Identifying sample mix-ups in eQTL data

Karl W Broman

Department of Biostatistics & Medical Informatics
University of Wisconsin – Madison

www.biostat.wisc.edu/~kbroman

Attie project

$\sim$500 B6 $\times$ BTBR intercross mice, all ob/ob

Genotypes at 2057 SNPs (Affymetrix arrays)

Gene expression in six tissues (Agilent arrays)

- adipose
- gastrocnemius muscle
- hypothalamus
- pancreatic islets
- kidney
- liver

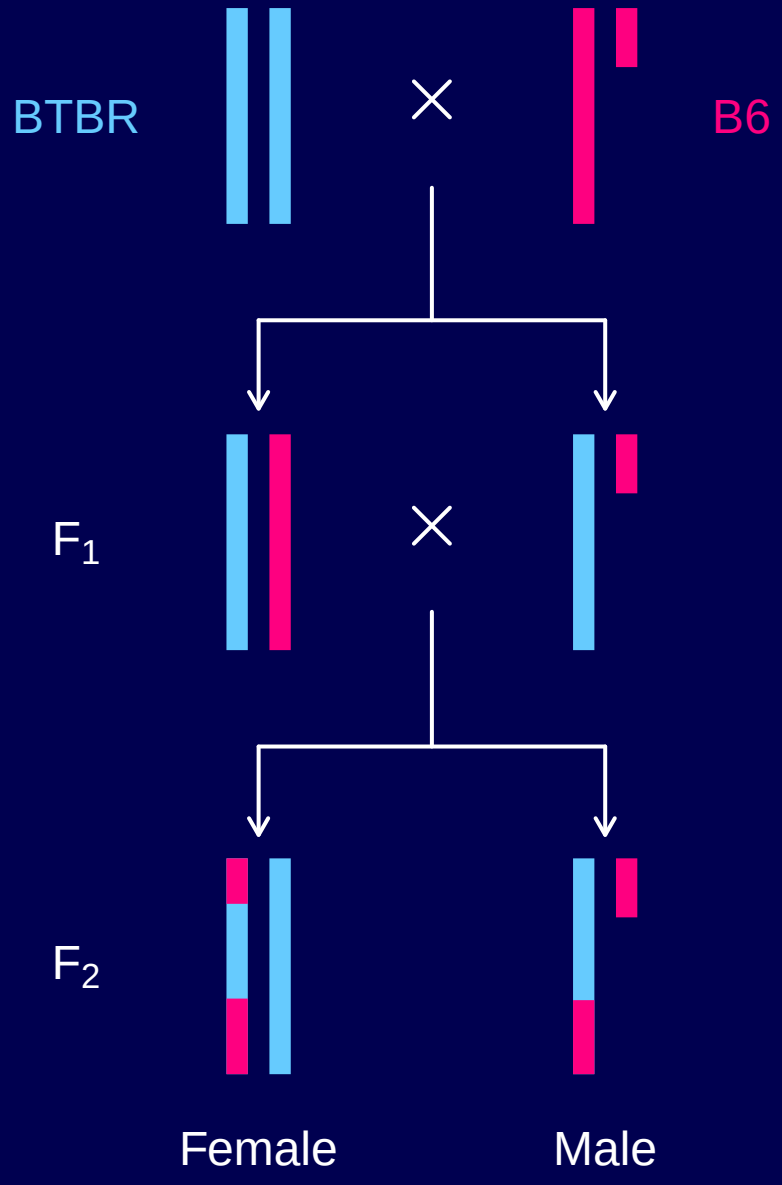
Numerous clinical phenotypes

(e.g., body weight, insulin and glucose levels)

Considerable brain power

Alan Attie, Karl Broman, Aimee Teo Broman, Danielle Greenawalt, Mark Keller, Christina Kendziorski, Amit Kulkarni, Eric Schadt, Brian Yandell, . . .

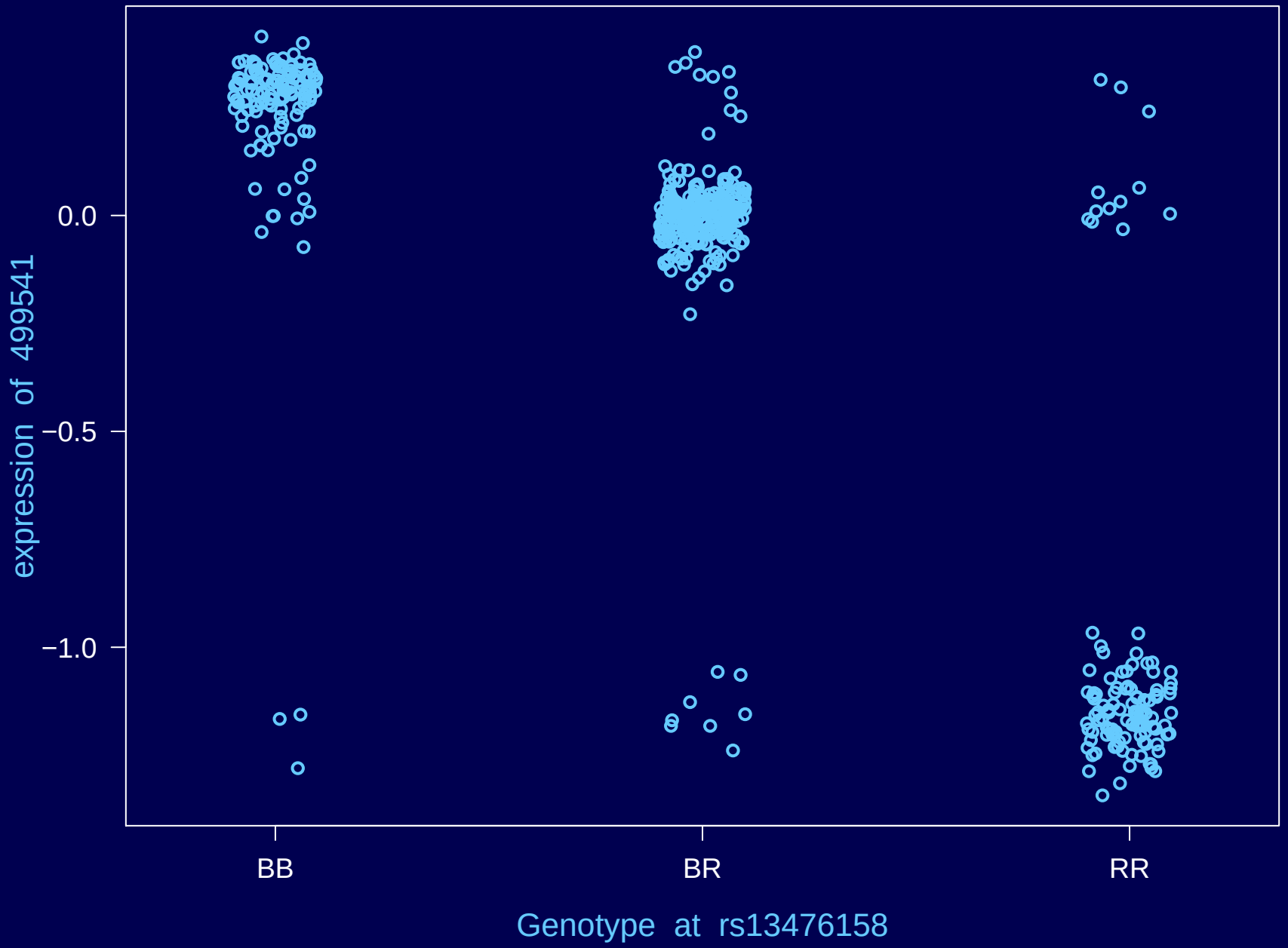
Sex and the X chr



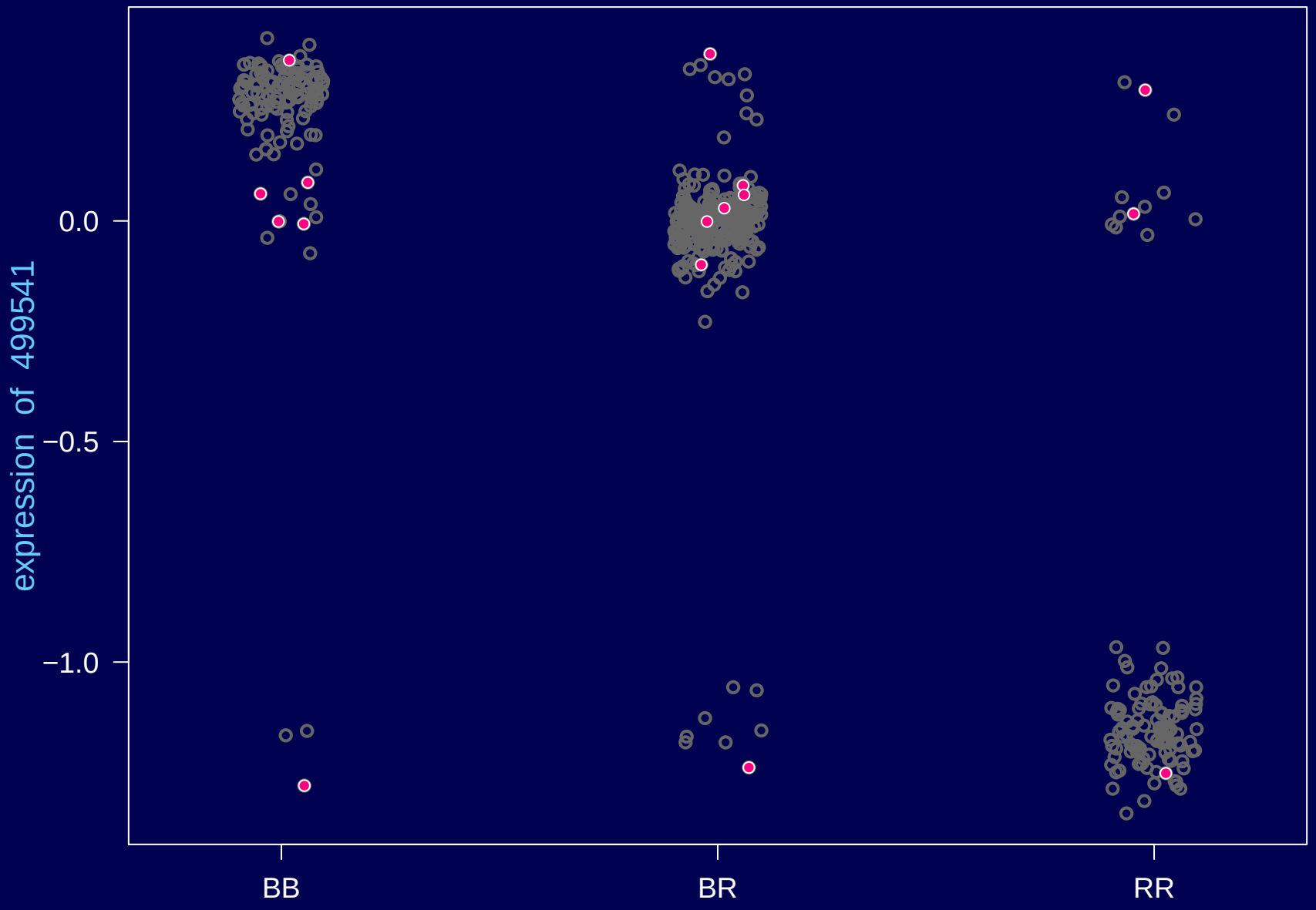
F_2 females: R/R or B/R

F_2 males: hemizygous B or R

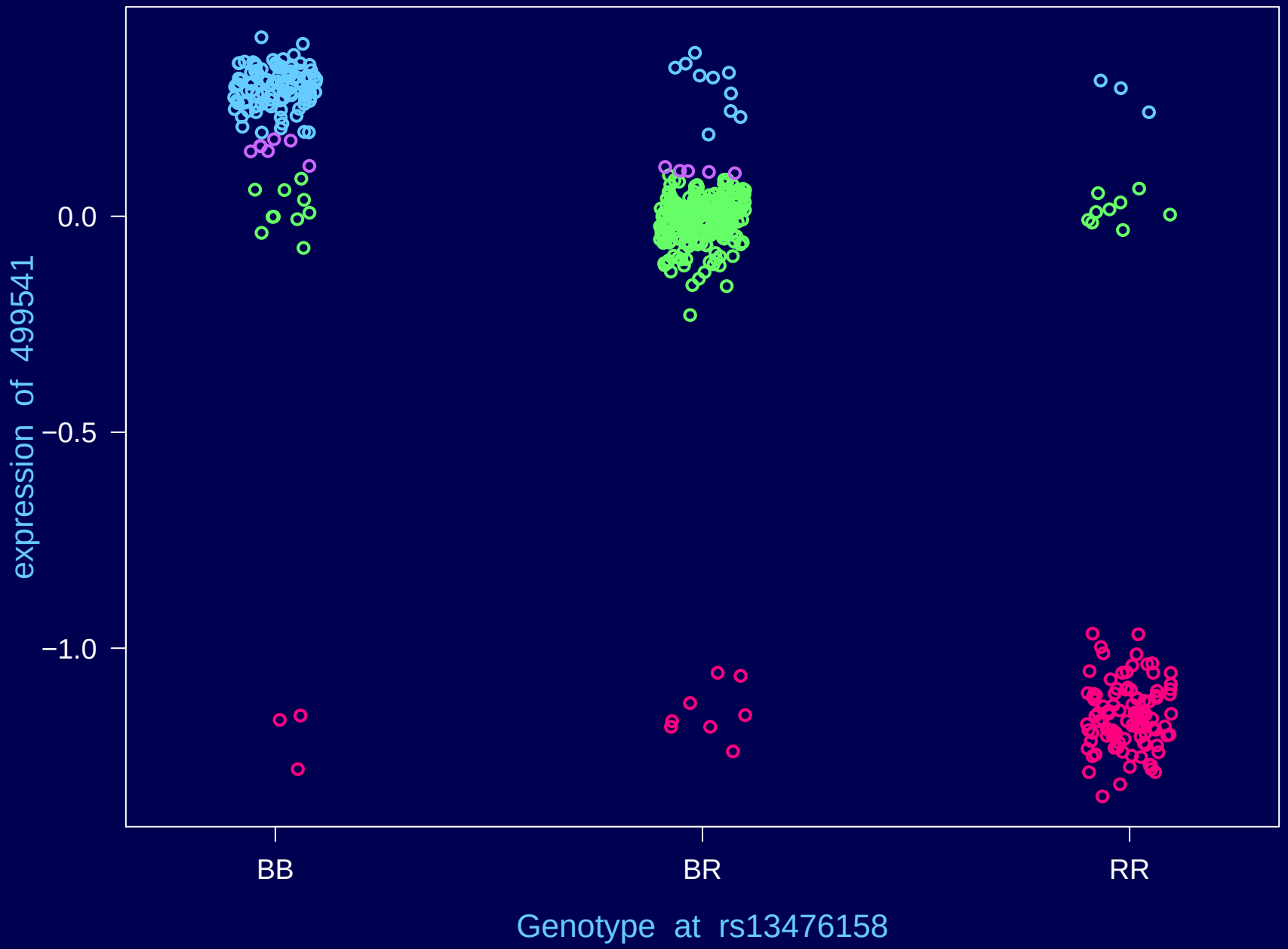
E vs G


0.0
−0.5
−1.0
expression of 499541
BB
BR
RR
Genotype at rs13476158

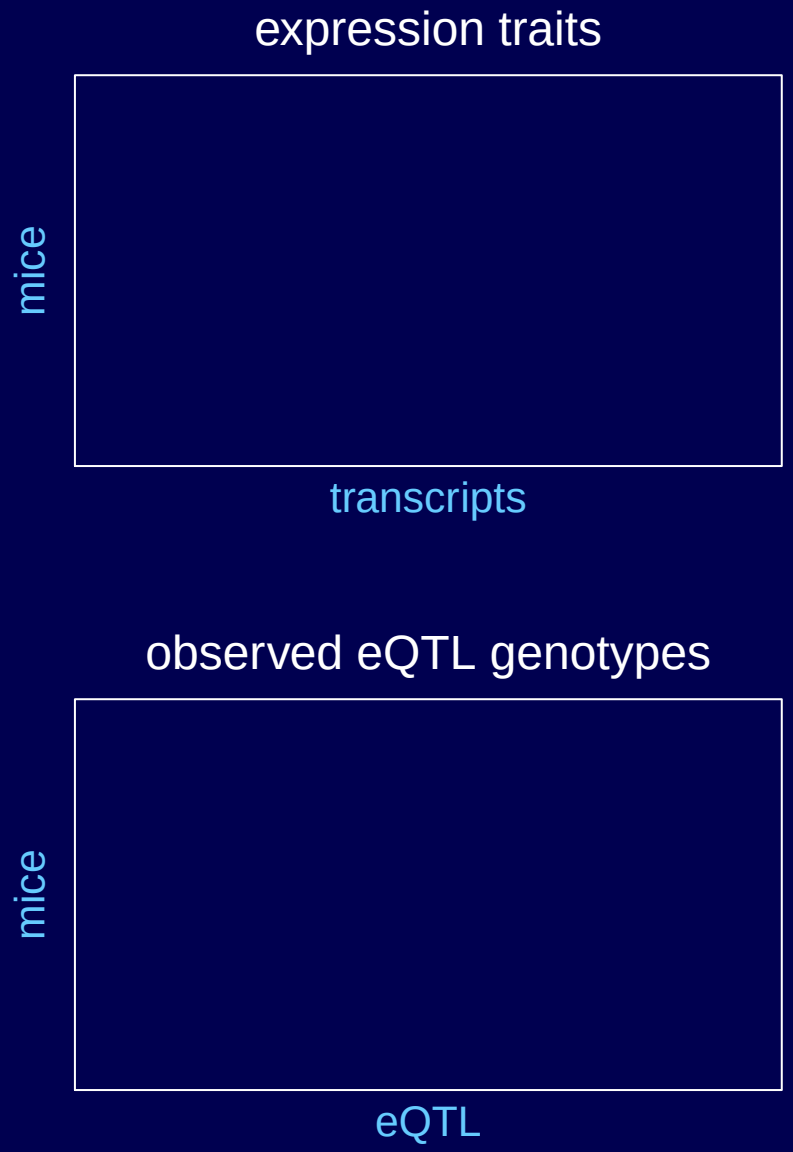
E vs G


0.0
−0.5
−1.0
expression of 499541
BB
BR
RR
Genotype at rs13476158

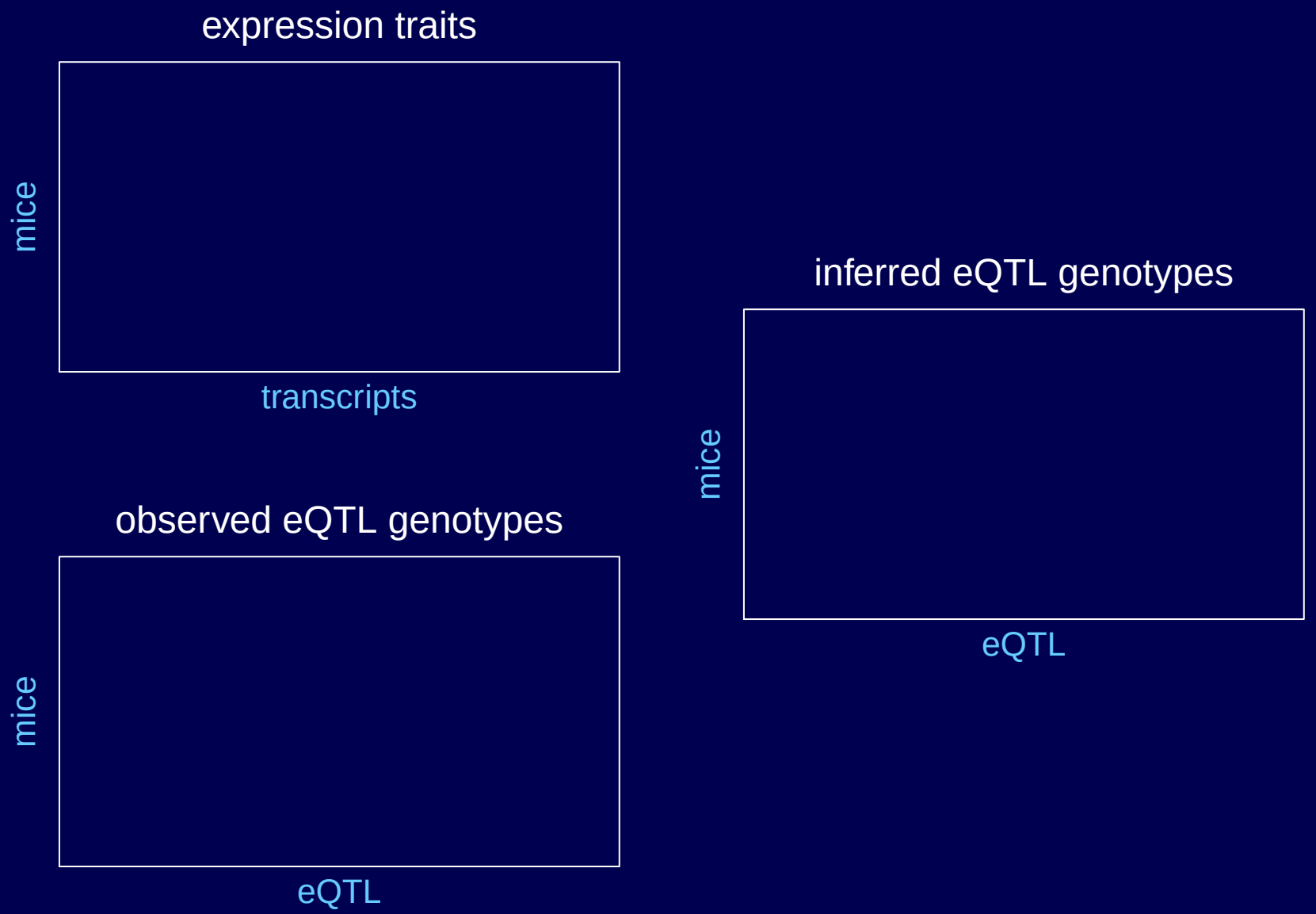
kNN classifier



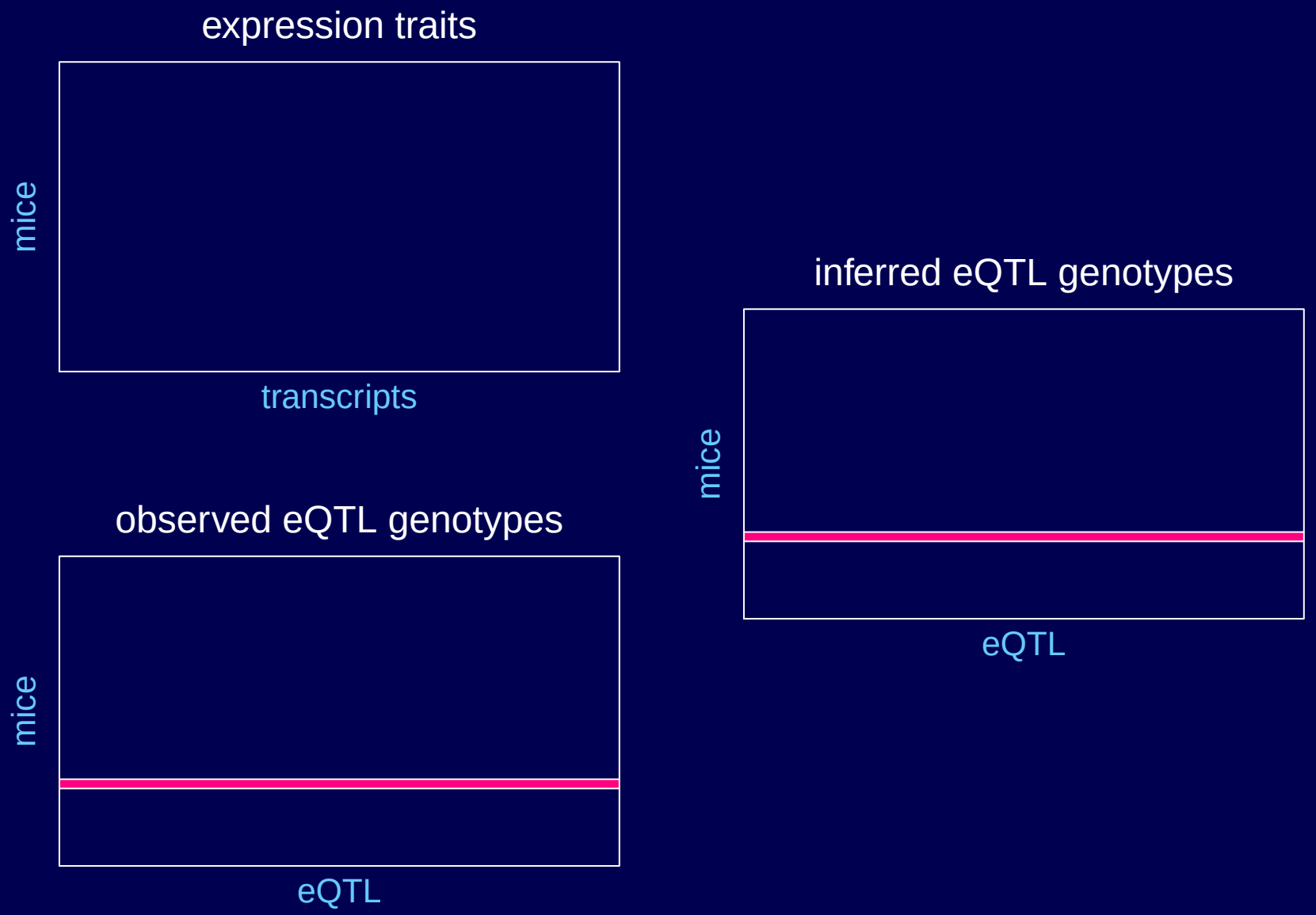
Basic scheme



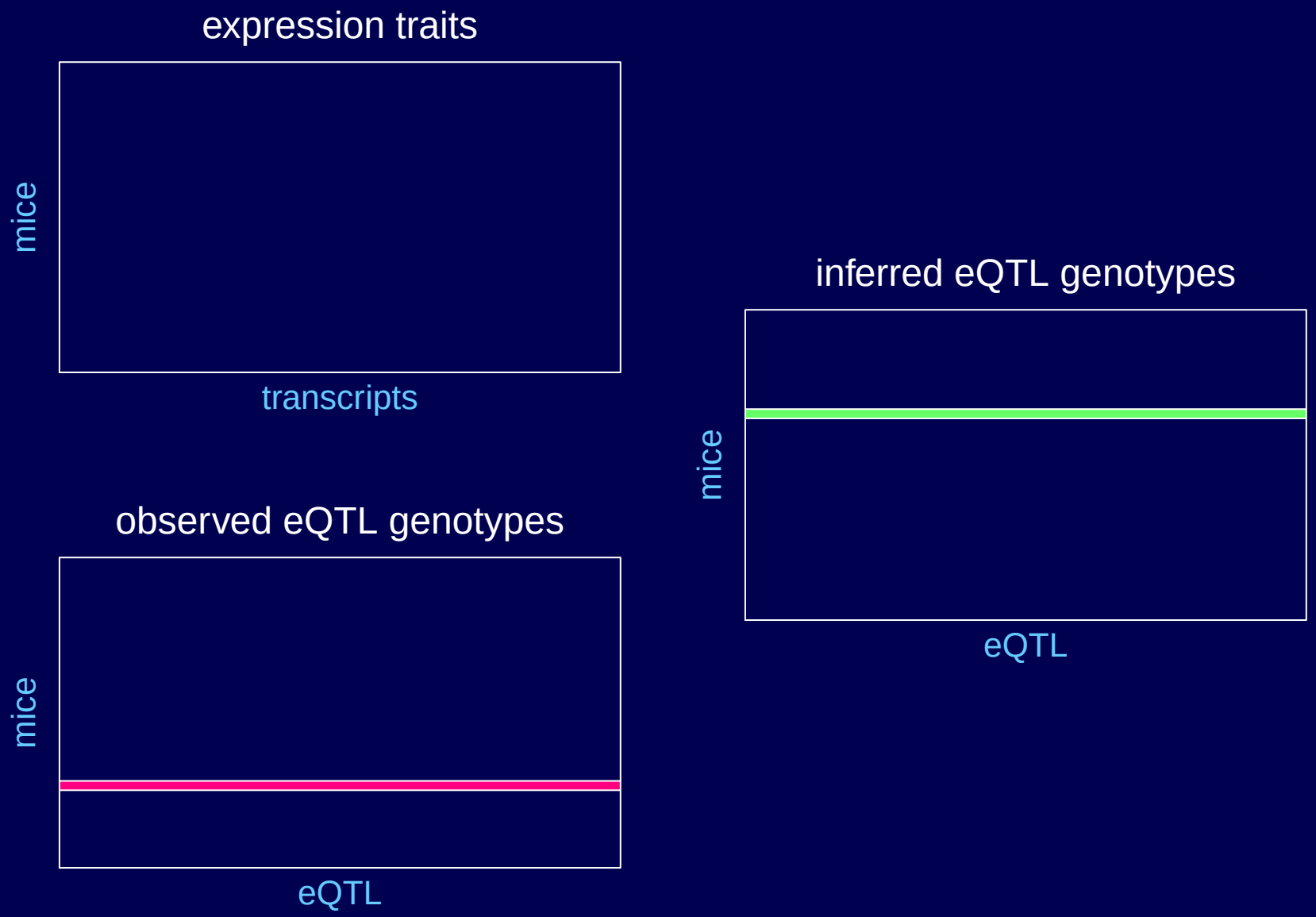
Basic scheme



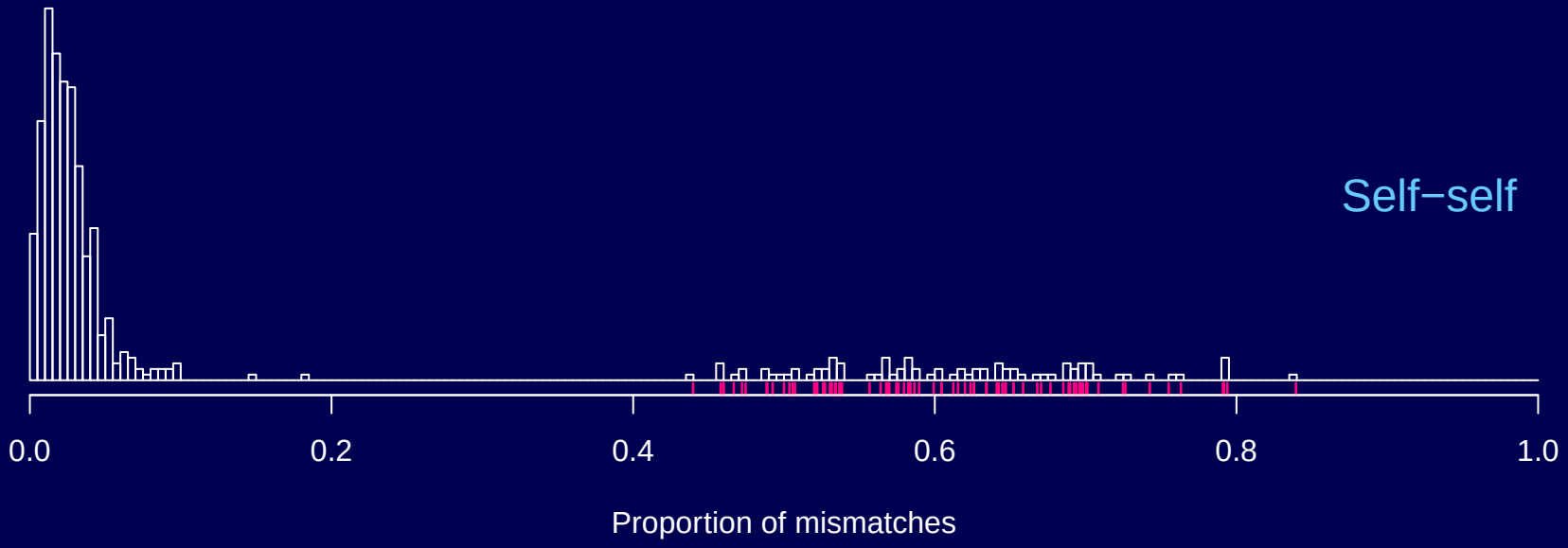
Basic scheme

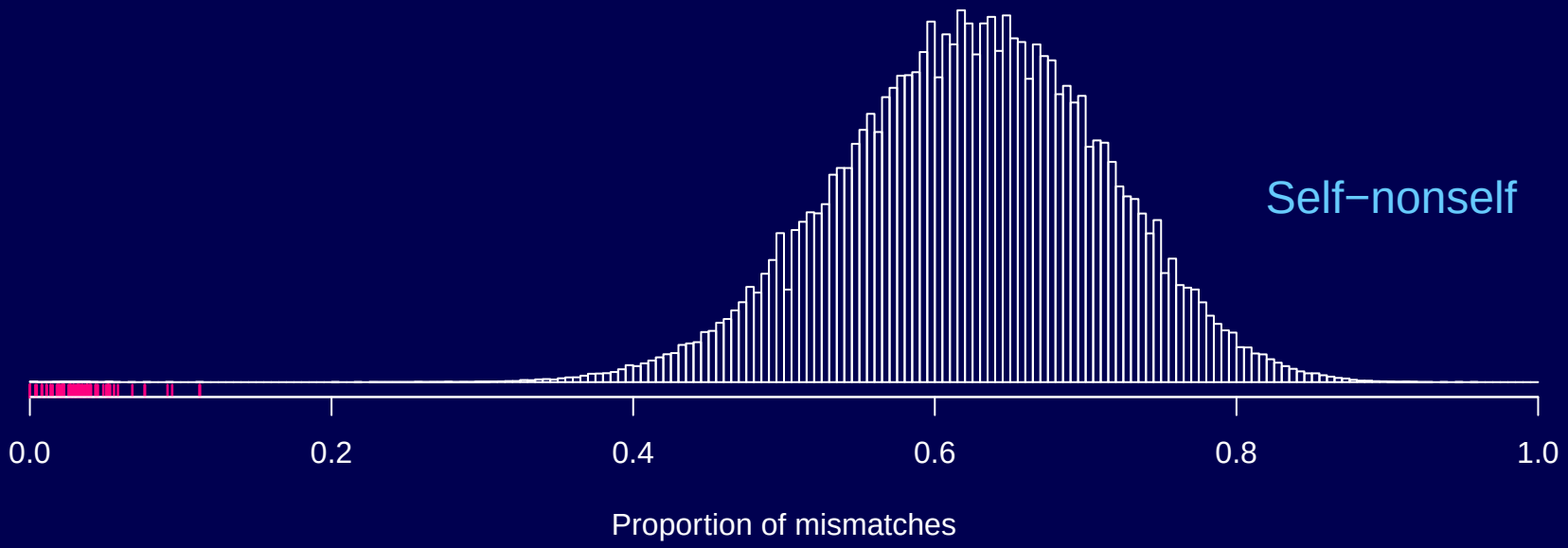


Basic scheme



Prop'n mismatches



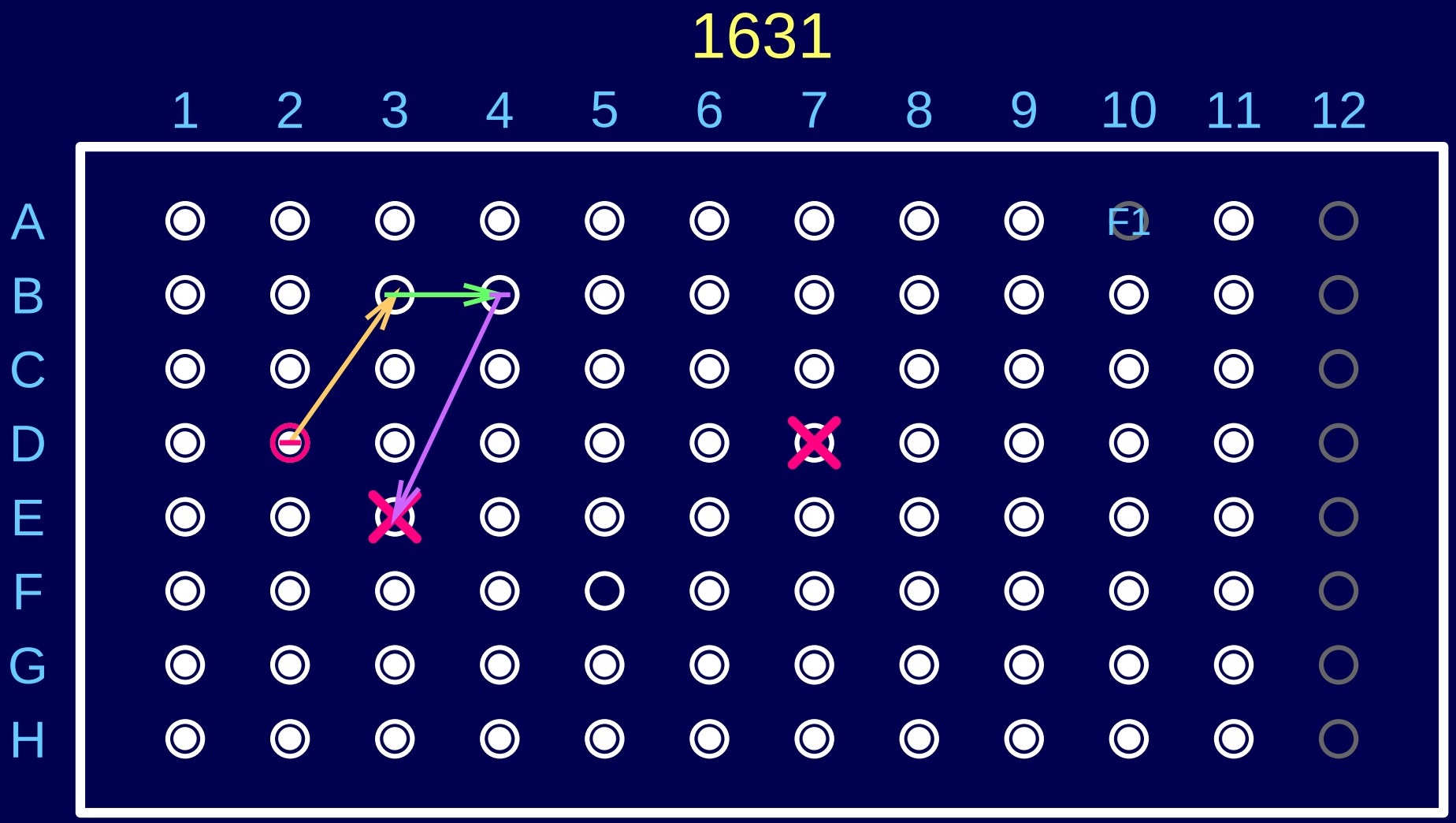


Decisions


Self vs best
Fixable
Not found
Good
self−self distance
minimum distance
0.0
0.2
0.4
0.6
0.8

Genotype mix-ups

Plate 1631



Plates 1632 and 1630

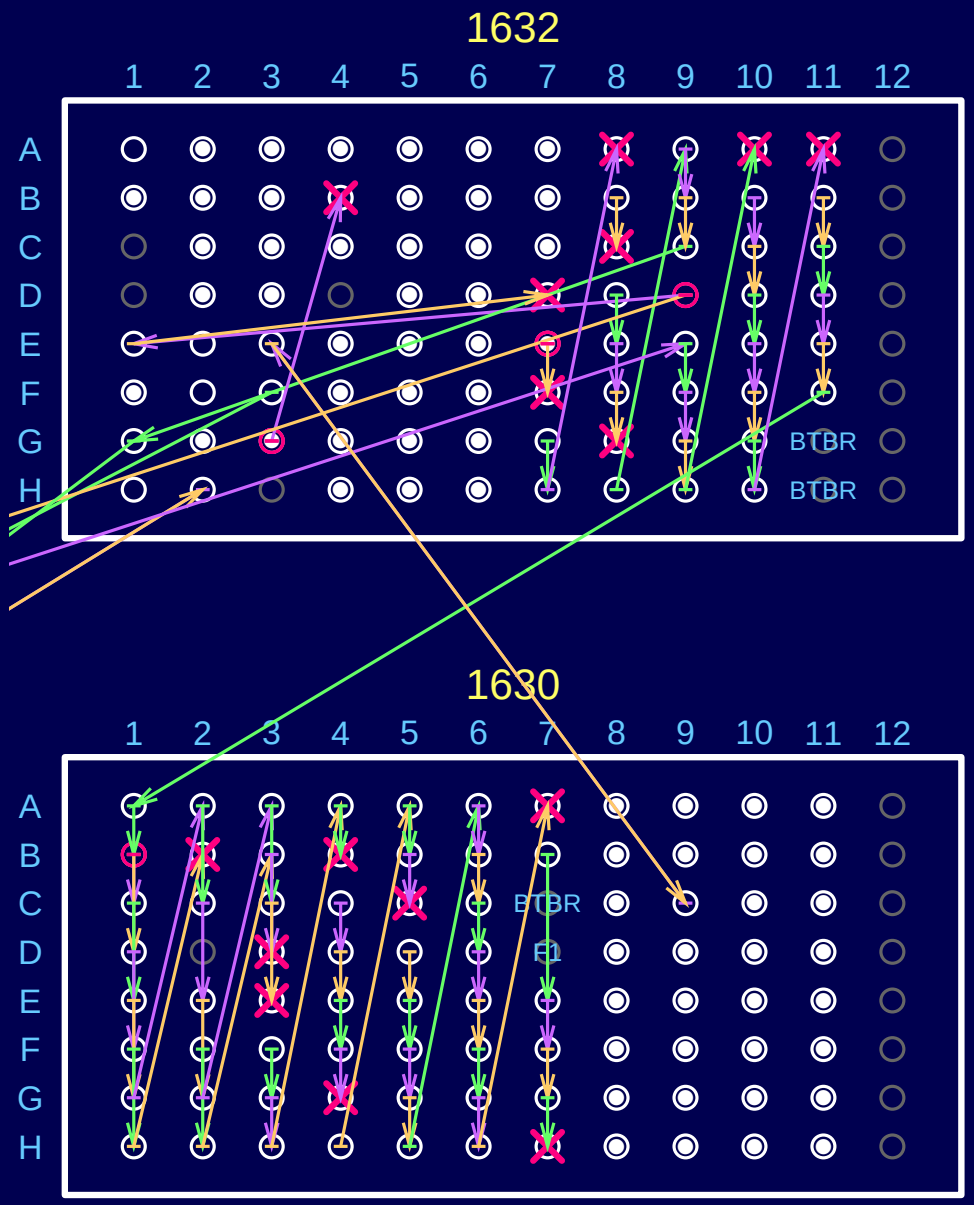
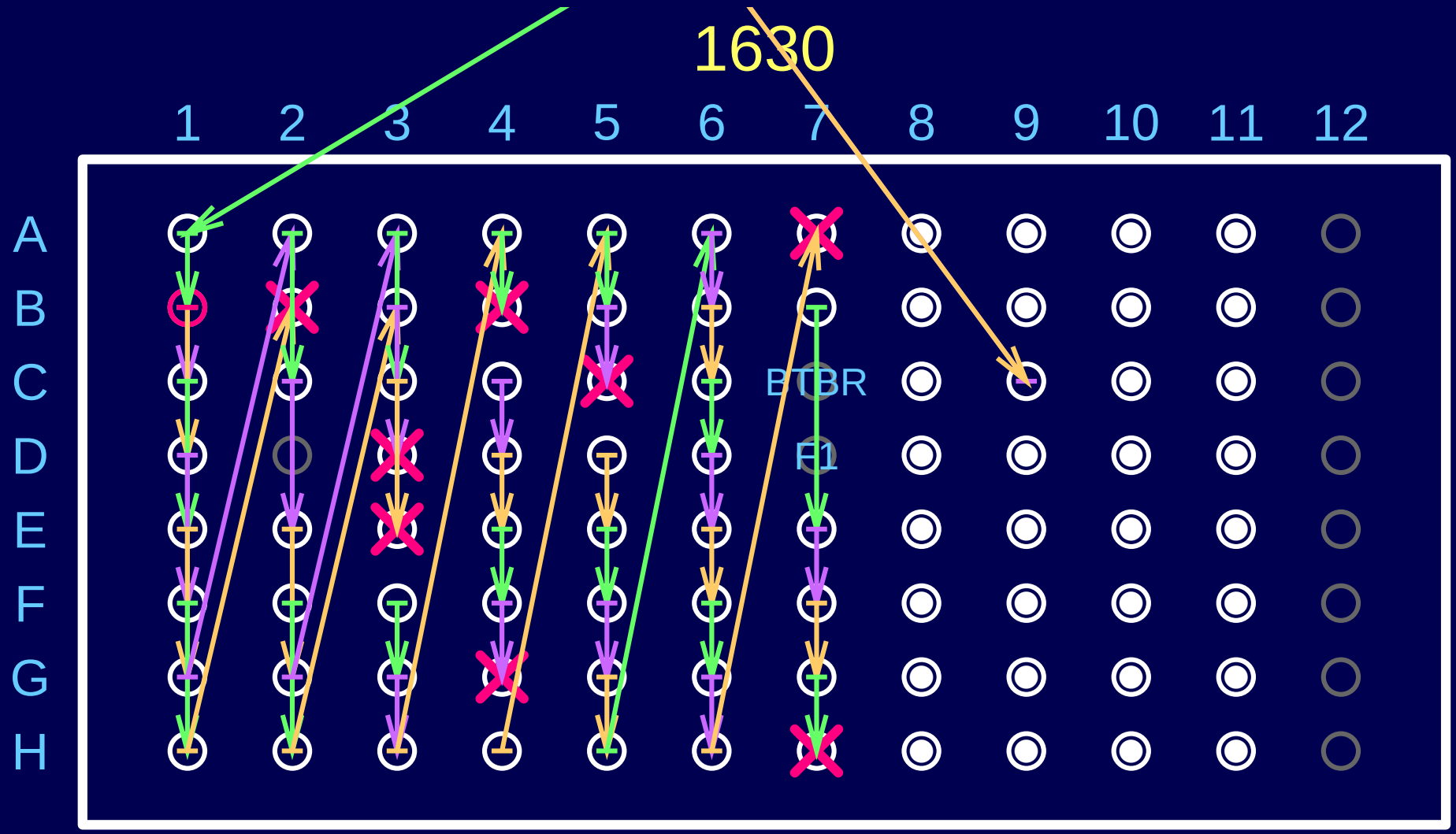
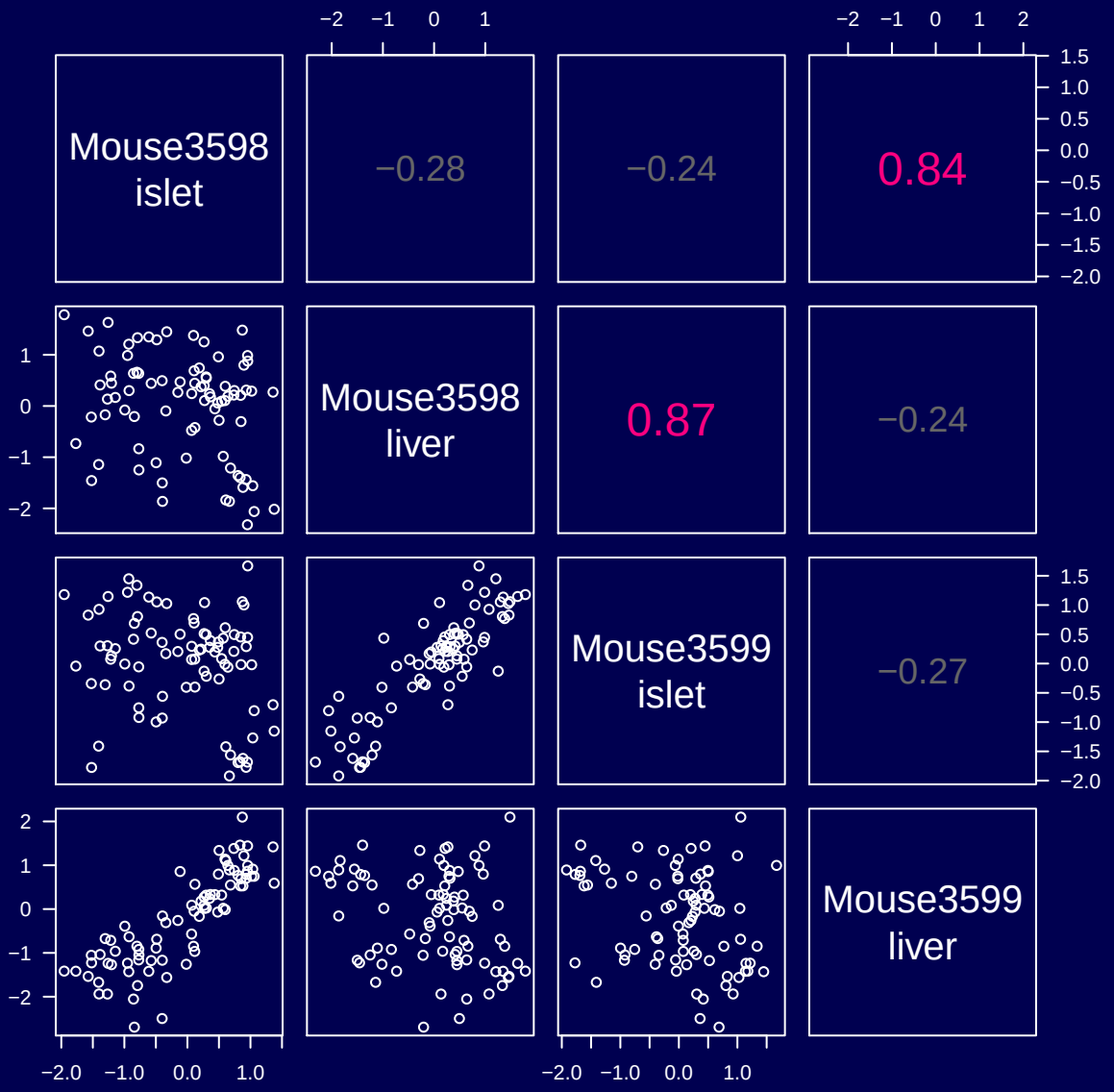


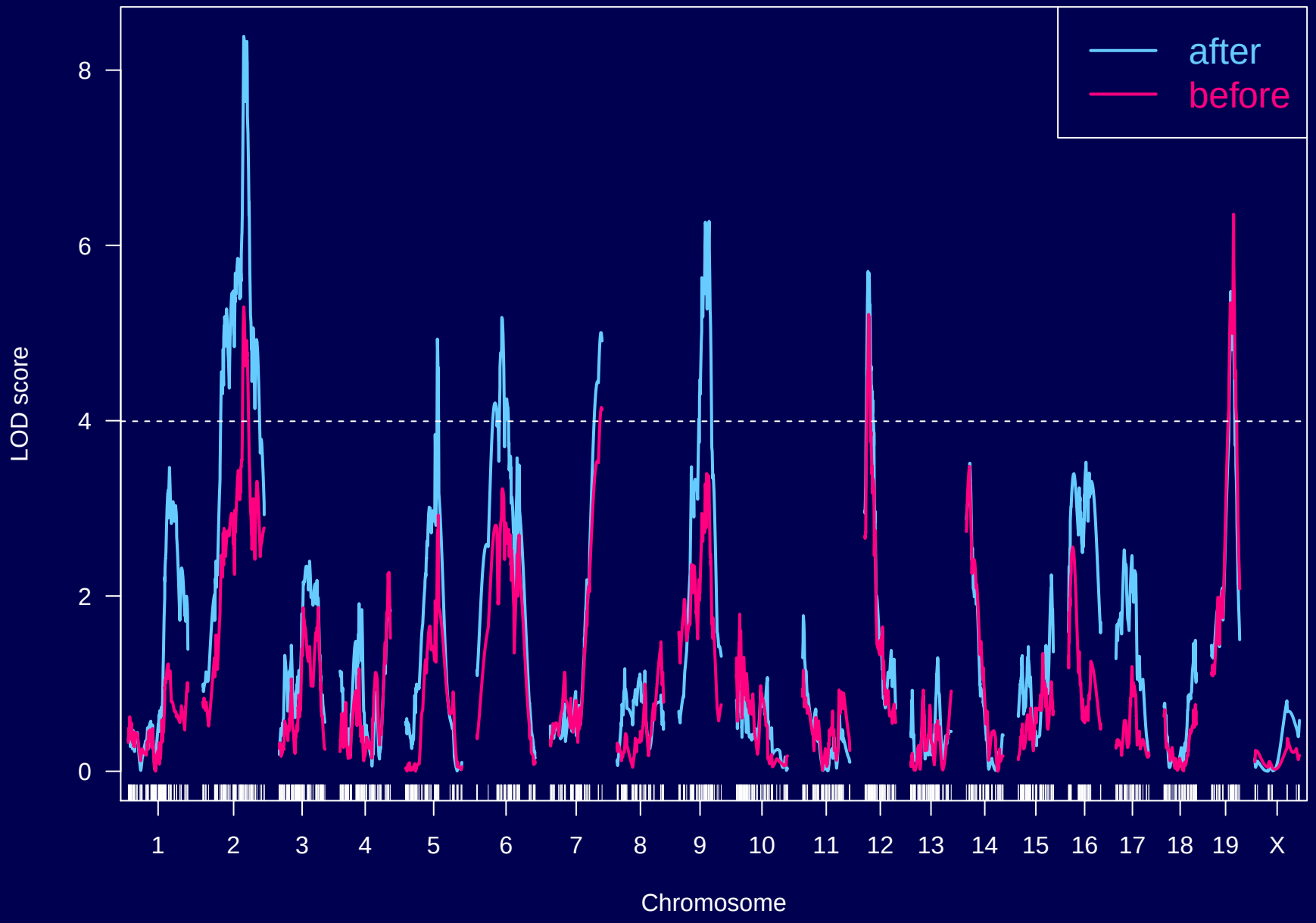
Plate 1630



E vs E


Mouse3598 islet
−0.28
−0.24
0.84
Mouse3598 liver
0.87
−0.24
Mouse3599 islet
−0.27
Mouse3599 liver

Insulin QTL


after
before
LOD score
0
2
4
6
8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome

Summary

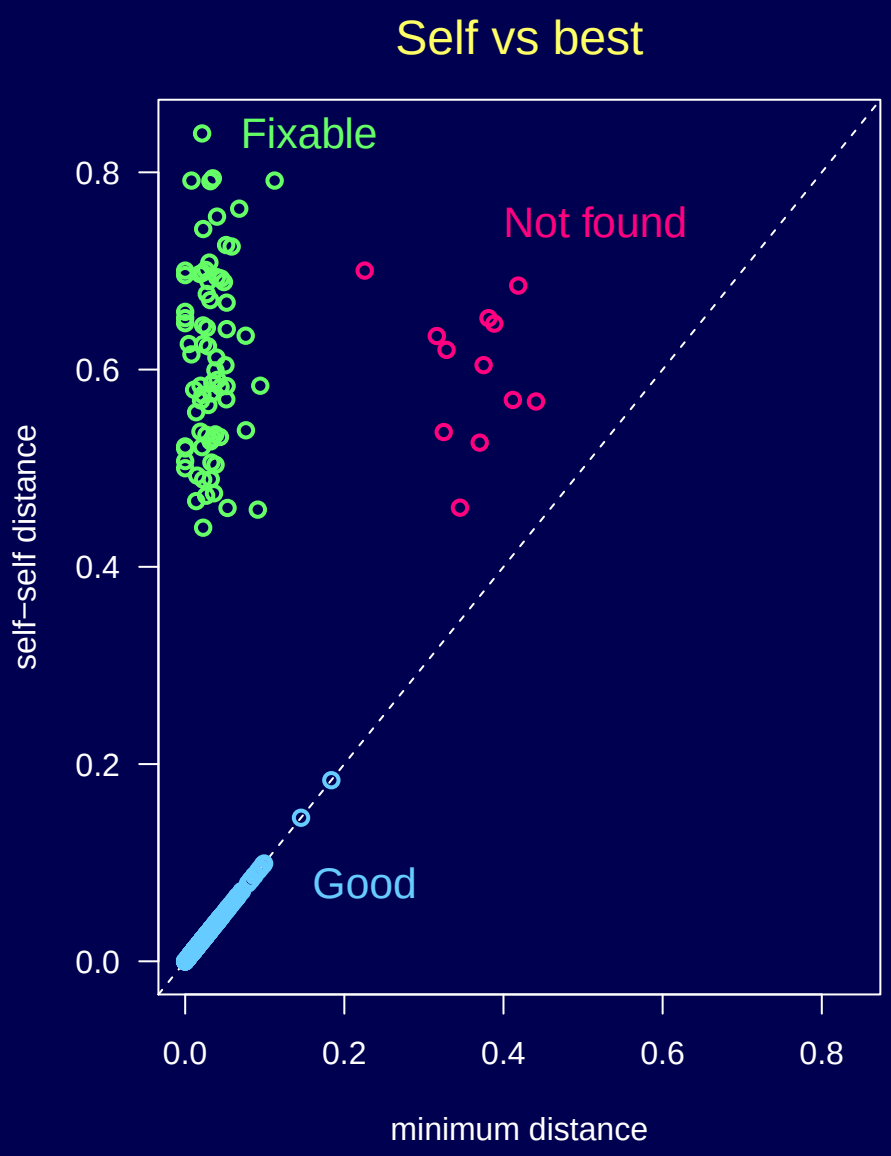
- Sample mix-ups happen
- With eQTL data, we can both identify and correct mix-ups
- There is great value in having expression on multiple tissues
- The general idea here has wide application for high-throughput data

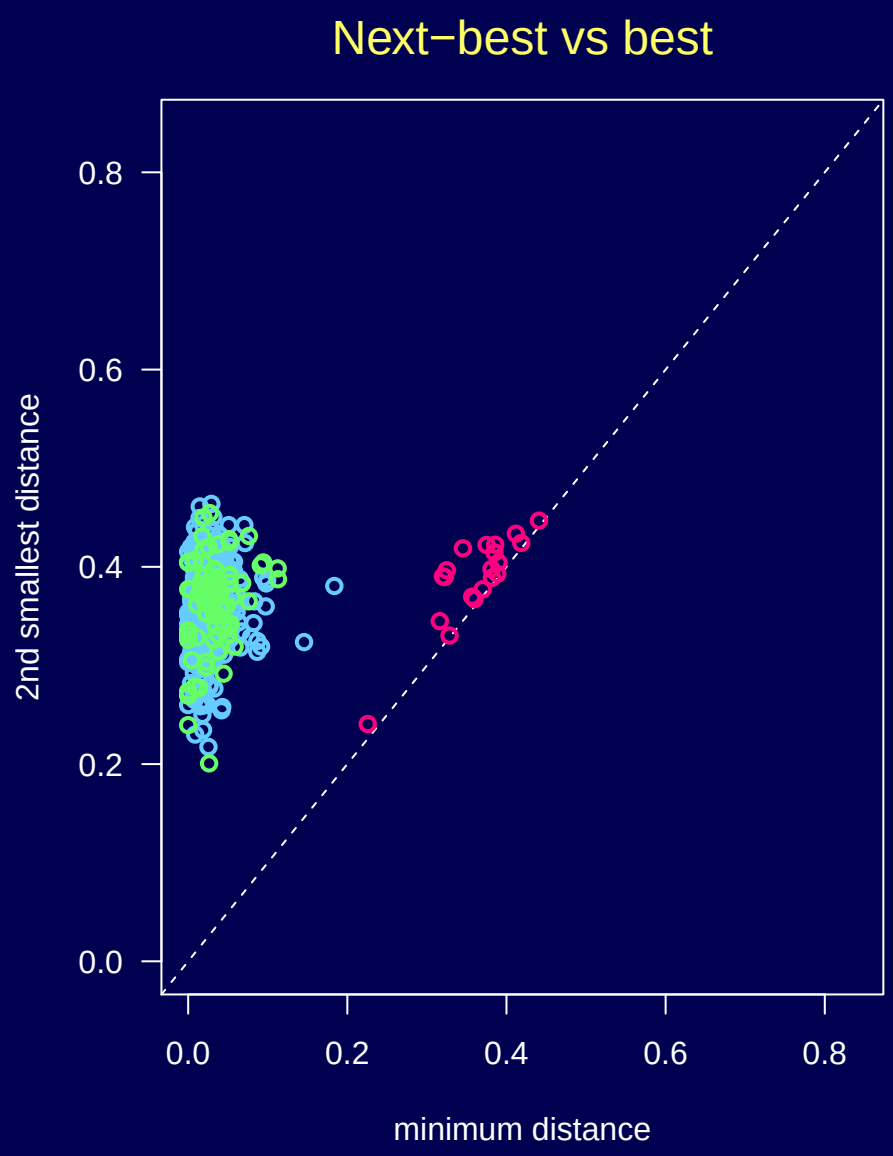
Acknowledgments

Alan Attie Mark Keller	Biochemistry, UW–Madison
Brian Yandell	Statistics and Horticulture, UW–Madison
Christina Kendziorski Aimee Teo Broman	Biostatistics & Medical Informatics, UW–Madison
Eric Schadt	Pacific Biosciences of California
Danielle Greenawalt Amit Kulkarni	Merck & Co., Inc.

NIH: R01 GM074244, R01 DK066369

Decisions


Self vs best
Fixable
Not found
Good
self−self distance
minimum distance
0.0
0.2
0.4
0.6
0.8


Next−best vs best
2nd smallest distance
minimum distance
0.0
0.2
0.4
0.6
0.8