

# *R/xqtl: High Throughput Modeling, Mapping and Exploration of Big Data*



- **Danny Arends** - Groningen Bioinformatics Center, University of Groningen
- **Pjotr Prins** - Department of Nematology, Wageningen University
- **Karl Broman** - Medical Informatics, University of Wisconsin-Madison



# DOGMA:

DN → carrying information

→ supporting information

→ factor molecules

→ Goal or function

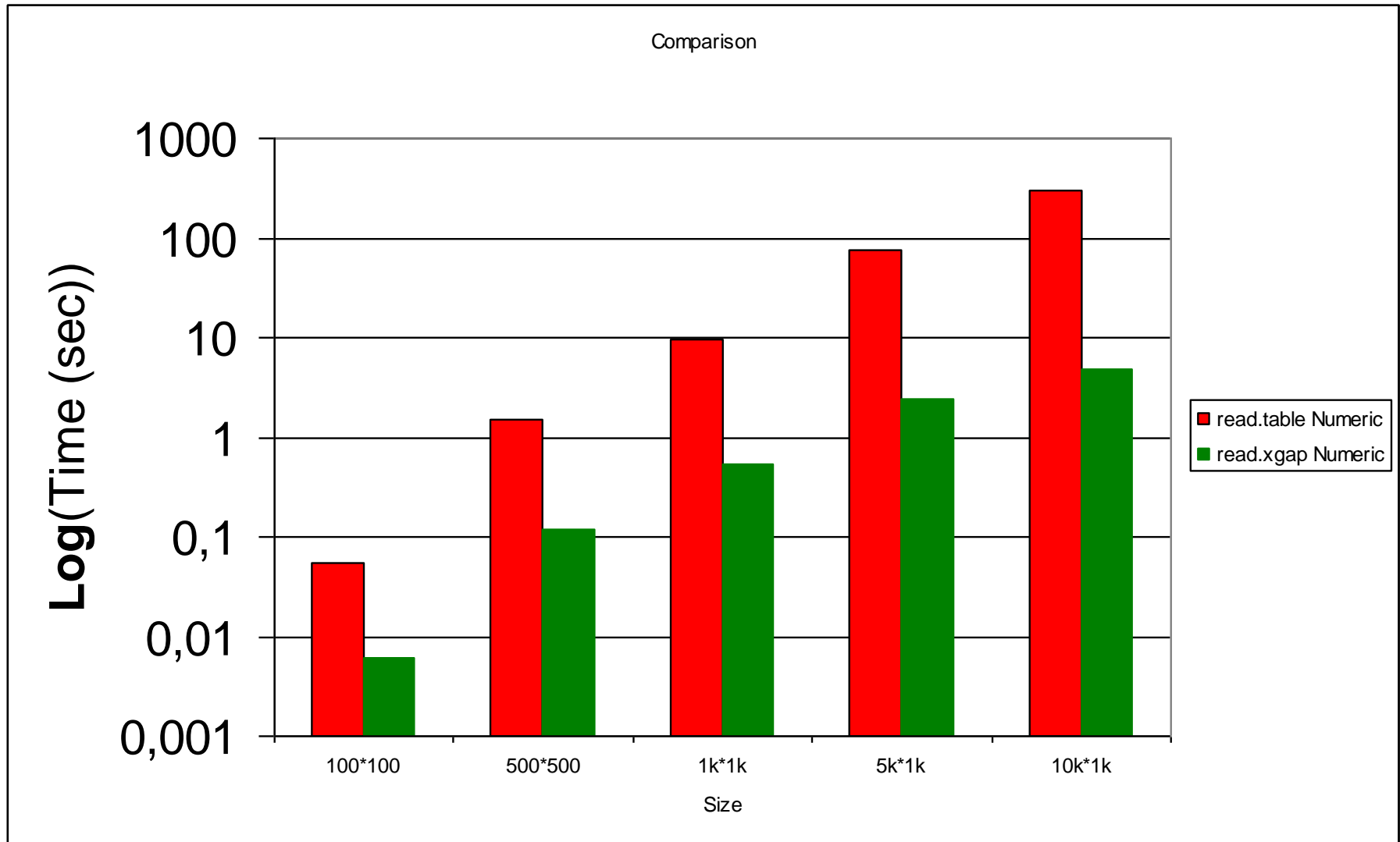
*extensible OR modular*

However this is not a

So how do we model what we know and room for  
anything we might forget?



# Storing this data



# DOGMA:

DNA -> Storing information

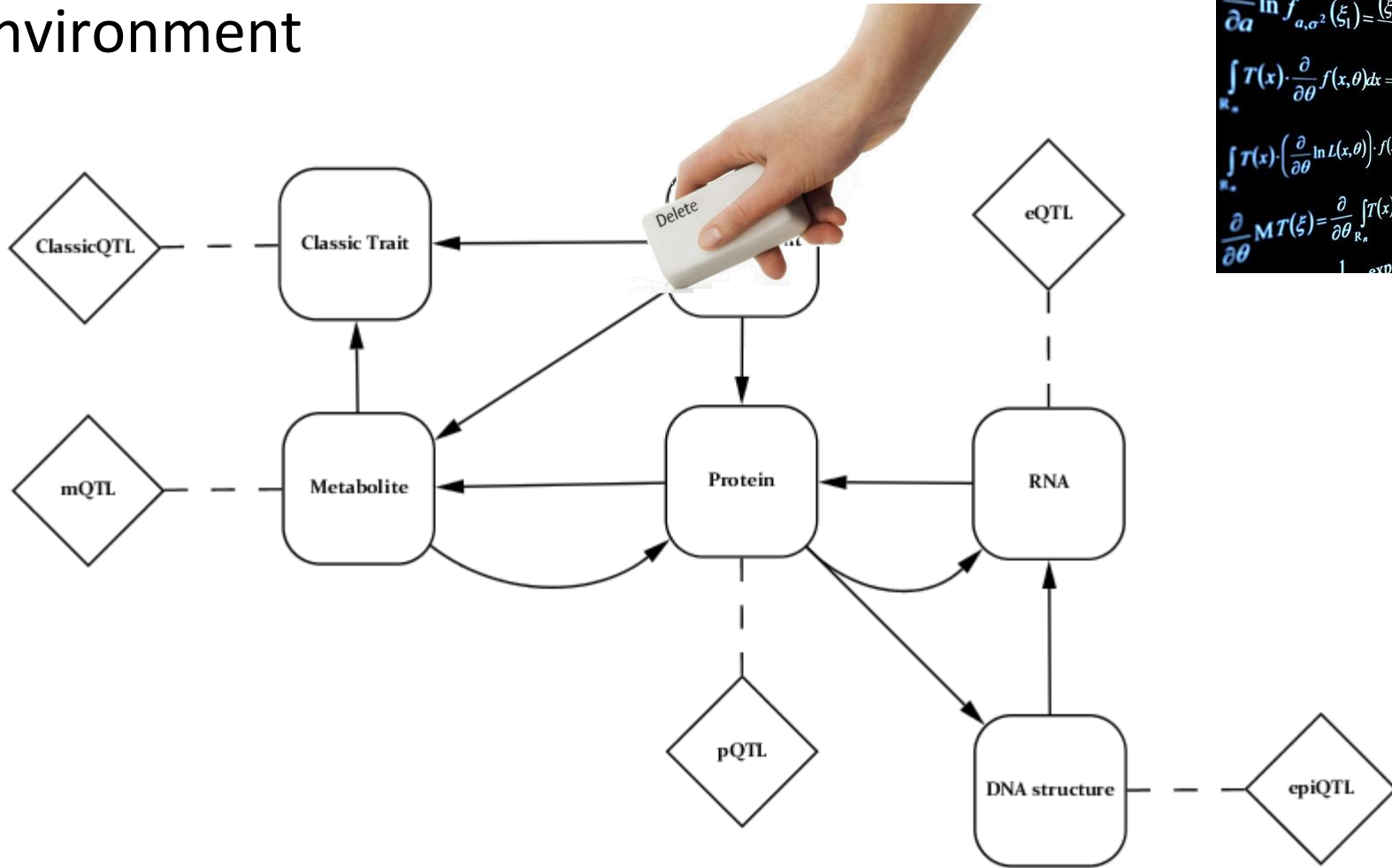
RNA -> Transporting information

PROTEIN -> Effector molecules

METABOLITE -> Goal or function

# Modeling: Quantitative traits

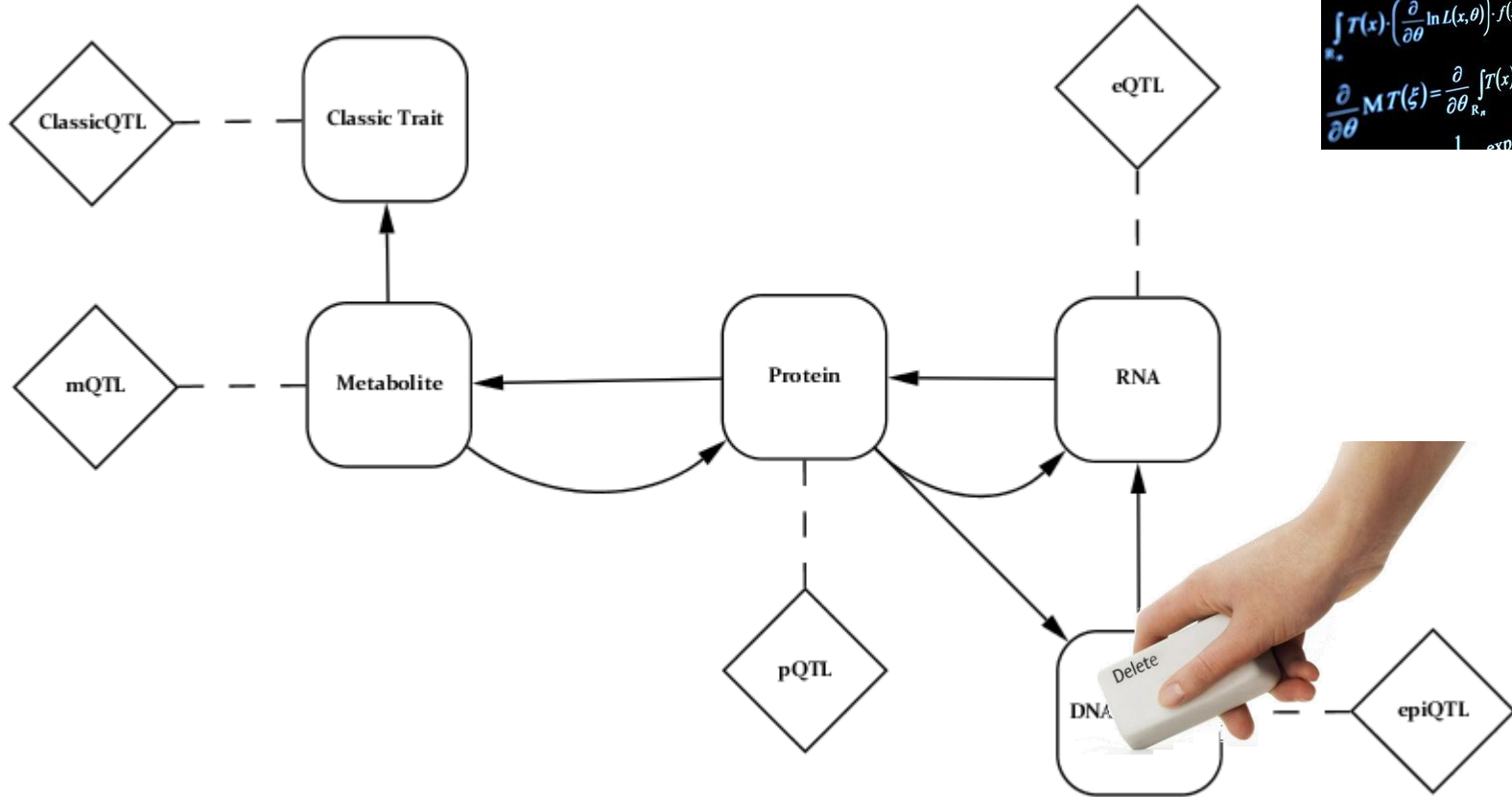
Environment



$$\frac{\partial}{\partial \theta} \ln f_{a, \sigma^2}(\xi_1) = \frac{(\xi_1 - a)}{\sigma^2} f_{a, \sigma^2}(\xi_1) - \frac{1}{\sqrt{2\pi\sigma^2}} \left\{ \frac{\xi_1 - a}{\sigma^2} \right\}$$
$$\int_{\mathbb{R}^n} T(x) \cdot \frac{\partial}{\partial \theta} f(x, \theta) dx = M \left( T(\xi) \cdot \frac{\partial}{\partial \theta} \ln L(\xi, \theta) \right)$$
$$\int_{\mathbb{R}^n} T(x) \cdot \left( \frac{\partial}{\partial \theta} \ln L(x, \theta) \right) \cdot f(x, \theta) dx = \int_{\mathbb{R}^n} \left( \frac{\partial}{\partial \theta} \ln L(x, \theta) \right) f(x, \theta) dx$$
$$\frac{\partial}{\partial \theta} \ln L(x, \theta) = \frac{\partial}{\partial \theta} \left[ -\frac{1}{2\sigma^2} (x - a)^2 \right] = -\frac{(x - a)}{\sigma^2}$$

# Reduced Model

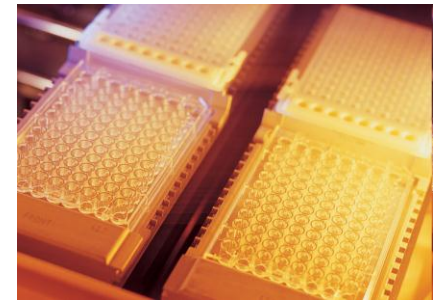
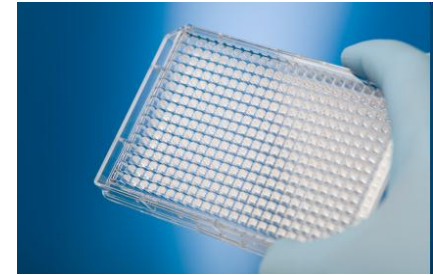
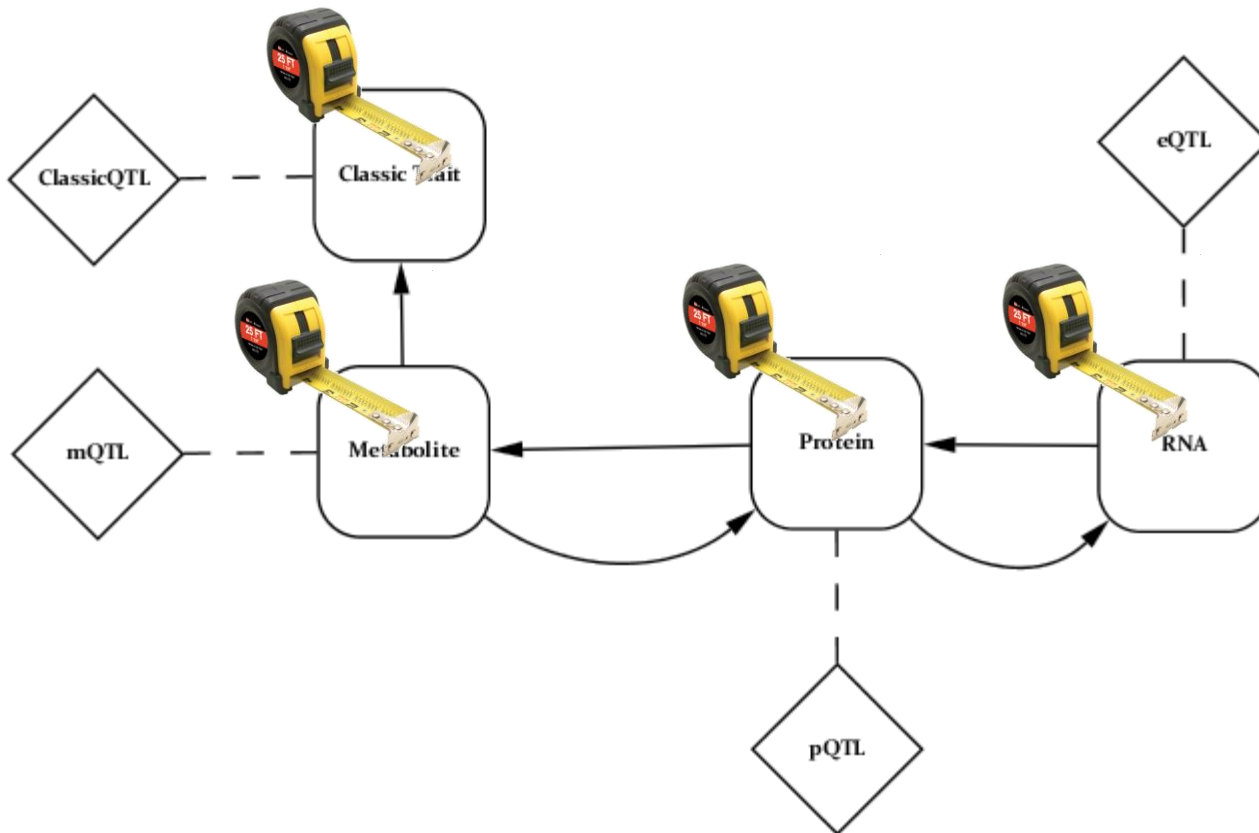
## DNA structure



$$\frac{\partial}{\partial \theta} \ln f_{a, \sigma^2}(\xi) = \frac{(\xi - a)}{\sigma^2} f_{a, \sigma^2}(\xi) - \frac{1}{\sqrt{2\pi\sigma^2}} \left[ \frac{\xi - a}{\sigma^2} \right]$$
$$\int_{\mathbb{R}^n} T(x) \cdot \frac{\partial}{\partial \theta} f(x, \theta) dx = M \left( T(\xi) \cdot \frac{\partial}{\partial \theta} \ln L(\xi, \theta) \right)$$
$$\int_{\mathbb{R}^n} T(x) \cdot \left( \frac{\partial}{\partial \theta} \ln L(x, \theta) \right) \cdot f(x, \theta) dx = \int_{\mathbb{R}^n} T(x) \cdot \left( \frac{\partial}{\partial \theta} \ln f(x, \theta) \right) \cdot f(x, \theta) dx$$
$$\frac{\partial}{\partial \theta} \ln f_{a, \sigma^2}(\xi) = \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} T(x) f(x, \theta) dx = \int_{\mathbb{R}^n} T(x) \cdot \frac{\partial}{\partial \theta} f(x, \theta) dx = \int_{\mathbb{R}^n} T(x) \cdot \left( \frac{\partial}{\partial \theta} \ln f_{a, \sigma^2}(\xi) \right) \cdot f_{a, \sigma^2}(\xi) dx$$

# Reduced Model

## Measure



# How to calculate

## Using R programming environment

- Statistical tools
- R/qtl (Karl Broman)
- SNOW (Tierney)

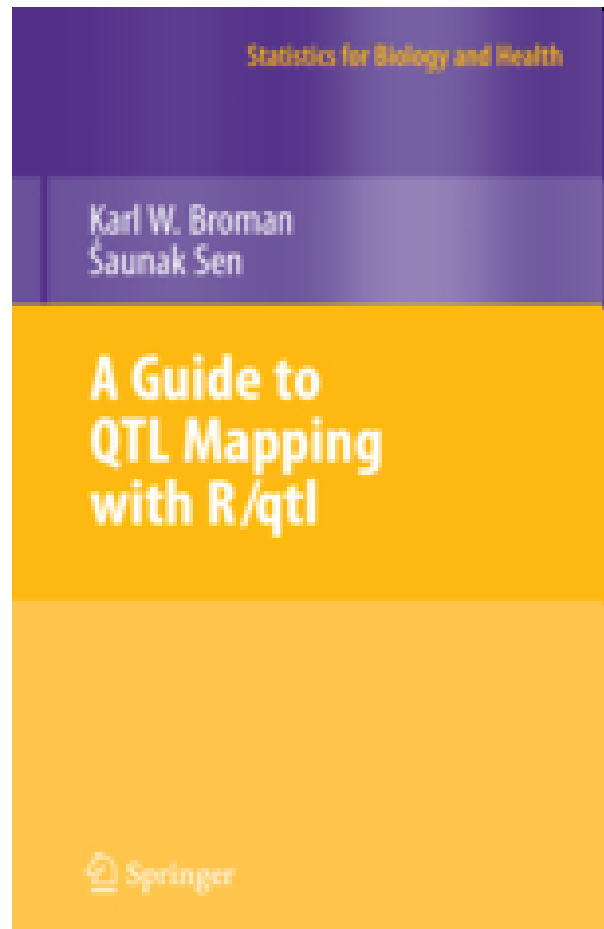
Computational effort: # traits and # permutations





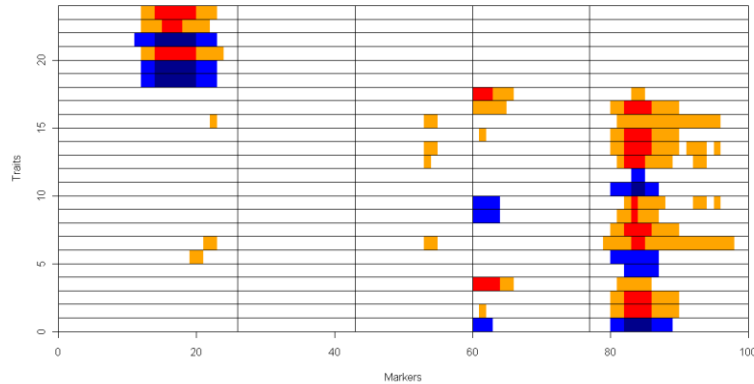
## Framework for QTL mapping in experimental crosses

Karl Broman, Saunak Sen, Gary Churchill, Danny Arends, Piotr Prins, Ritsert Jansen

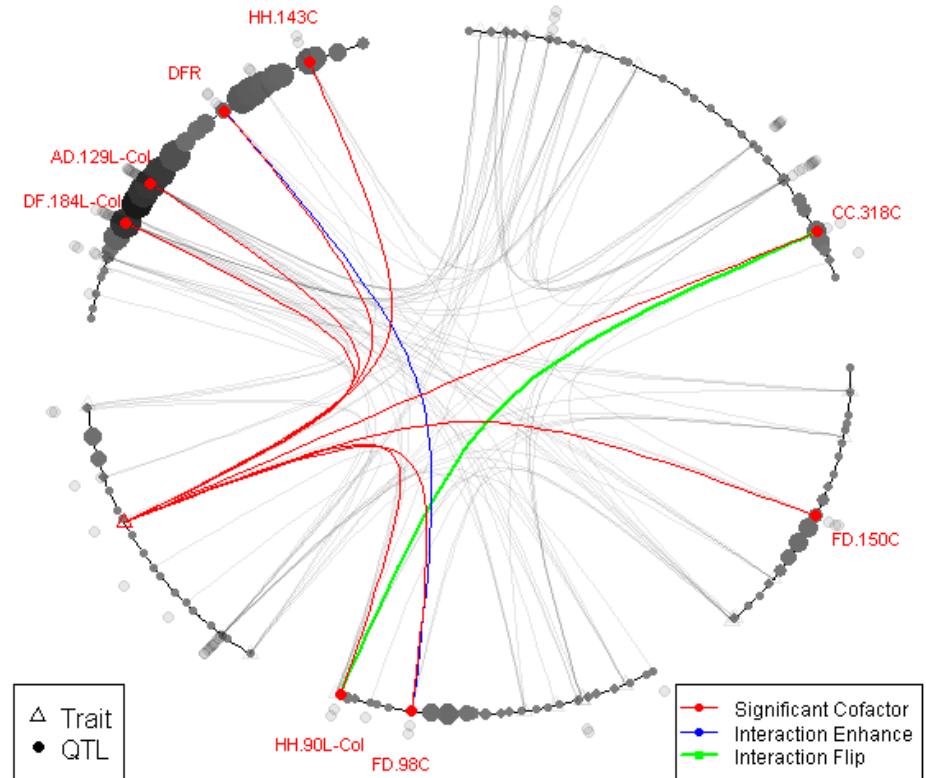


# Explore using visualizations

## Exploring using R



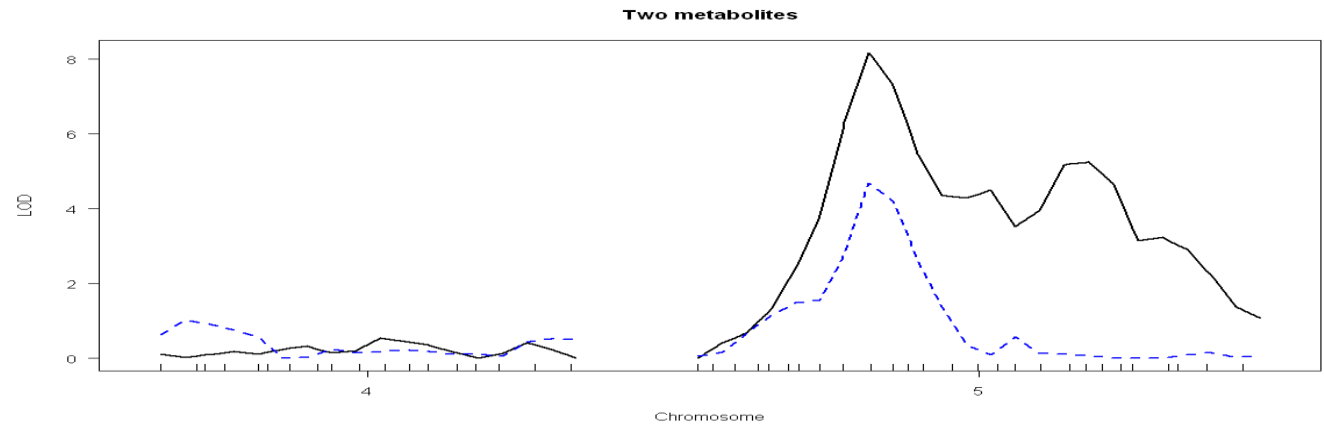
## Circular genome plot



eptyl

# Making sense out of it

## QTL analysis

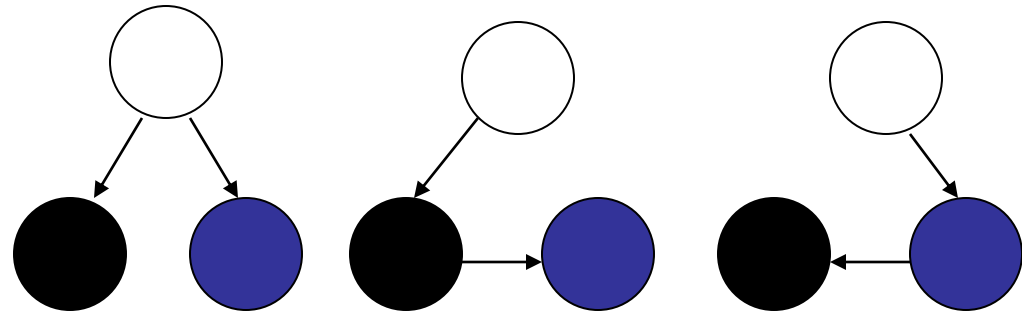


## Find regions affecting multiple traits

## Causality analysis

Within a phenotype level

Between phenotype levels



(not yet in R/qtl but will be added soon)

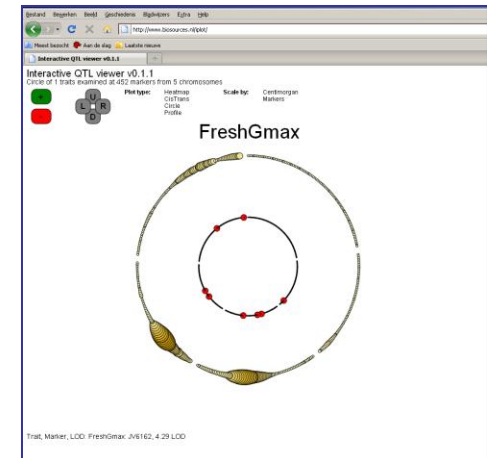
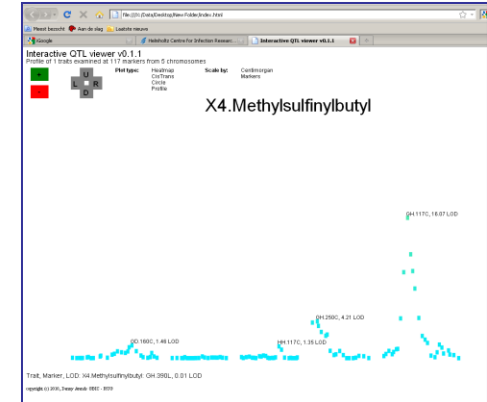
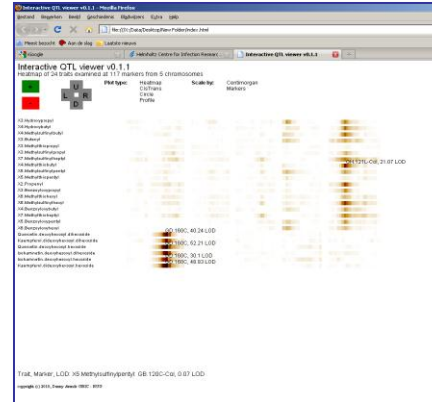


# Interactive plots

Showing it to the world:

Interactive plots

- Interactive SVG / VRML
  - Web [Click](#) and standalone
- Standalone java viewer ([Click](#))
  - E.g. High performance viewer



# Overview

## Our choices:

To Store Big Data	(Molgenis)
To Structure Big Data	(XGAP)
To Transport Big Data	(Binary)
Make sense of Big Data	(Desktop, Cluster, Cloud)
Extract knowledge from Big Data	(R and R/qtI)
To share this knowledge	(Interactive Visualizations)

# Questions ?

Thank you for your attention



university of  
groningen

Interactive SVG / VRML:  
[www.biosources.nl/iplot](http://www.biosources.nl/iplot)