eQTL mapping

Our plans for today

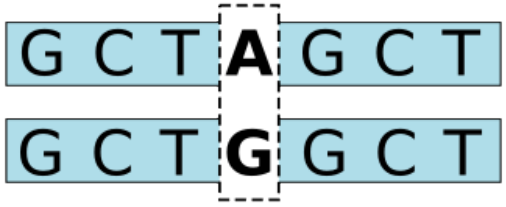
Learn the basics of eQTL analyses:

- Understand what an eQTL analysis is
- Revise some basics in genetics
- Understand how an eQTL analysis works
  - Step-by-step understand which data is needed
  - Understand what the data is means
  - Look at basic quality measures
- Look at a locus in more detail

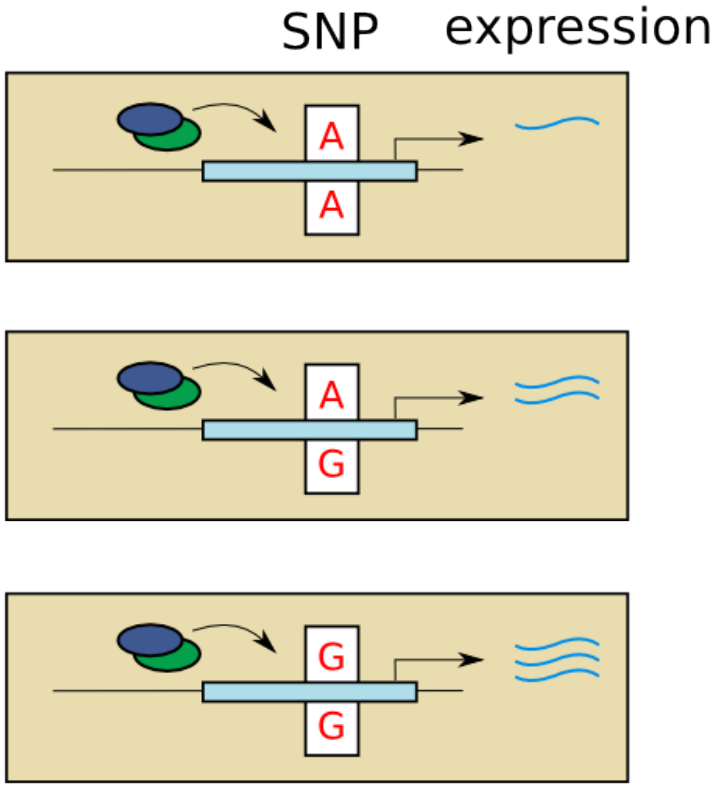
eQTL analysis

- mRNA levels are measured simultanously on thousands of genes using expression arrays or RNAseq
- Hundreds of thousands of SNPs are availble through genotyping arrays or genome sequencing
- The aim: Identify association of SNPs with molecular traits
- Advantage of molecular traits:
  - Quantifiable
  - Objectively measured
- Associated SNPs can then be overlapped with genome annotation.

eQTL analysis



Genotyping



Differences to other association studies

- Gene expression is a continuous value, not yes/no ("Quantitative trait")
- Quantitative trait locus: a chromosomal region that affects the levels of a heritable quantitative trait.
- Gene expression and mRNA abundance of a given gene—is also heritable.
- eQTL (expression Quantitative Trait Locus) is the regulatory region for an expression trait.
- eQTL mapping is QTL mapping with an expression trait used as the phenotype.

Which ingredients are needed?

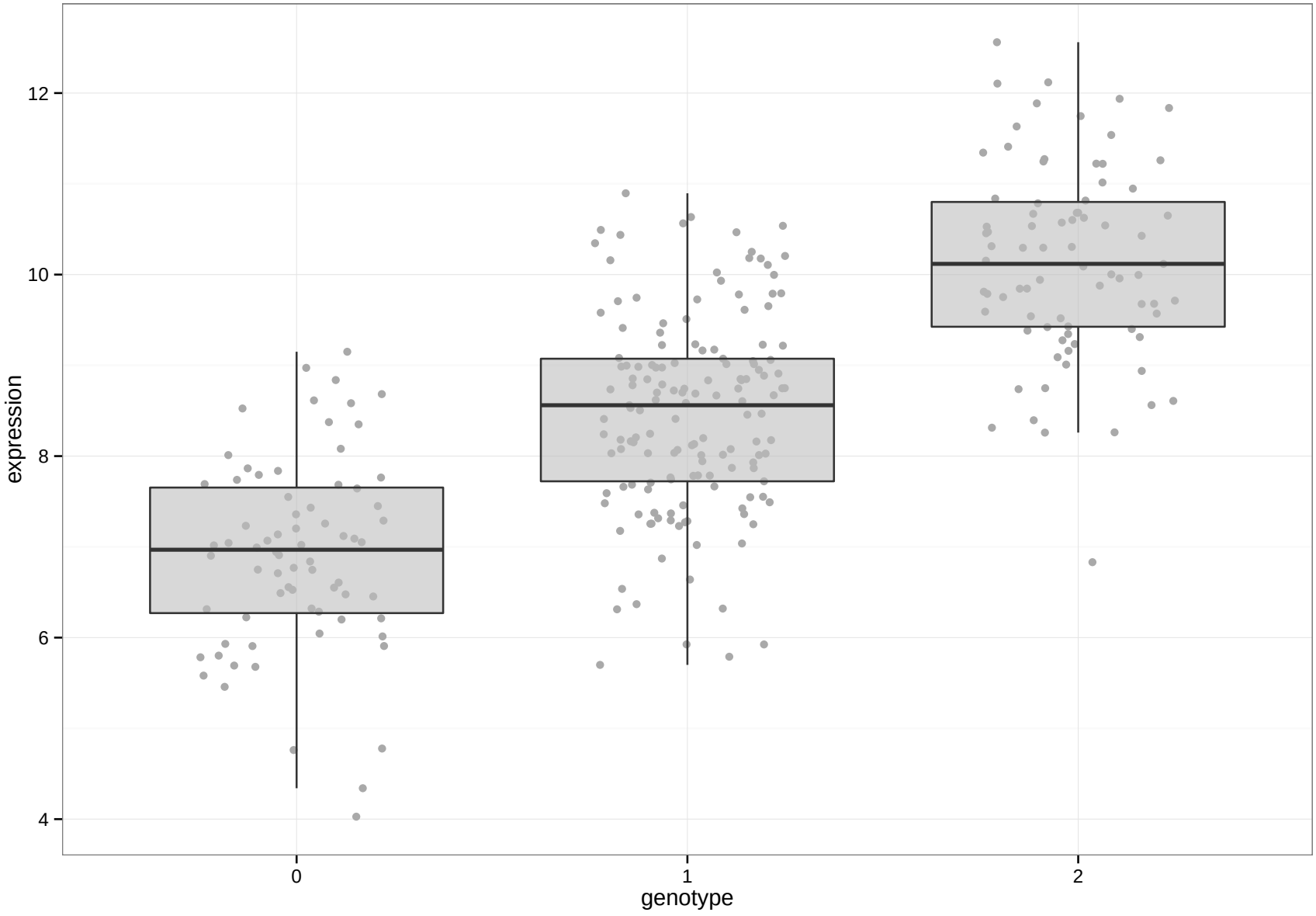
From each sample:

- Genotype data
- Gene expression data
- Covariates
  - Sex
  - Age
  - Smoking habits
  - Environmental effect sources
  - ...

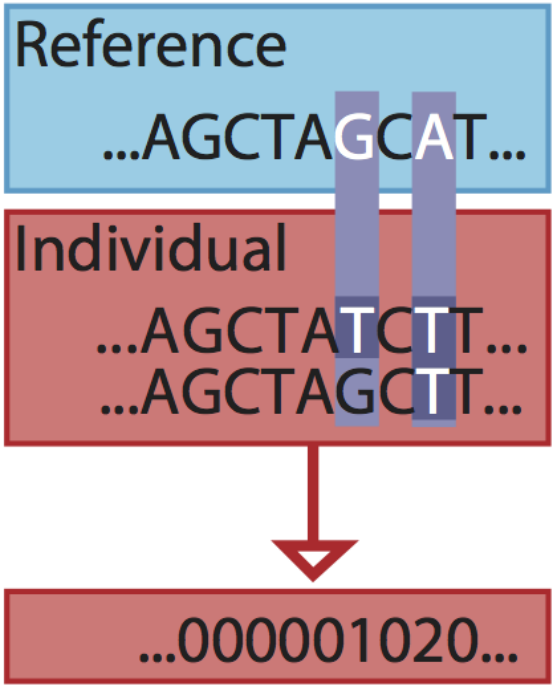
How does it work?

- Take expression levels of a gene
- Take genotype alleles at a certain position
- Measure correlation
- Estimate how trustworthy compared to other random SNP (= p-value)

Gene expression



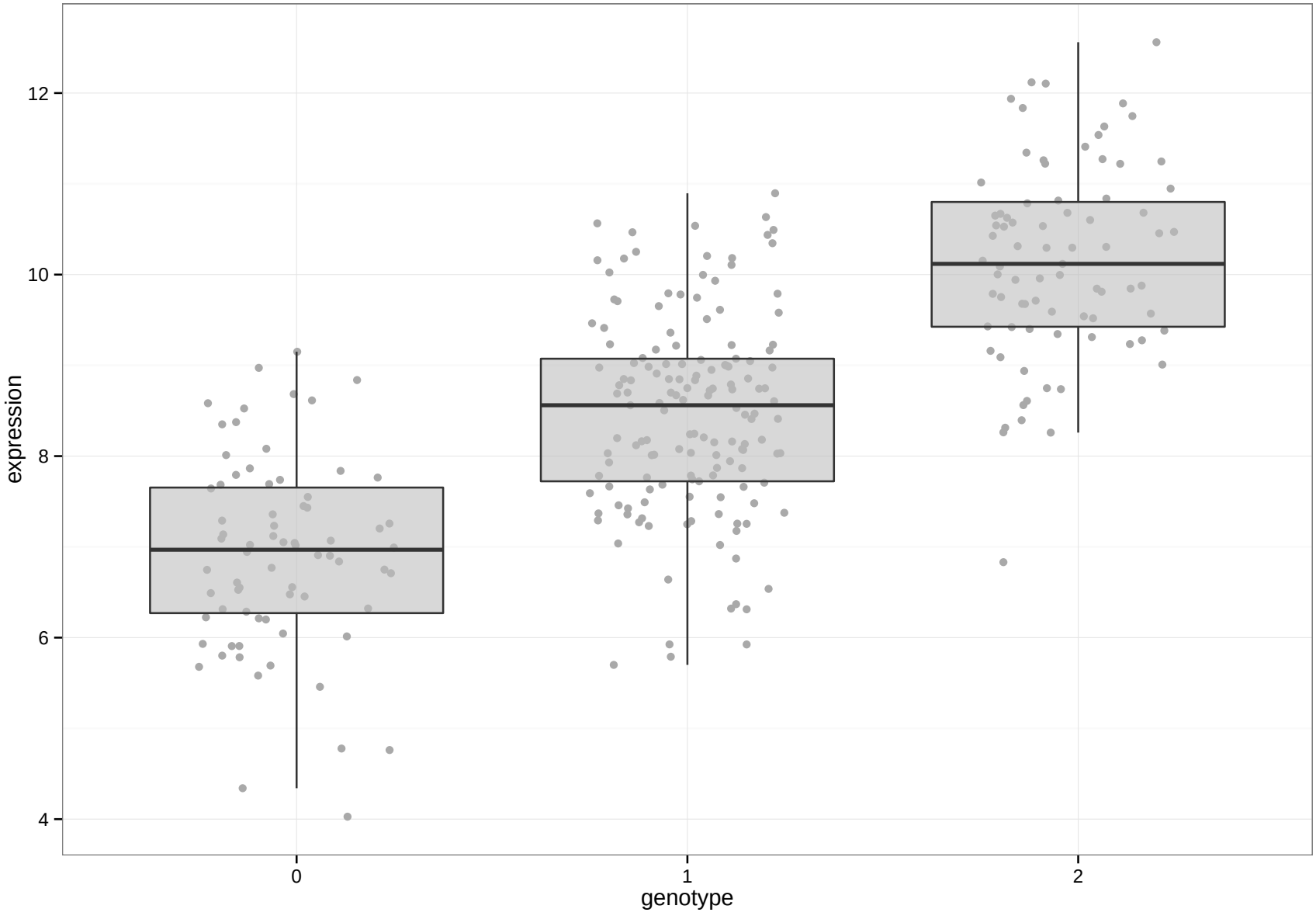
Genotype coding of indivuals



Spotting the differences:

- Individual has two SNPs
- In eQTL analyses SNPs can only have two alleles: The reference and the alternative.
- Each individual can have 0, 1 or 2 times the alternative allele.

Gene expression



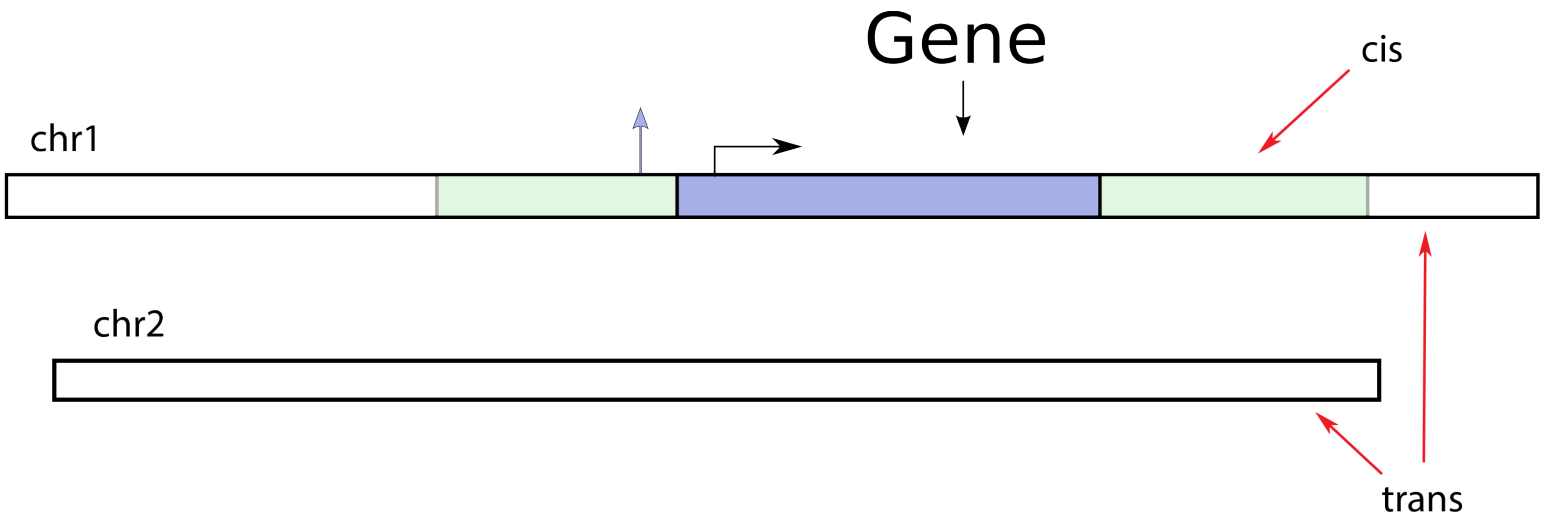
How to we select what to test?

- All genes are tested
- Genotype information is present for many locations ("SNP") within and between genes
- For each gene select a subset of those SNPs and test for association
  - Select SNPs close to a gene: cis
  - Select SNPs further away from the gene: trans
- Why not use the entire genome?
  - Multiple testing burden would hide many cis-associations
  - Differentiation between cis and trans effects

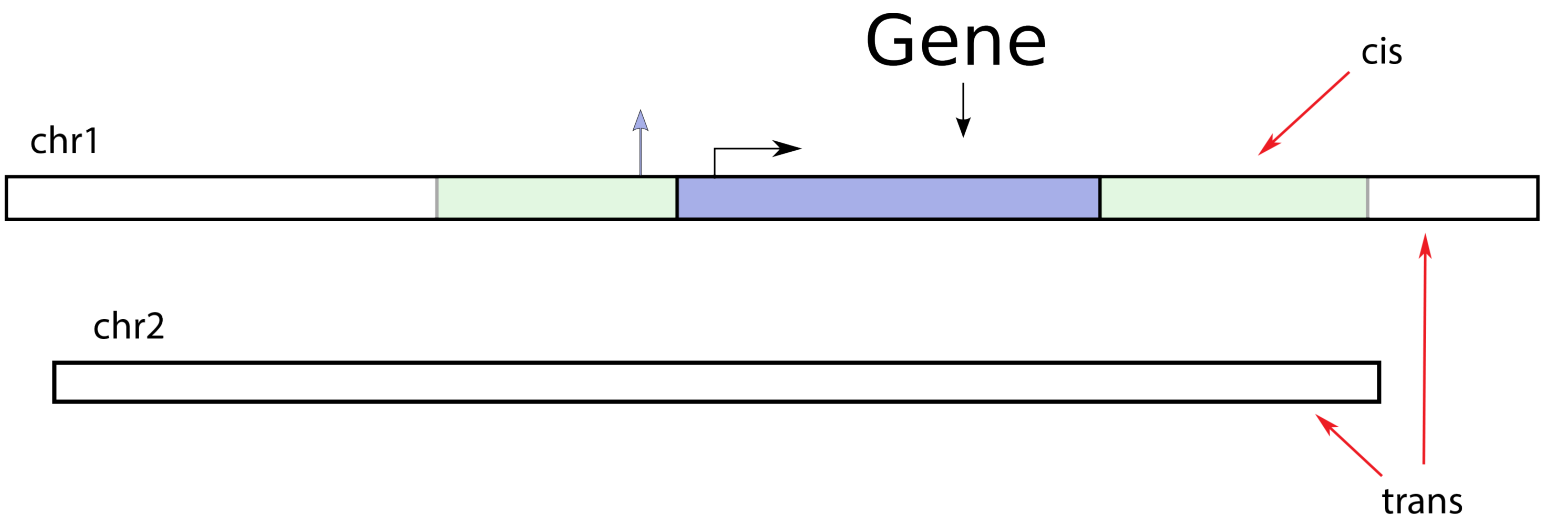
What is a cis / trans effect

- cis-effect:
  - The genetic variant in an individual (SNV) directly affects the expression of its associated gene
- trans-effect:
  - Gene expression is affected through possibly complex effects in which the genetic variant is involved
- How can we distinguish those?
  - Genomic linear distance of the tested SNP from the tested gene
    - Low distance: More likely cis-effect
    - High distance: More likely trans-effect

cis- and trans-eQTL testing



cis- and trans-eQTL testing



For each gene:

- cis: test all SNPs within 500kb around the gene (100s)
- trans: test all other SNPs in the genome (millions – 100s of millions)

Different models and programs

Linear (mixed) models

- MatrixEQTL (R package), FastQTL, LIMIX, PLINK, ...

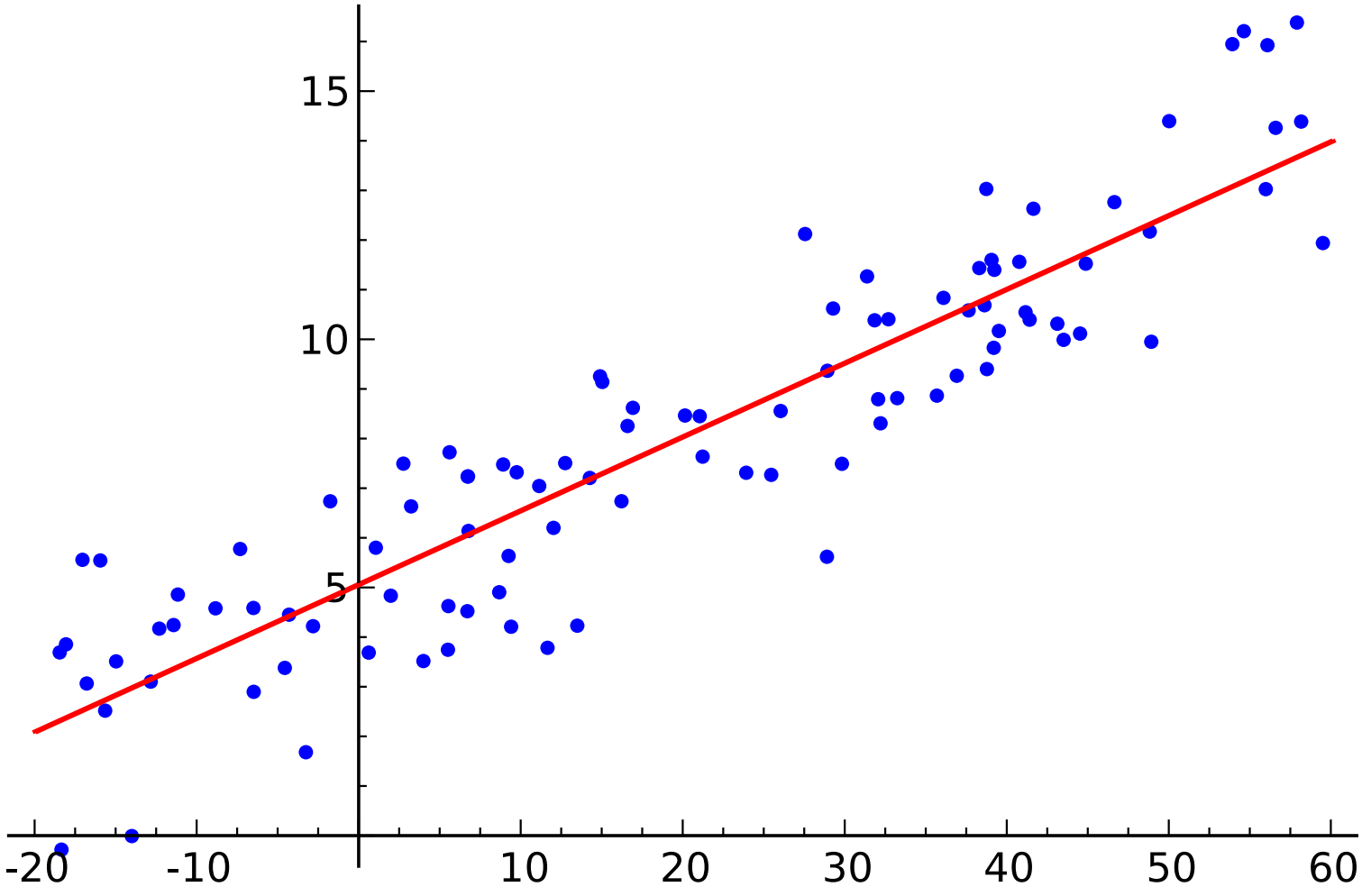
Random Forests

- LIMIX, ...

Bayesian approach

- GOAL (R package), ...

Linear regression



Linear regression in eQTLs