

# Complex trait genetics in mice

Karl Broman

Biostatistics & Medical Informatics  
Univ. Wisconsin–Madison

kbroman.org  
github.com/kbroman  
@kbroman  
Slides: [bit.ly/BMI2017-03](https://bit.ly/BMI2017-03)



These are slides for a talk that I gave to prospective graduate students on 31 March 2017.

The source for the slides is at GitHub:  
[https://github.com/kbroman/Talk\\_ProspStudents](https://github.com/kbroman/Talk_ProspStudents)

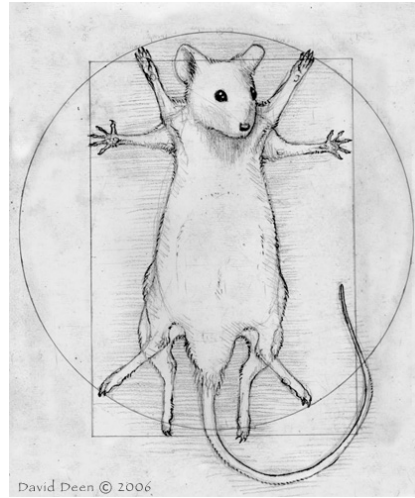
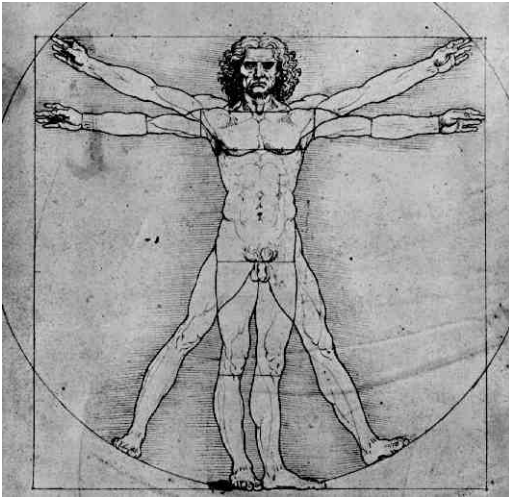
Slides and interactive graphs: [bit.ly/BMI2017-03](https://bit.ly/BMI2017-03)



I'll start with a bit of background.

I focus on genetics problems, and particularly on mouse genetics.

I think these are SWR mice; the photo is from David Threadgill.

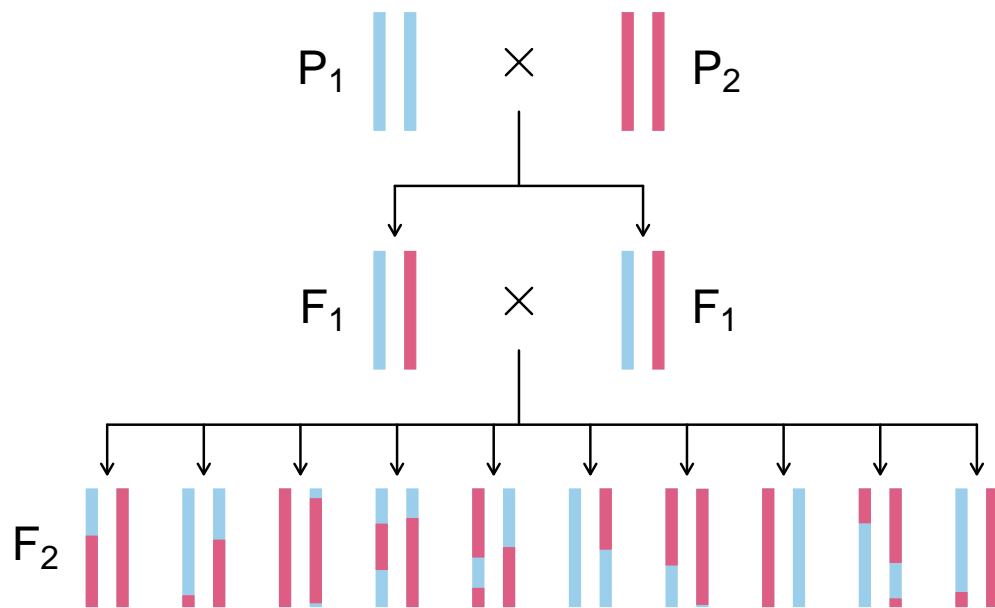


daviddeen.com

Mice are not humans, but you can learn a great deal about human biology and disease from mice.

The figure on the right is from David Deen.

# Intercross



4

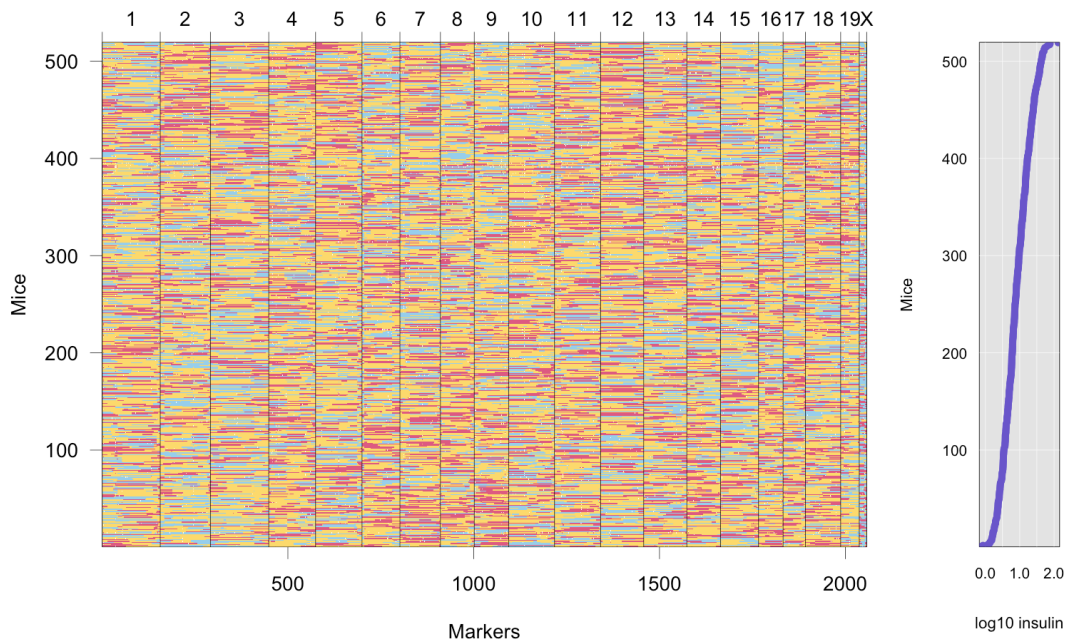
I've mostly focused on simple crosses between two inbred strains.

Say strain P<sub>1</sub> has low blood pressure and P<sub>2</sub> has high blood pressure. We cross the two strains to get the F<sub>1</sub> hybrid, and then intercross F<sub>1</sub> siblings to get a large set of F<sub>2</sub> individuals.

The F<sub>2</sub> mice may inherit a P<sub>1</sub> or P<sub>2</sub> chromosome intact, but generally their chromosomes are a mosaic of the two parental chromosomes as a result of recombination at meiosis. The points of exchange are called crossovers or recombination events.

At any one autosomal locus, the F<sub>2</sub> individuals will have genotype BB, BR, or RR. We'd generate many such mice and then determine their genotype along chromosomes as well as measure their phenotype (e.g., blood pressure). The simplest analysis is to look for genomic regions where genotype is associated with phenotype.

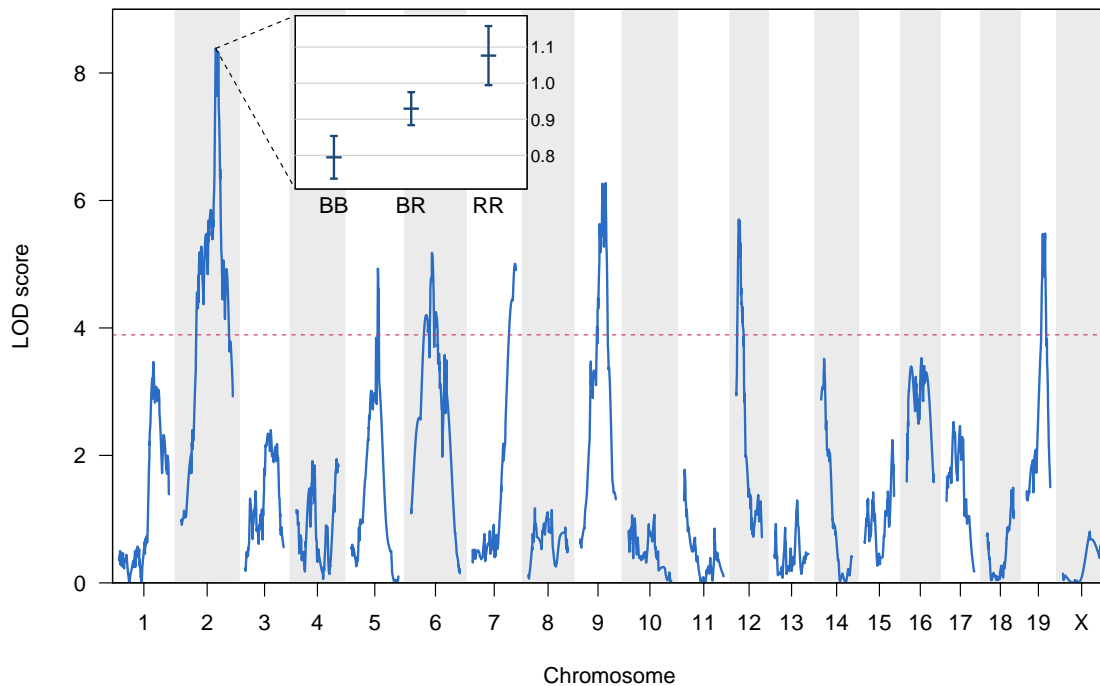
# Genotype & phenotype data



5

The data consist of genotypes at a set of markers across the genome, plus some quantitative phenotype for each mouse. The goal is to identify regions of the genome where the genotype is associated with the phenotype.

# QTL mapping



6

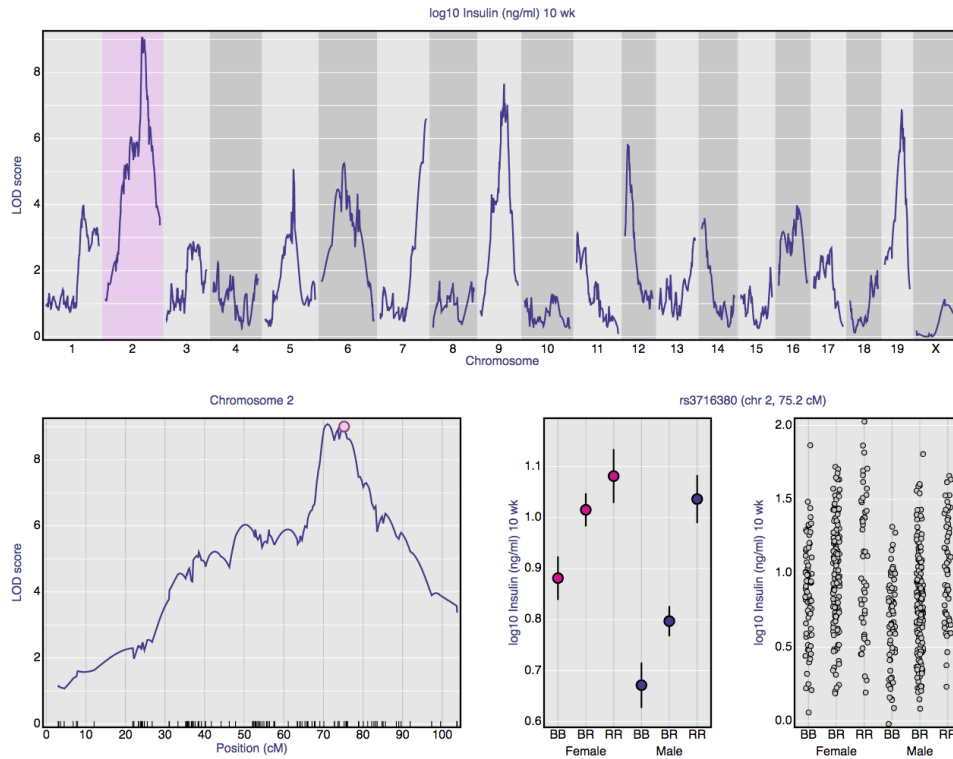
Our goal is to identify quantitative trait loci (QTL): regions of the genome for which genotype is associated with the phenotype.

The basic analysis is to consider each locus, one at a time, split the mice into the three genotype groups, and perform analysis of variance.

We then plot a test statistic that indicates the strength of the genotype-phenotype association. For historical reasons, we calculate a LOD score as the test statistic: the  $\log_{10}$  likelihood ratio comparing the hypothesis that there's a QTL at that position to the null hypothesis of no QTL anywhere.

Large LOD scores indicate evidence for QTL and correspond to there being a difference in the phenotype average for the three genotype groups.

# Interactive graph



7

I've become convinced of the importance of interactive data visualizations: basically all data visualizations are improved with some amount of interactivity. This enables a more thorough exploration of results, which is particularly important and valuable for high-dimensional data.

Click on the slide title to jump to the interactive version:

<http://bit.ly/D3lod>

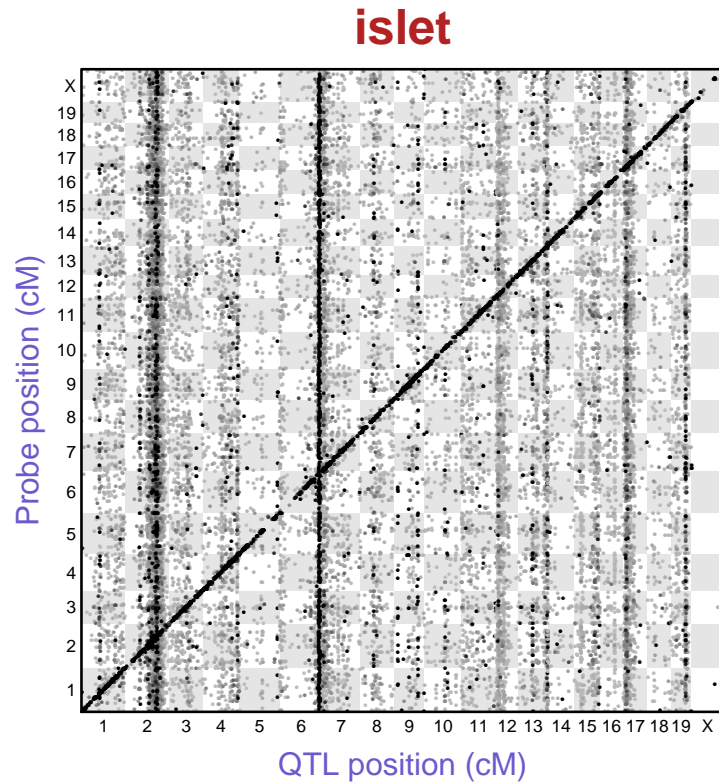


I've mostly focused on a mapping genes affecting a single phenotype, but in the past decade, I've become swamped with data.

This is a picture of a pile of gene expression arrays. More and more, we're seeing genome-scale phenotype information. For example, in one of my collaborations, we have data on 500 mice, each with gene expression microarrays for 6 different tissues.

We're interested in identifying genes that control the expression of other genes.

# eQTL analysis



9

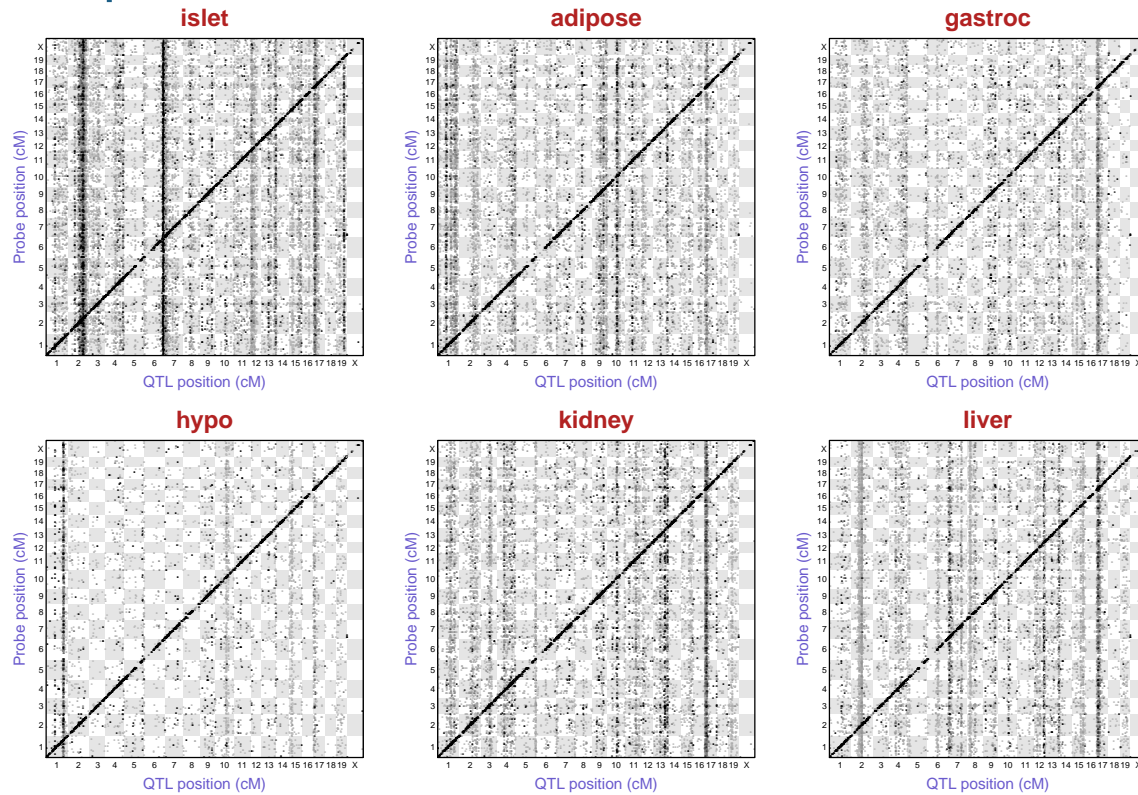
In collaboration with Alan Attie, we've been studying a B6×BTBR intercross, with all mice knocked out for leptin (and so obese), in order to understand obesity-induced diabetes.

There are 500 intercross mice, phenotyped at a large number of clinical traits, and also with gene expression microarray data on 6 tissues. These were custom two-color Agilent arrays.

This figure shows the basic result of single-QTL genome scan for each expression trait, one at a time, in pancreatic islets. Each dot is an inferred QTL. The y-axis is the location of the corresponding microarray probe, and the x-axis is the location of the QTL.

We see a prominent diagonal, of local-eQTL, and several prominent vertical bands: “trans-eQTL hotspots” where genotype at a give region is associated with the mRNA expression of numerous genes across the genome.

# Multiple tissues



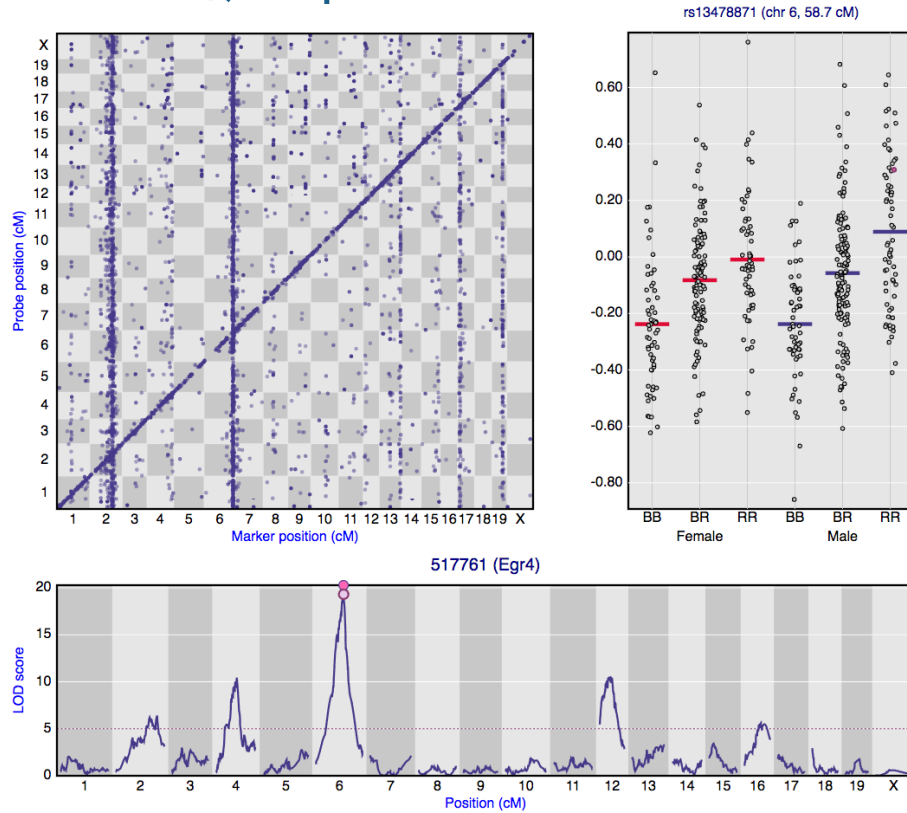
10

Here are the results for all six tissues.

There are numerous “trans-eQTL hotspots” (where genotype at a give region is associated with the mRNA expression of numerous genes across the genome). Some of these trans-eQTL hotspots are specific to a given tissue (e.g. islet chr 6) and some are seen in many tissues (e.g. chr 17).

We seek to fine-map these trans-eQTL hotspots, and to determine whether they involve one or multiple eQTL.

# Interactive eQTL plot



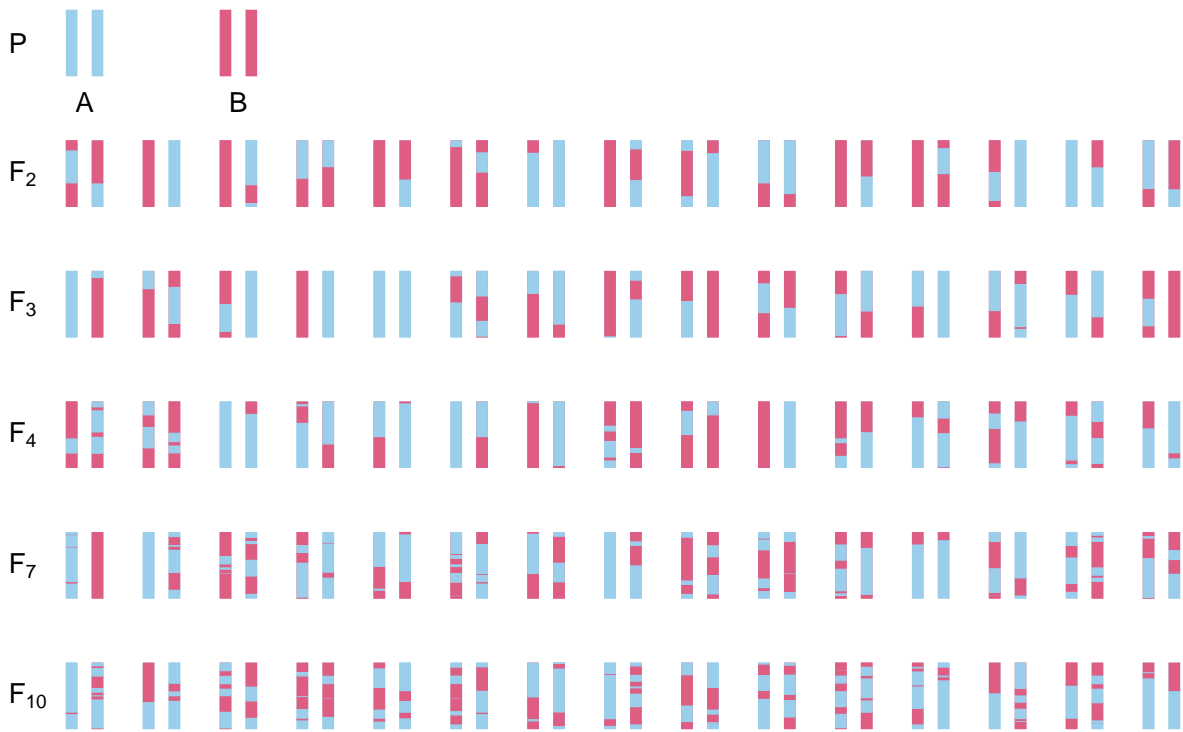
11

An interactive version of the eQTL plot allows you to get to the underlying detailed results.

Click on the slide title to jump to the interactive version:

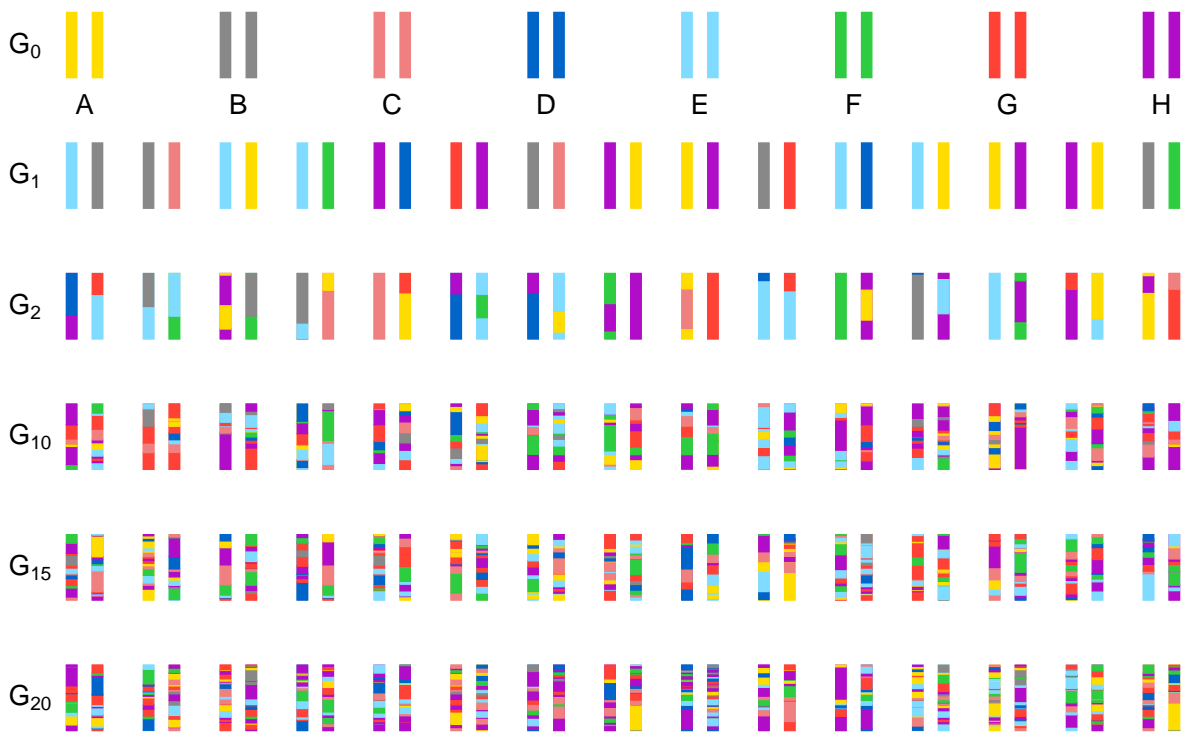
<http://bit.ly/D3cistrans>

# Advanced intercross lines



One strategy for dealing with the low-resolution mapping of QTL is to perform multiple generations of intercrossing, to break up the chromosomes into small chunks.

# Heterogeneous stock



13

The cost (in time and money) to create advanced intercross lines is generally prohibitive for a single lab. But for multiple labs to collaborate on this sort of effort, they'll want to ensure that the strains involved are relevant to their traits of interest.

Heterogeneous stock is an advanced intercross derived from multiple (say 8) founder strains.

Analysis issues include the reconstruction of the haplotypes along each chromosome in each individual, the treatment of the multiple genotypes at a single locus, and the need to account for population structure.

# Summary

- ▶ Gene mapping in mice is fun and useful
- ▶ Lots of open questions:
  - Multiple loci contributing to one trait
  - Joint consideration of many traits
  - Treatment of complex crosses with >2 strains
  - Prioritizing candidate genes
- ▶ Practicality > optimality
- ▶ Data analysis, statistical methods, software, data visualization

It's always good to have a summary slide.

Slides: [bit.ly/BMI2017-03](http://bit.ly/BMI2017-03)



[kbroman.org](http://kbroman.org)

[github.com/kbroman](https://github.com/kbroman)

@kbroman

Here's where you can find me, as well as the slides for this talk.