

# Reanalysis of

## The effects of genetic variation on gene expression dynamics during development

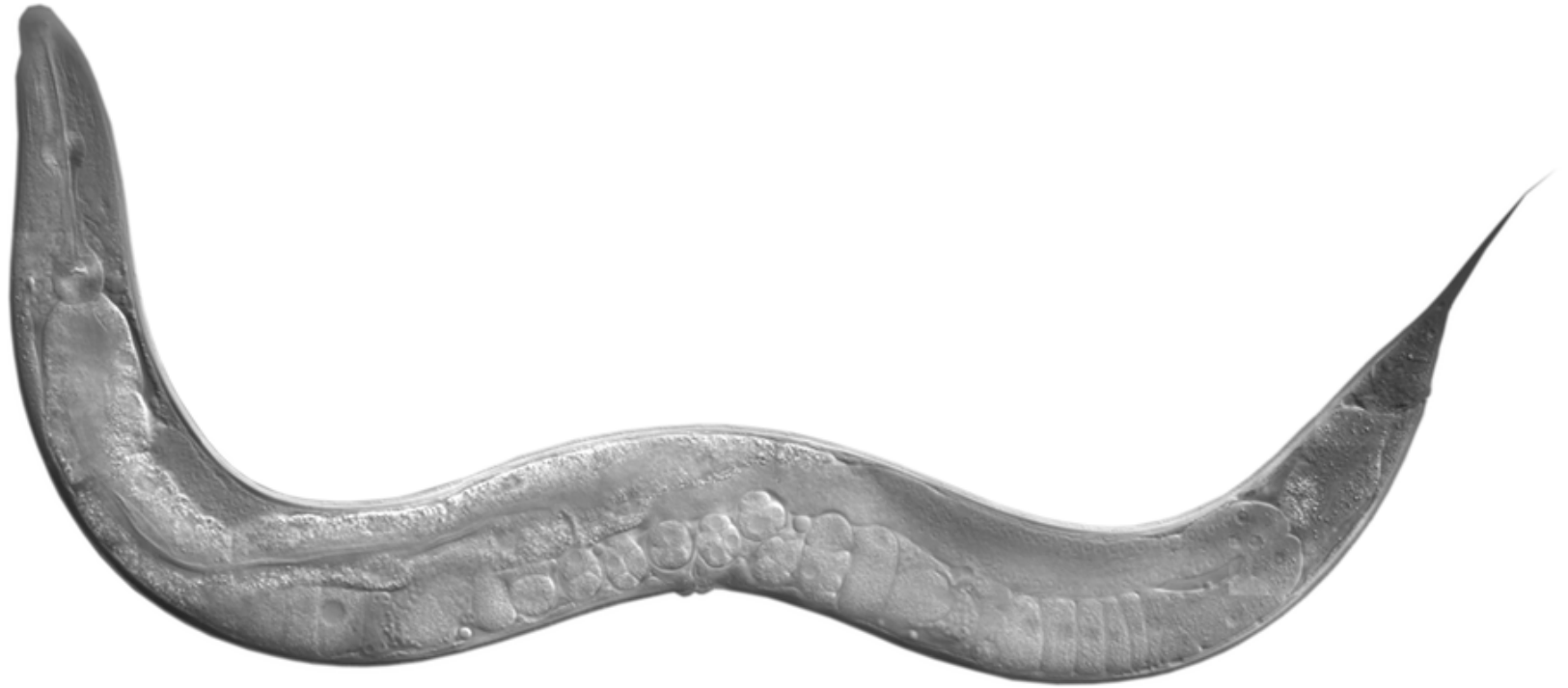
*Mirko Francesconi & Ben Lehner Nature 2014*

Stefan Zdraljevic  
Statistical Approaches for Bioinformatics

# Overview of paper

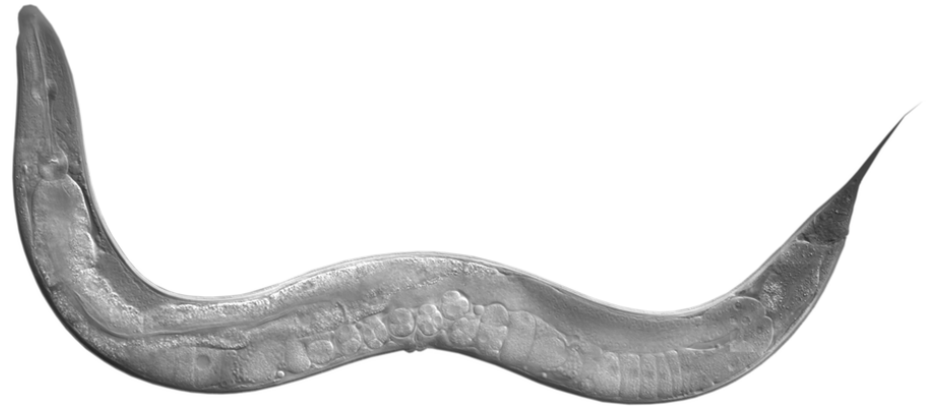
1. Take microarray data from two experiments (RIAIL and developmental time course)
2. Line up RIAILs to reference time course
3. Perform linkage mapping to identify QTL using original data set and taking RIAILs developmental time into account

## Background - *C. elegans* as a model

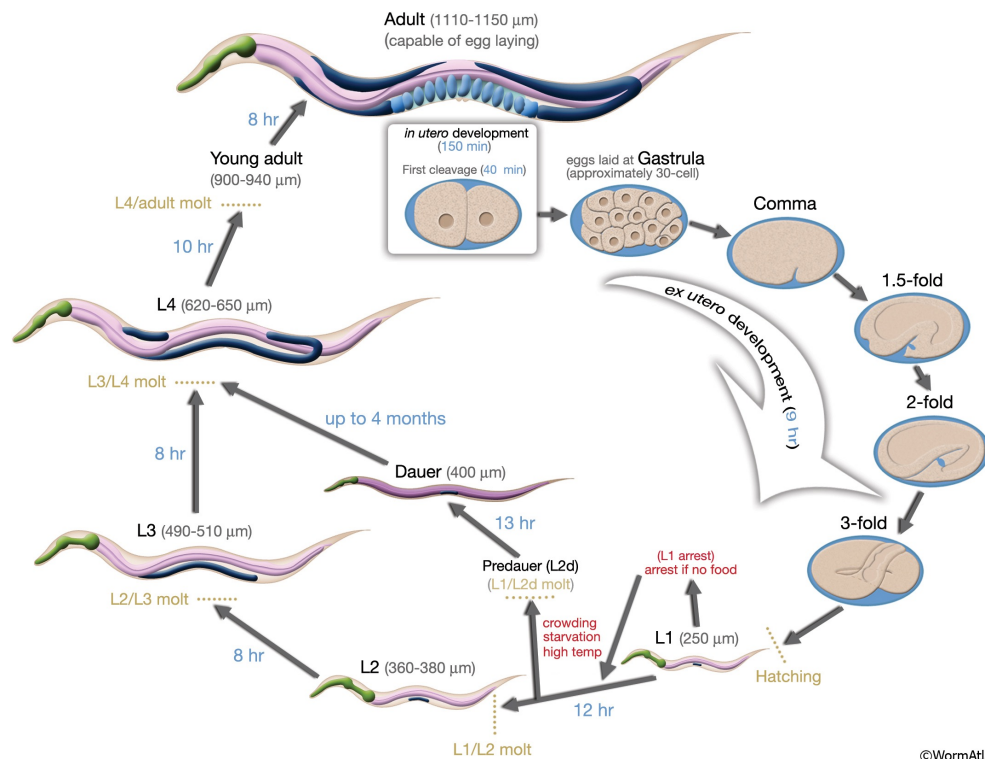


*C. elegans* is a free-living nematode that is used as a metazoan model due to its ease of handling and conservation of important human signaling pathways.

# Background



*C. elegans* is a free-living nematode that is used as a metazoan model due to its ease of handling and conservation of important human signaling pathways.



The *C. elegans* life cycle consists of embryonic development, four larval stages, and sexually mature adults.

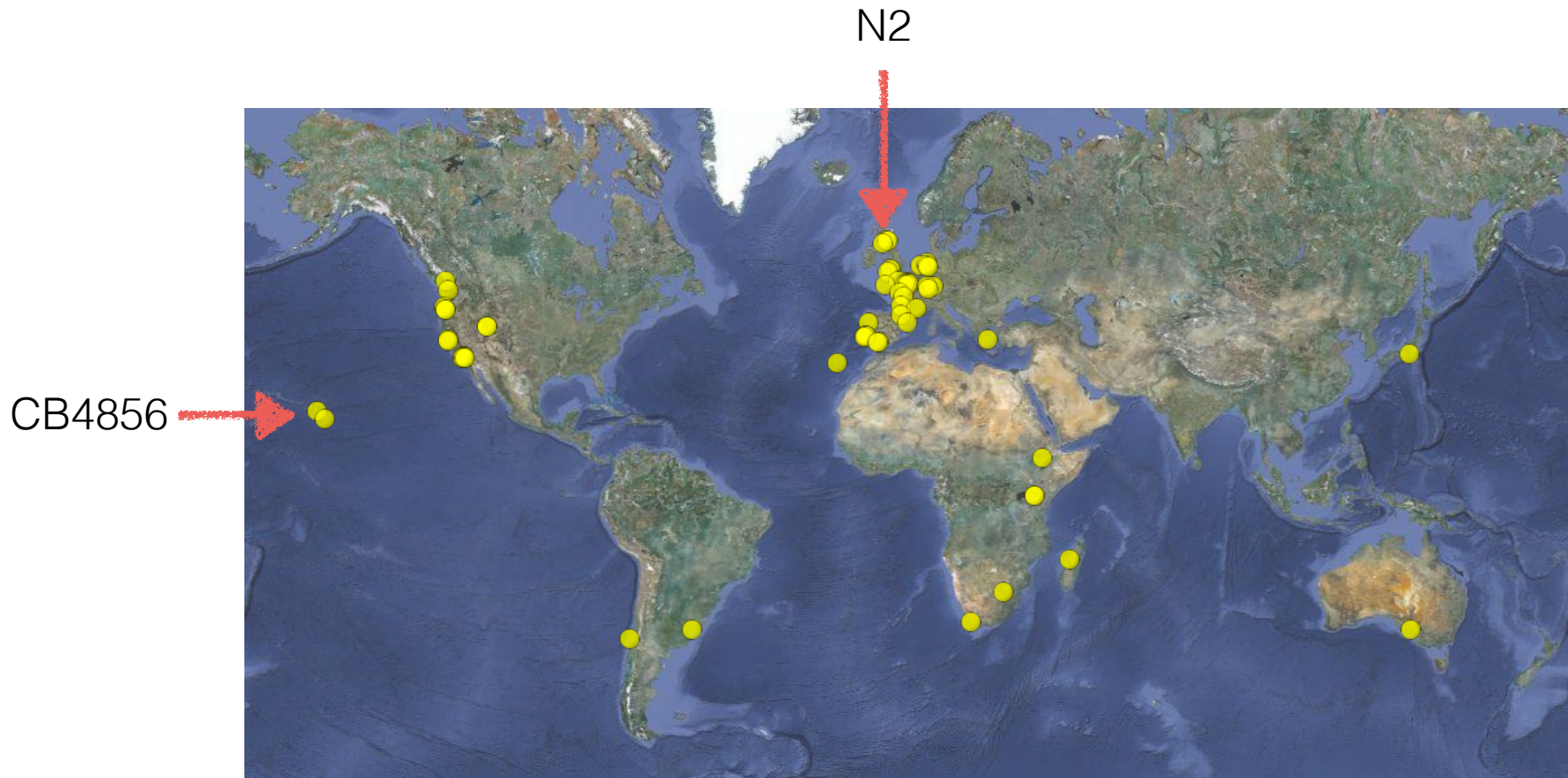
*C. elegans* mainly exists as a hermaphrodite and the main mode of reproduction is selfing

# Background



The global distribution of isolated *C. elegans* individuals.

## Background - CB4856 and N2



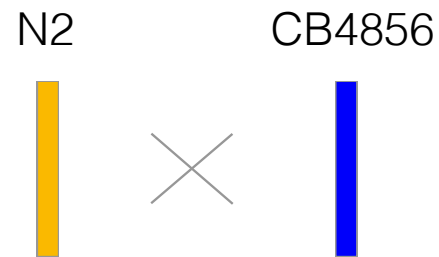
The two isolates used in the study are N2 and CB4856. N2 is the commonly used lab strain of *C. elegans* and CB4856 is a wild isolate from Hawaii. There is approximately one variant per 800 nucleotides between the two individuals (comparable to variation observed in humans).



# Background - RIALs

Recombinant Inbred Advanced Intercross Lines

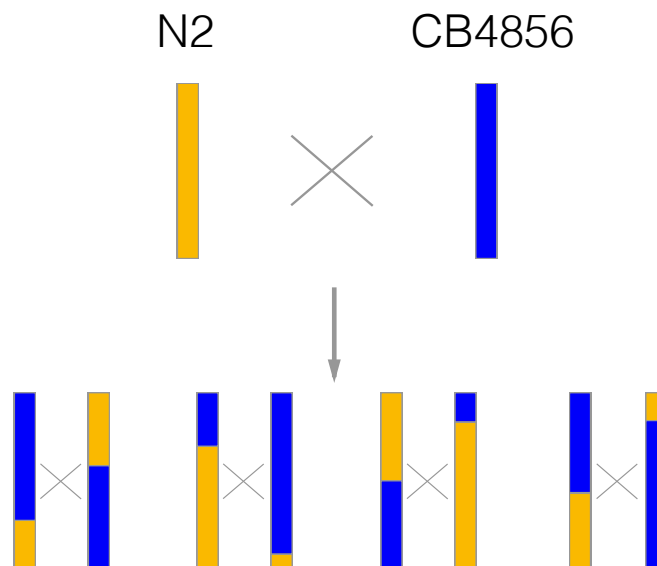
Mix up the genomes of two diverged strains



# Background - RIALs

Recombinant Inbred Advanced Intercross Lines

Mix up the genomes of two diverged strains

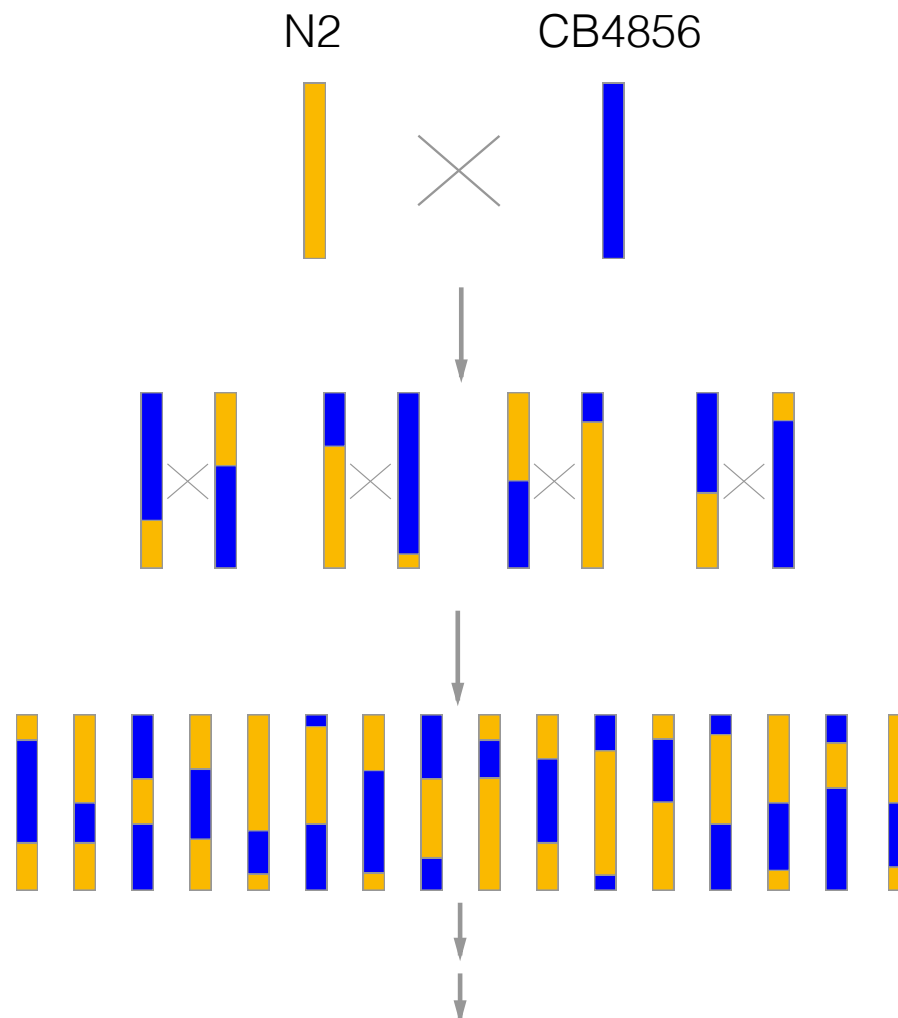




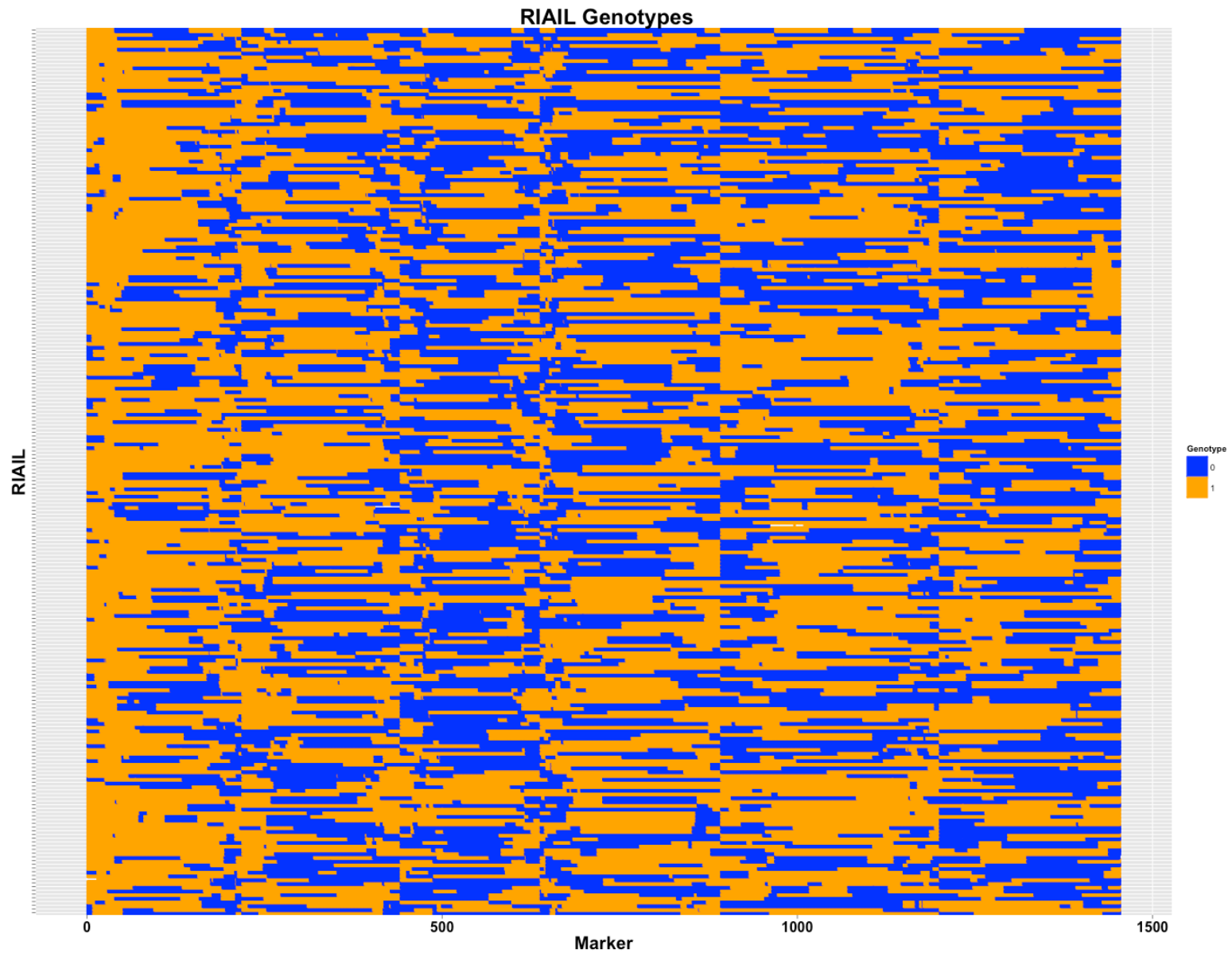
# Background - RIALs

Recombinant Inbred Advanced Intercross Lines

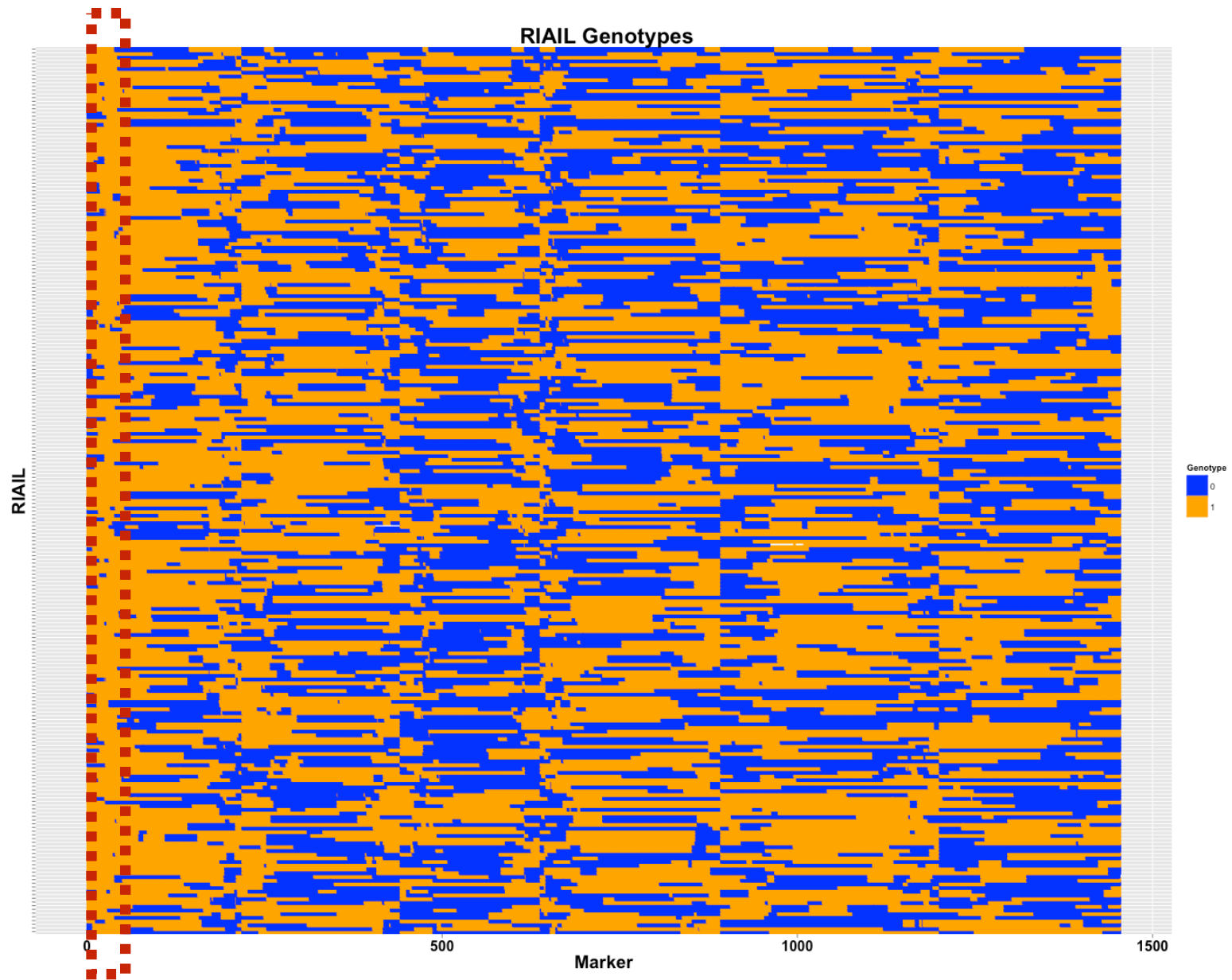
Mix up the genomes of two diverged strains



# Background - RIAILs



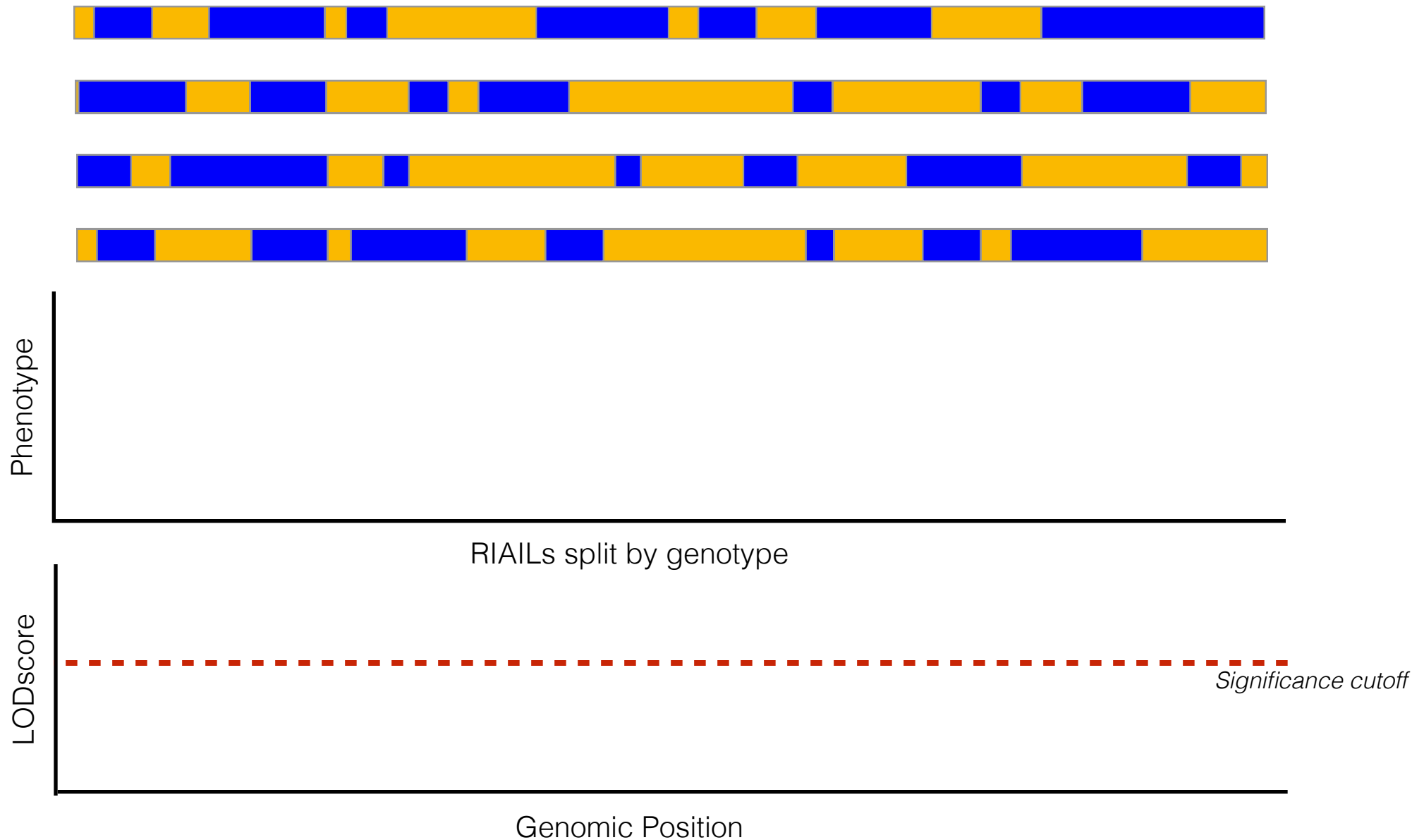
# Background - RIALs



*peel-1/zeel-1* genetic incompatibility locus

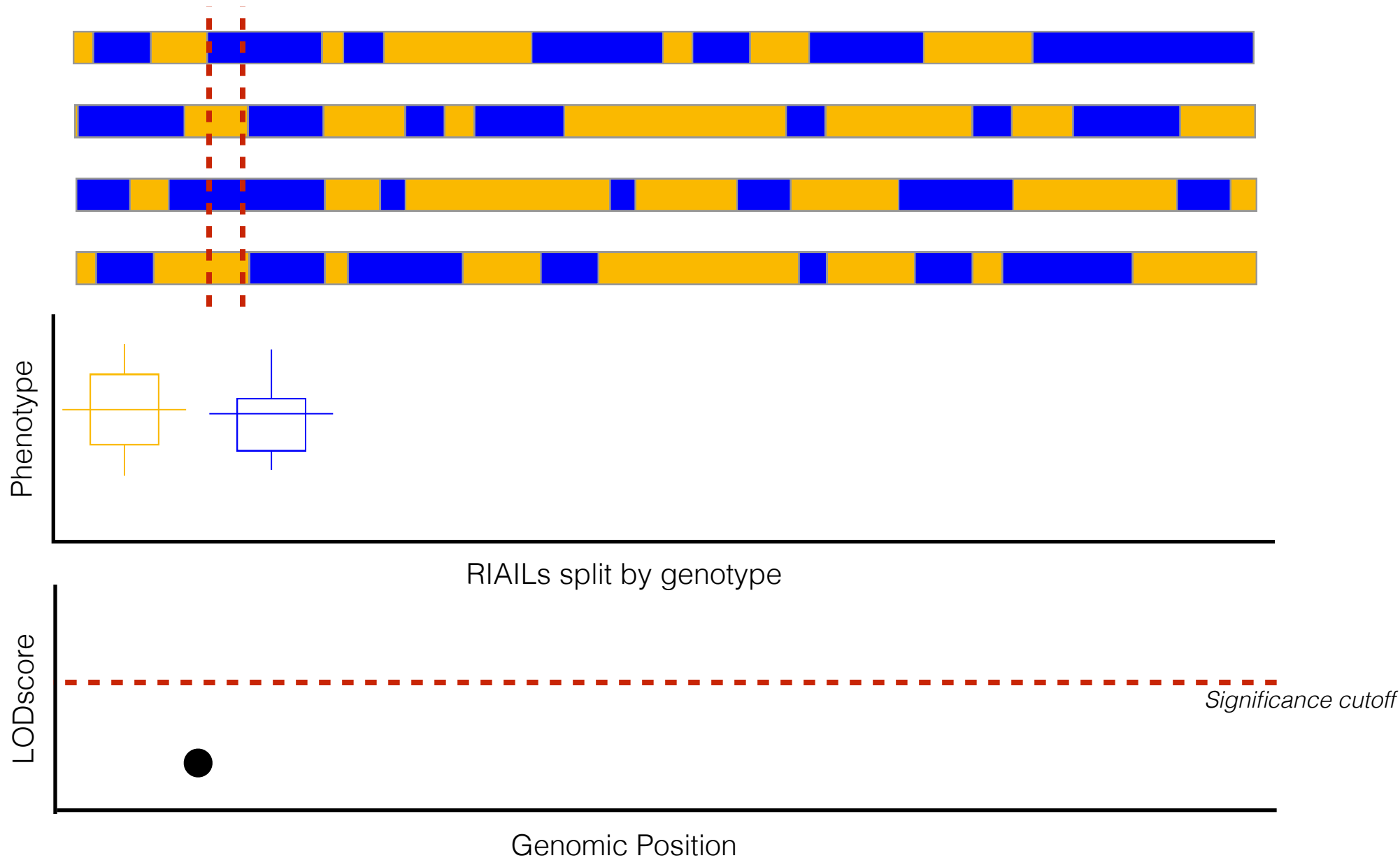
# Background - Linkage Mapping

Test if a phenotype is associated with a region of the genome



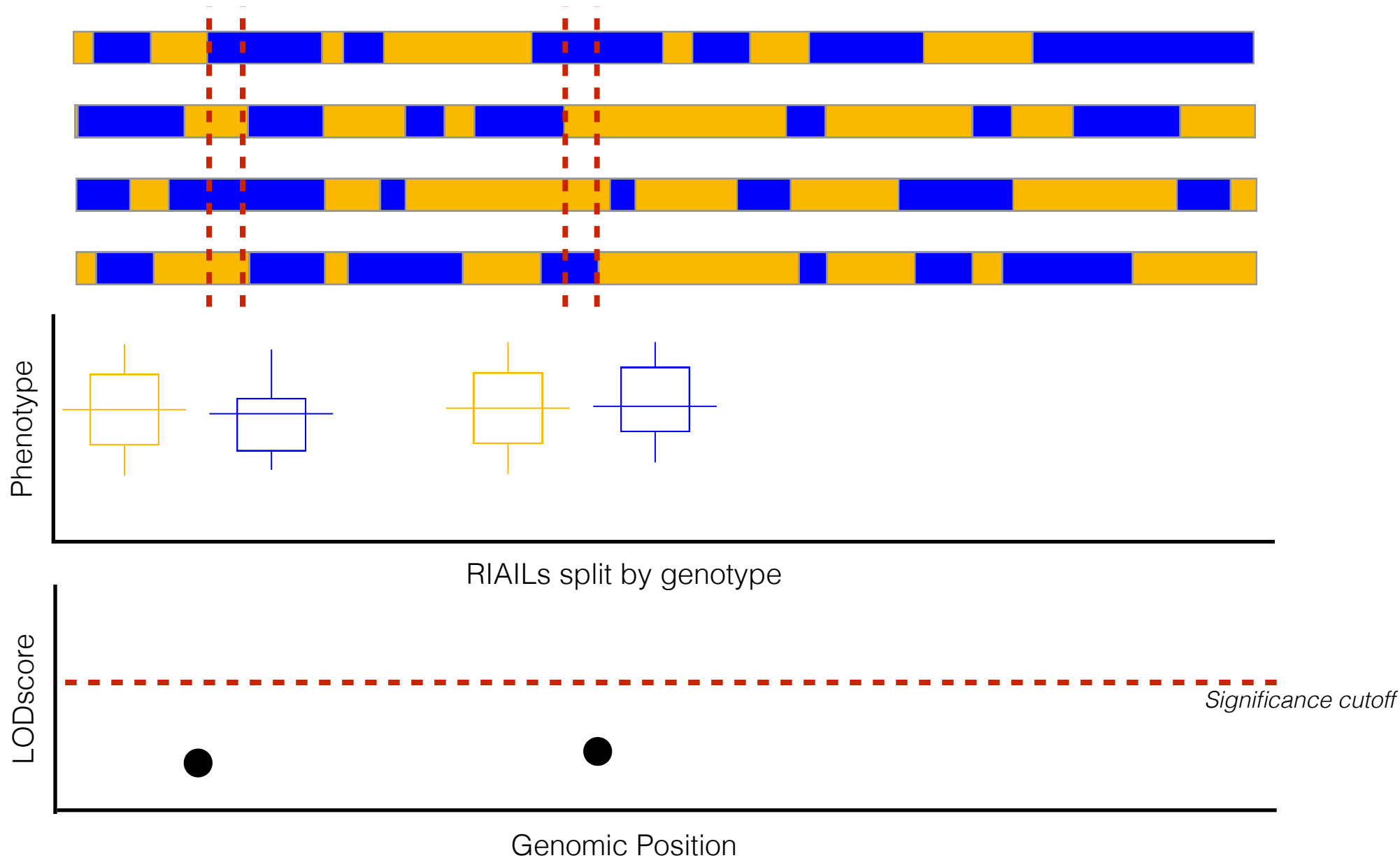
# Background - Linkage Mapping

Test if a phenotype is associated with a region of the genome



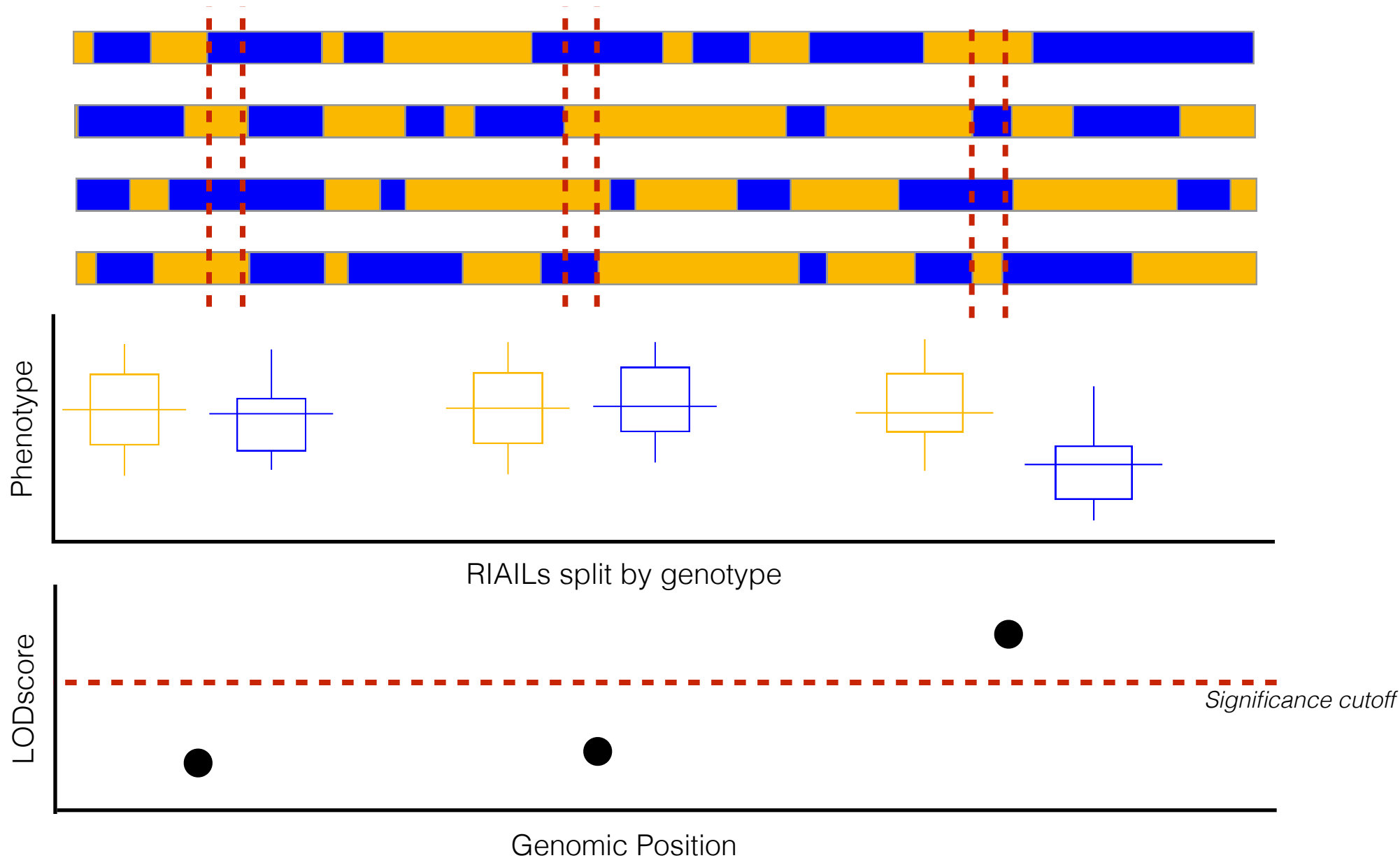
# Background - Linkage Mapping

Test if a phenotype is associated with a region of the genome



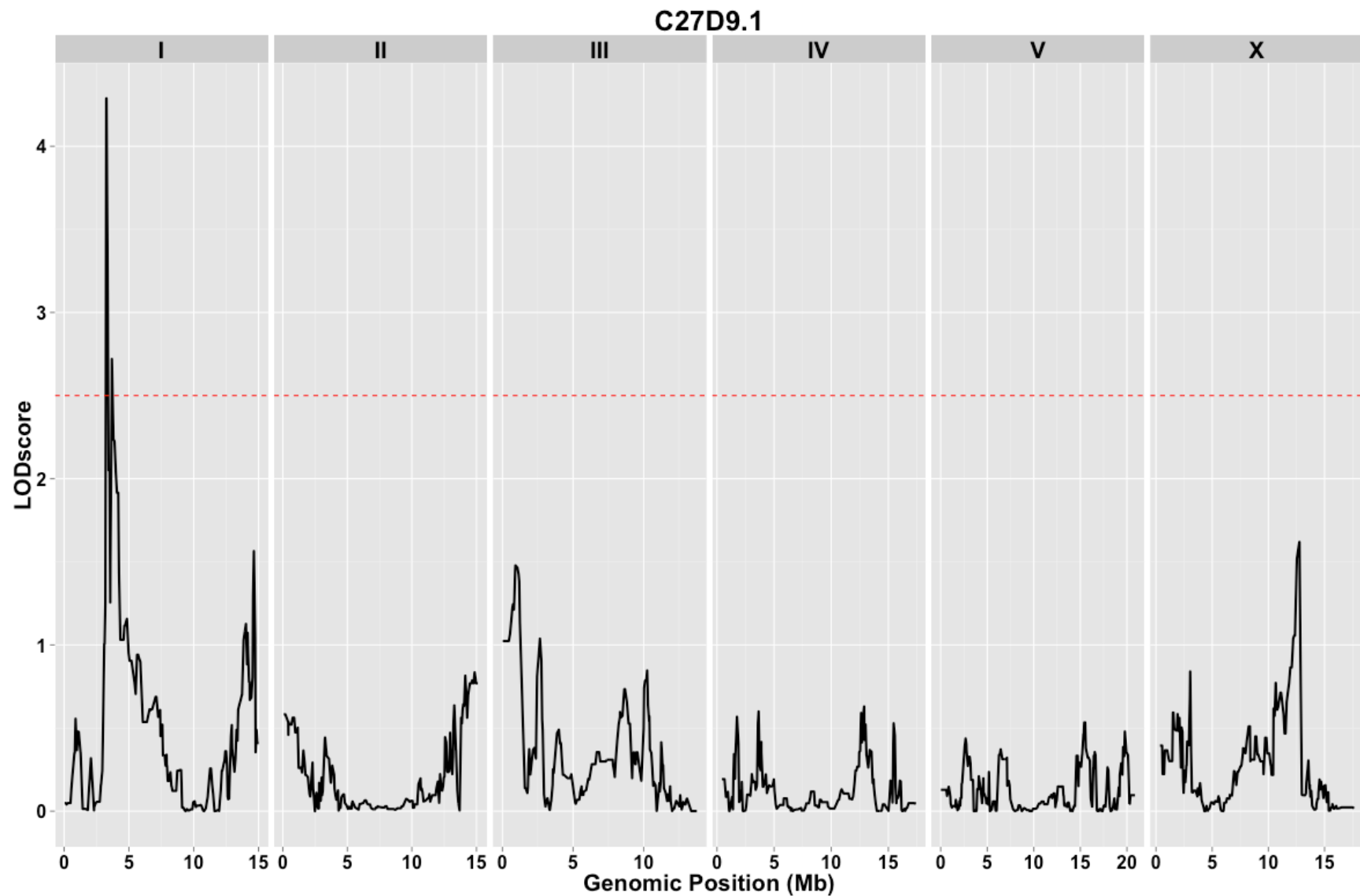
# Background - Linkage Mapping

Test if a phenotype is associated with a region of the genome





# Background - Example Mapping



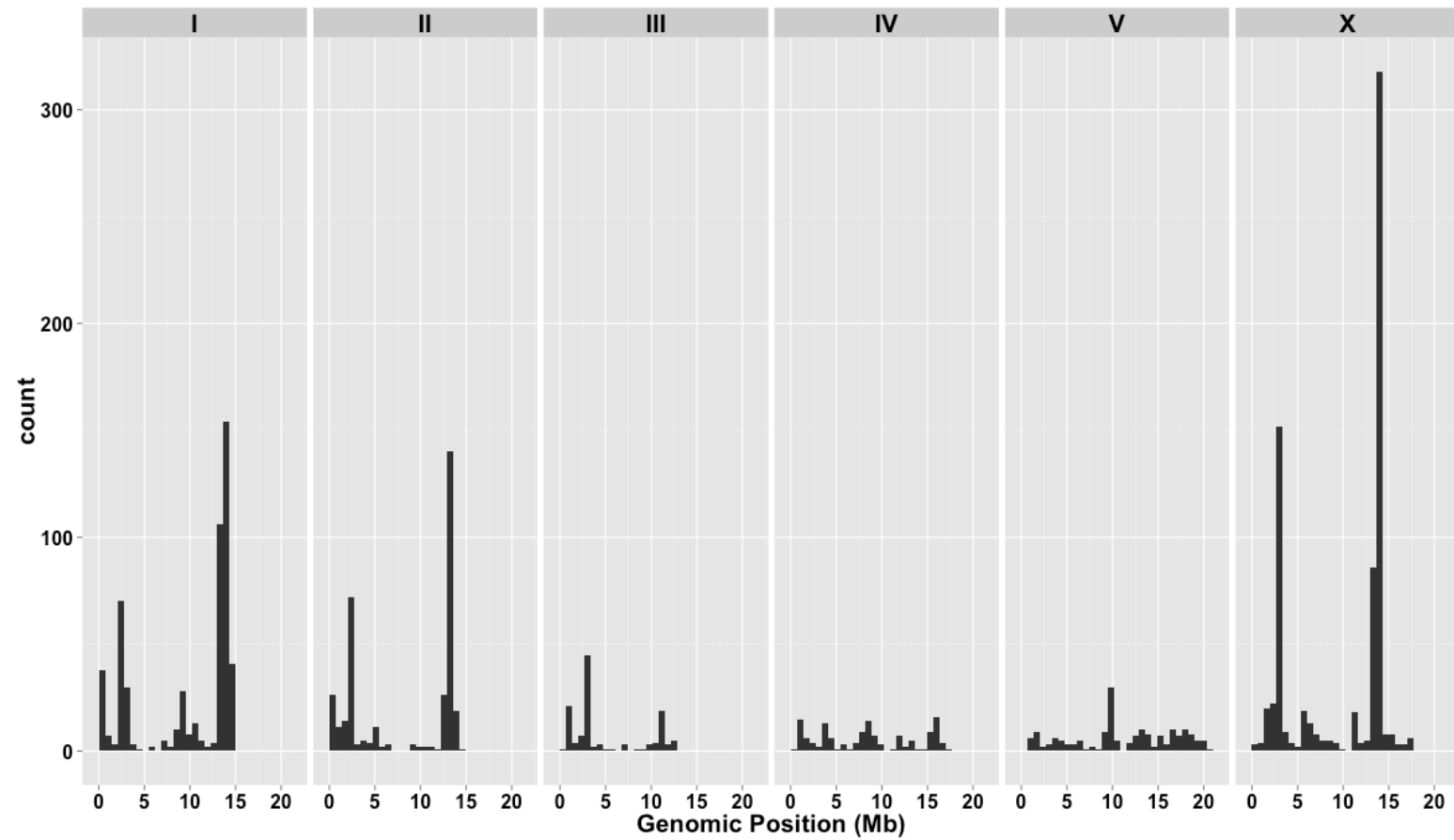
# Background - Experimental Setup

Quantification of gene expression in 206 CB-N2 RIAILs staged at L4

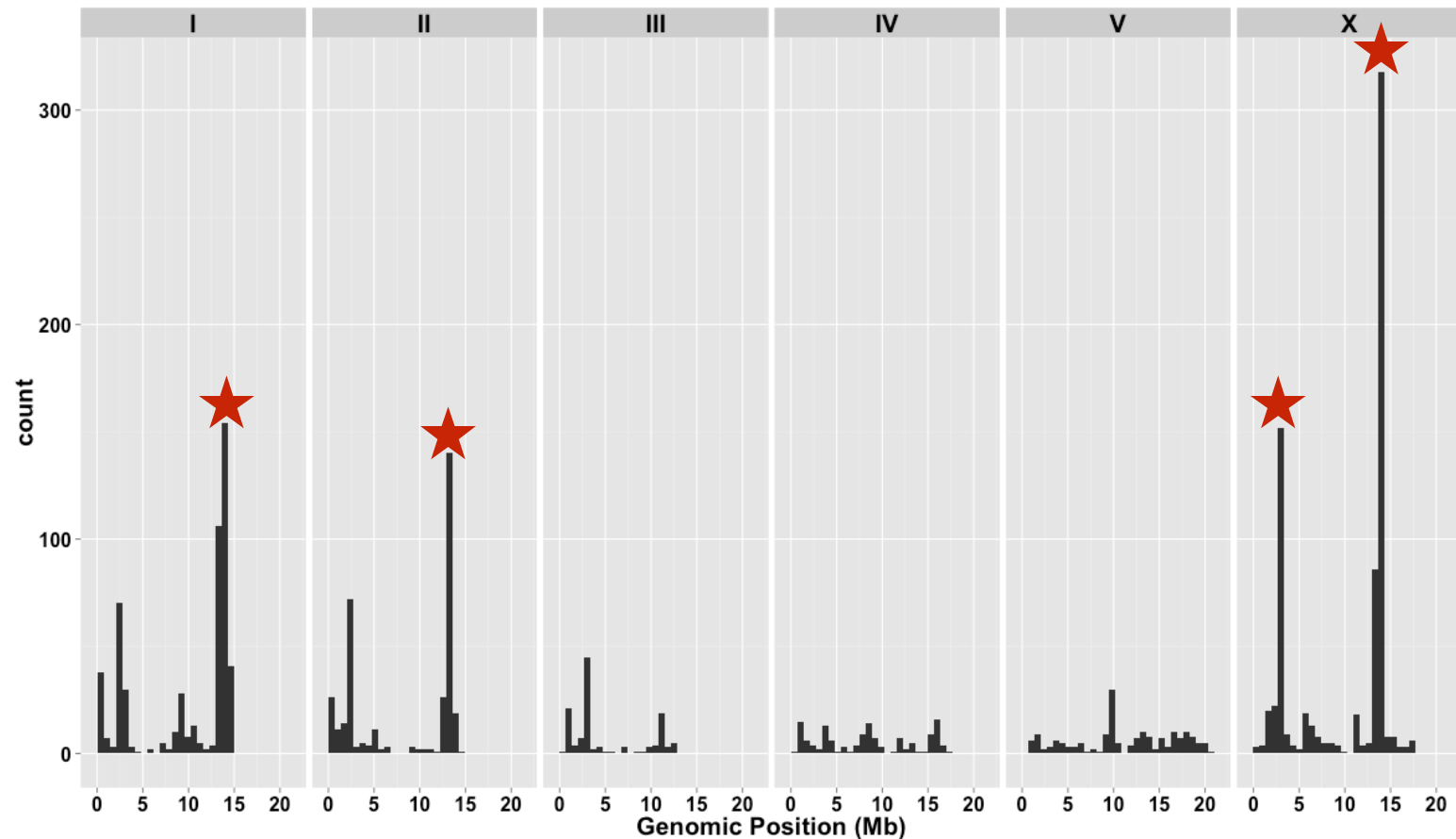
Quantification of gene expression in N2 - 12 time points between L3 and young adults

12200 genes overlap in this data set and were used for canonical correlation analysis

# Distribution of eQTL for raw expression data



# Distribution of eQTL for raw expression data



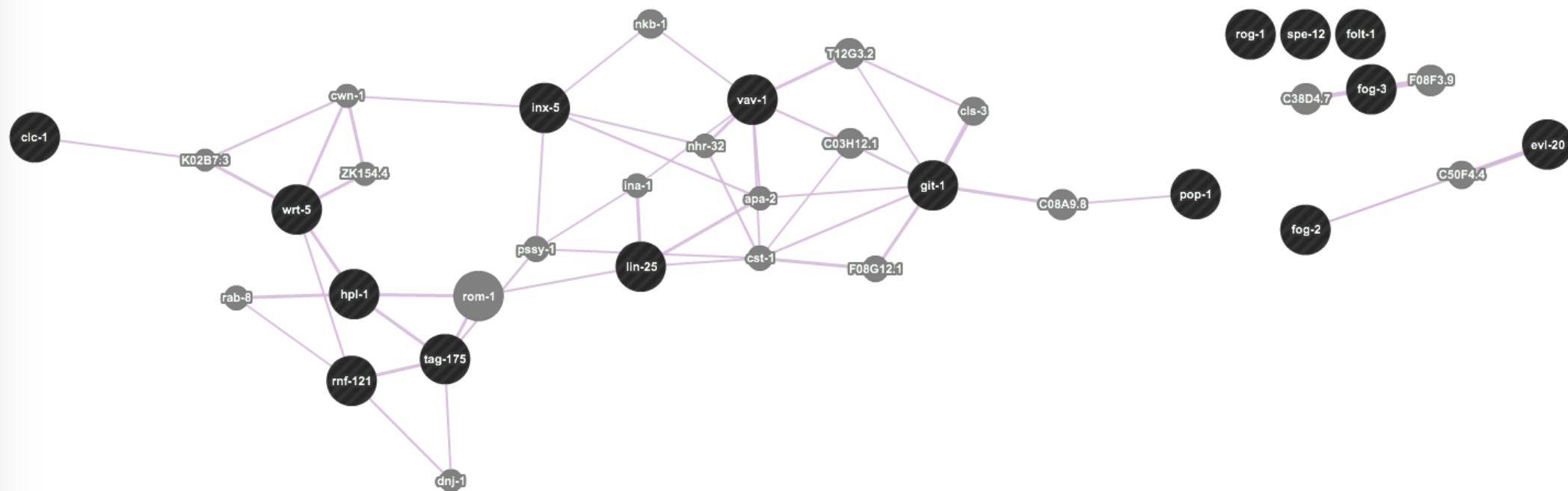
eQTL hotspots are indicative of genes that regulate the expression of multiple genes

# Identification of genes involved in gamete production

clc-1 evl-20 fog-2 fog-3 hpl-1 inx-5 lin-25 pop-1  
spe-12 tag-175 vav-1 wrt-5 folt-1 rnf-121 yap-1 git-1 rog-1

# Identification of genes involved in gamete production

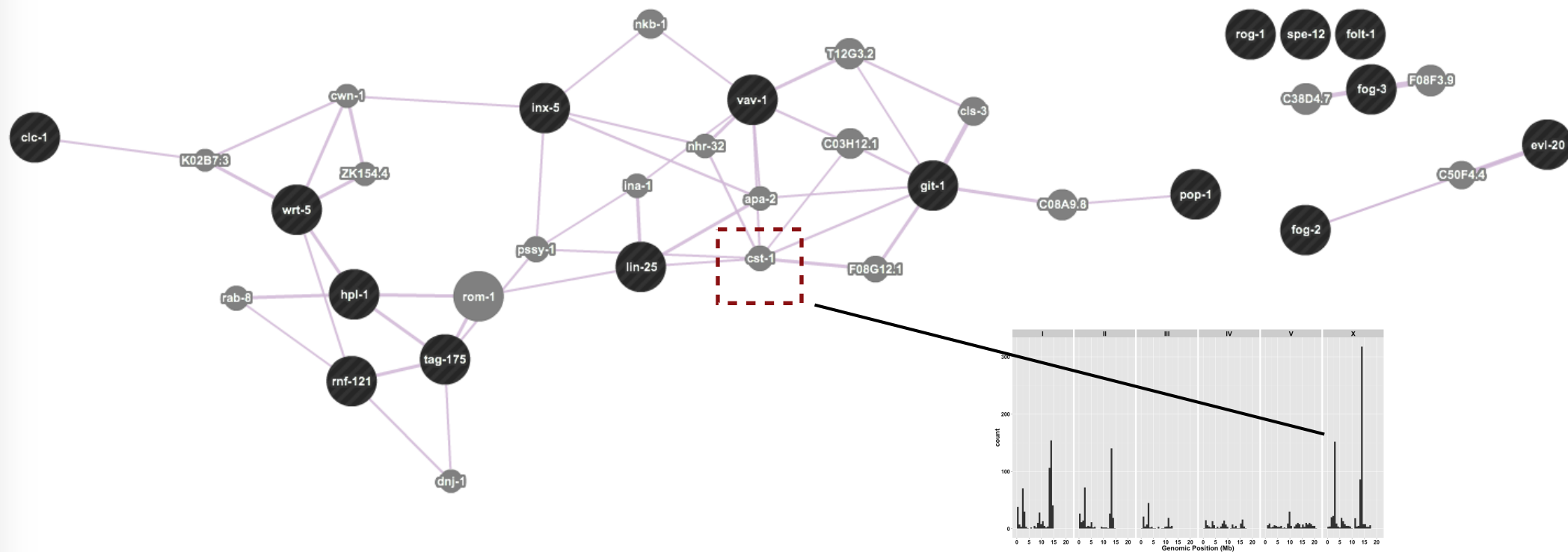
clc-1 evl-20 fog-2 fog-3 hpl-1 inx-5 lin-25 pop-1  
spe-12 tag-175 vav-1 wrt-5 folt-1 rnf-121 yap-1 git-1 rog-1



*purple indicates co-expression*

# Identification of genes involved in gamete production

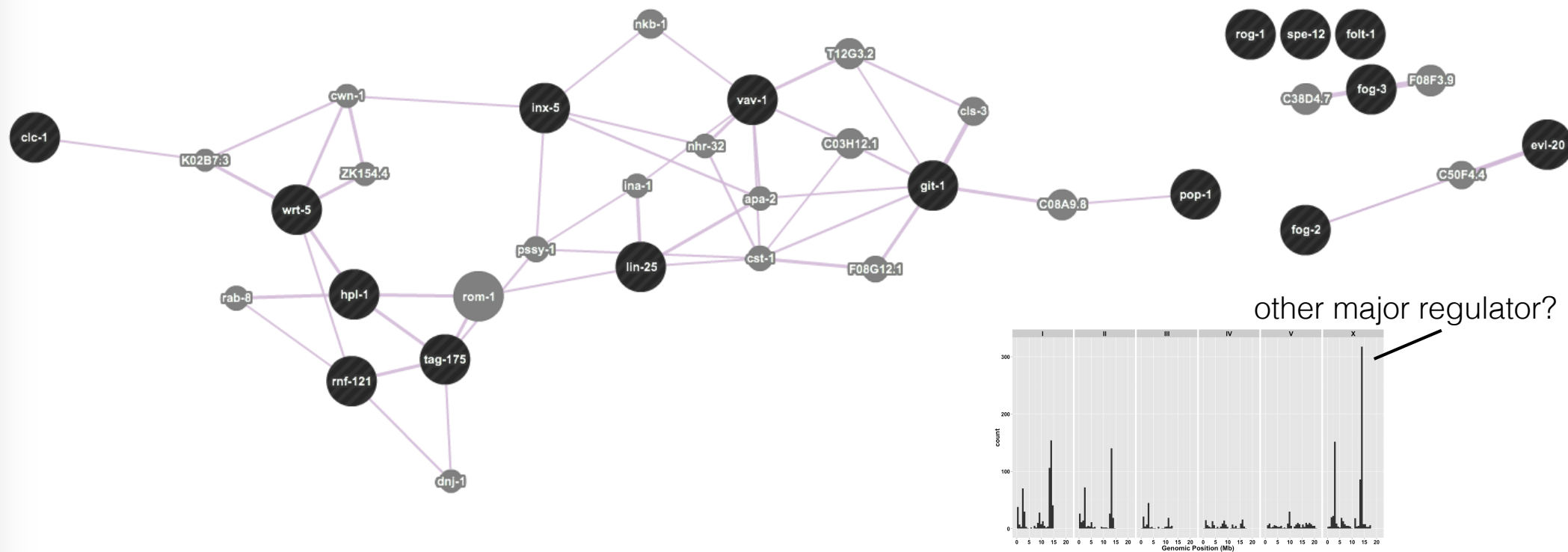
clc-1 evl-20 fog-2 fog-3 hpl-1 inx-5 lin-25 pop-1  
spe-12 tag-175 vav-1 wrt-5 folt-1 rnf-121 yap-1 git-1 rog-1



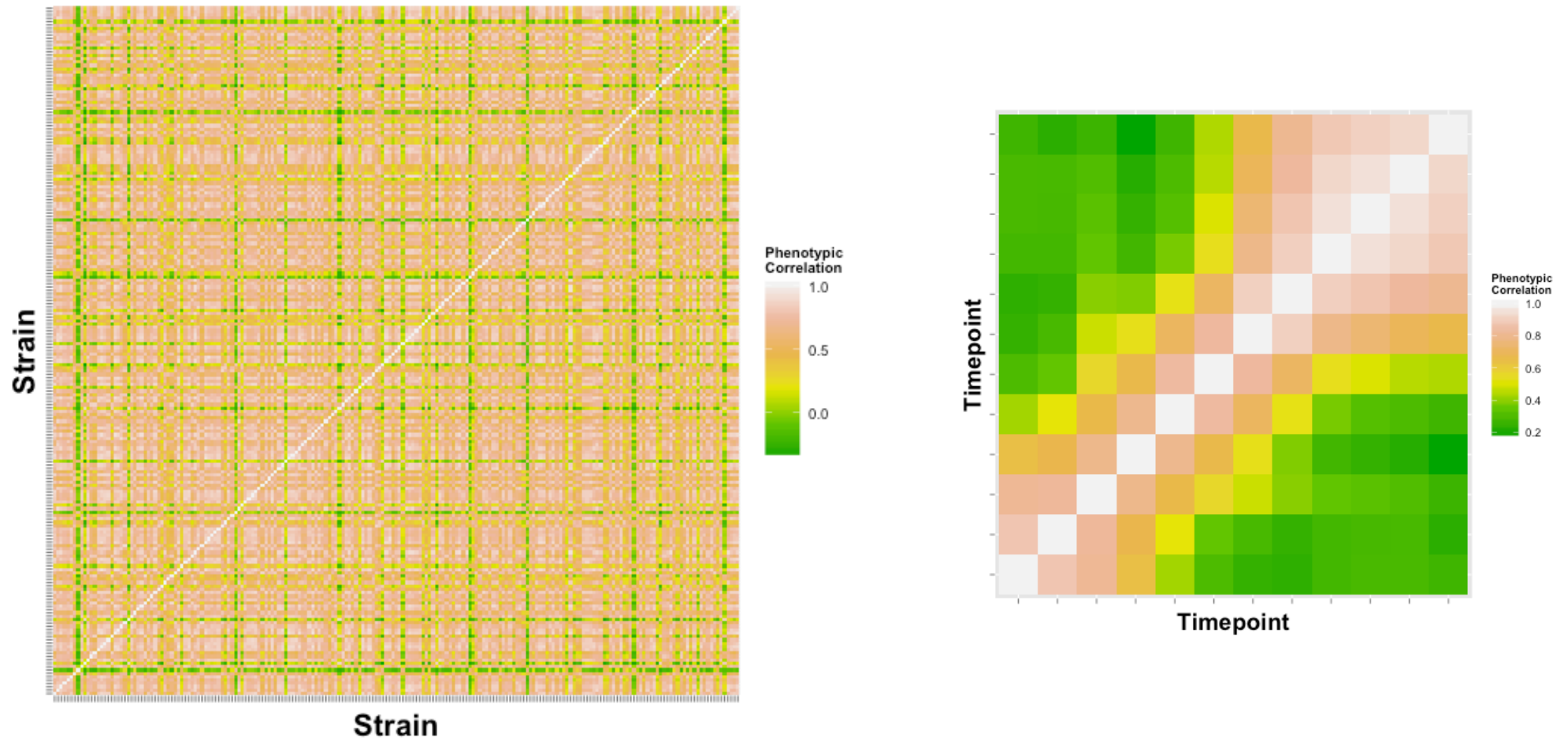


# Identification of genes involved in gamete production

clc-1 evl-20 fog-2 fog-3 hpl-1 inx-5 lin-25 pop-1  
spe-12 tag-175 vav-1 wrt-5 folt-1 rnf-121 yap-1 git-1 rog-1

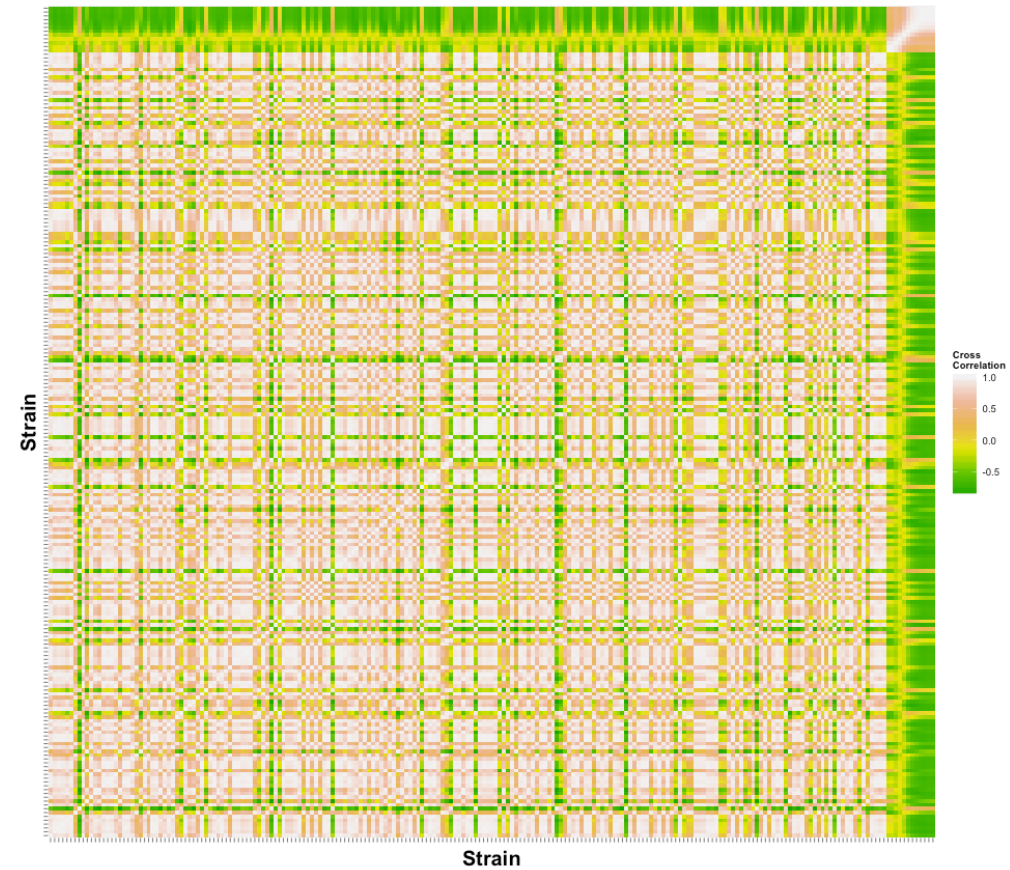
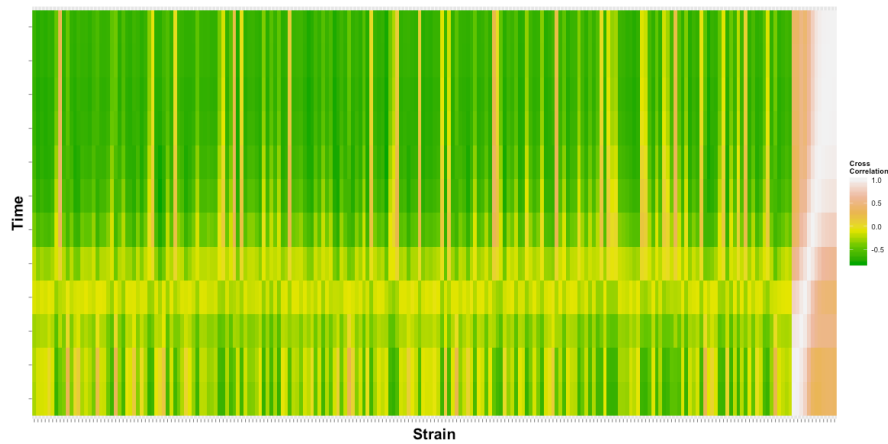


# Canonical Correlation Analysis



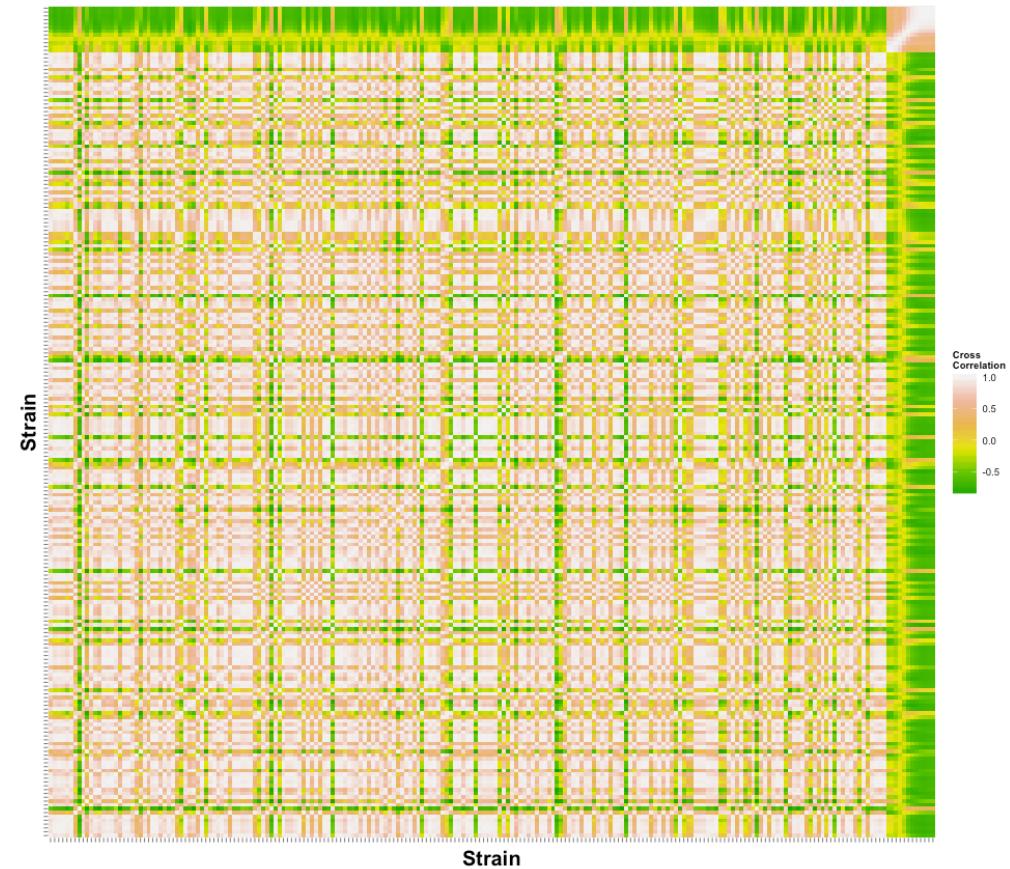
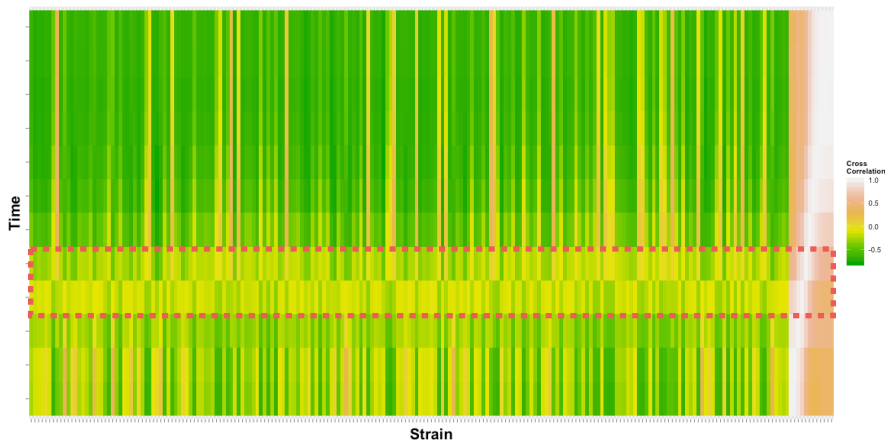
Phenotypic correlation of both expression sets

# Canonical Correlation Analysis



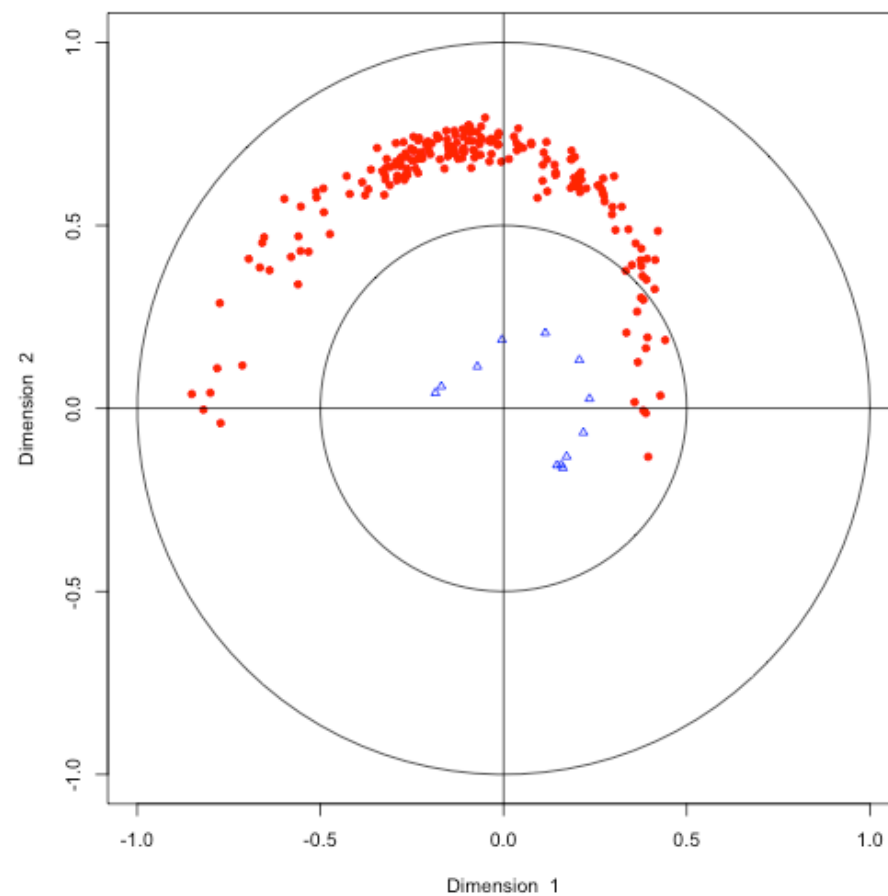
Cross correlation of expression sets

# Canonical Correlation Analysis



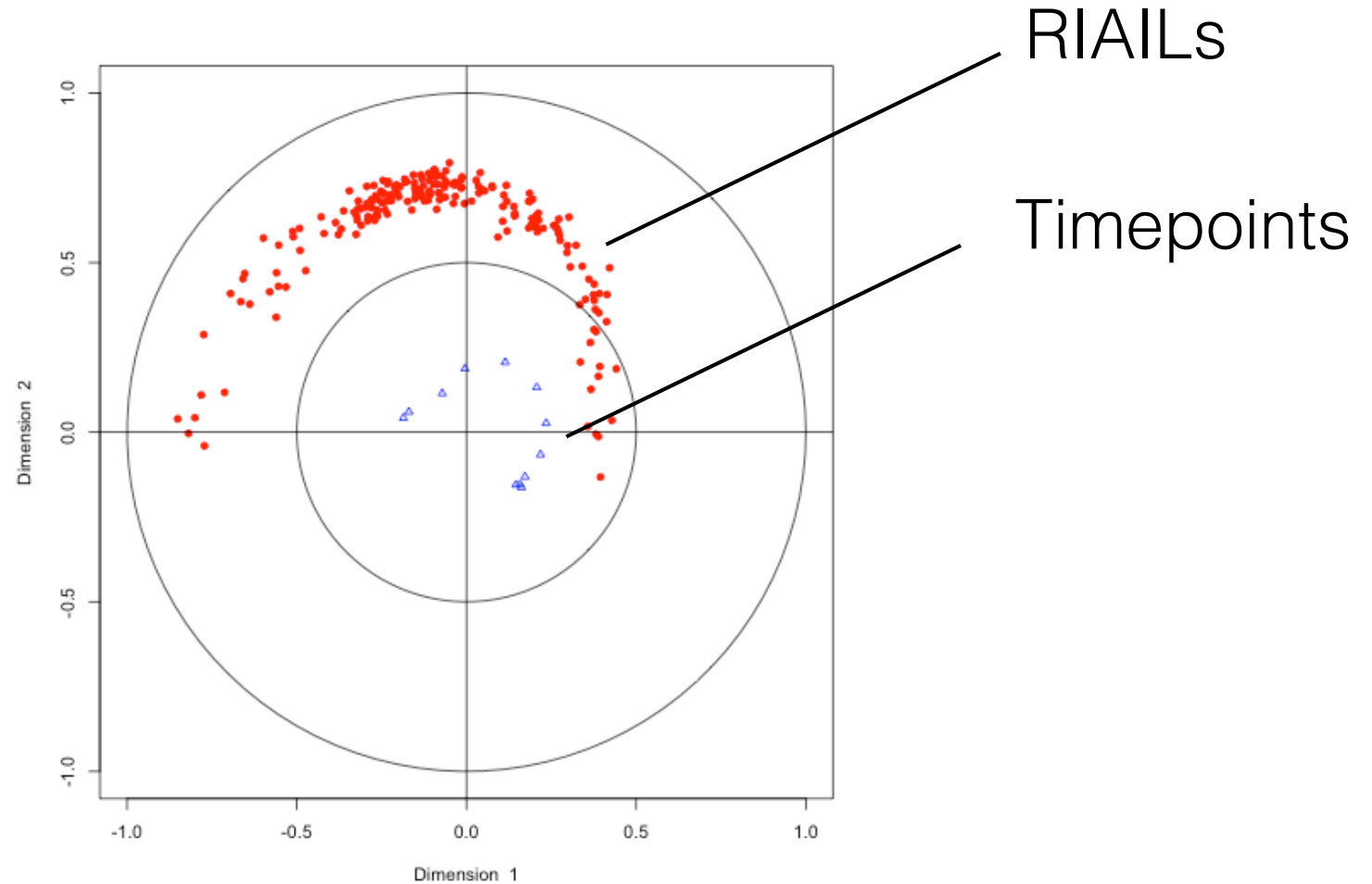
Identification of gene expression patterns in time course that are maximally correlated with gene expression in RIALs

# Canonical Correlation Analysis



Relative age of RIALs

# Canonical Correlation Analysis

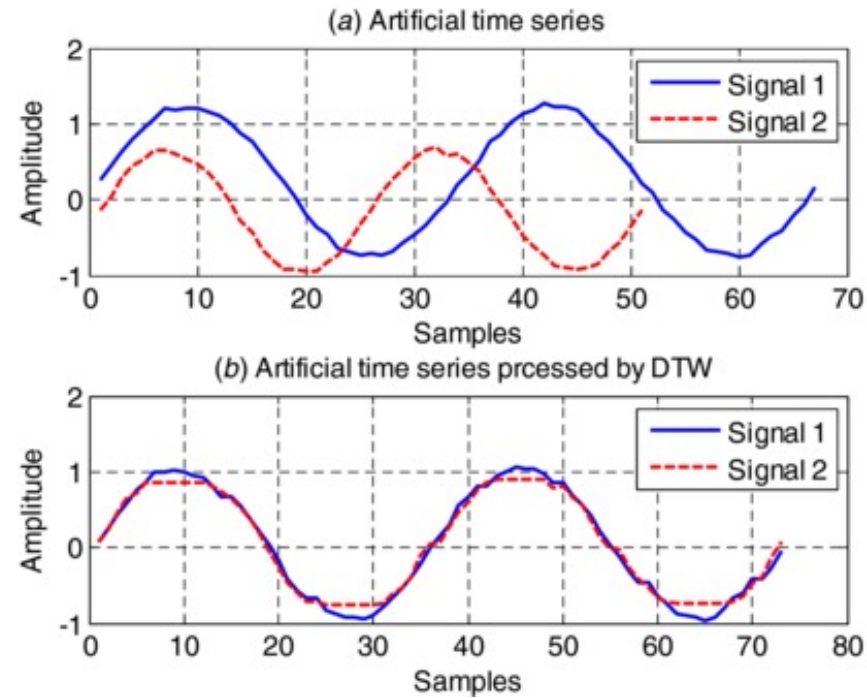


Relative age of RIALs

# Alignment of RIALs to developmental time course

## Dynamic Time Warping

Figure 1 from D Zhen et al 2012 Meas. Sci. Technol. 23 055601





# Alignment of RIALs to developmental time course

## Dynamic Time Warping

Figure 1 from D Zhen et al 2012 Meas. Sci. Technol. 23 055601

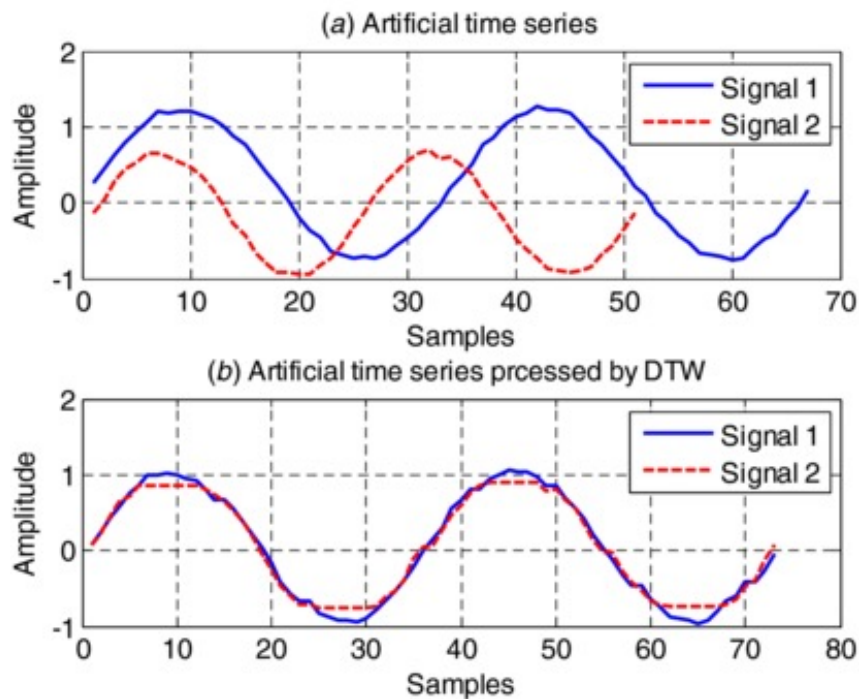
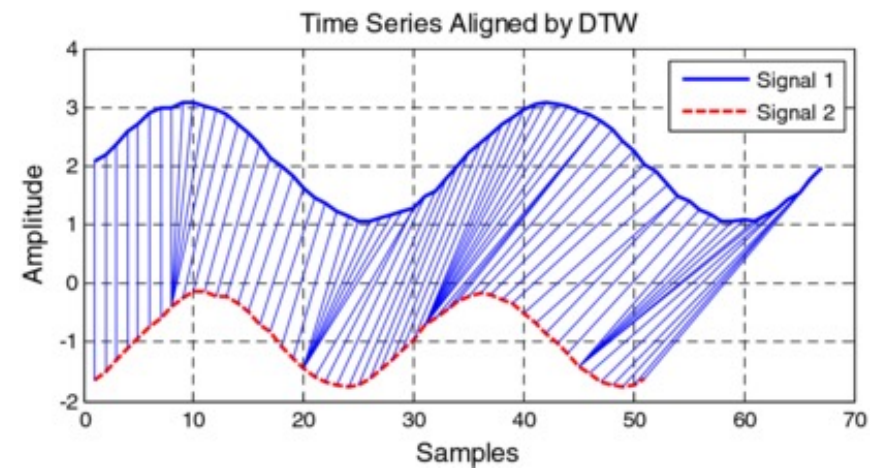
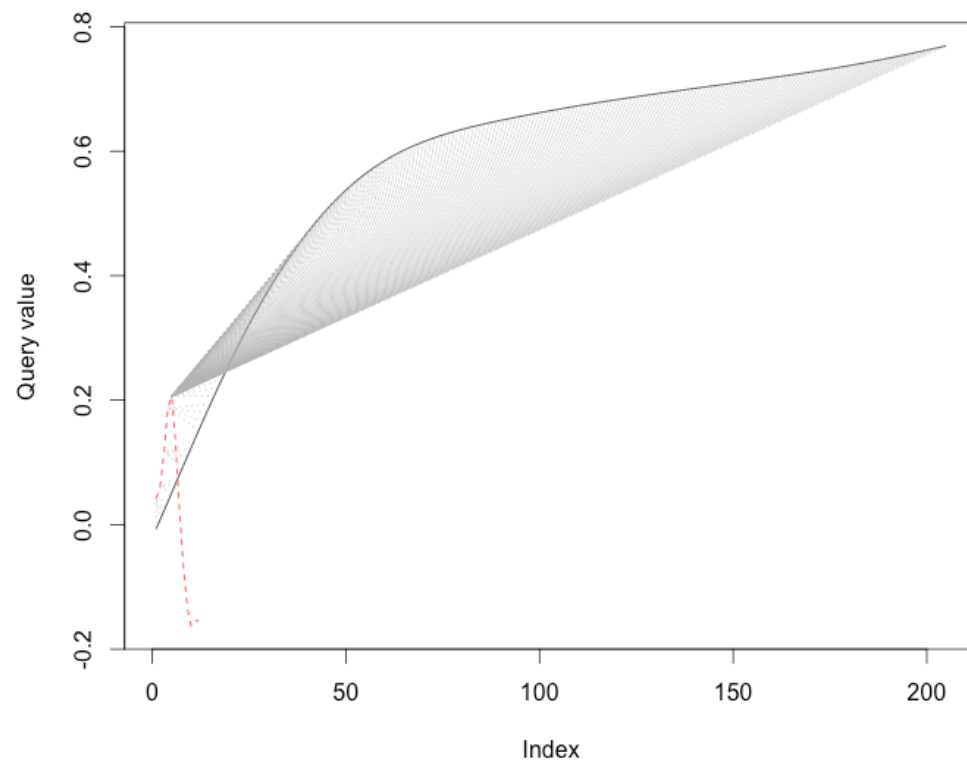
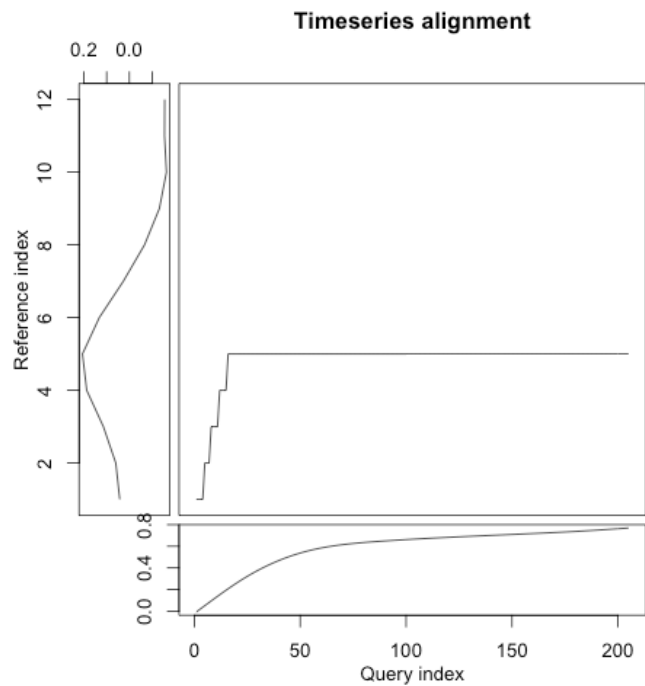


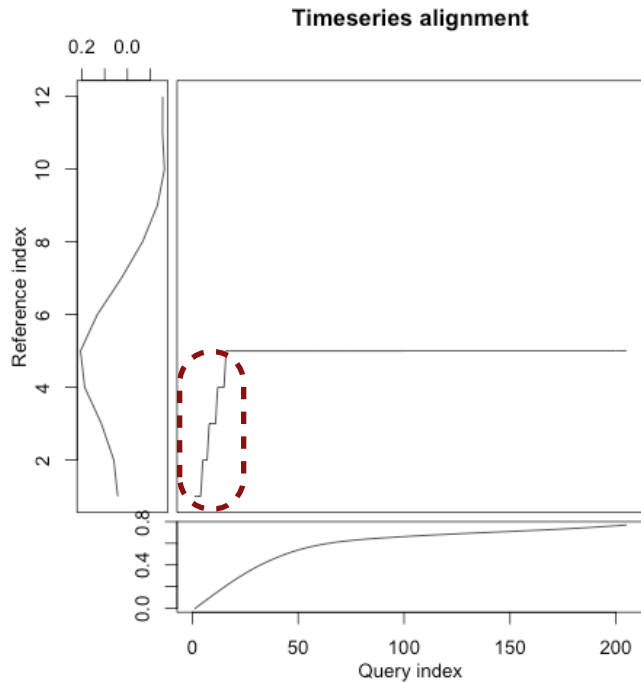
Figure 2 from D Zhen et al 2012 Meas. Sci. Technol. 23 055601



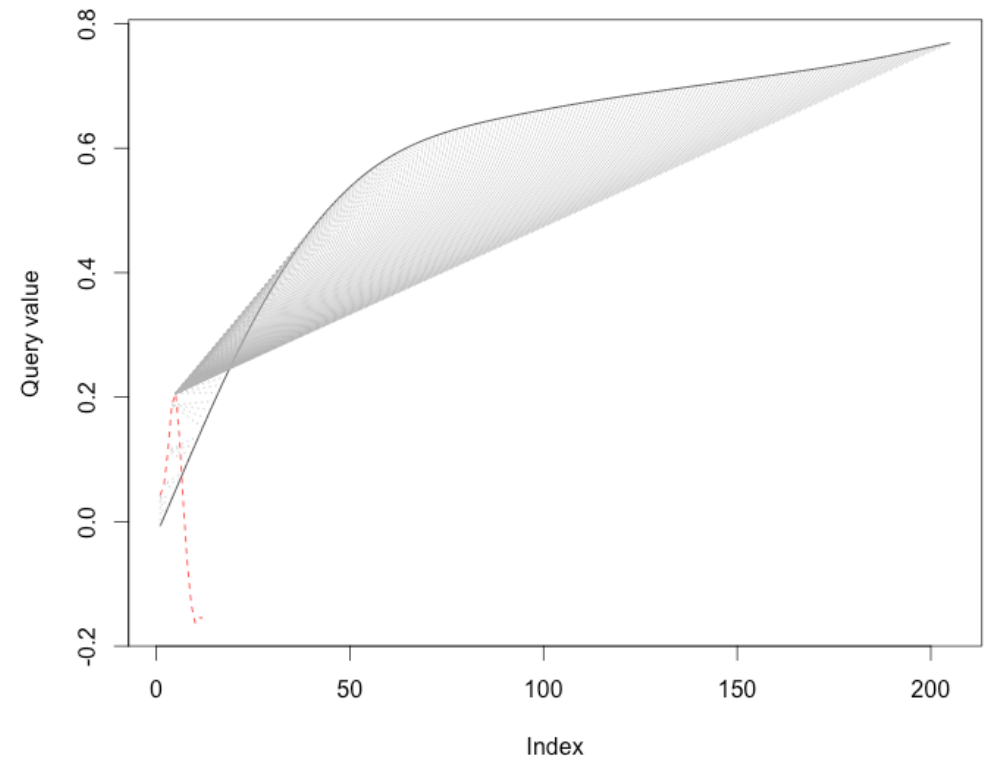
# Alignment of RIAILs to developmental time course



# Alignment of RIAILs to developmental time course

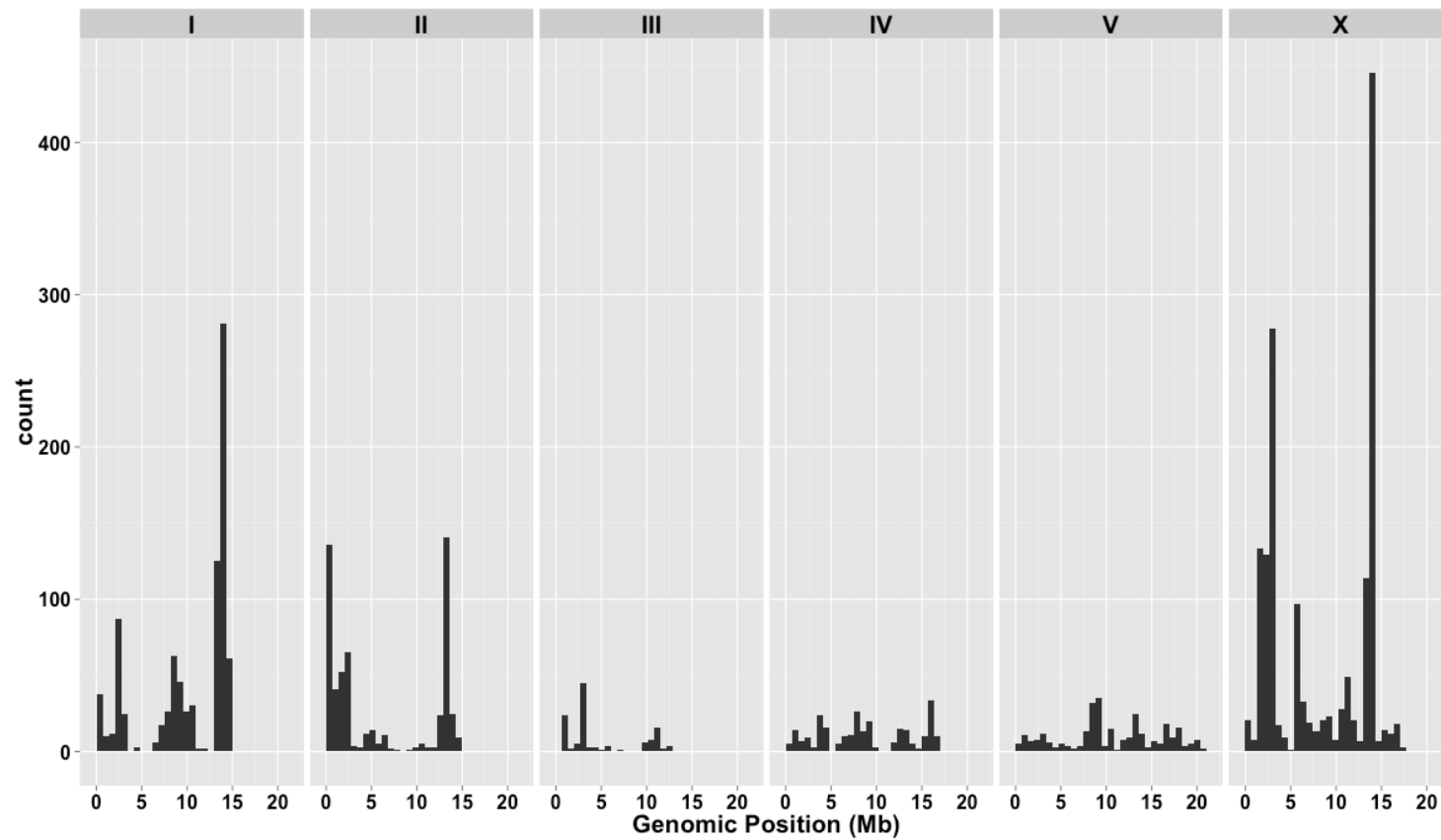


15 strains staged younger than L4

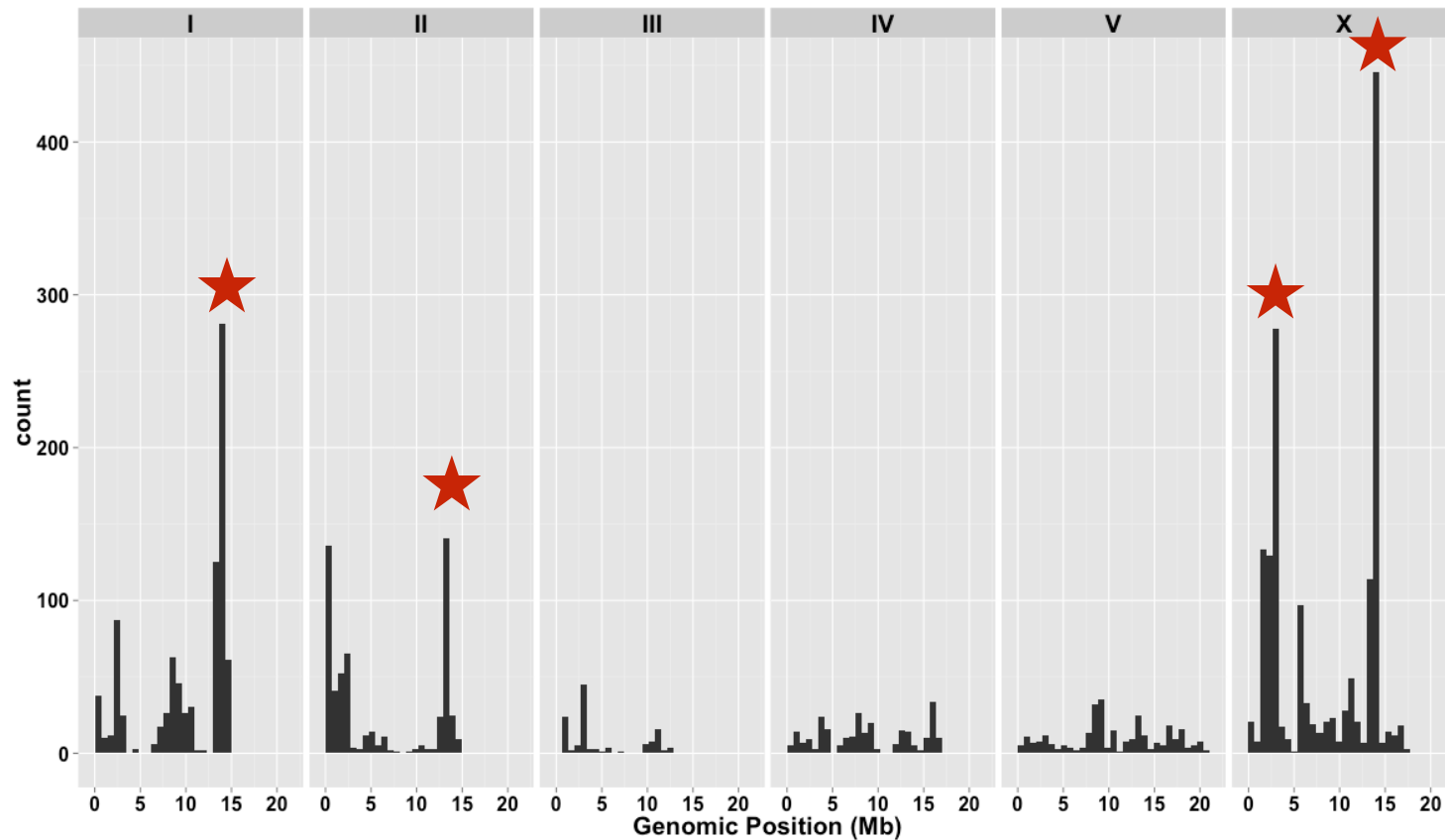


A majority of RIAILs appear to be L4s

# Linkage mapping with “developmental stage” as a covariate

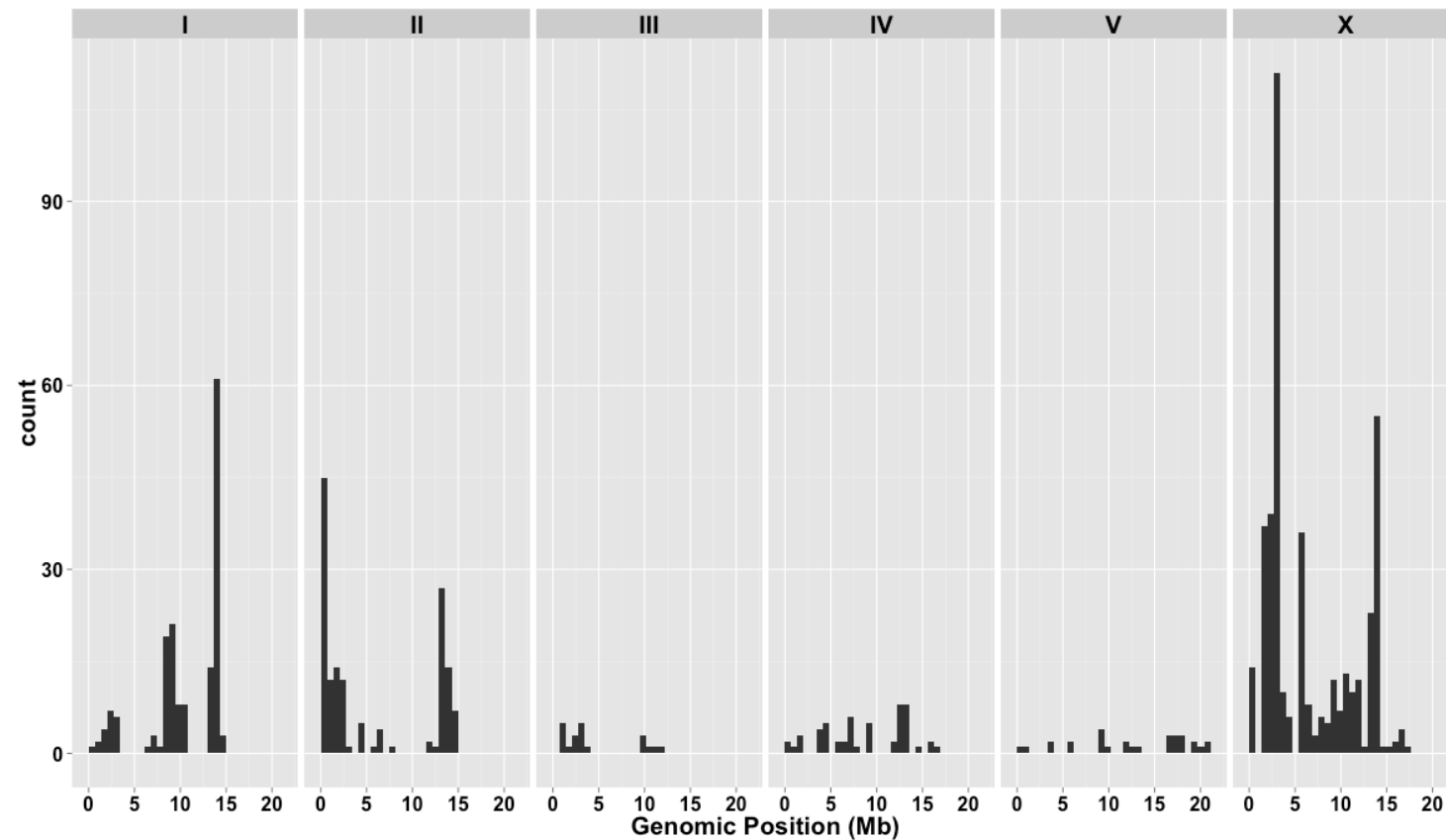


# Linkage mapping with “developmental stage” as a covariate



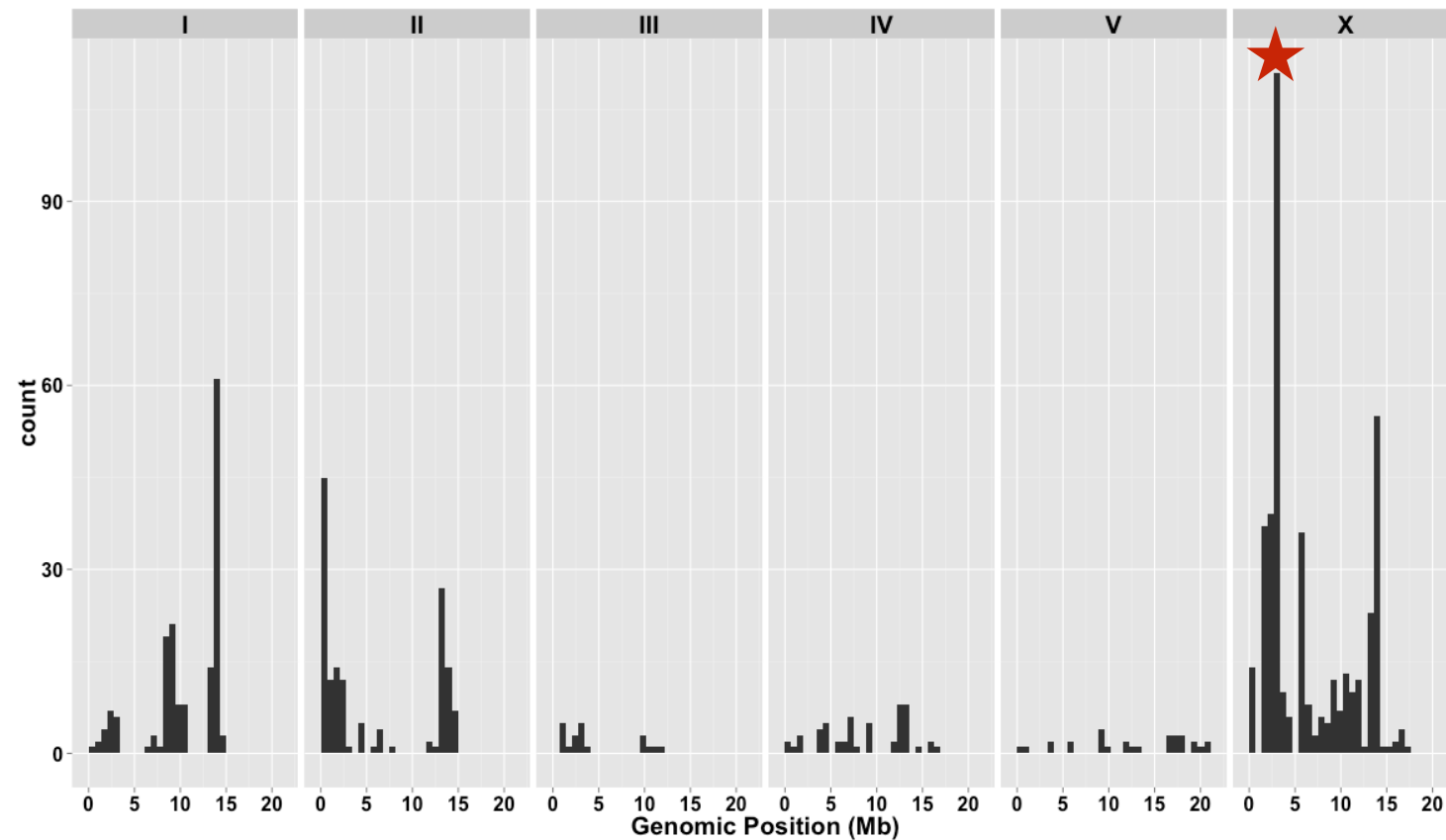
The same hotspots appear but regulate more genes

# Linkage mapping with “developmental stage” as a covariate



825 unique eQTL appear when time is a covariate

# Linkage mapping with “developmental stage” as a covariate



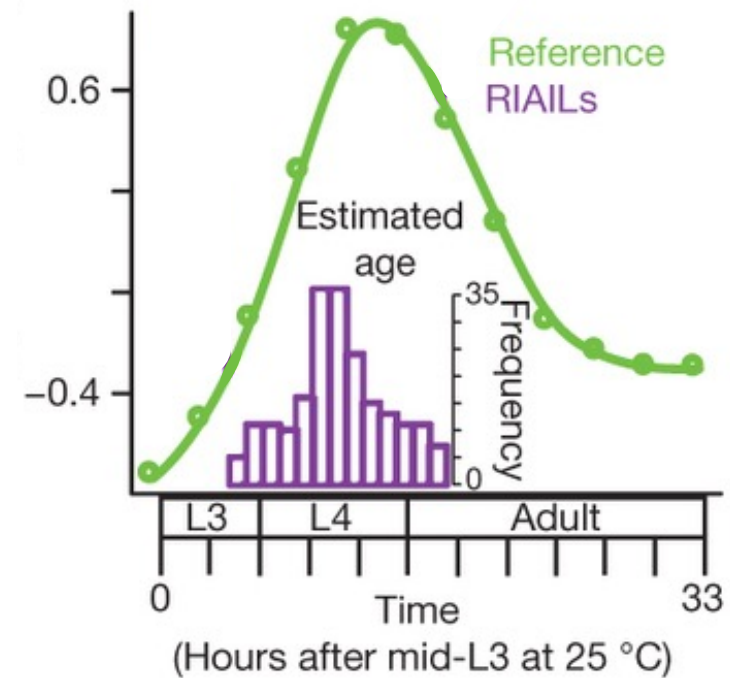
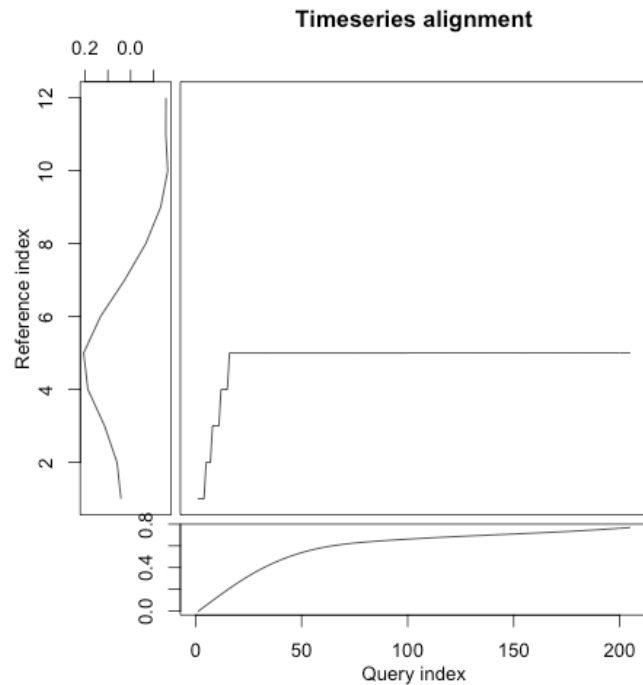
Major hotspot overlaps *npr-1*, which encodes for a GPCR responsible for many phenotypic differences between CB4856 and N2



# Identification of more genes associated with gamete production

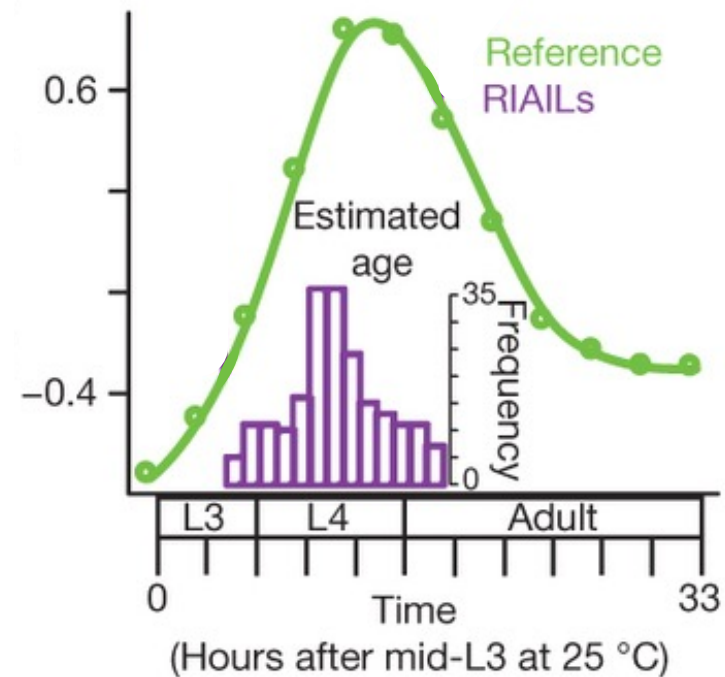
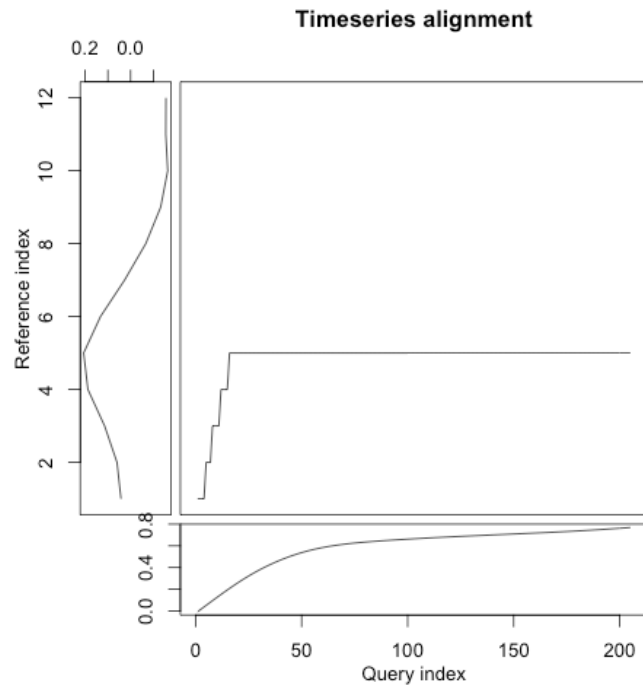
adm-2    gbh-2    hsp-12.3    mdl-1    nhr-111    pes-7    raga-1    vpr-1

# Discrepancy between publication



The author's identified RIAILs that grow faster than wild type strains

# Discrepancy between publication

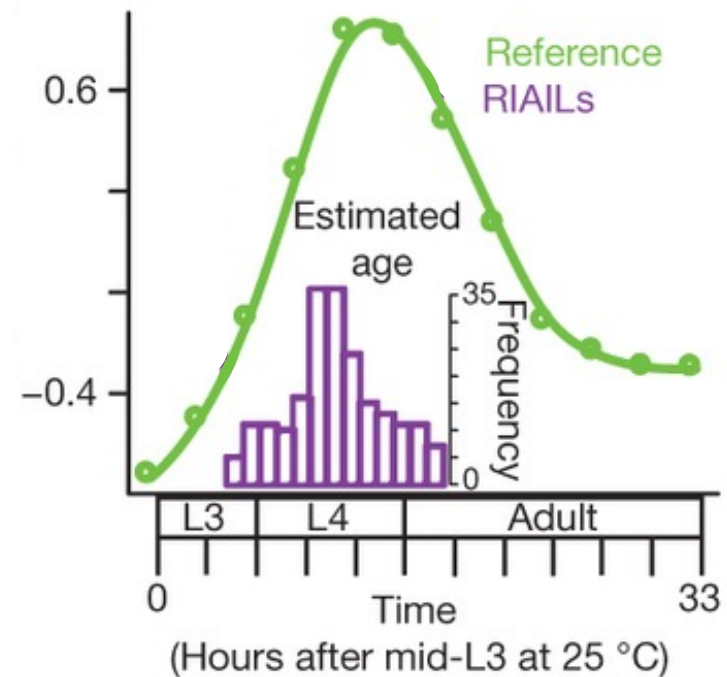
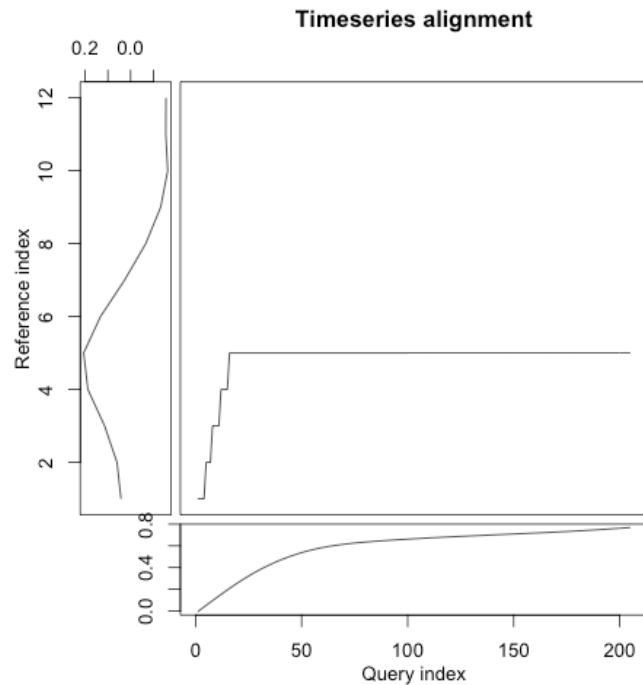


The author's identified RIAILs that grow faster than wild type strains

A possible reason for this -

I used 3000 fewer genes in my analysis. Though this might suggest that this type of analysis might not be valid. (e.g. randomly subsetting genes leads to drastically different results)

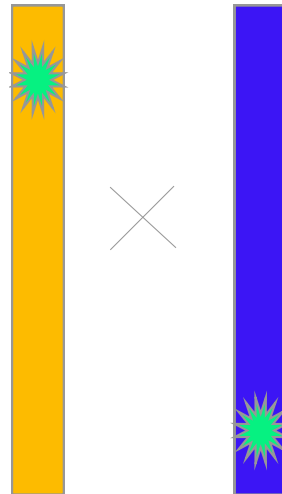
# Discrepancy between publication



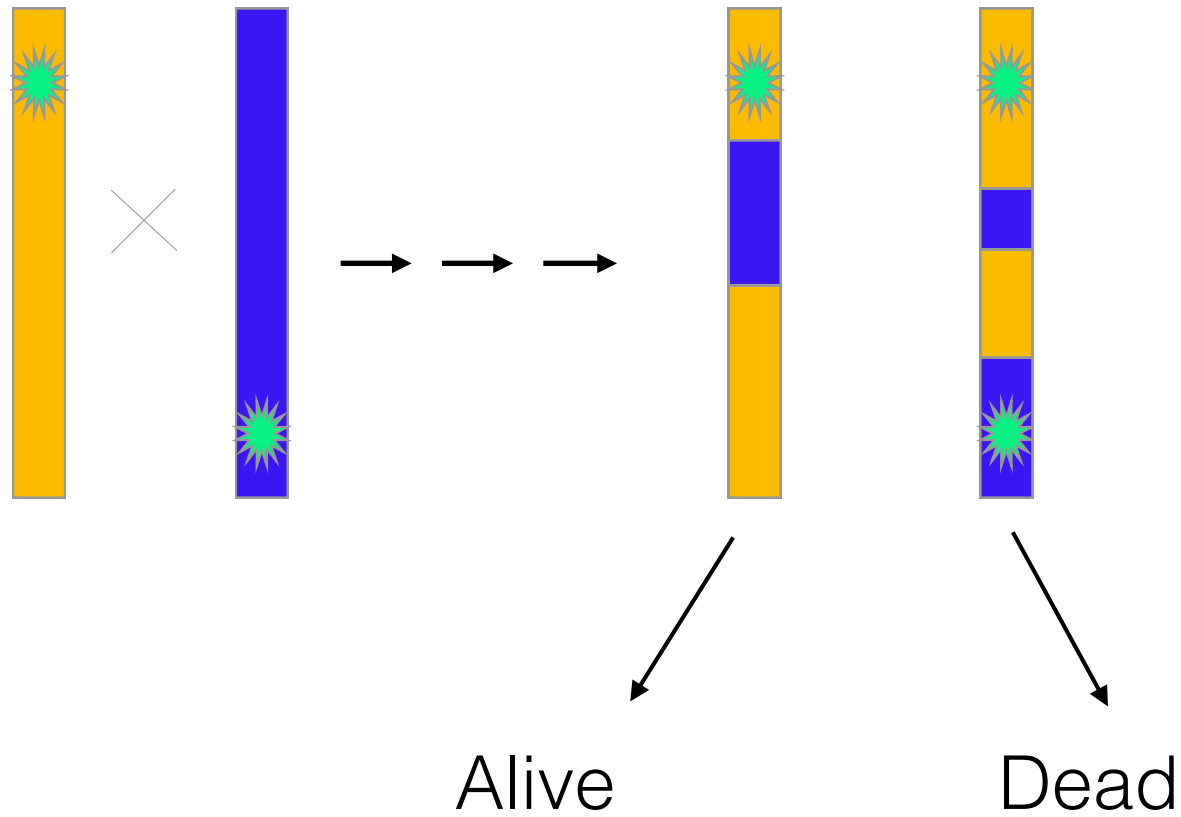
The author's identified RIAILs that grow faster than wild type strains

However there are widespread Bateson-Dobzhansky-Muller incompatibilities present in the genomes of CB4856 and N2

# Bateson-Dobzhansky-Muller incompatibilities

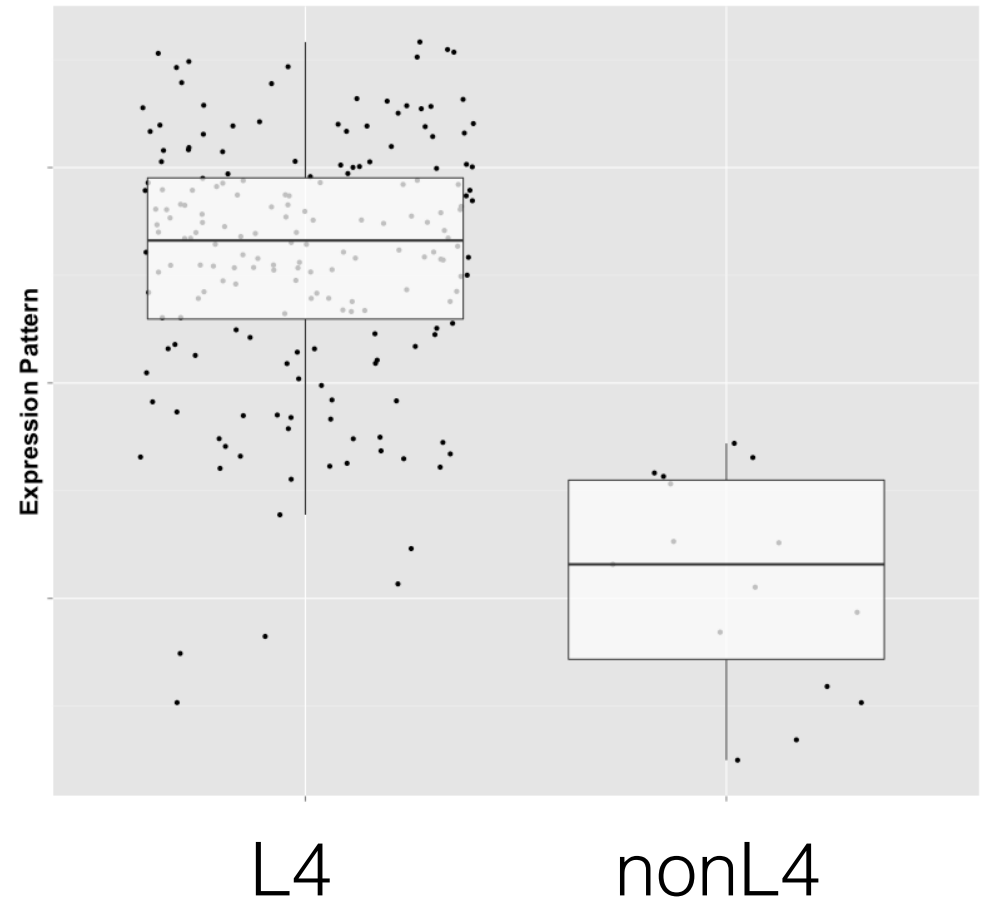
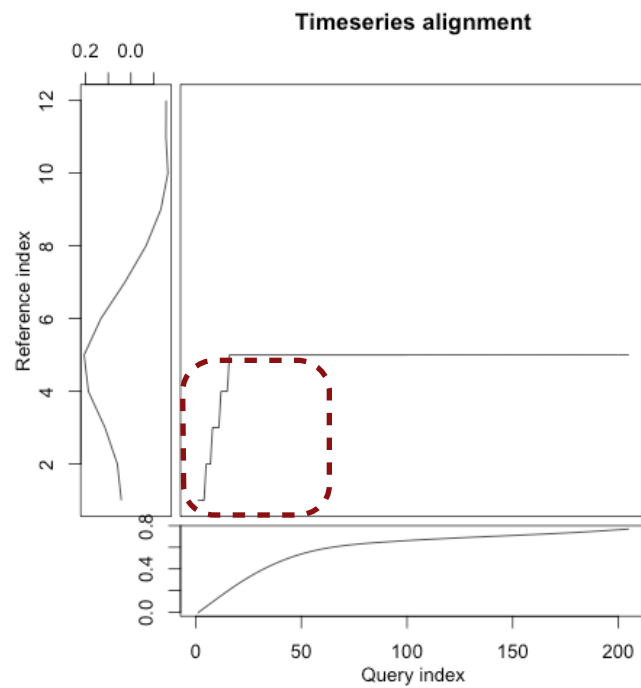


# Bateson-Dobzhansky-Muller incompatibilities

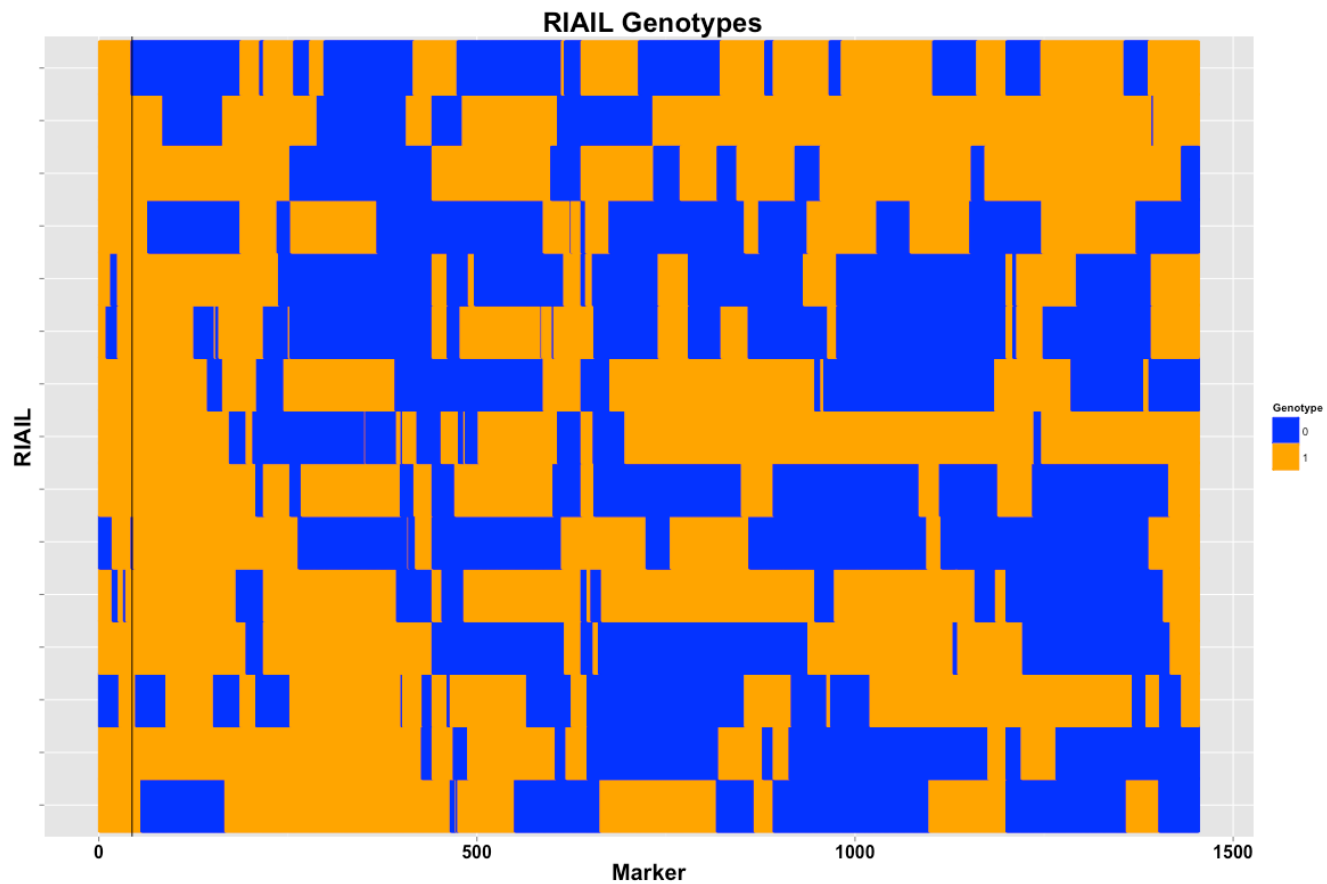


An extreme example. More subtle examples might include a reduction in fitness

# Distinct phenotypic distribution of slow growing strains



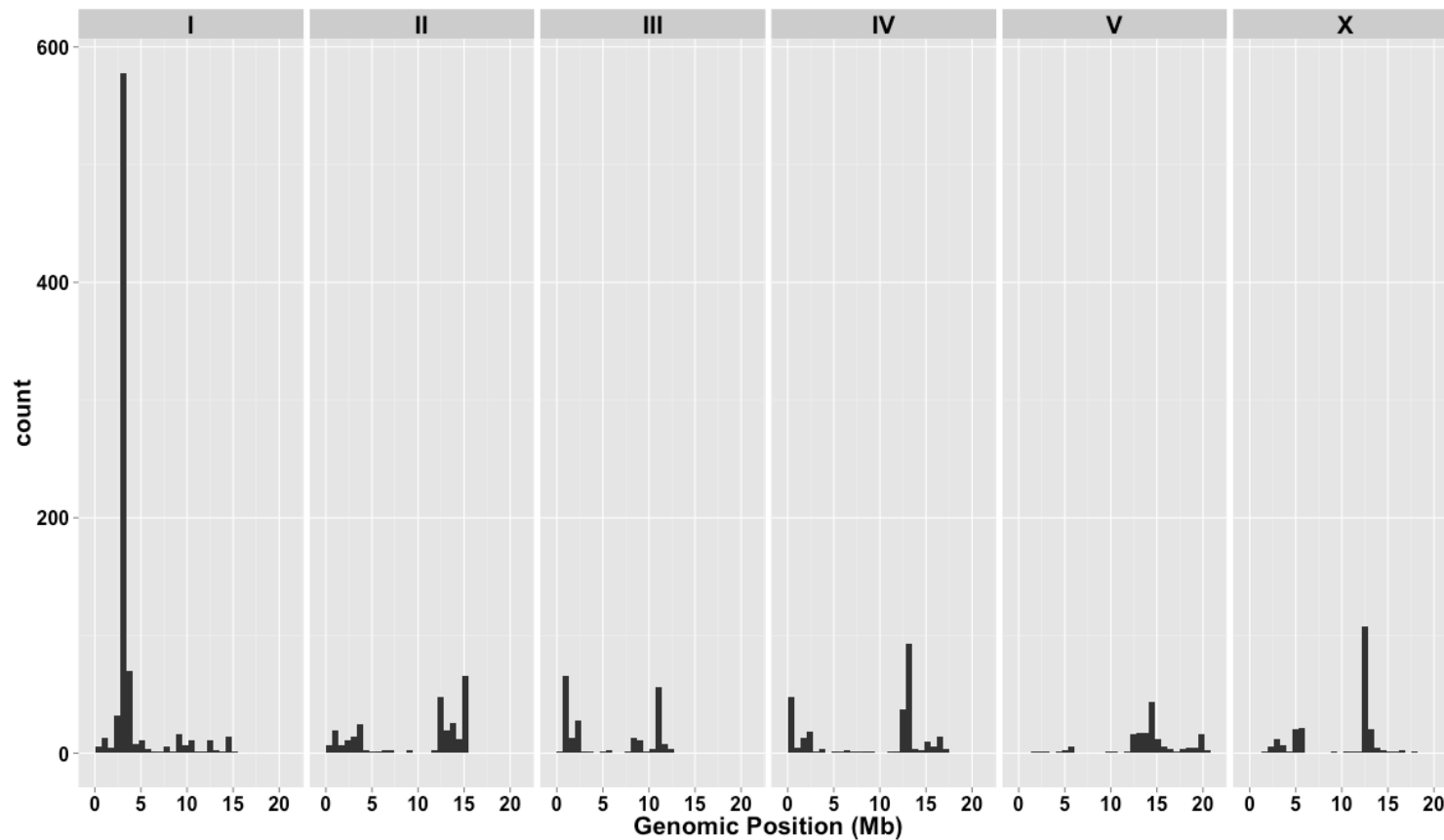
Slow growers share genotype



13/15 slow growers are CB4856



Adding slow growers as a covariate reveals an additional eQTL hotspot



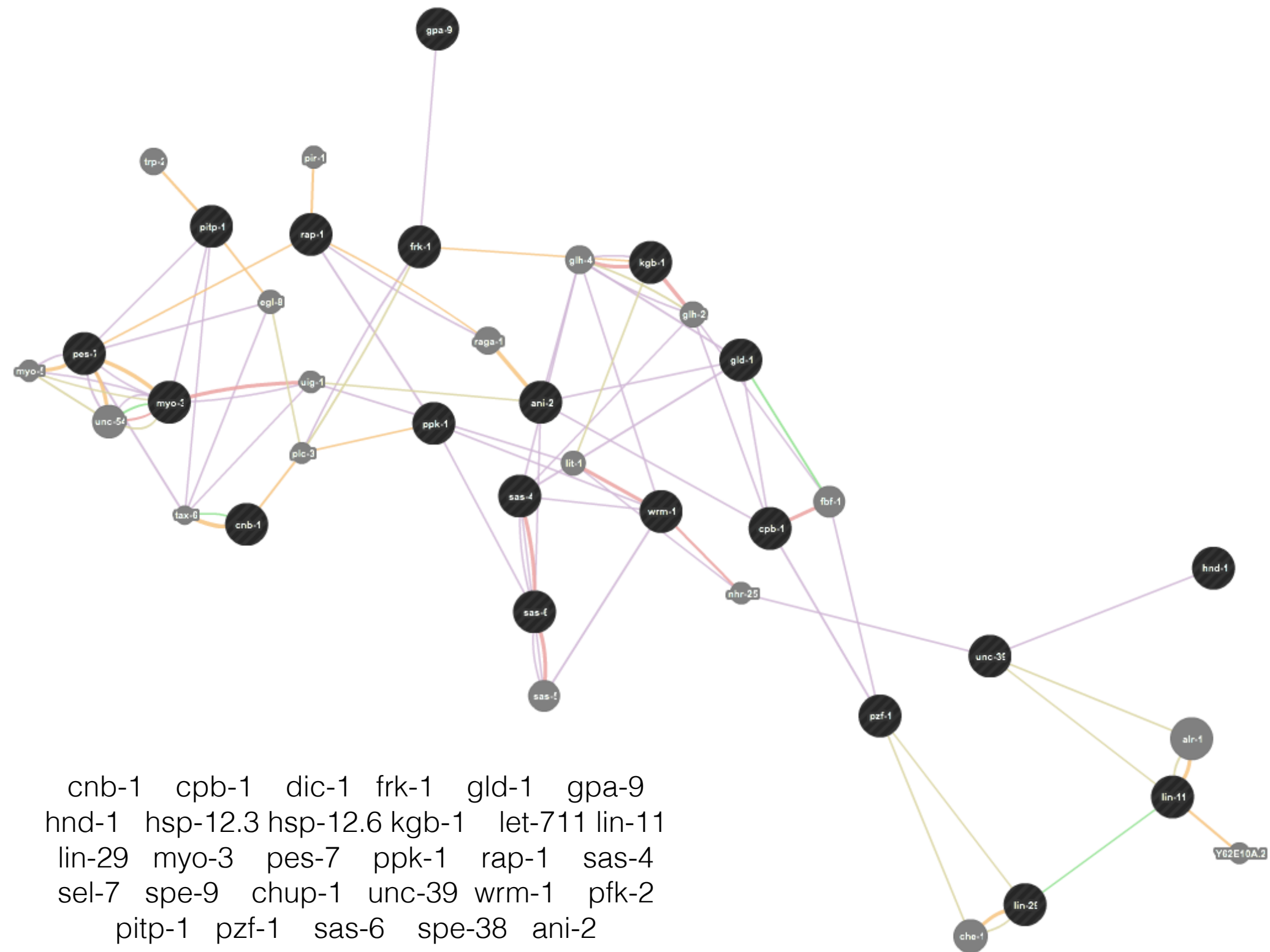
1700 unique eQTL

## GO annotations for top four gene classes

1. “integral to membrane”
2. “embryo development birth or egg hatching”
3. “reproduction”
4. “nematode larval development”

## Unique gamete production genes

cnb-1	cpb-1	dic-1	frk-1	gld-1	gpa-9	hnd-1	hsp-12.3	hsp-12.6
kgb-1	let-711	lin-11	lin-29	myo-3	pes-7	ppk-1	rap-1	sas-4
sel-7	spe-9	chup-1	unc-39	wrm-1	pfk-2	pitp-1	pzf-1	sas-6
			spe-38	ani-2				



# Conclusions

1. Staging RIALs to a reference developmental time course uncovers many more eQTL

# Conclusions

1. Staging RIAILs to a reference developmental time course uncovers many more eQTL

Using N2 as the reference developmental time course was not the best choice because RIAILs only share half of their genome with N2

A time course of CB4856 and N2 would have been interesting to compare, but I could not find a good data set.

# Conclusions

1. Staging RIALs to a reference developmental time course uncovers many more eQTL
2. Adding covariates to analysis may yield meaningful results, but still need to be confirmed with experiments.

# Conclusions

1. Staging RIALs to a reference developmental time course uncovers many more eQTL
2. Adding covariates to analysis may yield meaningful results, but still need to be confirmed with experiments.
3. eQTL mapping reveals already described gene interaction networks and may yield insights into yet unknown interactions

The end