Complex Trait Genetics in Animal Models

Will Valdar
Oxford University

Mapping Genes for Quantitative Traits in Outbred Mice

Will Valdar

Oxford University

What's so great about mice?

Share ~99% of genes with humans

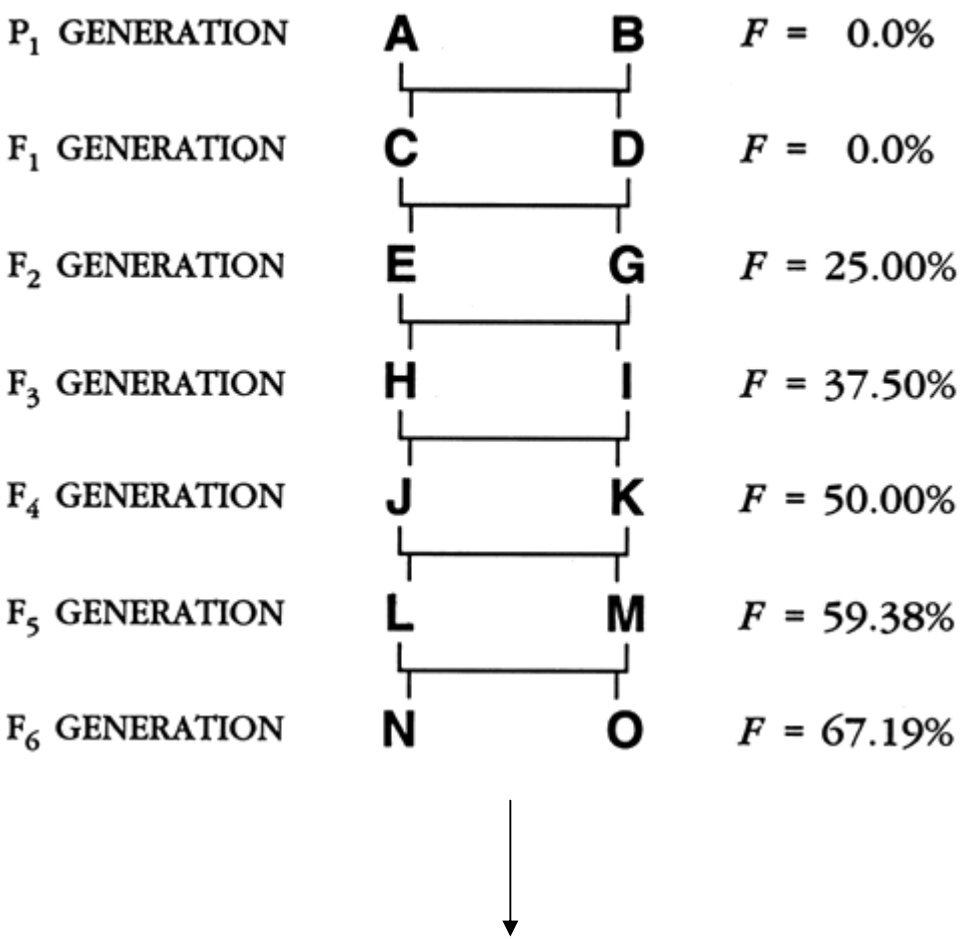
~90% of the two genomes can be portioned into regions of conserved synteny

Shorter lifespans

You can do invasive experiments

You can breed them as you like – control the genetics

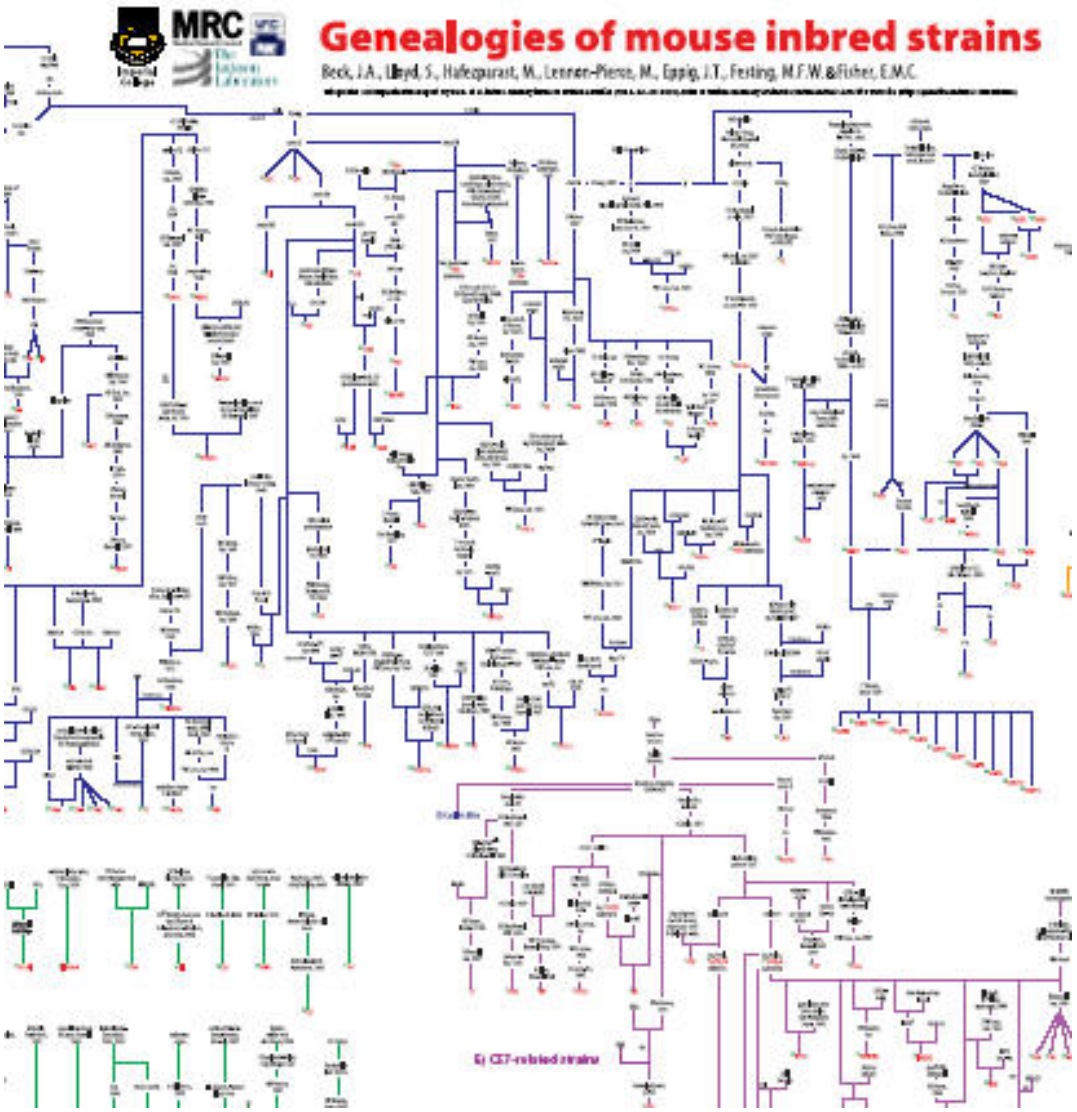
What is an inbred strain?



F20 generation

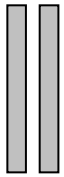
F = 99%

BALB/c

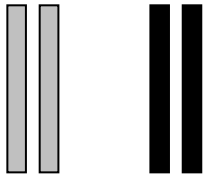


http://www.informatics.jax.org/mgihome/genealogy/

one inbred strain



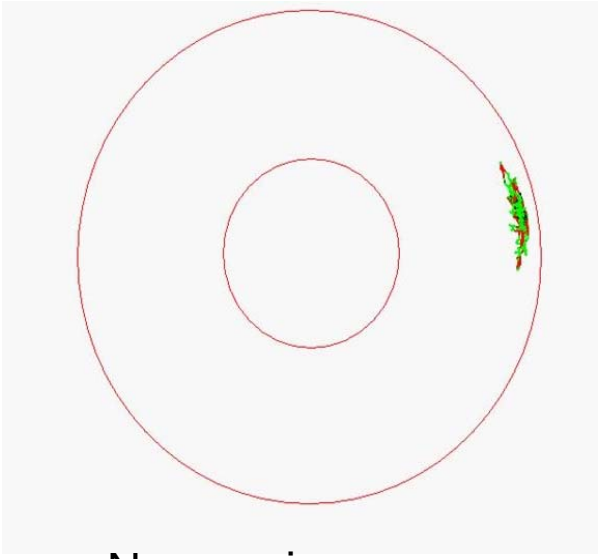
two inbred strains



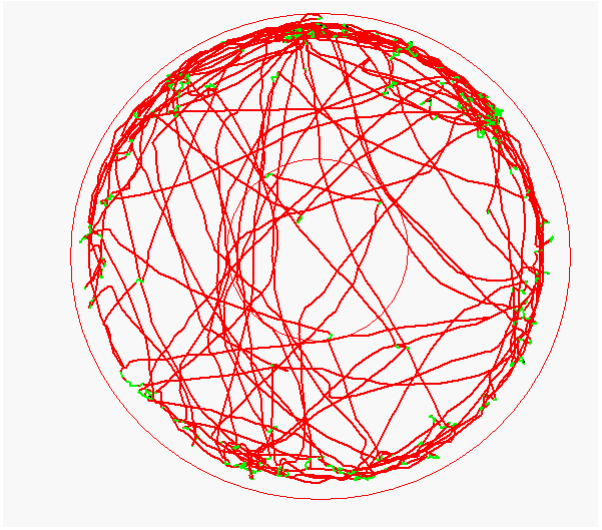
Mouse model of anxiety

Mouse model of anxiety

Anxious mouse



Non anxious mouse



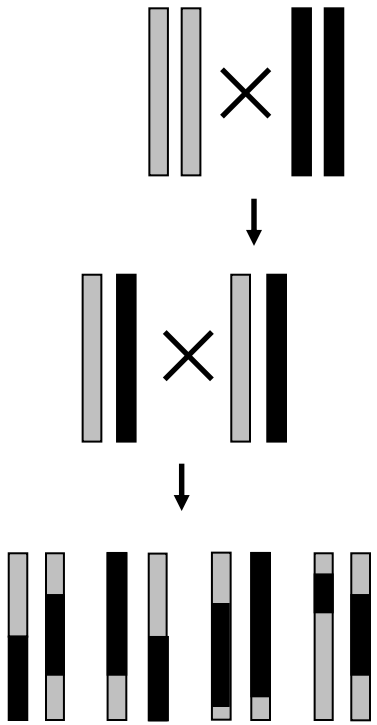
F2 cross

Generation

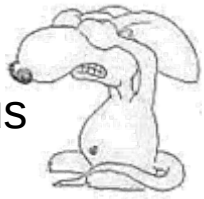
F0

F1

F2

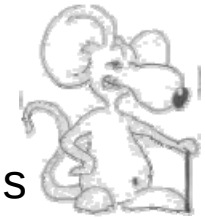


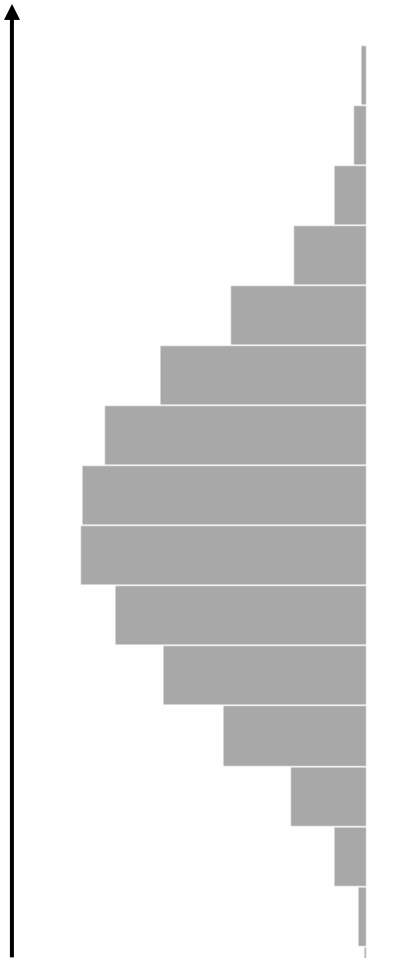
anxious



about average

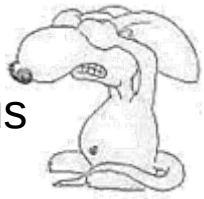
non-anxious





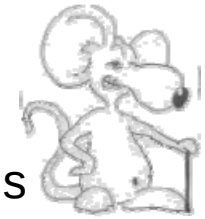
Quantitative Trait Locus

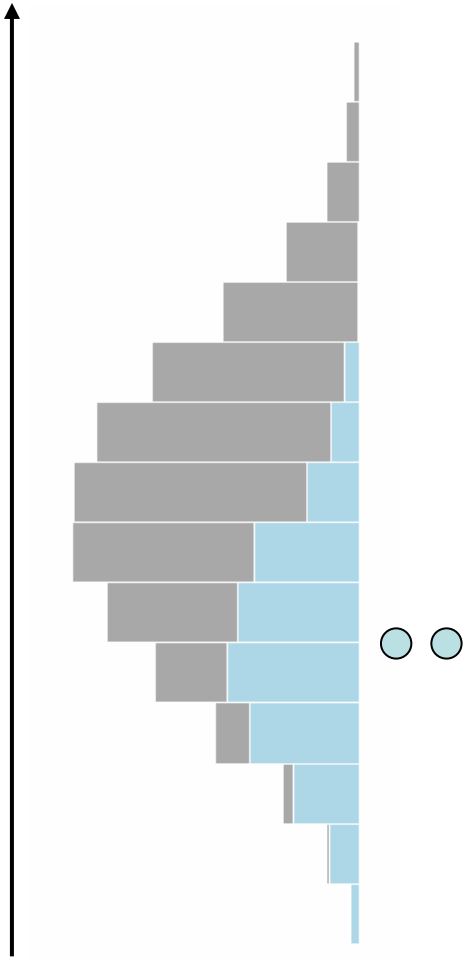
anxious



about average

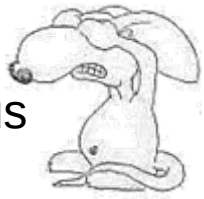
non-anxious





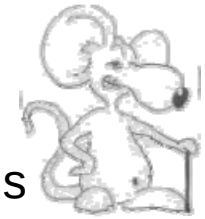
Quantitative Trait Locus

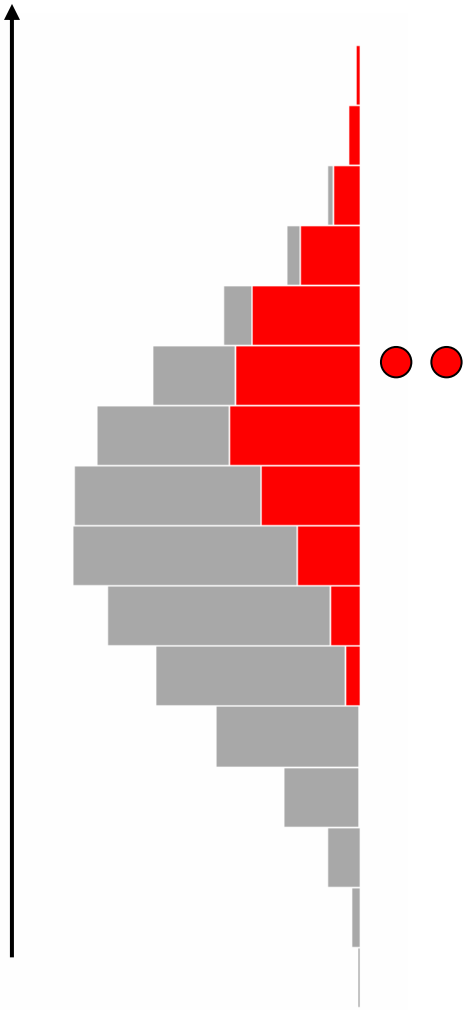
anxious



about average

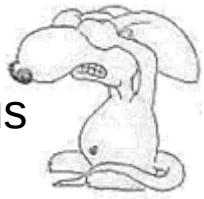
non-anxious





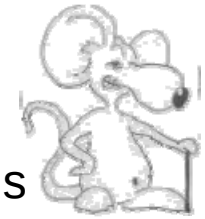
Quantitative Trait Locus

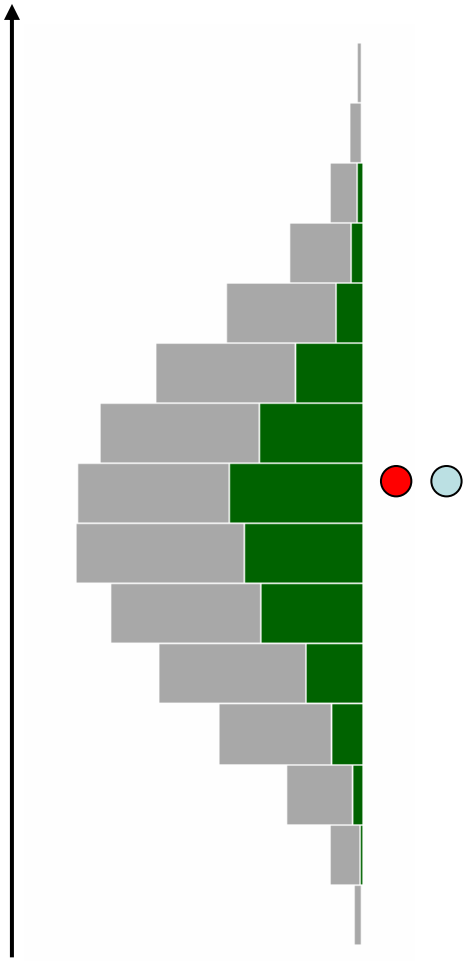
anxious



about average

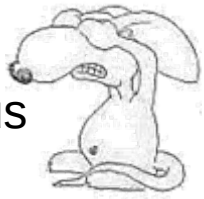
non-anxious





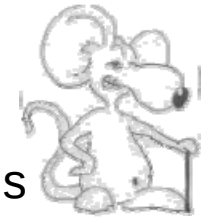
Quantitative Trait Locus

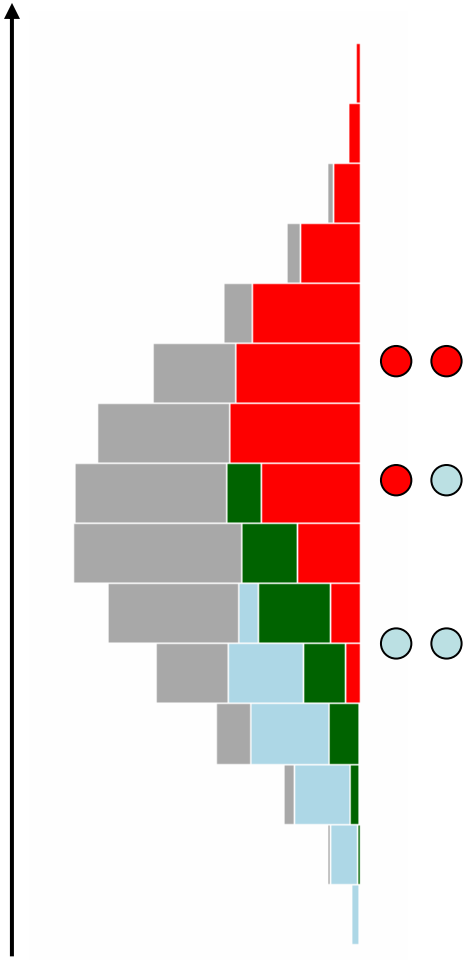
anxious



about average

non-anxious





QTL snp

Linear models

Also known as

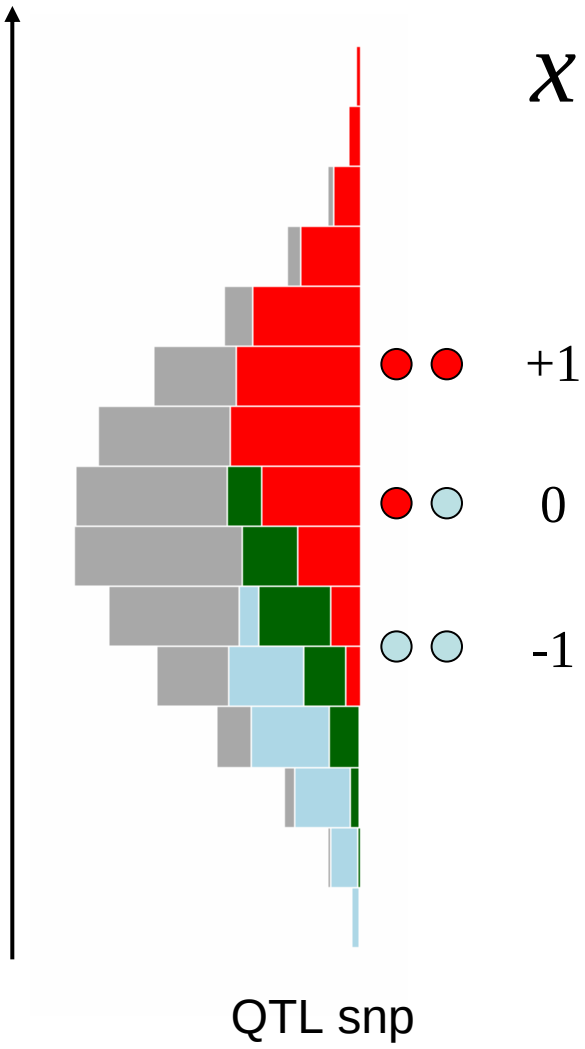
ANOVA

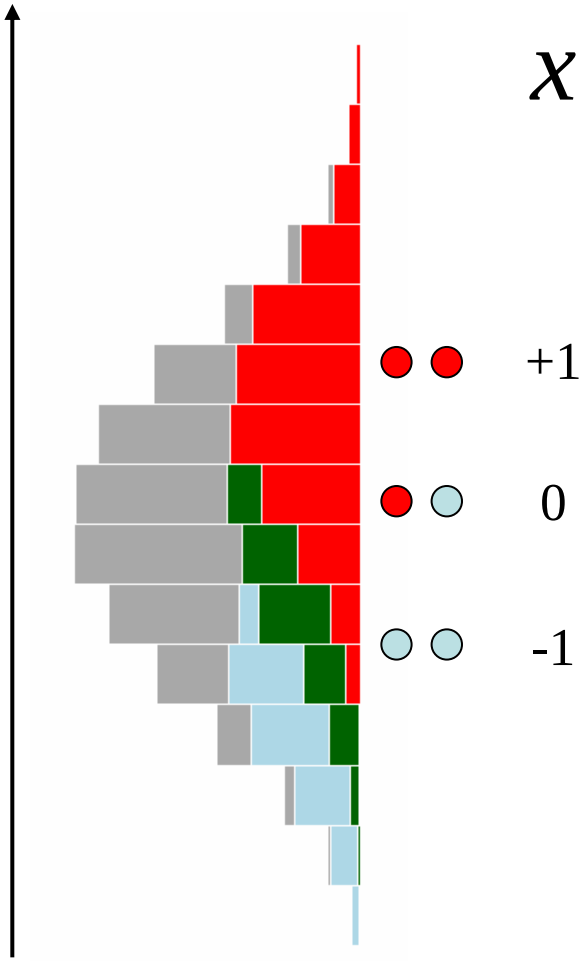
ANCOVA

regression

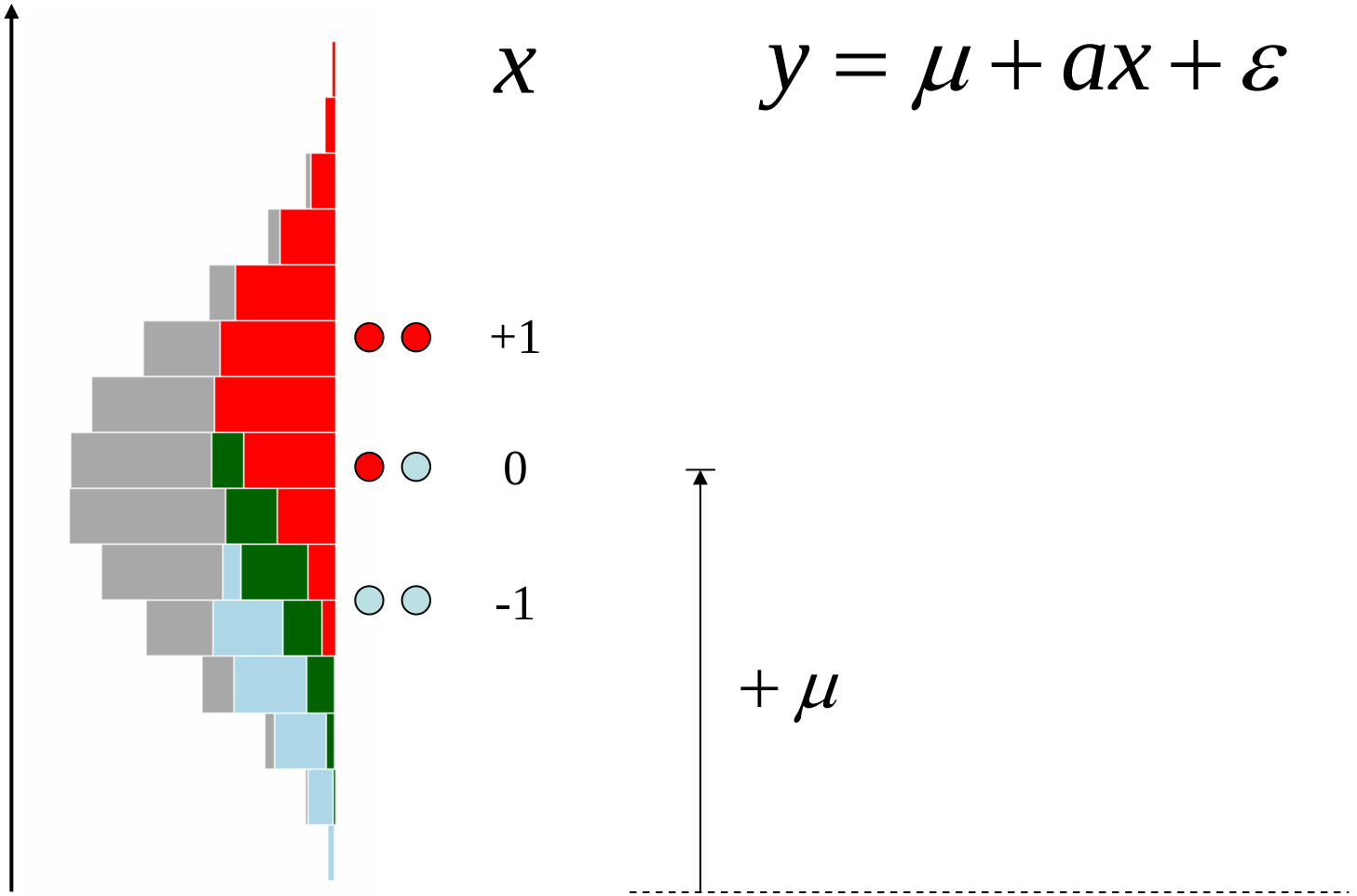
multiple regression

linear regression

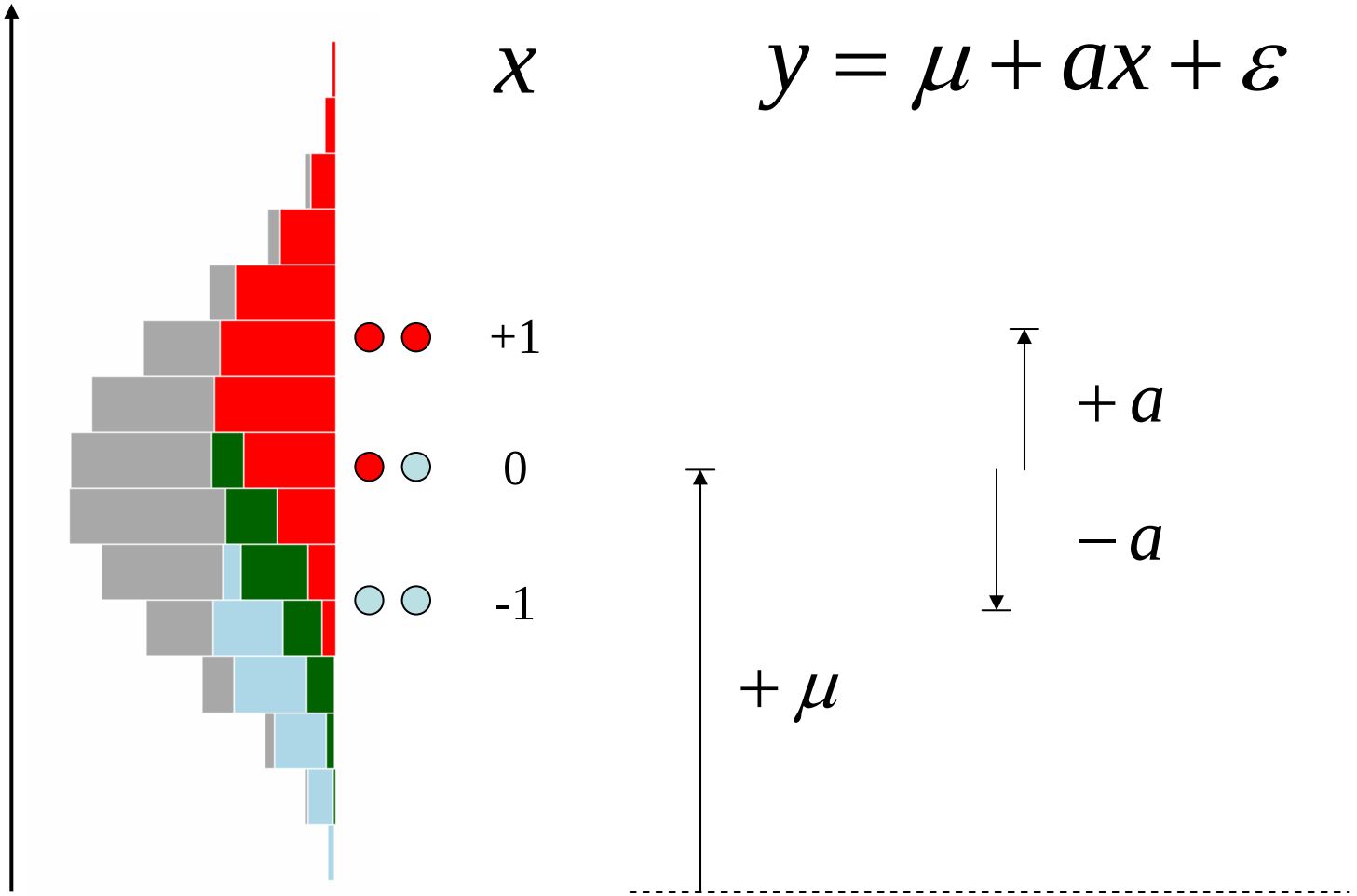

x
+1
0
-1
QTL snp



$$y = \mu + ax + \varepsilon$$



$$y = \mu + ax + \varepsilon$$


x
+1
0
-1
QTL snp
$y = \mu + ax + \varepsilon$
$+a$
$-a$
$+\mu$

Hypothesis testing

H_0: $y = \mu + \varepsilon$

H_1: $y = \mu + ax + \varepsilon$

Hypothesis testing

H_0: y ~ 1

$$y = \mu + \varepsilon$$

H_1: y ~ 1 + **x**

$$y = \mu + ax + \varepsilon$$

Hypothesis testing

H_0: y ~ 1

$$y = \mu + \varepsilon$$

H_1: y ~ 1 + **x**

$$y = \mu + ax + \varepsilon$$

H_1 vs H_0 : Does **x** explain a significant amount of the variation?

Hypothesis testing

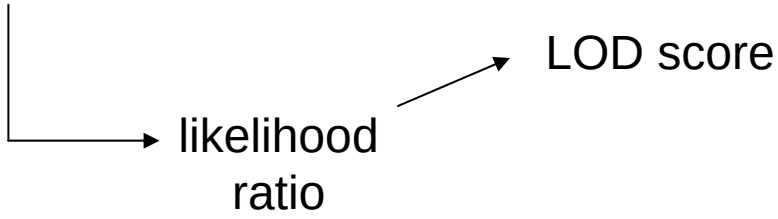
H_0: y ~ 1

$$y = \mu + \varepsilon$$

H_1: y ~ 1 + **x**

$$y = \mu + ax + \varepsilon$$

H_1 vs H_0 : Does **x** explain a significant amount of the variation?

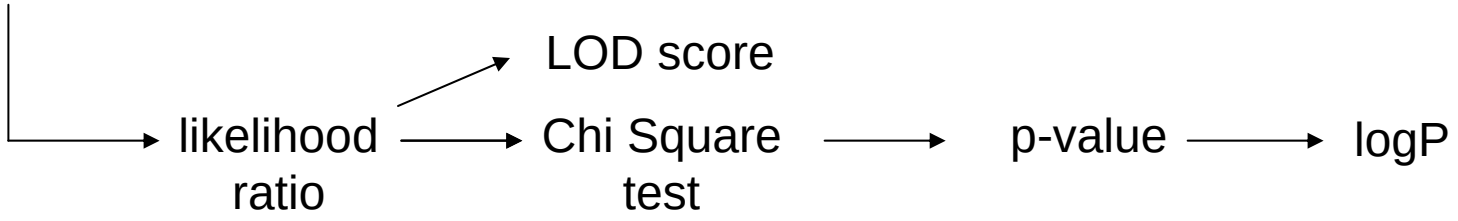


Hypothesis testing

H_0: y ~ 1 $\quad y = \mu + \varepsilon$

H_1: y ~ 1 + **x** $\quad y = \mu + ax + \varepsilon$

H_1 vs H_0 : Does **x** explain a significant amount of the variation?

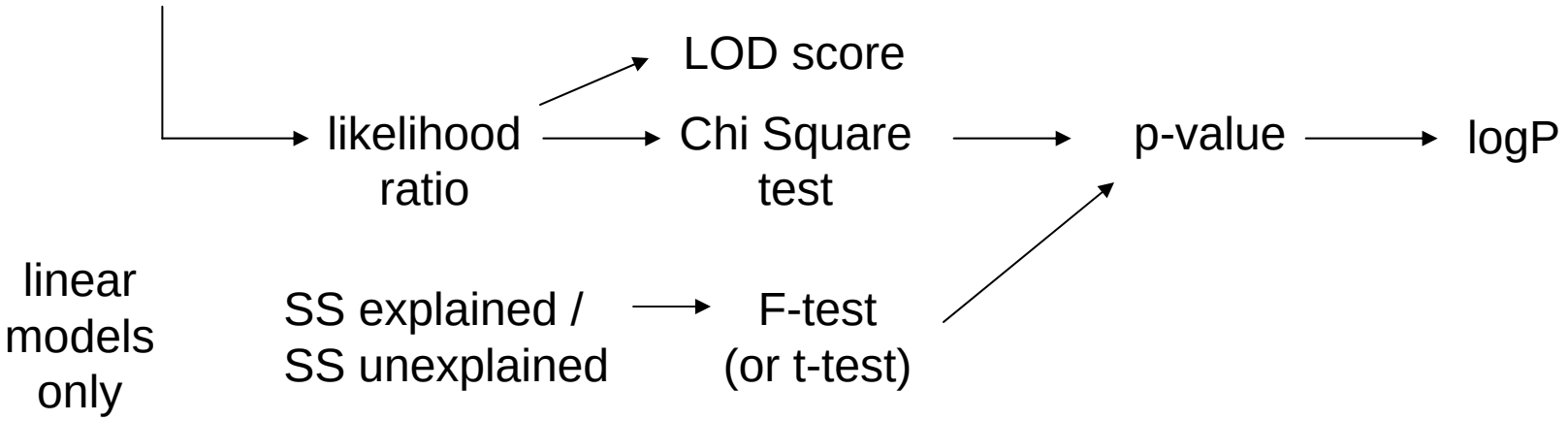


Hypothesis testing

H_0: y ~ 1 $\quad y = \mu + \varepsilon$

H_1: y ~ 1 + **x** $\quad y = \mu + ax + \varepsilon$

H_1 vs H_0 : Does **x** explain a significant amount of the variation?



Hypothesis testing

H_0: y ~ 1 + x1

$$y = \mu + a_1 x_1 + \varepsilon$$

H_1: y ~ 1 + x1 + **x2**

$$y = \mu + a_1 x_1 + a_2 x_2 + \varepsilon$$

H_1 vs H_0 : Does **x2** explain a significant amount of the variation after accounting for x1?

or

is **x2** significant conditional on x1?

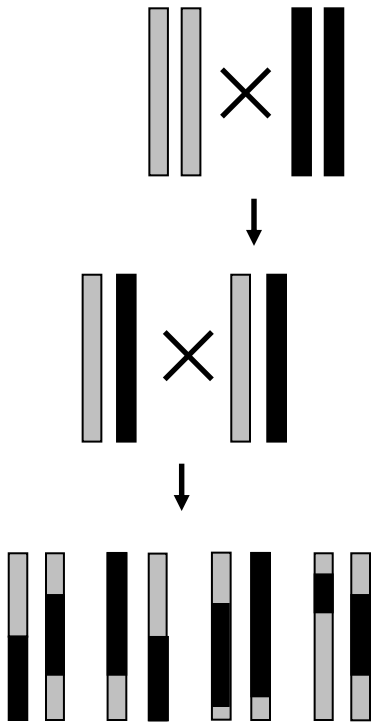
F2 cross

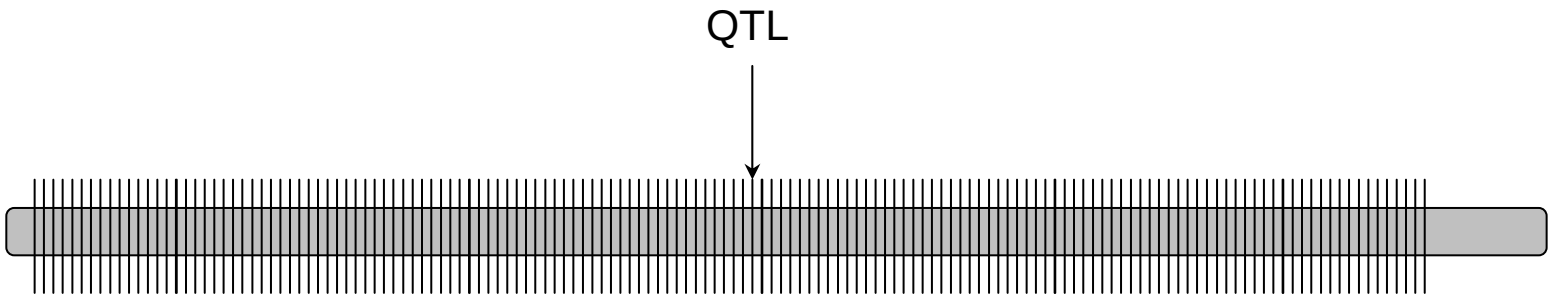
Generation

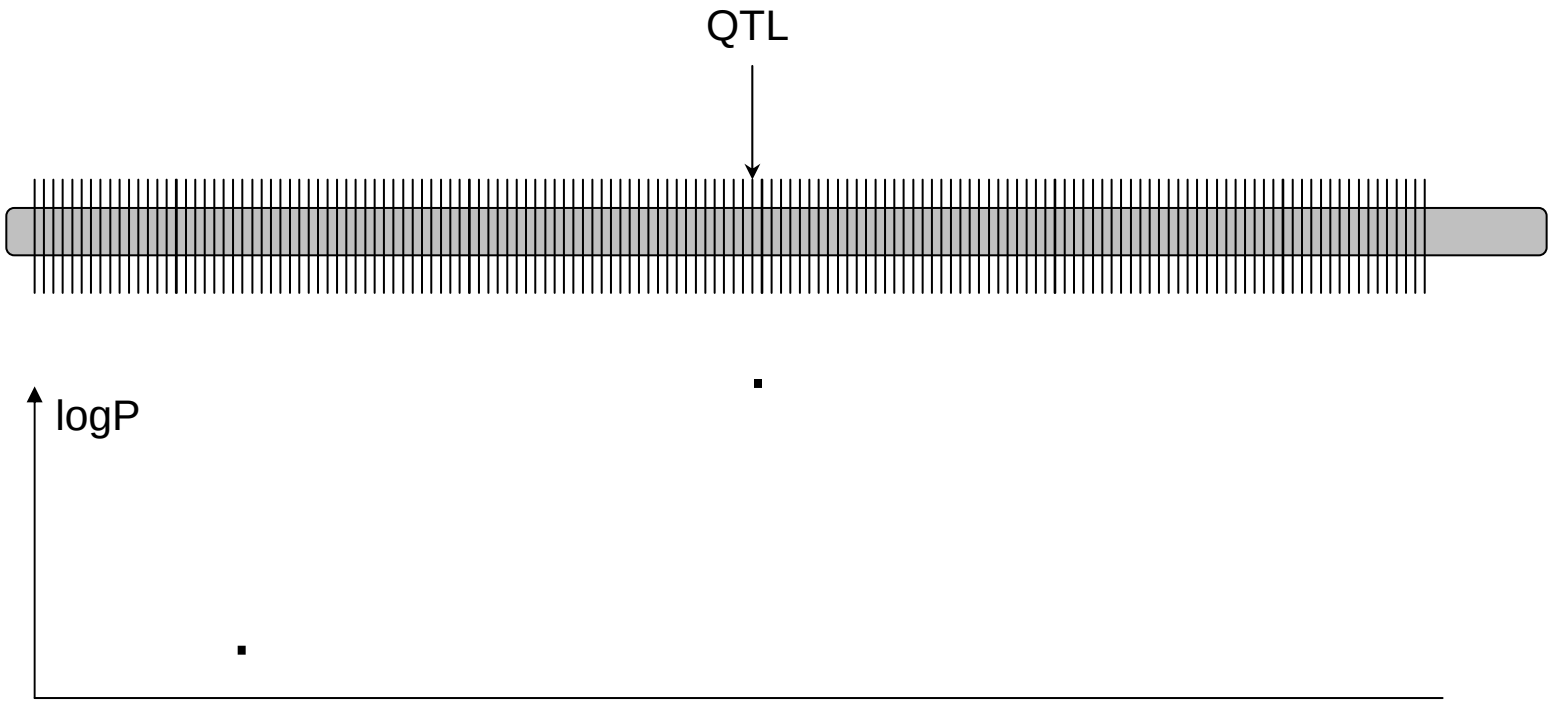
F0

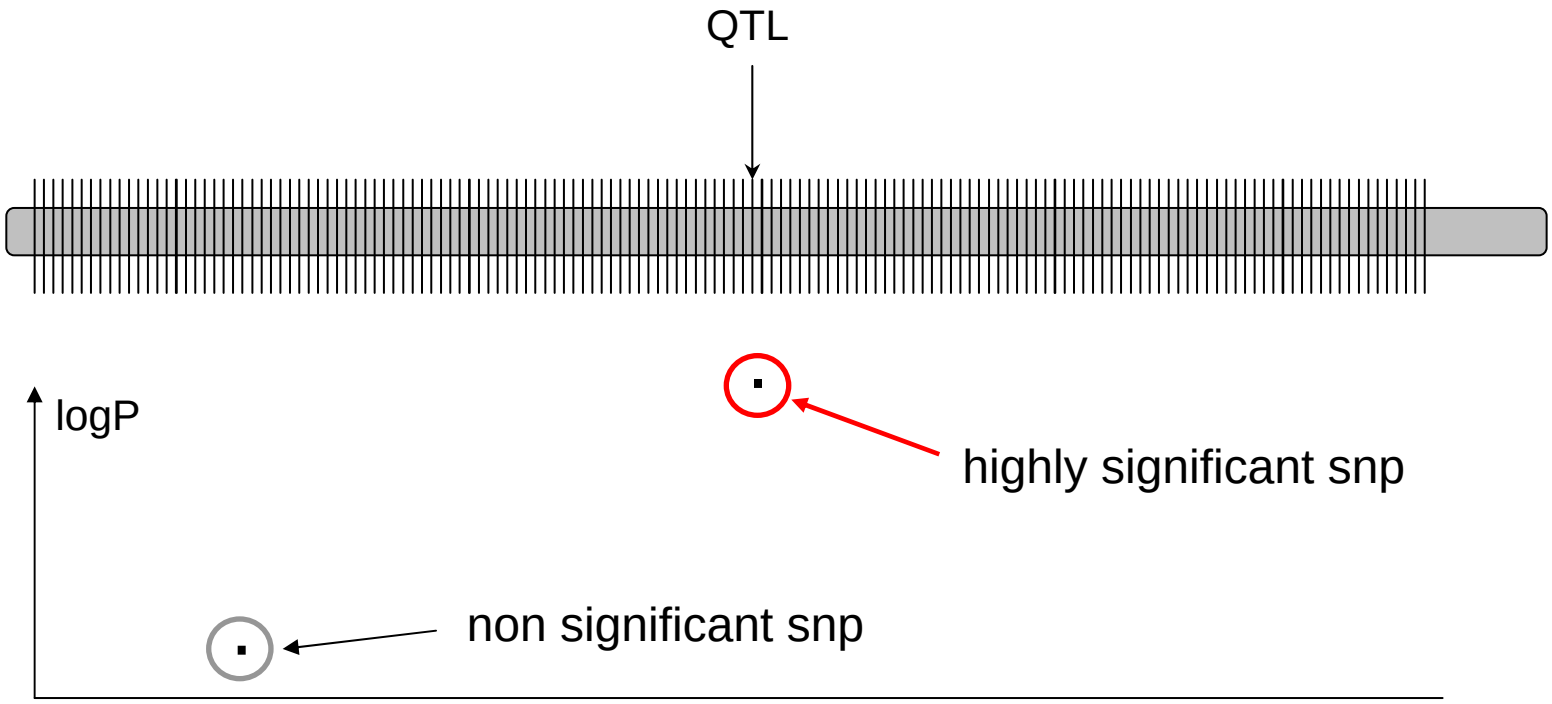
F1

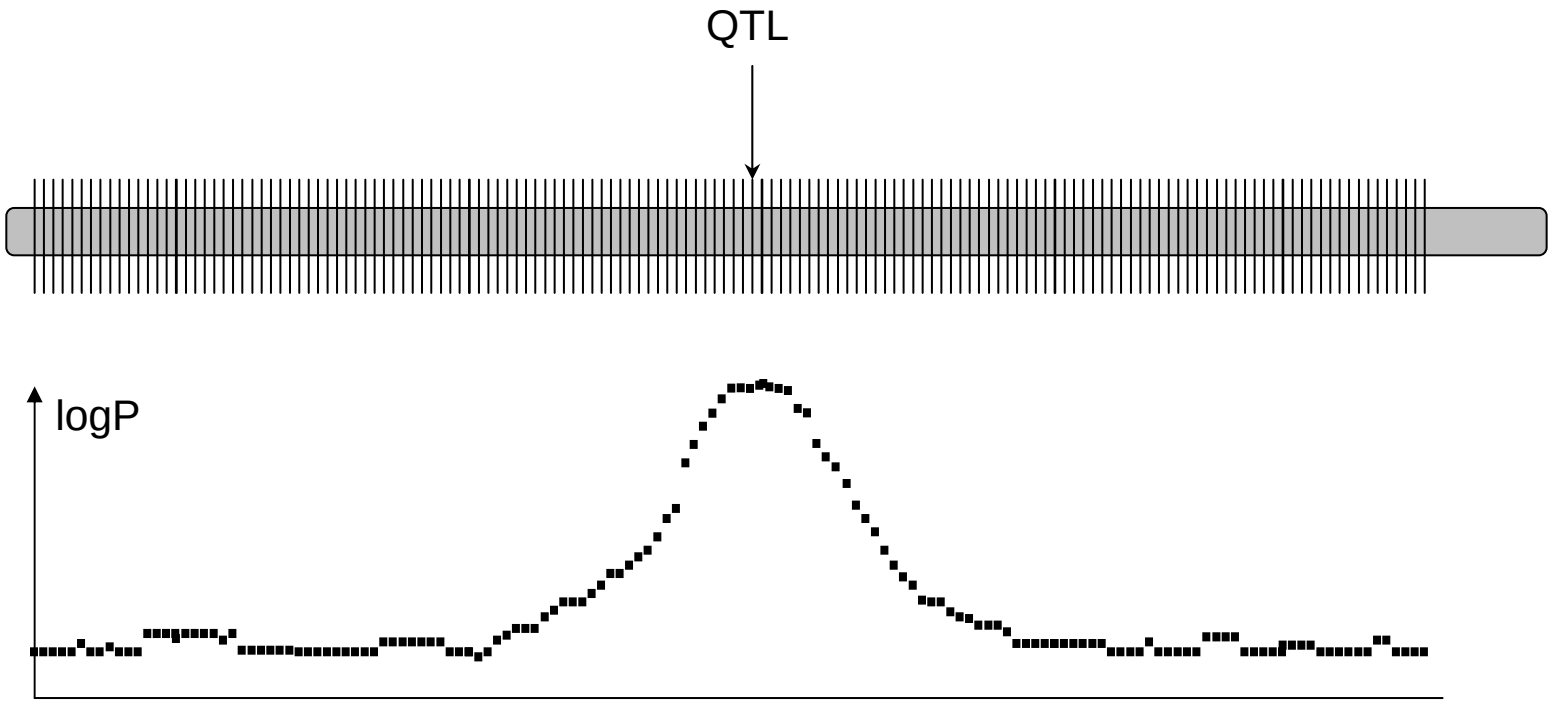
F2

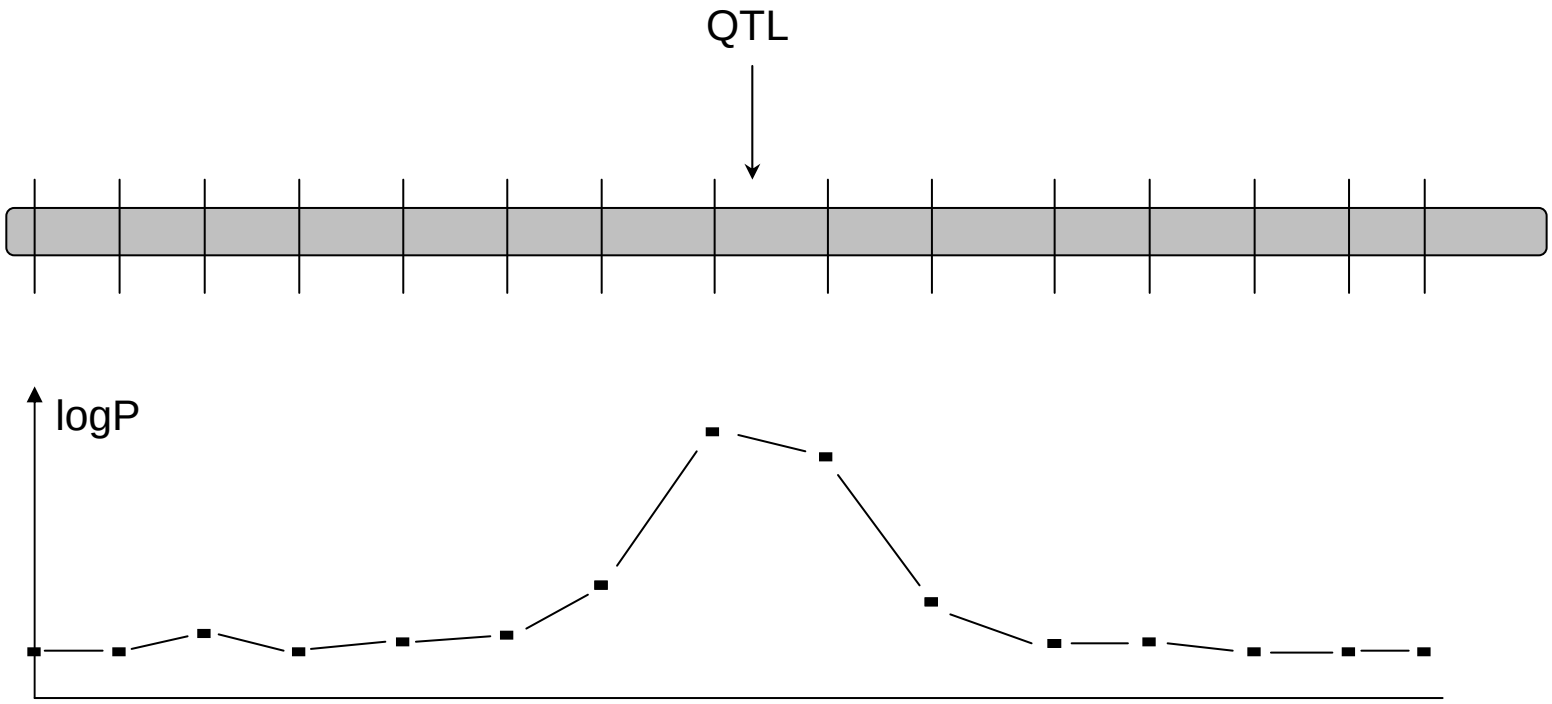



QTL

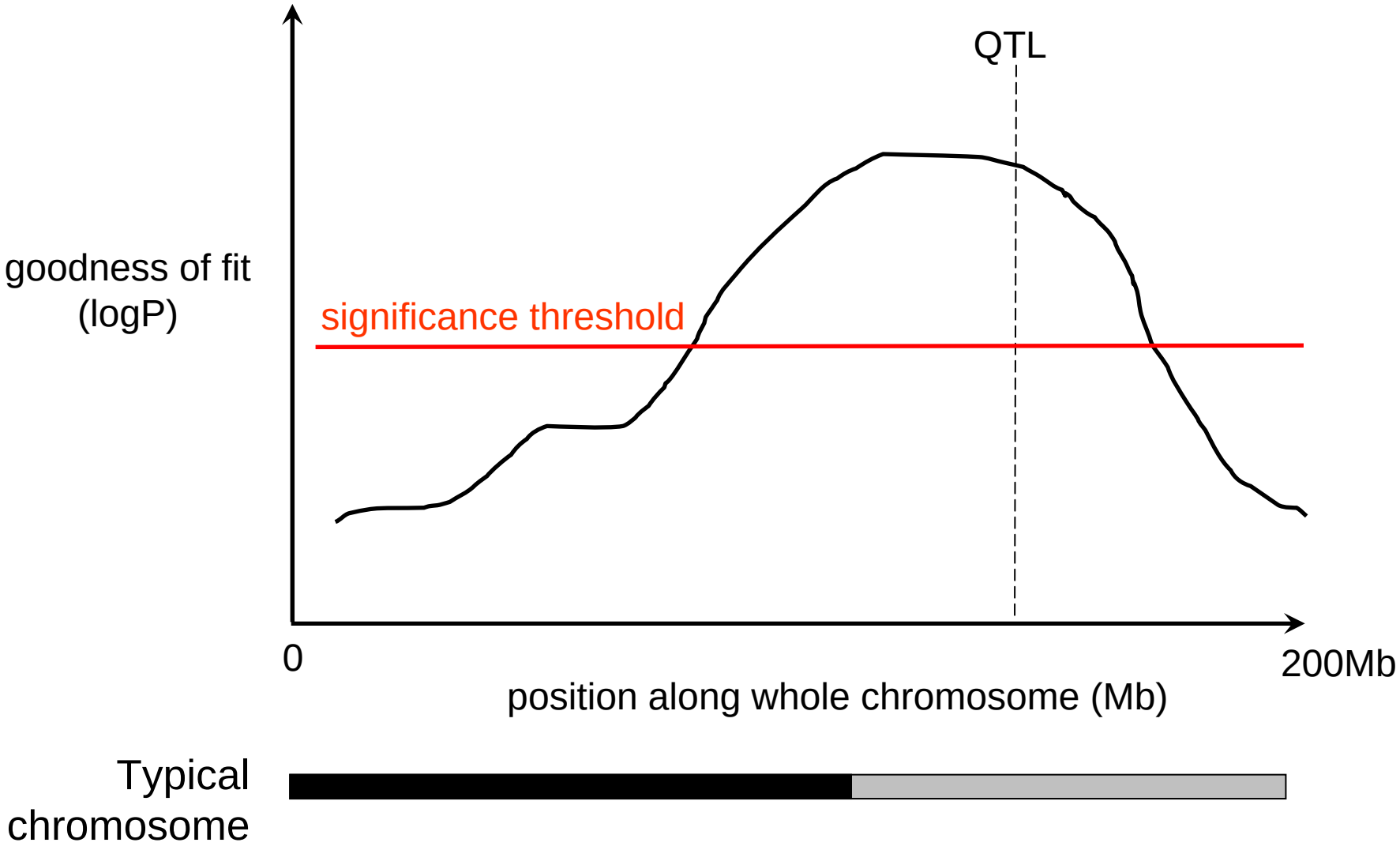

QTL
logP


QTL
logP
highly significant snp
non significant snp

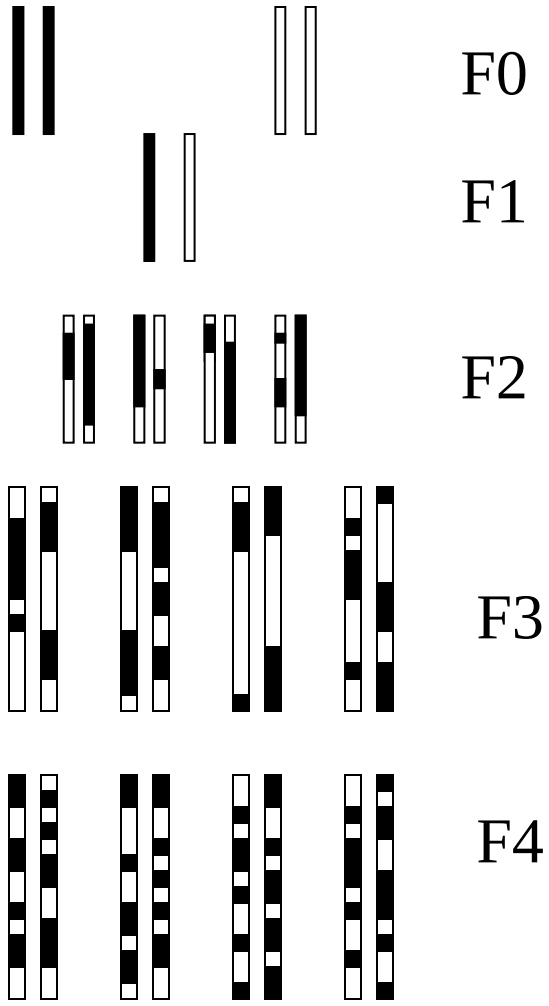

QTL
logP


QTL
logP

Chromosome scan for F2

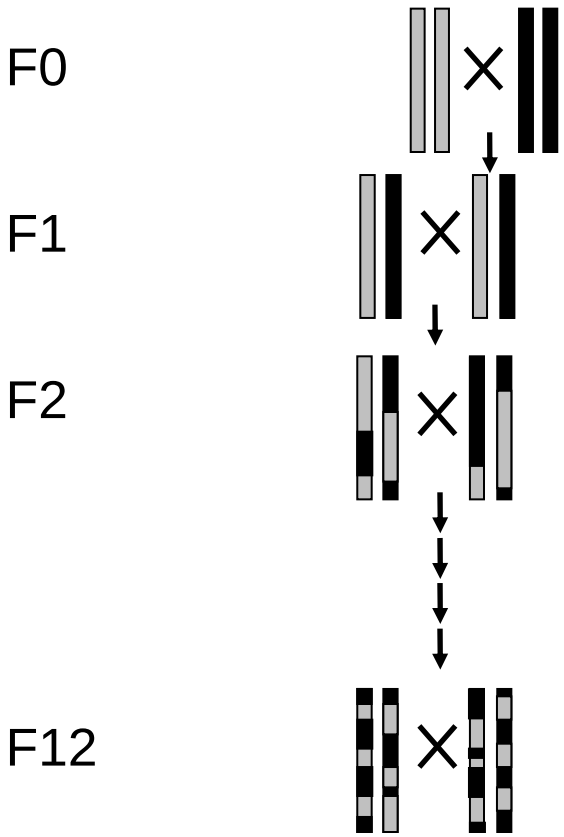


Advanced intercross lines (AILs)

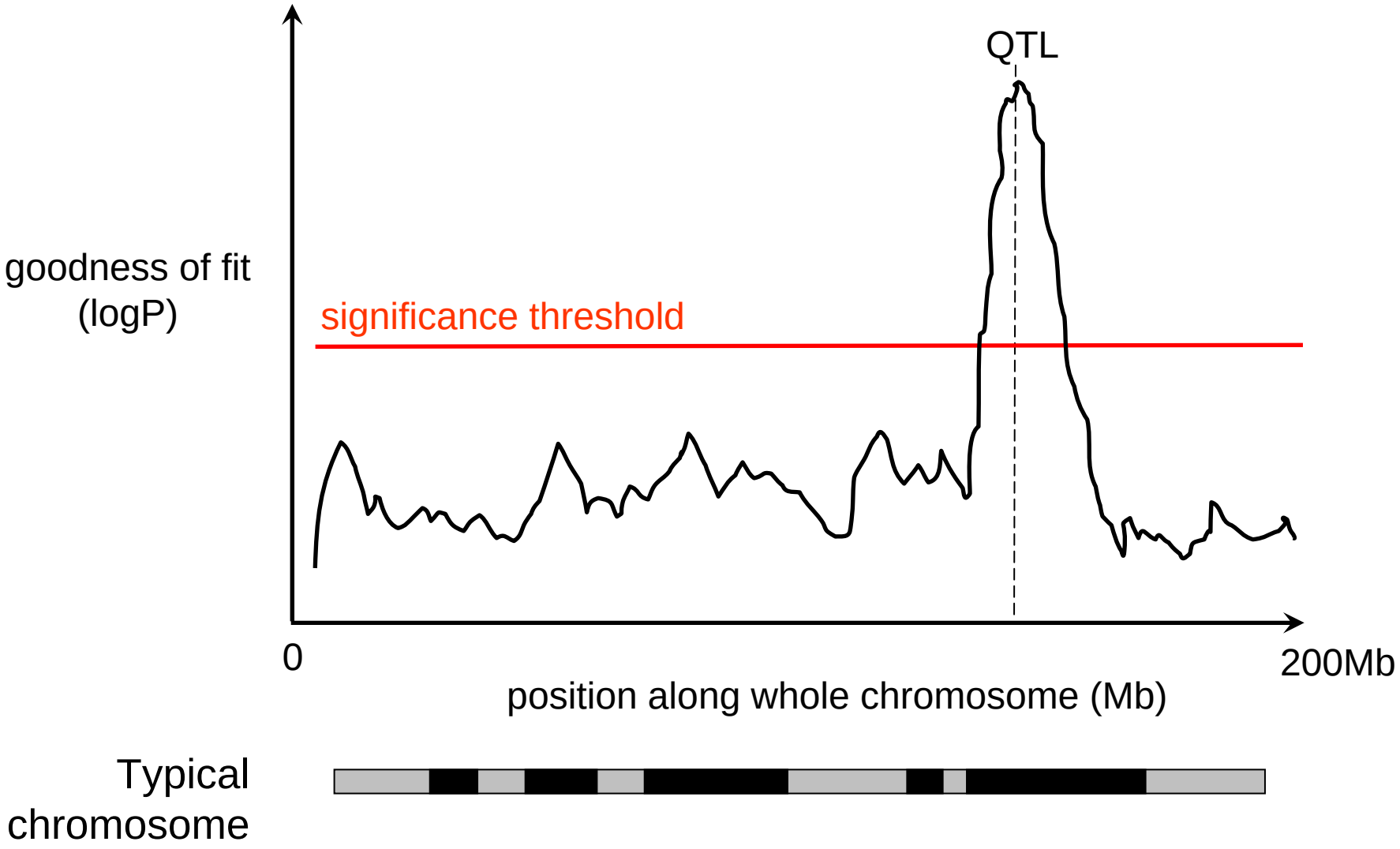


Darvasi & Soller (1995) Genetics

F12 cross



Chromosome scan for F12



Practical

1. Fitting a linear model to test a marker-phenotype association

2. Single marker association on an F2

3. Permutation test

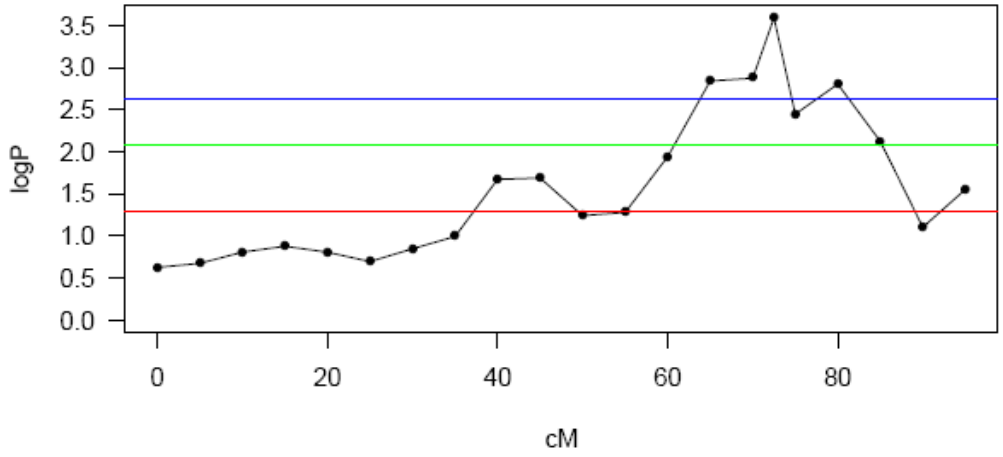
4. Single marker association on an AIL (F12)

5. Conditional modelling of loci

Start Firefox, File->Open and go to F:\valdar\ThursdayAfternoonAnimals\practical.R

Start R

Practical: F2 cross

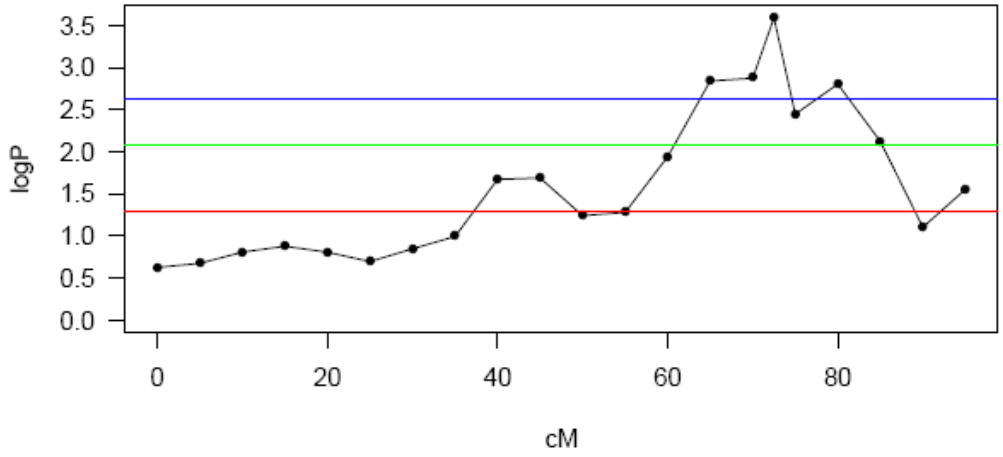


Bonferroni = 2.6

permutation ~ 2.1

uncorrected = 1.3

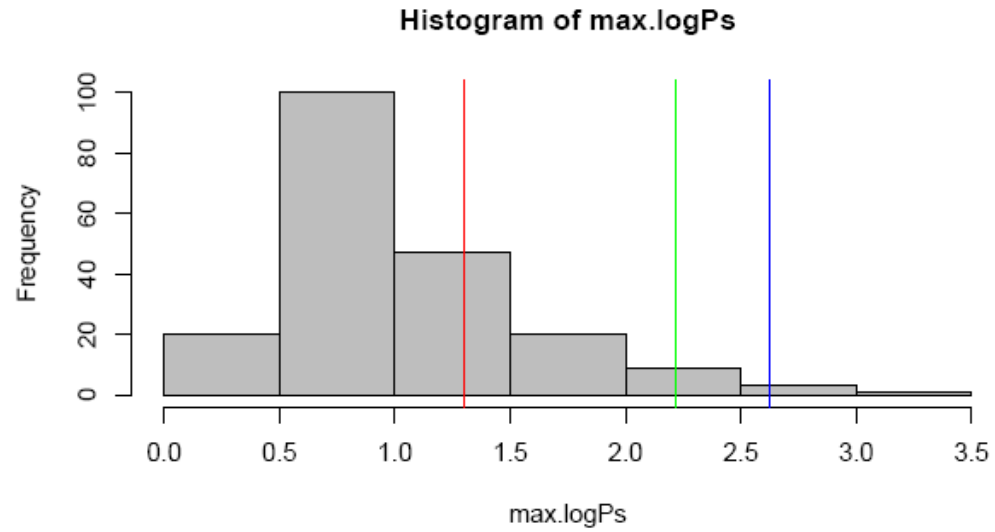
Practical: F2 cross



Bonferroni = 2.6

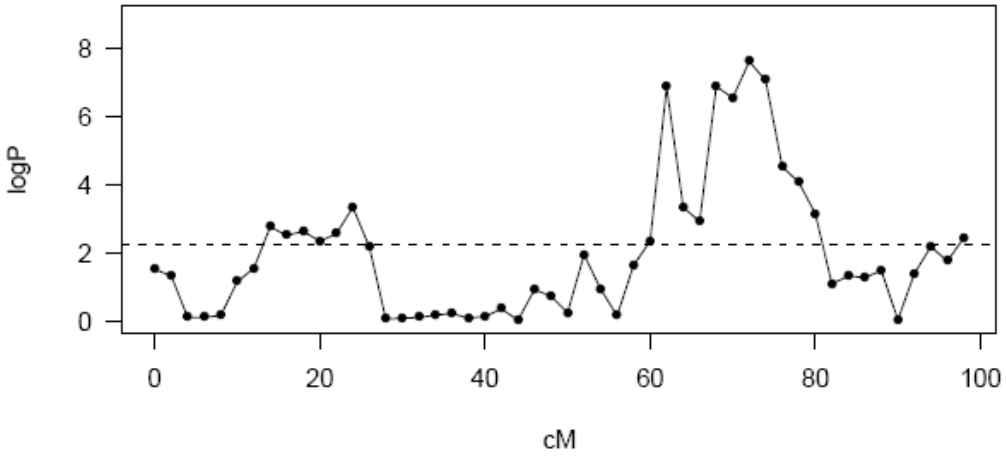
permutation ~ 2.1

uncorrected = 1.3



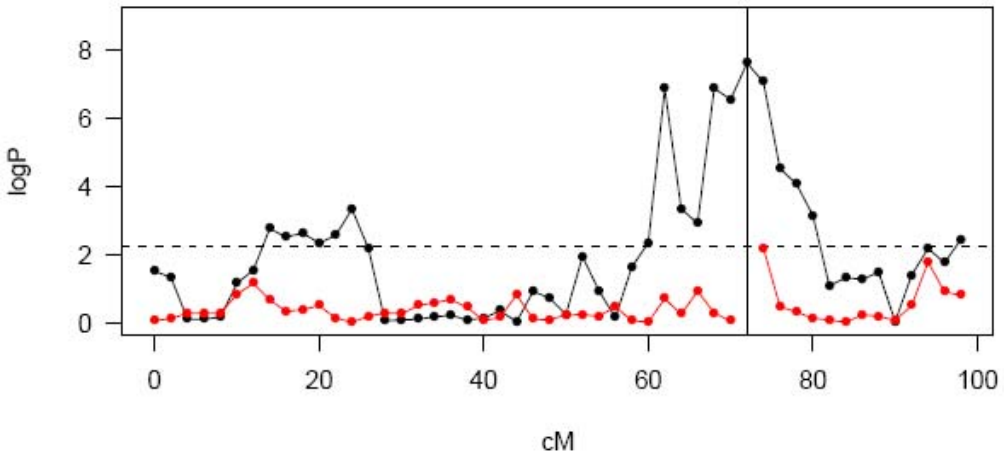
generalized extreme value (GEV) distribution

Practical: F12



phenotype ~ MARKER

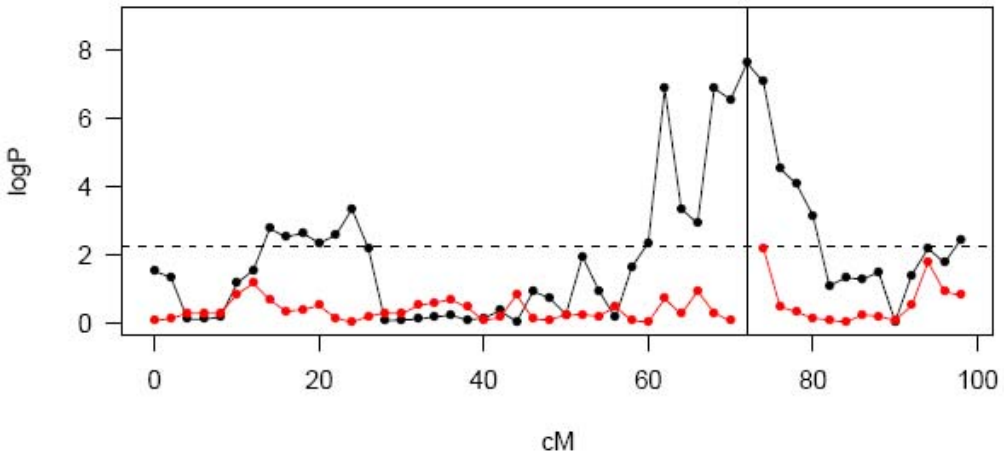
Practical: F12



phenotype ~ MARKER

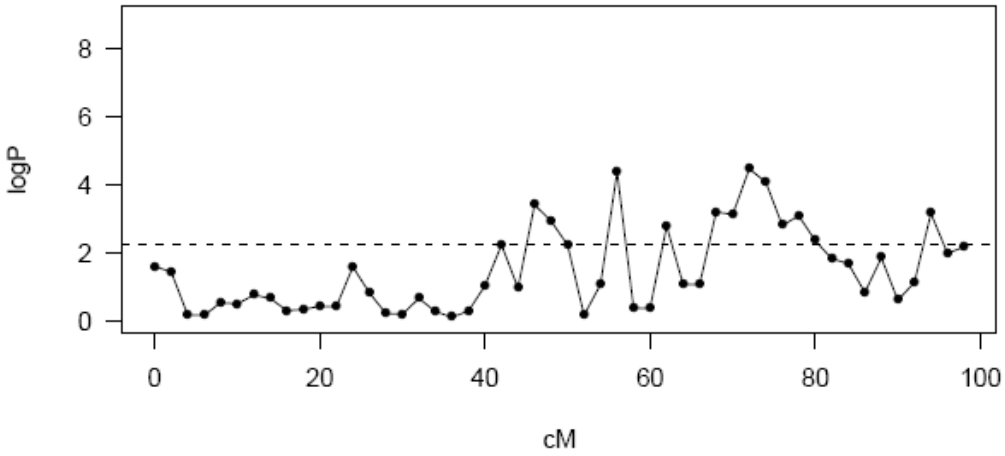
phenotype ~ m37 + MARKER

Practical: F12



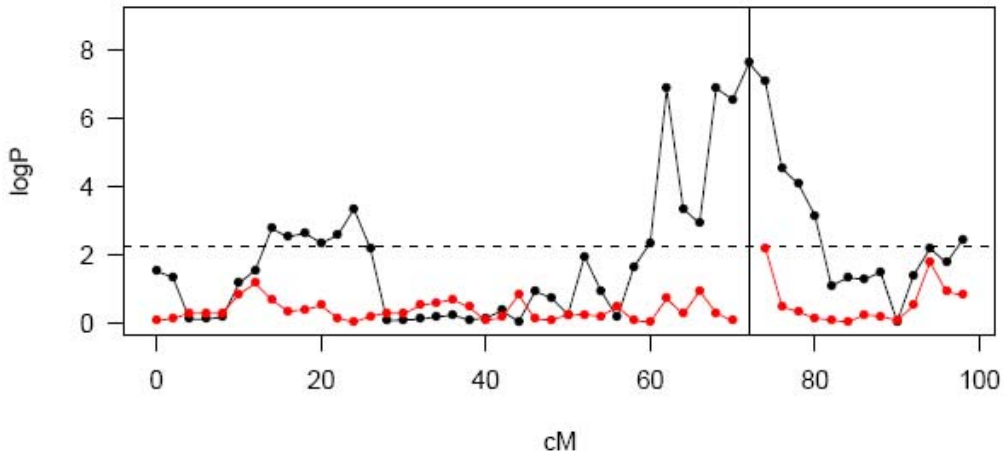
phenotype ~ MARKER

phenotype ~ m37 + MARKER



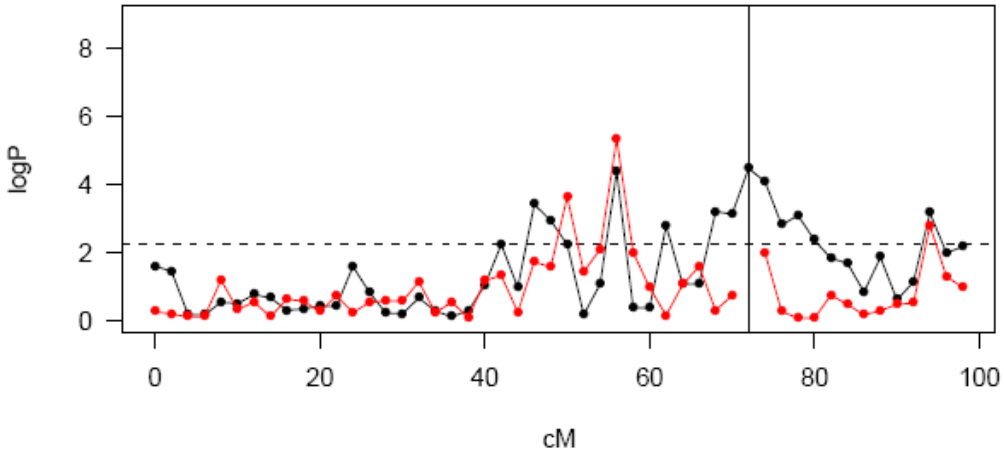
phenotype ~ MARKER

Practical: F12



phenotype ~ MARKER

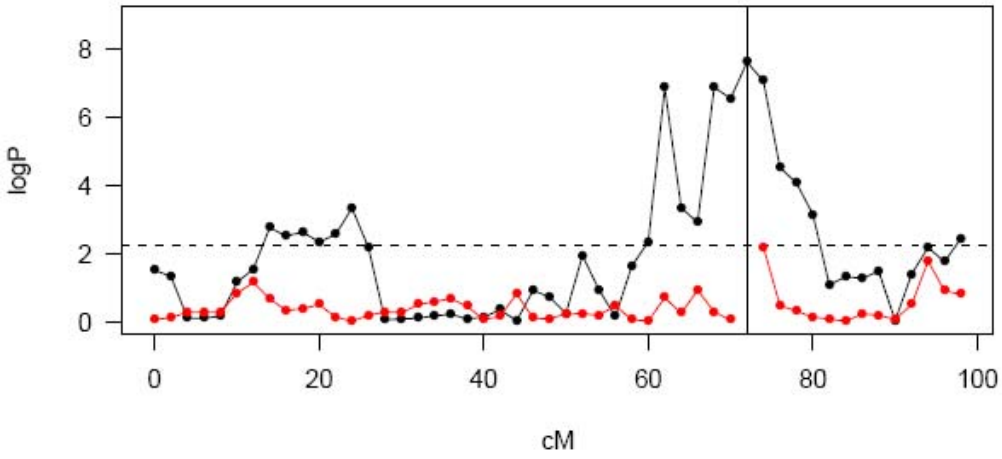
phenotype ~ m37 + MARKER



phenotype ~ MARKER

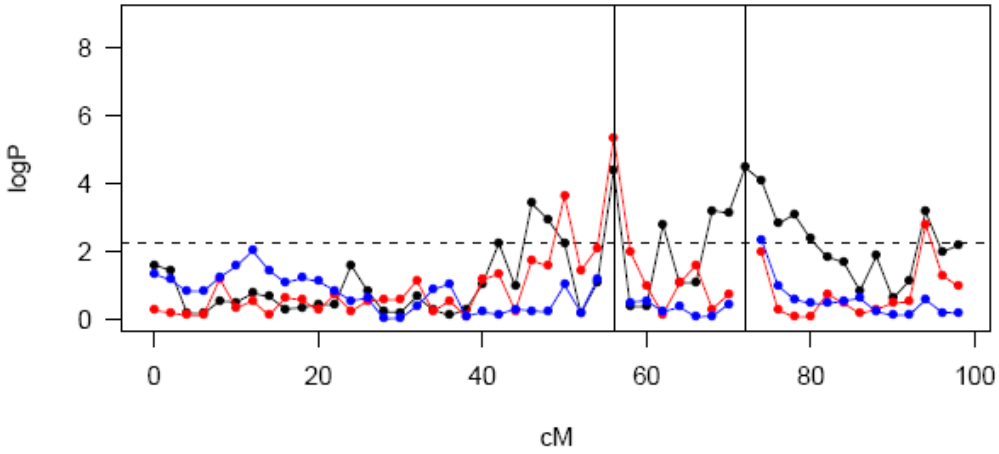
phenotype ~ m37 + MARKER

Practical: F12



phenotype ~ MARKER

phenotype ~ m37 + MARKER

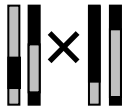


phenotype ~ MARKER

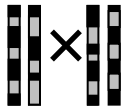
phenotype ~ m37 + MARKER

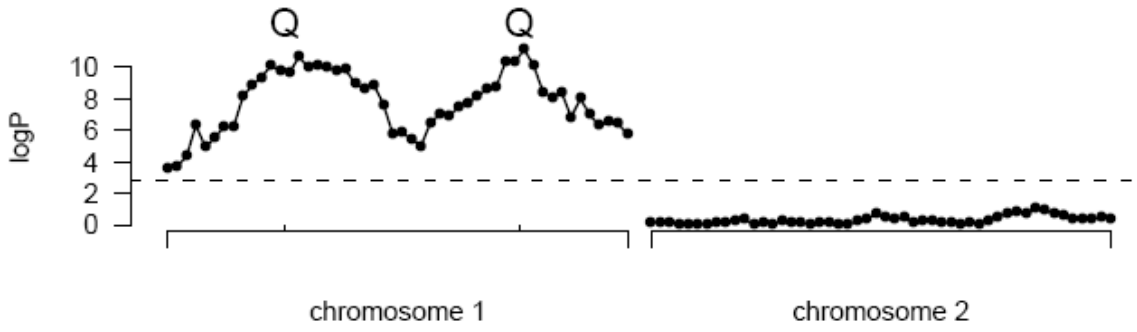
phenotype ~ m37 + m29 + MARKER

F2

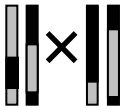


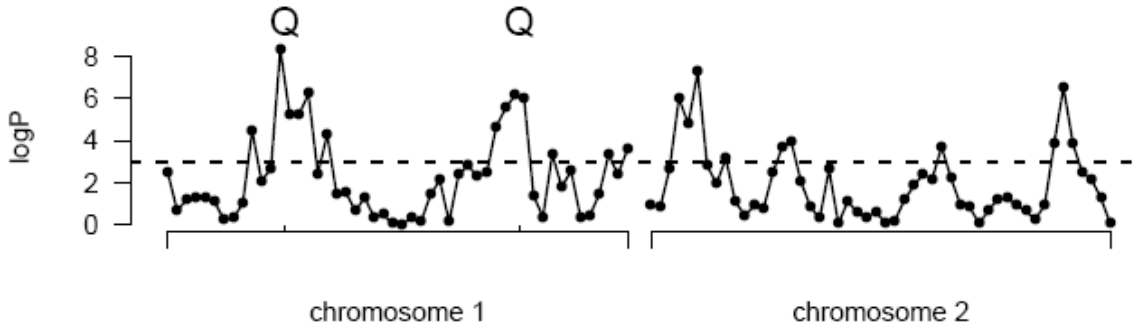
F18



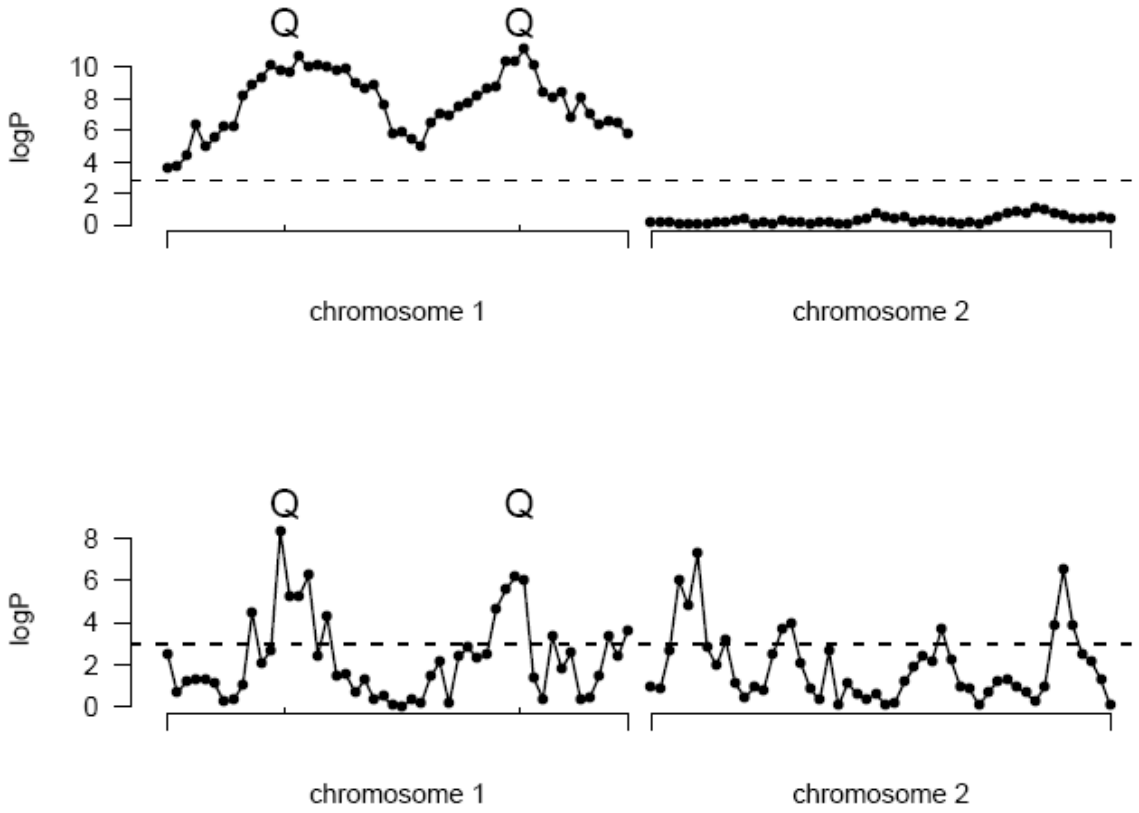

Q
Q
logP
10
8
6
4
2
0
chromosome 1
chromosome 2

F2

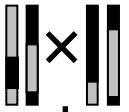



Q
Q
logP
8
6
4
2
0
chromosome 1
chromosome 2

F18

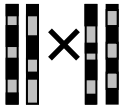


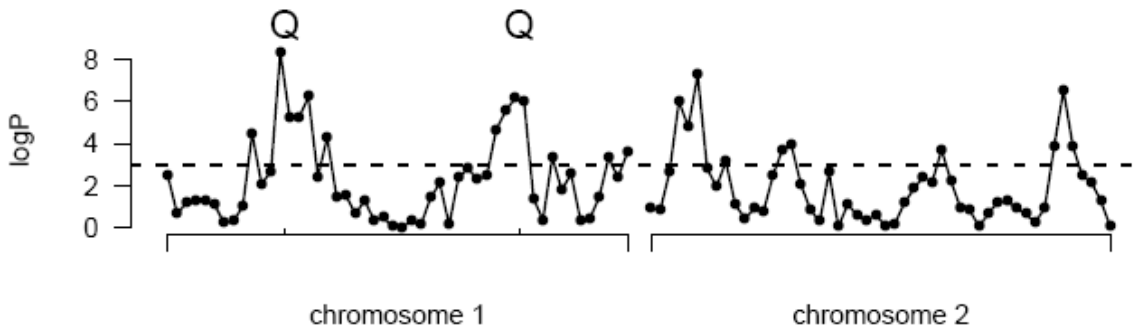
F2



breeding
in a small
population

F18

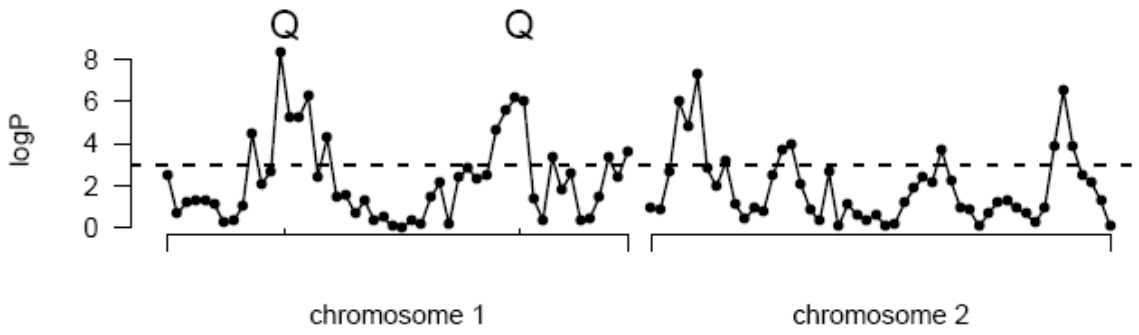


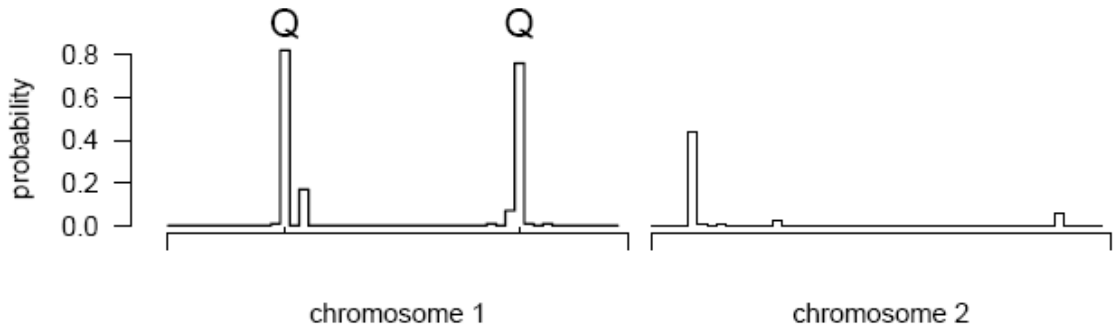


population structure

gross genetic differences between groups

where groups = families





population structure

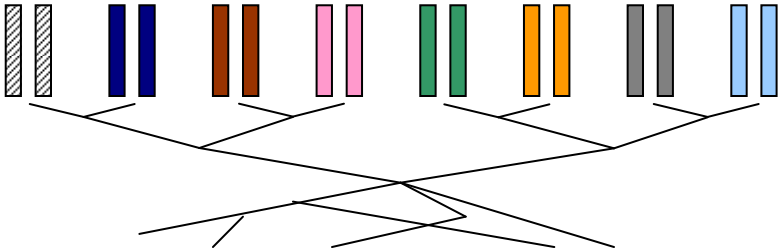
gross genetic differences between groups

where groups = families

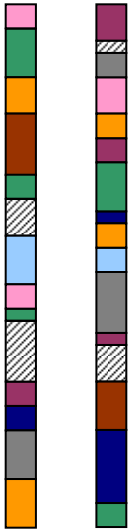
multilocus approach

Heterogeneous Stocks

Heterogeneous Stock



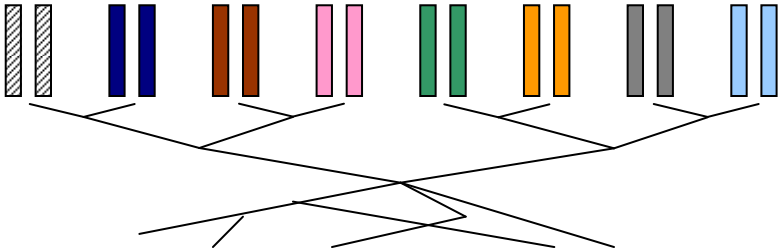
Pseudo-random mating
for 50 generations



Avg. Distance Between
Recombinations:

HS
~2 cM

Heterogeneous Stock

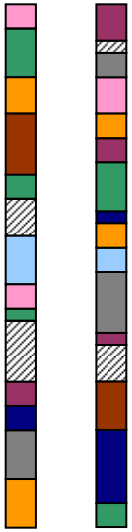


Pseudo-random mating
for 50 generations

F2 Intercross

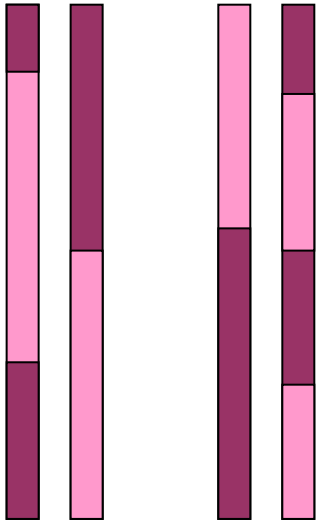
x

F1



Avg. Distance Between
Recombinations:

HS	F2 intercross
~2 cM	~30 cM



F2

124 Phenotypes

Anxiety [24]

Asthma [13]

Biochemistry [15]

Bone Morphology [23]

Diabetes [16]

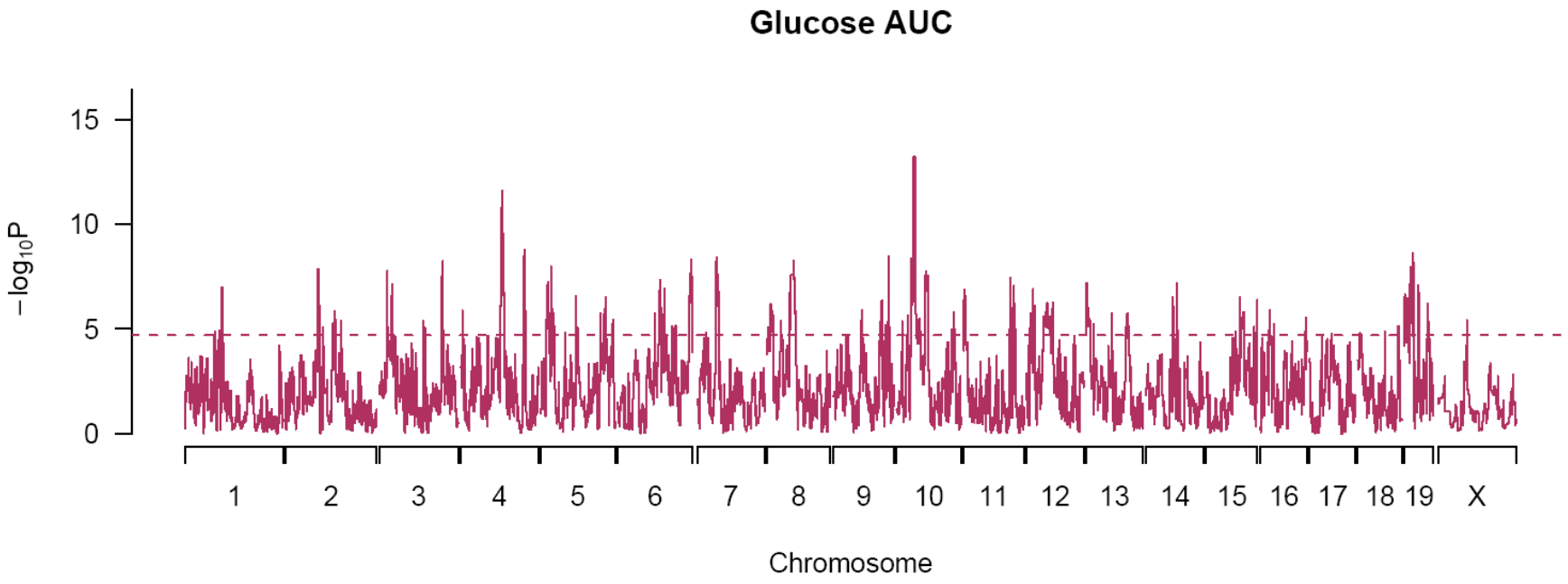
Haematology [15]

Immunology [9]

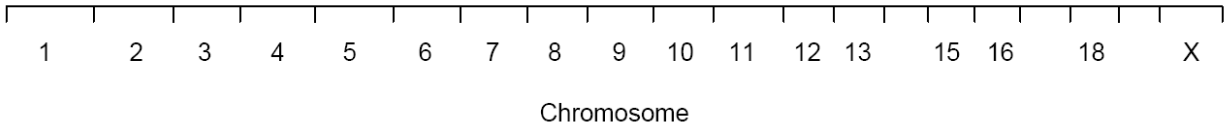
Weight/size related [8]

Wound Healing [1]

Intraperitoneal Glucose Tolerance Test


Glucose AUC
15
10
5
0
−log10P
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome

How to select peaks: a simulated example



How to select peaks: a simulated example

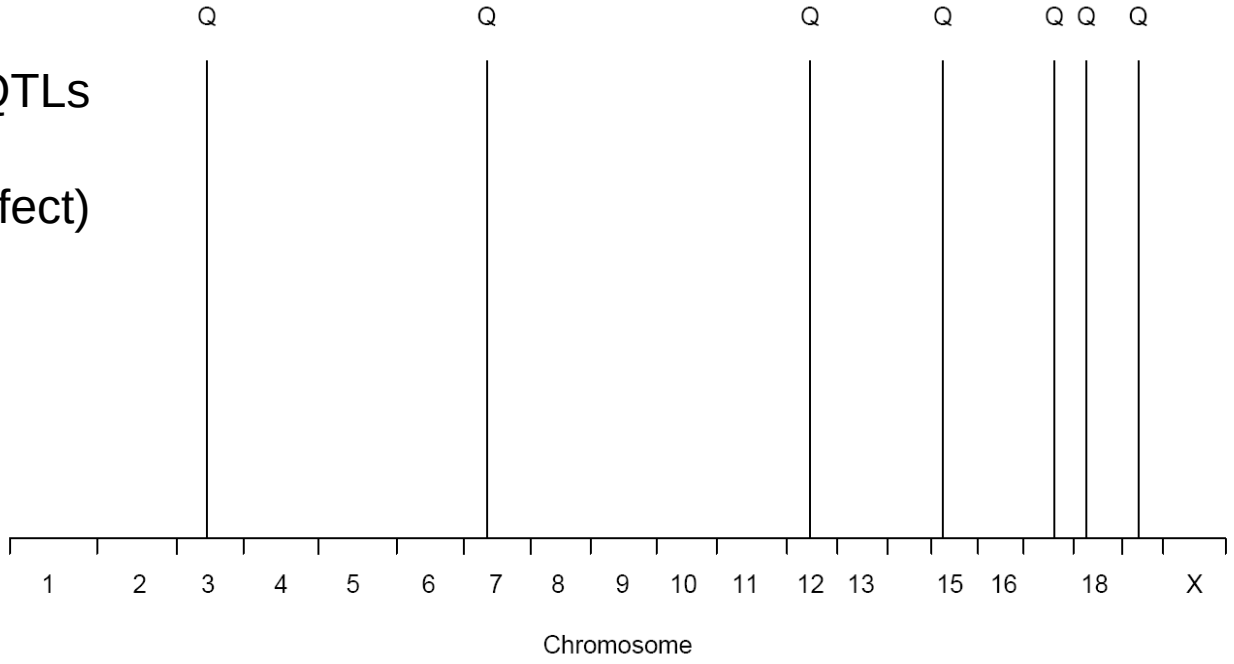
Simulate 7 x 5% QTLs

(ie, 35% genetic effect)

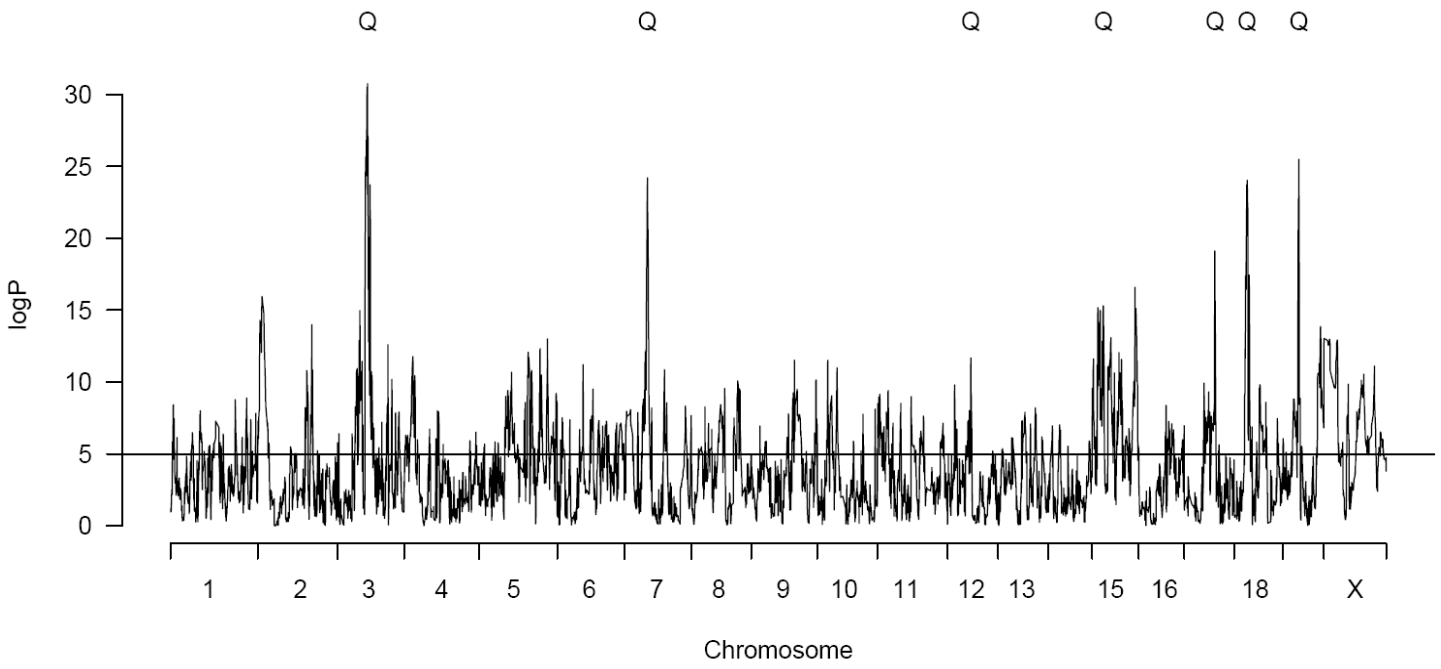
+ 20% shared environment effect

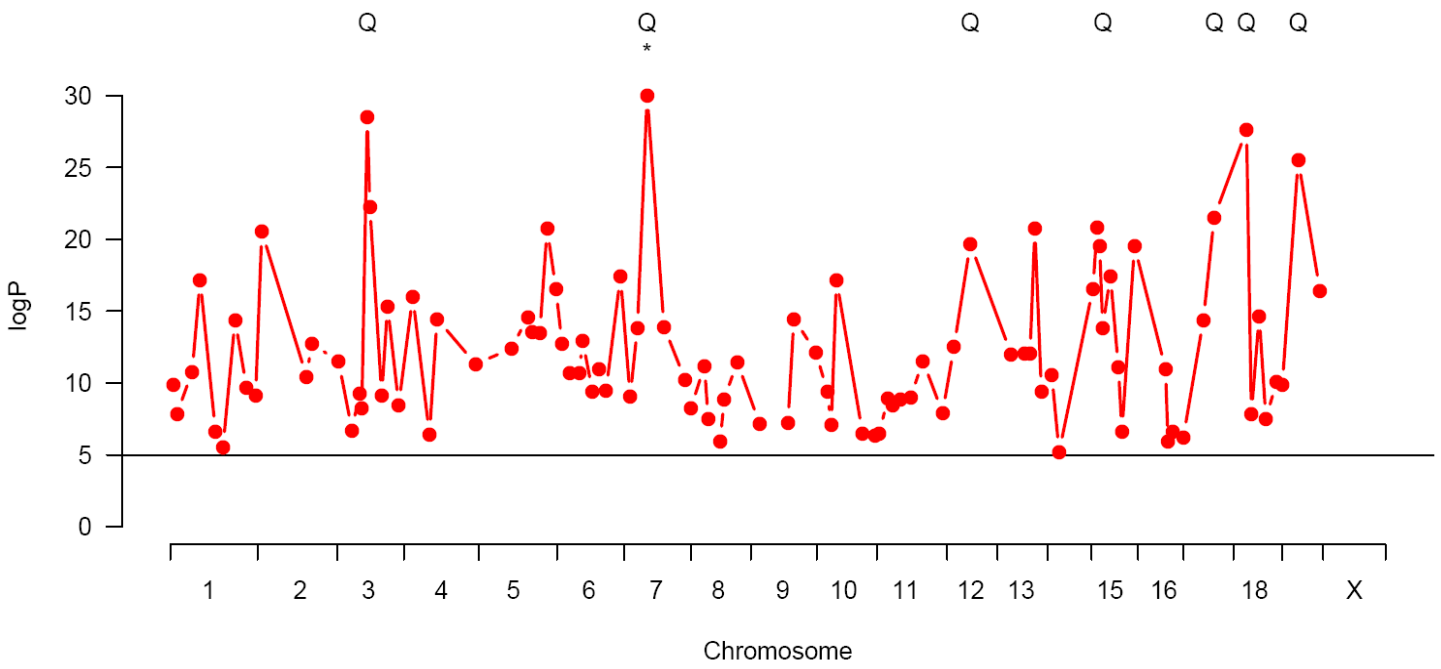
+ 45% noise

= 100% variance



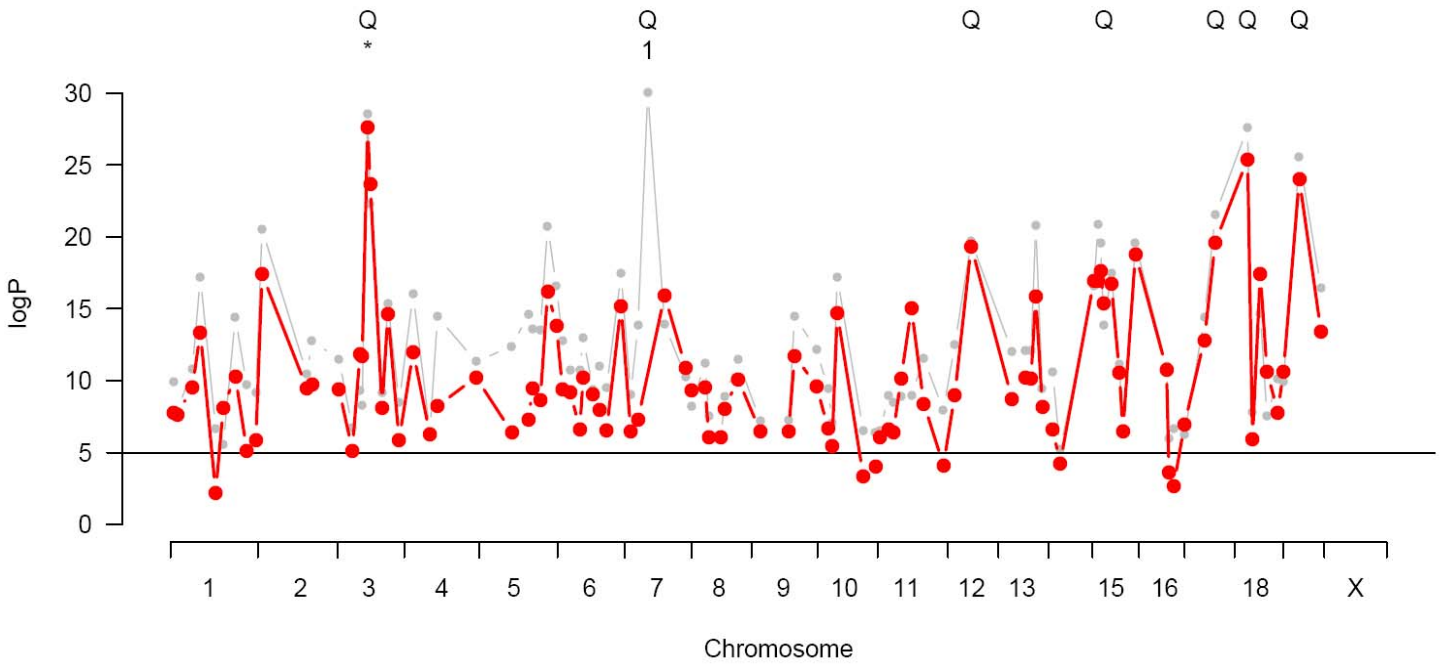
Simulated example




Q
Q
*
Q
Q
Q
Q
Q
30
25
20
15
10
5
0
logP
1
2
3
4
5
6
7
8
9
10
11
12
13
15
16
18
X
Chromosome

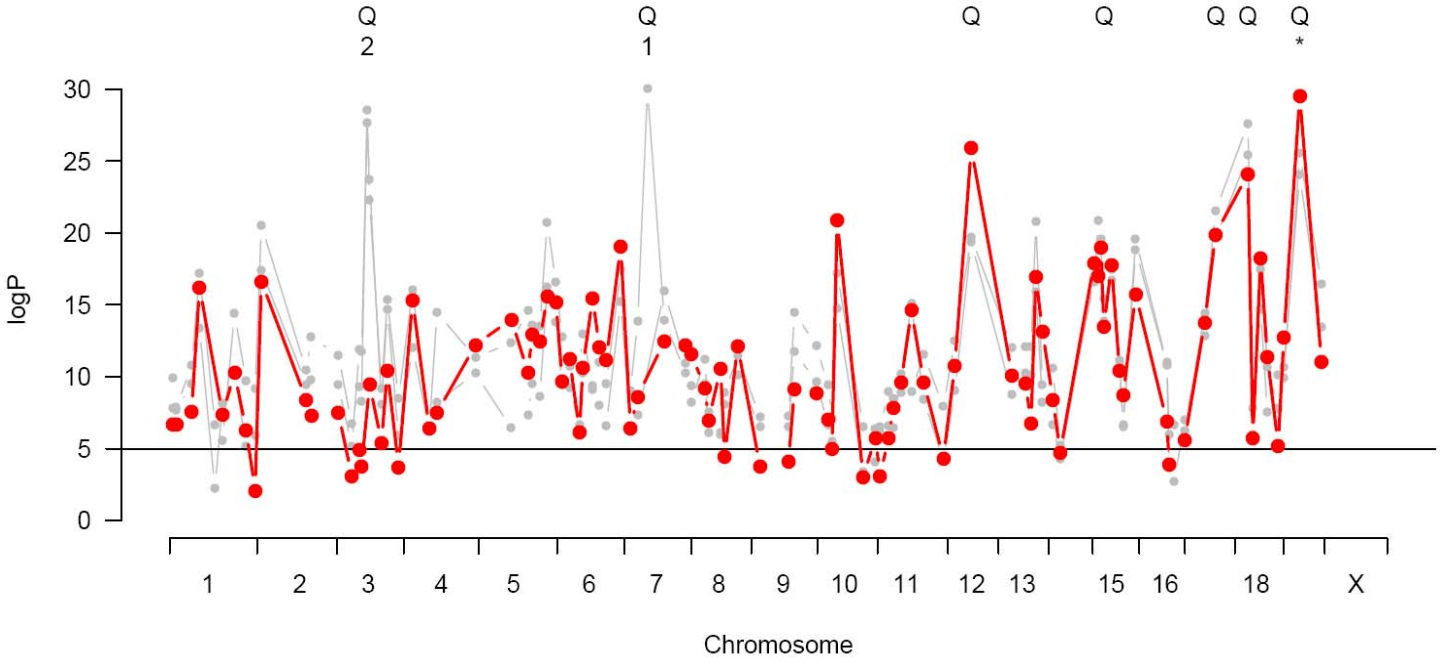
phenotype ~ ?

condition on 1 peak



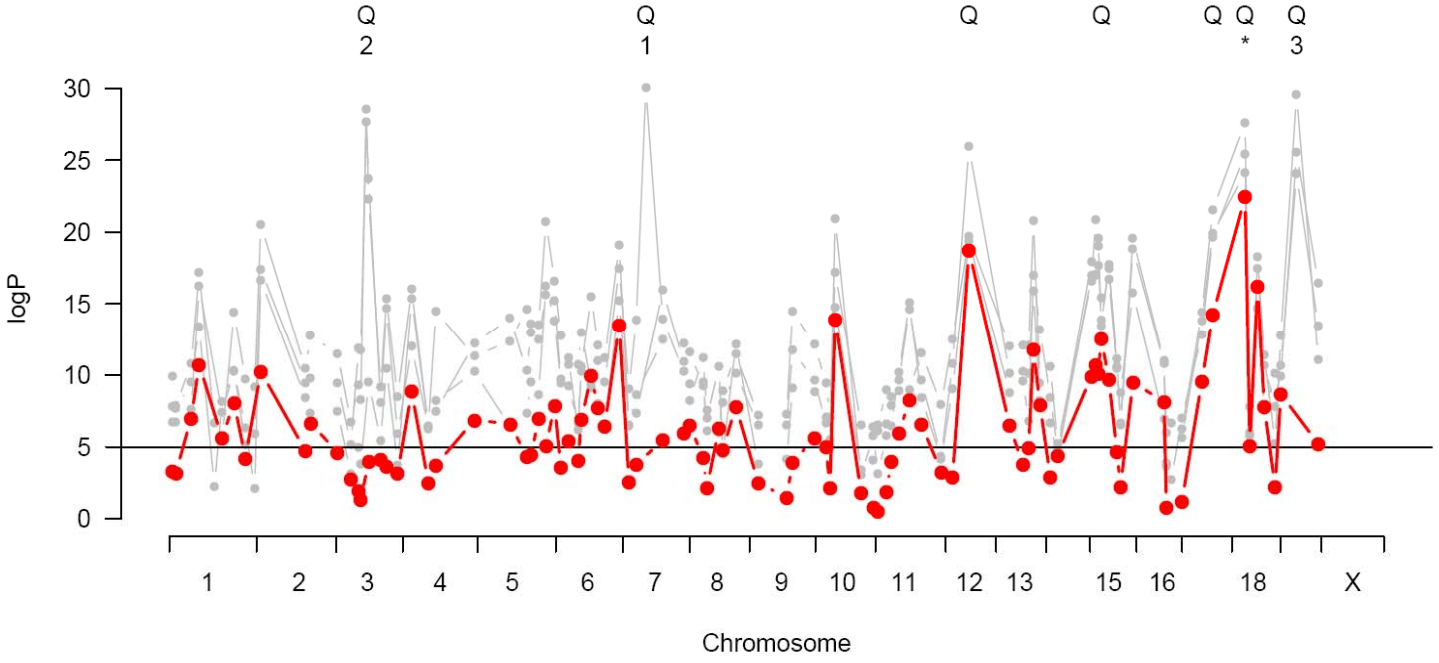
phenotype ~ peak 1 + ?

condition on 2 peaks



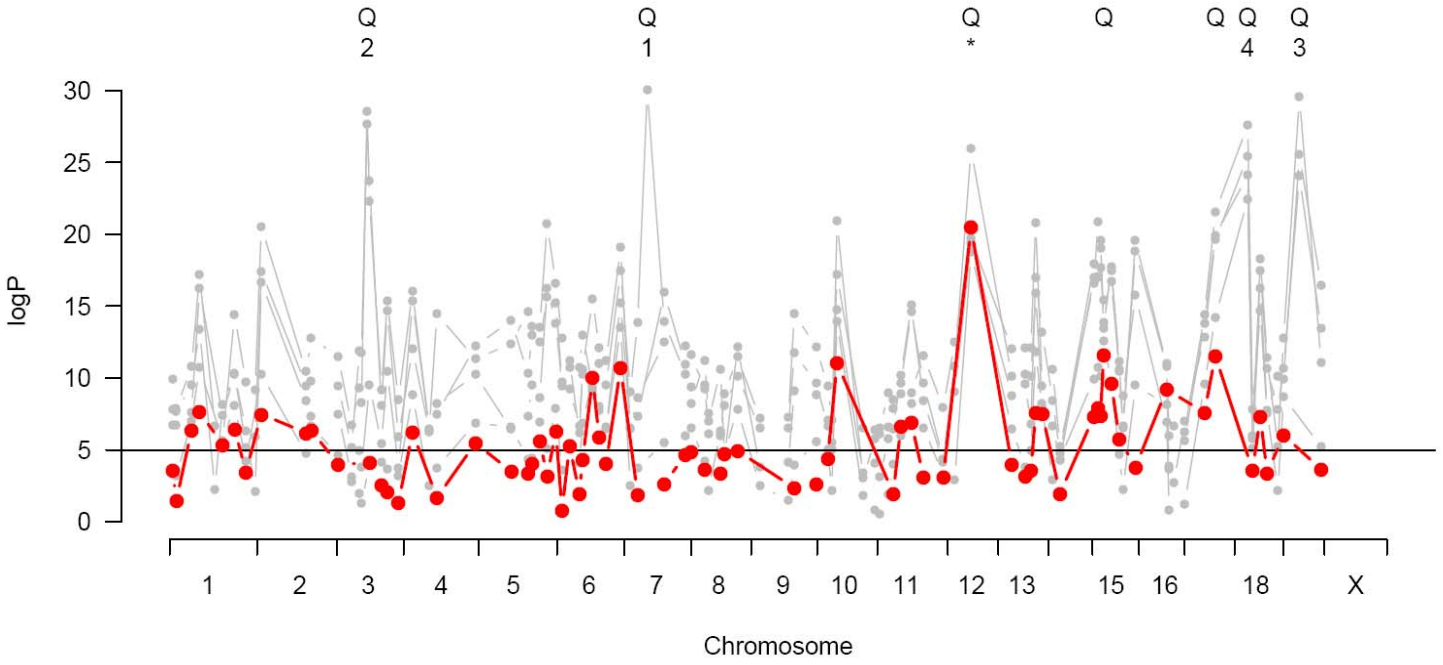
phenotype ~ peak 1 + peak 2 + ?

condition on 3 peaks



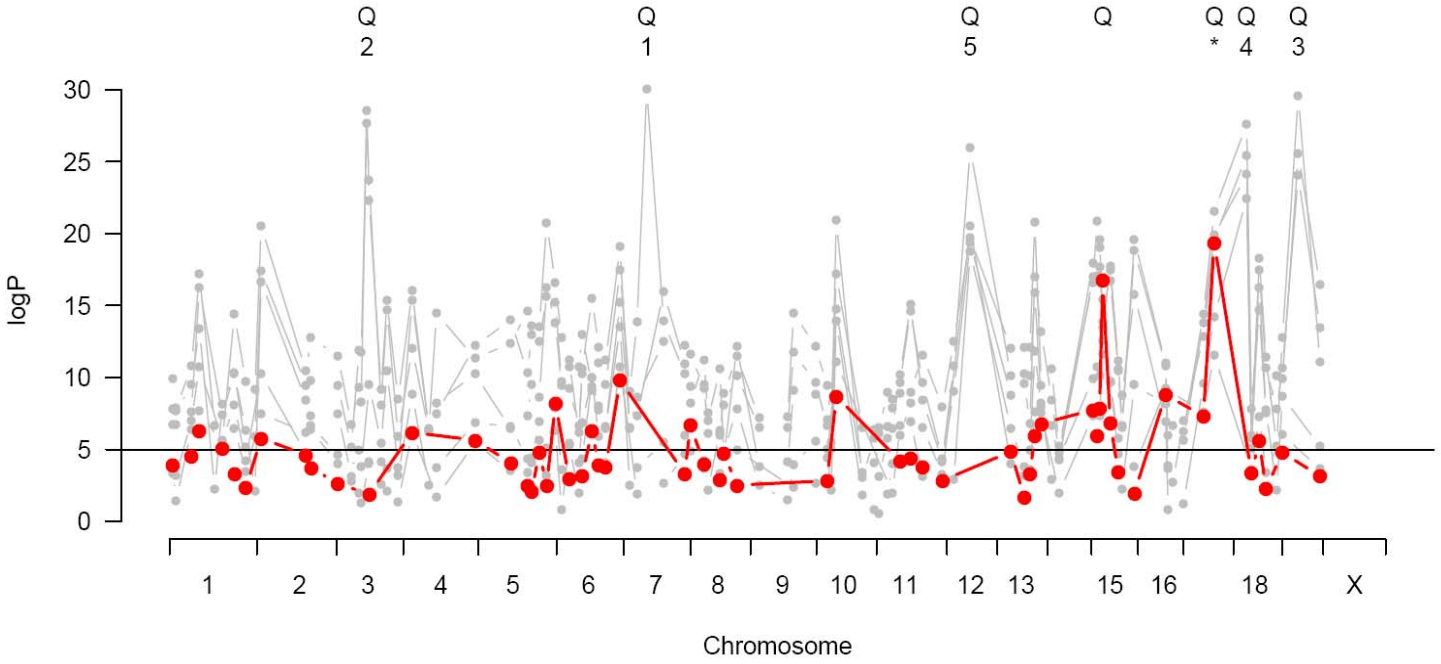
phenotype ~ peak 1 + peak 2 + peak 3 + ?

condition on 4 peaks



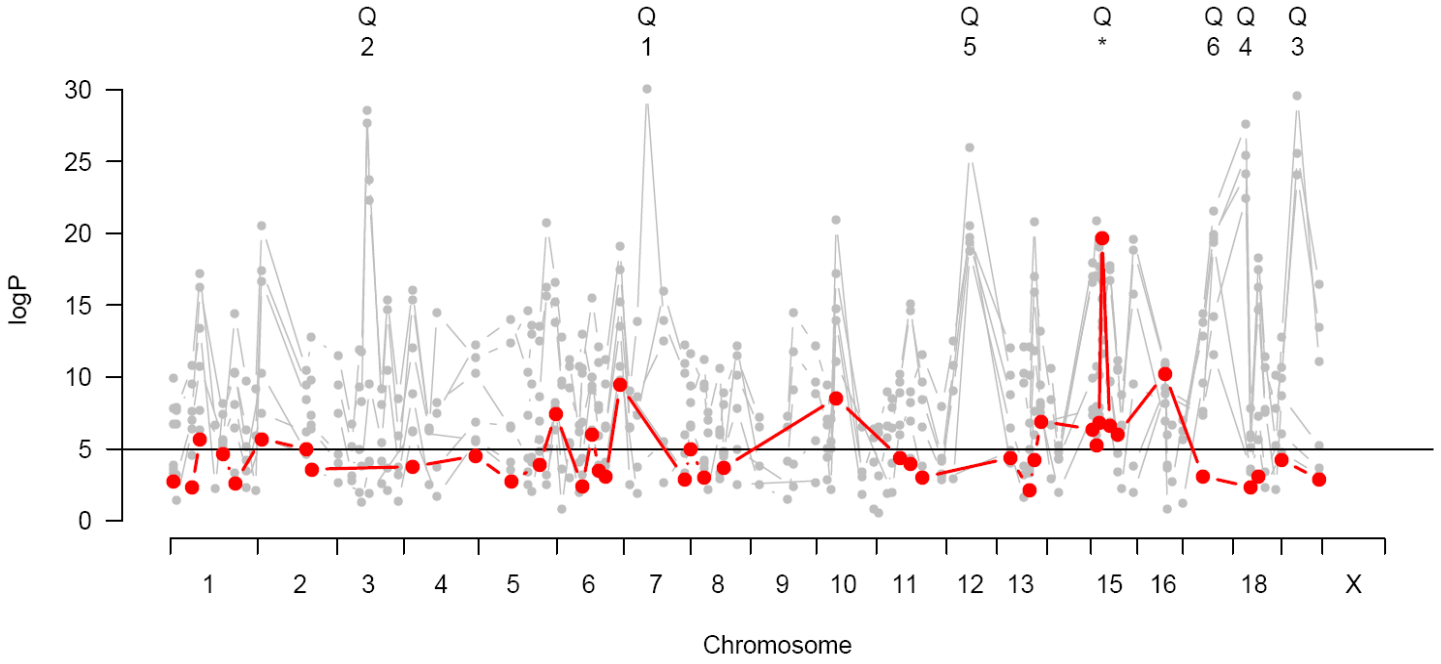
phenotype ~ peak 1 + peak 2 + peak 3 + peak 4 + ?

condition on 5 peaks



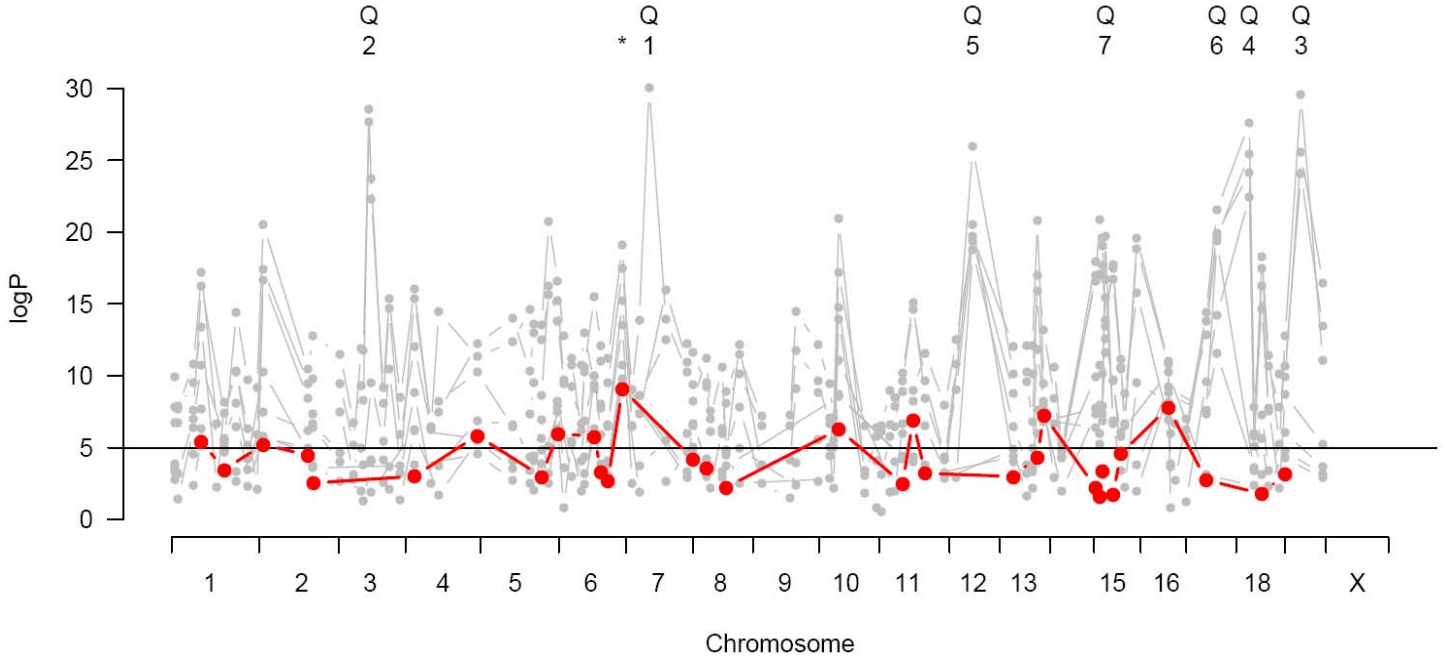
phenotype ~ peak 1 + peak 2 + peak 3 + peak 4 + peak 5 + ?

condition on 6 peaks



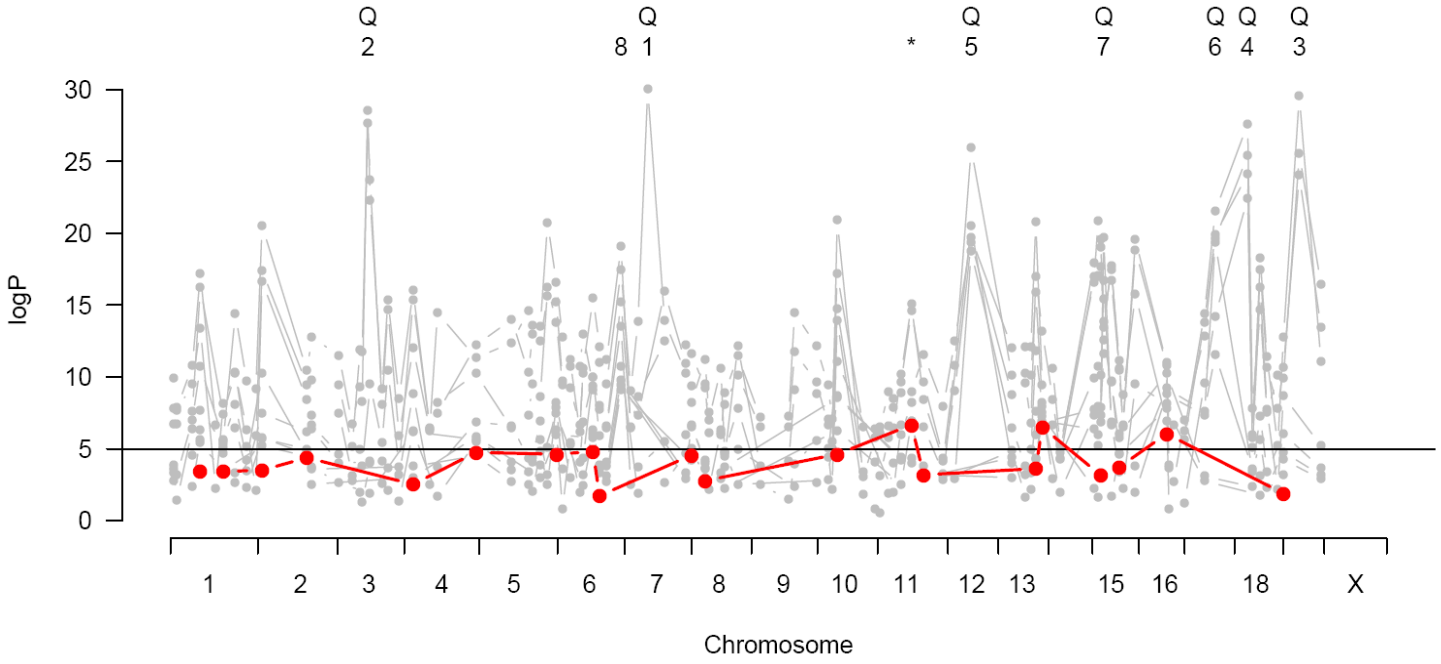
phenotype ~ peak 1 + peak 2 + peak 3 + peak 4 + peak 5 + peak 6 + ?

condition on 7 peaks



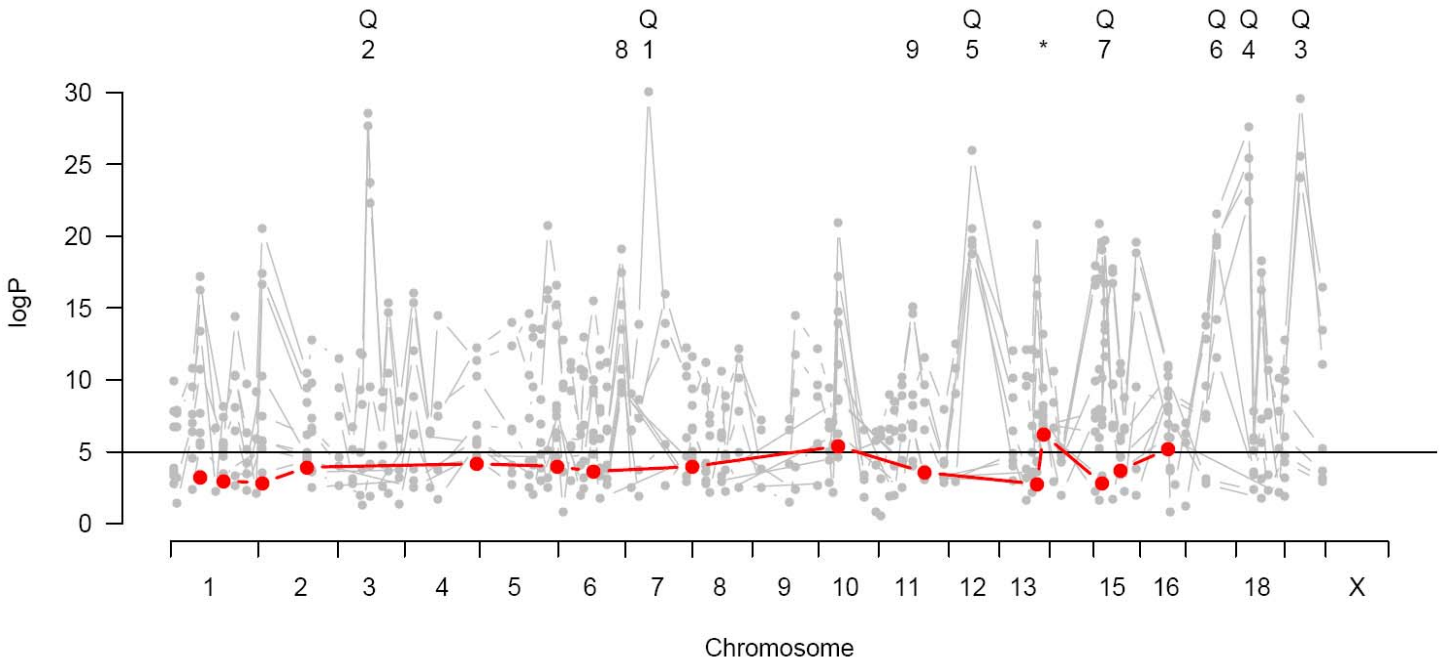
phenotype ~ peak 1 + peak 2 + peak 3 + peak 4 + peak 5 + peak 6 + peak 7 + ?

condition on 8 peaks



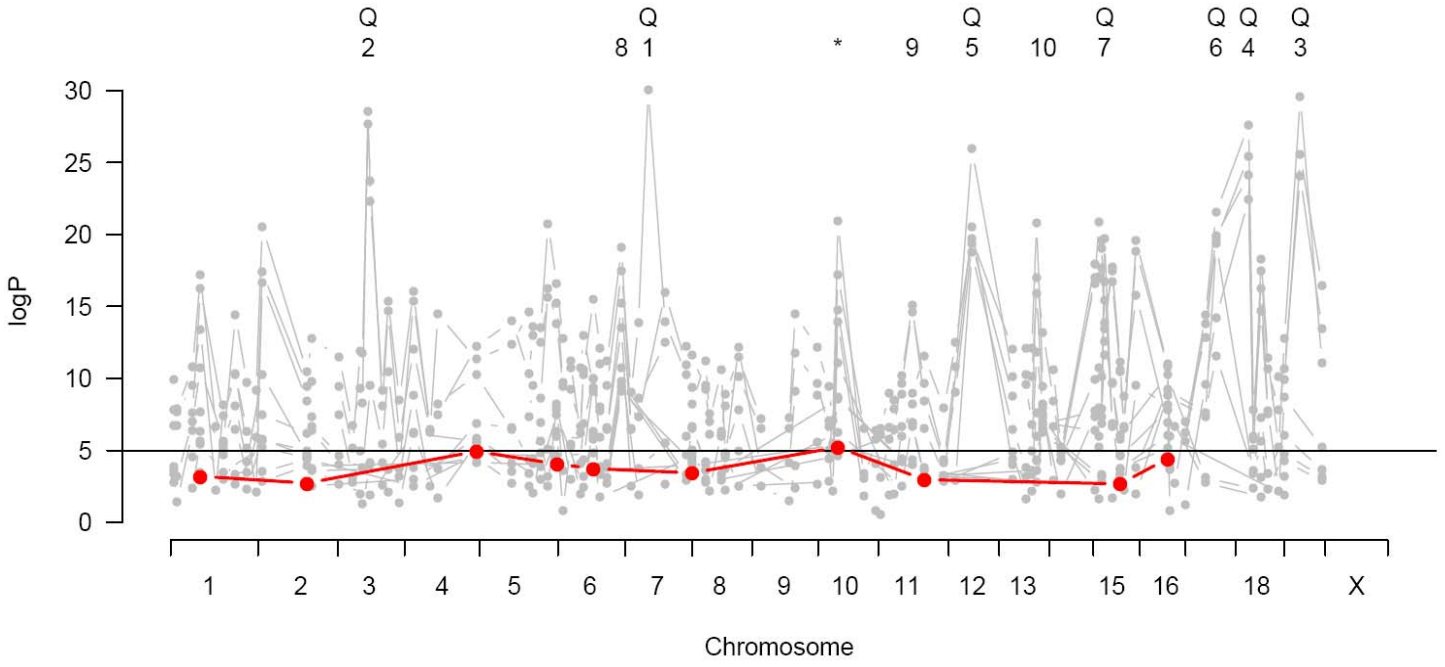
phenotype ~ peak 1 + peak 2 + peak 3 + peak 4 + peak 5 + peak 6 + peak 7 + peak 8 + ?

condition on 9 peaks



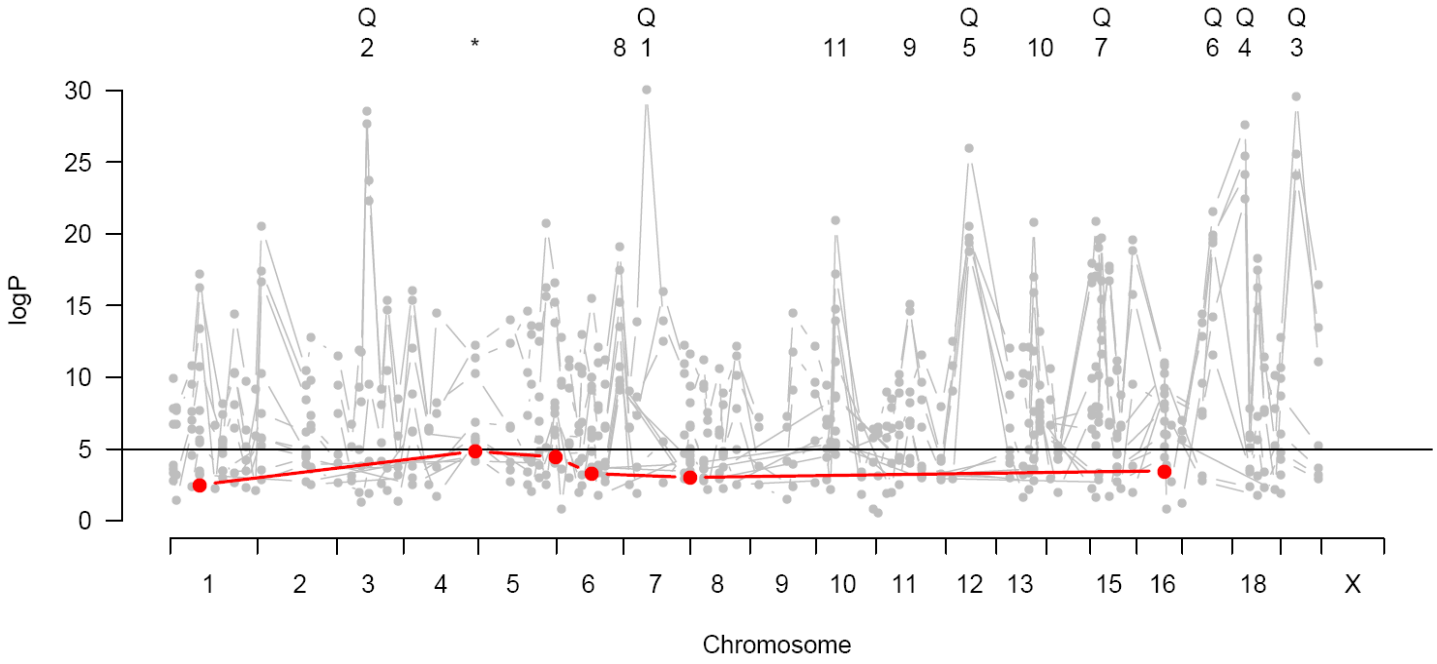
phenotype ~ peak 1 + peak 2 + peak 3 + peak 4 + peak 5 + peak 6 + peak 7 + peak 8 + peak 9 + ?

condition on 10 peaks



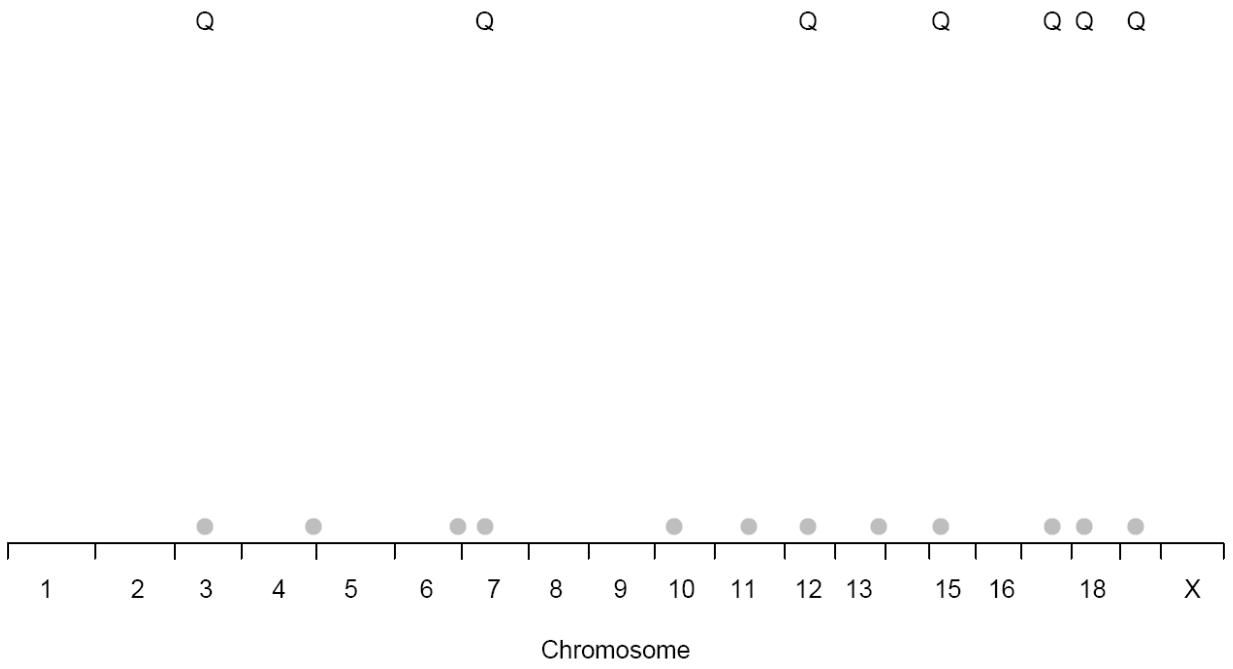
phenotype ~ peak 1 + peak 2 + peak 3 + peak 4 + peak 5 + peak 6 + peak 7 + peak 8 + peak 9 + peak 10 + ?

condition on 11 peaks



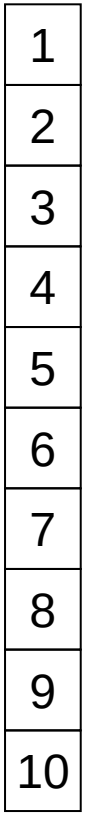
phenotype ~ peak 1 + peak 2 + peak 3 + peak 4 + peak 5 + peak 6 + peak 7 + peak 8 + peak 9 + peak 10 + peak 11 + ?

Peaks chosen by forward selection

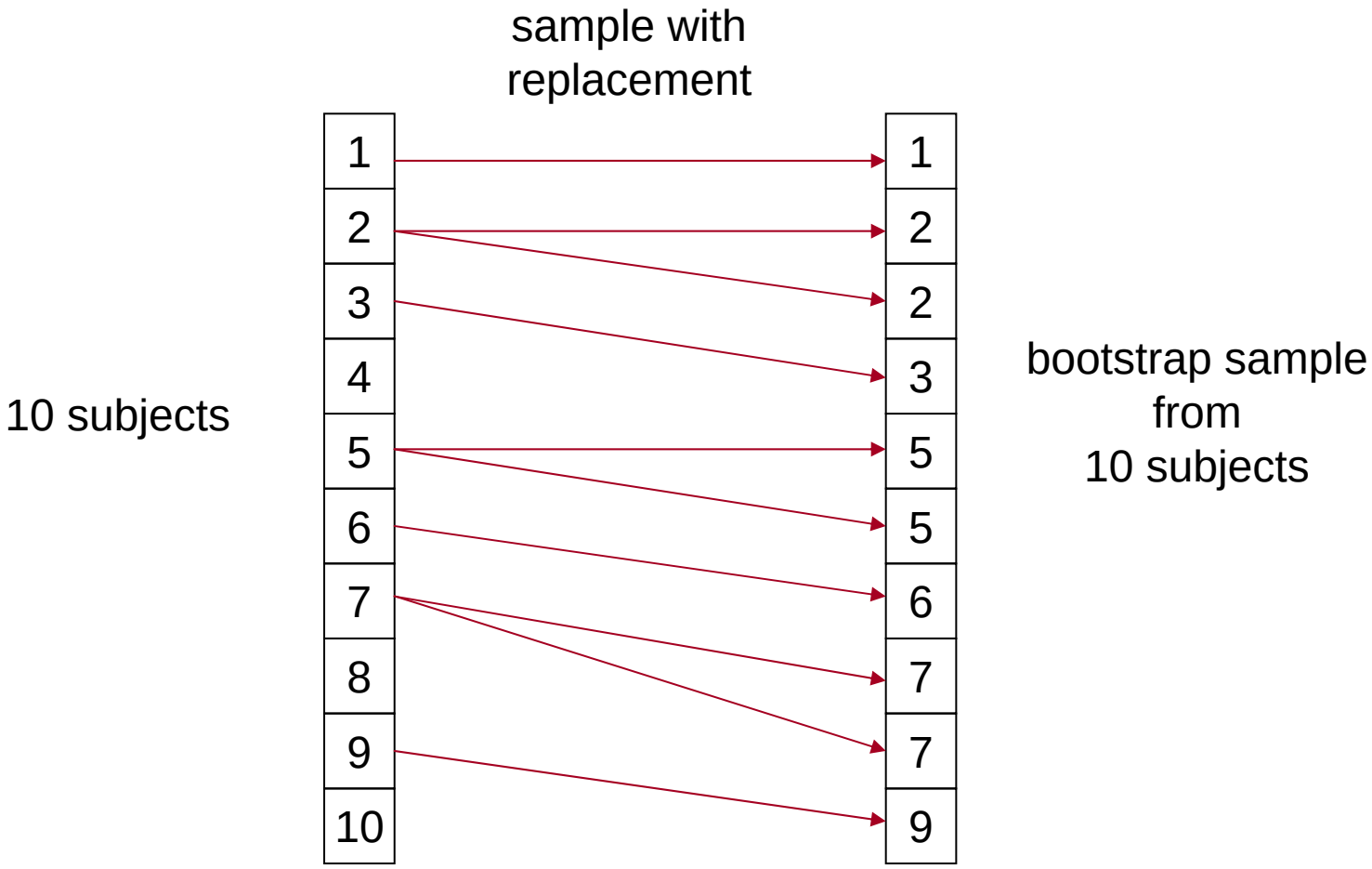


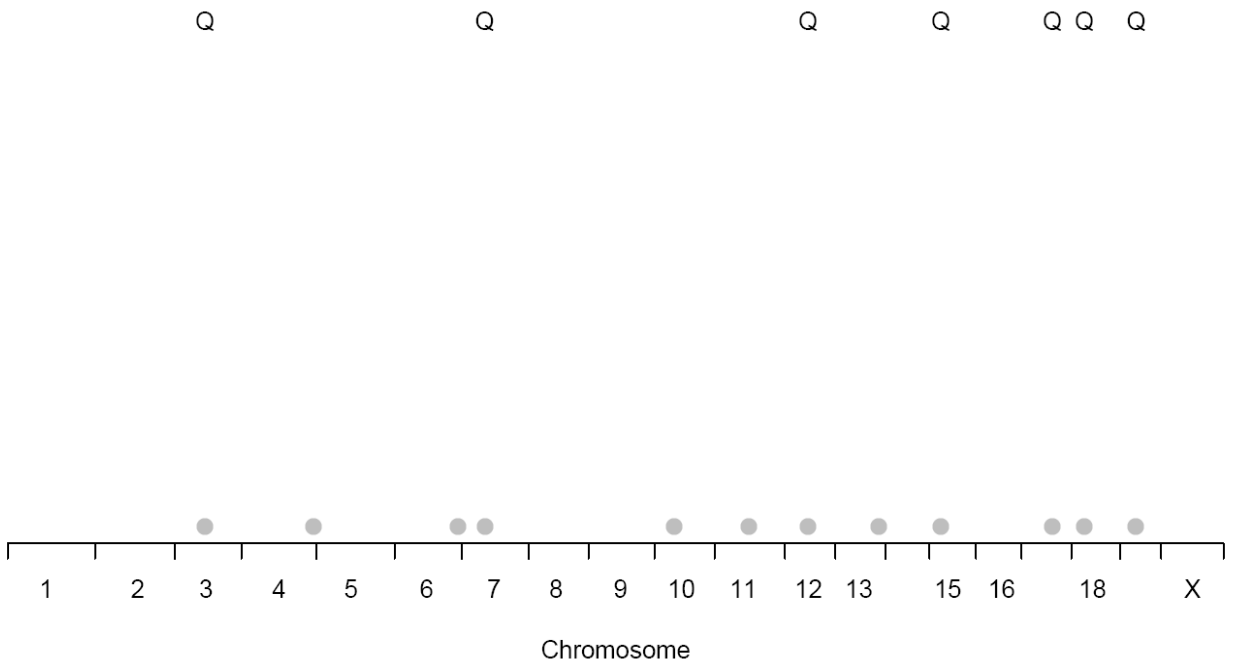
Bootstrap sampling

10 subjects

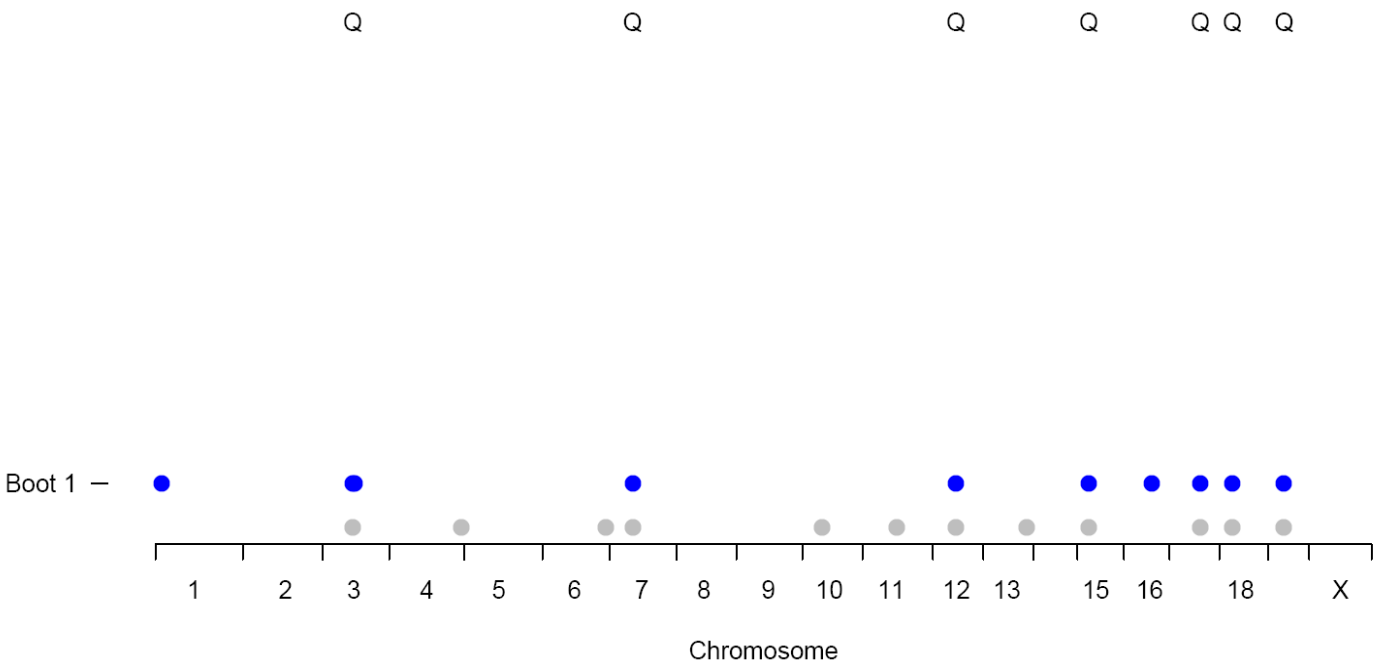


Bootstrap sampling

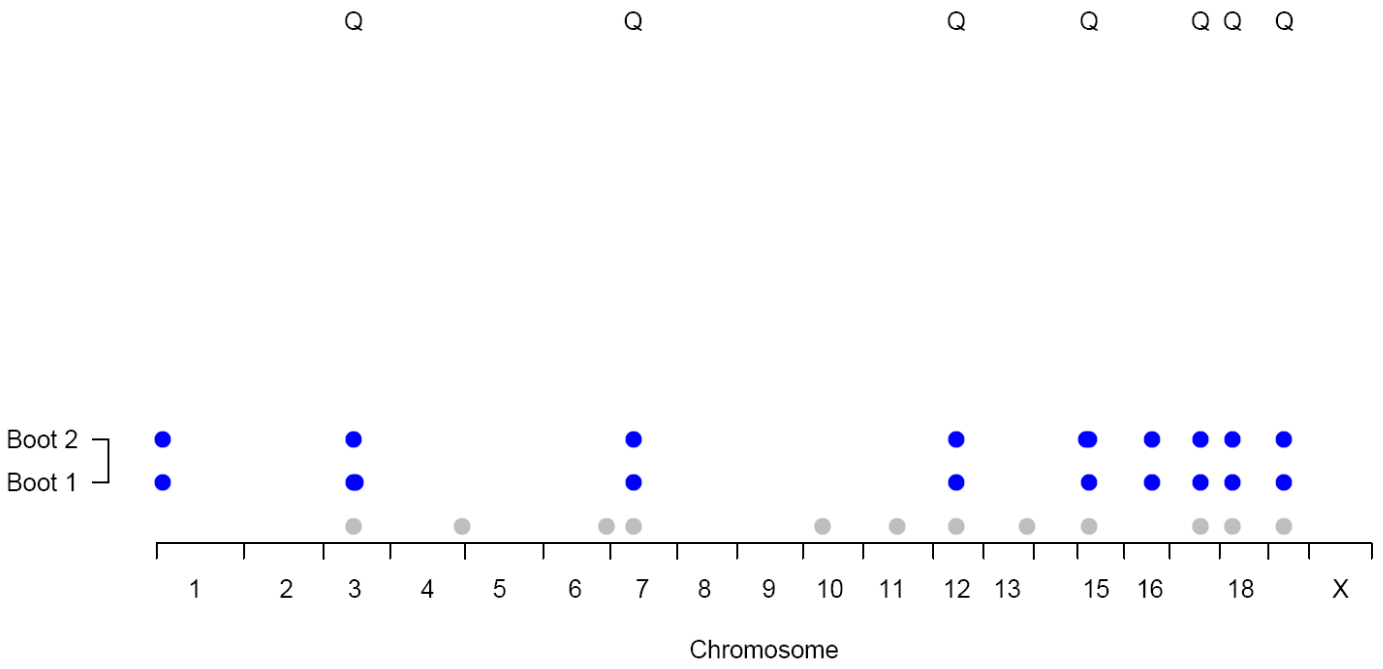



Q
Q
Q
Q
Q
Q
Q
1
2
3
4
5
6
7
8
9
10
11
12
13
15
16
18
X
Chromosome

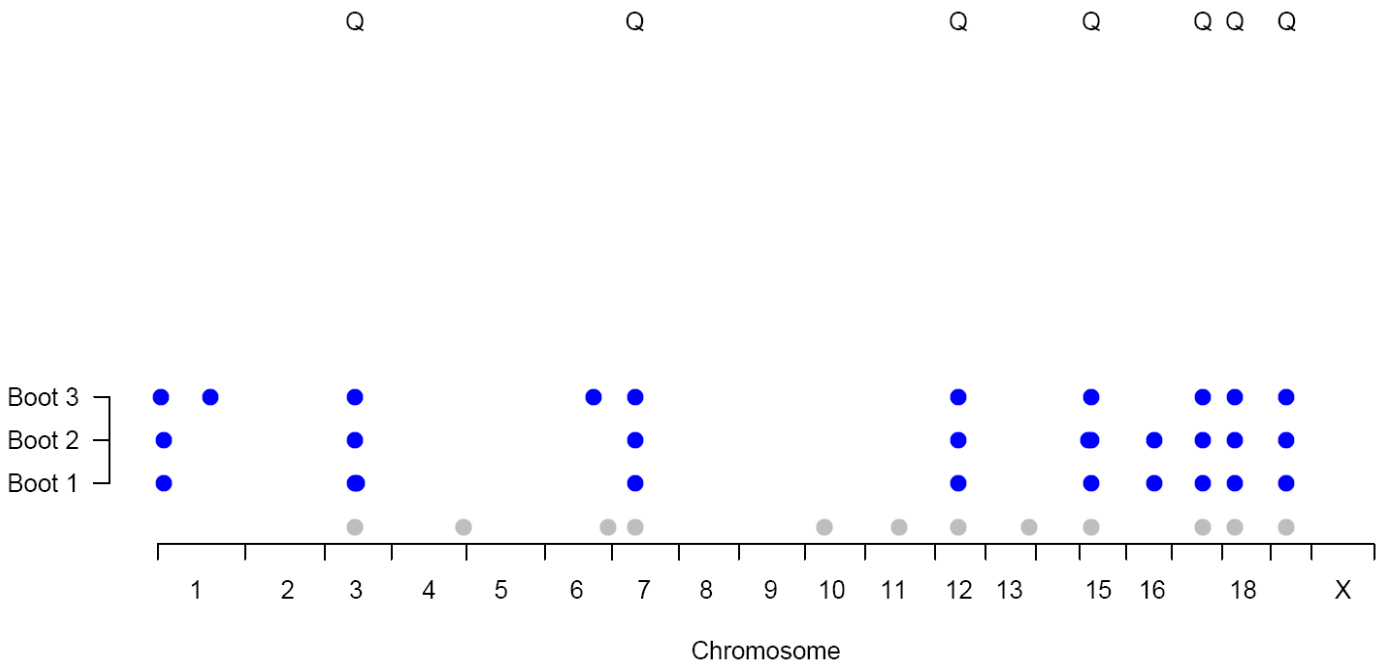
Forward selection on a bootstrap sample



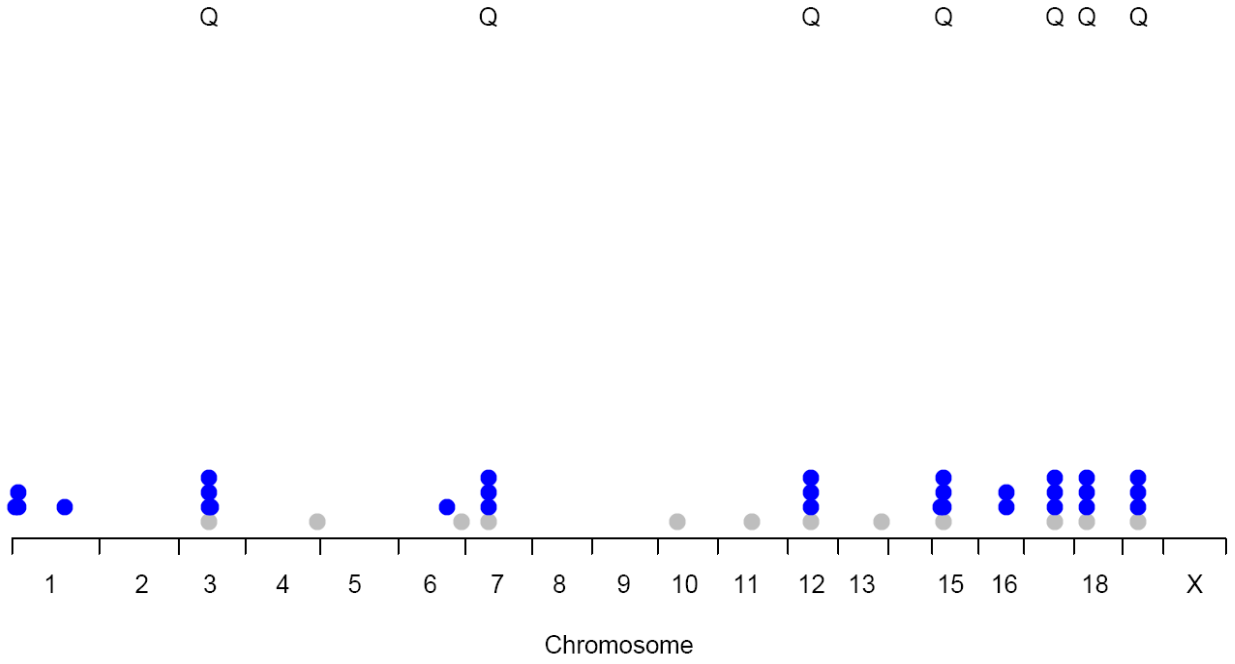
Forward selection on a bootstrap sample



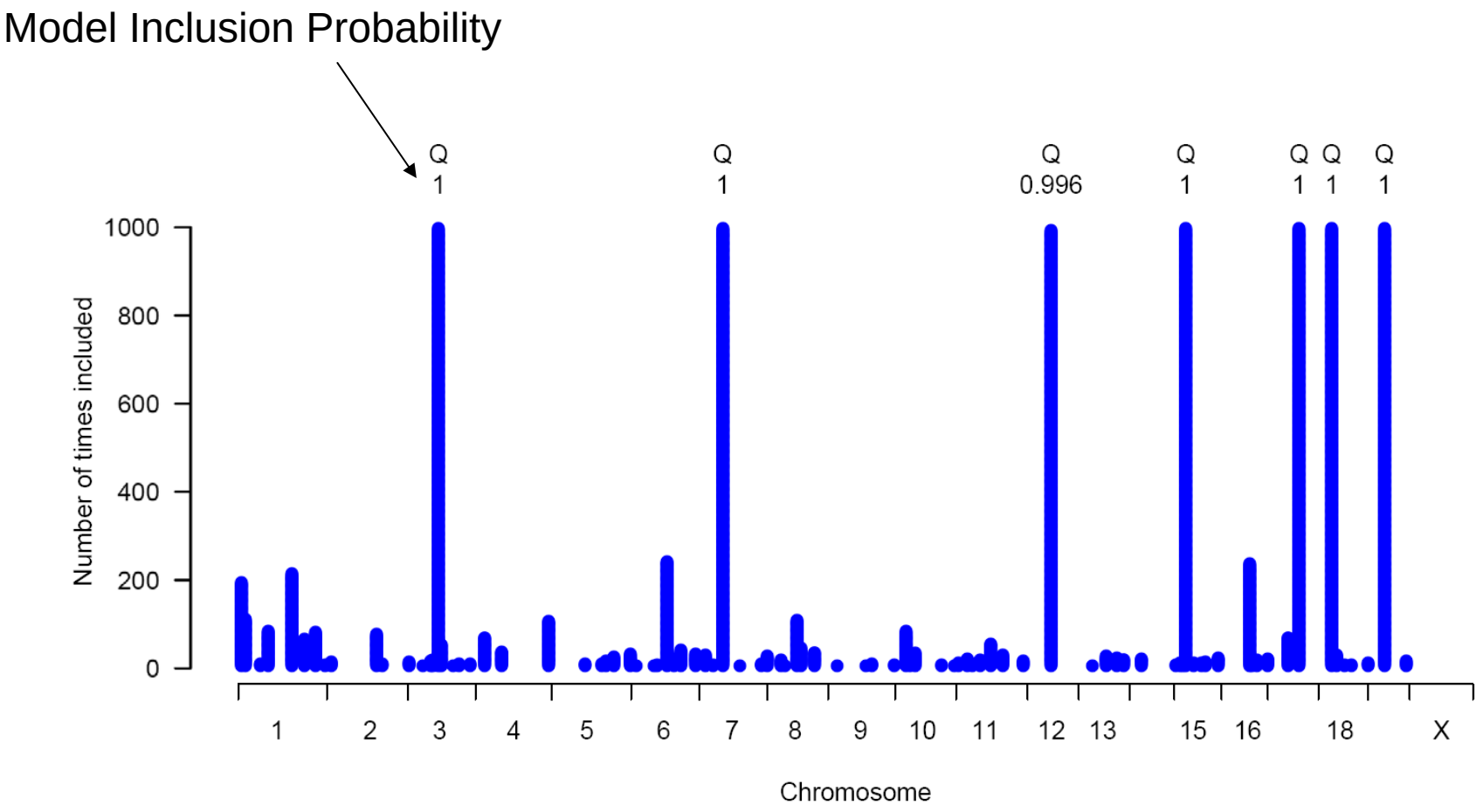
Forward selection on a bootstrap sample

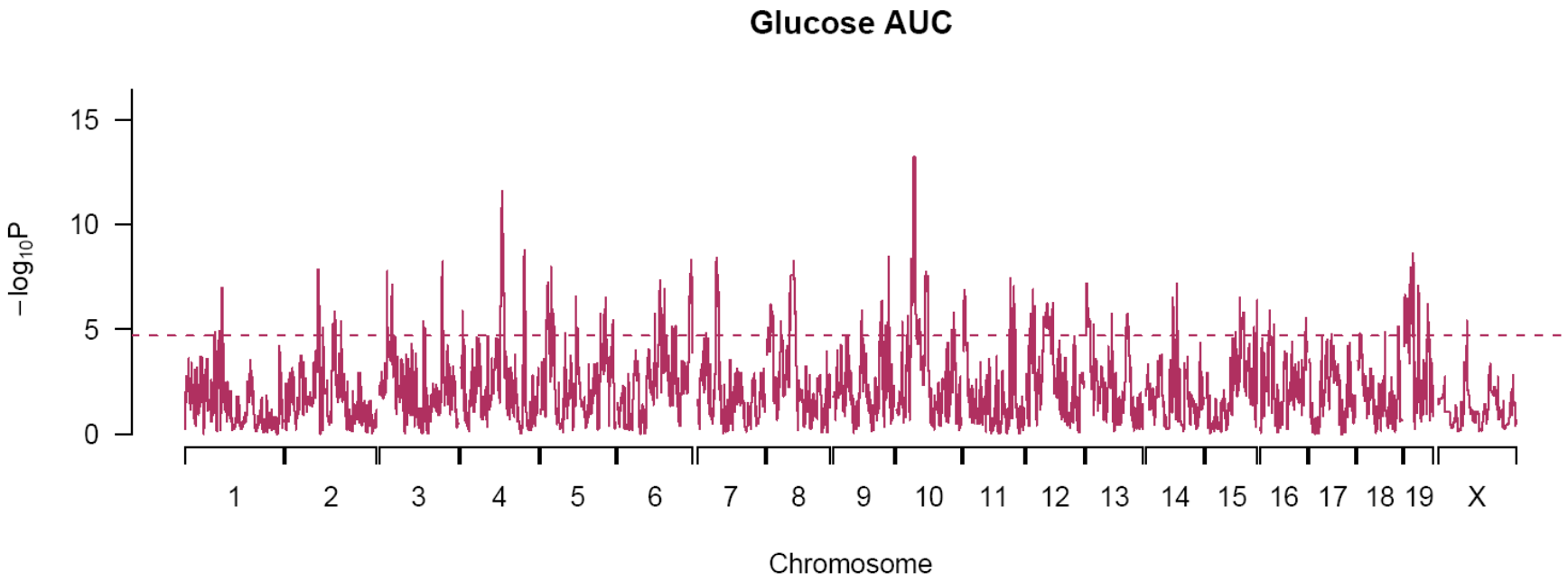


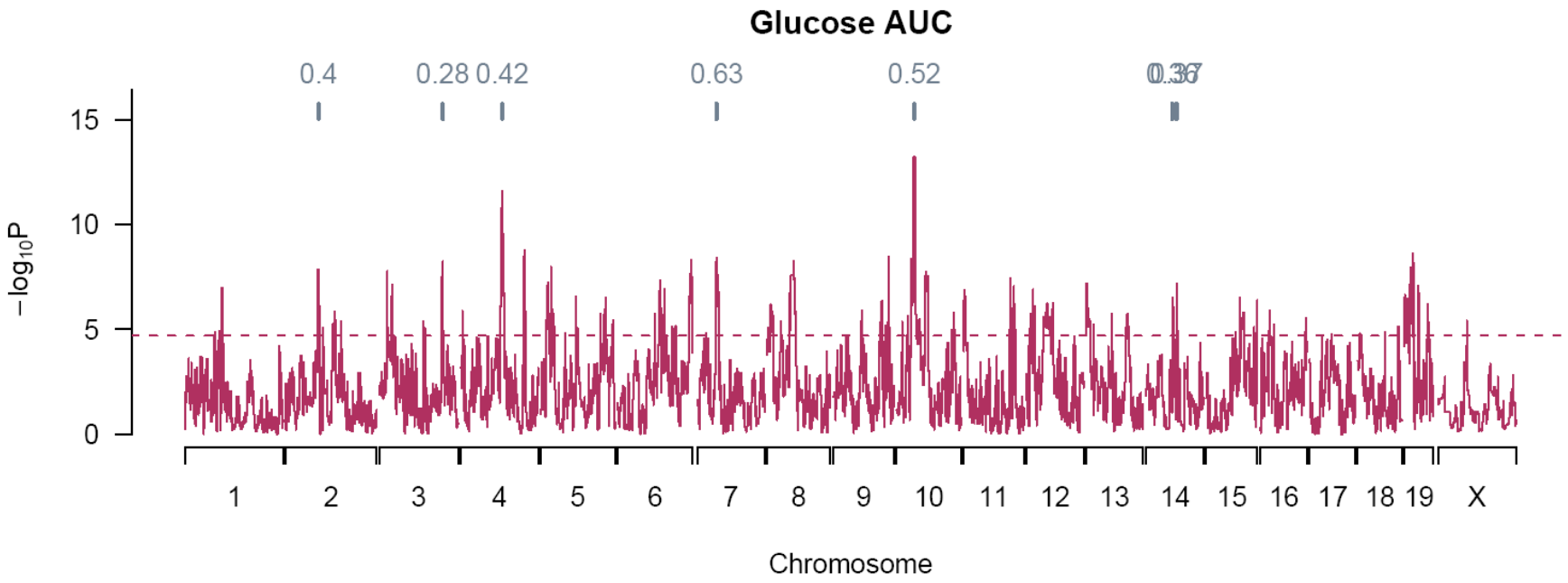
Bootstrap evidence mounts up…

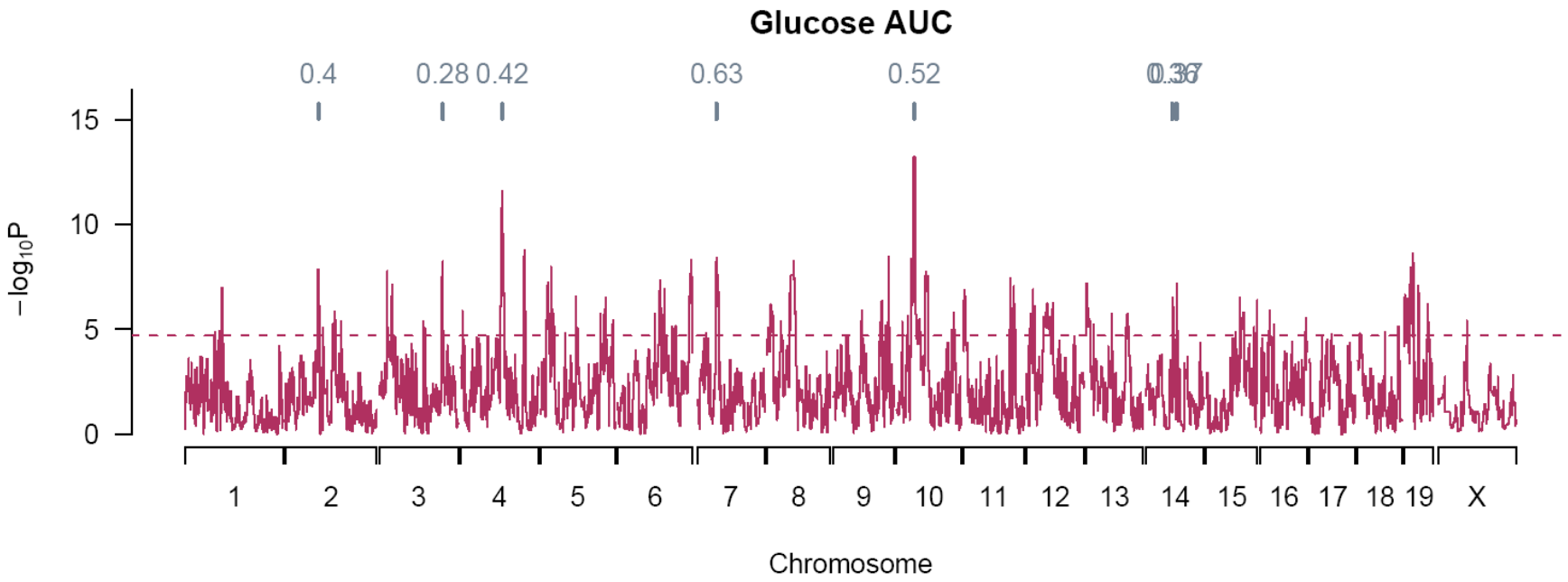


In 1000 bootstraps…

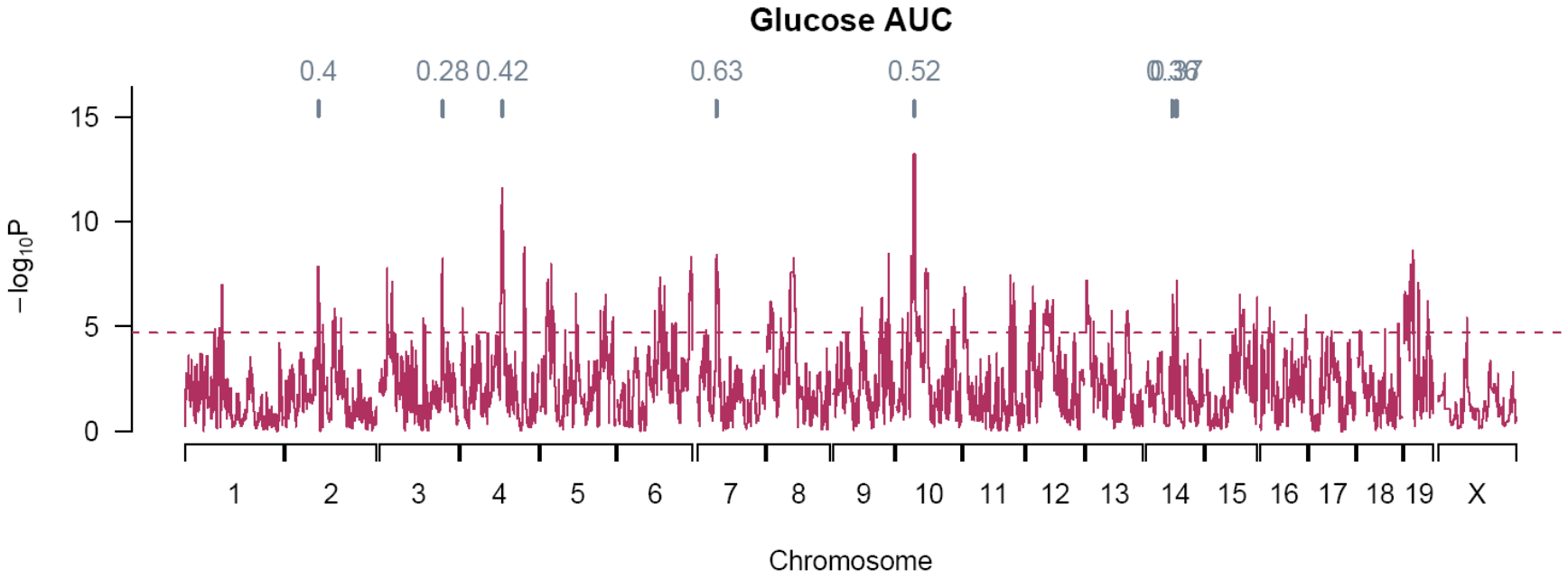



Glucose AUC
15
10
5
0
−log10P
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome


Glucose AUC
0.4
0.28
0.42
0.63
0.52
15
10
5
0
$-\log_{10}P$
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome



854 loci in all phenotypes, 84 diabetes loci



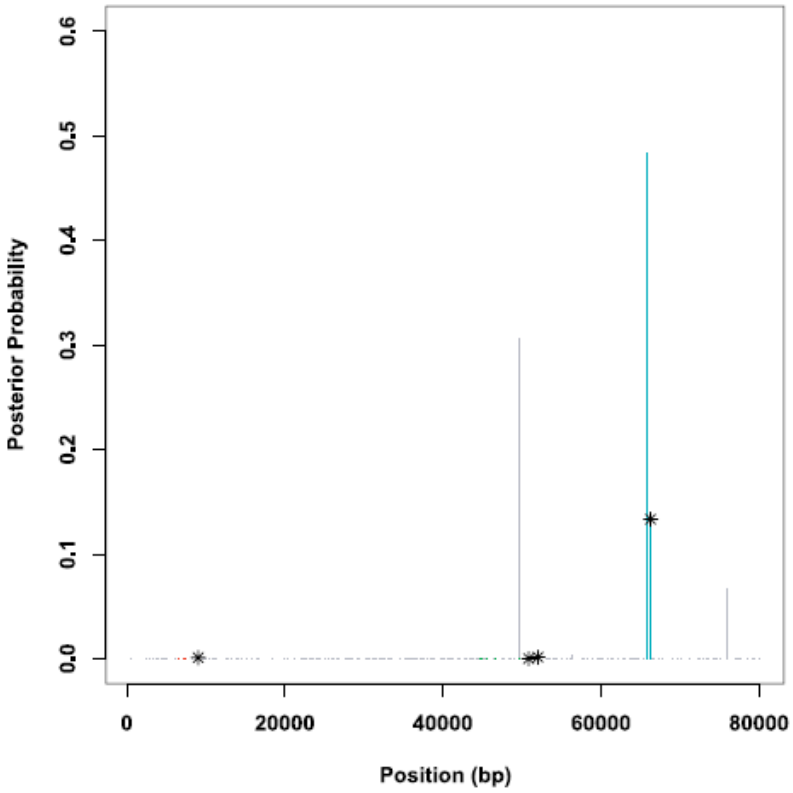
854 loci in all phenotypes, 84 diabetes loci

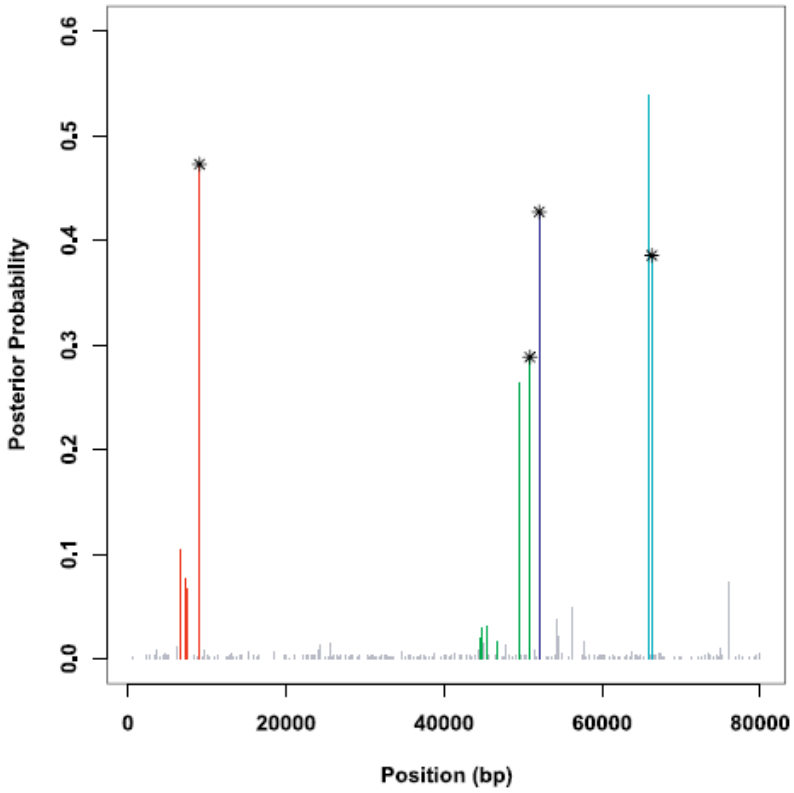
ARTICLES

nature genetics

Genome-wide genetic association of complex traits in heterogeneous stock mice

William Valdar[1], Leah C Solberg[1,4], Dominique Gauguier[1], Stephanie Burnett[1], Paul Klenerman[2], William O Cookson[1], Martin S Taylor[1], J Nicholas P Rawlins[3], Richard Mott[1] & Jonathan Flint[1]





Servin B, Stephens M (2007)

Bayesian Multiple QTL modelling

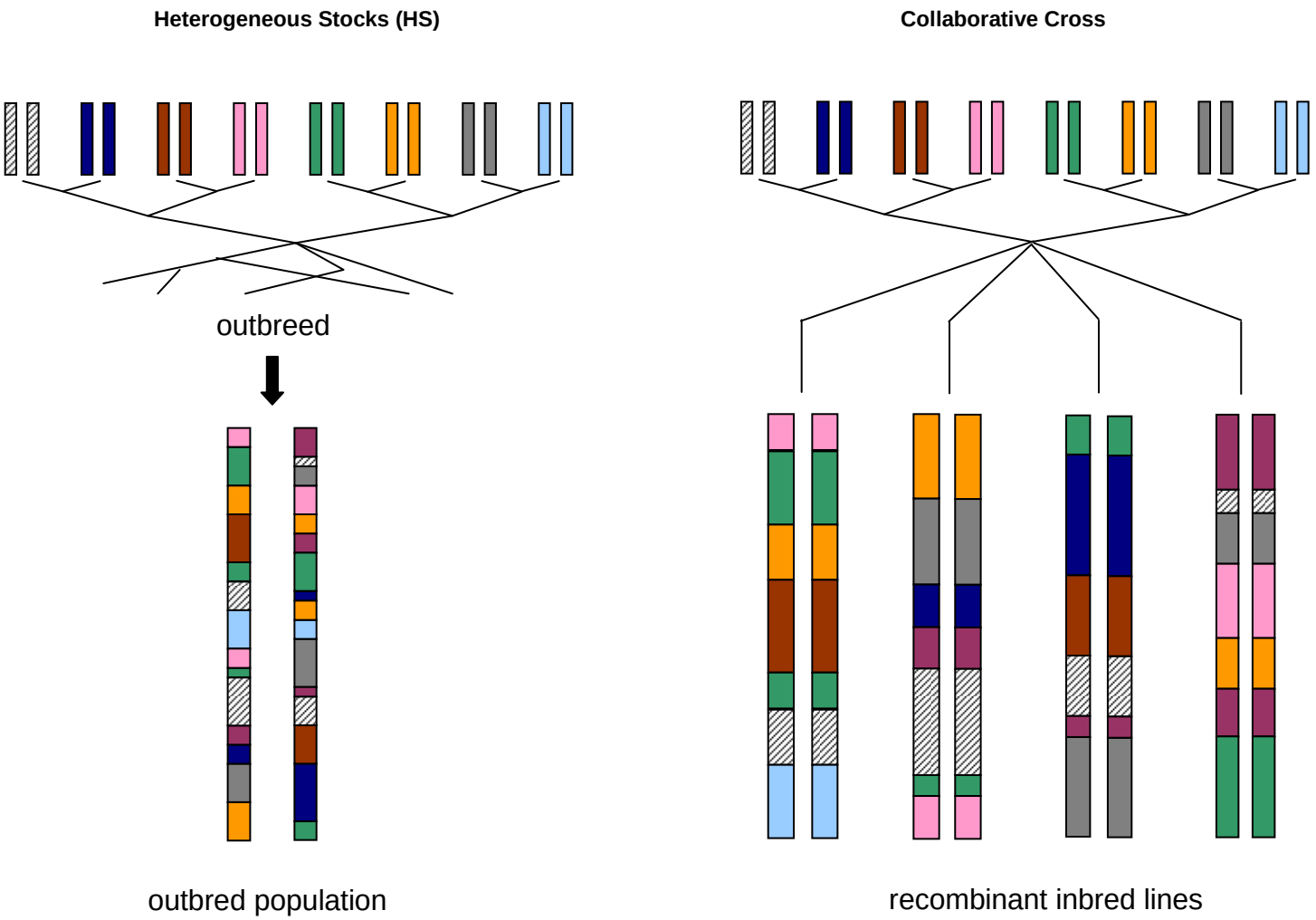
Kilpikari R, Sillanpaa MJ (2003) Bayesian analysis of multilocus association in quantitative and qualitative traits. Genet Epidemiol 25: 122-135

Yi N (2004) A unified Markov chain Monte Carlo framework for mapping multiple quantitative trait loci. Genetics 167: 967-975

Servin B, Stephens M (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. PLoS Genet 3: e114

Fridley BL (2008) Bayesian variable and model selection methods for genetic association studies. Genet Epidemiol.

The Collaborative Cross



Churchill et al 2004; Broman 2005; Valdar et al 2006