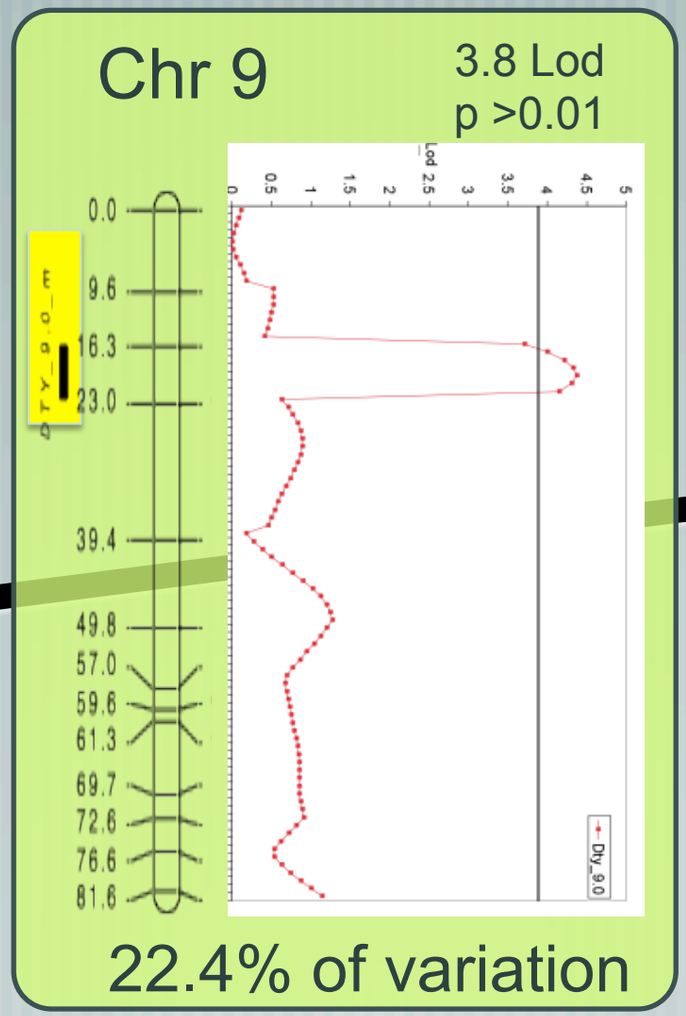
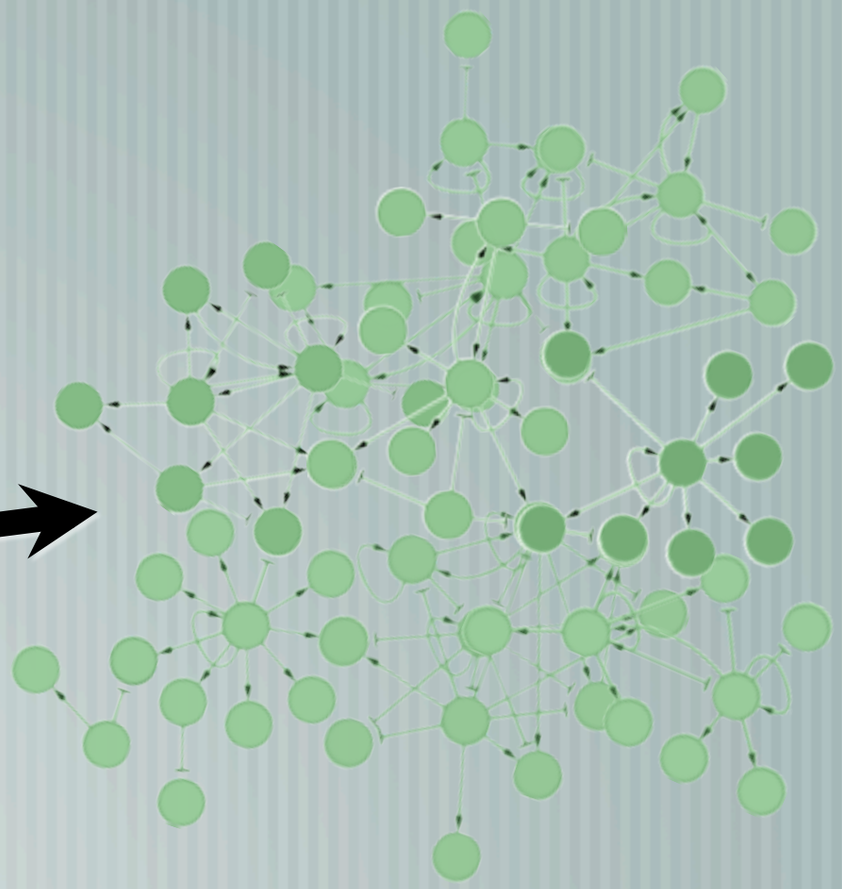
(4) QTLs to networks: eQTL analysis





Lecture objectives

By the end of this lecture you should be able to explain:

- How QTL analysis/genetics has been fused with genomics
- What eQTL analysis can do compared to standard expression analysis using microarrays
- What the different types of eQTLs are
- What the challenges associated with eQTL mapping are

What is expression QTL (eQTL) analysis?

- Recent development

- Classical QTL analyses combined with gene expression profiling

- Identify genomic locations that significantly influence the amount of protein or RNA made by a particular gene (gene expression)

- Expression of each gene = a trait; genes affecting it = eQTLs

How does eQTL analysis differ from genomic analysis?

- Large-scale gene expression data from microarray experiments:

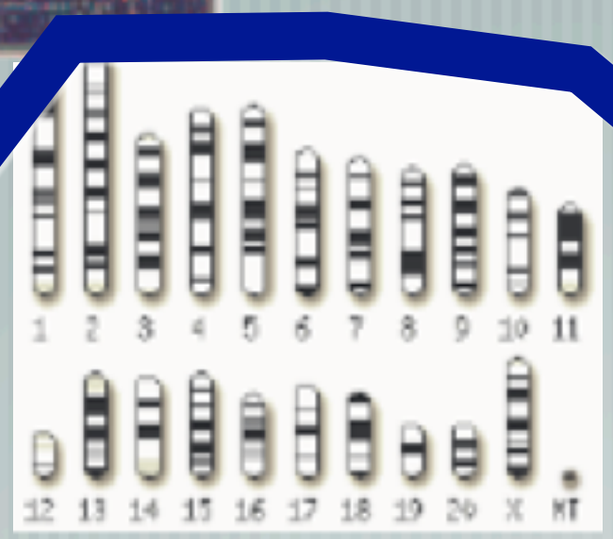
- information about regulatory relationships between genes

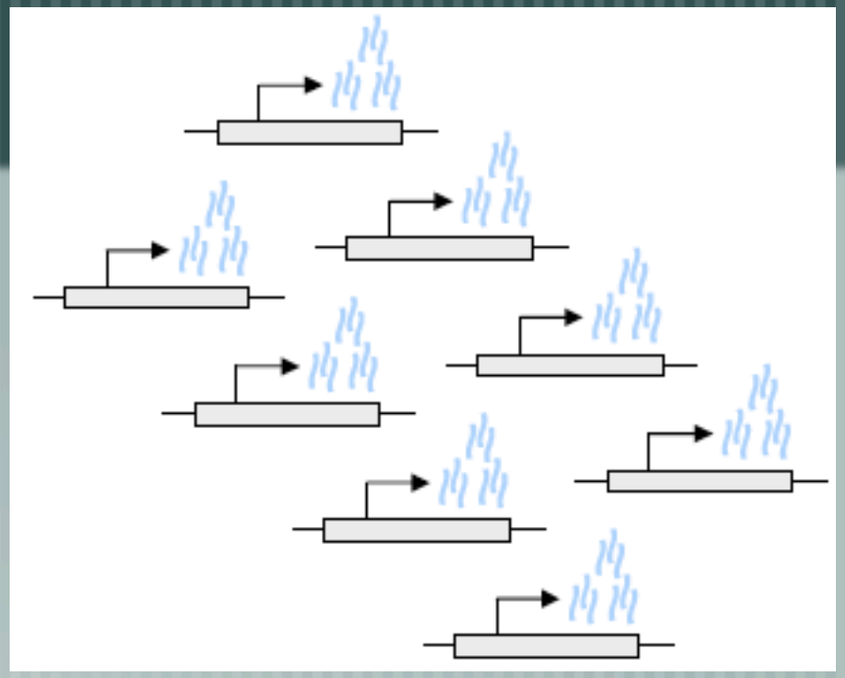
- Most approaches to transcriptional network inference rely on this alone

- based on external environment perturbations

- single-gene perturbations in homogeneous genetic background

eQTL overview



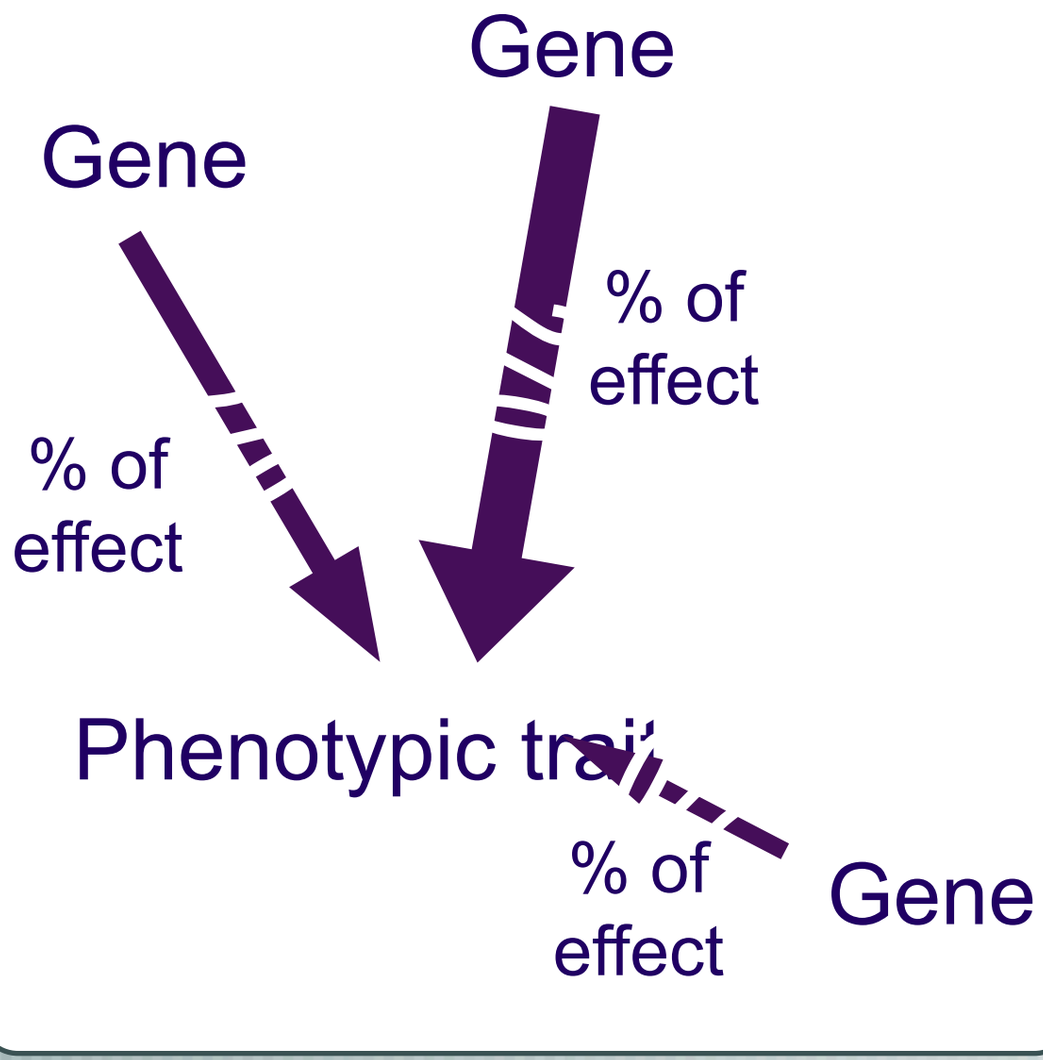


quantitative variation of mRNA levels in a segregating population

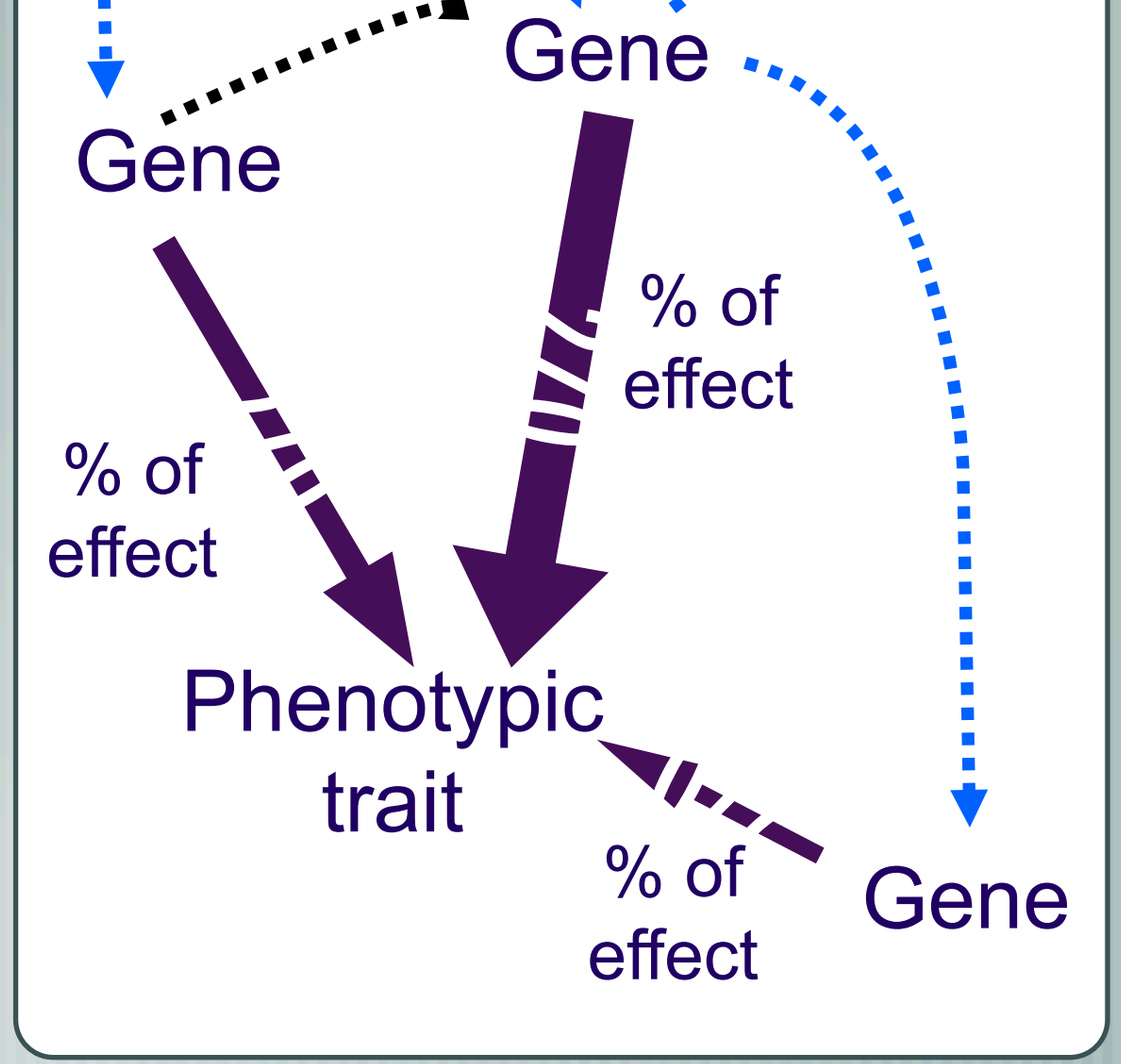
Expression QTLs
genetic determinants
of gene expression

Additional level of knowledge from eQTL analysis

QTL analysis
Association mapping



eQTL analysis



Some things to consider with eQTL mapping analysis

- Identify & map genomic regions that significantly affect gene expression levels

- Statistical methods and power to map eQTLs

- Common design: use RILs and examine a number of microarrays across a modest set of lines (10-100)

- Some improvement in power (over an F2 design) occurs because of being able to replicate within each RIL and the expanded map distances (4 fold) found in RILs vs. F2

- eQTL mapping faces many of the same lack of power issues as QTL mapping

Some issues with eQTL mapping analysis

- Identify & map genomic regions that significantly affect gene expression levels

- Measurement error/noise (could be high for microarrays)
- Justification of mapping procedures and therefore results

  For markers: evaluate linkage statistics for each genetic marker and use permutation testing to provide genome-wide corrected p-values

  For microarrays: determine the expected proportion of false positives called significant in the linkage analysis (FDR)

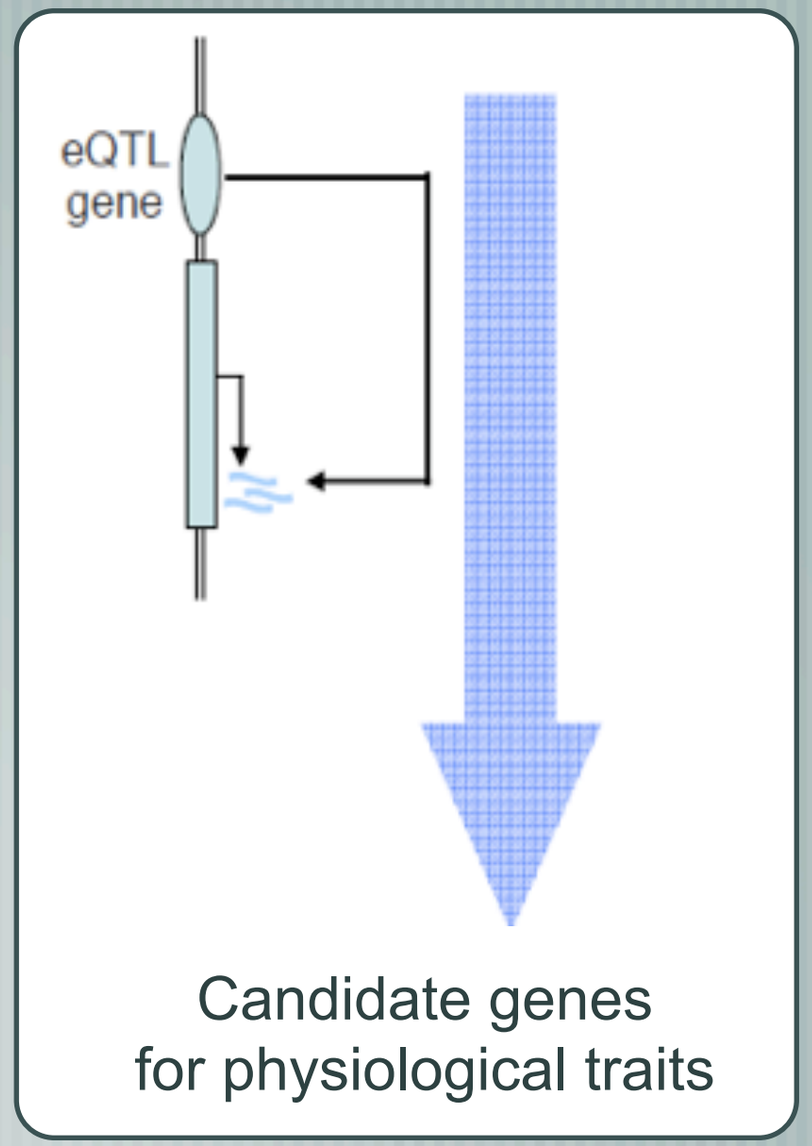
- Aim to identify cis-and trans-regulation of eQTLs

cis- and trans-acting eQTLs

- Used in a variety of species, including humans and rodents, to help elucidate the molecular basis of complex disease

  - Cis-acting eQTLs:
    - caused by genomic sequence variants that reside within or close to the gene being regulated

    - attractive candidate genes for (patho)physiological QTLs (pQTLs) mapped to the same location



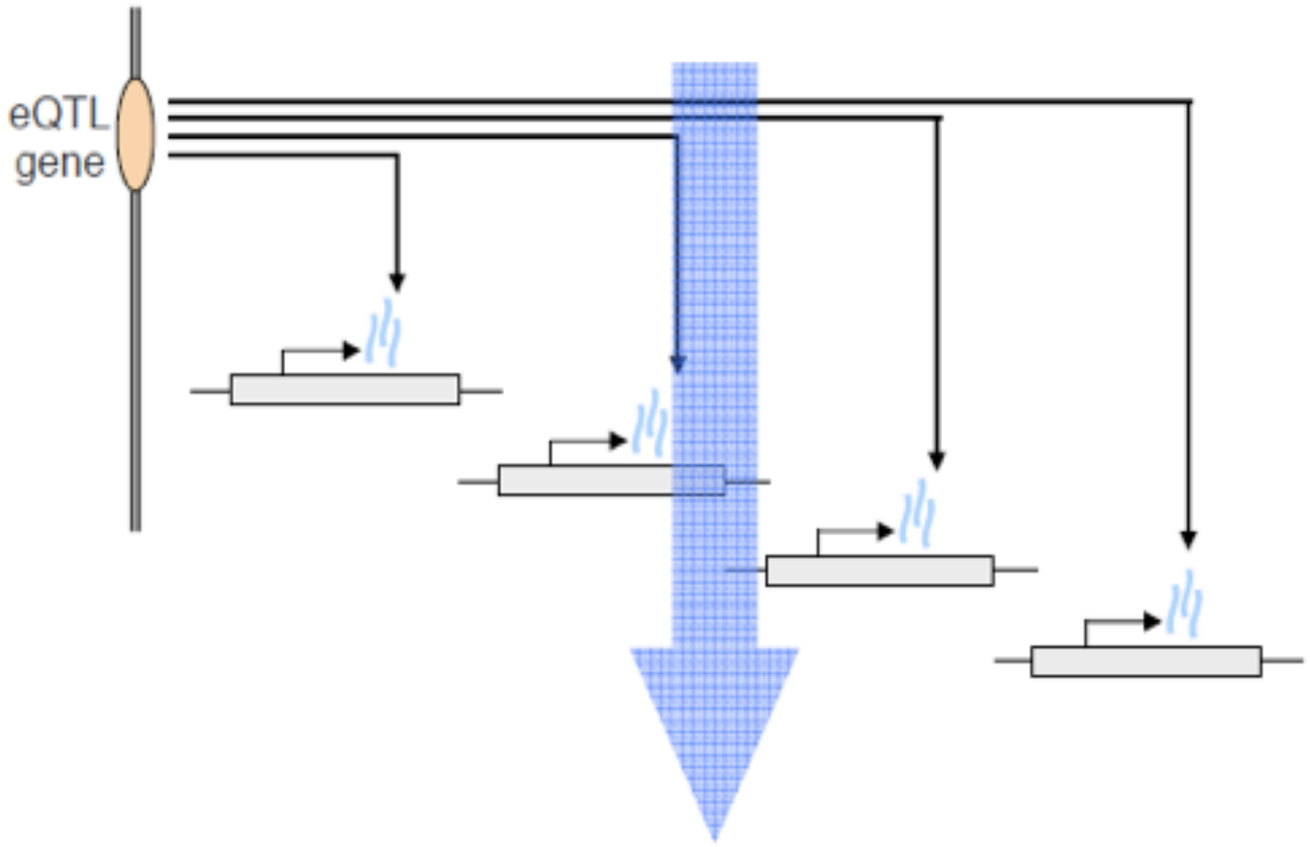
cis- and trans-acting eQTLs

- Trans-regulated eQTLs:

- reflect differences in remotely regulated gene expression and

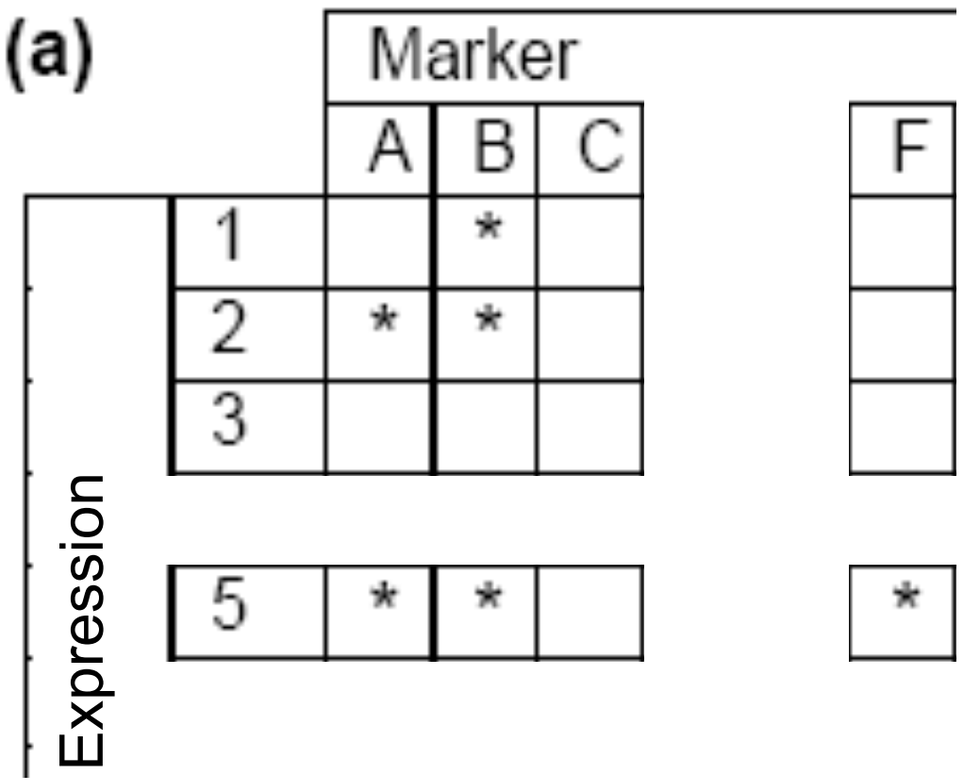
- often occur in clusters: co-ordinated ‘master’ regulation of many genes

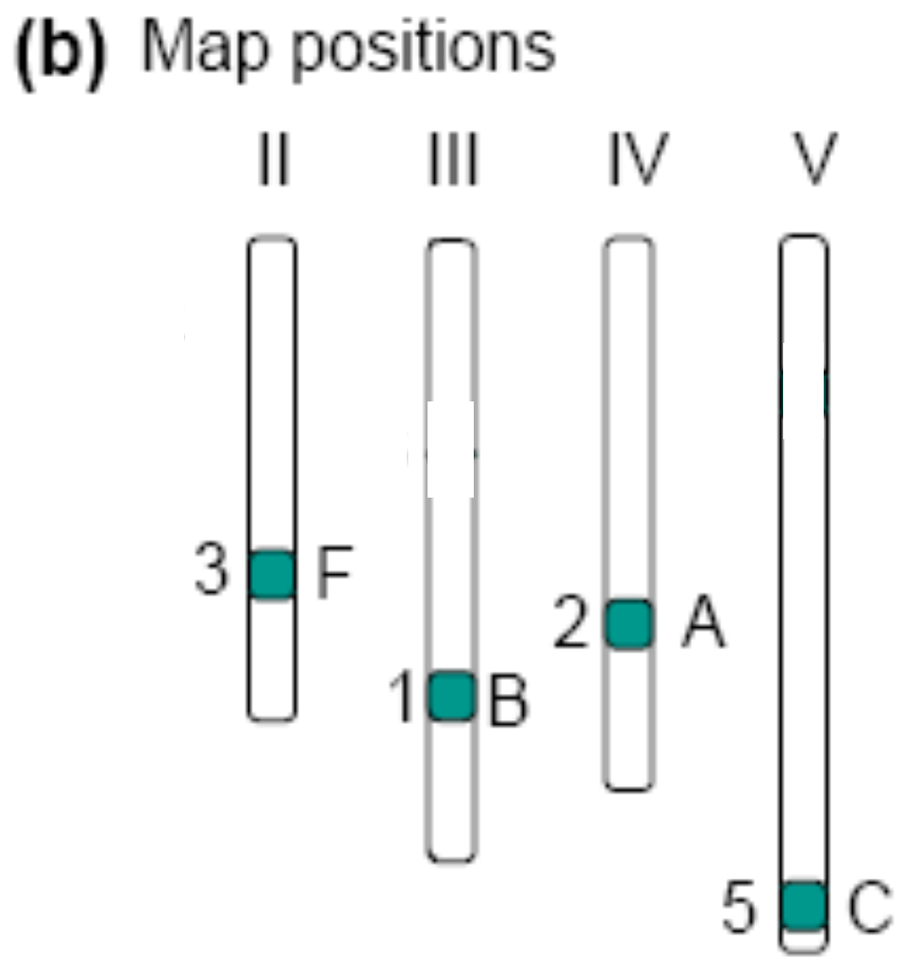
- Master regulators of gene expression - key control points in gene networks



Regulatory gene networks

Identifying cis and trans gene regulation



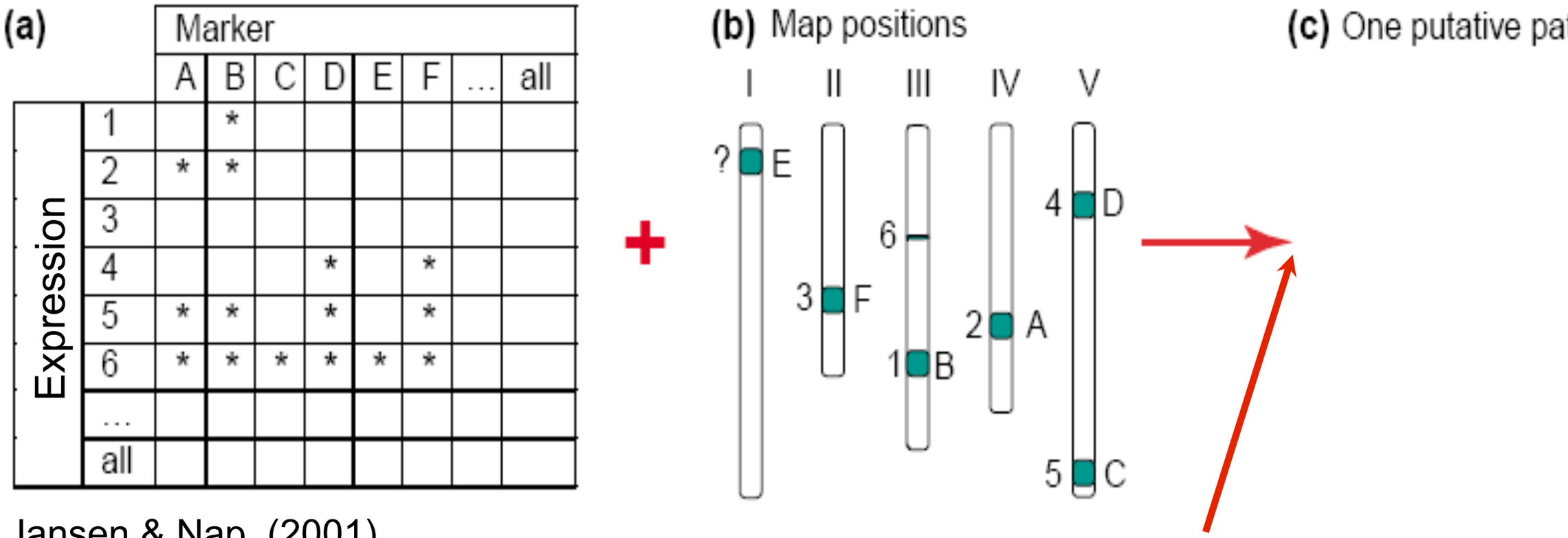


Jansen & Nap (2001)
Trends in Genetics: 17, 388-391

- Gene 3 = not regulated in cis or by any other gene
- Gene 1 = only controlled by ‘B’ = cis regulation
- Gene 2 = controlled by ‘A’ = cis regulation & by gene 1 ‘B’ = trans regulation
- Gene 5 = not cis (C) controlled but controlled in trans by 1 ‘B‘ & 2 ‘A‘ & 3 ‘F’

Moving to eQTL networks

Constructing regulatory networks (hypothetical example)



Jansen & Nap (2001)
Trends in Genetics: 17, 388-391

- Gene 3 = not regulated in cis or by any other gene
- Gene 1 = only controlled by 'B' = cis regulation
- Gene 2 = controlled by 'A' = cis regulation & by gene 1 'B' = trans regulation
- Gene 5 = not cis 'C' controlled but controlled in trans by 1 'B' & 2 'A' & 3 'F' & 4 'D'
- Gene 4 = controlled by 'D' = cis regulation & in trans by 3 'F' regulation

Challenges to eQTL mapping analysis

- Low levels of gene expression
- Fine mapping:

- Due to the spacing of genetic markers and/or linkage disequilibrium, several genes can reside near each marker

- Identifying the true causative gene requires additional data, since all genes at a locus are indistinguishable based on the eQTL measurements alone

- Lack of mechanistic explanation:

- A gene-phenotype association = little insight into underlying molecular mechanism for the association

Some ways to solve the problem of fine mapping

- Simple fine-mapping techniques

- linkage disequalibrium mapping
- additional genotyping of loci and individuals in regions of interest

- Main bioinformatic focus has been on predicting which genes within a given locus are the true regulators of expression of the target

e.g. Kulp and Jagalur (2006) sought to infer the true causal genes using a Bayesian network model constructed from expression correlations detected within the eQTL profiles

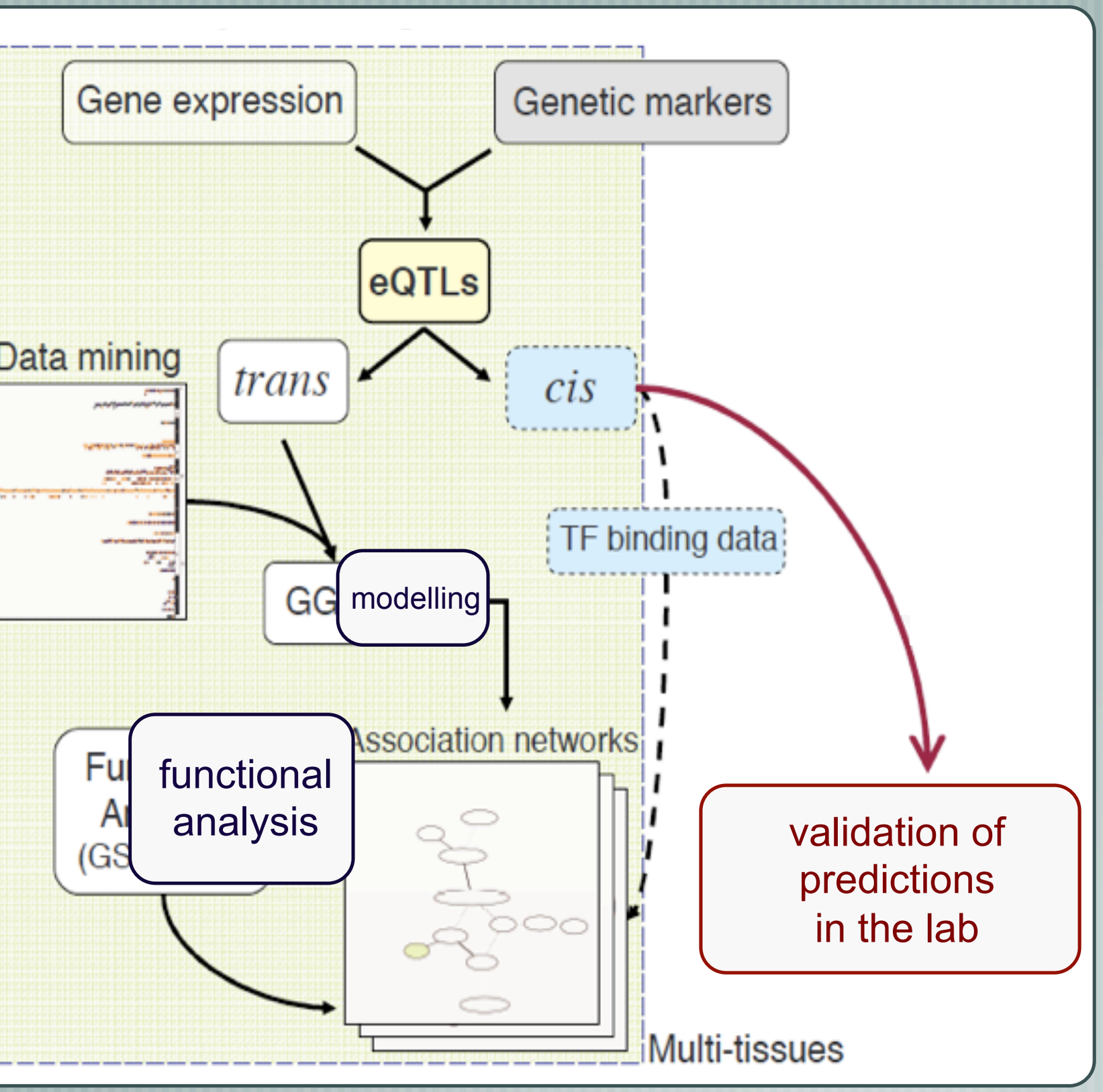
More ways to solve the problem of fine mapping

- Complement eQTLs with data on physical molecular interactions

e.g. Tu et al (2006) modeled each eQTL association as a sequence of transcriptional and protein–protein interactions (PPIs) that transmits signals from the locus to the affected target

Integrating eQTL data with additional independent information may significantly reduce the noise and improve the statistical power of the analysis

e.g. a strategy to identify master transcriptional regulators



Model for master transcriptional regulator

genetic variant

↓

cis-linked gene

↓

Transcription Factor (TF) activity profile

↓

Expression of *trans*-linked genes