AIL, RIL, RIX, HS, CC, CSS

Karl W Broman

Department of Biostatistics
Johns Hopkins University

www.biostat.jhsph.edu/~kbroman

[→ Teaching → Miscellaneous lectures]

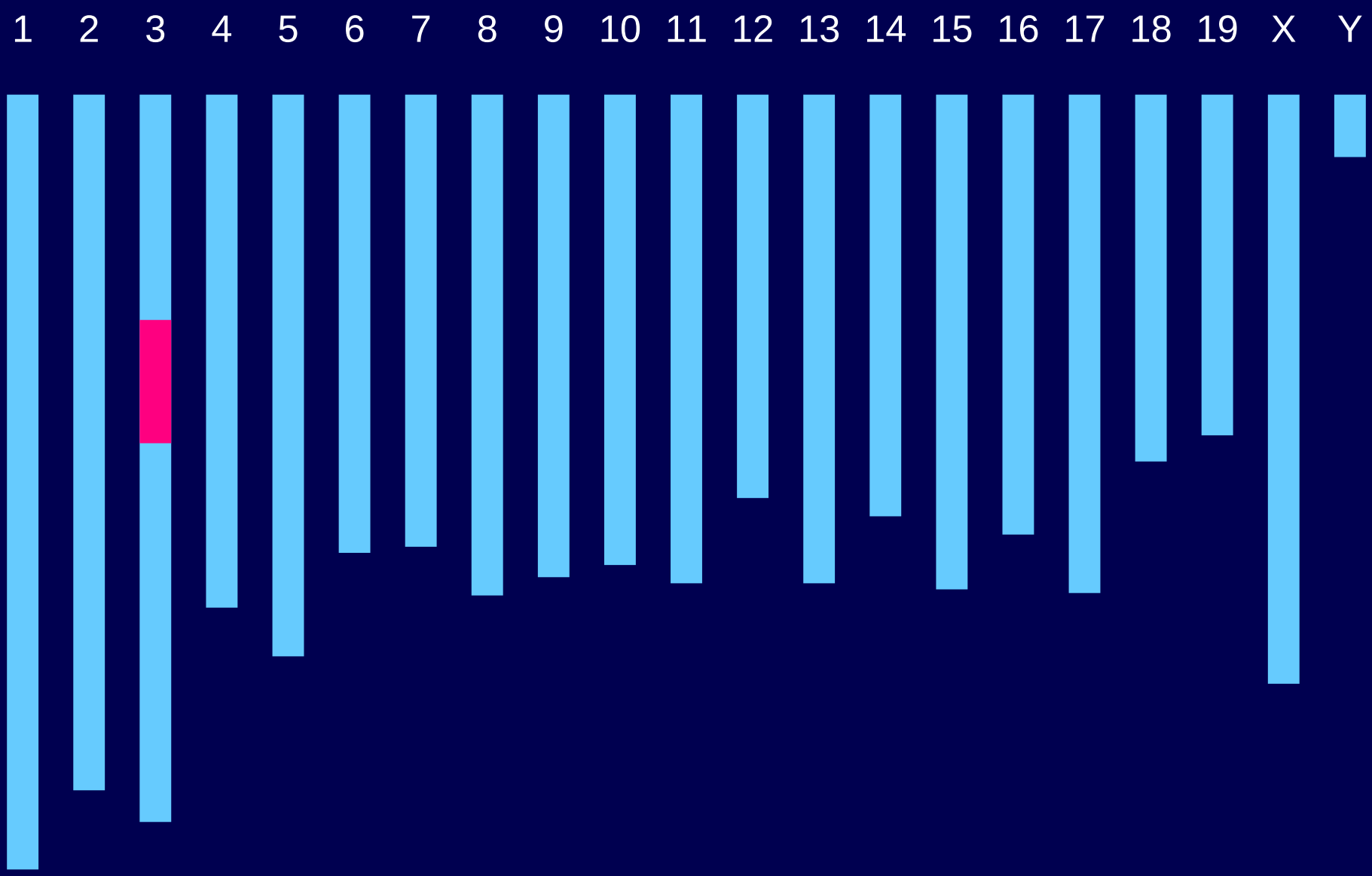
Traditional approach

F2/BC $\longrightarrow$ QTL $\longrightarrow$ congenic $\longrightarrow$ subcongenics

$\longrightarrow$
- coding/regulatory polymorphism
- expression/function difference
- knock-in / transgenic
- knock-out
- homology to other species

Issues:
- Large QTL regions
- Time consuming and expensive

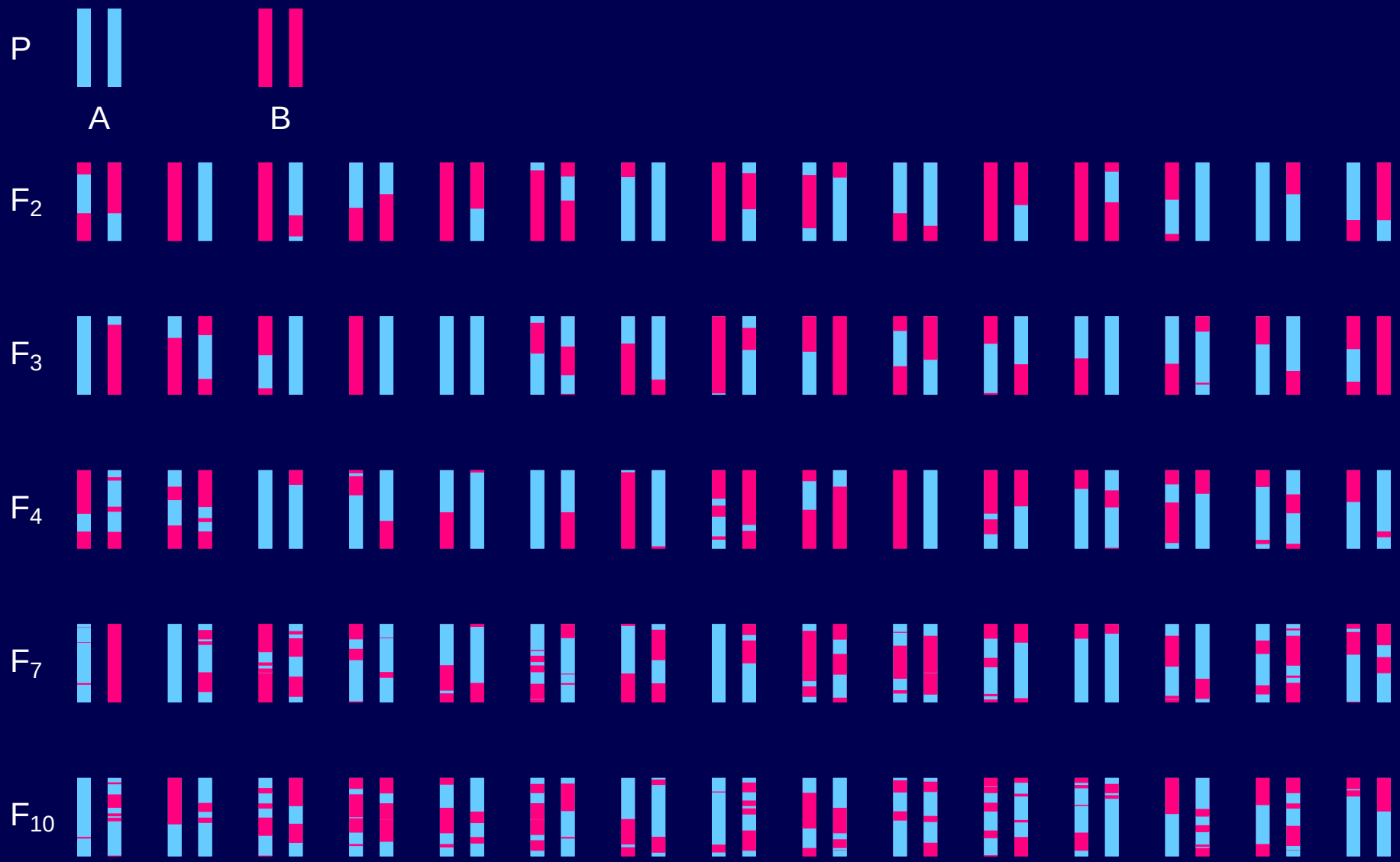
A congenic line



Advanced intercross lines

- Perform intercross with two inbred strains.
- Perform several generations of random mating, avoiding inbreeding.
- Genotype and phenotype individuals from the final generation.

Advanced intercross lines



AIL

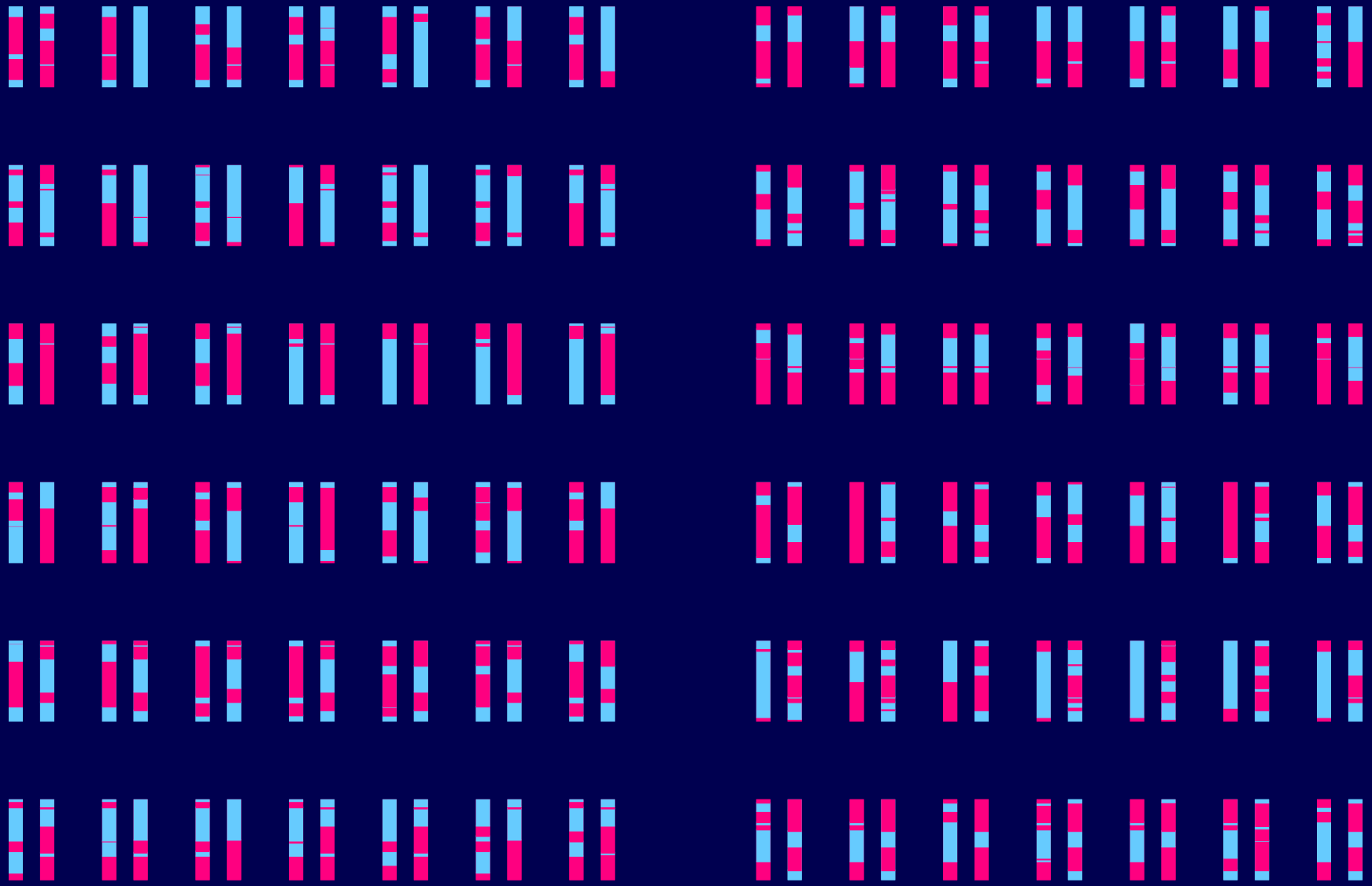
Advantages

- Many more breakpoints $\Longrightarrow$ more precise mapping
- Straightforward to create

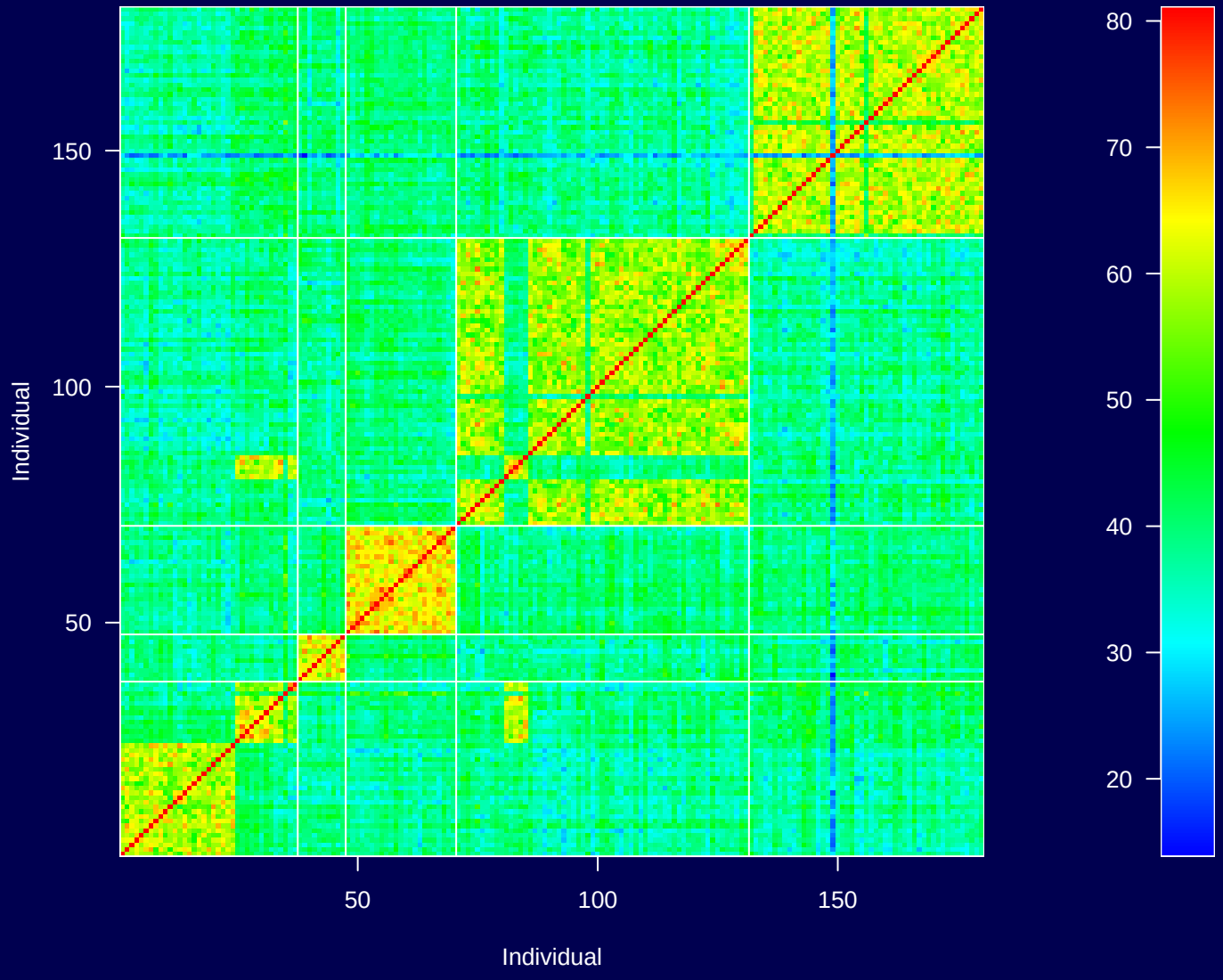
Disadvantages

- Time and cost
- Each individual genetically distinct
- Useful largely for fine-mapping known loci
- Relationships among individuals in final generation

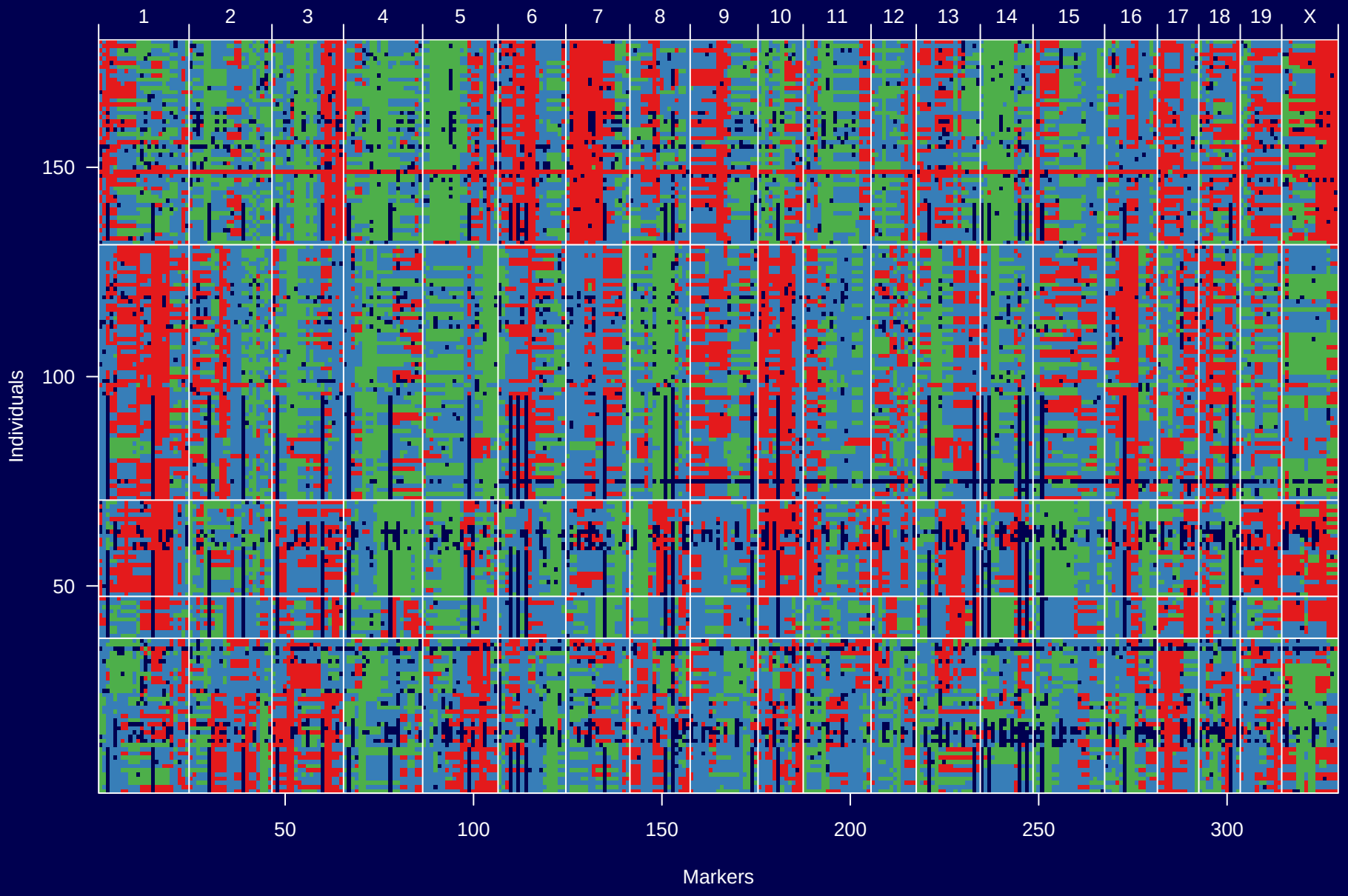
Sibships at F_8



AIL: percent matching genotypes



AIL: genotype data



Analysis of AIL

- One generally treats the AIL as an intercross, but with an expanded genetic map. [Longer map $\Longrightarrow$ larger significance threshold.]

- However, the relationships among the AIL individuals makes analysis tricky.

- To account (at least) for the sibships at the final generation:

$$y_{ij} = \mu + \beta q_{ij} + \epsilon_{ij}$$

  for sib j in sibship i, with $\text{cov}(\epsilon_{ij}, \epsilon_{ij'}) = \tau^2$ and $\text{var}(\epsilon_{ij}) = \tau^2 + \sigma^2$.

- Standard permutation tests are no longer appropriate.

  If possible, reconstruct the parental haplotypes (for the previous generation) and permute them. Alternatively, simulate the whole pedigree.

Analysis of AIL

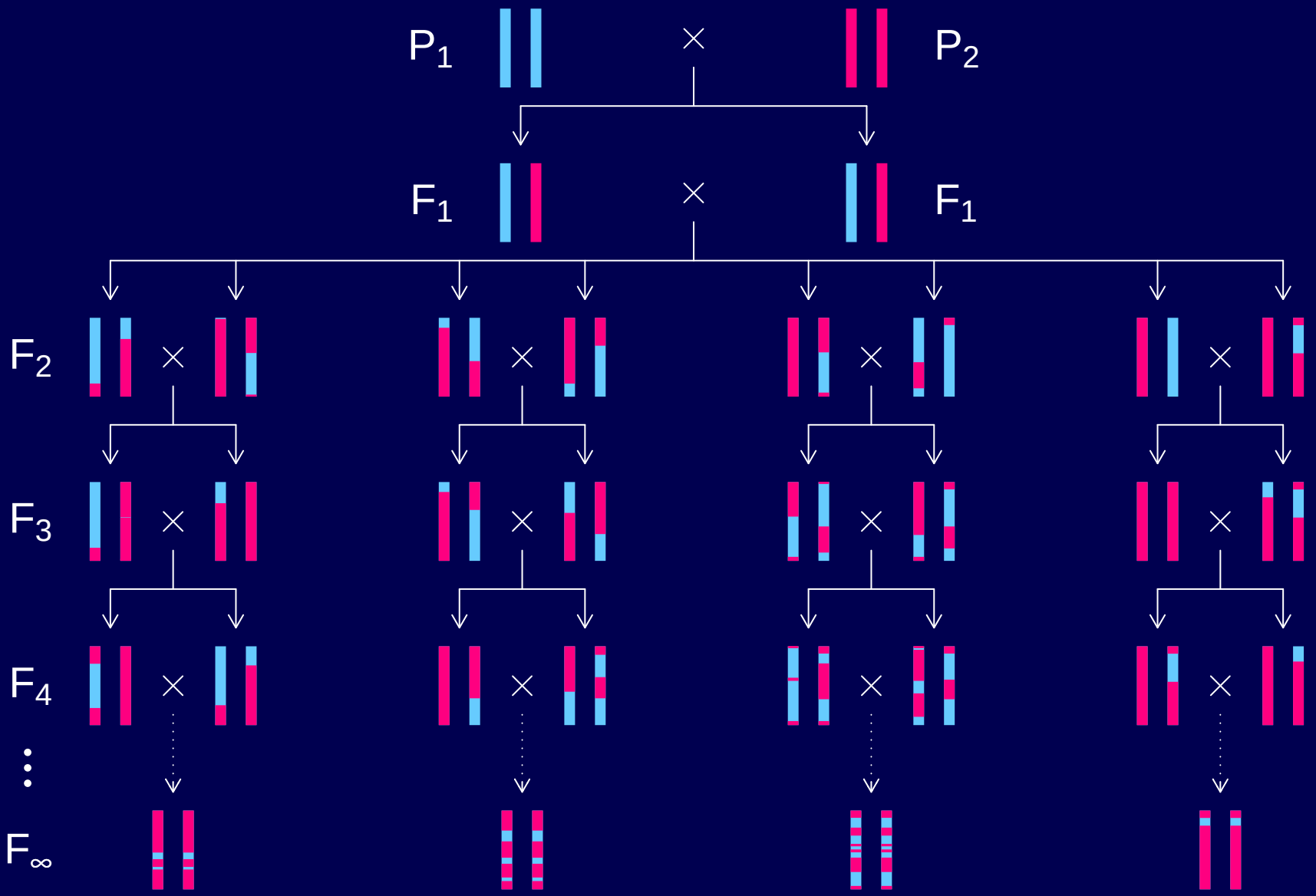
Imagine numerous, small-effect, unlinked, additive, background QTL.

$$\begin{aligned} y_{ij} &= \mu + \beta q_{ij} + \sum_k \alpha_k a_{ijk} + \epsilon_{ij} \\ &= \mu + \beta q_{ij} + \gamma_{ij} + \epsilon_{ij} \end{aligned}$$

γ_{ij} = combined effect of all background QTL

$\longrightarrow$ correlated between siblings (or half-siblings)

Recombinant inbred lines



RIL

Advantages

- High density of breakpoints
- Just genotype once
- Phenotype multiple individuals to reduce environmental/individual variation
- Multiple phenotypes on the same genomes
- Longitudinal phenotypes
- Genotype $\times$ environment interactions

Disadvantages

- Time-consuming, expensive to create
- Available panels generally too small
- Only homozygotes

Analysis of RIL

Available methods for analysis of RIL are rather rudimentary:

Treat like a backcross, working with the line averages.

Map expansion: $R = 4r/(1 + 6r)$

Let y_{ij} be the phenotype for individual j in line i.

Let $x_i = 1$ or 0, genotype of line i at putative QTL.

Assume $y_{ij} = \mu + \beta x_i + \gamma_i + \epsilon_{ij}$ where $\epsilon_{ij} \sim N(0, \sigma_e^2)$, $\gamma_i \sim N(0, \sigma_p^2)$.

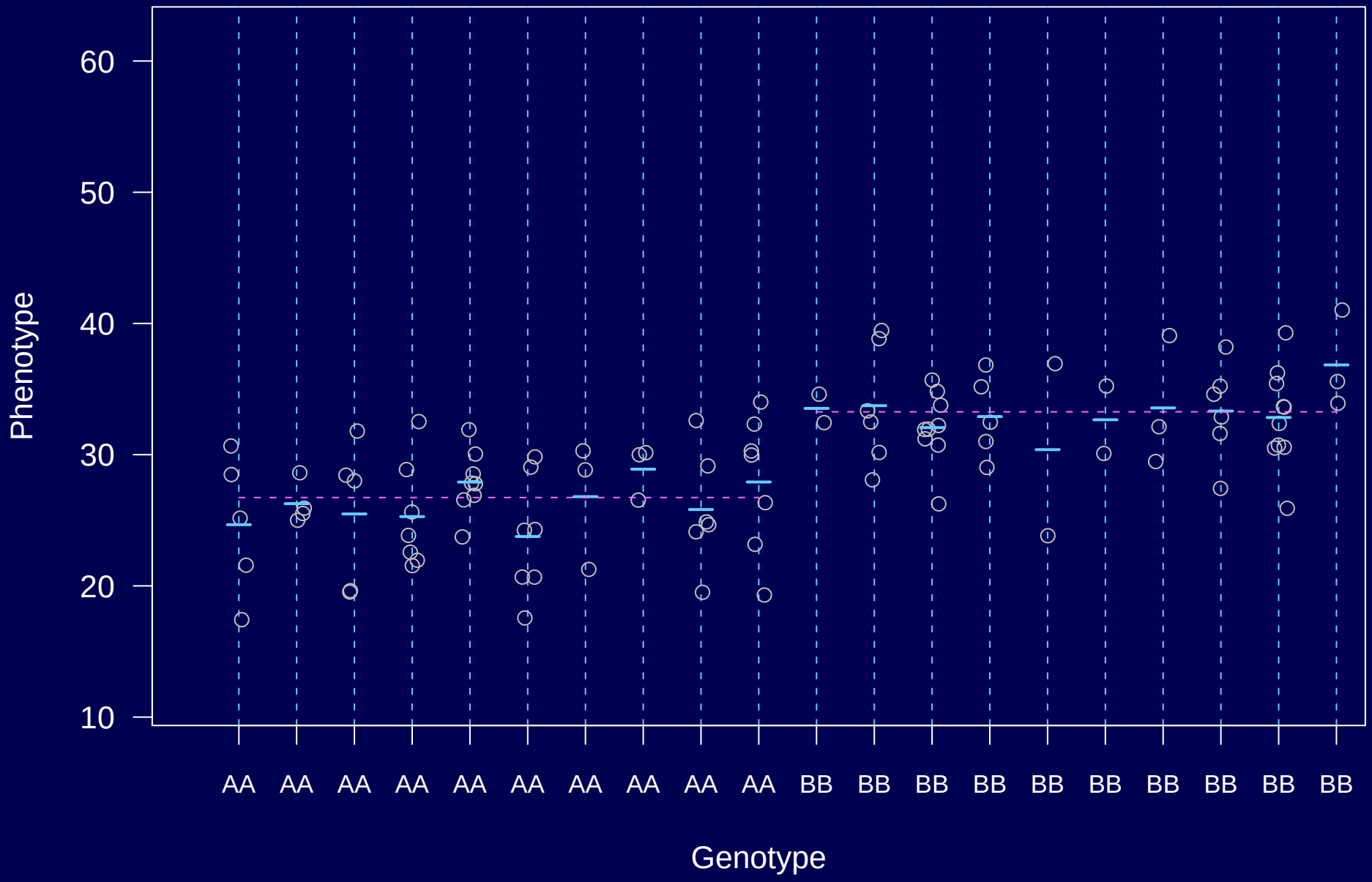
Residual heritability: $h_p^2 = \sigma_p^2/(\sigma_p^2 + \sigma_e^2)$

Then we can work with the strain averages,

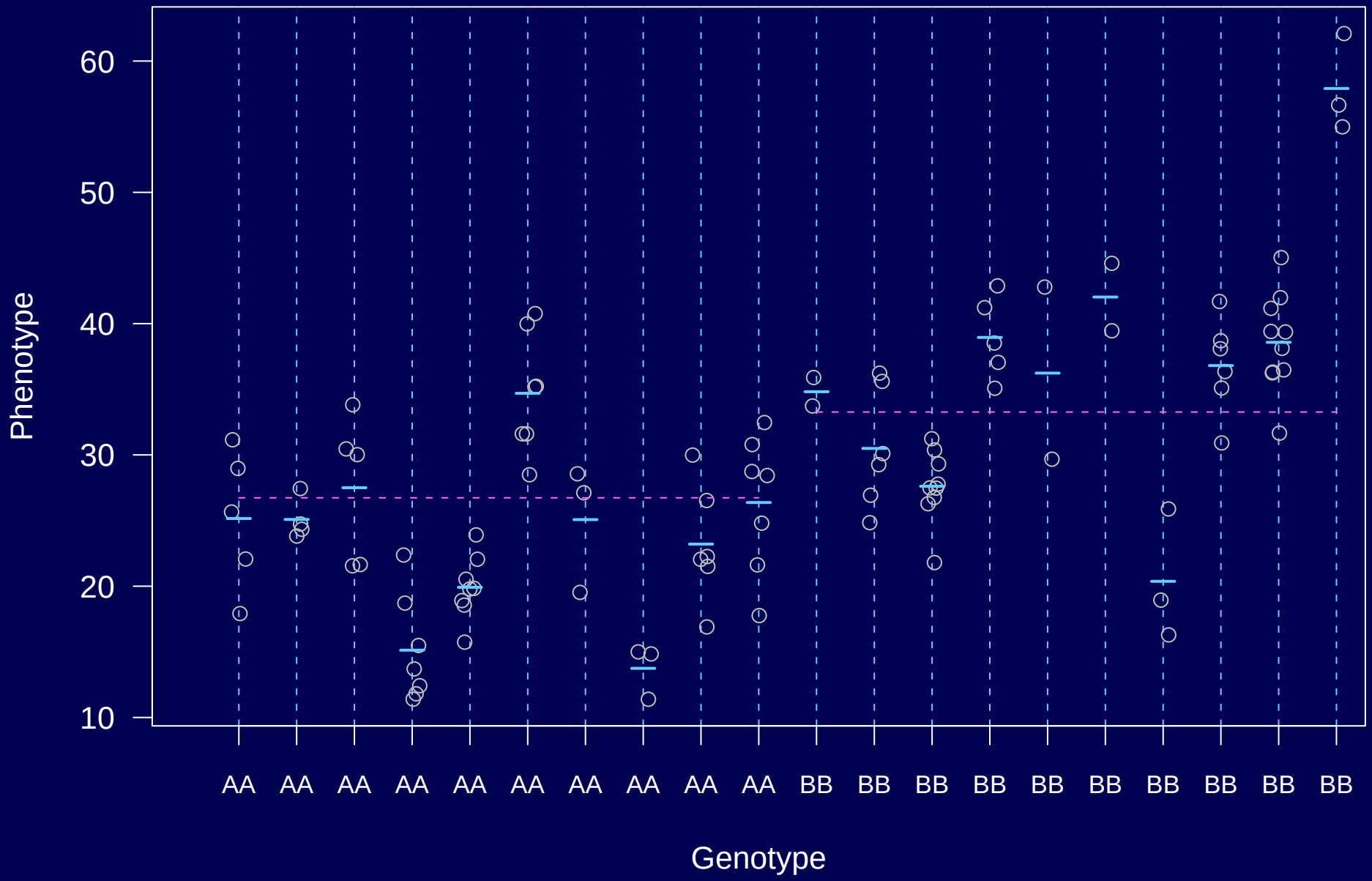
but they should be weighted by the sample sizes.

$\text{var}(\bar{y}_i) = \sigma_p^2 + \sigma_e^2/n_i$ $\Longrightarrow$ weight inversely proportional to $h_p^2 + (1 - h_p^2)/n_i$

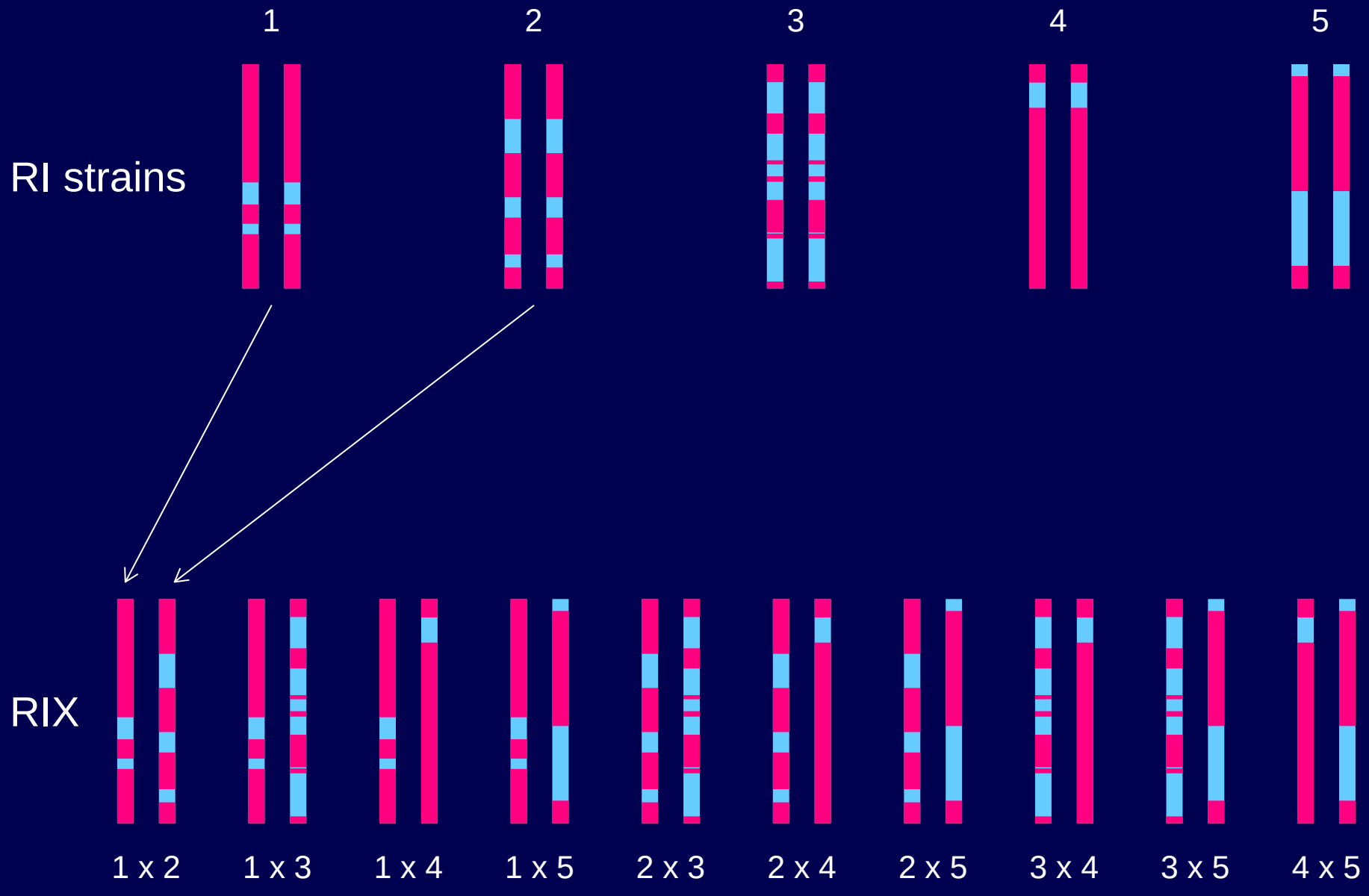
RIL: One QTL


Phenotype
60
50
40
30
20
10
AA AA AA AA AA AA AA AA AA AA BB BB BB BB BB BB BB BB BB BB
Genotype

RIL: Many QTL


Phenotype
60
50
40
30
20
10
AA AA AA AA AA AA AA AA AA AA BB BB BB BB BB BB BB BB BB BB
Genotype

RIX



Analysis of RIX

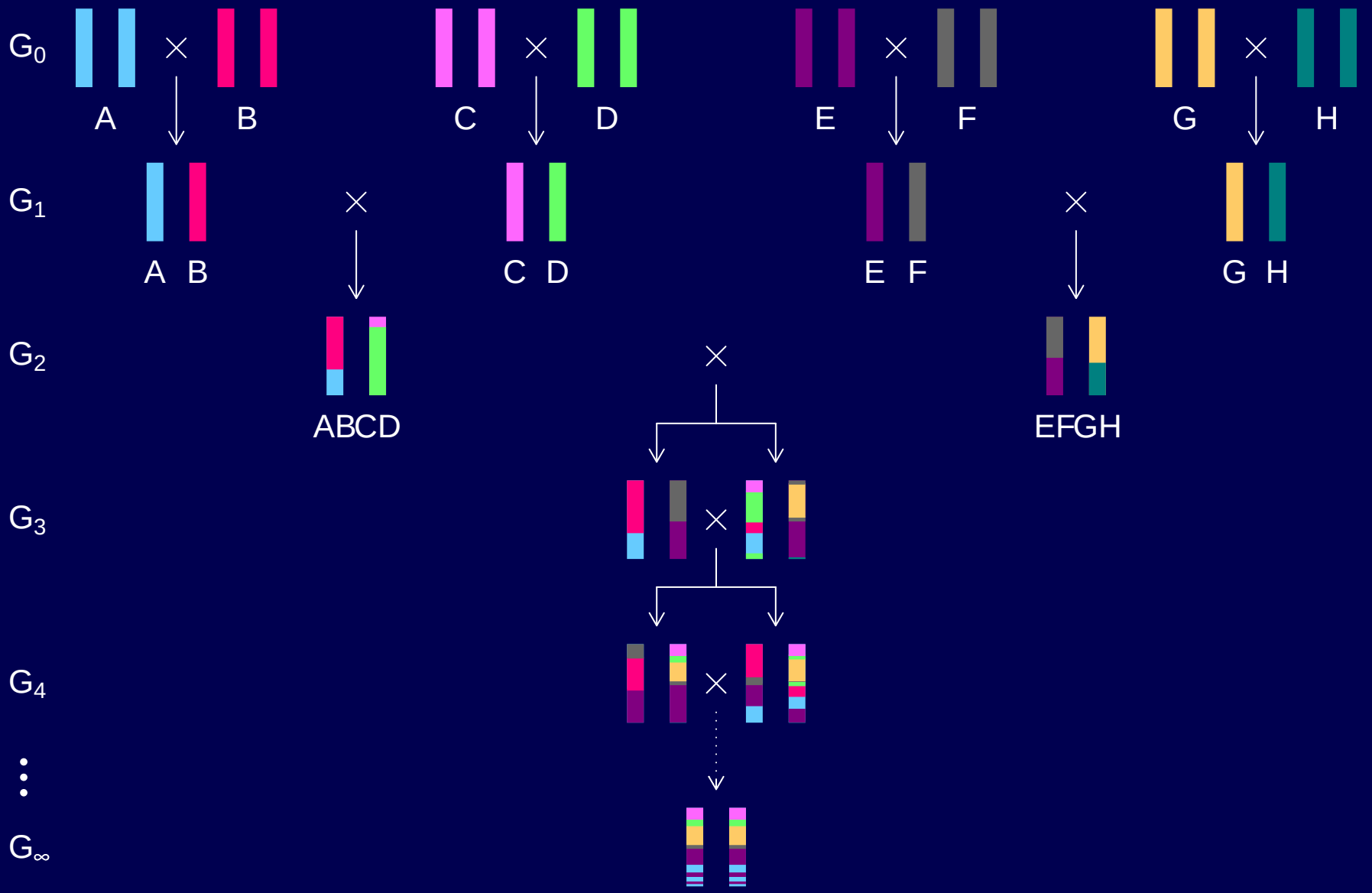
- Correlations among RIX: shared paternal/maternal chromosomes

  $\Longrightarrow$ Mixed-effects model, as with AIL (but it's more clear here).

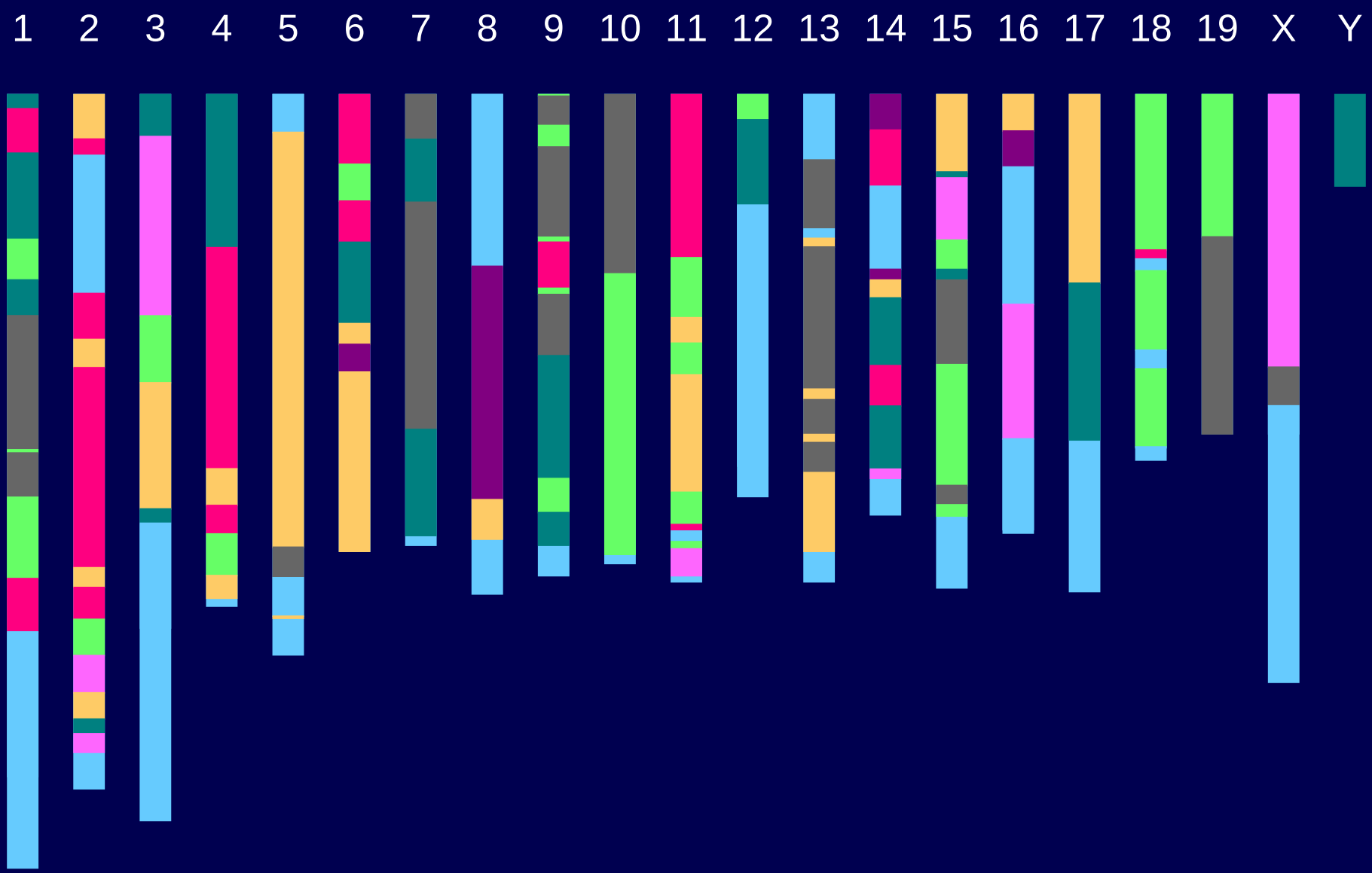
- Permutation test:

  Permute the parental genotypes and pass them down to the RIX.

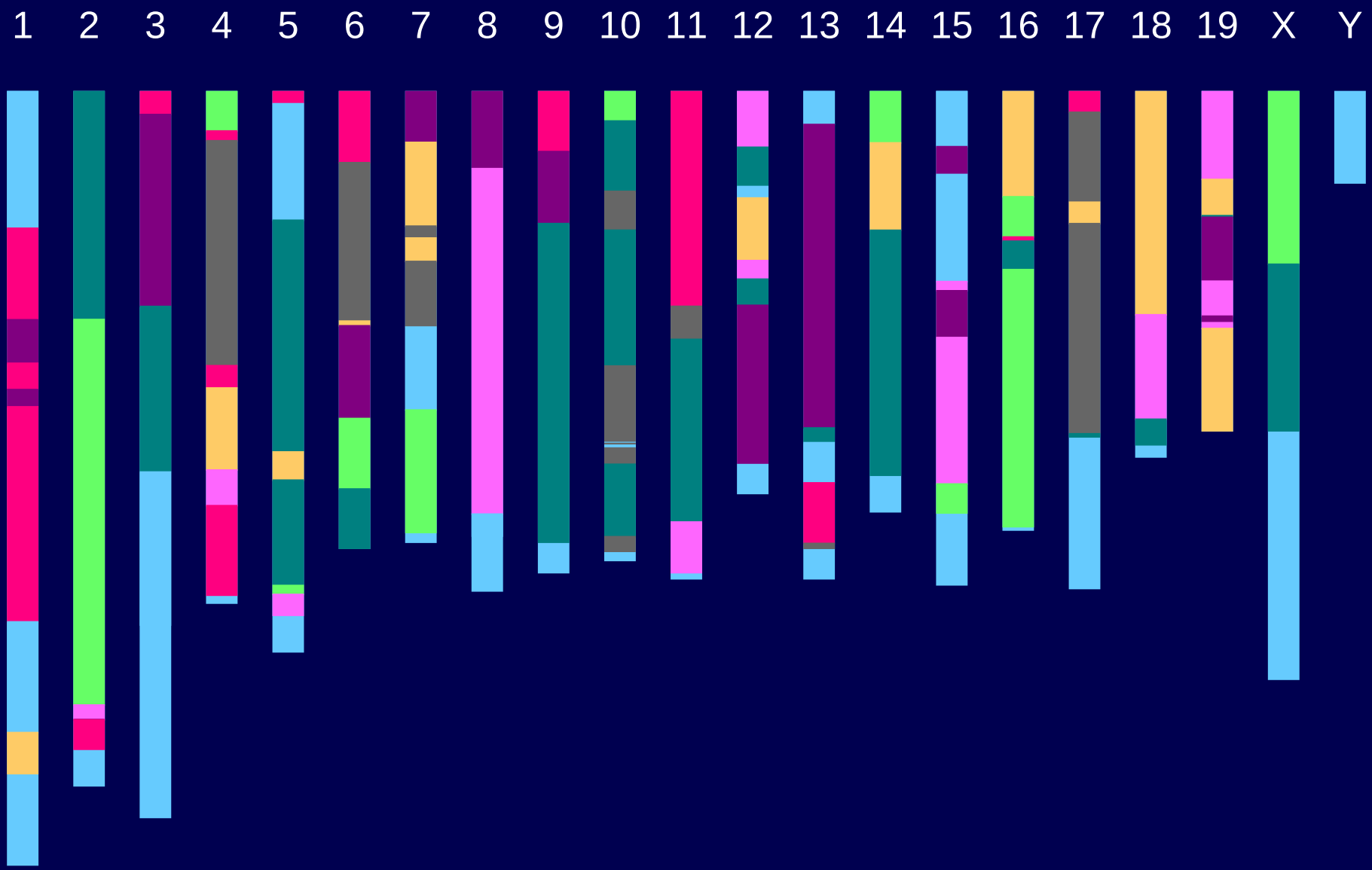
Collaborative Cross



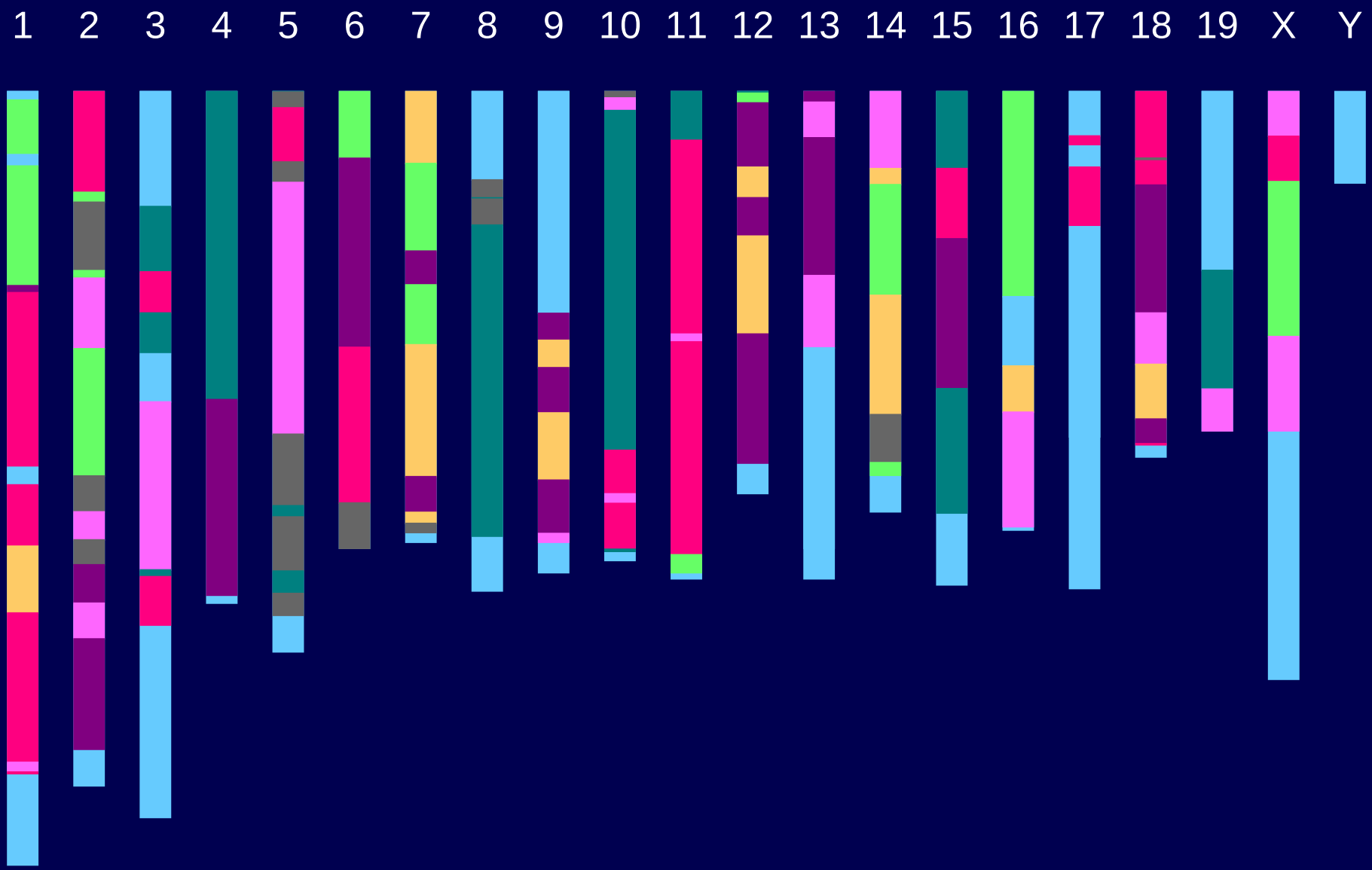
CC genome



CC genome



CC genome



Analysis of CC

- First, reconstruct haplotypes.

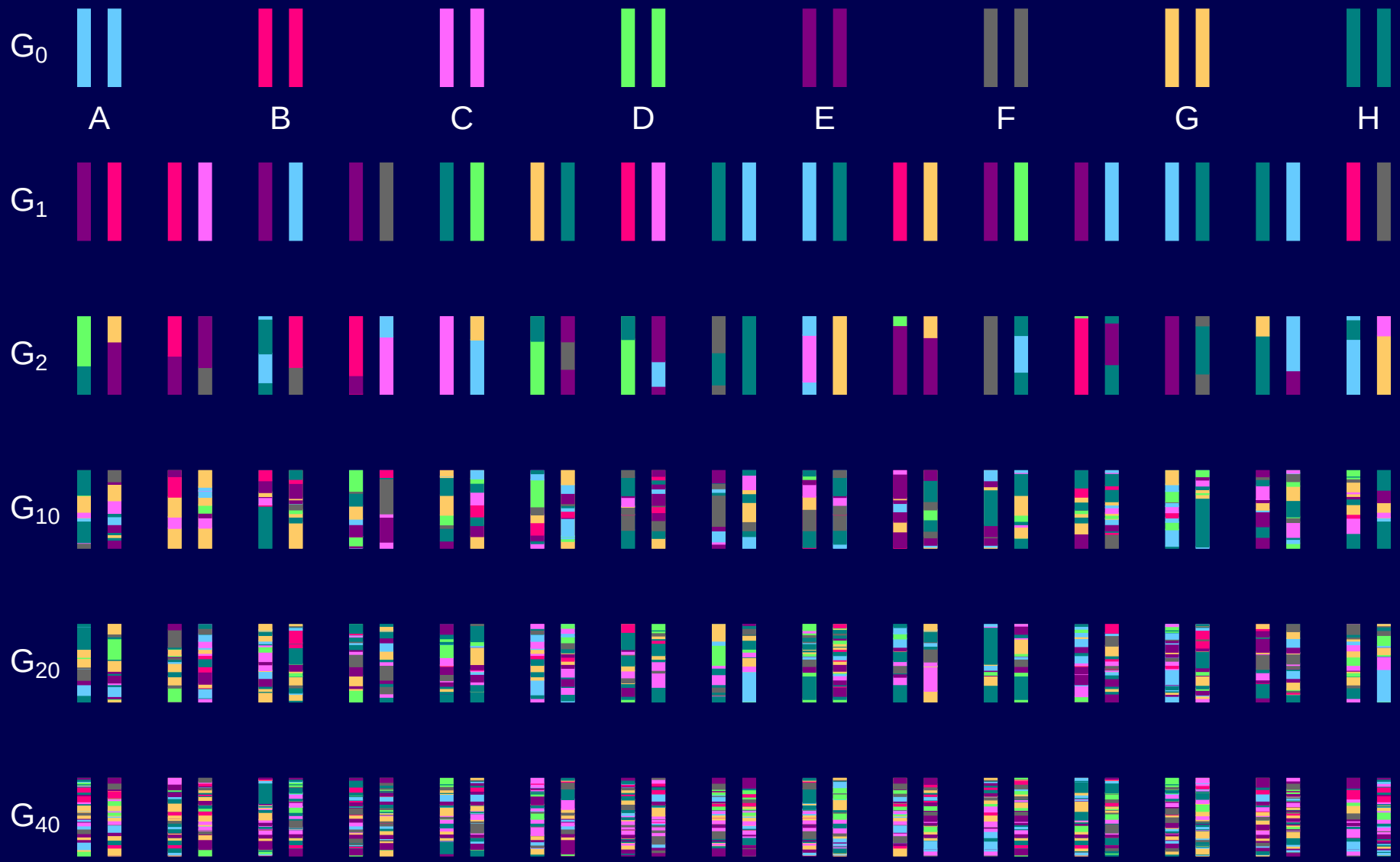
  Map expansion: $R = 7r/(1 + 6r)$

- Further analysis just like RIL, but now 8 alleles.

- Epistasis: consider 64 possible two-locus genotypes.

  Random effects model?

Heterogeneous stock



HS

Advantages

- Super-dense breakpoints
- Many alleles
- Heterozygous

Disadvantages

- Must be satisfied with what is available
- Inbreeding: loss of alleles
- Each individual unique
- Like AIL, maybe best for fine-mapping known loci
- Like AIL, relationships at last generations

Analysis of HS

- Reconstruction of haplotypes
- Treatment of the 8 alleles (and so 36 genotypes)
  - Additive alleles
  - Random effect
- Dealing with the relationships
- Establishing statistical significance?

Association mapping

- Phenotype available inbred strains
- Make use of available SNP data

- Need to account for the correlations among strains
- Likely want to work with haplotypes rather than just individual SNPs
- Be careful about wild-derived strains

Association mapping

Advantages

- Once you've done a strain survey, no further data needed
- Potentially very high resolution

Disadvantages

- All the usual problems with association mapping
- Power is unpredictable
- How to account for relationships among strains?

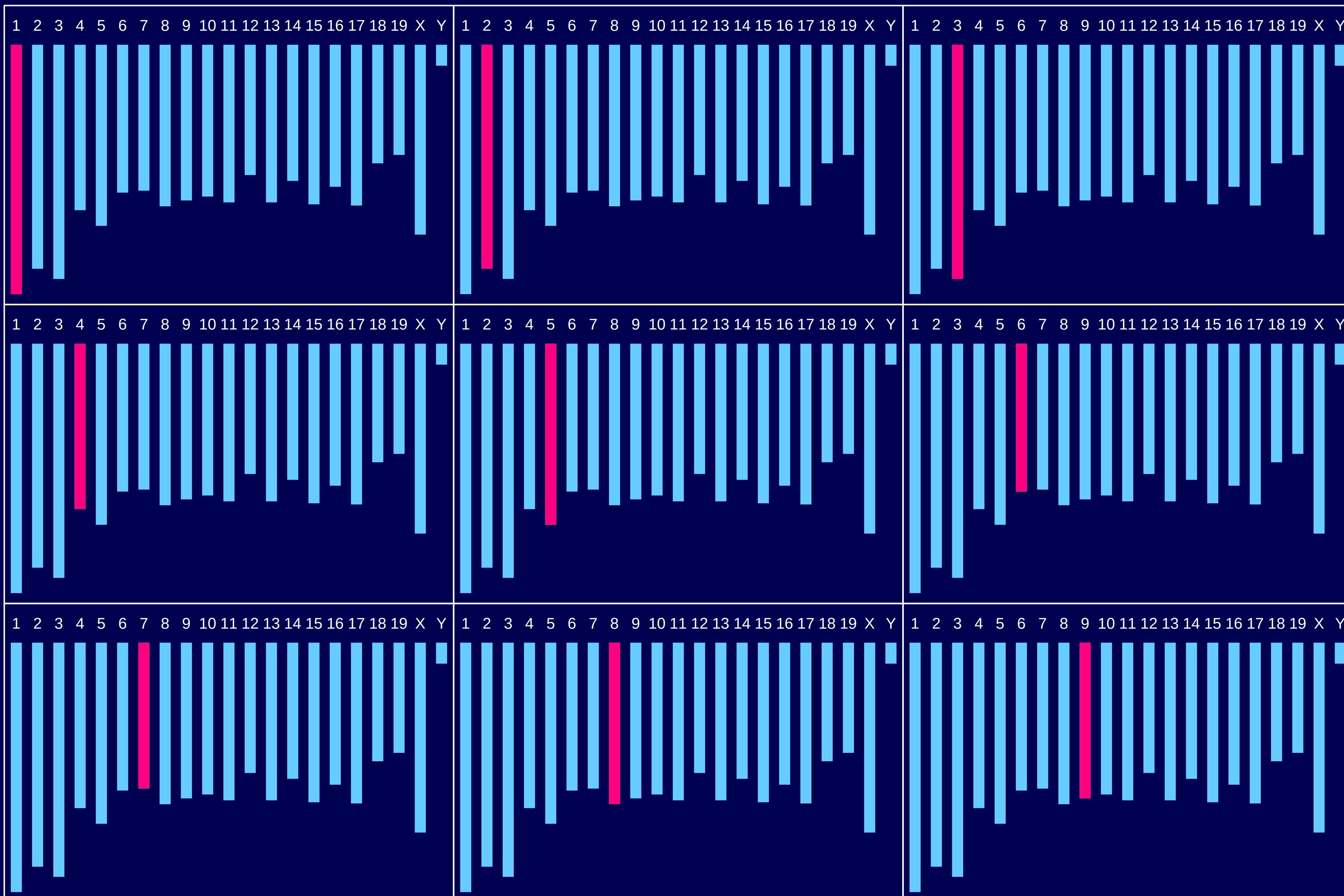
CC vs HS vs association mapping

These approaches have many similarites.

Key differences:

- CC, HS: pattern of association along chromosomes by design
- HS: each individual unique

Chromosome substitution strains



CSS

Advantages

- Just phenotyping can get you to the chromosomes
- Eliminate the effects of other QTL
- Easy to create congenics

Disadvantages

- Time-consuming, expensive to create
- Lots of phenotyping required
- Cannot see interactions

Analysis of CSS

- Compare each CSS to recipient strain via t test.
- Account for multiple (dependent) tests
  - Easy with a permutation test

Summary

- Traditional approach
- Advanced intercross lines (AIL)
- Recombinant inbred lines (RIL)
- RIX
- Collaborative cross (CC)
- Heterogeneous stock (HS)
- Association mapping
- Chromosome substitution strains (CSS)

References

- Silver L (1995) Mouse Genetics: Concepts and applications. Oxford University Press. (`http://www.informatics.jax.org/silver`)

  A book everyone doing mouse genetics should own.

- Complex Trait Consortium (2003) The nature and identification of quantitative trait loci: a community's view. Nat Rev Genet 4:911–916

  Describes a view on sufficient proof that a candidate locus is resposible for the effect of a QTL.

- Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199–1207

  Proposed AIL.

- Belknap JK (1998) Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. Behav Genet 28:29–38

  Discusses choice of number of individuals per strain to phenotype in RIL.

- Threadgill DW, Hunter KW, Williams RW (2002) Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. Mamm Genome 13:175–178

  Proposes the collaborative cross; discusses RIX.

- Zou F, et al. (2005) Quantitative trait locus analysis using recombinant inbred intercrosses (RIX): theoretical and empirical considerations. Genetics 170:1299–1311

  Analysis issues concerning RIX.

- Tsaih SW, et al. (2005) Quantitative trait mapping in a diallel cross of recombinant inbred lines. Mamm Genome 16:344–355

  Analysis issues concerning RIX.

- Complex Trait Consortium (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nature Genetics 36:1133–1137

  Advocates Collaborative Cross.

- Broman KW (2005) The genomes of recombinant inbred lines. Genetics 169:1133–1146

  Describes map expansion in multiple-strain RIL.

- Mott R, Flint J (2002) Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. Genetics 160:1609–1618

  Describes use of HS; fine-mapping application.

- Valdar W, et al. (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat Genet 38:879–887

  High-density genome-wide scans with HS.

- Grupe A, et al. (2001) In silico mapping of complex disease-related traits in mice. Science 292:1915–1918

  Proposes the use of association mapping in mice.

- Pletcher MT, et al. (2004) Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. PLoS Biol 2:e393

  Further improvements regarding association mapping in mice.

- Wade CM, Daly MJ (2005) Genetic variation in laboratory mice. Nat Genet 37:1175–1180

  Patterns of linkage disequilibrium in mice.

- Nadeau JH, Singer JB, Matin A, Lander ES (2000) Analysing complex genetic traits with chromosome substitution strains. Nat Genet 24:221–225

  Proposes the use of CSS.

- Singer JB, et al. (2004) Genetic dissection of complex traits with chromosome substitution strains of mice. Science 304:445–448

  Application of CSS.

- Belknap JK (2003) Chromosome substitution strains: some quantitative considerations for genome scans and fine mapping. Mamm Genome 14:723–732

  Analysis issues in CSS.