

Introduction to expression (e)QTL and their role in connecting QTL to genes and molecular networks

3rd Webinar for Quantitative Genetics Tools for Mapping Trait Variation to Mechanisms, Therapeutics, and Interventions

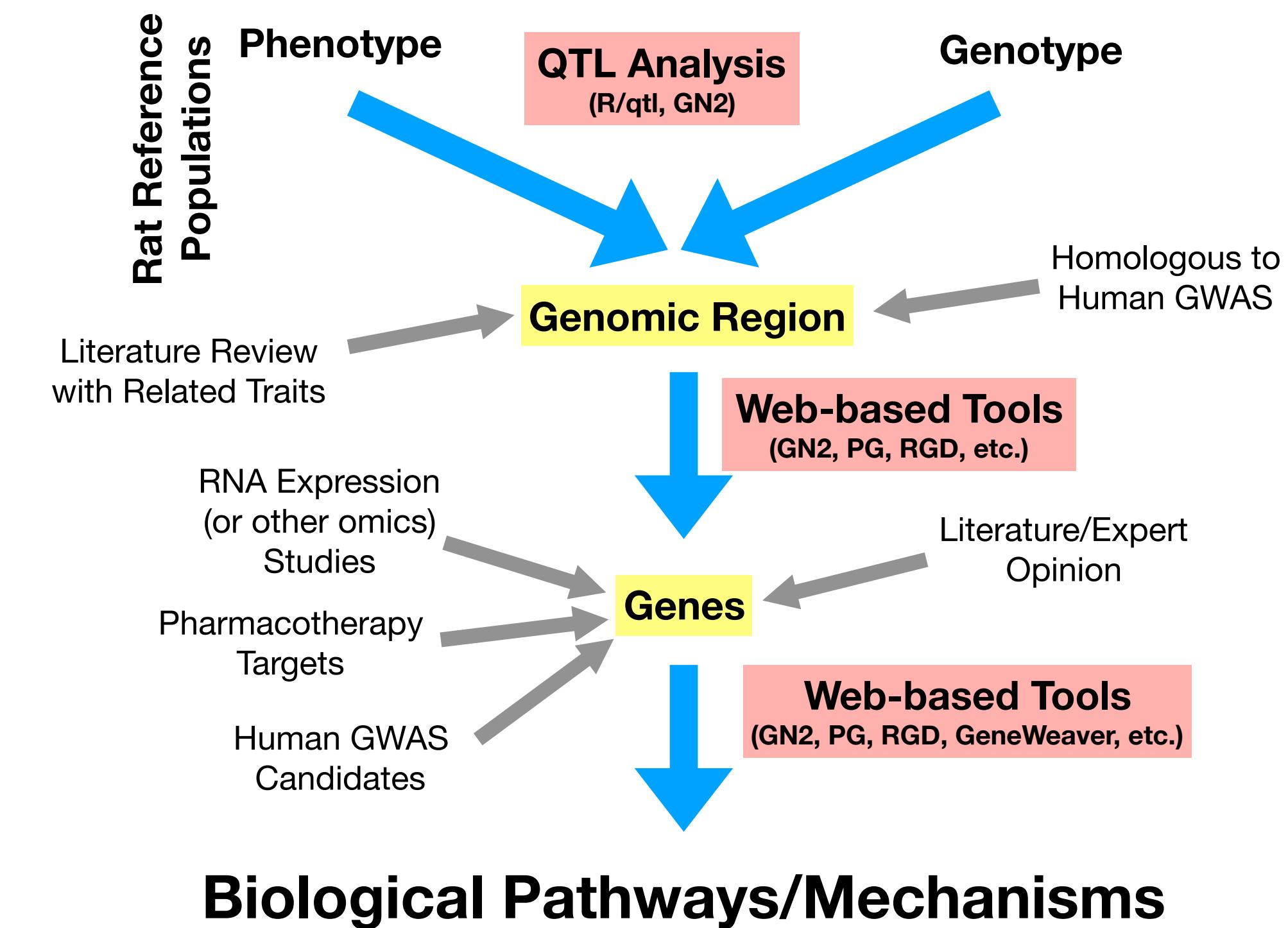
Laura Saba, PhD
University of Colorado Anschutz Medical Campus
NIDA Center of Excellence in Omics, Systems Genetics and the Addictome

Quantitative Genetics Tools for Mapping Trait Variation to Mechanisms, Therapeutics, and Interventions Webinar Series

Goal of the Series:

Transverse the path from trait variance to QTL to gene variant to molecular networks to mechanisms to therapeutic and interventions

Forward Genetics in Model Organisms



Recap of Webinar 1 - Quantitative Trait Loci Analysis

Presented by Dr. Saunak Sen

- **What is a quantitative trait locus (QTL)? Why are we interested in them?**

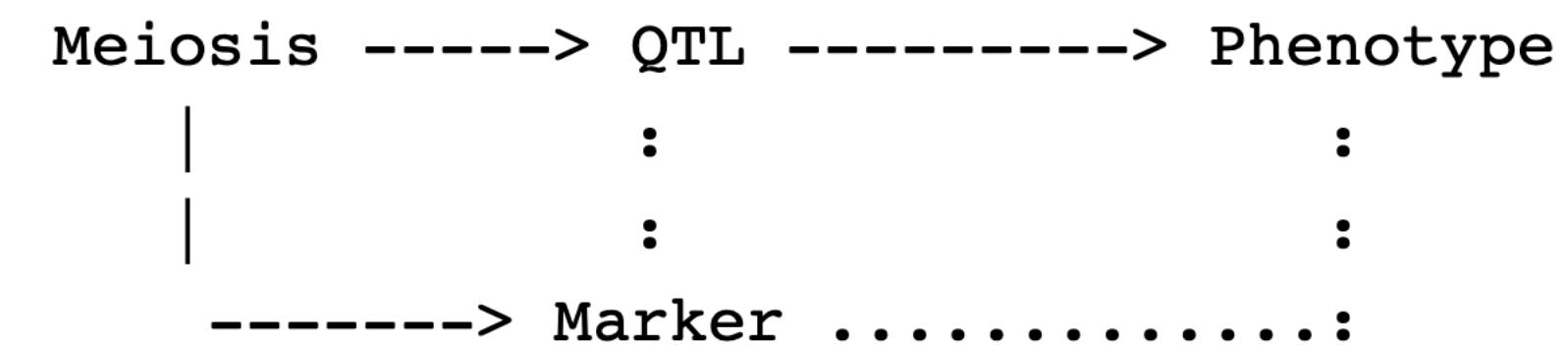
- A region of the genome (locus) that contributes to variation in a quantitative trait.
- Experimental crosses provide strong evidence of causation.
- Finding a QTL is usually the first step in dissecting the cause of the trait variation.

- **What are genome scans and how do they help find QTL?**

- Meiosis randomizes the (unobserved) QTL and (observed) marker.
- The QTL causes the phenotype.
- This leads us to observe an apparent association between marker and phenotype.
- Since the QTL is not directly observable, we test for association between genetic markers and phenotype (trait). The markers associated with the trait are likely to be physically close to the QTL (idea of genome scan).

- **What are the strengths and limitations of QTL mapping?**

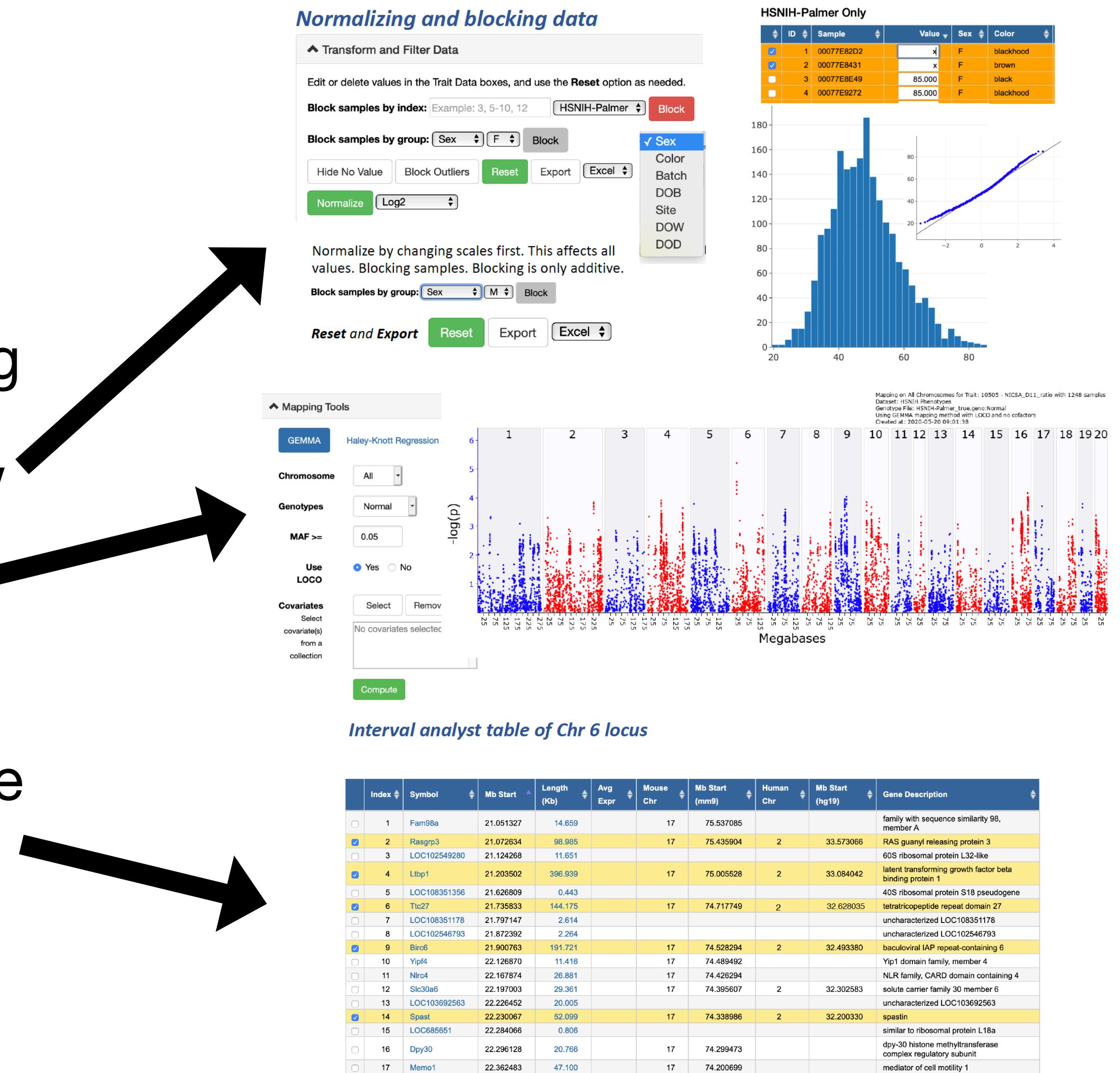
- Randomized genomes of experimental crosses allow us to infer causal genetic elements (QTL)
- We infer QTL by scanning the genome for genetic markers associated with a trait of interest



Recap of Webinar 2 - Mapping traits using GeneNetwork.org

Presented by Dr. Rob Williams

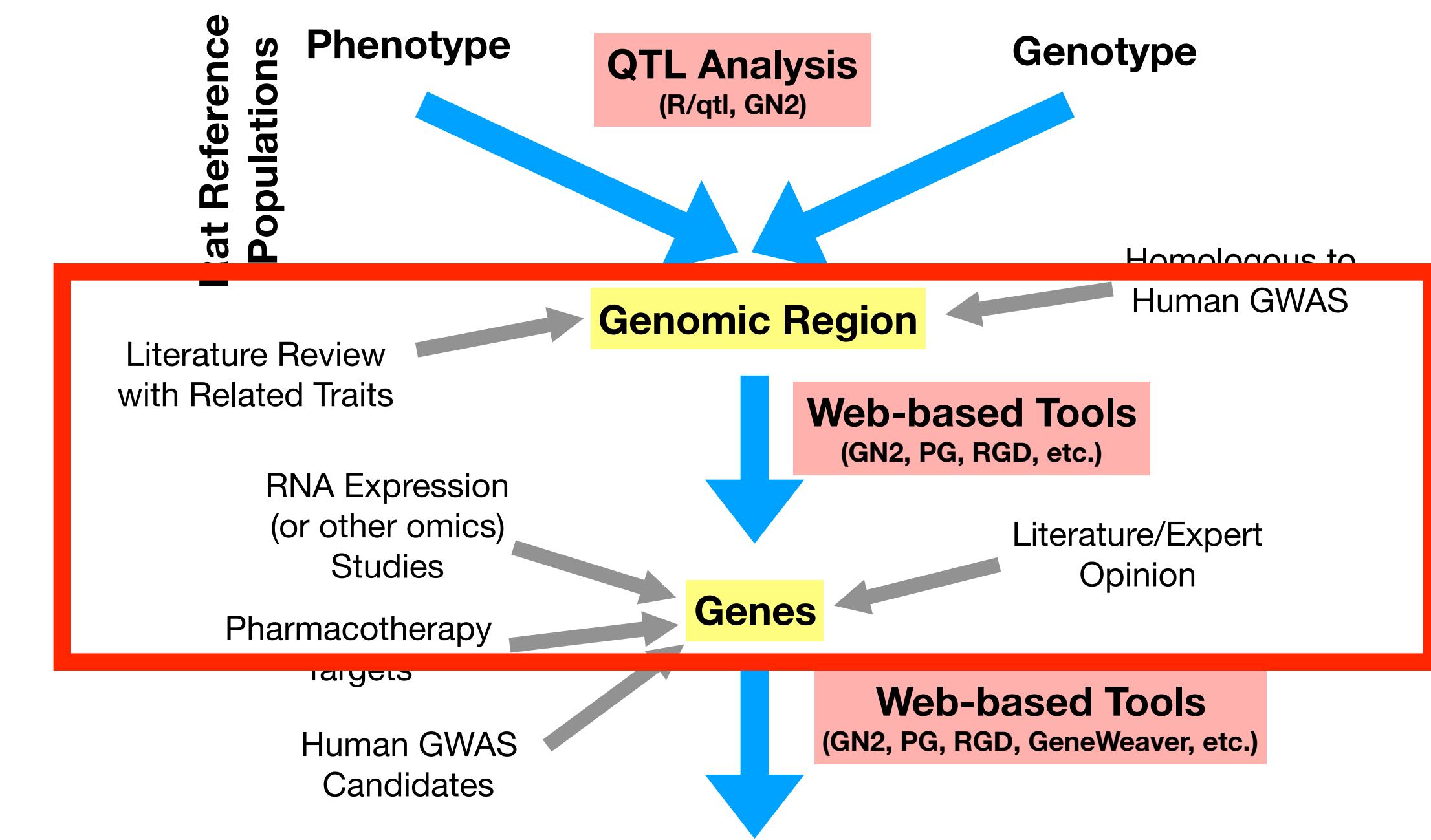
- Quick review of rat, mouse and human molecular, SUD, and behavioral data in GeneNetwork.org
- How to use GeneNetwork to review and edit trait data
- How to map QTLs
- How to evaluate and rank candidate genes



Outline

- Introduction to expression QTL
- How eQTL can be used to link phenotypic QTL to individual genes
- Introduction to RNA co-expression network analysis
- Linking co-expression networks to phenotypic QTL through module eQTL

Forward Genetics in Model Organisms



Introduction to eQTL

Strategies for linking QTL to genes of interest

Gene of interest physically located in/near the QTL

- Direct effect on character of protein - dichotomous outcome (on/off switch)
 - Causal DNA variant in protein coding sequence that changes the protein being produced
 - Causal DNA variant in transcription or translation start site that prevents the protein from being produced
- Direct effect on amount of RNA/protein produced - quantitative difference (dimmer switch)
 - Causal DNA variant in transcription factor binding site binding site that potentially alter levels RNA transcripts
 - Causal DNA variant in untranslated regions that alters levels of proteins produced

Strategies for linking QTL to genes of interest

RNA expression levels of a gene are controlled by the same QTL that controls the phenotype (i.e., eQTL/pQTL overlap)

- Direct effect on amount of RNA/protein produced
 - Causal DNA variant in transcription factor binding site binding site that potentially alter levels RNA transcripts
 - Causal DNA variant in untranslated regions that alters levels of proteins produced
- Indirect effect on the amount of RNA/protein produced
 - Causal DNA variant in transcription factor or microRNA targeting the gene
 - Causal DNA variant that affects chromatin structure

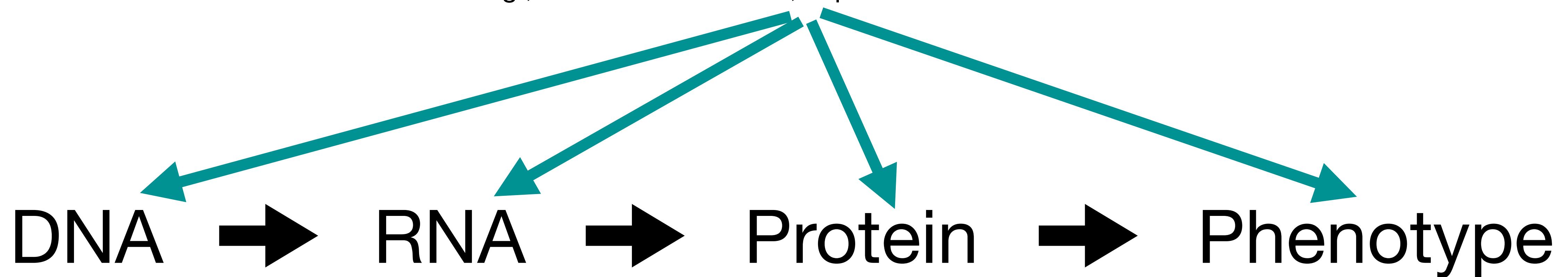
Simplistic view of central dogma of biology

DNA → RNA → Protein → Phenotype

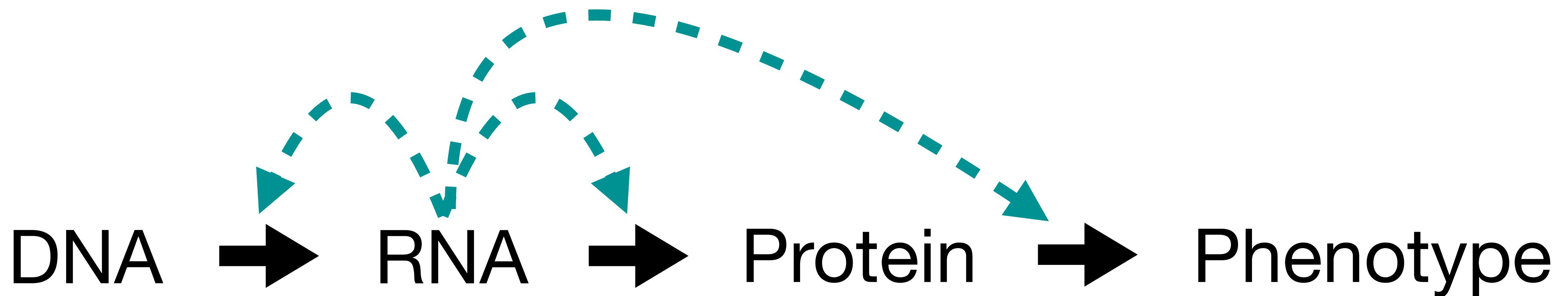
Role of the environment

Environmental effects

e.g., cellular environment, exposure to xenobiotics



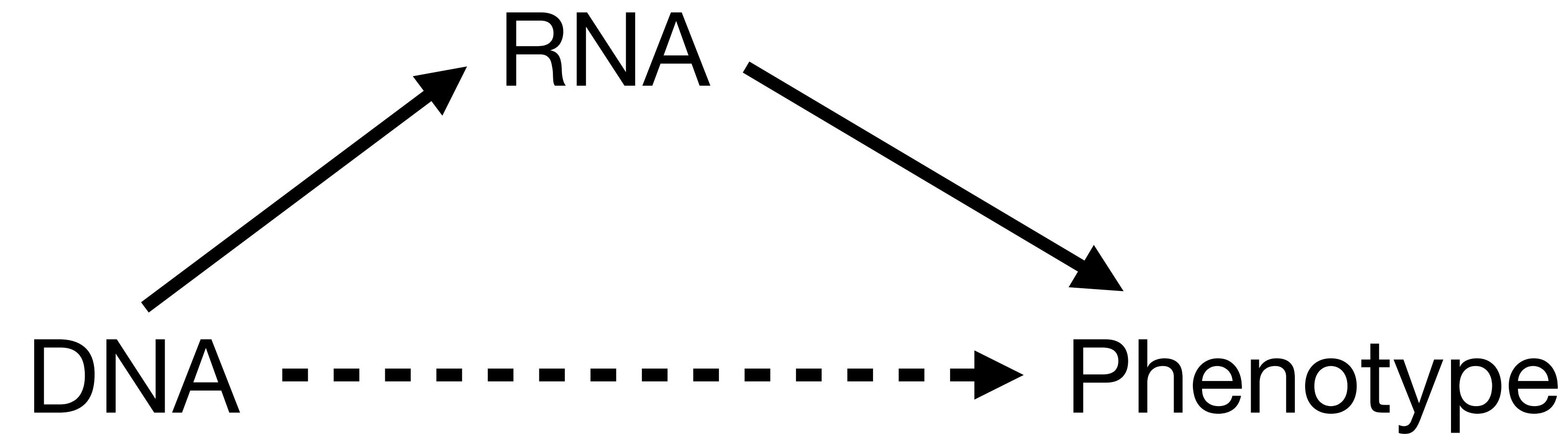
Various roles of RNA



Why RNA expression?

- One of the first quantitative links between DNA sequence and phenotype
- First step where DNA sequence and environment interact
- RNA expression levels differ between cell types and tissues within an individual, so incorporating RNA expression can help identify relevant cell types/tissues
- Network analyses at the transcript level provides insight into genetic/environmental interactions that are the basis for susceptibility to complex diseases

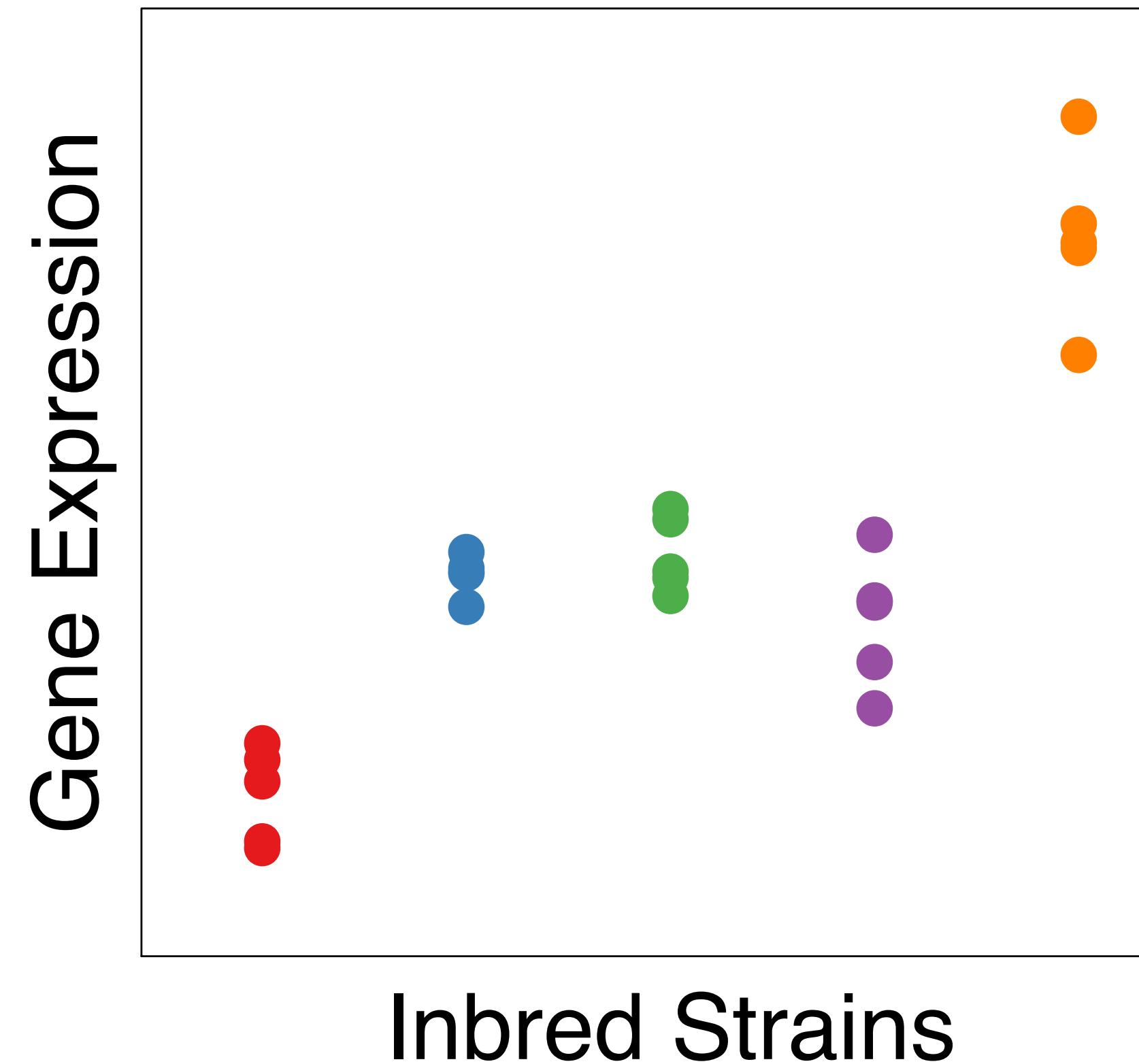
RNA as a mediator of the effect of a DNA variant on a phenotype



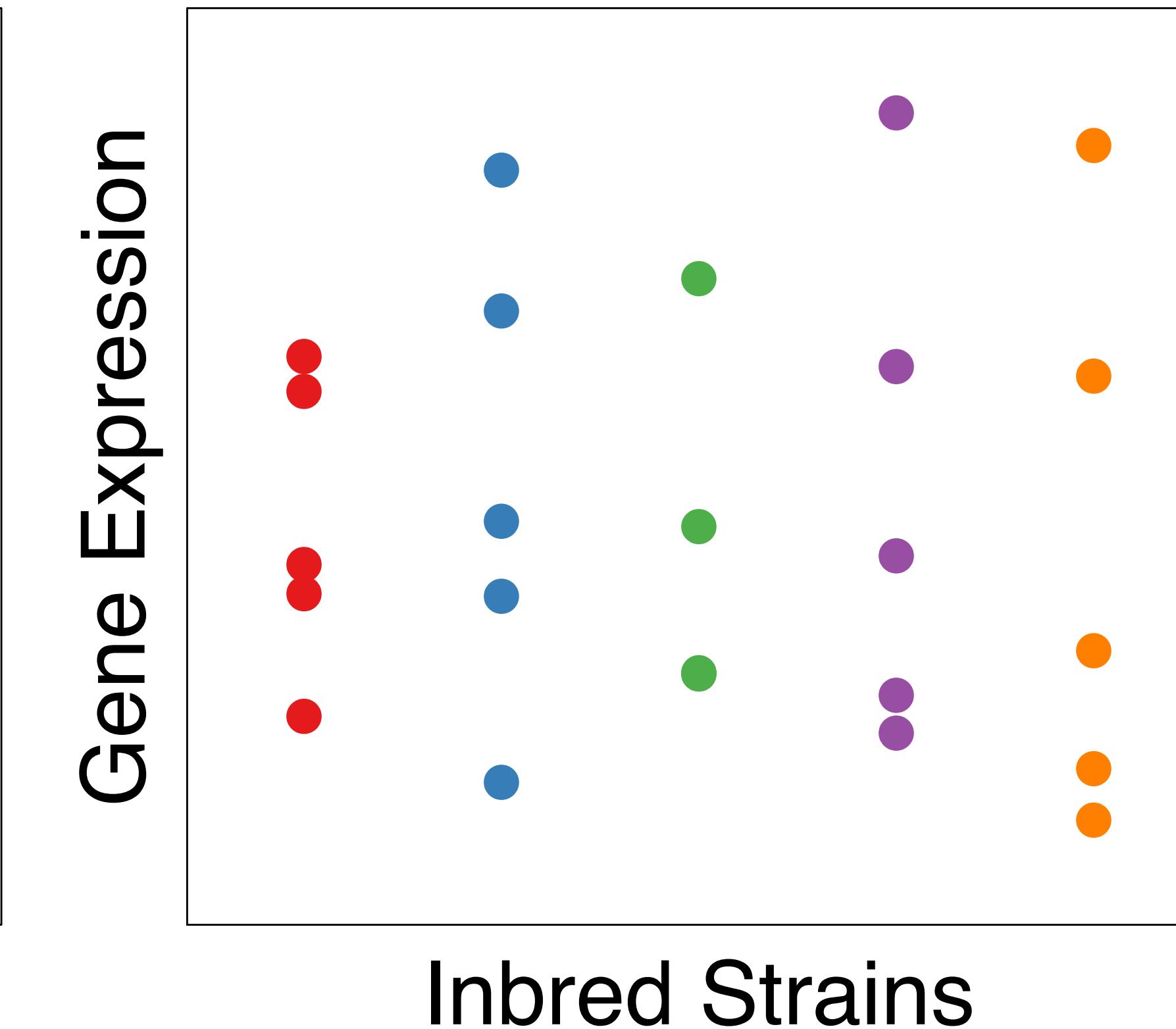
Genetic differences in RNA expression

Heritability - ratio of genetic variance to total variance

Strong Heritability

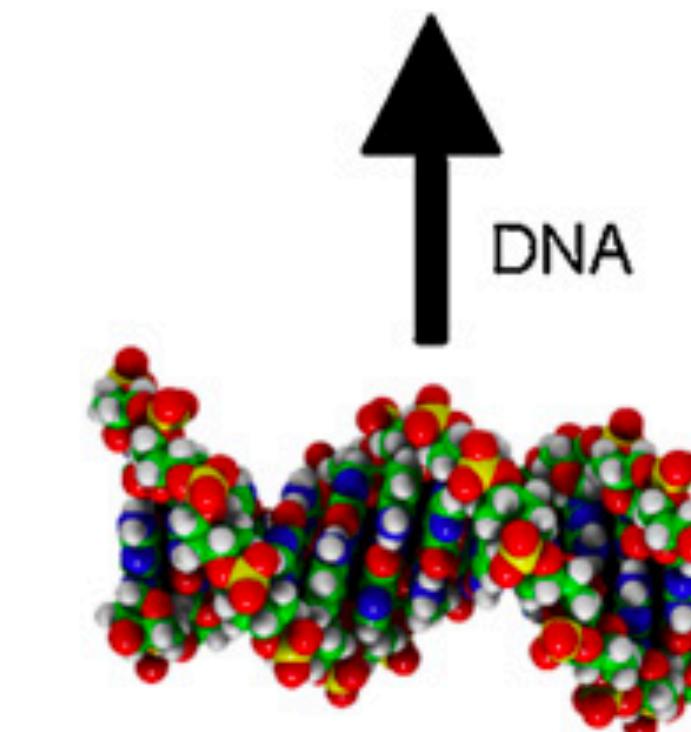
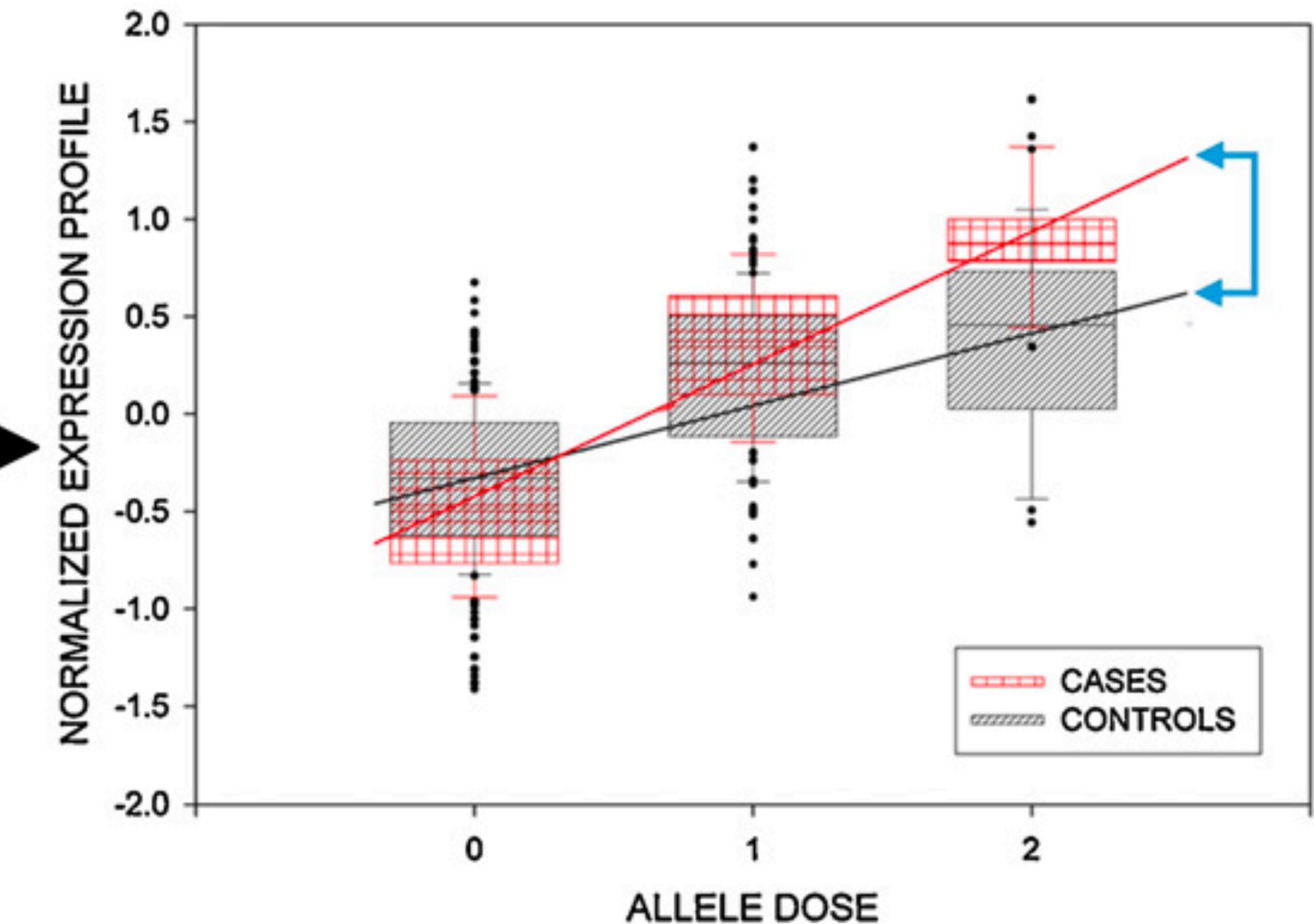
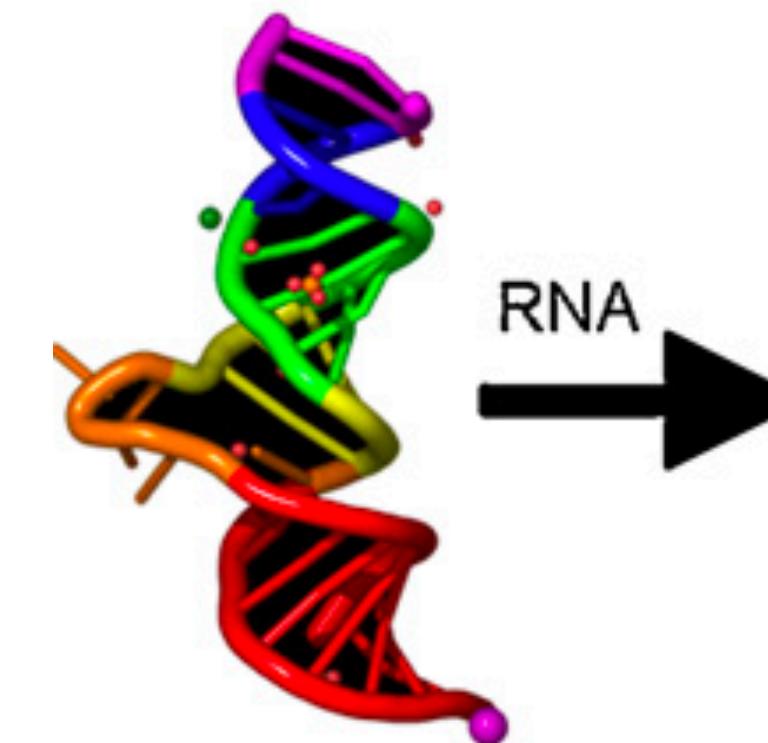


No Heritability



Mapping expression (e)QTL

- RNA expression levels can be treated like any other quantitative trait in QTL mapping.
- 30,000 genes by 10,000 SNPs = 300,000,000 comparisons!
- eQTL studies are sometimes called genetical genomics



cis vs. *trans* eQTL

- *cis*-eQTL (or local eQTL) – the locus controlling transcription is “near” the physical location of the gene in the genome
- *trans*-eQTL (or distal eQTL) – the locus controlling the transcription is NOT “near” the physical location of the gene in the genome

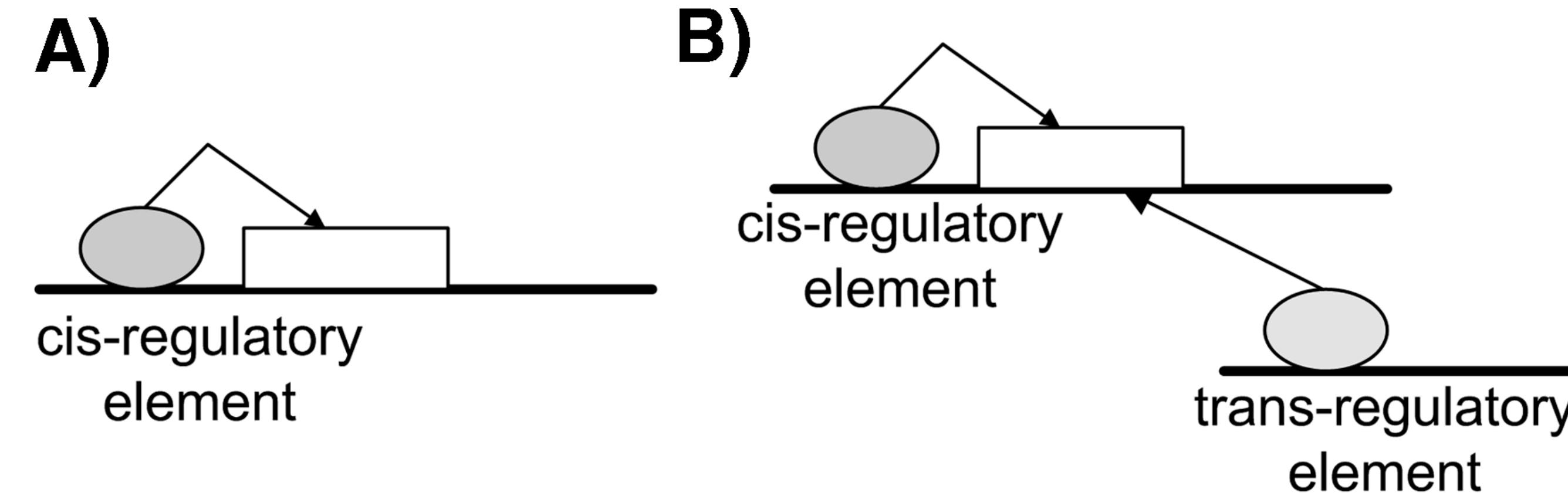
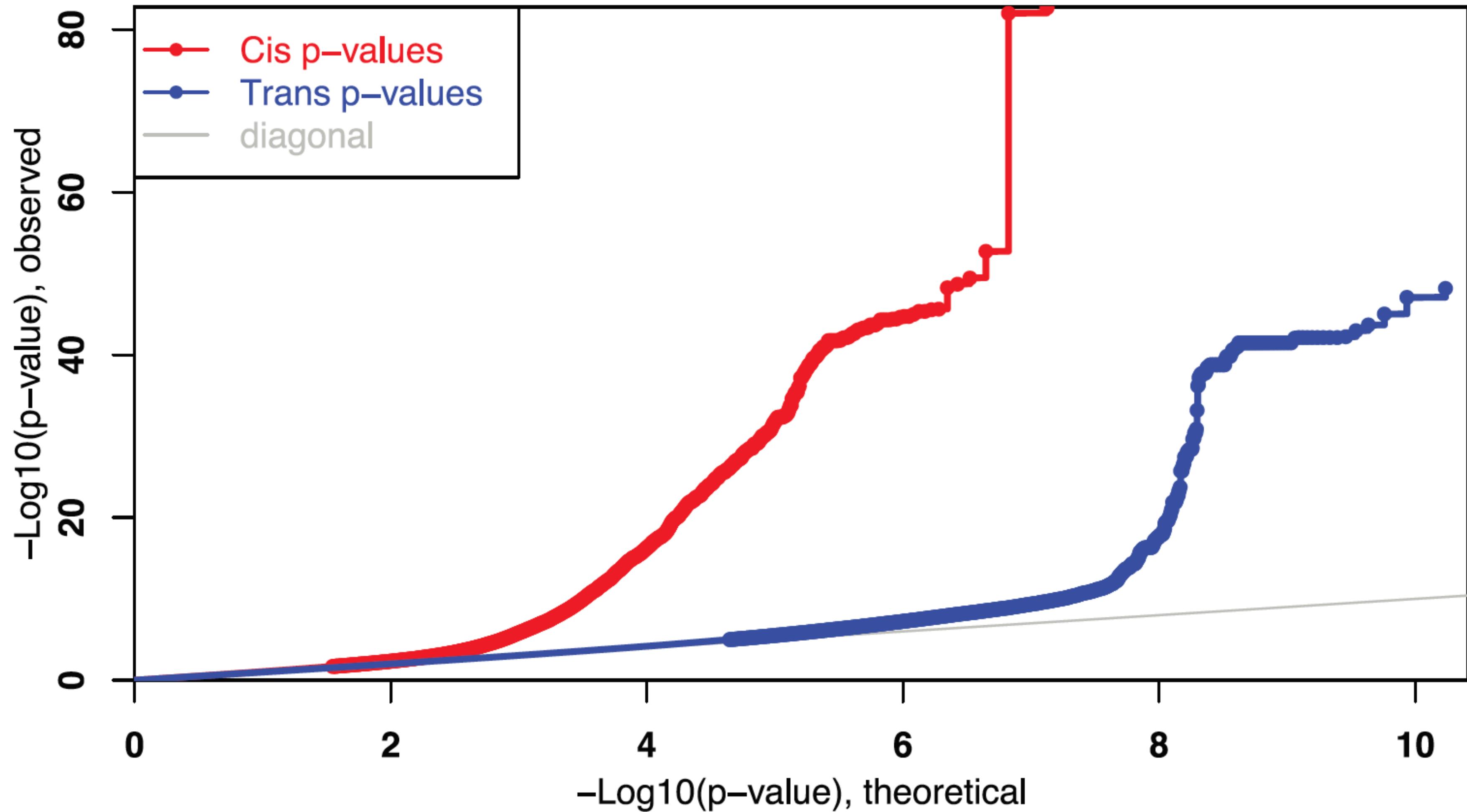


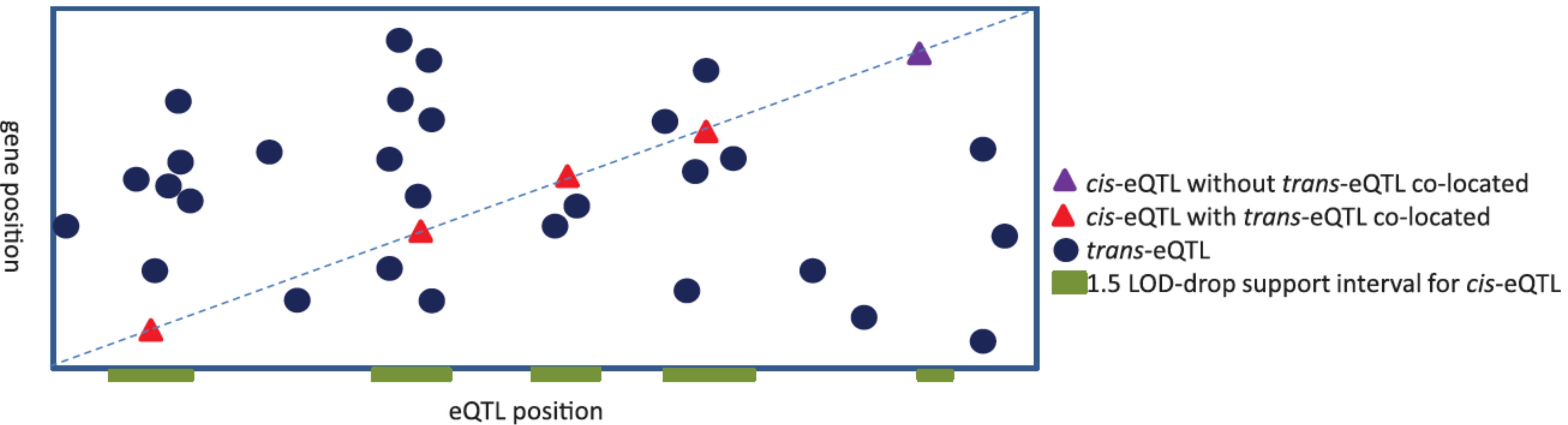
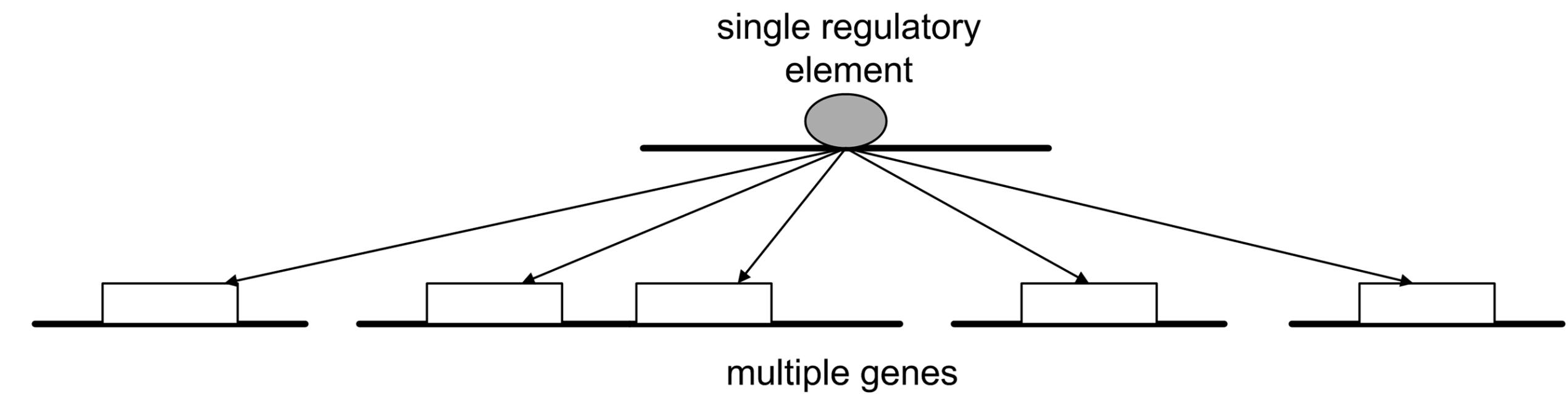
Image copied from Sieberts SK and Schadt EE (2007). Moving toward a system genetics view of disease. Mammalian Genome 18(6): 389-401.

Cis vs. trans eQTL

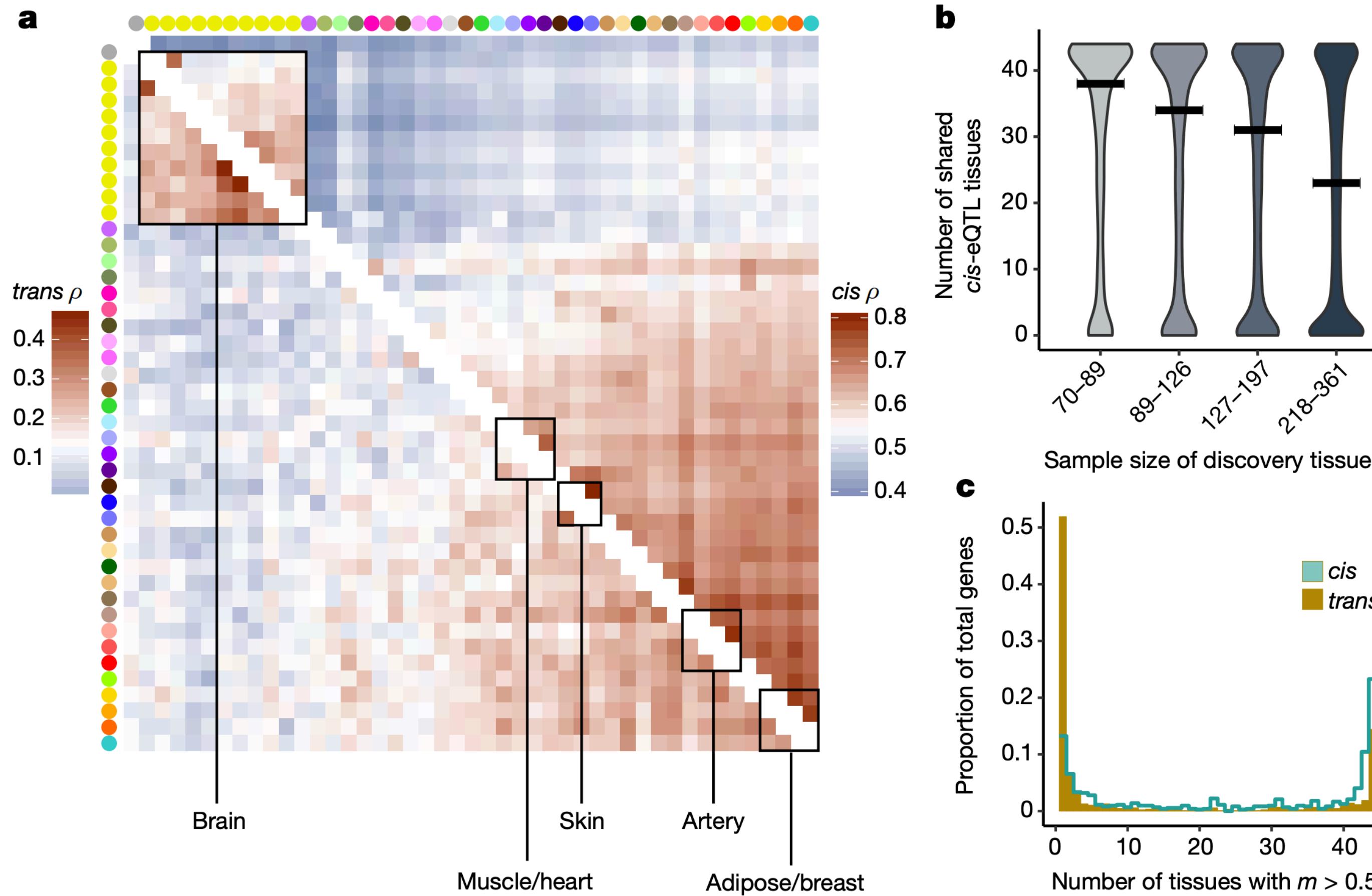
- *cis*-eQTL tend to have stronger statistical results than *trans*-eQTL
- *cis*-eQTL tend to represent a direct effect of a genetic variant on expression while a *trans*-eQTL usually exerts its influence on expression through another transcript or molecular element



eQTL hot spots



Tissue-specificity of eQTL



- *cis* eQTL are more likely to be shared across tissues.
- *cis* and *trans* eQTL tend to be similar across related tissues, i.e., different brain regions
- *cis* eQTL tend to either be replicated in almost all tissues or specific to only 1 or 3 tissues
- *cis* eQTL with large effect sizes are more likely to be replicated across tissues

Tools for mapping eQTL

eQTL can be mapped by any software that maps behavioral or physiologic QTL

Efficiency is of utmost importance!

Standard QTL analysis (i.e., no population structure)

- R/qtl; QTLReaper

QTL analysis with population structure

- R/qtl2; GEMMA; Fast-LMM; GCTA

Databases/Websites for Rat eQTL

- PhenoGen (<https://phenogen.org>)
 - Hybrid Rat Diversity Panel- brain; liver; heart; brown fat; kidney (coming soon)
- GeneNetwork (<http://genenetwork.org/>)
 - Hybrid Rat Diversity Panel - adipose; adrenal gland; brain; heart; hippocampus; kidney; liver; peritoneal fat
 - Heterogeneous Stock - nucleus accumbens; lateral habenula; prelimbic cortex; infralimbic cortex; orbitalfrontal cortex

Summary of intro to eQTL

- Expression (e)QTL are regions of the genome that influence the RNA expression levels in a tissue
- Cis (or local) eQTL tend to have a bigger effect size and are more likely to be retained across tissues than trans (or distal) eQTL
- eQTL are mapped using similar methods/software as other types of QTL. However efficiency and computational time may become a deciding factor on what methods/tools will be used.

**Phenotypic QTL to eQTL to
genes**

Motivation for Genetical Genomics/Phenomics Approach

When only identifying candidate genes as genes that are physically located in within a phenotypic QTL you may:

- Miss genes that are not physically located within the QTL but have a molecular element such as a transcription factor or a binding protein that is essential/relevant to the function or the expression level of the protein/RNA that gene codes for
- Identify genes that are not expressed in the tissue(s) of interest
- Miss critical information about how the causal variant influences a phenotype

When identifying candidates genes through the overlap of eQTL with the phenotypic QTL you will likely:

- Focus attention on genes that expressed in the tissue of interest
- Identify candidate genes that may influence the phenotype through variation in RNA expression
- Be able to identify broader biological/mechanistic themes through shared function/shared control of multiple genes with overlapping eQTL and pQTL

Definition of Genetical Genomics/Phenomics Approach

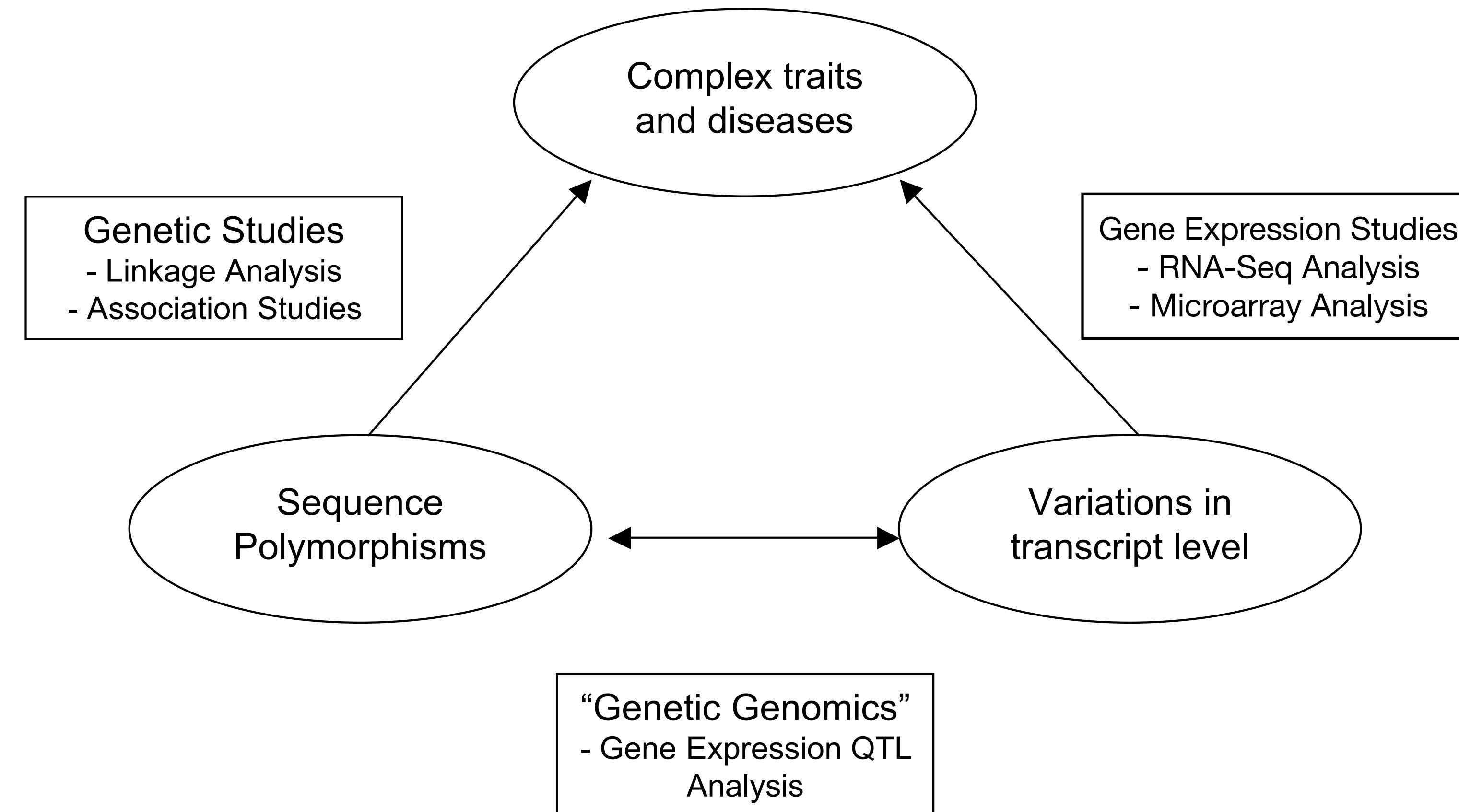


Image copied from Saba et al, The Marriage of Phenomics and Genetical Genomics: A Systems Approach to Complex Trait Analysis. In Systems Biology in Psychiatric Research: From High-Throughput Data to Mathematical Modeling, edited by Tetter F, Winterer G, Gebicke-Haerter PG, and Mendoza E. Wiley-VCH 2010.

Application of GGP approach

BMC Biology

BioMed Central

Research article

Open Access

Genetical genomic determinants of alcohol consumption in rats and humans

Boris Tabakoff^{*1}, Laura Saba¹, Morton Printz², Pam Flodman³, Colin Hodgkinson⁴, David Goldman⁴, George Koob⁵, Heather N Richardson^{5,16}, Katerina Kechris⁶, Richard L Bell⁷, Norbert Hübner⁸, Matthias Heinig⁸, Michal Praveneč⁹, Jonathan Mangion^{10,17}, Lucie Legault¹¹, Maurice Dongier¹¹, Katherine M Conigrave¹², John B Whitfield¹³, John Saunders¹⁴, Bridget Grant¹⁵, Paula L Hoffman¹ and WHO/ISBRA Study on State and Trait Markers of Alcoholism

Address: ¹Department of Pharmacology, University of Colorado, Denver, Aurora, CO, USA, ²Department of Pharmacology, University of California San Diego, La Jolla, CA, USA, ³Department of Pediatrics, University of California Irvine, Irvine, CA, USA, ⁴Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD, USA, ⁵Committee on the Neurobiology of Addictive Disorders, The Scripps Research Institute, La Jolla, CA, USA, ⁶Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Denver, Aurora, CO, USA, ⁷Institute of Psychiatric Research, Indiana University School of Medicine, Indianapolis, IN, USA, ⁸Max Delbrück Center for Molecular Medicine, Berlin, Germany, ⁹Institute of Physiology, Czech Academy of Sciences, Prague, Czech Republic, ¹⁰MRC Clinical Sciences Centre, London, UK, ¹¹Department of Psychiatry, McGill University, Montreal, Quebec, Canada, ¹²Drug Health Services, Royal Prince Alfred Hospital, Sydney Medical School, University of Sydney, New South Wales, Australia, ¹³Queensland Institute of Medical Research, Queensland, Australia, ¹⁴School of Medicine, University of Queensland, Brisbane, Queensland, Australia, ¹⁵Division of Epidemiology, National Institute on Alcohol Abuse and Alcoholism, Rockville, MD, USA, ¹⁶Department Psychology-Neuroscience, University of Massachusetts Amherst, Amherst, MA, USA and ¹⁷Applied Biosystems, Lingley House, 120 Birchwood Blvd., Warrington, Cheshire, WA3 7QH, UK

Email: Boris Tabakoff* - boris.tabakoff@ucdenver.edu; Laura Saba - laura.saba@ucdenver.edu; Morton Printz - mprintz@ucsd.edu; Pam Flodman - pflozman@uci.edu; Colin Hodgkinson - ch32z@nih.gov; David Goldman - davidgoldman@mail.nih.gov; George Koob - gkoob@scripps.edu; Heather N Richardson - hrichardson@cnr.umass.edu; Katerina Kechris - katerina.kechris@ucdenver.edu; Richard L Bell - ribell@iupui.edu; Norbert Hübner - nhuebner@mdc-berlin.de; Matthias Heinig - m.heinig@mdc-berlin.de; Michal Praveneč - praveneč@biomed.cas.cz; Jonathan Mangion - jonathan.mangion@eur.appliedbiosystems.com; Lucie Legault - lucie.legault@douglas.mcgill.ca; Maurice Dongier - maurice.dongier@douglas.mcgill.ca; Katherine M Conigrave - katec@med.usyd.edu.au; John B Whitfield - John.Whitfield@qimr.edu.au; John Saunders - mail@jbsaunders.net; Bridget Grant - bgrant@mail.nih.gov; Paula L Hoffman - paula.hoffman@ucdenver.edu; WHO/ISBRA Study on State and Trait Markers of Alcoholism - boris.tabakoff@ucdenver.edu

* Corresponding author

Published: 27 October 2009

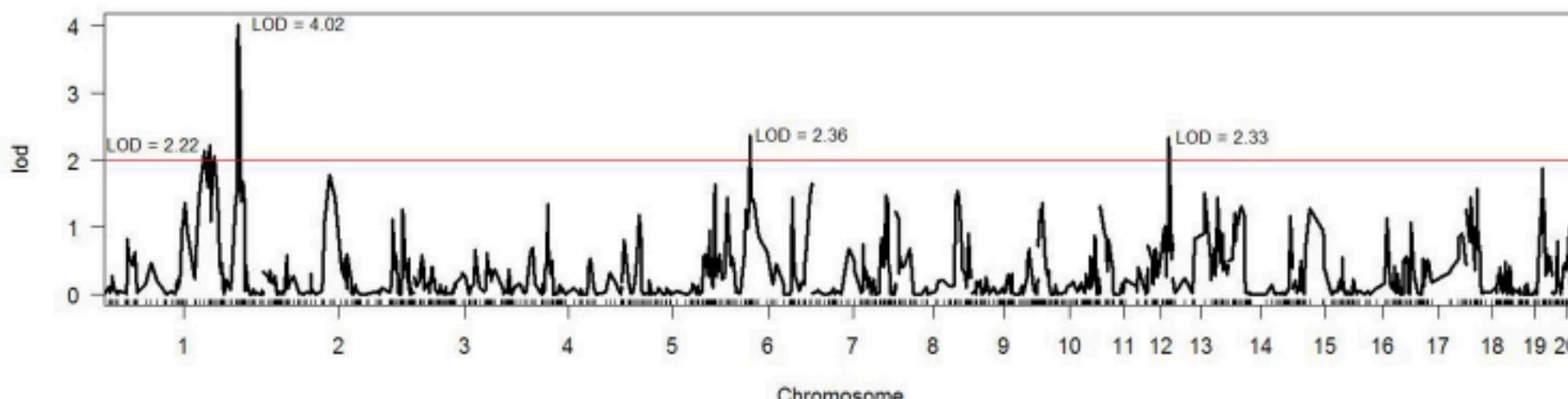
Received: 29 September 2009

Accepted: 27 October 2009

This article is available from: <http://www.biomedcentral.com/1741-7007/7/70>

© 2009 Tabakoff et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



QTLs for alcohol consumption:

Chr 1: 159-194 Mb
Chr 1: 213-238 Mb
Chr 6: 33-53 Mb
Chr 12: 25-46 Mb

Table I: Candidate Genes for Alcohol Consumption by HXB/BXH RI Rats.

Gene Symbol	Gene Description	Gene Location Chr(Mb)	eQTL Location Chr(Mb)	eQTL Mb Range	eQTL LOD (p-value)	bQTL Location Chr(Mb)	Correlation Coefficient (p-value)	Broad Sense Heritability	Percent Present
Cckbr	cholecystokinin B receptor	I (163.17)	I (168.95)	139.73-179.97	6.3 (<0.001)	I (159-194)	0.61 (0.0023)	0.82	100%
Coq7	demethyl-Q 7 (Coenzyme q (ubiquinone) biosynthetic enzyme (DHPB methyltransferase) 7)	I (176.76)	I (177.64)	152.25-179.75	9.0 (<0.001)		0.55 (0.0059)	0.93	100%
Fgrf2	fibroblast growth factor receptor 2 isoform c	I (189.48)	I (184.63)	173.38-184.63	20.7 (<0.001)		0.58 (0.0029)	0.99	93%
Ptpre	protein tyrosine phosphatase receptor type E	I (195.16)	I (195.13)	184.63-196.57	13.1 (<0.001)		-0.43 (0.0348)	0.95	100%
Abat	4-aminobutyrate aminotransferase	10 (7.04)	I (132.04)	12.84-249.11	4.2 (0.027)		-0.51 (0.0267)	0.56	100%
Tpst1	tyrosylprotein sulfotransferase 1 (predicted)	12 (27.59)	I (15.55)	9.02-205.67	3.1 (0.096)		0.51 (0.0265)	0.54	99%

Summary of pQTL to eQTL to gene

- The overlap of eQTL with pQTL presents a different aspect of how a causal DNA variant can influence a phenotype and can be used in conjunction with looking for genes physically located within the pQTL.
- Utilizing RNA expression information to link DNA variants to phenotype provides additional insight into possible mechanisms and relevant tissues.

Introduction to RNA co-expression networks

Why networks?

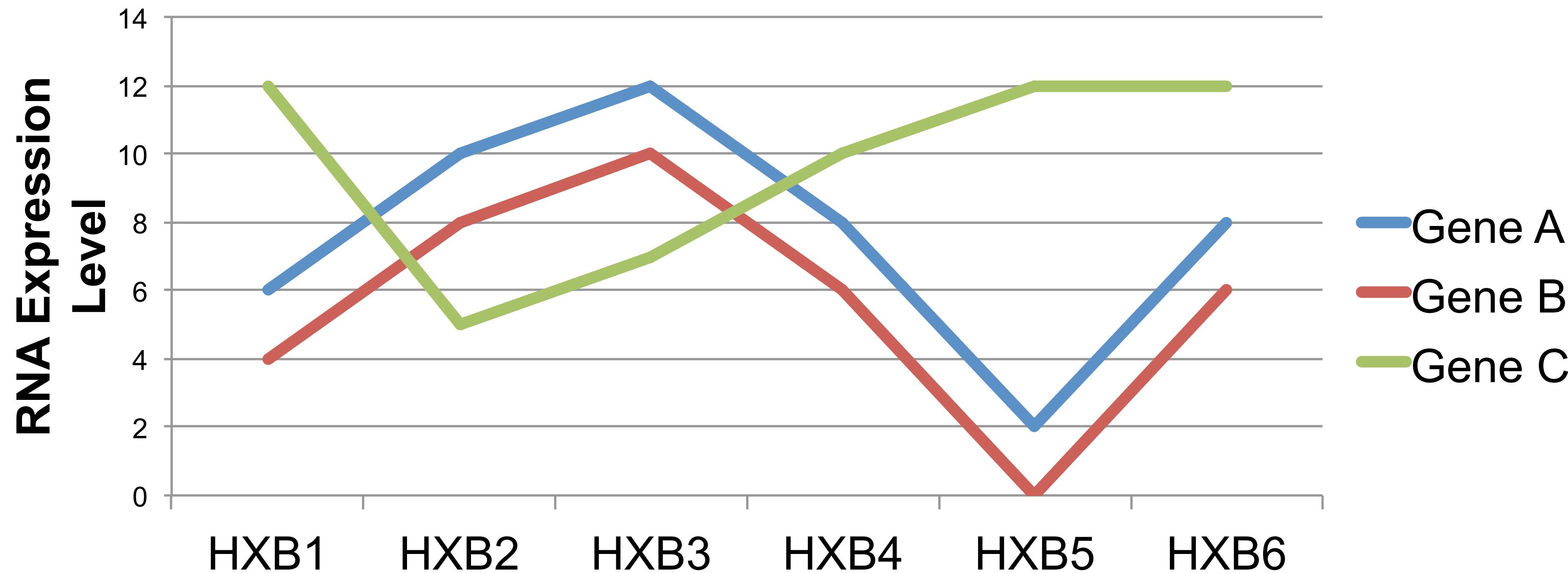
1. No gene product acts independently in the cell.
2. Most tissues, especially brain, are **complex hierarchical networks** that are spatio-temporally linked through structure and functions.
3. Many diseases, including substance use disorders, are often conceptualized as **failure in network regulation**.
4. The generation of a network **provides insight for understanding** predisposition to disease, etiology of organ or behavioral pathology, and response to medications or toxins.

Methods for defining networks of genes

- Protein-protein interactions
 - e.g., STRING database (<https://string-db.org/>), bioGRID (<https://thebiogrid.org/>)
- Annotated pathways/gene ontology terms
 - e.g., KEGG Pathways (<https://www.genome.jp/kegg/>), PANTHER Pathways (<http://www.pantherdb.org/>), Gene Ontology (<http://geneontology.org/>)
- RNA co-expression
 - e.g., Weighted Gene Co-Expression Network Analysis, k-means clustering, Bayesian Networks, Gaussian graphical models

Co-expression as a measure of “connection”

Theory - if the magnitude of RNA expression of two transcripts correlates over multiple “environments” (genomes), then the two transcripts are involved in similar biological processes.



What do we gain by building networks and identifying co-expression modules instead of considering each candidate gene individually?

- Inferred biological function of a gene from other co-expressed genes
 - Helpful for under-annotated or unannotated genes
- Context in which the gene exerts its effect
 - Helpful for genes with multiple functions within the cell
- Identification of multiple therapeutic targets within the same pathway
 - Helpful for finding druggable targets
- Implicate cell types or regions of interest in heterogeneous tissues

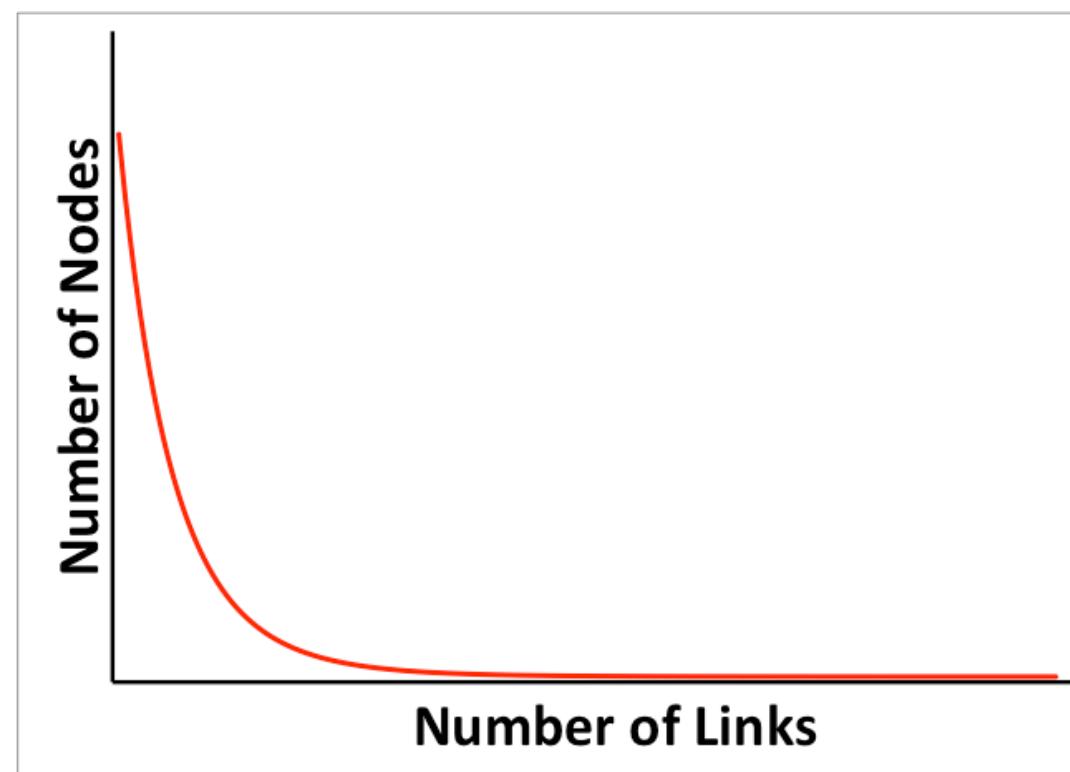
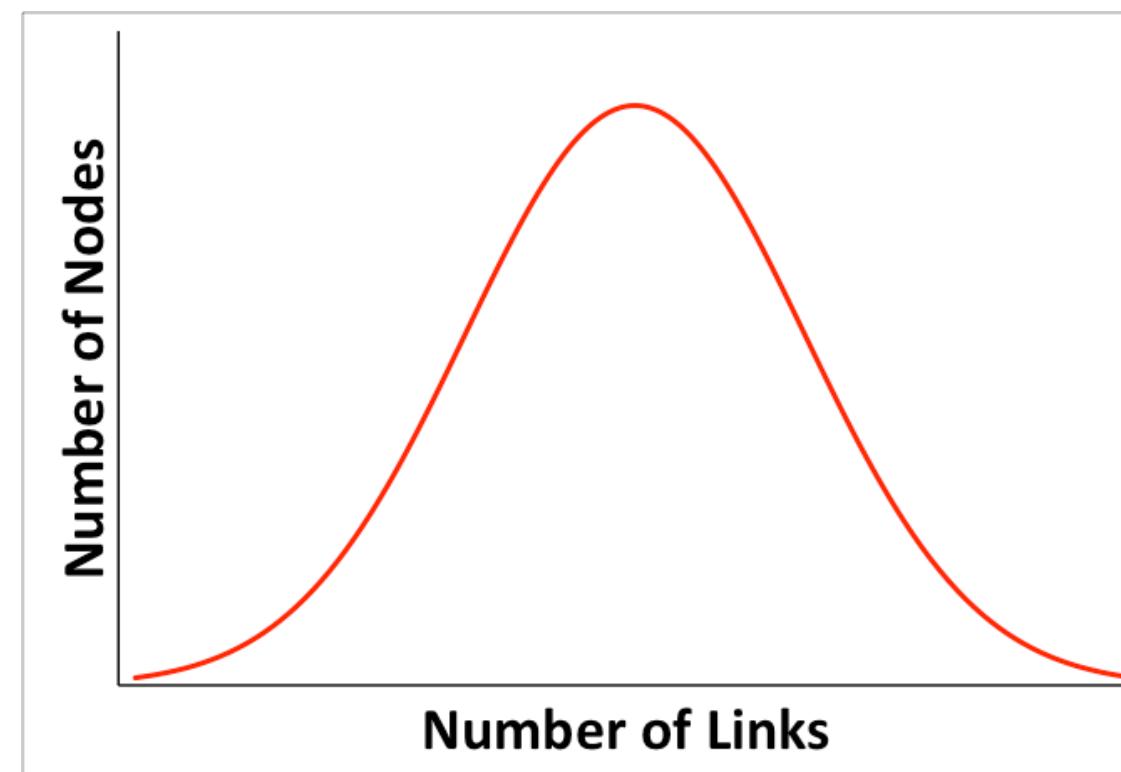
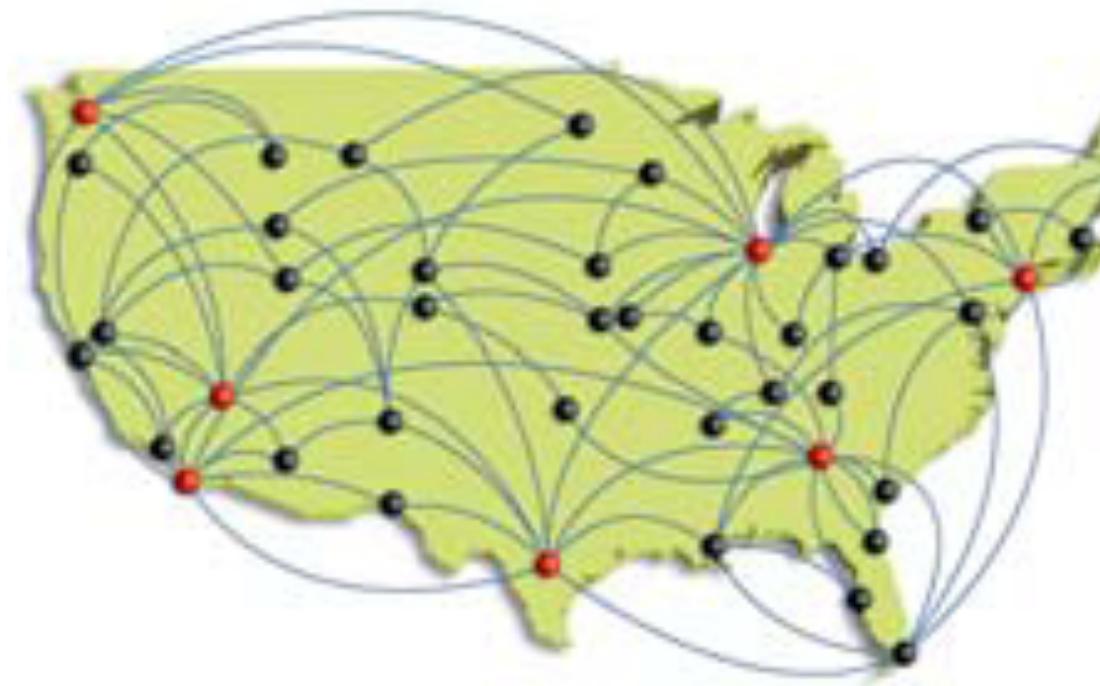
Weighted gene co-expression network analysis (WGCNA)

Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4

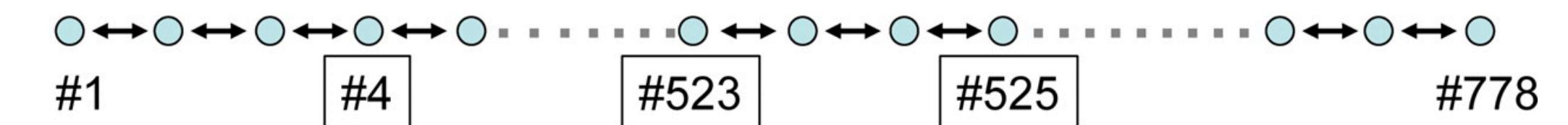
Why not just correlation?

1. Relationships between genes better described using a **scale-free network**
2. We can get a more **robust** measure of co-expression by including a measure of how many “friends” two genes have in common (indirect relationships)

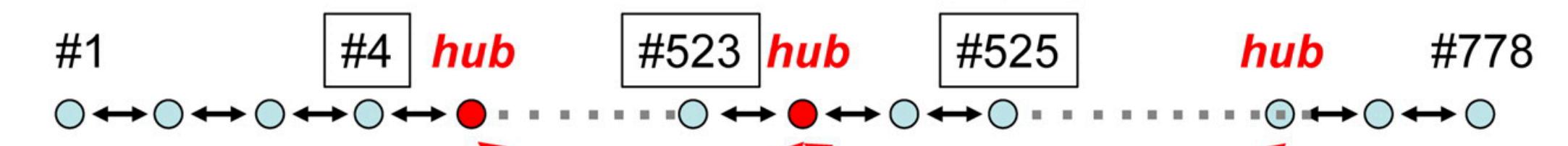
Scale-free network assumption



A Linear array



B Network with hubs

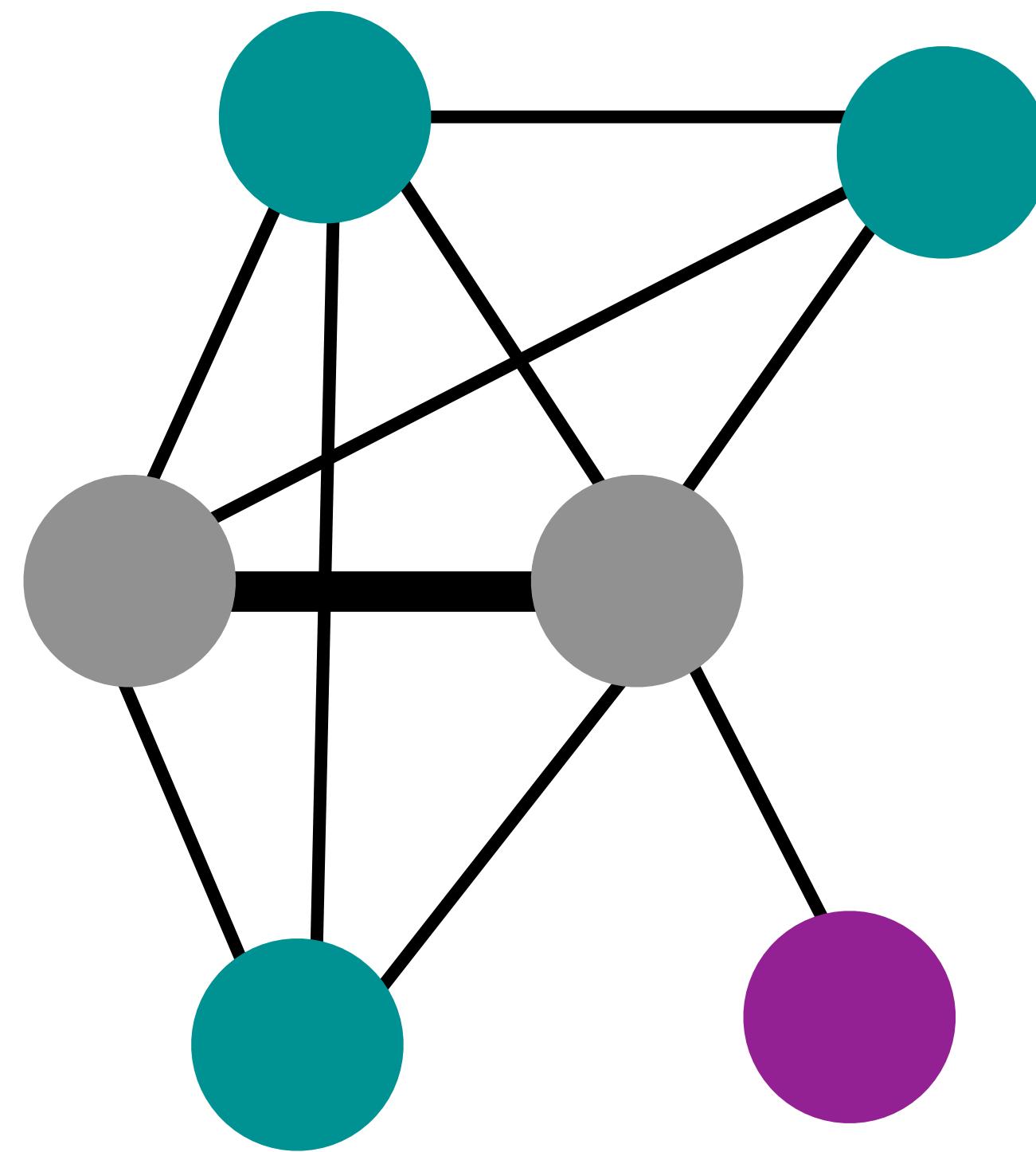


Small-world properties in metabolic networks. A, Hypothetical arrangement of E. coli's 778 metabolites in a linear chain. If No. 525 is uridine (required for DNA synthesis and replication), its synthesis from nearby precursor metabolites such as No. 523 is highly efficient, requiring only a few steps. However, if only distant precursors such as No. 4 are available, hundreds of steps are now required, making uridine synthesis slow, inefficient, and energetically costly. B, In contrast, highly connected hub nodes (red) in a scale-free network allow efficient conversion of either No. 523 or No. 4 to uridine in only a few steps.

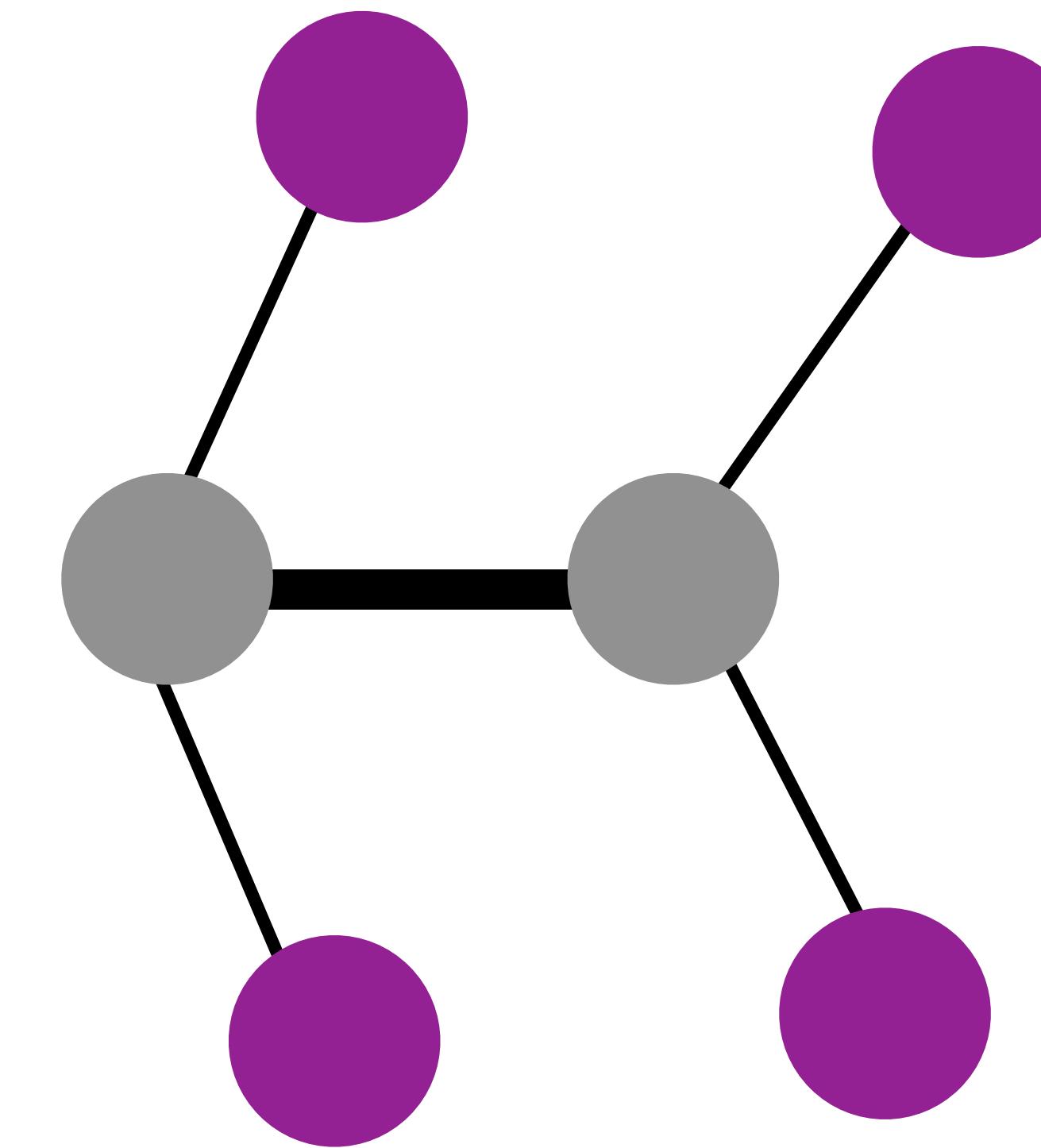
Weiss et al. Circulation Research. 2012;111:493-504

Images taken from *Scale-Free Networks*, Scientific American, May 2003

Integration of indirect and direct correlations



High Indirect Correlations

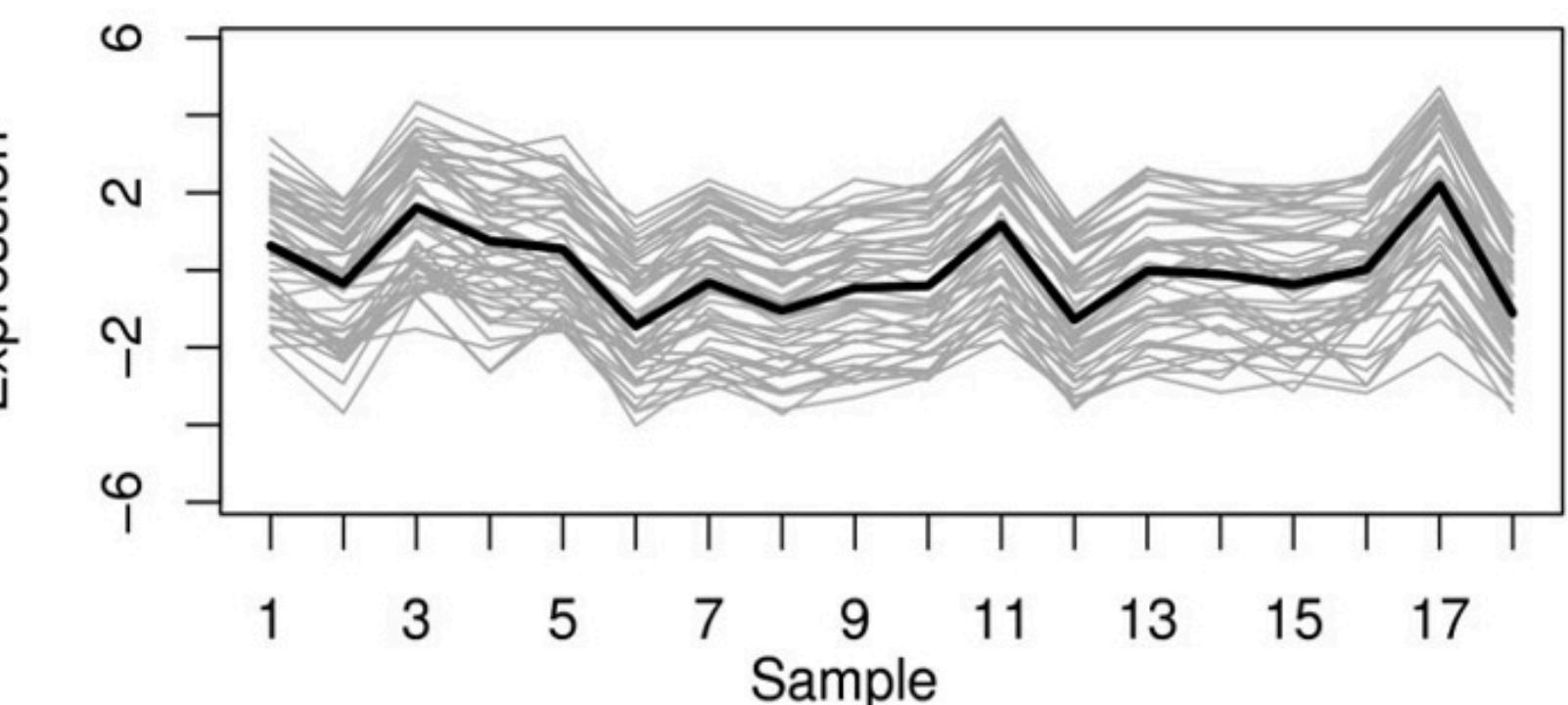


Low Indirect Correlations

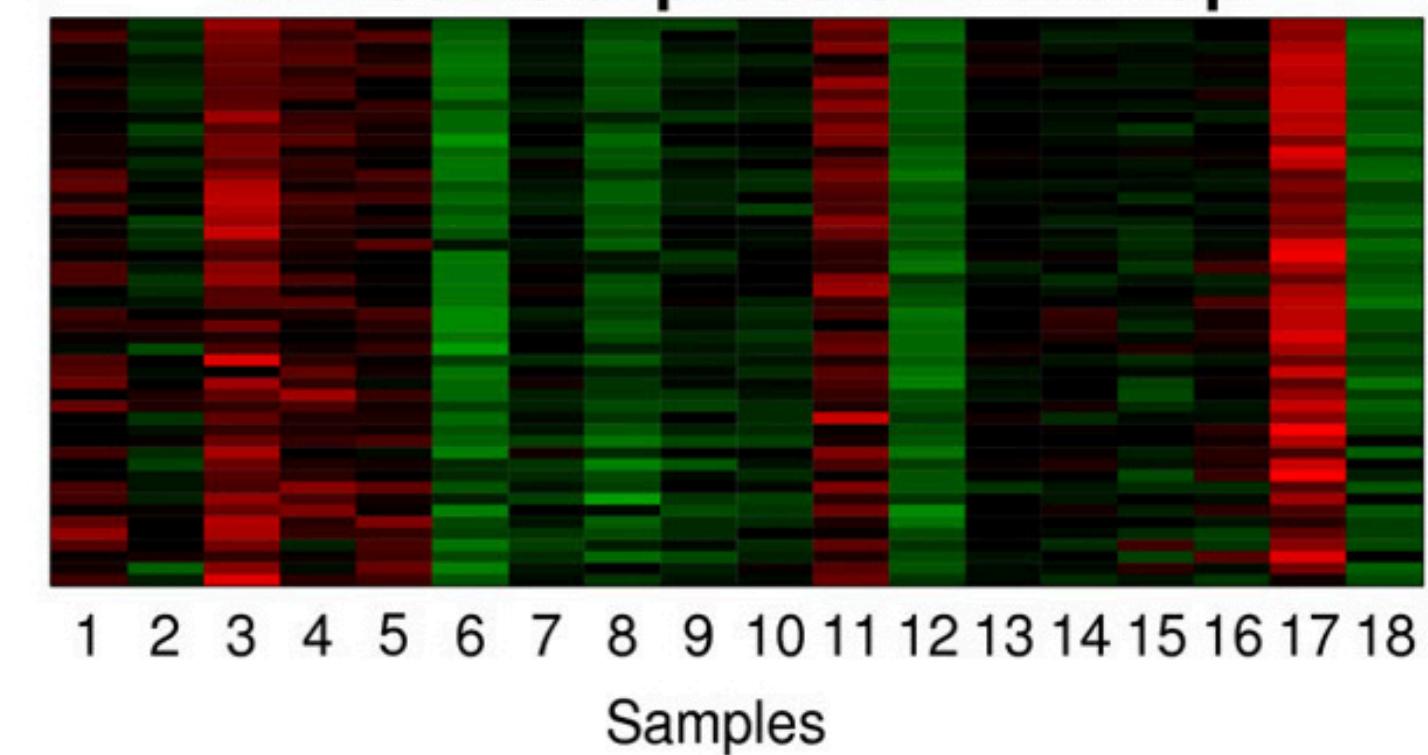
Module eigengenes

- Module eigengene is used to summarize the expression patterns of the module genes across samples
- First principal component from PCA
- Often the proportion of total variation explained by the module eigengene is used as a measure of robustness
- End results is one quantitative value per sample

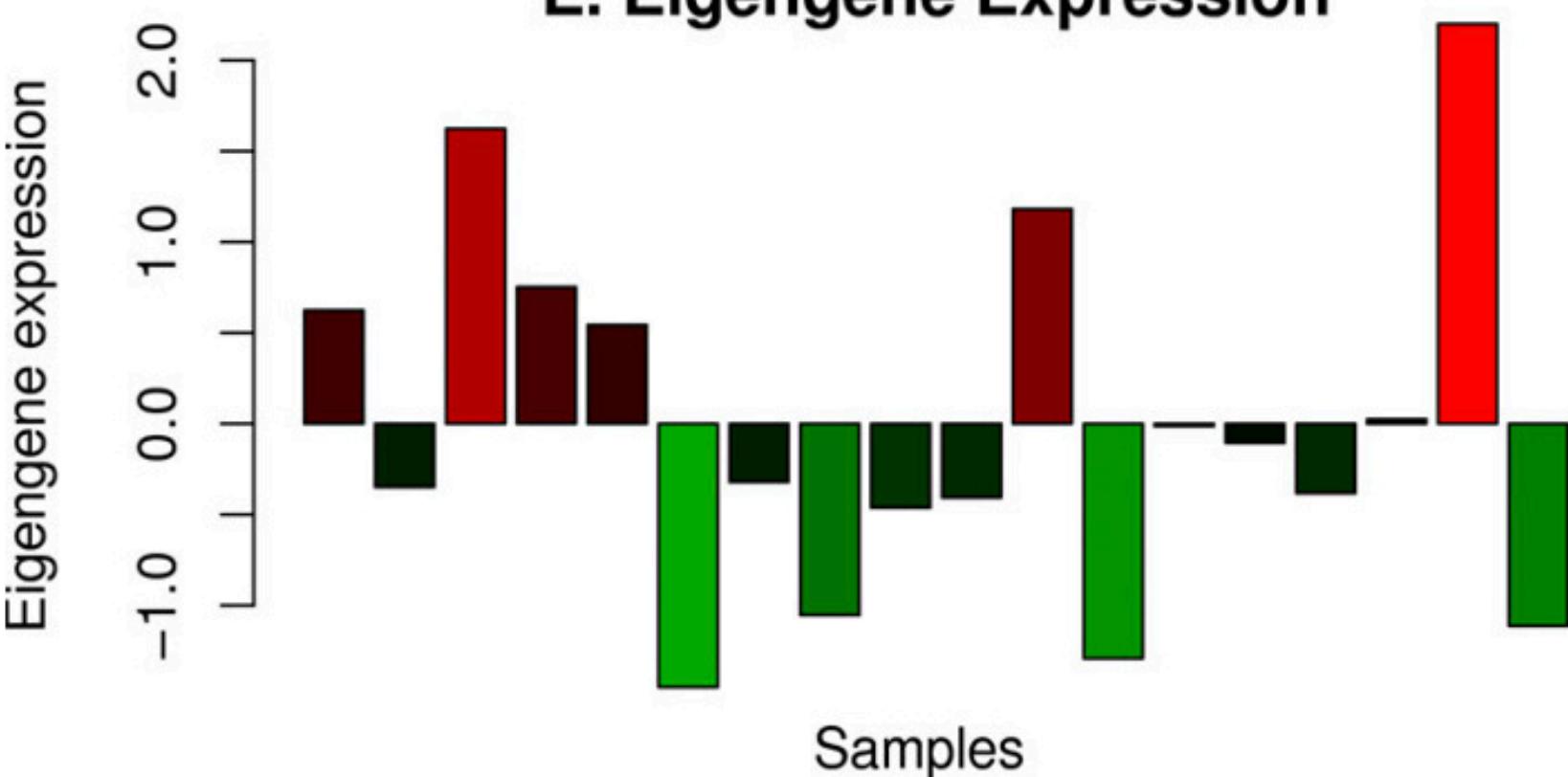
C. Expression levels of genes and eigengene



D. Module expression heatmap

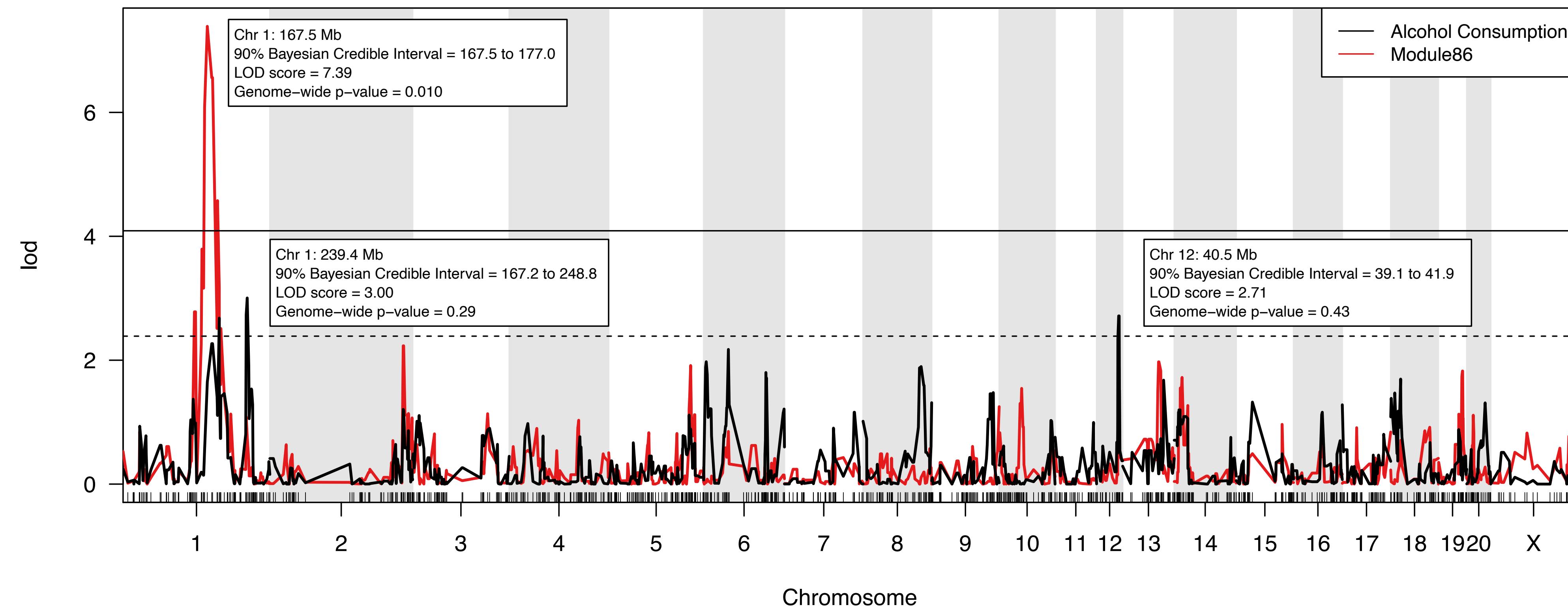


E. Eigengene Expression



Module eigengene QTL

Just like a behavioral or physiological phenotype, a module eigengene can be mapped to the genome using a QTL analysis.

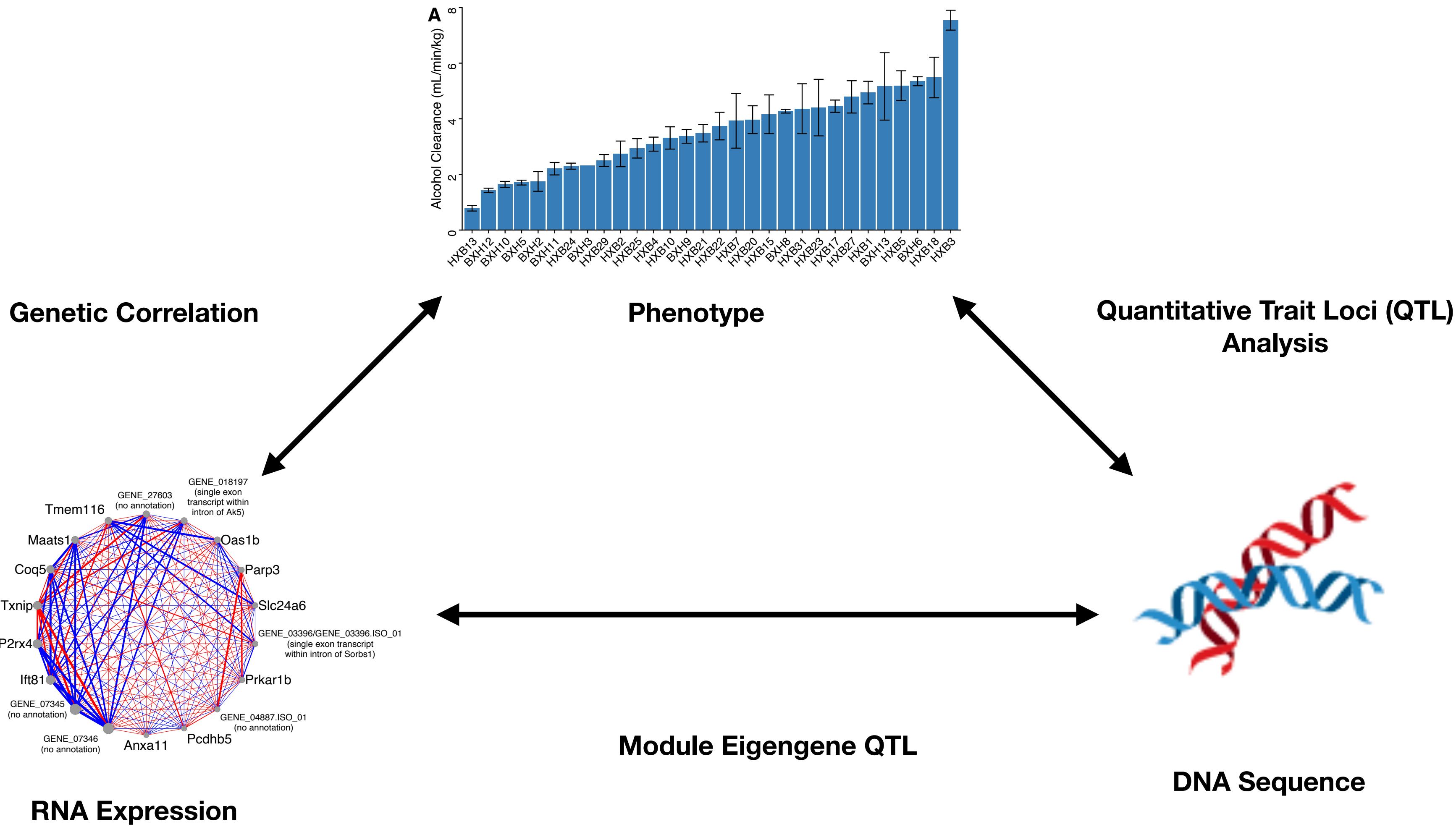


Summary of co-expression networks

- Networks can add context and insight into biological function of candidate genes.
- Co-expression is one way of mathematically describing a network. This data driven approach allows for the inclusion of under-annotated or unannotated RNA transcripts.
- WGCNA is just one method that can be used for this type of analysis.
- Like individual transcripts, the expression levels of a co-expression module can be mapped to the genome.

Linking co-expression networks to phenotypic QTL through module eQTL

Extension of GGP approach to co-expression networks



Example application of GGP with co-expression networks

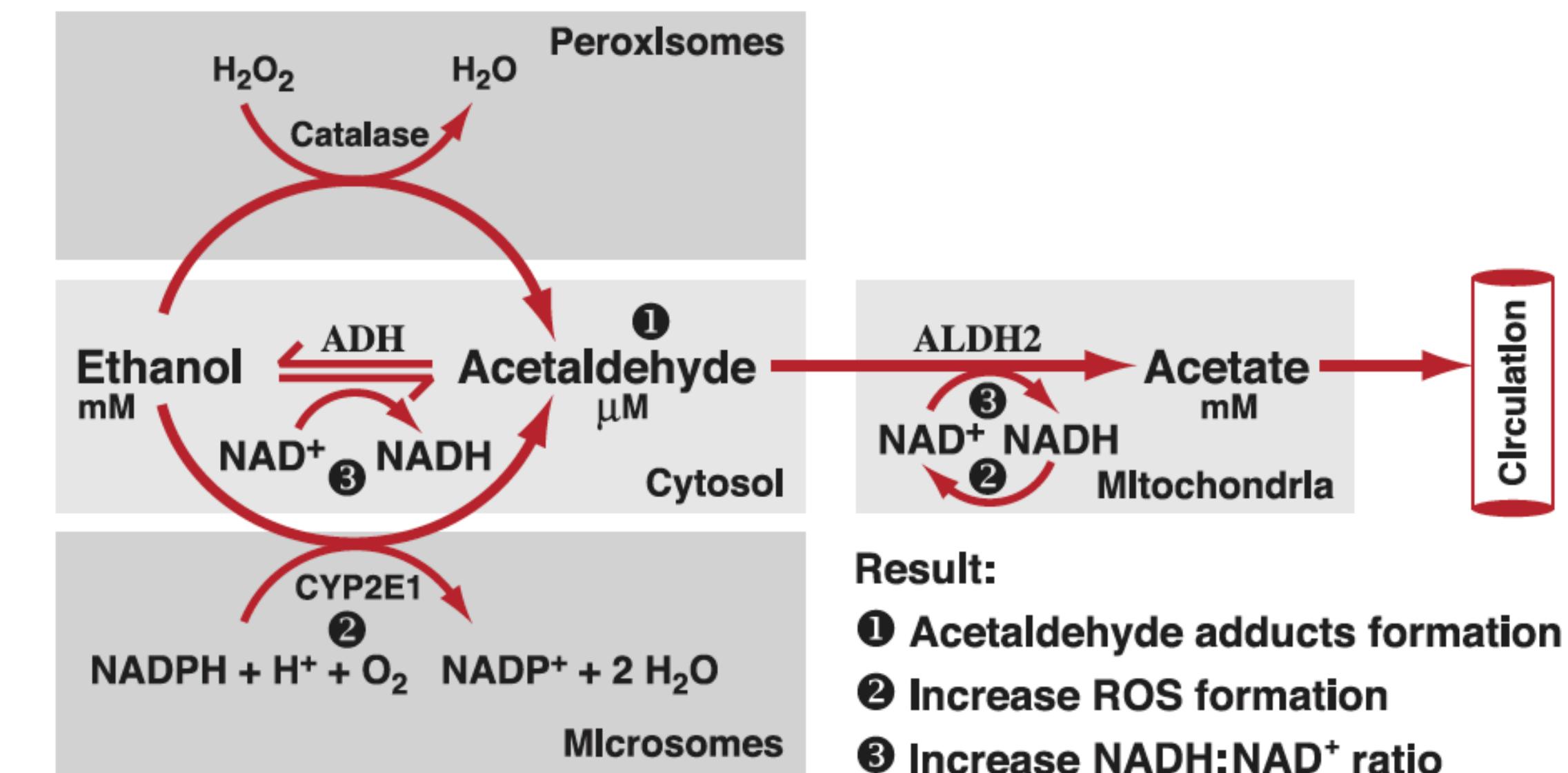
Alcohol Metabolism and Liver Expression

ALCOHOLISM: CLINICAL AND EXPERIMENTAL RESEARCH

Vol. 42, No. 7
July 2018

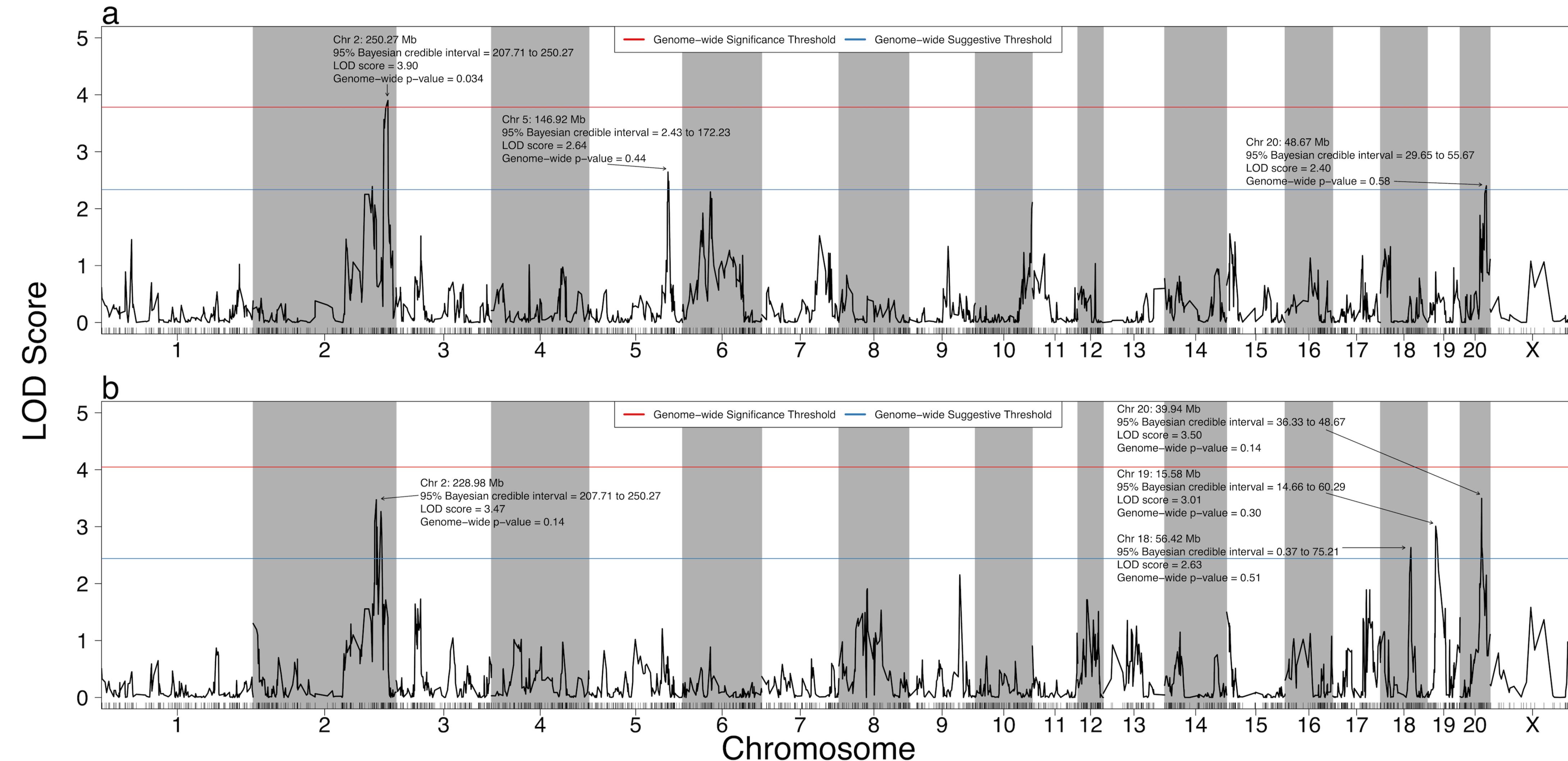
Unsupervised, Statistically Based Systems Biology Approach for Unraveling the Genetics of Complex Traits: A Demonstration with Ethanol Metabolism

Ryan Lusk, Laura M. Saba, Lauren A. Vanderlinden, Vaclav Zidek, Jan Silhavy,
Michal Pravenec, Paula L. Hoffman, and Boris Tabakoff 



Zakhari S. Overview: how is alcohol metabolized by the body? Alcohol Res Health. 2006;29(4):245-54.

Alcohol Clearance and Circulating Acetate Levels QTL



Alcohol
Clearance

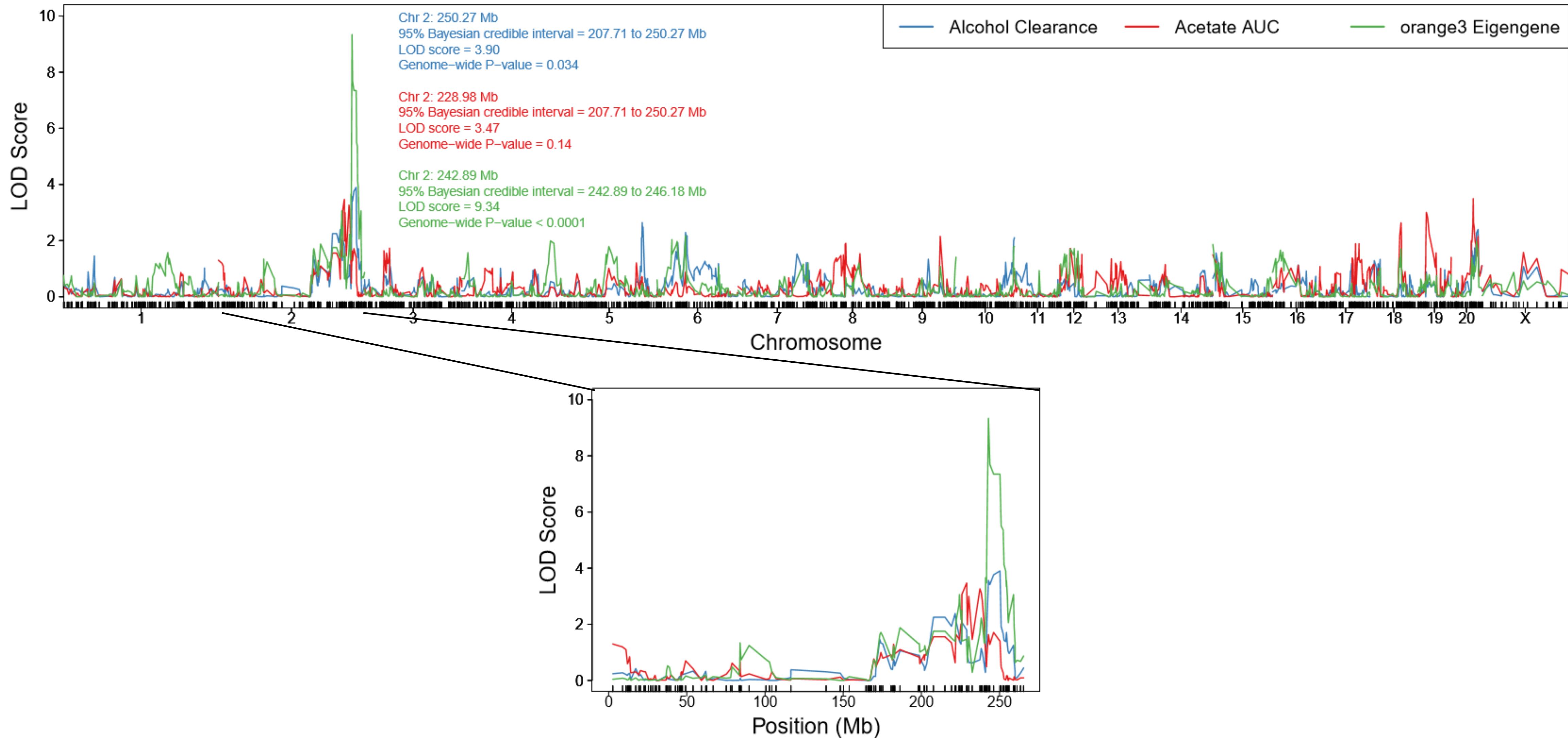
Circulating
Acetate
Levels

Association between measures of alcohol metabolism and co-expression networks

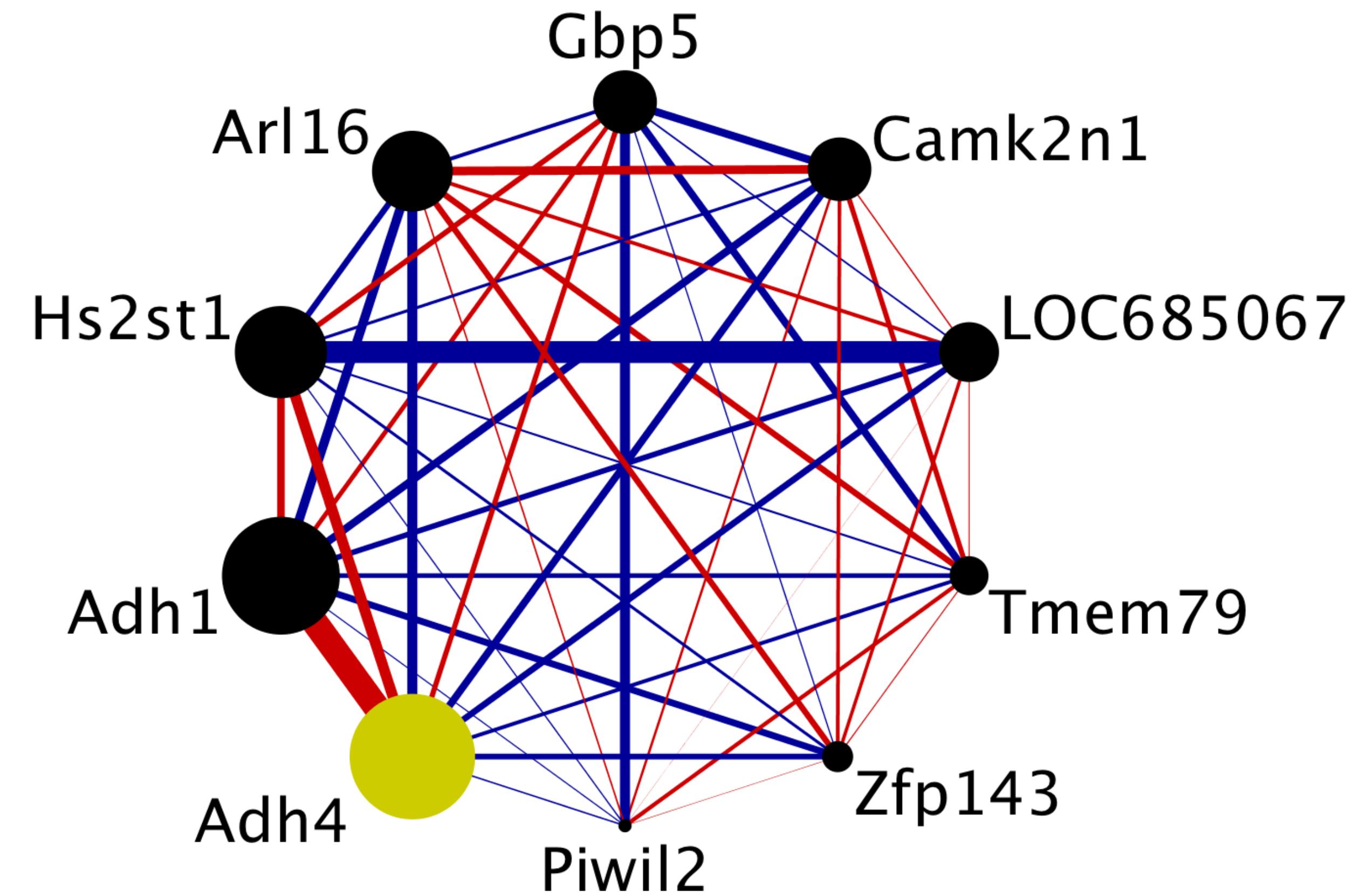
Module	Number of Genes in Module	Proportion of Variance in Module Explained By Eigengene	Correlation with Alcohol Clearance in HXB/BXH	
			Correlation Coefficient	p-value
orange3	10	0.60	-0.75	<0.001
darkslateblue.1	6	0.66	-0.56	0.0016
peachpuff3	9	0.59	-0.55	0.0018
palegreen1	8	0.62	-0.53	0.0030
maroon4	8	0.63	0.50	0.0055
lightcyan.1	6	0.65	-0.49	0.0065
mistyrose3.1	5	0.66	0.49	0.0075
palegreen3	8	0.66	0.49	0.0076
darkorange3	7	0.68	0.48	0.0084
goldenrod2	8	0.62	-0.48	0.0091

Module	Number of Genes in Module	Proportion of Variance in Module Explained By Eigengene	Correlation with Acetate AUC in HXB/BXH	
			Correlation Coefficient	p-value
lightslateblue	14	0.54	0.70	<0.001
sienna2.1	5	0.68	0.62	0.0003
bisque3.1	5	0.65	-0.58	0.0009
orange3	10	0.60	-0.57	0.0014
antiquewhite4	22	0.55	0.51	0.0050
goldenrod2	8	0.62	-0.51	0.0051
cadetblue4	7	0.70	0.50	0.0063
maroon4	8	0.63	0.49	0.0073
deepskyblue2	8	0.64	0.47	0.0095
cadetblue	8	0.62	-0.47	0.0100

Overlap of phenotypic QTL and meQTL



Candidate Co-expression Module for Alcohol Clearance



Summary of pQTL to meQTL to network

- Similar to the genetical genomics/phenomics approach that was used to identify candidate genes from the convergences of DNA, RNA, and phenotypic data, co-expression modules can be used in a similar manner.
- The ‘unsupervised’ approach was used to evaluate alcohol metabolism related phenotypes and liver RNA expression and the resulting network contained alcohol dehydrogenase, a well-known alcohol metabolism enzyme.

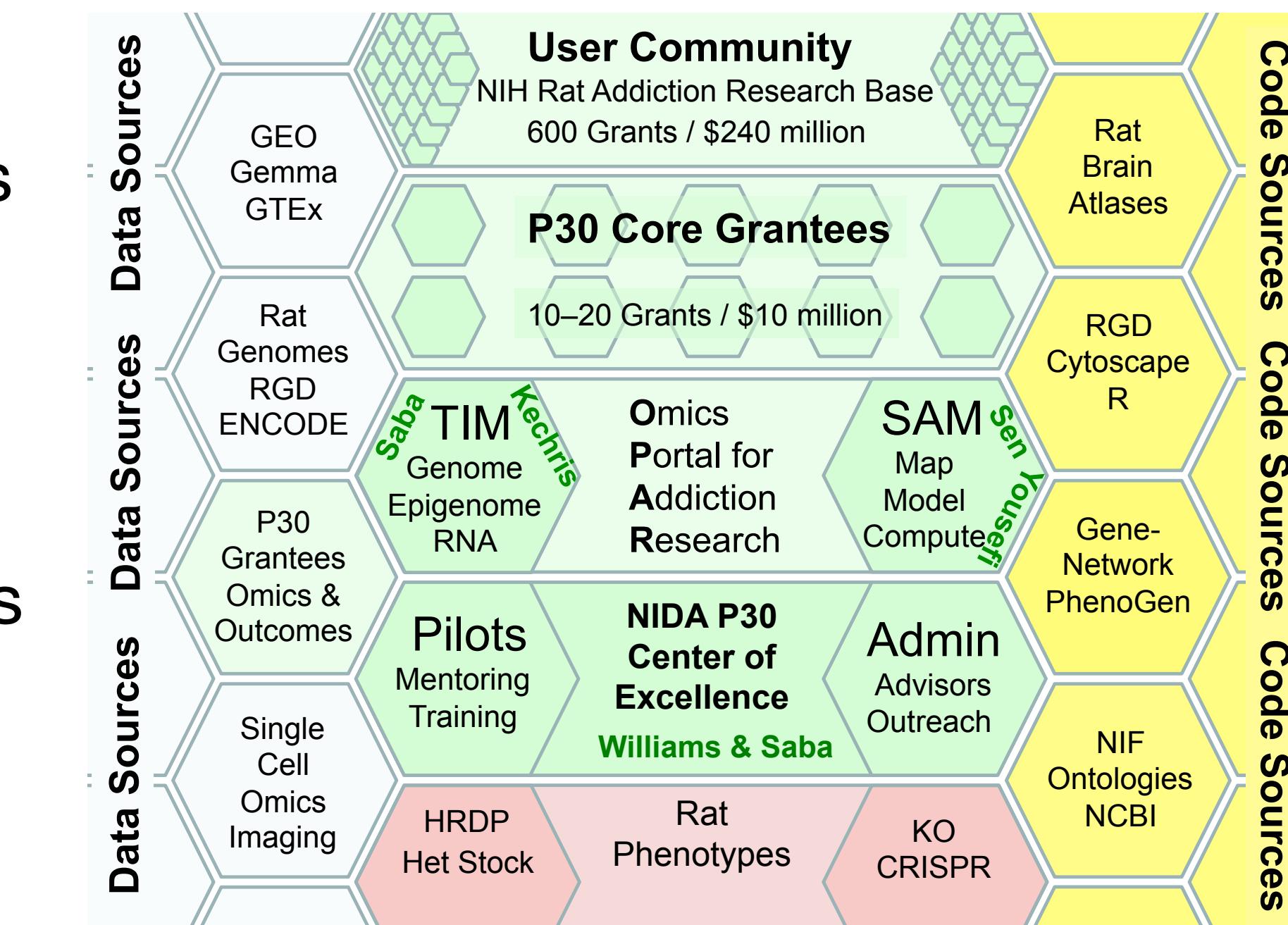
Conclusions

- RNA expression is a natural mediator between DNA variants and behavioral and physiological phenotypes.
- Incorporating knowledge about variation in RNA expression into our search for the biological mechanism underlying a phenotypic QTL can better capture genes that influence the phenotype in a quantitative fashion.
- Co-expression networks and other types of networks provide additional information about biological context and help identify possible functional roles of unannotated or under annotated genes.

NIDA Core Center of Excellence in Omics, System Genetics, and the Addictome

Co-Directors: Rob Williams (UTHSC) and Laura Saba (CU-AMC)

The **purpose** of the NIDA P30 Core Center of Excellence in Omics, Systems Genetics, and the Addictome is to empower and train researchers supported by NIH, NIDA, NIAAA, and other federal and state institutions to use more quantitative and testable ways to analyze genetic, epigenetic, and the environmental factors that influence drug abuse risk and treatment.



Our Approach:

- Omics Portal for Addiction Research (OPAR)
- Study design and RNA-Seq analysis services
- Training in Systems Genetics, RNA-Seq, and OPAR usage
- Funding for pilot grants

Acknowledgements

- Saba Lab:
 - Ryan Lusk, Lauren Vanderlinden, Harry Smith (former member), Sean Hickey, and Cheyret Wood
- Boris Tabakoff, Paula Hoffman, and their lab
 - Spencer Mahaffey and Jenny Mahaffey
- Financial Support:
 - NIDA Core “Center of Excellence” in Omics, Systems Genetics and the Addictome (NIDA - P30DA044223; MPIs - Williams, Saba)
 - The heritable transcriptome and alcoholism (NIAAA - R24AA013162; MPIs - Tabakoff, Hoffman, Saba)

