The genetic dissection of complex traits

Karl W Broman

Department of Biostatistics
Johns Hopkins University

http://www.biostat.jhsph.edu/~kbroman

Goal

Identify genes that contribute to complex human diseases

Complex disease = one that's hard to figure out

Many genes + environment + other

QTL = quantitative trait locus

Genomic region that affects a quantitative trait

2

The genetic approach

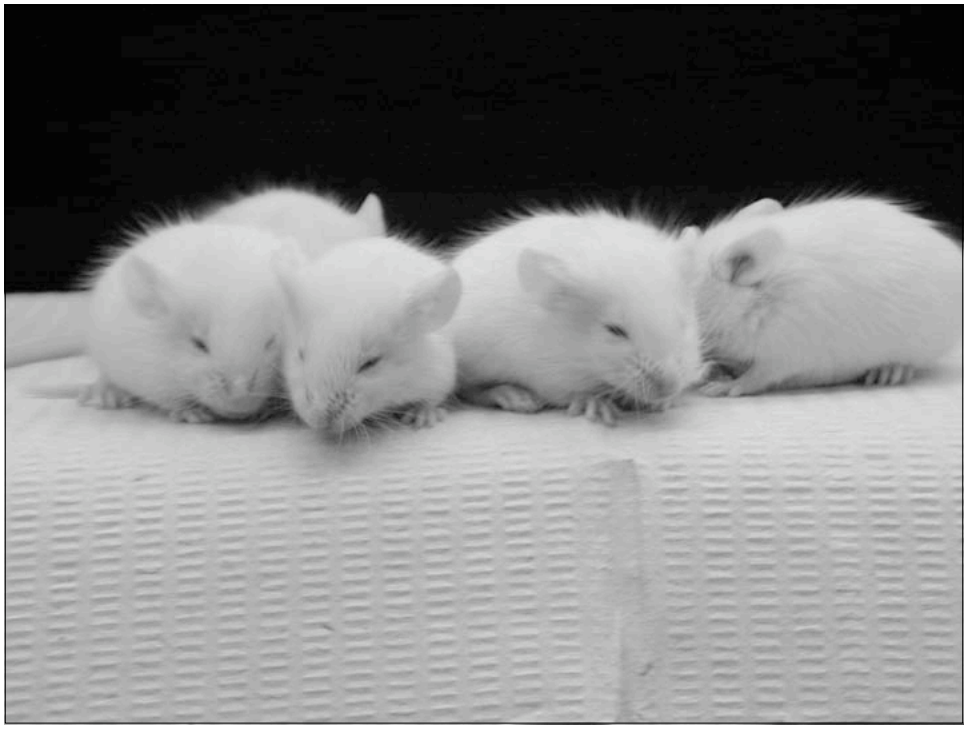
- Start with the trait; find genes the influence it.
  - Allelic differences at the genes result in phenotypic differences.
- Value: Need not know anything in advance.
- Goal
  - Understanding the disease etiology (e.g., pathways)
  - Identify possible drug targets

3

Approaches

- Experimental crosses in model organisms
- Mutagenesis in model organisms
- Linkage analysis in human pedigrees
  - A few large pedigrees
  - Many small families (e.g., sibling pairs)
- Association analysis in human populations
  - Isolated populations vs. outbred populations
  - Candidate genes vs. whole genome

4



Advantages of the mouse

- Small and cheap
- Inbred lines
- Disease has simpler genetic architecture
- Controlled environment
- Large, controlled crosses
- Experimental interventions
- Knock-outs and knock-ins

6

Disadvantages of the mouse

- Is the model really at all like the corresponding human disease?
- Still not as small (or as fast at breeding) as a fly.

7

The mouse as a model

- Same genes?
  - The genes involved in a phenotype in the mouse may also be involved in similar phenotypes in the human.
- Similar complexity?
  - The complexity of the etiology underlying a mouse phenotype provides some indication of the complexity of similar human phenotypes.
- Transfer of statistical methods.
  - The statistical methods developed for gene mapping in the mouse serve as a basis for similar methods applicable in direct human studies.

8

Mutagenesis

Advantages

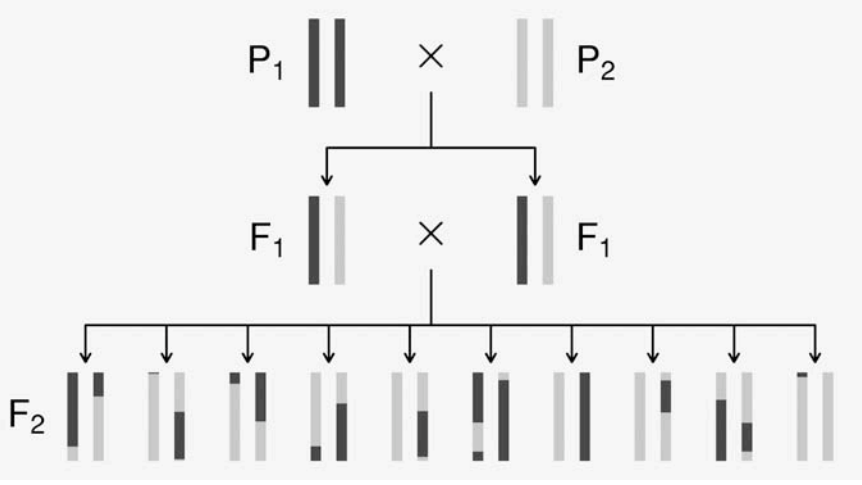
+ Can find things
+ Genes at least indicate a pathway

Disadvantages

– Need cheap phenotype screen
– Mutations must have large effect
– Genes found may not be relevant
– Still need to map the mutation
– Mutations with recessive effects are hard to see

9

The intercross



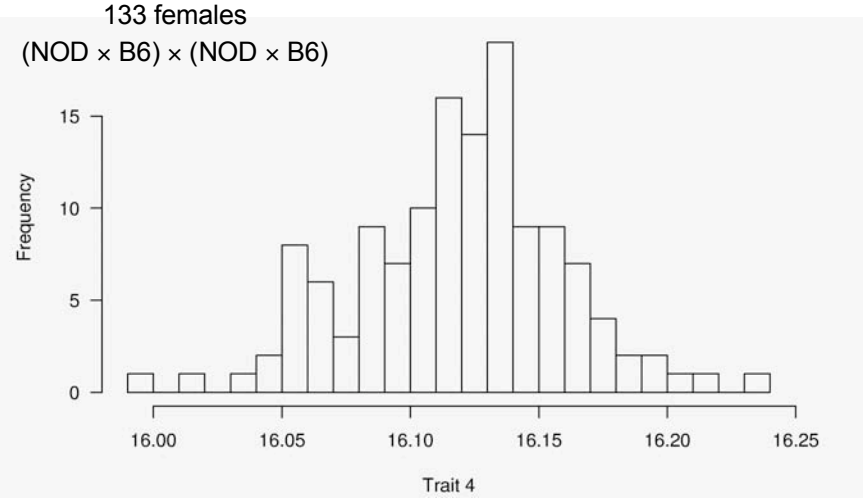
10

The data

- Phenotypes, y_i
- Genotypes, x_{ij} = AA/AB/BB, at genetic markers
- A genetic map, giving the locations of the markers.

11

Phenotypes



12

NOD

13

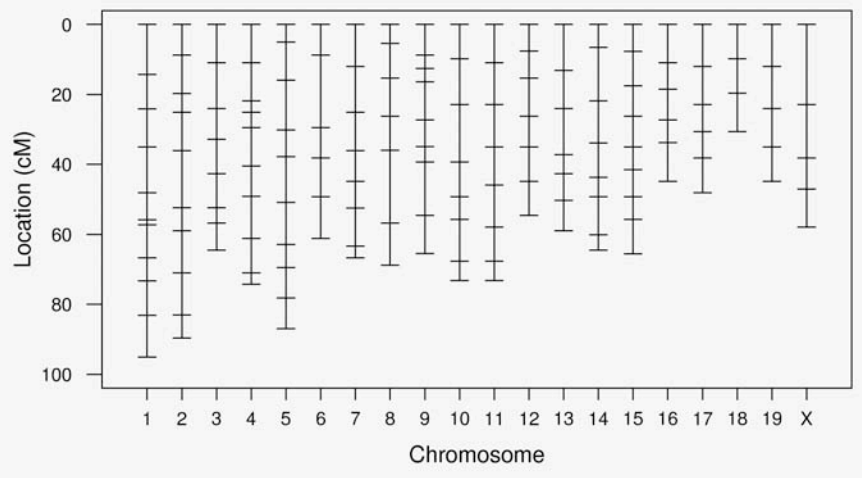
C57BL/6

14

Agouti coat

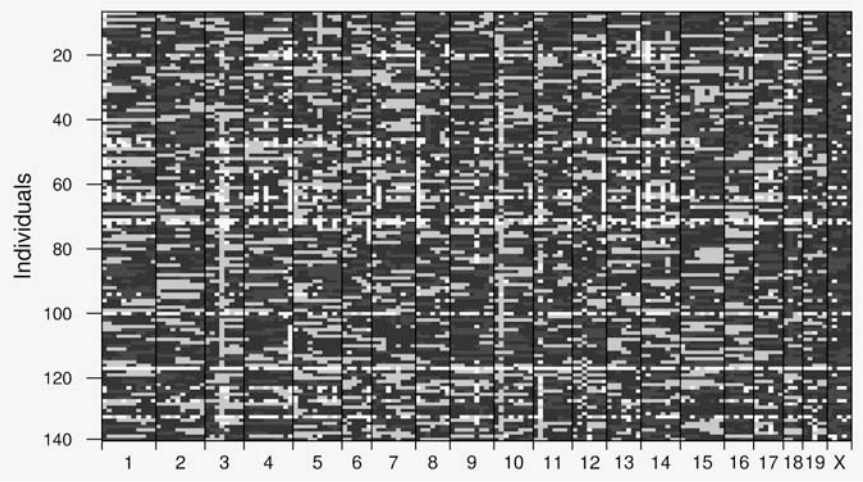
15

Genetic map



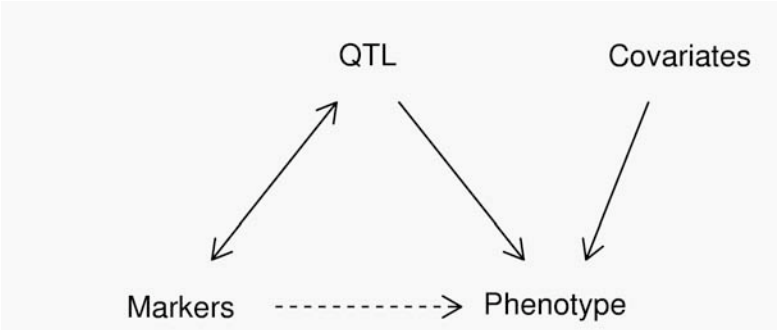
16

Genotype data



17

Statistical structure



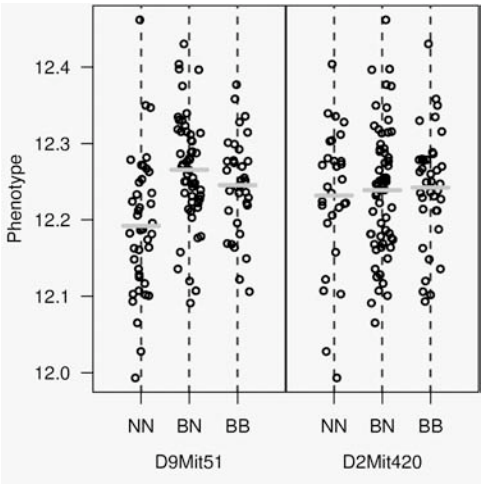
- Missing data: markers ↔ QTL
- Model selection: genotypes ↔ phenotype

18

The simplest method

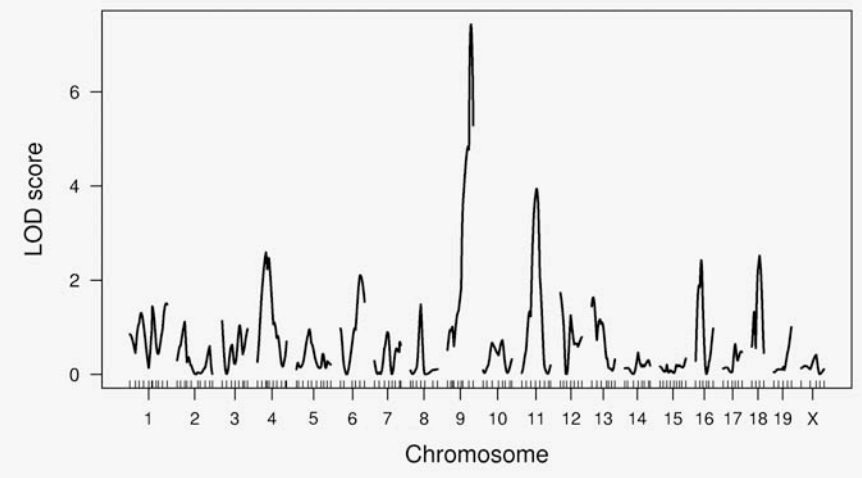
"Marker regression"

- Consider a single marker
- Split mice into groups according to their genotype at a marker
- Do an ANOVA (or t-test)
- Repeat for each marker



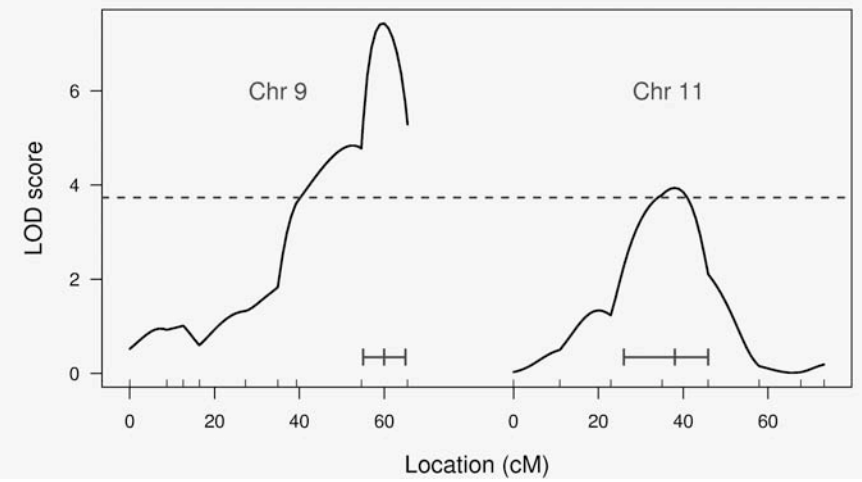
19

LOD curves



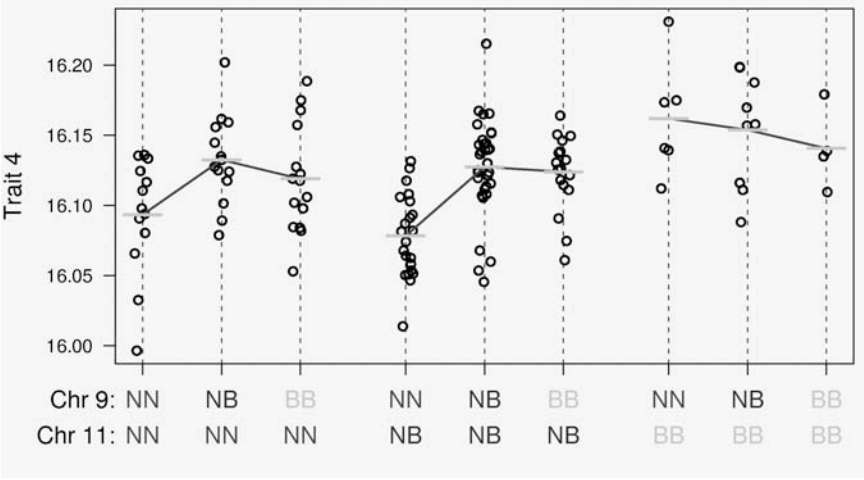
20

Chr 9 and 11



21

Epistasis



22

Back to the strategy

- First: QTL mapping results in a 10-20 cM region
- Next step: create congenics
- Then: subcongenics
- Then: test candidates
- Finally: prove a gene is the gene

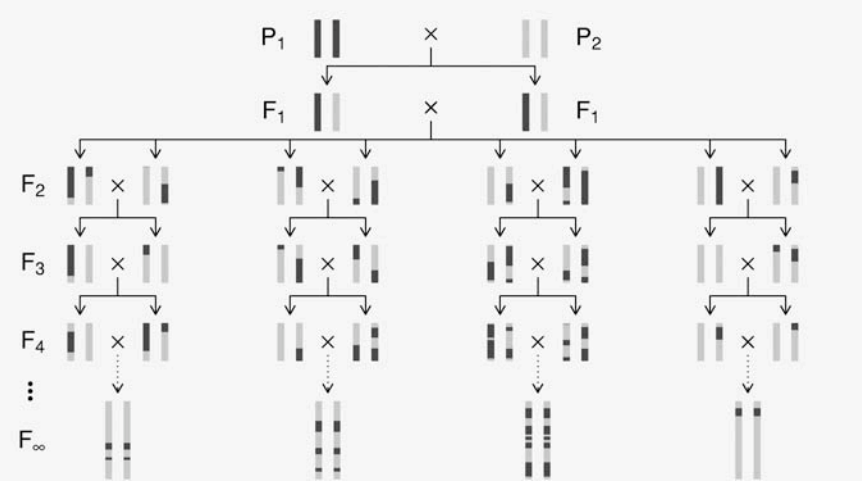
23

"Modern" approaches

- Recombinant inbred lines (RILs)
- Advanced intercross lines (AILs)
- Heterogeneous stock (HS)
- The Collaborative Cross (CC)
- Partial advanced intercross (PAI)
- Association mapping across mouse strains
- Combining crosses, accounting for the history of the inbred strains
- Gene expression microarrays

24

Recombinant inbred lines



25

RI lines

Advantages

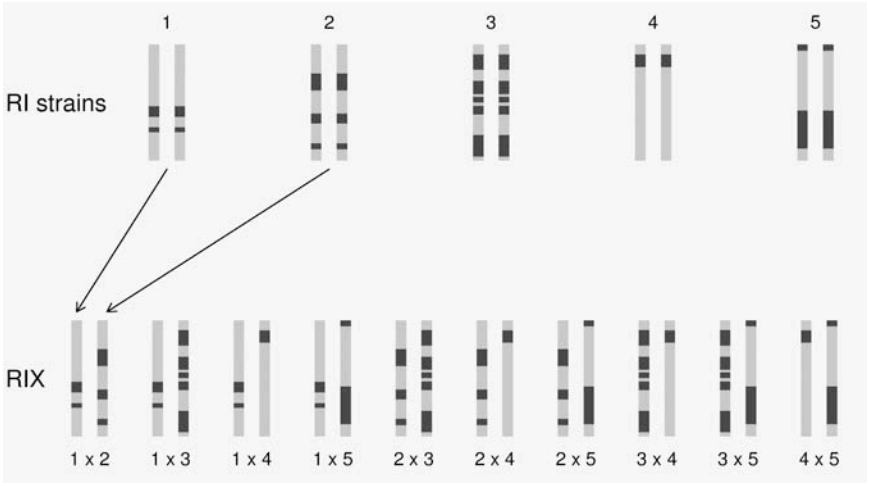
- Each strain is a eternal resource.
  - Only need to genotype once.
  - Reduce individual variation by phenotyping multiple individuals from each strain.
  - Study multiple phenotypes on the same genotype.
- Greater mapping precision.

Disadvantages

- Time and expense.
- Available panels are generally too small (10-30 lines).
- Can learn only about 2 particular alleles.
- All individuals homozygous.

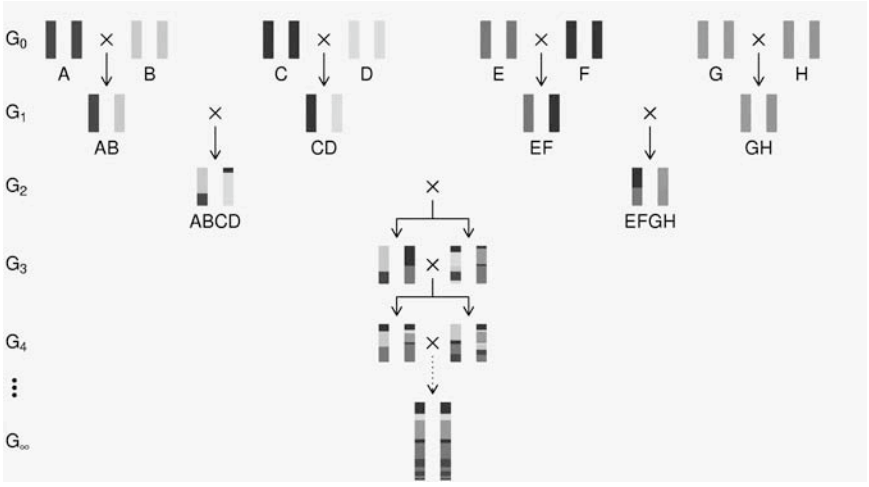
26

The RIX design



27

The "Collaborative Cross"



28

Genome of an 8-way RI

29

Heterogeneous stock

McClearn et al. (1970)

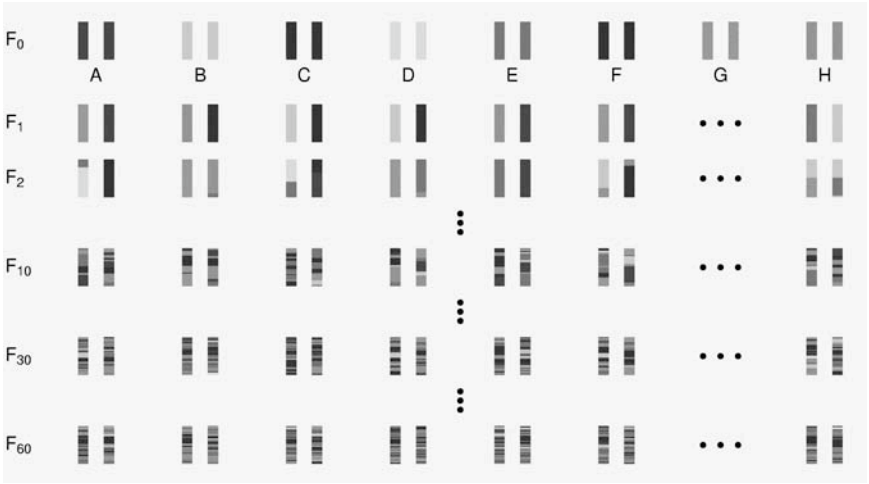
Mott et al. (2000); Mott and Flint (2002)

- Start with 8 inbred strains.
- Randomly breed 40 pairs.
- Repeat the random breeding of 40 pairs for each of ~60 generations (30 years).
- The genealogy (and protocol) is not completely known.

Note: AILs are similar, but start with 2 strains and don't go as many generations

30

Heterogeneous stock



31

"Modern" approaches

- Recombinant inbred lines (RILs)
- Advanced intercross lines (AILs)
- Heterogeneous stock (HS)
- The Collaborative Cross (CC)
- Partial advanced intercross (PAI)
- Association mapping across mouse strains
- Combining crosses, accounting for the history of the inbred strains

32

Towards proof

- Gene has nonsynonymous mutation
- Gene shows difference in expression between parental strains
- Expression variation correlated with QTL genotype
- RNA interference
- Knock out/knock in

33

Summary

- Experimental crosses in model organisms
  - + Cheap, fast, powerful, can do direct experiments
  - – The model may relevant to the human disease
- Standard QTL mapping results in large regions with many genes
- Fine mapping
  - – Congenics, AILs, RILs, HS, PAI, association mapping
  - – Expression differences
- Proof
  - – RNA interference
  - – Knock outs/knock ins

34