Linkage Analysis for Gene Identification

Chiara Sabatti

www.stat.ucla/~sabatti

csabatti@mednet.ucla.edu

sabatti/home/teaching/teach.html
Genetic therapy –
treatment of different forms of disease

Sophisticated diagnostic tools suggest the correct
disease and the same drugs may be useful for other stuff

This same pathway could be affected in other forms of the

develop drugs that restore functionality of pathways

Learn about the biological pathways affected

Screening and counseling

Relevant for medicine

Curiosity

Why do we want to find a gene?
What does it mean to find a gene?

1. Locate a chromosomal region where the gene seems to be studied in detail.
2. Refine the possible region to a length where it can be studied in detail.
3. Identify one gene that is mutated in affected individuals.
   - Study of polymorphism
   - Study of functions of genes involved
   - Sequence
4. Understand the function of that one gene and the effect of
   the mutation.

And not in controls.
Grand Canyon while driving in the desert.
DNA is polymorphic. (ex. street signs, Las Vegas and the
recognize, of which we know the relative positions and where
Markers are landmarks in the genome: places that we

The first step: locate gene on a map
How do we construct maps?

distance between beads proportional to number of times they
are together
Cutting processes and map types

(see book) number of bases
(More later) Morgans
radiation recombinatation
physical maps genetic maps
The average number of crossover per gamete is 1.

Two chiasmata:

Crossover: the hidden process
The distance between two positions \( x \) and \( y \) on a chromosome is measured in Morgans.

\[
d(x, y) = E(\text{crossover between } x \text{ and } y)
\]

The distance between \( x \) and \( y \) is defined as the expected number of crossovers per gamete.
Recombination: the observable process

0 recombinant gametes between marker I and III
1 recombinant gametes between markers II and III
2 recombinant gametes between markers I and II
Recombination and distance

We can go from Recombination to distance with mathematical formulas.
(Marshfield, CEPH families)

By looking at genotypes of pedigrees

How do we measure Recombination?
Estimate recombination between disease gene and markers

- Genotyphes markers that cover the entire genome
- Collect families with affected individuals

Mapping a Disease Gene
\[
B_{\text{prob}}(\text{observation} | \theta) = \frac{1/2}{\text{prob}(\text{observation} | \theta)} < 3 < \log_{10} \text{prob}(\text{observation} | \theta) \leq \text{prob}(\text{observation} | \theta)
\]

This result is statistically significant.

And marker is $\gt \frac{1}{2}$

The estimated recombination fraction $\theta$ between disease

Linked = recombination fraction $\theta$ $\gt \frac{1}{2}$

When is a disease „linked” to a marker?
\{(−, −)\} = M \quad \{(−, +)\} = 1

Generation II haplotypes are

\((−, −)\)

\((+^+, +)\) or \((−^+, +)\)

\((p^d)\) normal

\((D^d)\) or \((D,d)\) disease

An example
there is no evidence for linkage

\[
\text{Lod score} = \log_{10} \left( \frac{8/1}{8/(8-1)(8/1)} \right) = 1.1
\]
markers I have to worry about multiple comparisons

if I calculate the possible linkage with a big number of
collect very strong evidence

it is very unlikely that two things are linked, so one has to

observation under recombination = 1 / 2

best estimate of recombination over probability of

lodscore is log of ratio of probability of observations under

not enough data to be conclusive

why?
markers as one piece of information: multipoint analysis

It is important to consider the information from proximal models

many cases inheritance is not Mendelian: non parametric

The above was done assuming dominant inheritance, but in

In most cases we have to impute phases

More sophistication
how well defined is a phenotype
how many recombination events can be observed per family
how informative are the markers
how good is the map information
how well the markers span the genome

The success of a linkage screen will depend on

Some issues
Recombination.

individuals as a big family and observe the effects of
In some cases we can treat the population of diseased
need a narrower map interval!
sequencing machines give you a reliable read for 350-400
I CM \( \approx 1000000 \) bases
of crossover per gamete = 0.01.
between markers that are less than 1 cm apart (probability
In families is quite difficult to observe recombination
The power of resolution comes from recombination.

From family to population data
allel and the disease status. 

Generations by looking at the association between one 
observe the effects of recombination on hundreds of 
if the population has derived from 1 ancestor, we can 

populations.

allel at linked markers, this is not true in general for 
In a given family, the disease status is associated with one 

eneration, so we do not directly observe recombinations. 
population of diseased individuals, we are looking at only 1 
recombination events. When we look at the current 
By looking at pedigrees, we can actually observe 

Family vs Population data
distribution from the general population. The markers really close to the disease locus will tend to inherit this haplotype; initially all the chromosomes that inherit the disease.

Suppose that in a population of 100 chromosomes, 1 undergoes a mutation in a gene that causes a disease. Founder effect
dense markers.

It may be applied as a genome-screen technique with very

It is quite complicated to model mathematically.

It is connected to population history.

Works well in population isolates.

Status.

Marker whose allele distribution varies the most with disease

state at various markers and locate the disease near the

We look for association between allele distribution and disease

Linkage Disequilibrium
complex diseases.
Gene expression data may help defining the phenotype of
lots of markers and very good maps.
Consequences of the human genome project: we will have
more a promise than a reality.
Linkage disequilibrium may be more powerful, but it is still
For complex diseases, no strategies has yet won the battle.
Linkage has been very successful for Mendelian diseases.

The Future of Gene Localization