant disease
age-dependent penetrance = penetrance increases with age, e.g. Alzheimer's and Huntington's disease

Genetic Heterogeneity
many disorders are caused by mutations in any of the multiple causal genes
LOD score over all families may not show evidence of linkage, but significant linkage may occur within a single large family or a subset of families
restrict analysis to homogenous subset of families

Identity By Descent (IBD)

alleles that are copies of the same allele from a common ancestor

if an allele is causing the uncommon disease, two affected relatives will most likely share the allele IBD

region around the disease locus will likely also be IBD

look for regions of IBD sharing that are higher than expected in order to locate the disease locus

Degree of Relationship	**Relative Pair**	**Expected Probability of Number of IBD Alleles**		
		0	**1**	**2**
Zero Degree	monozygotic twins	0	0	1
First Degree	dizygotic twins	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
	siblings	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
	parent, child	0	1	0
Second Degree	half-siblings	$\frac{1}{2}$	$\frac{1}{2}$	0
	aunt/uncle, niece/nephew	$\frac{1}{2}$	$\frac{1}{2}$	0
	grandparent, grandchild	$\frac{1}{2}$	$\frac{1}{2}$	0
	double first cousins	$\frac{9}{16}$	$\frac{6}{16}$	$\frac{1}{16}$
Third Degree	first cousins	$\frac{3}{4}$	$\frac{1}{4}$	0
	great-grandparent, great-grandchild	$\frac{3}{4}$	$\frac{1}{4}$	0

Linkage Analysis Using Affected Sibling Pairs

H_0: There is no linkage between the disease and marker.

Probability of IBD Marker Alleles		
0	**1**	**2**
0.25	0.50	0.25

H_A: There is linkage between the disease and marker, therefore excess sharing of IBD.

Probability of IBD Marker Alleles		
0	**1**	**2**
<0.25	>0.50	>0.25

Dominant Trait

marker allele on the same haplotype as the disease gene from affected parent will be shared by the siblings

allele from the other parent has expected chance of being shared by the siblings

IBD status for a marker very close to the gene will be at least 1

Recessive Trait

marker allele on the same haplotype as the disease gene from each parent will be shared by the siblings

IBD status for a marker very close to the gene must be 2

Goodness of Fit Test / Likelihood Ratio Test

$$n_i = \#\ sibling\ pairs\ sharing\ i\ IBD\ alleles$$
$$N = total\ number\ of\ sibling\ pairs$$

IBD	Observed	Expected
0	n_0	$\frac{N}{4}$
1	n_1	$\frac{N}{2}$
2	n_2	$\frac{N}{4}$

$$\chi^2 = \frac{4\left(n_0 - \frac{N}{4}\right)^2 + 2\left(n_1 - \frac{N}{2}\right)^2 + 4\left(n_2 - \frac{N}{4}\right)^2}{N}, 2\ df$$

LOD Score

$$\chi^2 = 2\ln(10) \times LOD\ Score$$
$$LOD\ Score = \frac{\chi^2}{2\ln(10)}$$
$$z - score^2 \approx t - score^2 \approx \chi^2$$

calculate corresponding one-sided p-value

only interested in a situation where the distribution suggests excess IBD sharing $\left(\theta < \frac{1}{2}\right)$, not a deficit of IBD sharing $\left(\theta > \frac{1}{2}\right)$

H_0: $\theta = \frac{1}{2}$ The disease locus and marker segregate independently and are not linked. There is no excess of IBD sharing.

H_A: $\theta < \frac{1}{2}$ The disease locus and marker are close together on the same chromosome so are linked. There is an excess of IBD sharing.

Quantitative Trait Locus (QTL) Mapping

relatives who have similar trait values should have higher than expected levels of sharing of genetic material near the genes that influence those traits

there should be less sharing of genetic material among relatives with dissimilar trait values

Haseman-Elston Approach

$$E\left[Y_j^{\,D}\right] = \alpha + \beta\pi_j$$

linear regression of $Y_j^{\,D}$ on π_j using mean squared difference

$$Y_j^{\,D} = \left(Y_{1j} - Y_{2j}\right)^2$$

Y_{1j} = phenotype of sibling 1 of the j^{th} sibpair

Y_{2j} = phenotype of sibling 2 of the j^{th} sibpair

π_j = proportion of IBD marker alleles by the j^{th} sibpair

$0, \frac{1}{2}, or\ 1$ when IBD is known with certainty

ranges from 0 to 1 if not known with certainty

H_0: $\theta = \frac{1}{2}$ $\beta = 0$ The disease locus and marker segregate independently and are not linked. There is no relationship between IBD marker status and the trait similarity between siblings.

H_0: $\theta < \frac{1}{2}$ $\beta < 0$ The disease locus and marker are close together on the same chromosome so are linked.

negative slope implies linkage

relatives with similar trait values have smaller squared differences and higher IBD sharing

calculate corresponding one-sided p-value to t-score

Haseman Elston Revisited

$$E\left[Y_j^{\,D}\right] = \alpha^* + \beta^*\pi_j$$

linear regression of $Y_j^{\,D}$ on π_j using mean-adjusted product

$$Y_j^{\,P} = \left(Y_{1j} - \mu\right)\left(Y_{2j} - \mu\right)$$

Y_{1j} = phenotype of sibling 1 of the j^{th} sibpair

Y_{2j} = phenotype of sibling 2 of the j^{th} sibpair

μ = population mean phenotype

H_0: $\theta = \frac{1}{2}$ $\beta = 0$ The disease locus and marker segregate independently and are not linked. There is no relationship between IBD marker status and the trait similarity between siblings.

H_0: $\theta < \frac{1}{2}$ $\beta < 0$ The disease locus and marker are close together on the same chromosome so are linked.

positive slope implies linkage
siblings with similar trait values have larger mean-adjusted product and higher IBD sharing
calculate corresponding one-sided p-value to t-score