# Genetic analysis of high-throughput phenotypes

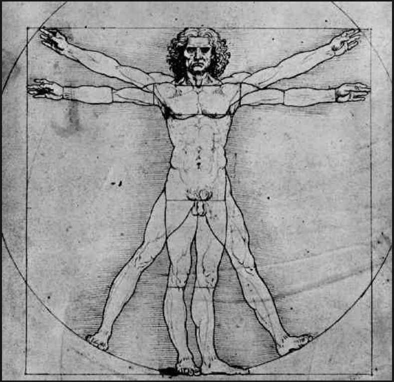
## Challenges and opportunities

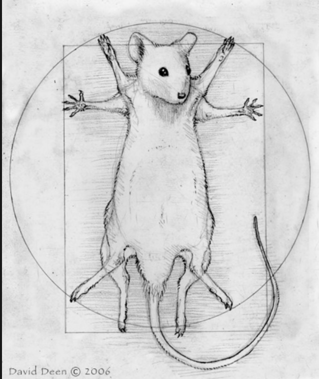
**Karl Broman**

Biostatistics & Medical Informatics, UW–Madison

`kbroman.org`
`github.com/kbroman`
`@kwbroman`

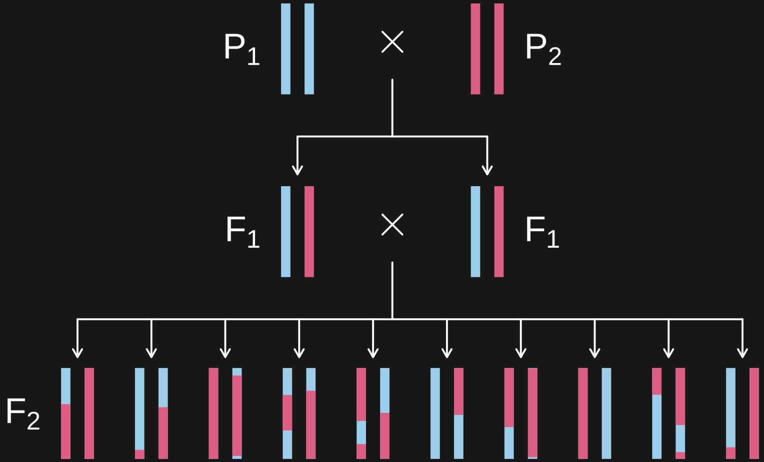
Slides: `bit.ly/pbpg2018intro`



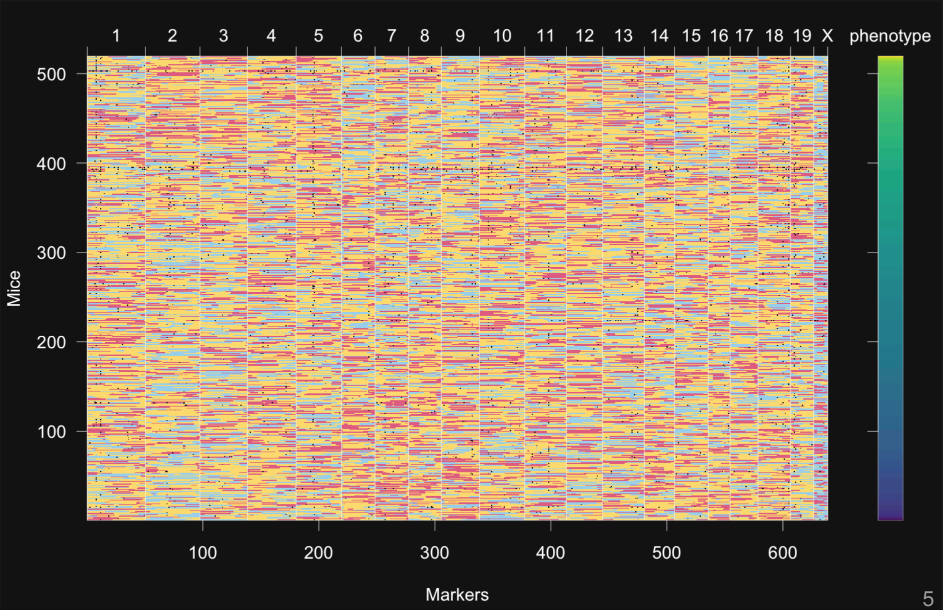


daviddeen.com

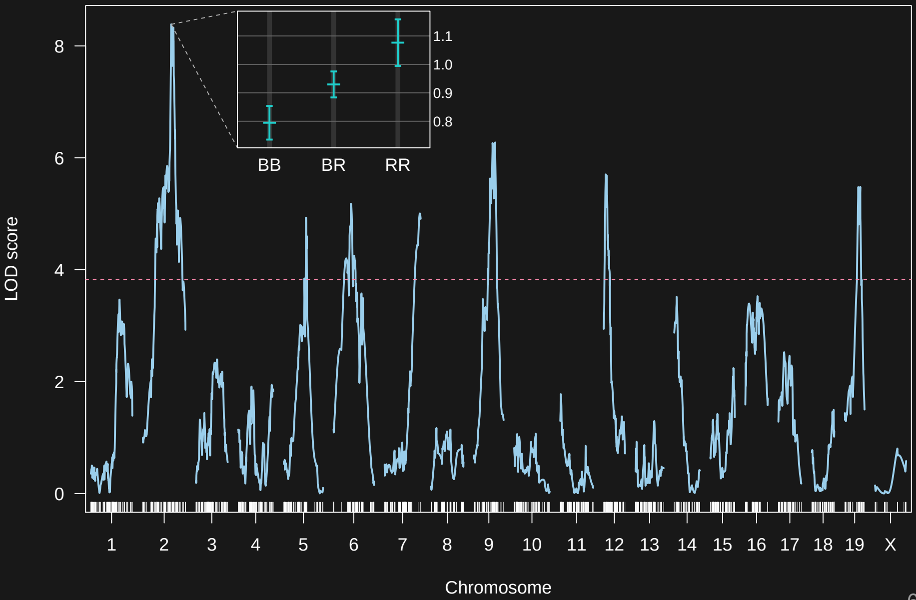
# Intercross



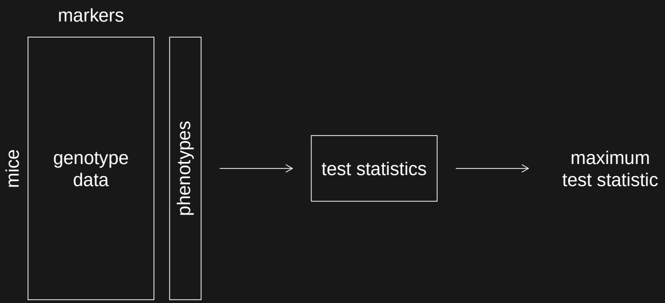
# Data


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
phenotype
500
400
300
200
100
Mice
100
200
300
400
500
600
Markers

# QTL mapping


8
6
4
2
0
LOD score
1.1
1.0
0.9
0.8
BB
BR
RR
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome

# Permutation test


markers

mice

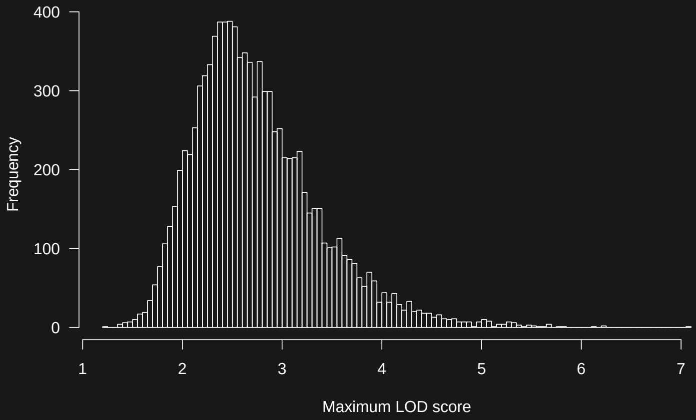
genotype data

phenotypes

test statistics

maximum test statistic

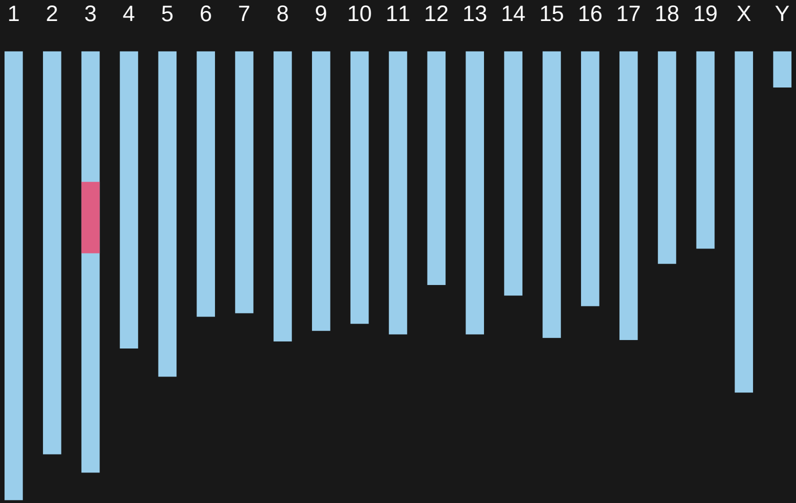
# Histogram of permutation results



# Modeling multiple QTL

- Reduce residual variation ⟶ increased power
- Separate linked QTL
- Identify interactions among QTL (epistasis)

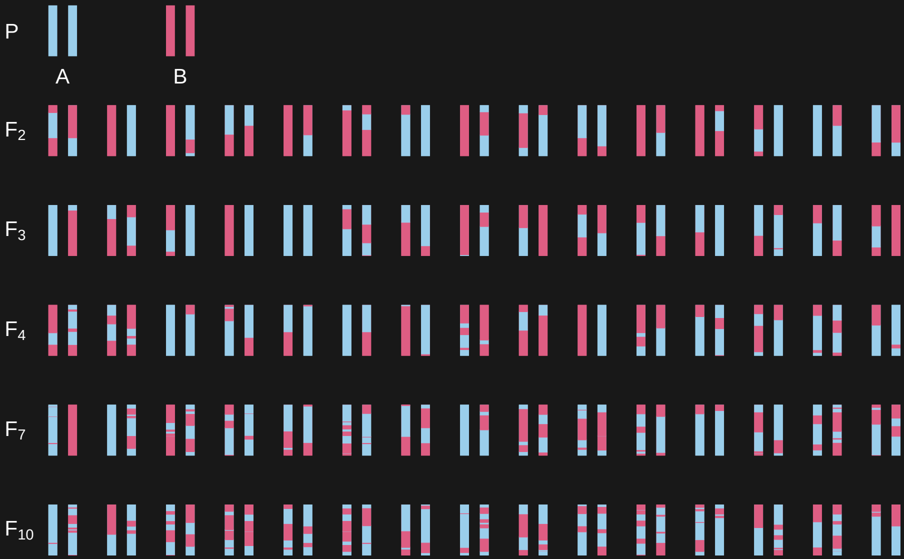
# Congenic line (NIL)



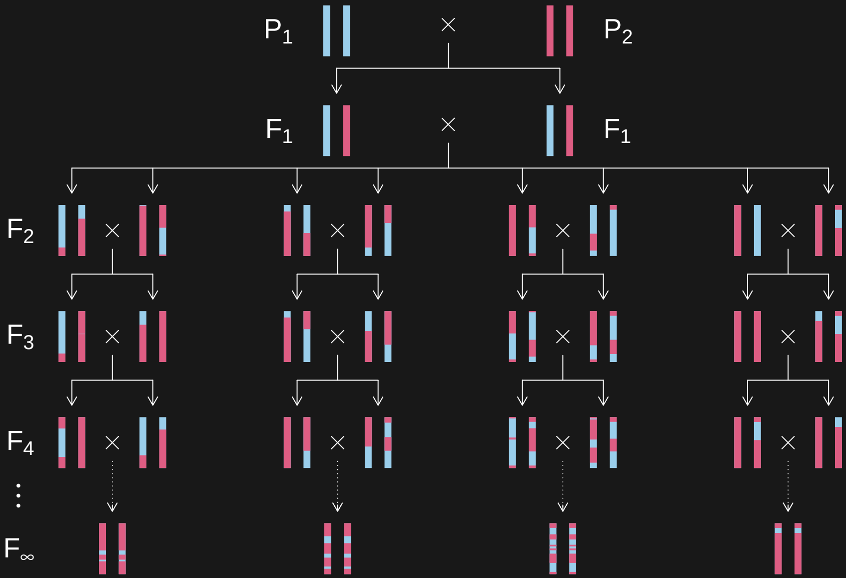
# Improving precision

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

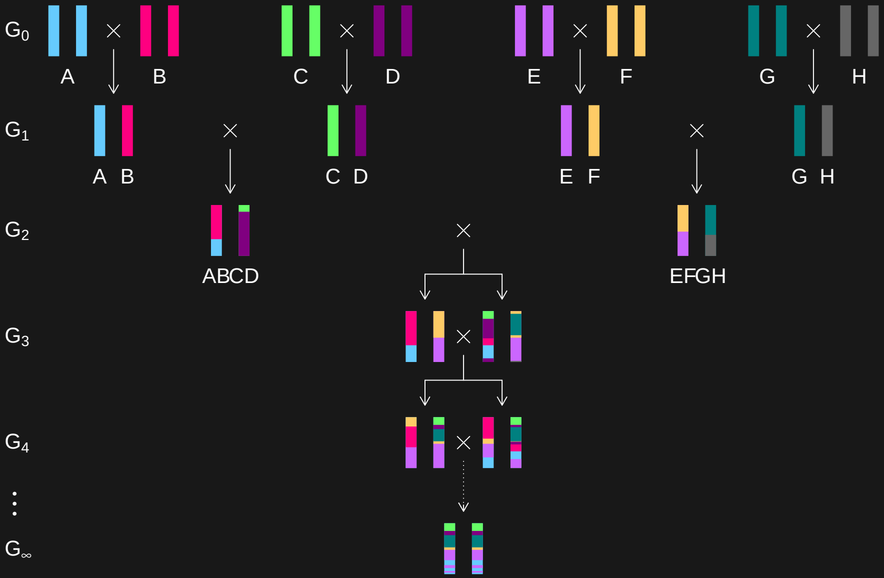
# Advanced intercross lines



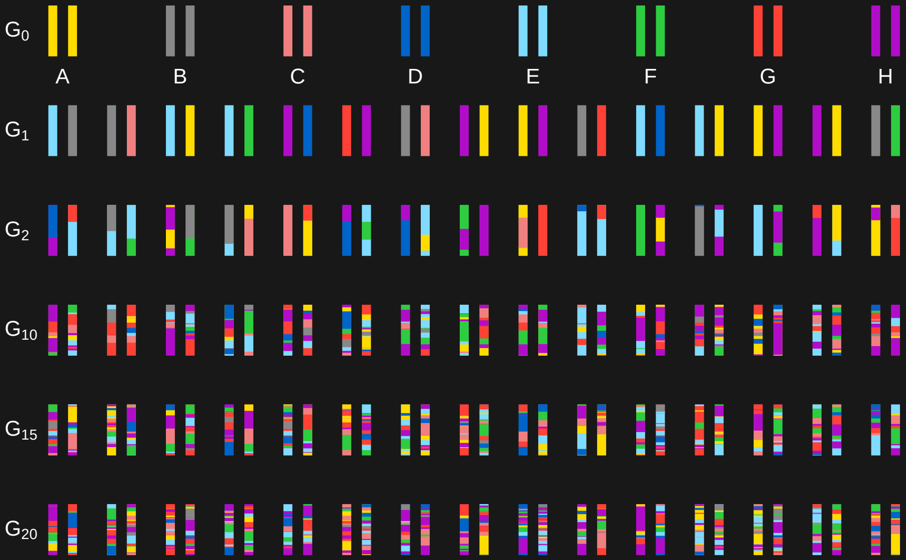
# Recombinant inbred lines



# Collaborative Cross / MAGIC

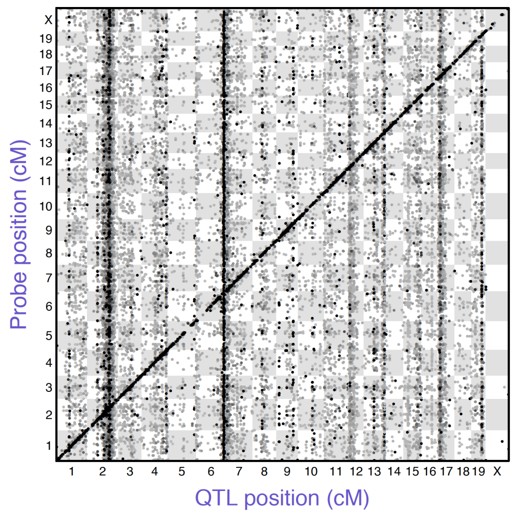


# Heterogeneous stock



# Genome-scale phenotypes

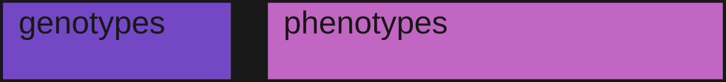
# eQTL


Probe position (cM)
X
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
QTL position (cM)

# Challenges: diagnostics

- What might have gone wrong?
- How might it be revealed?
- Make lots of graphs
- Follow up artifacts

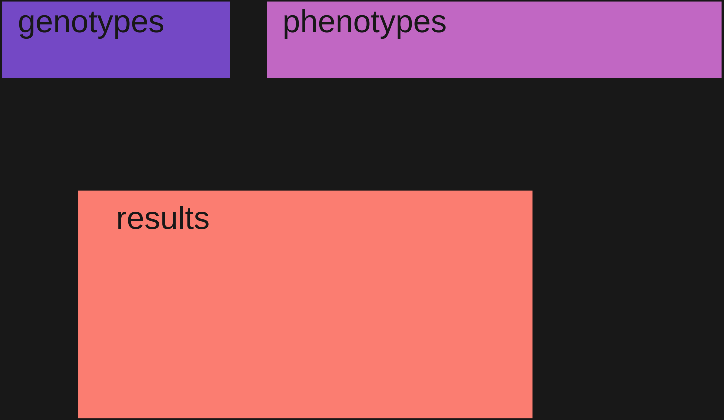
# Challenges: scale of results



genotypes

phenotypes

# Challenges: scale of results



genotypes

phenotypes

results

# Challenges: organizing, automating

genotypes

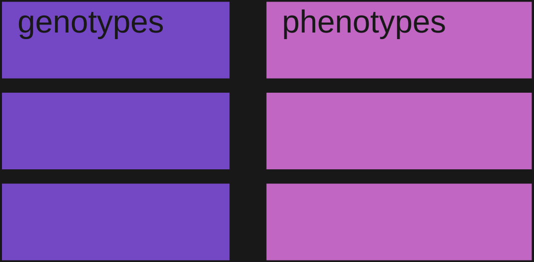
phenotypes

# Challenges: organizing, automating

genotypes

phenotypes

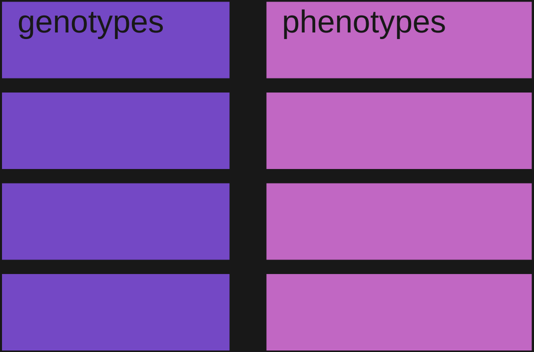
# Challenges: organizing, automating



genotypes

phenotypes

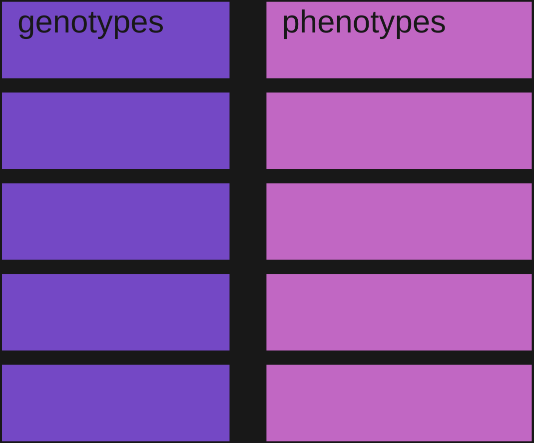
# Challenges: organizing, automating



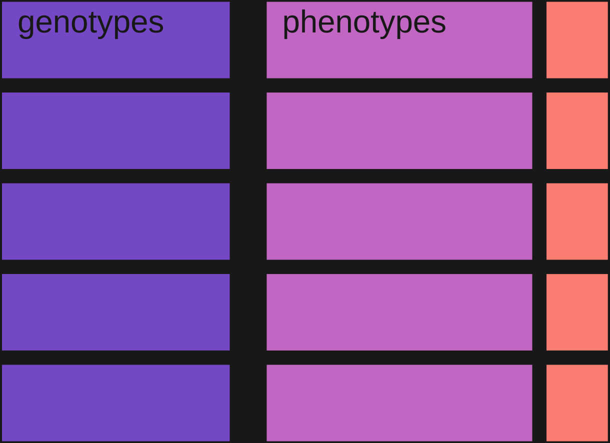
genotypes

phenotypes

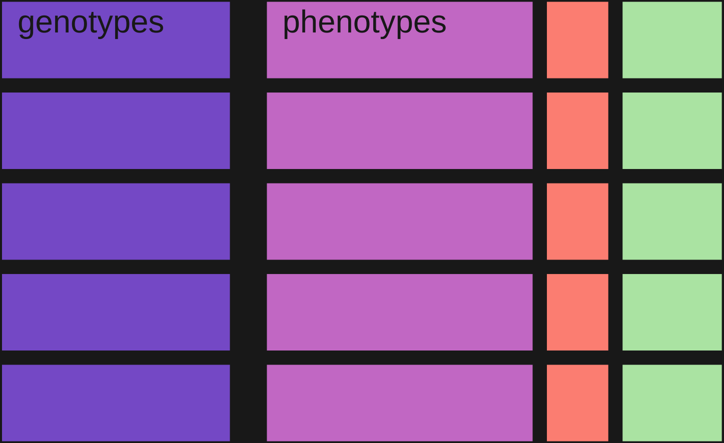
# Challenges: organizing, automating



# Challenges: organizing, automating



# Challenges: organizing, automating

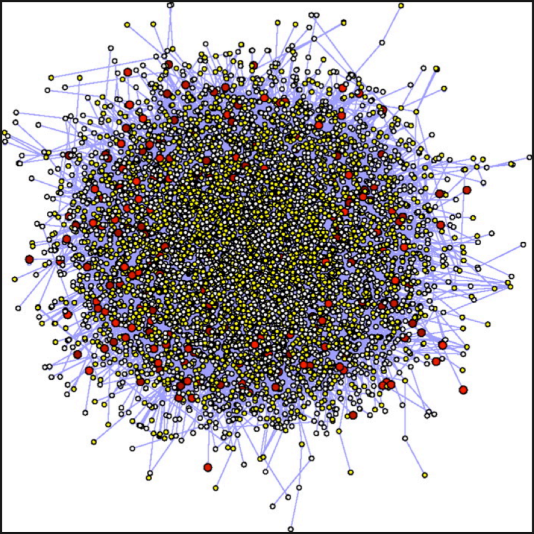


# Challenges: metadata

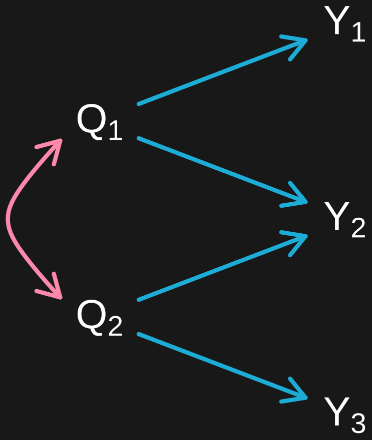
What the heck is `"FAD_NAD SI 8.3_3.3G"`?

What was the question again?

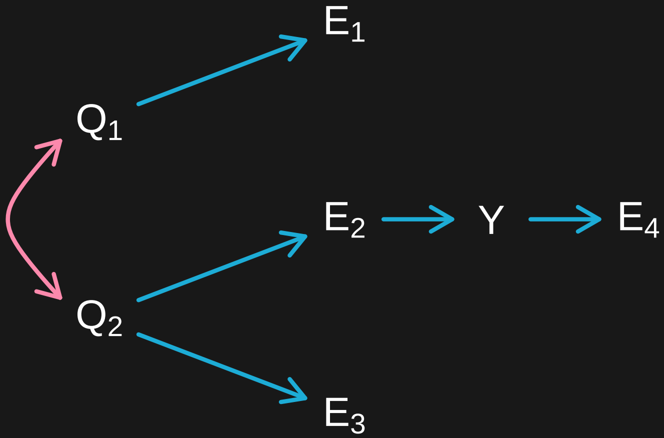
# The ridiculome



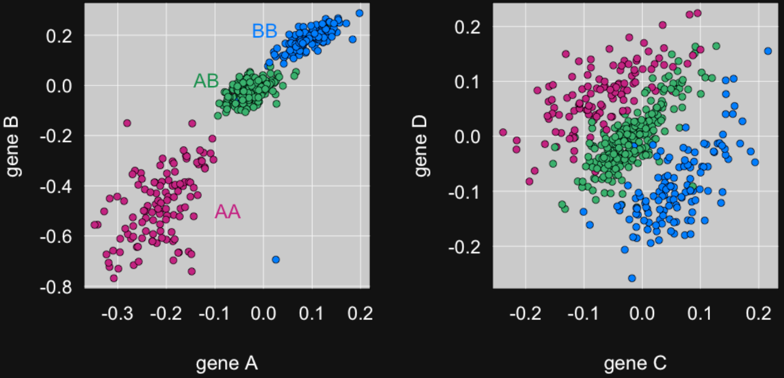
# Pleiotropy?



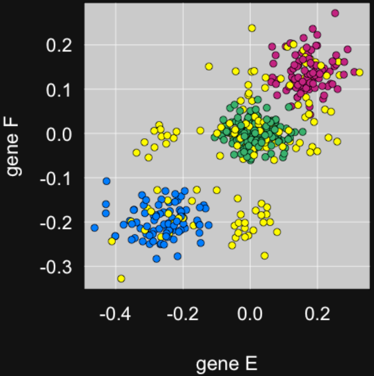
# Causal?



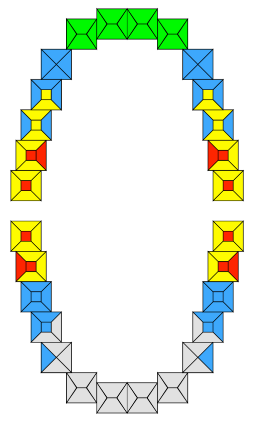
# Multivariate phenotypes



# Multivariate phenotypes



# Composite phenotypes



Shaffer et al. (2013) J Dent Res 92:32-37 share more data, sooner

Are your results reproducible?

cf Baggerly & Coombes (2009)

`projecteuclid.org/euclid.aoas/1267453942`

# Reproducibility

```
Karl -- this is very interesting,
however you used an old version of
the data (n=143 rather than n=226).

I'm really sorry you did all that
work on the incomplete dataset.

Bruce
```

# Steps toward reproducible research

1. Organize your data & code
2. Everything with a script
3. Automate the process (GNU Make)
4. Turn scripts into reproducible reports
5. Turn repeated code into functions
6. Create a package/module
7. Use version control (git/GitHub)
8. Pick a license, any license

`kbroman.org/steps2rr`

# Key challenges in QTL mapping

- Multiple phenotypes
  - Consider jointly to refine QTL location
  - Common or separate QTL?
  - Tease apart cause
- QTL × environment interactions
- Analysis methods for multi-parent populations
- Data diagnostics
- Data, software, and results management
- Data visualization

Slides: `bit.ly/pbpg2018intro`

`kbroman.org`

`github.com/kbroman`

`@kwbroman`