

Multiparent populations & R/qt12

Karl Broman

Biostatistics and Medical Informatics
University of Wisconsin – Madison

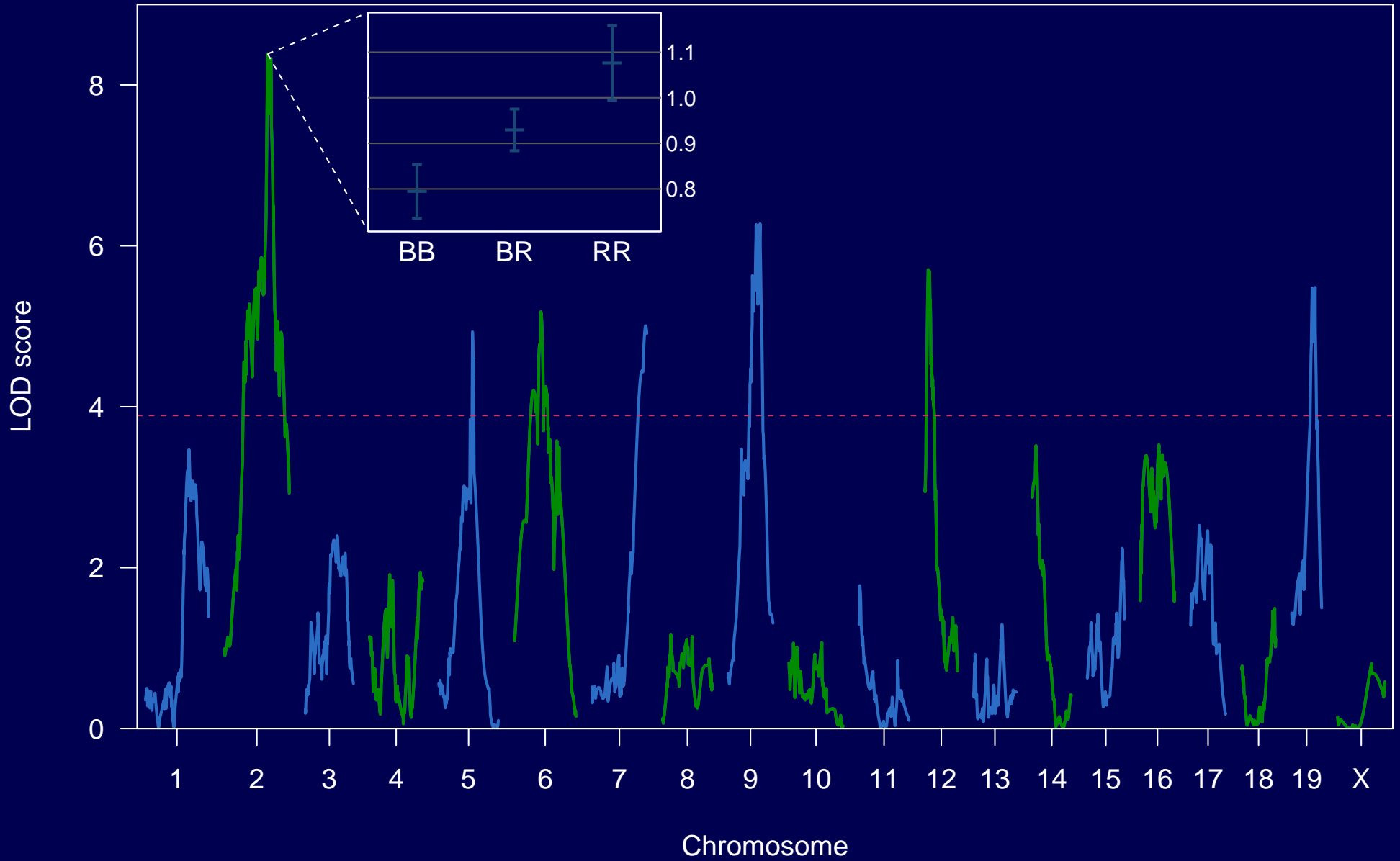
kbroman.org/qt12

kbroman.org

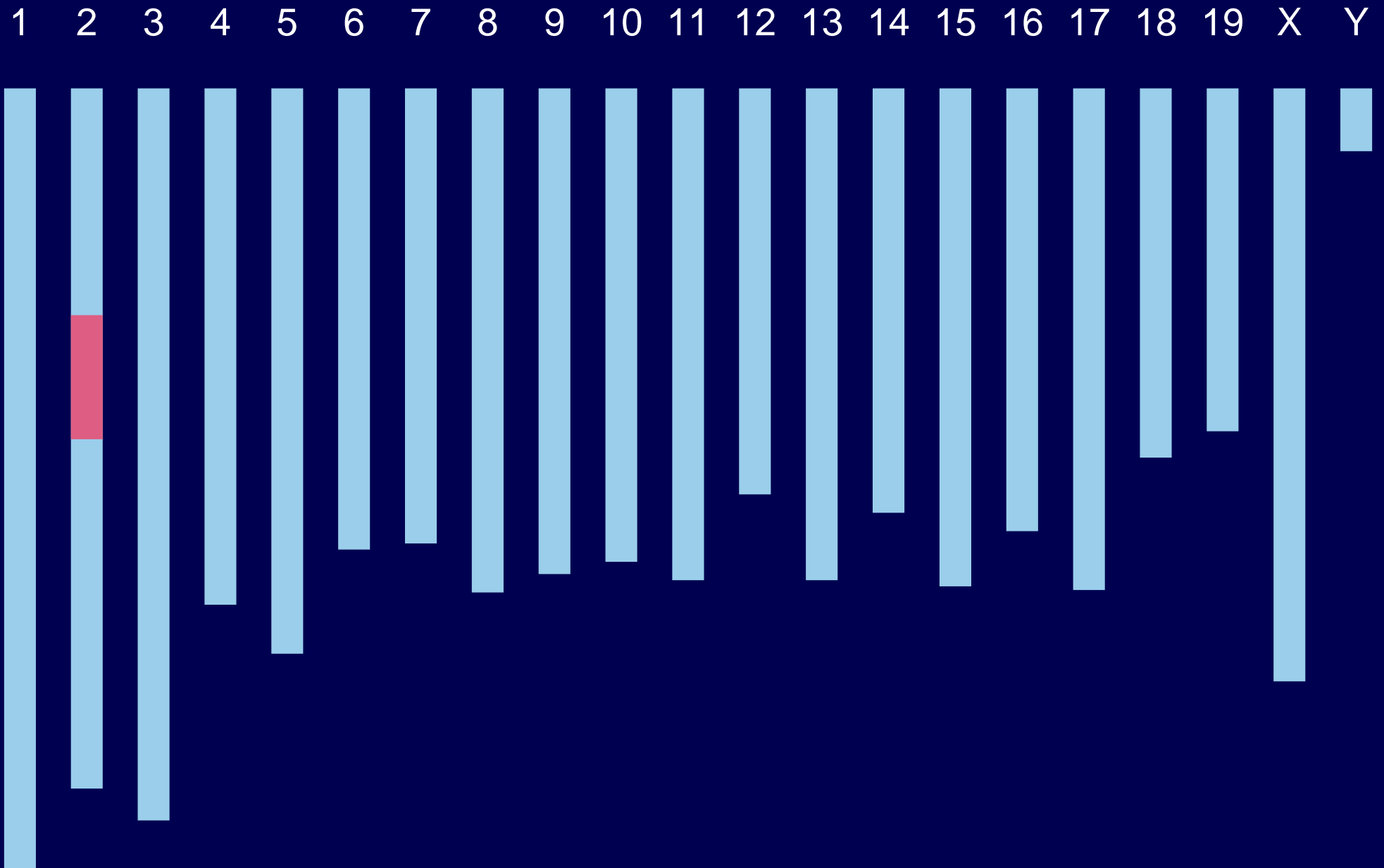
github.com/kbroman

@kwbroman

QTL mapping



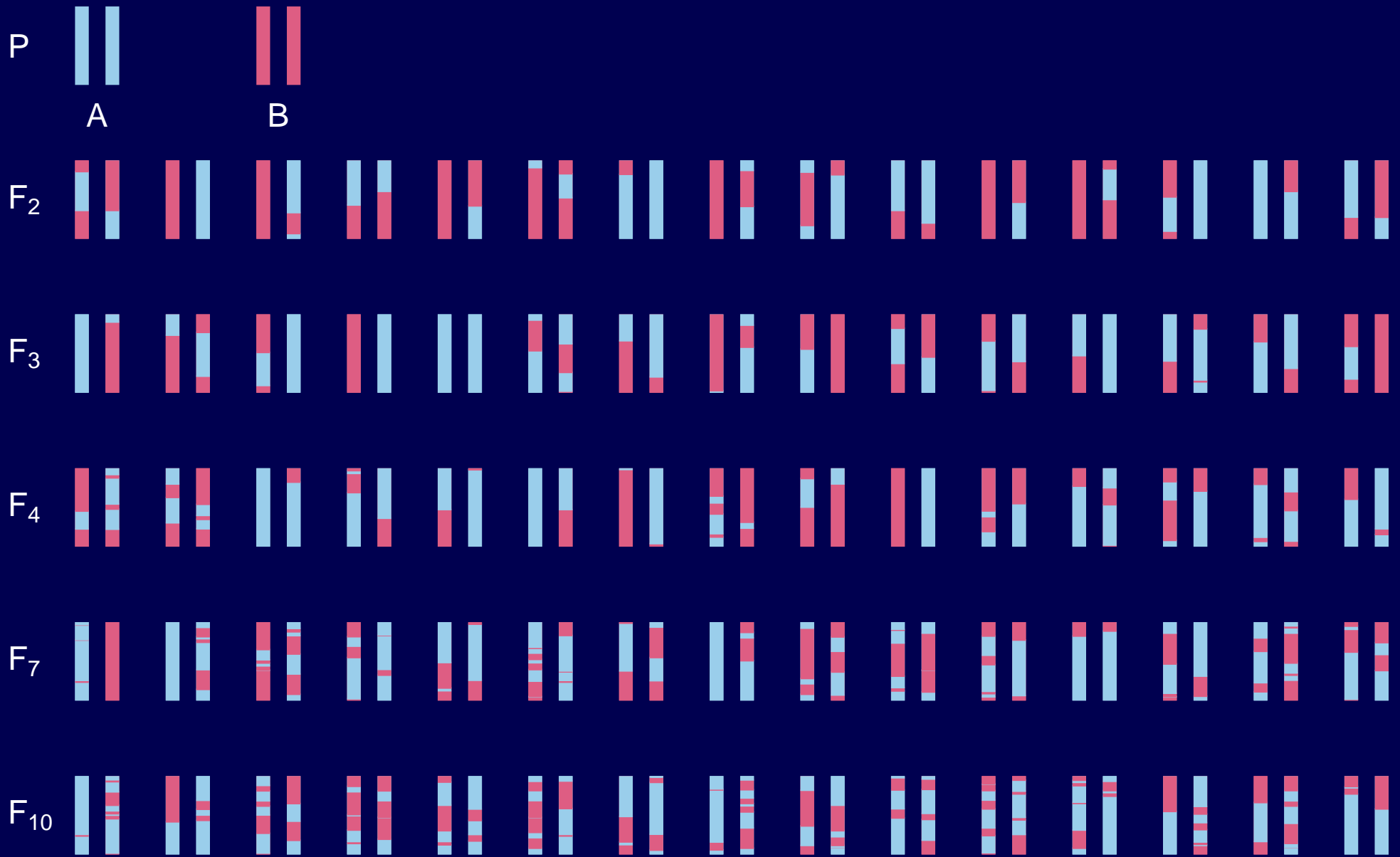
Congenetic line / NIL



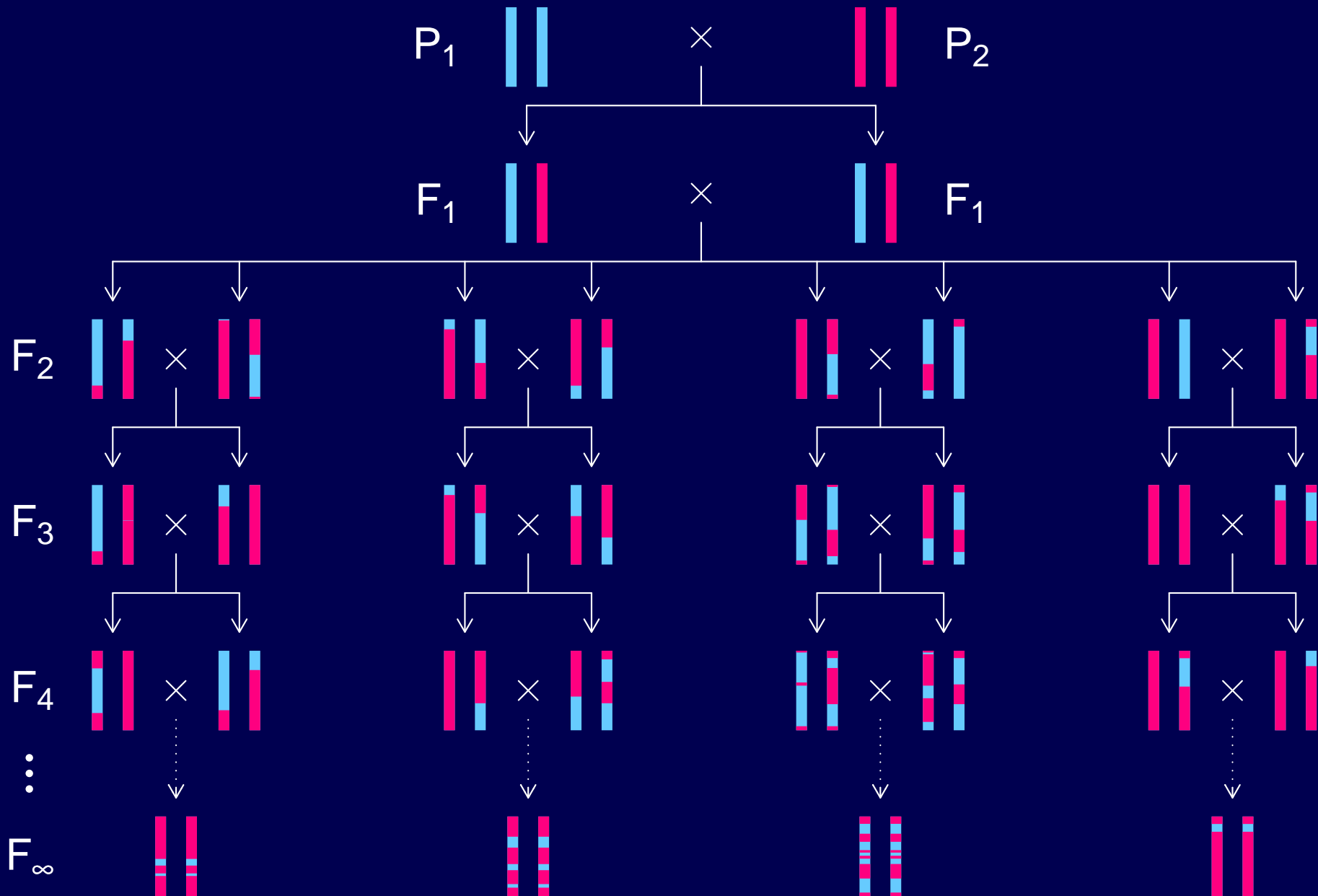
Improving precision

- more recombinations
- more individuals
- more precise phenotypes
- lower-level phenotypes
transcripts, proteins, metabolites

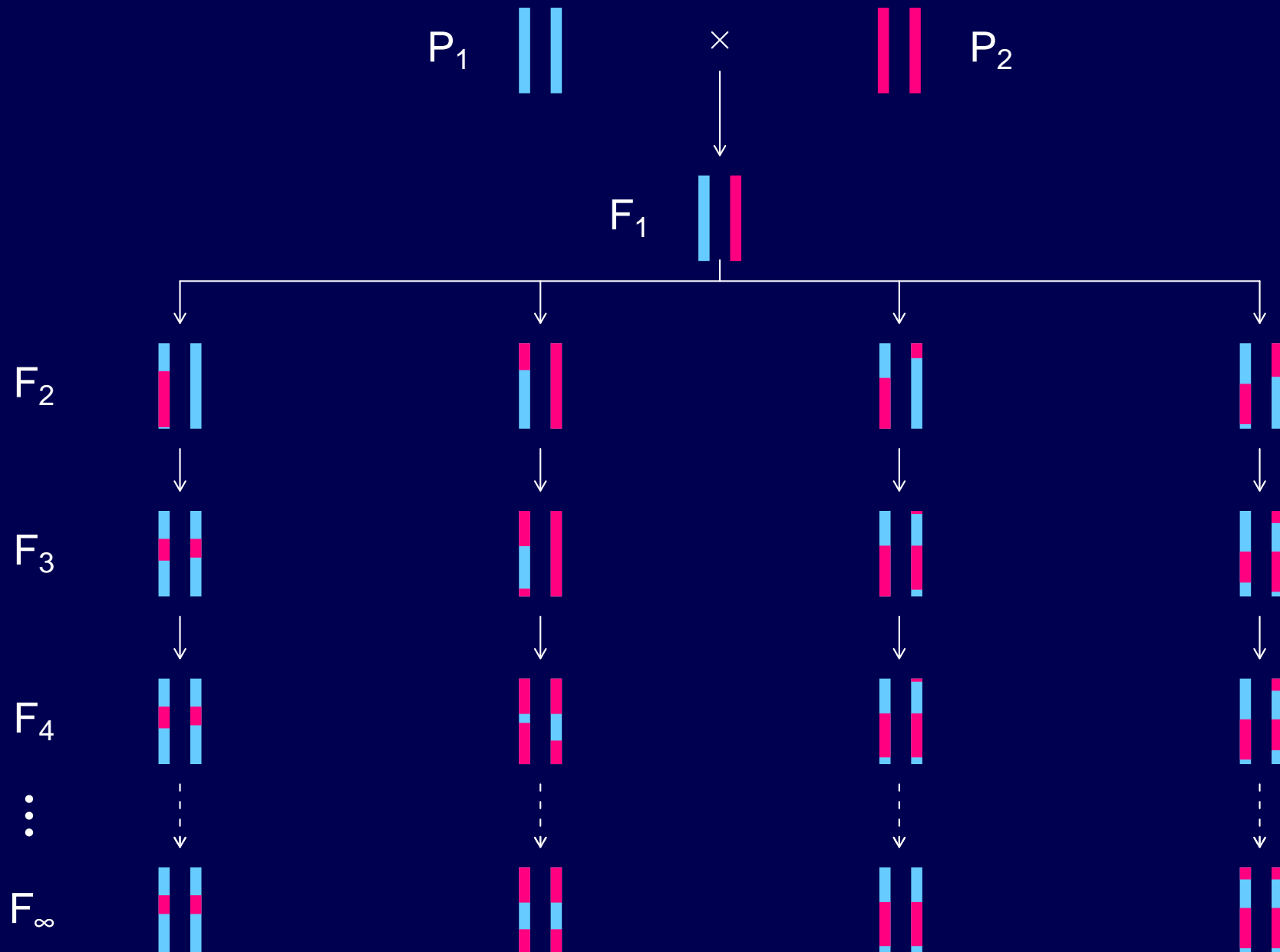
Advanced intercross lines



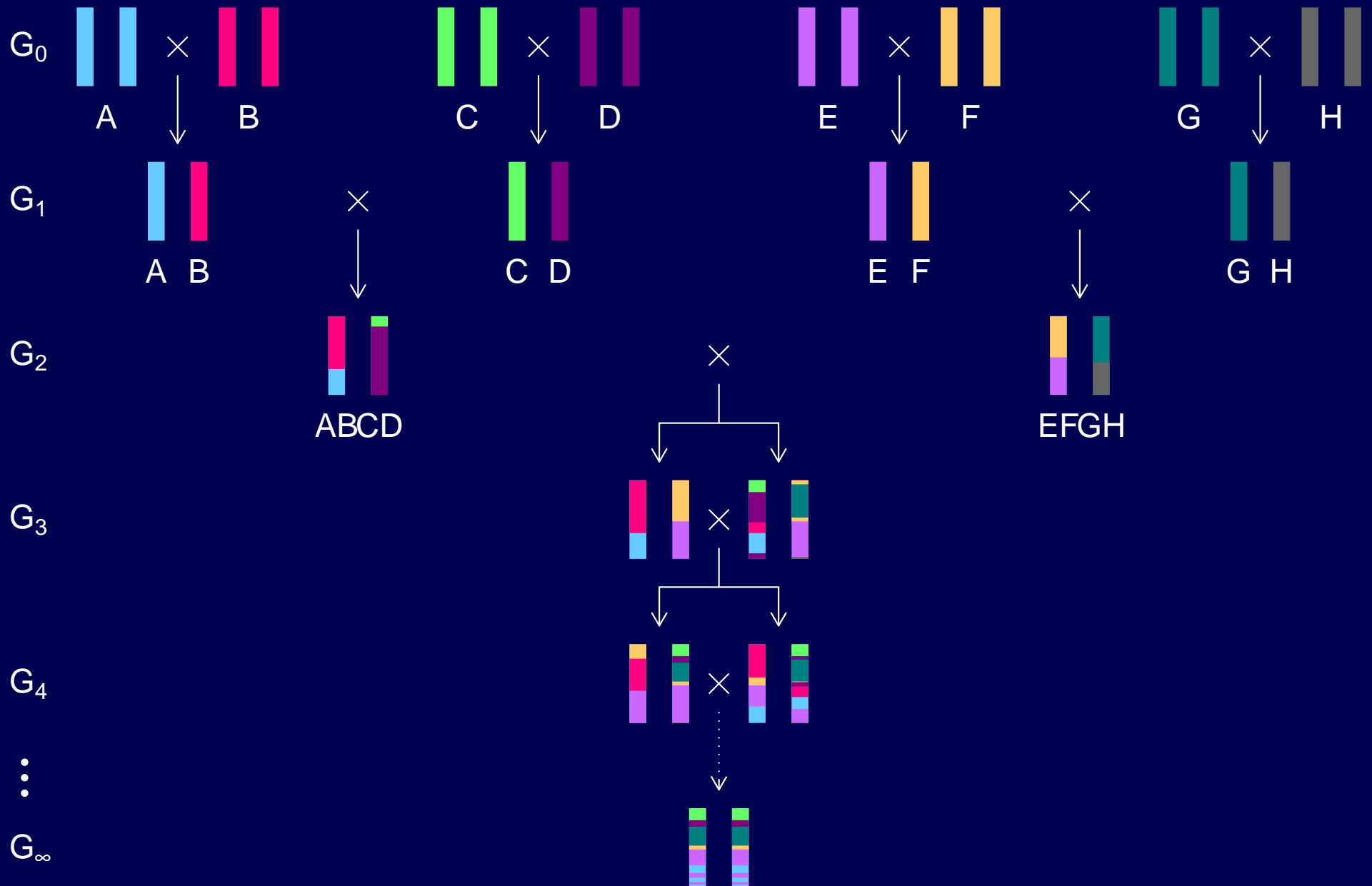
Recombinant inbred lines



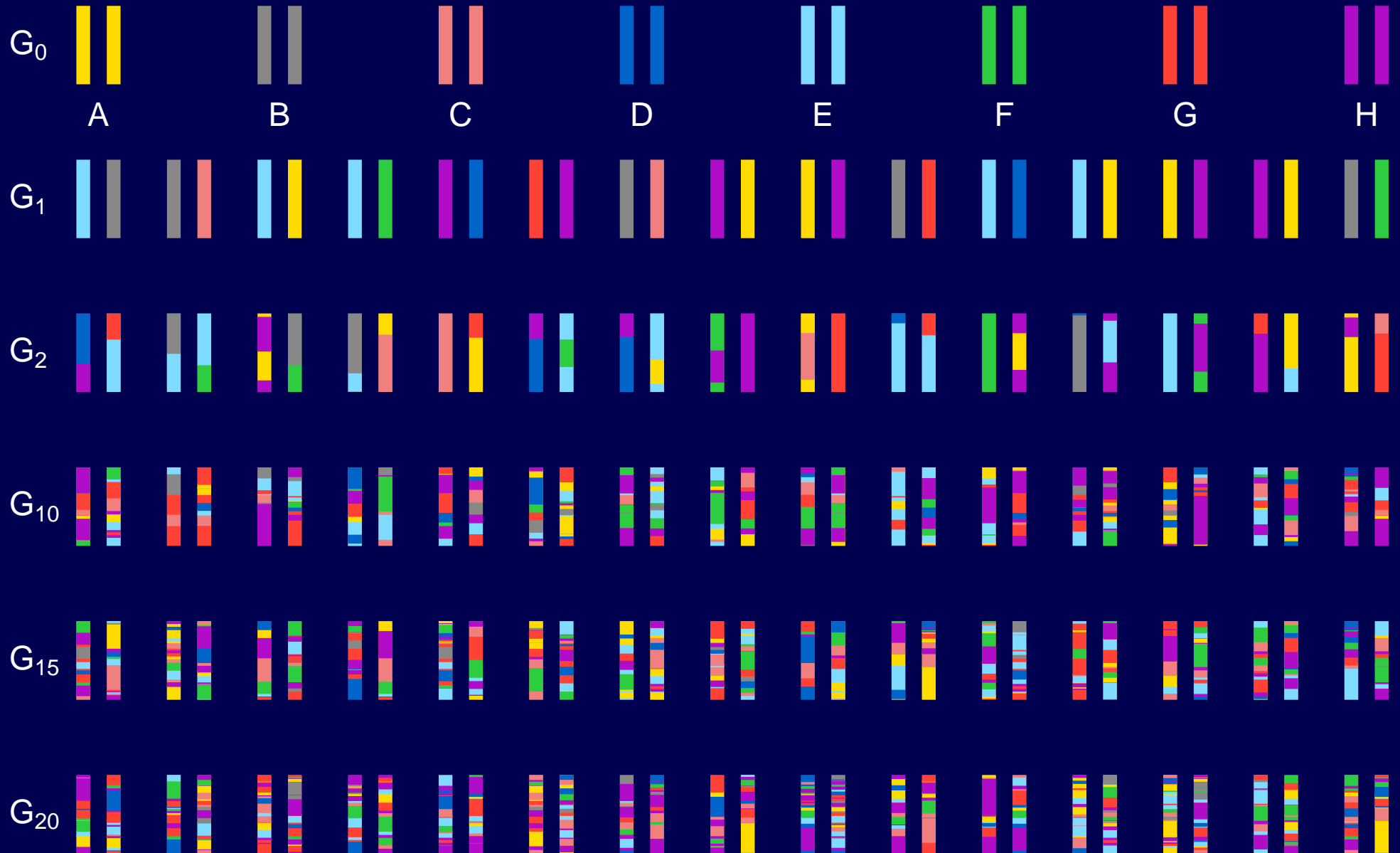
Recombinant inbred lines



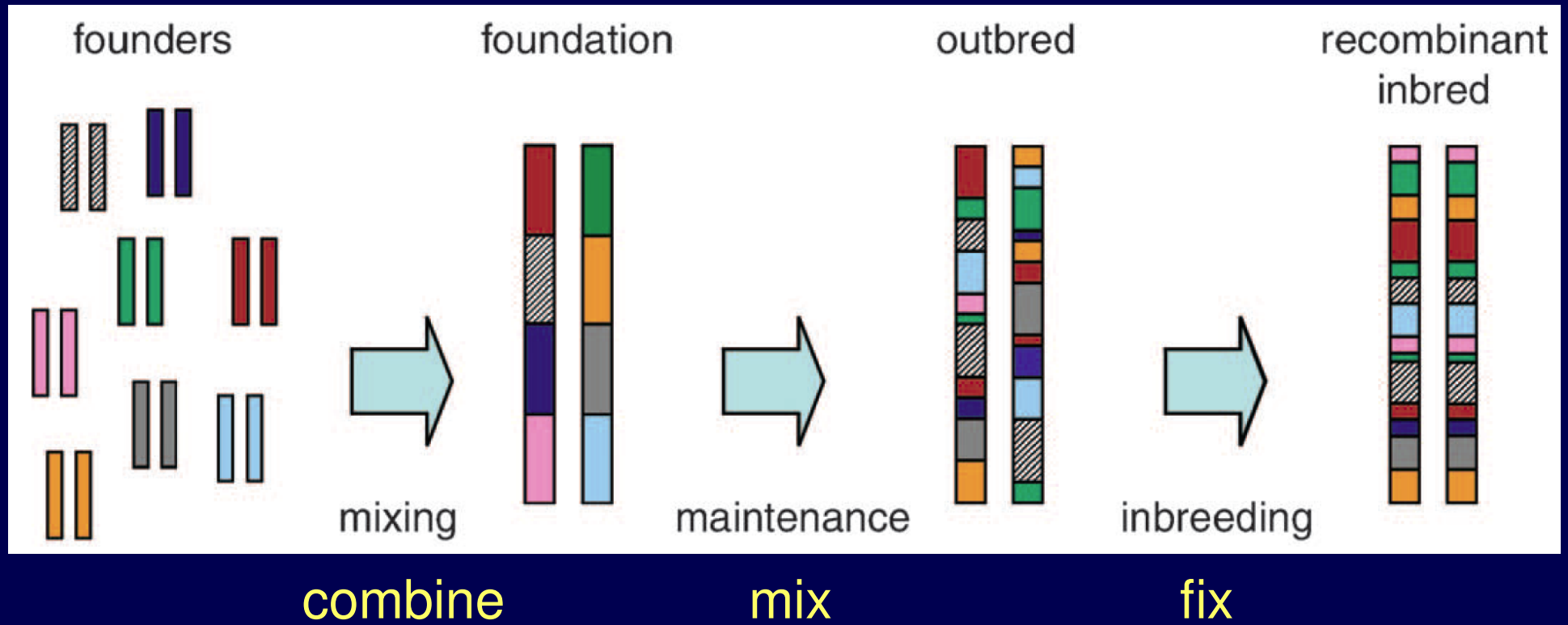
Collaborative Cross/MAGIC



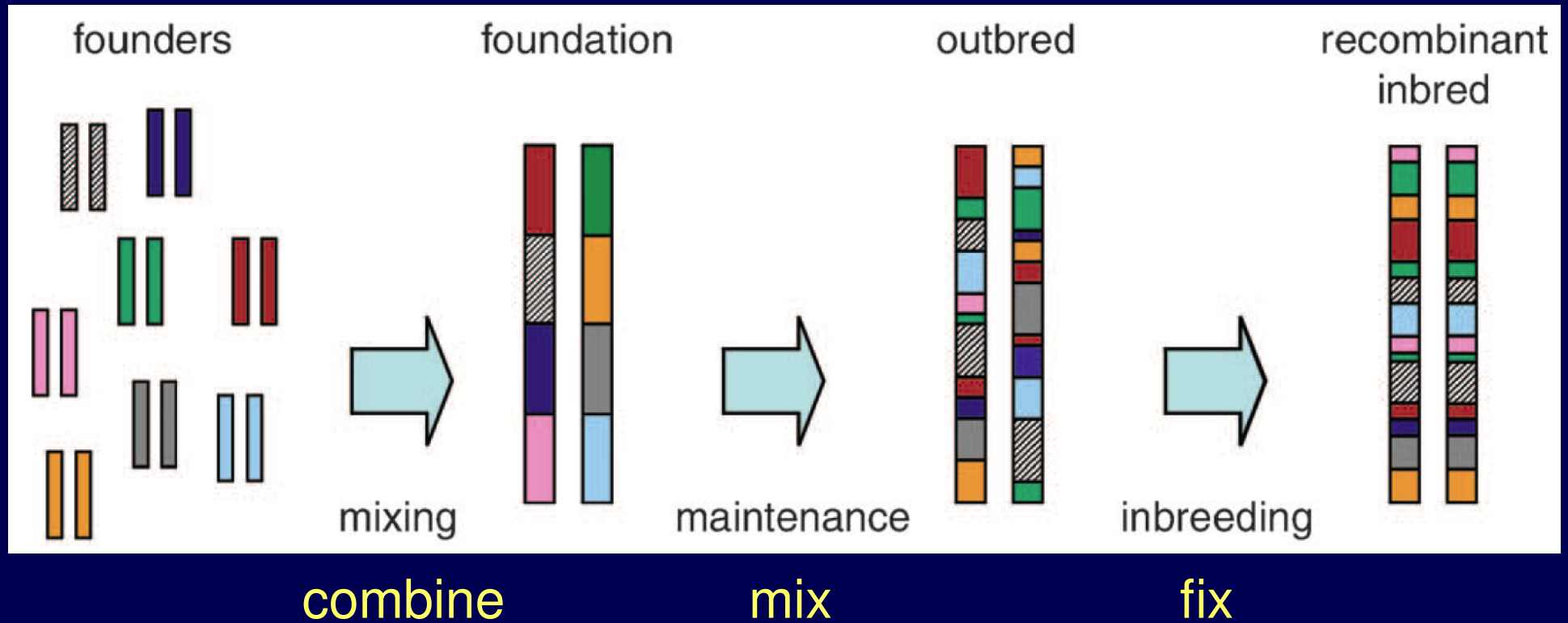
Heterogeneous stock



MAGIC lines

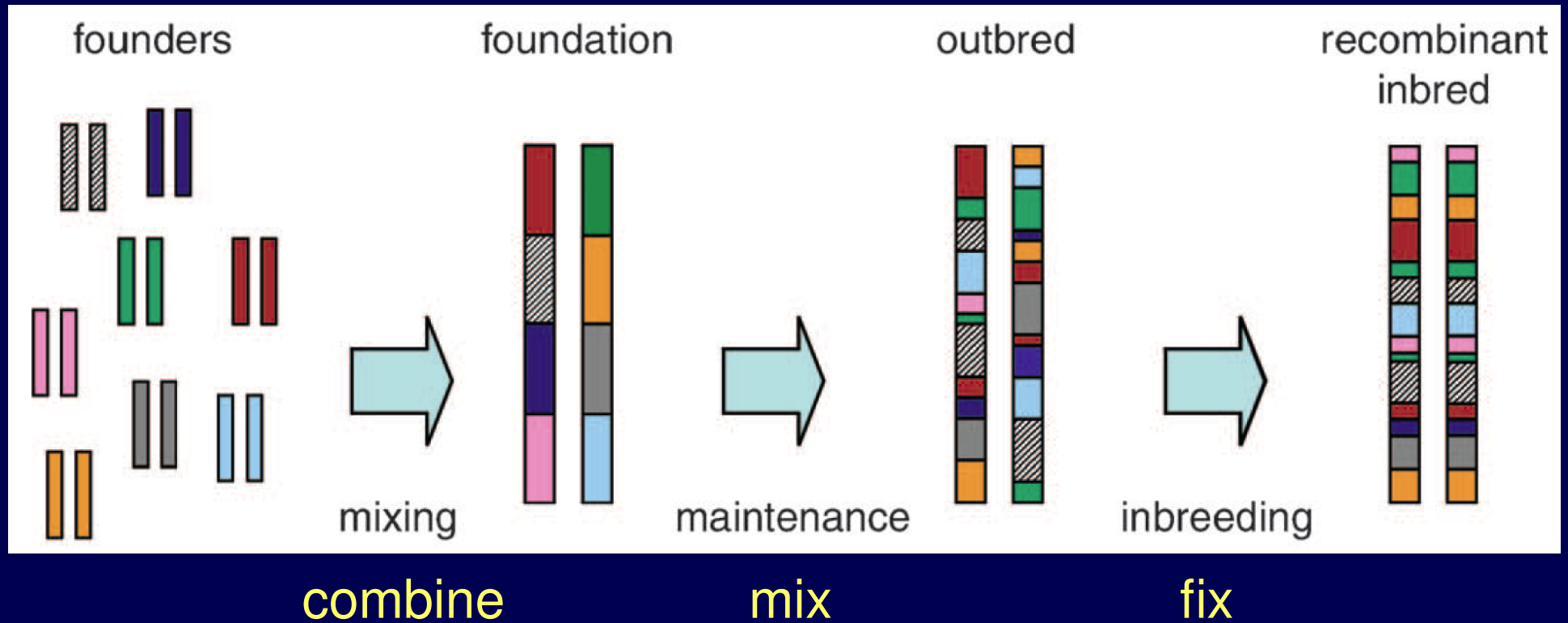


MAGIC lines



How many?

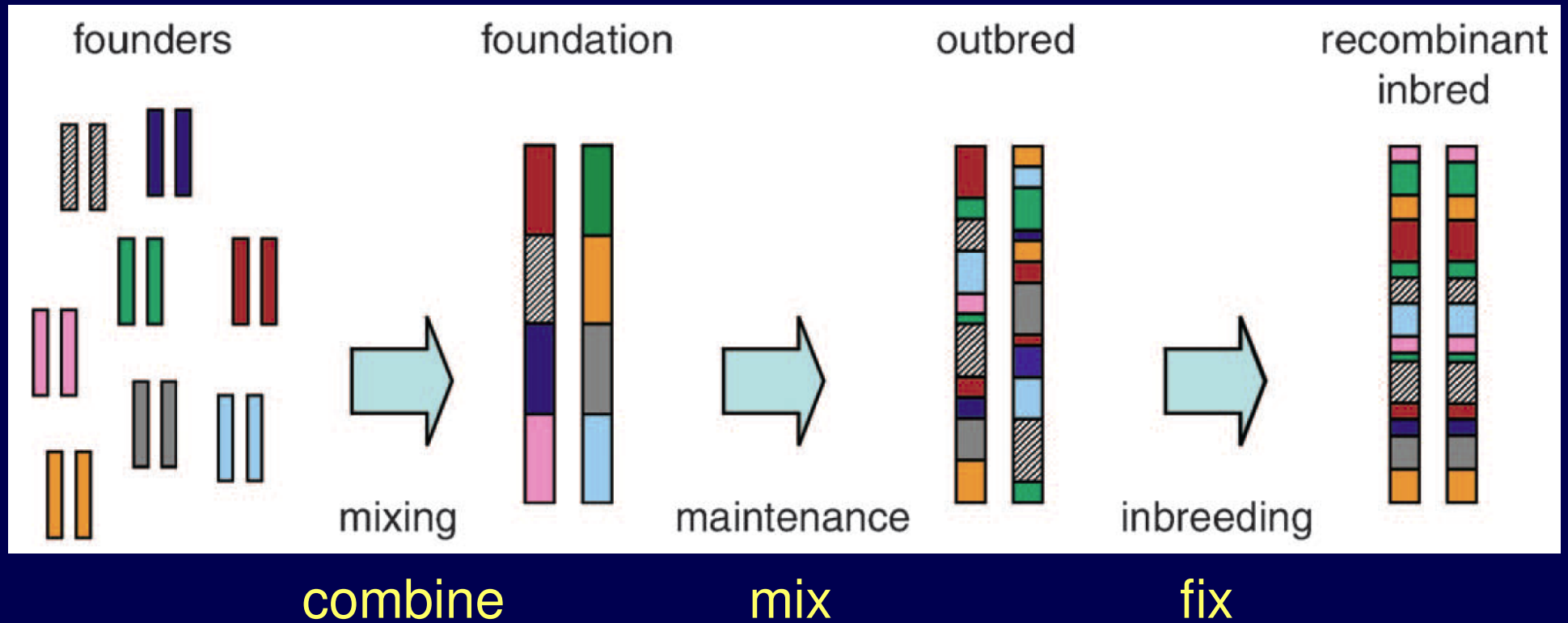
MAGIC lines



How many?

Which?

MAGIC lines

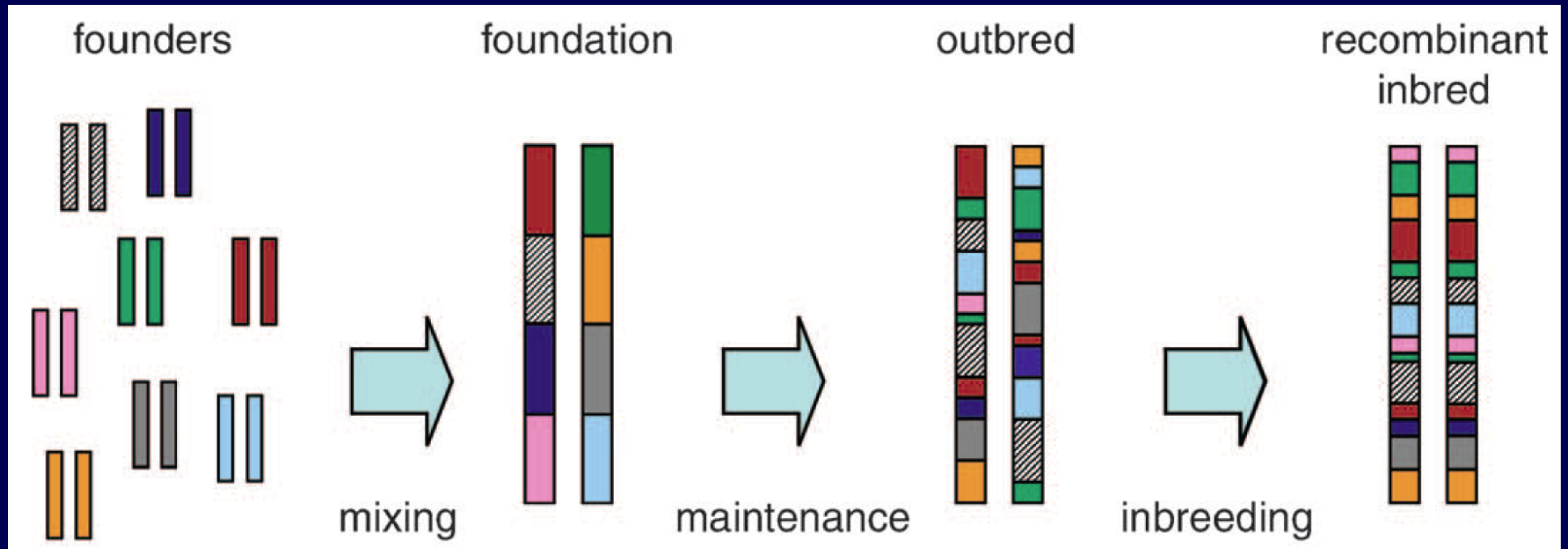


How many?

Which?

How long?

MAGIC lines



combine

mix

fix

How many?

How long?

How?

Which?

MAGIC is magic

- Genetic diversity
- High-precision mapping
- Predictable linkage disequilibrium
- Phenotype replicates to reduce individual variation
- Pool phenotypes from multiple labs, environments, treatments
- Genotype once

The goal

Identify QTG

- Power
- Mapping precision
- Estimate QTL allele frequencies

Principles

- Avoid population structure
- Tradeoff between power for *de novo* discovery and mapping precision
- More QTL to find \Rightarrow more QTL getting in the way?
- More QTL alleles \Rightarrow less information about each
- Are QTL alleles common or rare?

How many founders?

More

- More general use
- More QTL
- Greater precision
- Estimate allele frequencies
- Haplotype analysis in founders

Fewer

- Lower residual variance
- Greater power for a particular QTL?
- Better power for epistasis
- Rare alleles are less rare

Which founders?

- Diverse
- Interesting
- No breeding problems
- Balanced: star phylogeny

How much mixing?

- More mixing \Rightarrow Greater mapping precision
- ...but lower power for *de novo* mapping
- Potential for population structure, missing alleles
- Random mating or curated mating?
- Start with many random cross directions?

Key analysis issues

How to deal with the multiple alleles?

- Full model (an effect for each allele)
- Diallelic QTL model
- Random effects model (like BLUP)

How to account for multiple QTL?

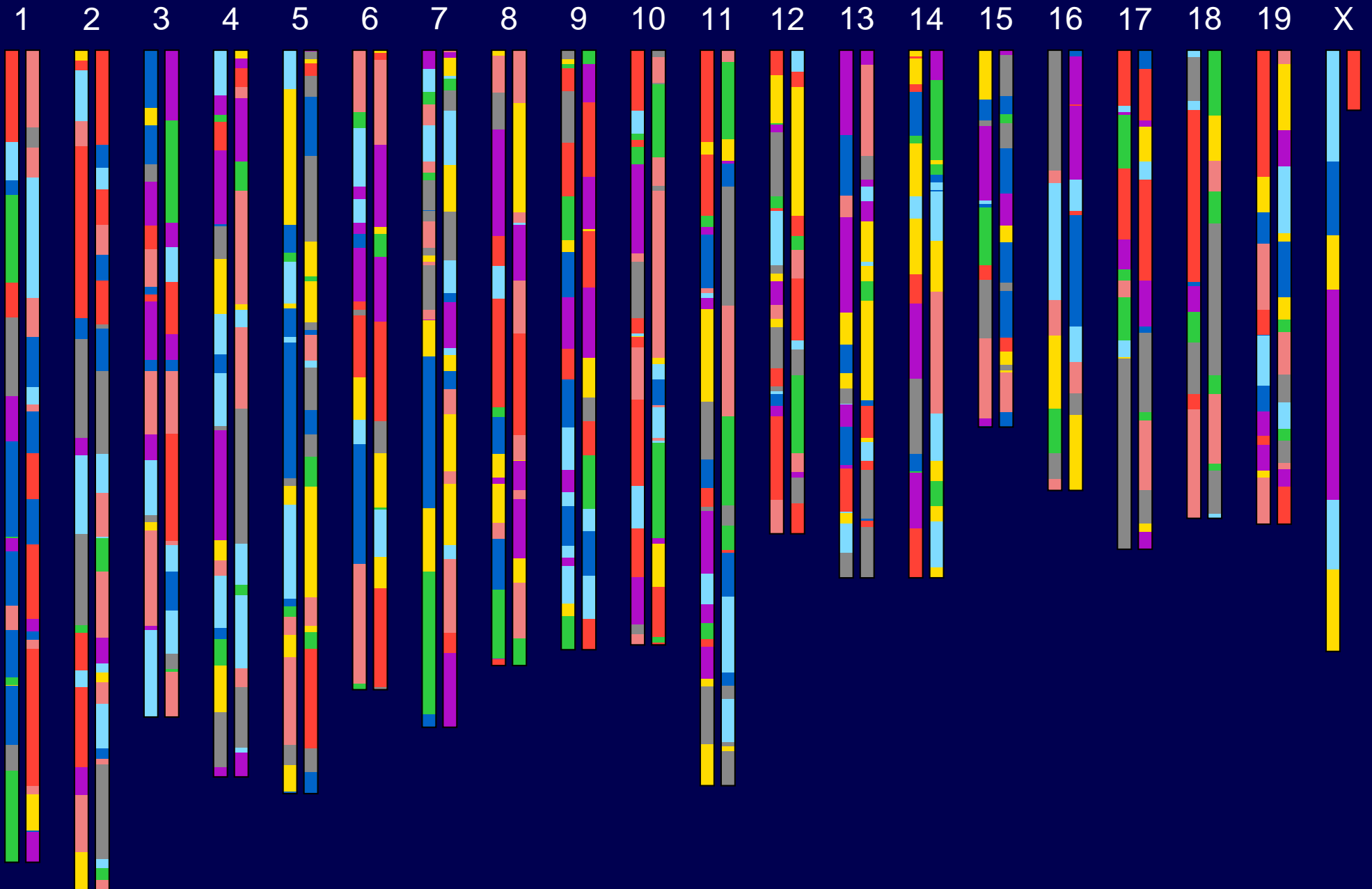
- Stepwise selection
- Bayesian model averaging
- Random effect for polygenes

Linear mixed models

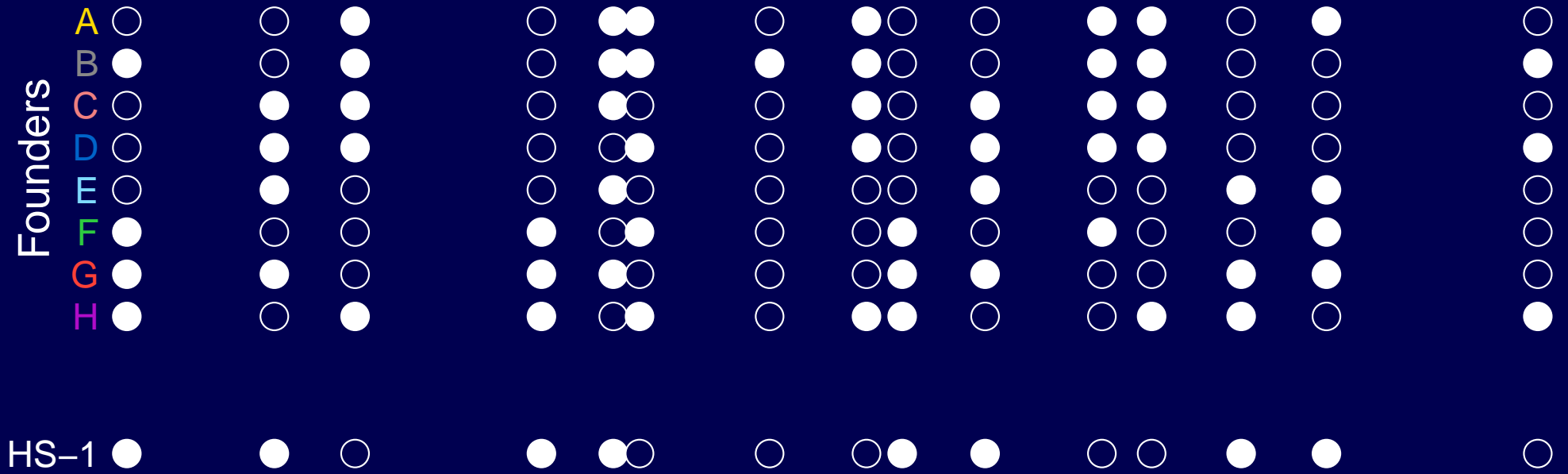
$$\begin{aligned} y_i &= \mu + \sum_k \beta_k q_{ik} + \epsilon_i & \epsilon_i &\sim \mathbf{N}(0, \sigma_e^2) \\ &= \mu + \eta_i + \epsilon_i & \eta_i &\sim \mathbf{N}(0, \sigma_p^2) \end{aligned}$$

$$\mathbf{COV}(\eta_i, \eta_j) = \sigma_p^2 (2k_{ij})$$

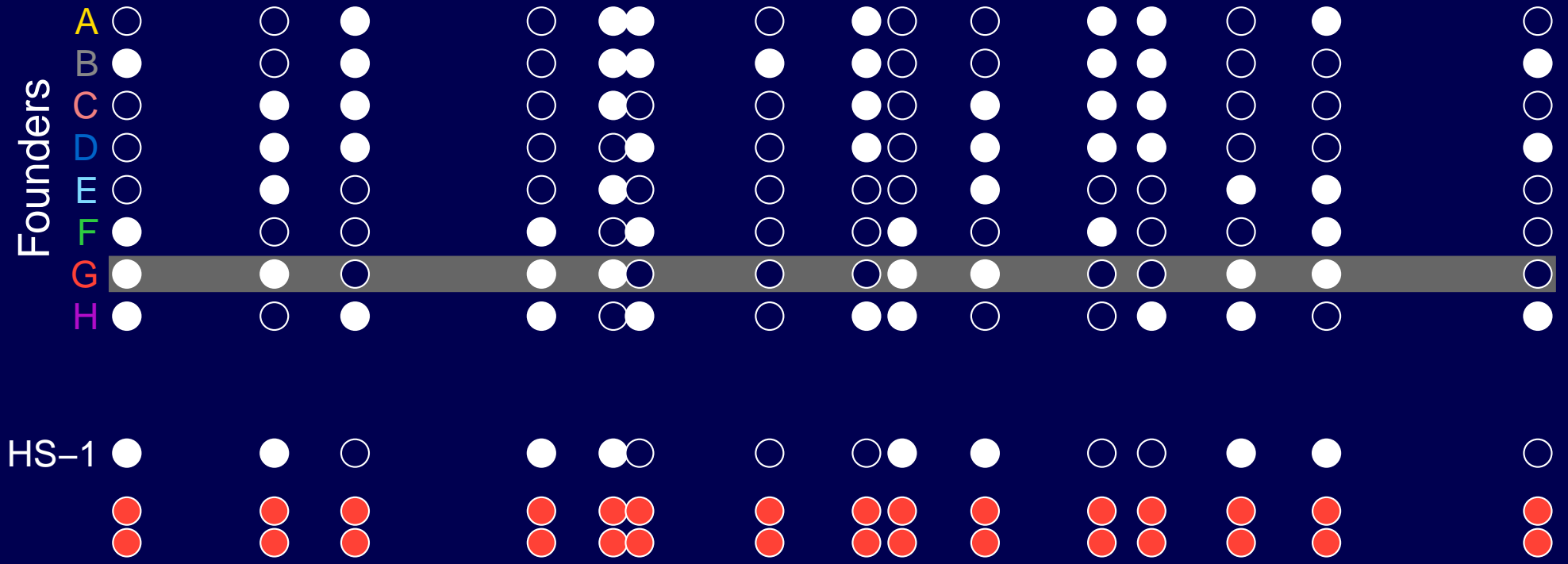
HS genome



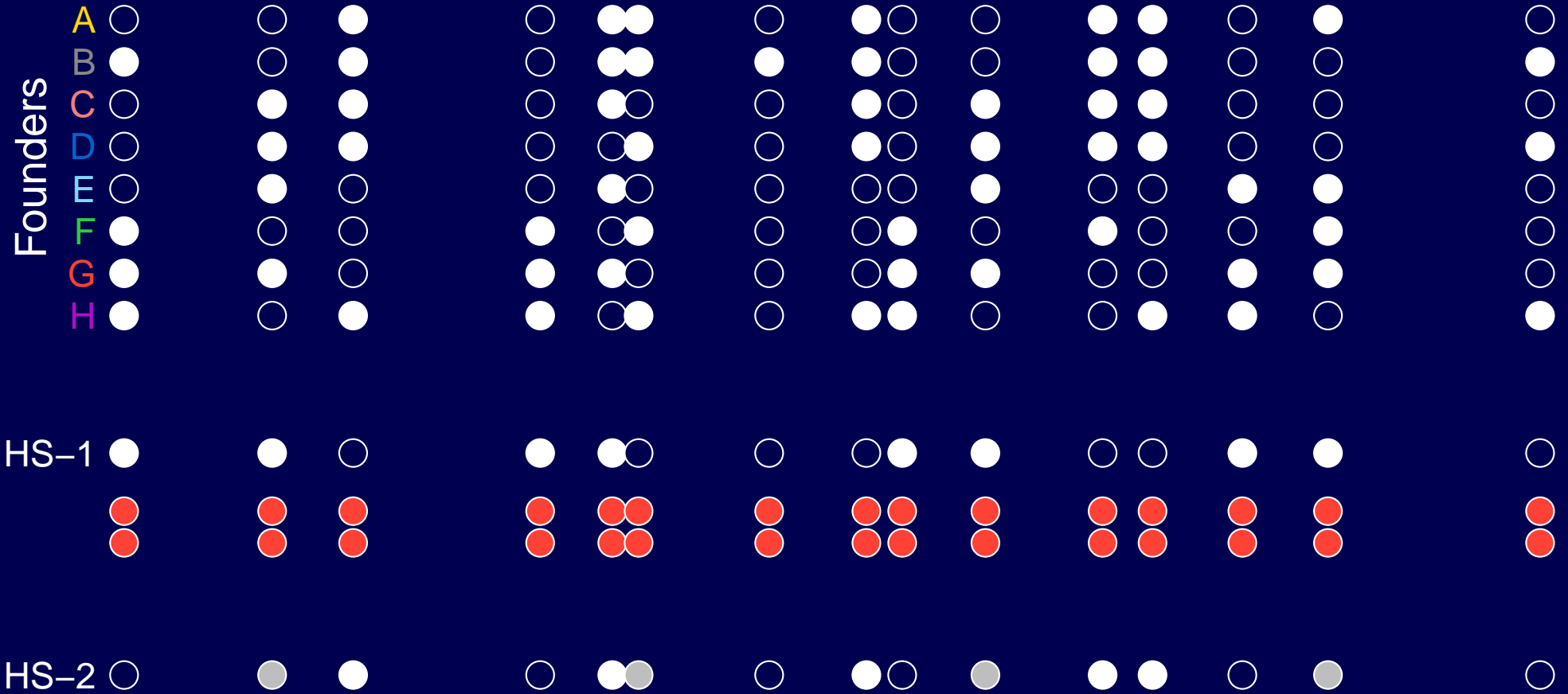
HS genotype reconstruction



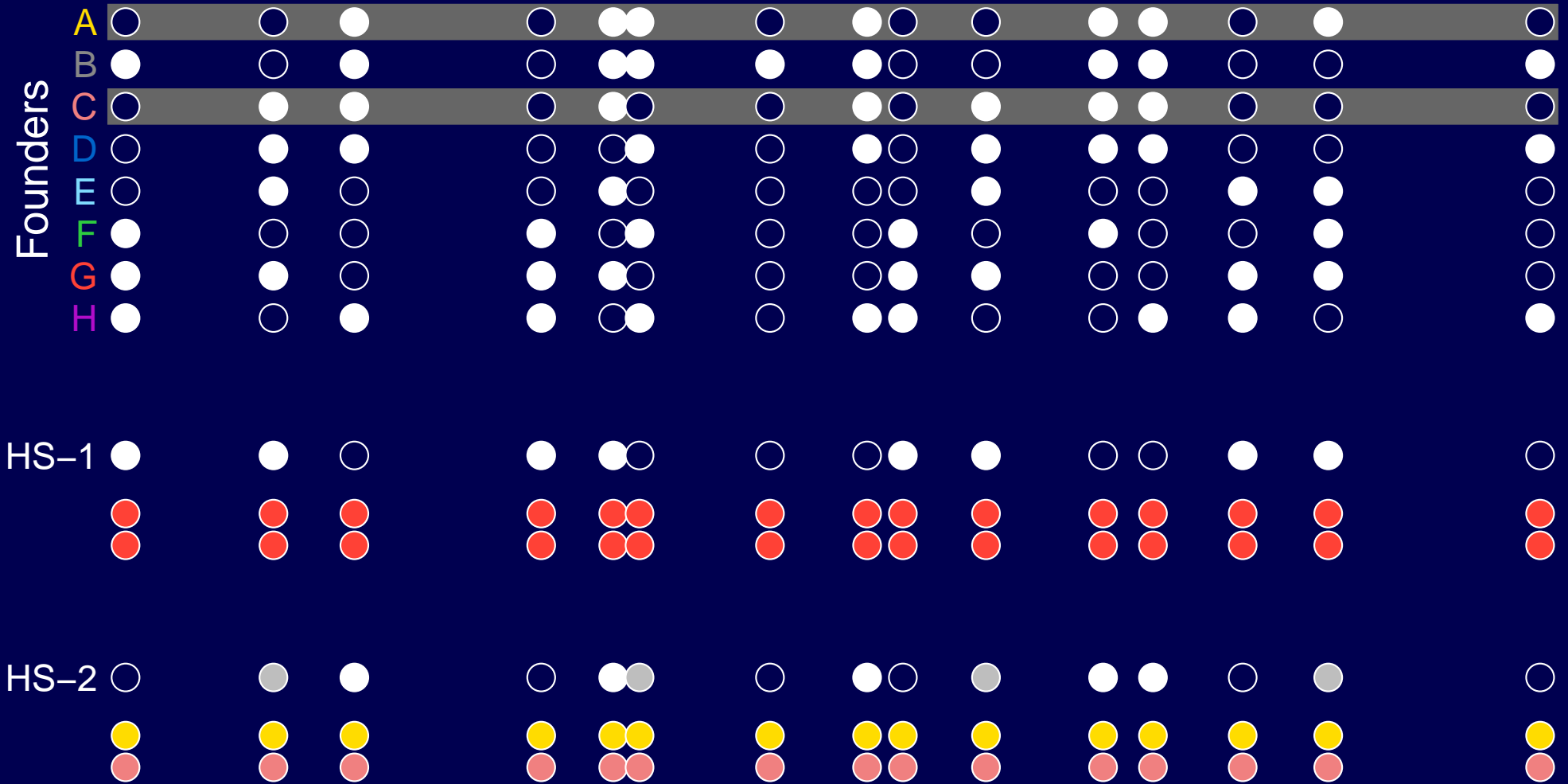
HS genotype reconstruction



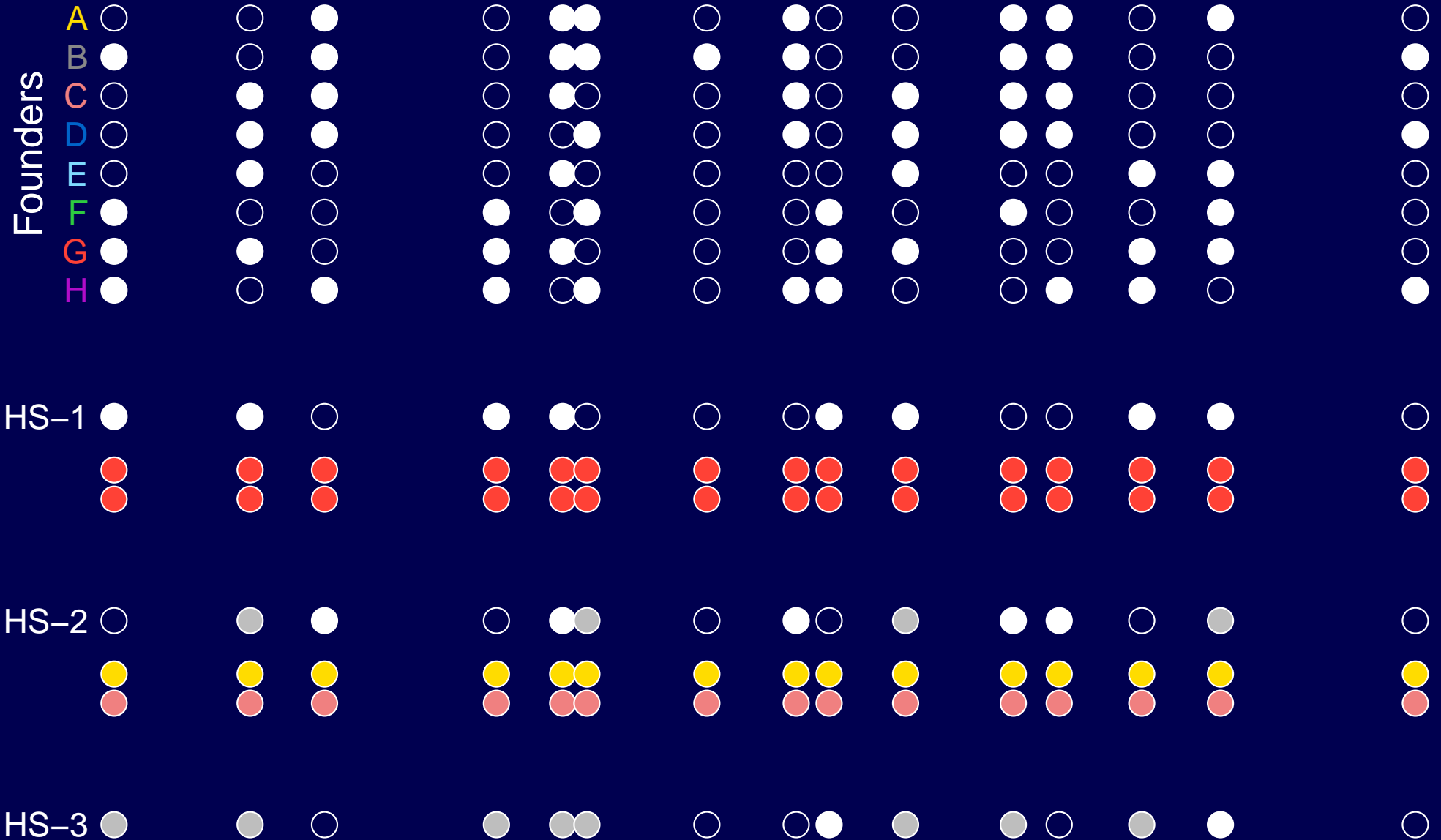
HS genotype reconstruction



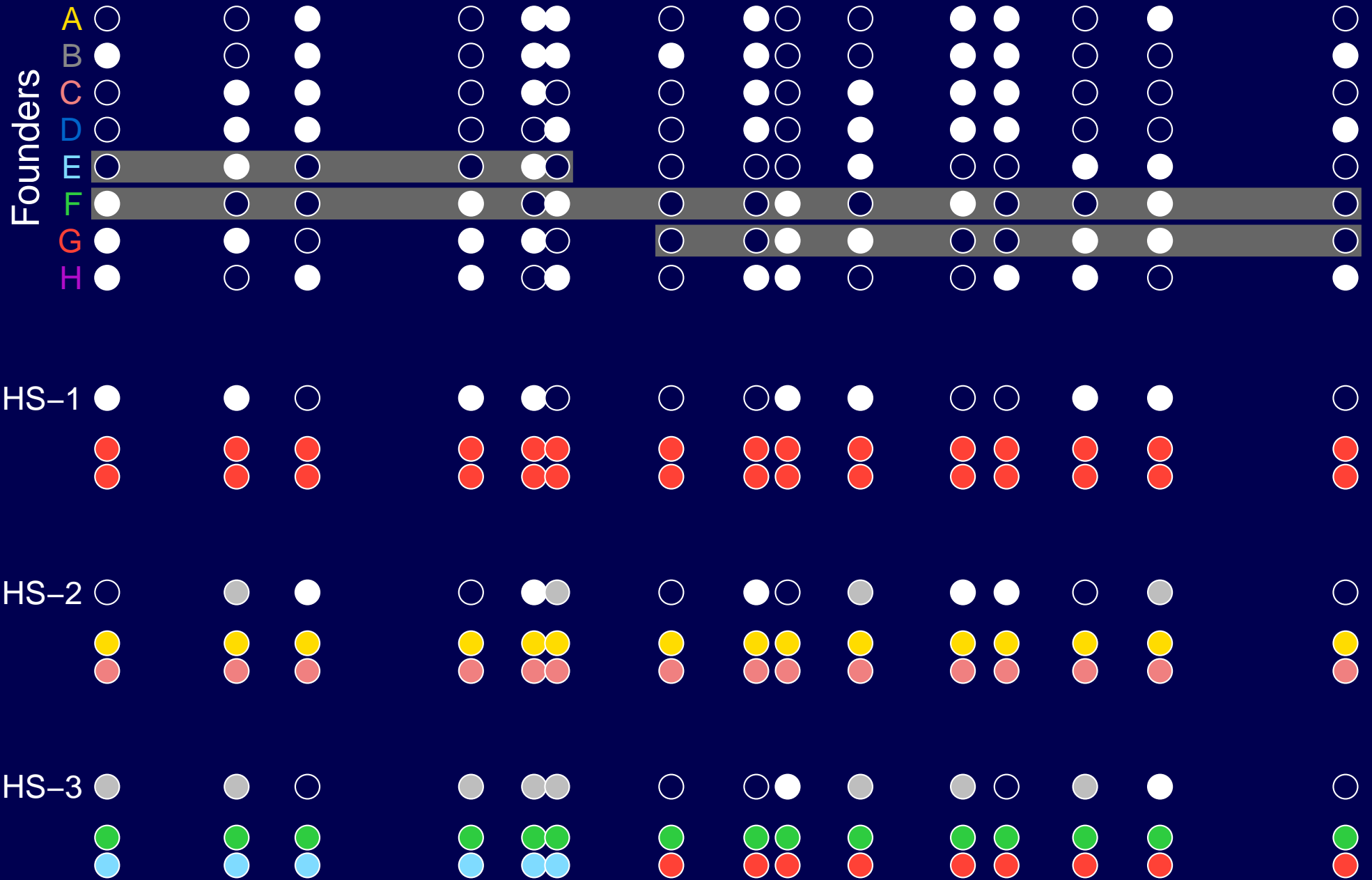
HS genotype reconstruction



HS genotype reconstruction

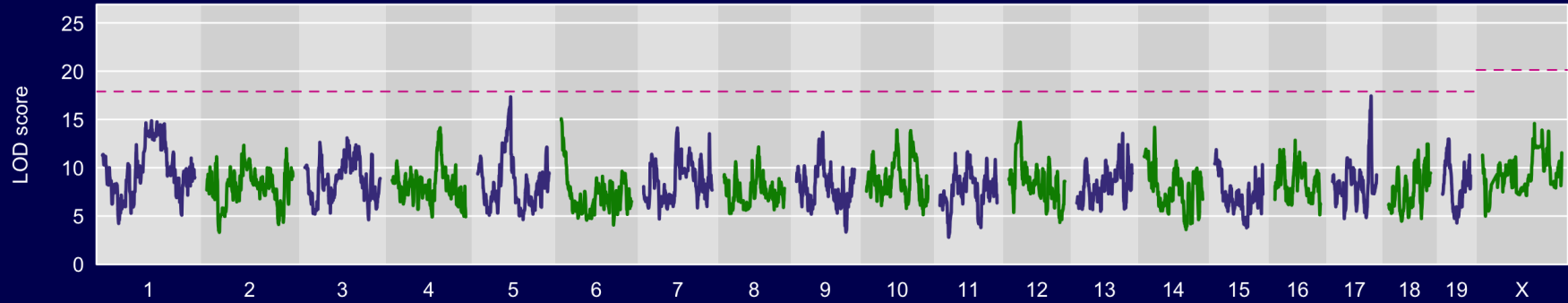


HS genotype reconstruction

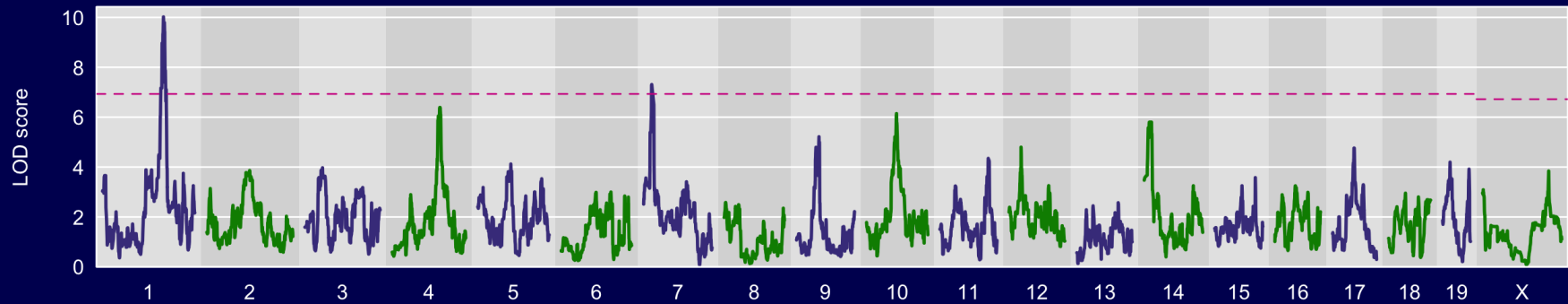


HS genome scans

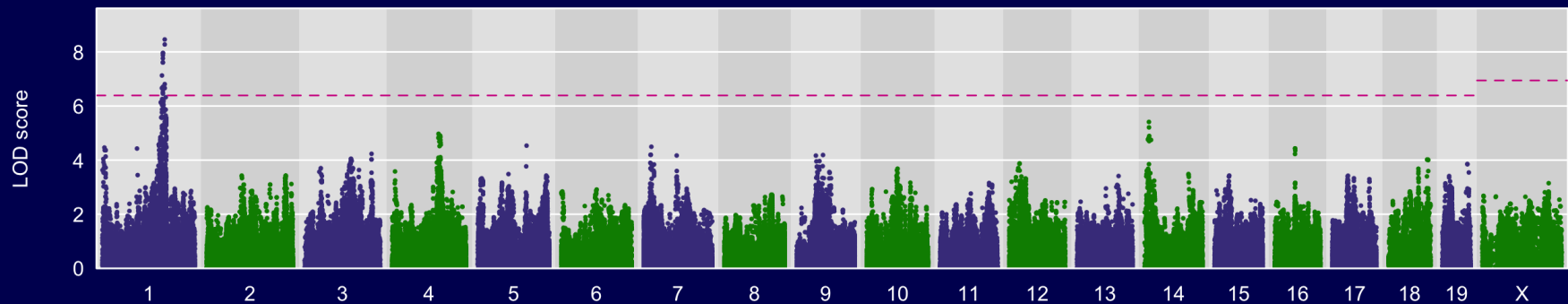
Full model



Additive alleles



SNP associations



Why R/qtl2?

- High-dimensional data
 - genotypes and phenotypes
- More diverse crosses
 - especially multi-parent populations
- Linear mixed models
 - especially in HS/AIL

Open problems

- Analysis of multi-parent populations
 - SNP vs haplotype vs random QTL effect
 - Identifying causal polymorphisms
- Joint analysis of high-dimensional phenotypes
 - gene expression, proteins, metabolites, etc.
- QTL × environment interactions