

Recombinant Inbred Rats: Genetics,
Transcriptomes and Use for
Identifying Phenotypic Determinants

(The Transcriptional Connectome)

**Boris Tabakoff, Laura Saba, Paula Hoffman and the
Phenogen Group, Anschutz Medical Campus, University of
Colorado Denver**

Acknowledgements

Collaborators:

Laura Saba, PhD; Stephen Flink, PhD; Lauren Vanderlinden, PhD; Paula L. Hoffman, PhD; Morton Printz, PhD; Michal Pravenec, PhD

Technical Support:

Adam Chapman; Jenny Yu; Spencer Mahaffey; James Huntley; Laura Breen; Donna Moye

Grant Support:

National Institute on Alcohol Abuse and Alcoholism; the Banbury Fund (Robertson Family)

Website: phenogen.ucdenver.edu for data and details

Outline

- Why rats?
- Rat Hybrid Diversity Panel
- Why networks?
- Transcriptional Connectome
- Applications of network topology information
 - Candidate gene approach for toxicity
 - Predispositional analysis of genetic contributors to phenotype (susceptibility analysis)

Why Rat?

An animal model most utilized for pre-clinical medication development

- Toxicology (including environmental exposure to toxic substances)
- Anatomy
- Biochemistry
- Physiology
- Pharmacology
- Genetic information (constantly expanding)
- Outbred, inbred and recombinant inbred and HS populations
- Genetic differences in phenotype at many levels
 - Acetaminophen, CYP3A1 and CYP2E1

Rat Resources

1. **Truly “outbred” colonies**: Different polymorphic versions of (alleles) transcripts embedded in a highly variable genetic background (e.g., Hsd Hot: Holtzman SD).
2. **Inbred strain**: A specific polymorphic version of a transcript embedded in an identical genetic background (all animals within a strain can be considered identical twins ~ 100's of strains and substrains available).
3. **A panel of inbred strains**: Each strain can be considered as a genetically unique individual and the number of strains will represent the number of individuals in an experiment. A **heterogeneous stock** is derived from mating of inbred animals.
4. **Recombinant inbred (RI) strain panel**: A group of strains derived usually from two progenitor strains wherein the genomic structure of each is scrambled with the other by recombination and then alleles are fixed by inbreeding. This produces a situation where particular alleles can be examined on a collage of backgrounds composed of the genomes of the progenitors.
5. **Hybrid panel**: A combination of RI strains and inbred strains providing for examination of a particular allele on both a known collage of particular backgrounds, and alleles of the same gene on a variety of stable backgrounds.

2, 3, 4, 5 constitute renewable and genetically stable resources for data collection over long periods of time in particular age epochs.

Rat Hybrid Diversity Panel

The development of the Phenogen resource

1. 30 RI strain (HXB/BXH panel)
2. 30 inbred strains (chosen for genetic diversity)

Total 60 strains

Mapping power adequate

Correlational power excellent ($\alpha=0.05$ power = 80% $r^2 > 0.35$)

Power to detect transcript variation (advantage of using both inbred and recombinant)

Origin of HXB/BXH RI Panel

PD/Cub (outbred Wistar)
Spontaneous mutation
(Lx), create congenic on
BN/Cub

WKY $\xrightarrow{\text{selective breeding}}$ Spontaneously Hypertensive SHR/Kyoto
(inbred)

BN-Lx/Cub X SHR/Ola
(Prague)

HXB RI strains

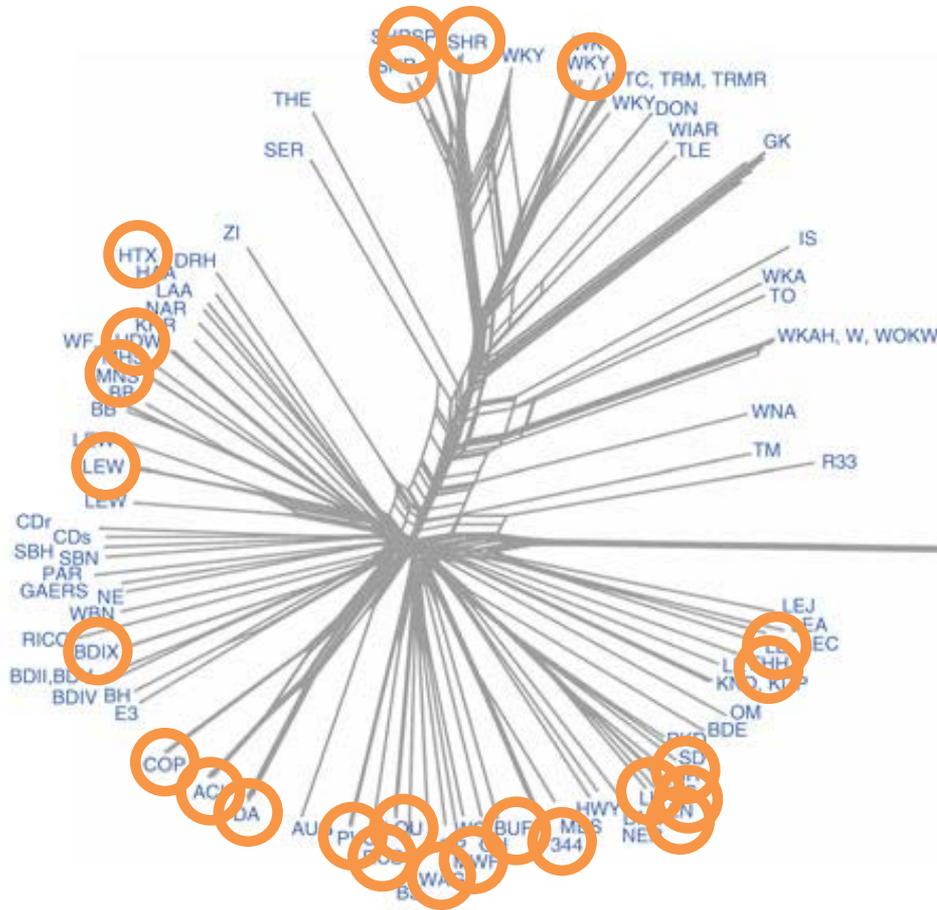
lpcv: 1, 2, 3, 4, 5, 7, 10,
13, 15, 17, 18, 20, 21, 22,
23, 24, 25, 27, 29, 31

BXH RI strains

Cub: 2, 3, 5, 6, 8, 9, 10,
11, 12, 13

- > 4,500,000 SNPs/indels between progenitors
 - Haplotype map generated

Genetic Relationship Between Inbred Strains



List of chosen strains

Strain	Source	Breeding Status (09/14)
LE/CpbHsd	Harlan	available
F344/NHsd	Harlan	available
BN/SsNOlaHsd	Harlan	available
SR/JrHsd	Harlan	available
SHRSP/A3NCrI	Charles River	available
LEW/CrI	Charles River	available
COP/CrCrI	Charles River	available
SS/JrHsd	Harlan	available
DA/OlaHsd	Harlan	available
WAG/RijHsd	Harlan	available
SHR/NCrI		available
SHR/OlaIpcv	Pravenec	available
WKY/NCrI		available
BN-Lx/Cub	Pravenec	available
PD/Cub		available
FHH/EurMcwiCrI	Charles River	cryopreserved
BDIX/HanHsd	Harlan	cryopreserved
LOU/CNimrOldHsd	Harlan	cryopreserved
PVG/OlaHsd	Harlan	cryopreserved
LUDW/OlaHsd	Harlan	cryopreserved
BUF/SimRijHsd	Harlan	cryopreserved
ACI/N	RRRC	cryopreserved
MWF/Ztm	RRRC	cryopreserved
RCS/N	RRRC	cryopreserved
MNS/N	RRRC	cryopreserved
LH/MavRrrc	RRRC	cryopreserved
LN/MavRrrc	RRRC	cryopreserved
HTX/HcjRrrc	RRRC	cryopreserved
M520/N	RRRC	cryopreserved
MR/N	RRRC	cryopreserved

Why Networks?

1. The brain is a complex hierarchical network spacio-temporally linked through structure and function.
2. Complex pathologic traits can be conceptualized as systems disorders of failed network regulation.
3. The brains structural and functional systems have features of complex networks that can be described through application of “graph theory” (small world topology).
4. Several canonical networks already described through use of fMRI (Resting state network, dorsal attention network, executive control network salience network).
5. Resting State connectivity can be a predictor of response to chemical and other environmental perturbations.

*****The generation of a “Transcriptional Connectome” representing the Resting State transcriptional networks provides power for understanding predisposition to disease, etiology of organ or behavioral pathology and response to medications or toxins.**

Why Study the RNA Dimension

Transcriptome links DNA and complex traits/diseases

- A. RNA is the first quantitative link between DNA sequence and phenotype (an endophenotype).
- B. Transcriptome information addresses part of the GWAS Gap: how does an identified DNA polymorphic locus contribute to disease?
- C. First step where DNA sequence and environment interact.
- D. Implementation of graph theory at the transcript level provides insight into genetic/environmental interactions that are the basis for susceptibility to complex diseases.

Goal for the PhenoGen Project

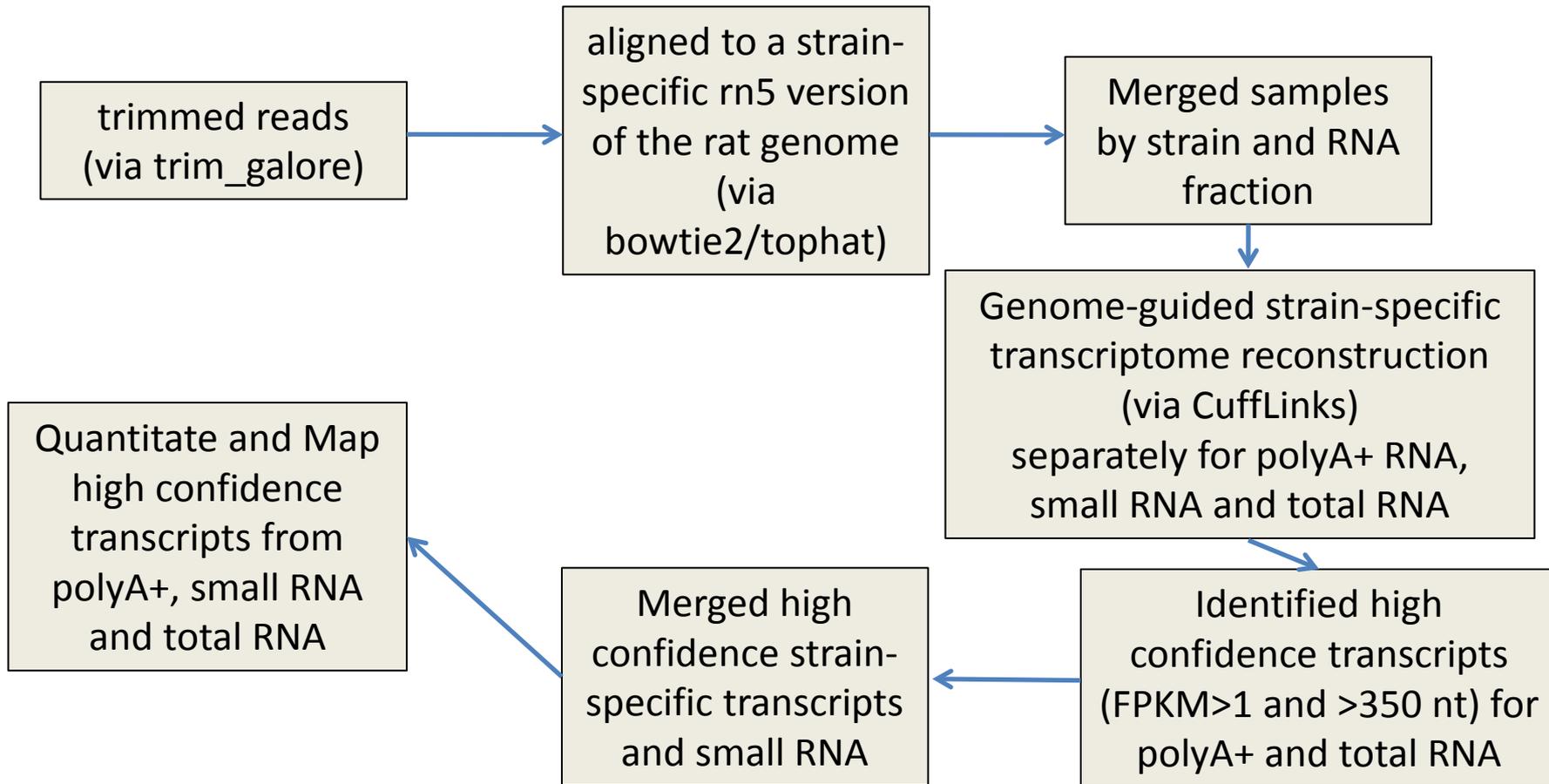
Transcriptional Connectome

To generate a new image of organs as networks of interacting elements (transcripts)

- Collect genome sequence and full transcriptome information for organs (brain, liver, heart).
 - We have completed exon array analysis for these three organs
 - RNA-Seq for brain and liver in 30 strains of the HXB/BXH panel completed
 - RNA-Seq for heart of the RI progenitor strains completed

Processing of RNA-Seq Data

RNA from Ribosome Depleted PolyA+ RNA fraction and total RNA fraction



FPKM = fragments per kilobase per million reads

Weighted Gene Coexpression Network Analysis

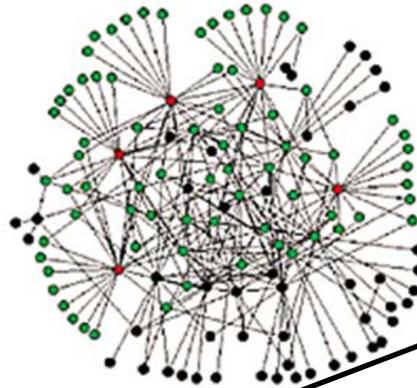
Construct a weighted gene co-expression network

Make use of interaction patterns between transcripts; correlate expression of each transcript with every other transcript and model a scale-free network



Identify Modules

Pathway-based analysis: modules are clusters of highly interconnected transcripts (nodes), corresponding to highly correlated gene expression. Minimum module size = 5 transcripts



Find Module QTLs

Module eigengene corresponds to first principal component of a module and can be considered representative of gene expression in the module. Determine chromosomal location (QTL, SNP) associated with module expression profile (eigengene).



Relate Modules to External Information

Utilize GO or other ontology information
Find biologically interesting modules; correlation with phenotype of interest



Use Module Membership, QTLs and correlations to prioritize interesting modules

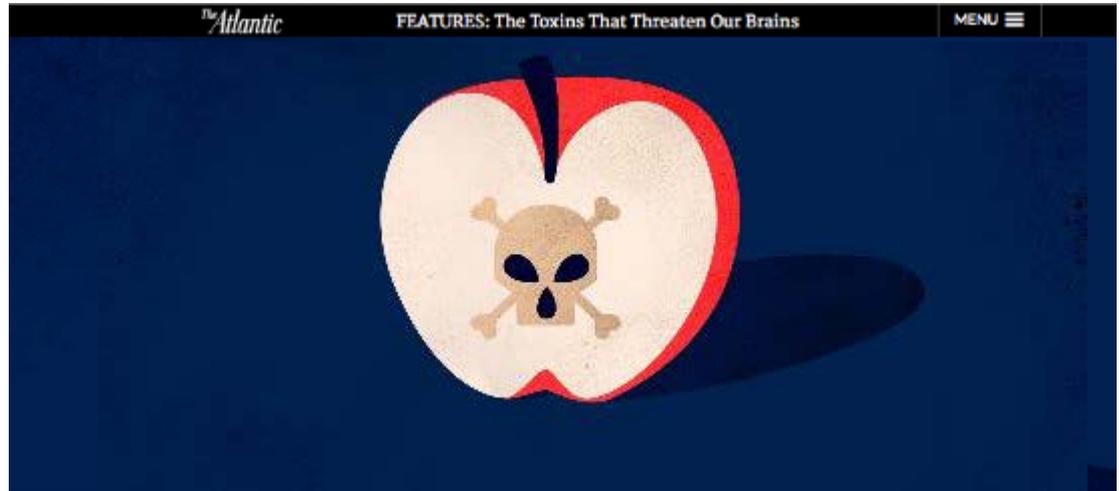
Co-expression as a measure of the “connectome”

- **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
- Caveats when multiple environments are multiple genetic backgrounds (false positive correlations)
 - Linkage Disequilibrium – two genes are physically located near one another in the genome or the loci that control expression of two genes are located near one another in the genome
 - Environment-dependent correlation
 - Cell-type mixing proportions - in heterogeneous tissue, differences in the composition of cell types within a sample can present as correlations between transcripts that are cell type specific (actually this is an informative caveat-relating module to particular cell type and/or tissue location)

A Toxicologic Example for Using PhenoGen Data

Start with gene product of interest and learn about its network partners and pathways.

e.g., [Lead Poisoning](#)



The Toxins That Threaten Our Brains

Leading scientists recently identified a dozen chemicals as being responsible for widespread behavioral and cognitive problems. But the scope of the chemical dangers in our environment is likely even greater. Why children and the poor are most susceptible to neurotoxic exposure that may be costing the U.S. billions of dollars and immeasurable peace of mind.

By James Hamblin

Illustrations by Jackie Lay

MARCH 18, 2014

Characteristics of Lead Neurotoxicity in Children and Adults

- **Morphologic Effects of Lead Exposure**
 1. Disruption of neuronal migration/differentiation
 2. Interference with synapse formation
 3. Aberant differentiation of glial cells
- **Cognitive/Behavioral Outcomes**
 1. Dose related decrease in intellectual ability (lower overall cognitive and intelligence scores)
 2. Impaired language skills (verbal concept formation, grammatical reasoning, command following)
 3. Impaired fine motor coordination
 4. Links to epilepsy, schizophrenia, autism
 5. Visual hyperresponsivity (increased proliferation of retinal progenitor cells)

One Target for Lead During Pre- and Post-natal Neural Development

Perinatal Lead Exposure Alters the Expression of Neuronal Nitric Oxide Synthase in Rat Brain

C. S. Chetty,¹ G. R. Reddy,¹ K. S. Murthy,² J. Johnson,¹ K. Sajwan,¹ and D. Desai³

¹*Savannah State University, Savannah, Georgia, USA*

²*Medical College of Virginia, Richmond, Virginia, USA*

³*University of Mississippi Medical Center, Jackson, Mississippi, USA*

International Journal of Toxicology, 20:113–120, 2001



ELSEVIER

Neuroscience Letters 236 (1997) 75–78

Neuroscience
Letters

The nitric oxide synthase expression of rat cortical and hippocampal neurons changes after early lead exposure

Asia Selvín-Testa*, Francisco Capani, C. Fabián Loidl, Jorge Pecci-Saavedra

*Instituto de Biología Celular y Neurociencias 'Dr. Eduardo De Robertis', School of Medicine,
University of Buenos Aires, Paraguay 2155, Buenos Aires 1121, Argentina*

Received 23 June 1997; received in revised form 6 September 1997; accepted 26 September 1997

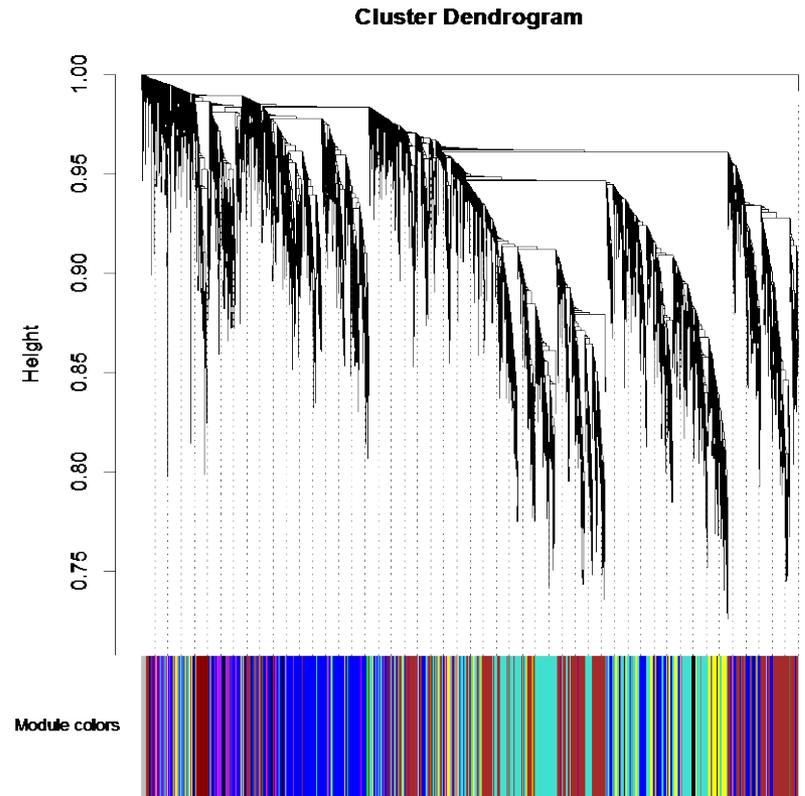
But how does this lead to the phenotype?

Collecting Data for WGCN Analysis

- RNA-Seq Gene Expression Analysis on 29 HXB/BXH Recombinant Inbred Strains and Parental Strains
 - 1 to 2 biological replicates per strain for RNA-Seq
 - Included parental strains in data processing (e.g., normalization), but not in WGCNA due to population structure
- 17,293 Ensembl genes (transcripts used for analysis)
 - 25,077 Ensembl genes were quantified using RSEM
 - Only Ensembl genes with more than 5 estimated read count in at least 50% of samples were retained

Description of Network

- 17,293 Ensembl genes
- 437 modules
- 1,520 genes not included in a module
- Median module size = 9 genes (range = 5 to 2141)



Co-expression Module for Nos1

- Pink Module
 - 351 genes
 - Hub gene – Capicua transcriptional repressor (Cic)
 - Other highly connected genes – proline-rich coiled-coil 2A (Prrc2a); syndecan-3 precursor (SDV3); BCL2-associated athanogene 6 (Bag6); N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 1 (Ndst1)
 - Nos1 is ranked 75th in intramodular connectivity

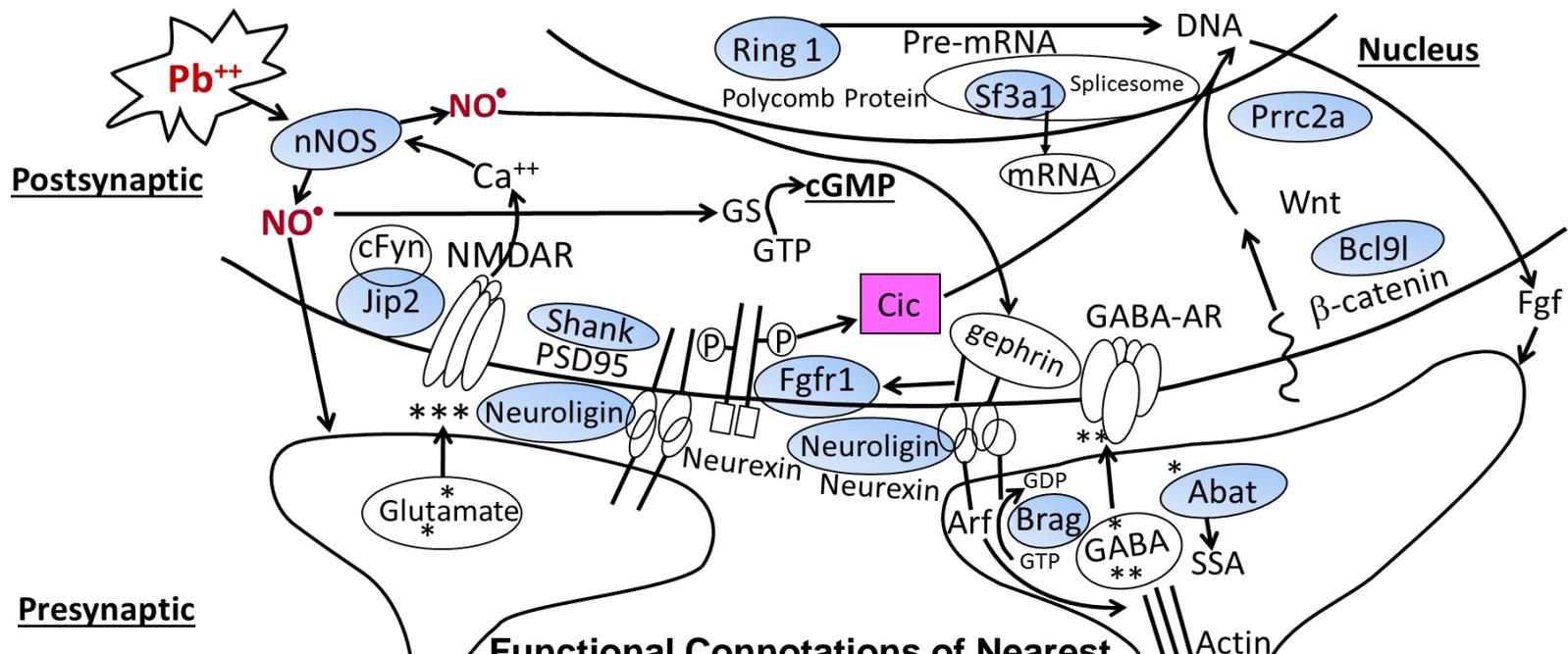
Functional Enrichment in Pink Module

- **Enriched Gene Ontology Categories (FDR<0.10 and fold enrichment > 1.5)**
 - **Molecular Functions**
 - **Adenyl-nucleotide exchange factor activity** (3 genes from pink; >5 fold enrichment)
 - **Transforming growth factor beta receptor, inhibitory cytoplasmic mediator activity** (2 genes from pink; >5 fold enrichment)
 - **Biological Processes (selected from top 20 relevant to brain)**
 - **Regulation of signaling** (80 genes from pink; 1.67 fold enrichment)
 - **Nervous system development** (60 genes from pink; 1.61 fold enrichment)
 - **Embryonic organ morphogenesis** (15 genes from pink; 2.83 fold enrichment)
 - **Social behavior** (6 genes from pink; >5 fold enrichment)
 - **Cellular Components**
 - **Photoreceptor inner segment** (6 genes from pink; >5 fold enrichment)
 - **Synapse** (29 genes from pink; 1.99 fold enrichment)
 - **Neuron part** (41 genes from pink; 1.66 fold enrichment)
- **PANTHER Pathways (FDR<0.10 and fold enrichment > 1.5)**
 - **Ionotropic glutamate receptor pathway** (5 genes from pink; >5 fold enrichment)

Nearest Neighbors of Nos1

Gene Symbol	Description	Correlation Coefficient
Nlgn2	neuroligin 2	0.90
Iqsec1	IQ motif and Sec7 domain 1	0.89
Mapk8ip2	mitogen-activated protein kinase 8 interacting protein 2	0.87
Fgfr1	fibroblast growth factor receptor 1	0.87
Prrc2a	proline-rich coiled-coil 2A	0.87
Abat	4-aminobutyrate aminotransferase	0.86
Sf3a1	splicing factor 3a, subunit 1	0.86
Bcl9l	B-cell CLL/lymphoma 9-like	0.86
Shank1	SH3 and multiple ankyrin repeat domains 1	0.86
Ring1	ring finger protein	0.86

Summary of Functions of Nearest Neighbors



Functional Connotations of Nearest Neighbors to NOS

1. Synaptic Adhesion and Anchoring Proteins
2. Growth factor signaling
3. Transcriptional control and splicing
4. GABA/Glutamate signaling
5. Stem cell proliferation/differentiation/synaptogenesis

Predisposition to Phenotype (Susceptibility)

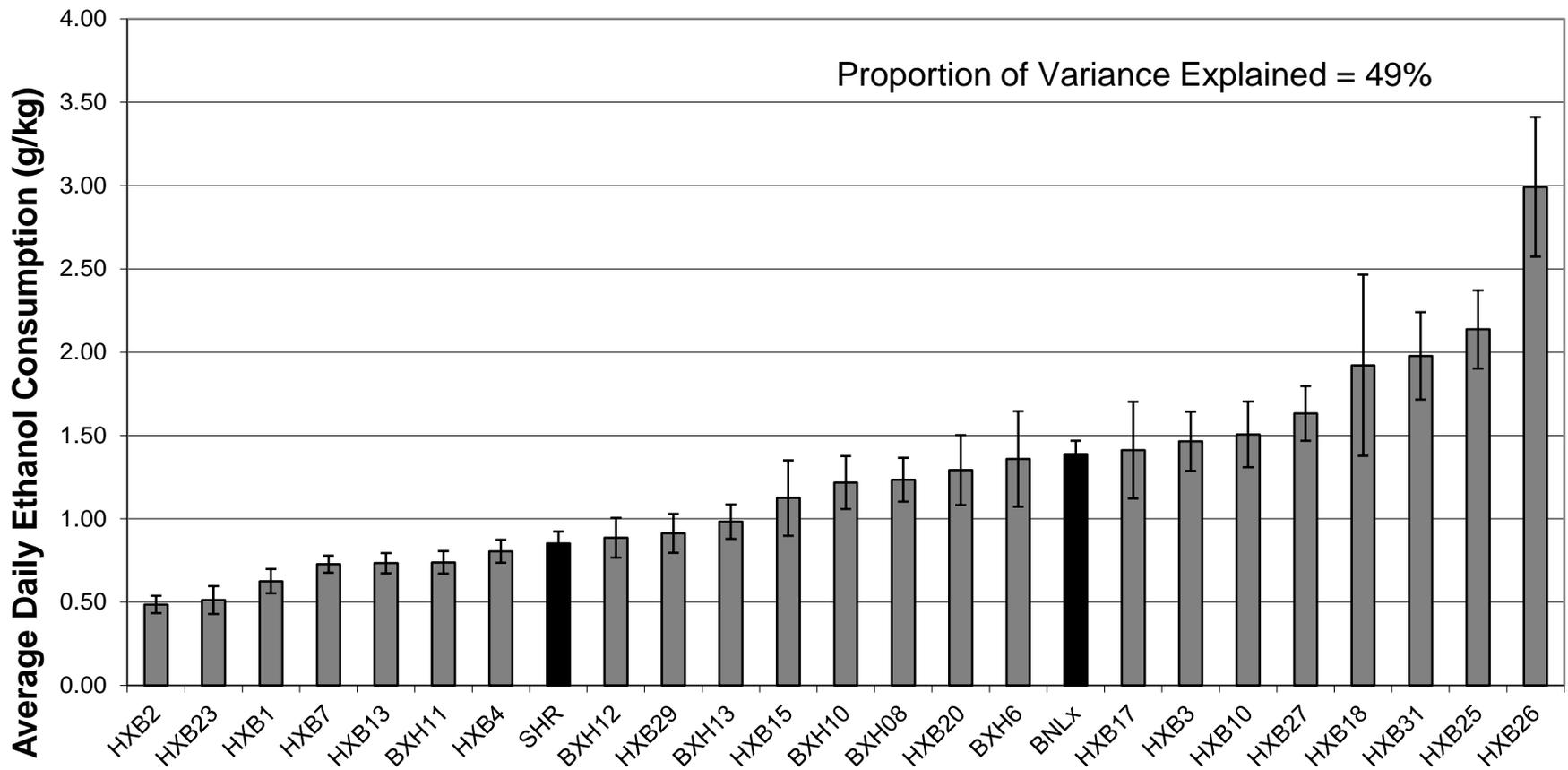
Genetics and Alcohol Consumption

- You can't be an alcoholic if you have never tried alcohol, i.e., it is an **etiologic essential**.
- Research shows a strong **genetic influence** on levels of alcohol consumption.
- Voluntary alcohol consumption is a **complex polygenic trait** that manifests through the complex interaction of many biological entities.

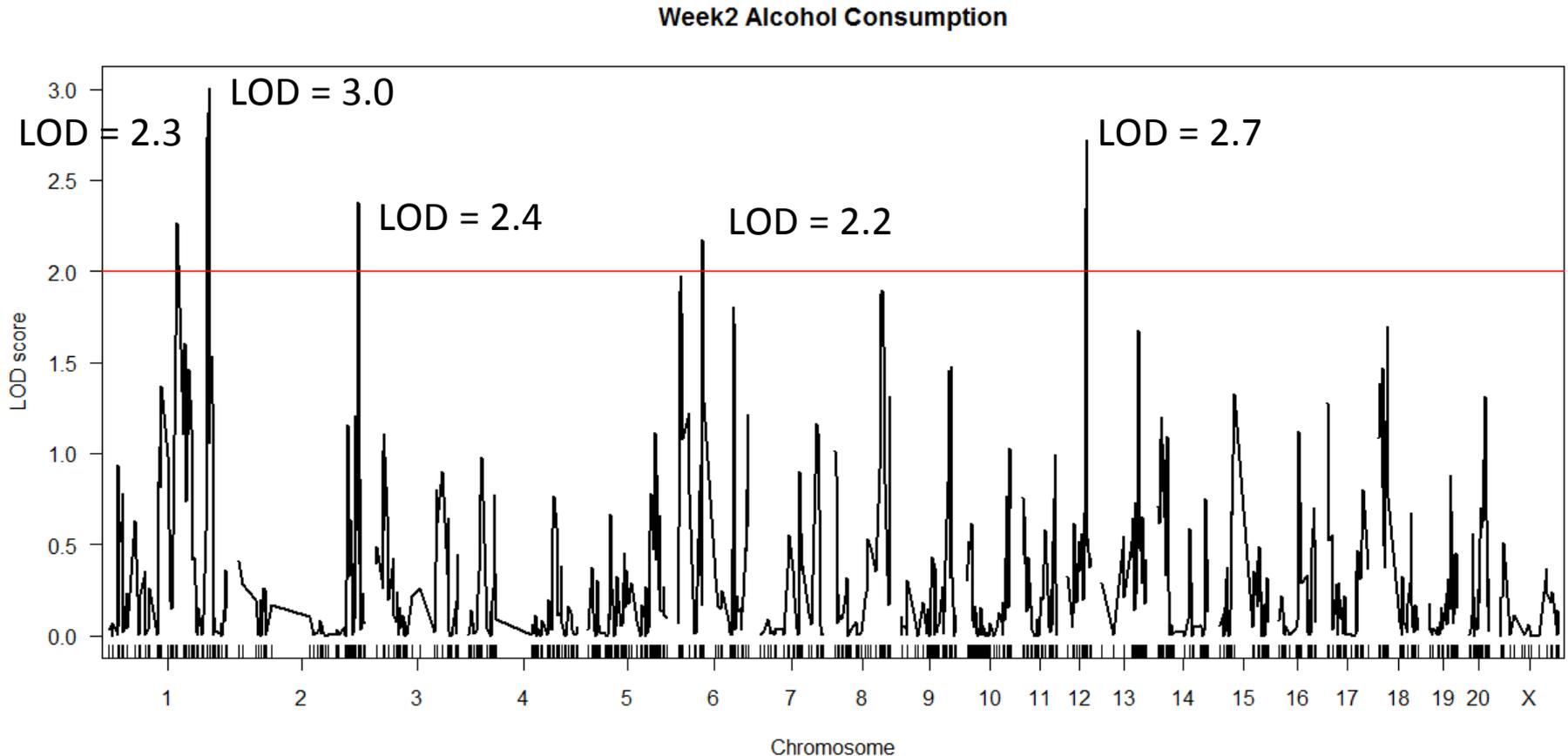


Distribution of Alcohol Consumption Across 25 HXB/BXH RI Strains

Strain Distribution of Average Daily Ethanol Consumption in Week 2



QTL Analysis for Alcohol Consumption in the HXB/BXH RI Rat Panel



Suggestive QTLs for alcohol consumption:

Chr 1: 167 – 187 Mb

Chr 1: 235 – 259 Mb

Chr 2: 256 – 276 Mb

Chr 6: 43 – 63 Mb

Chr12: 31 – 53 Mb

Quantify Transcription Levels of Genes and Individual Isoforms Expressed in Rat Brain Using Exon Array Information

Using RNA-Seq Data to “Clean” Hybridization Arrays

1. Eliminate probes/probe sets from Exon Array using DNA-Seq data

Out of 4.1M probes removed probes that:

Aligned to the rat genome (rn5) more than once or did not align to the rat genome at all (approx. 500K probes)

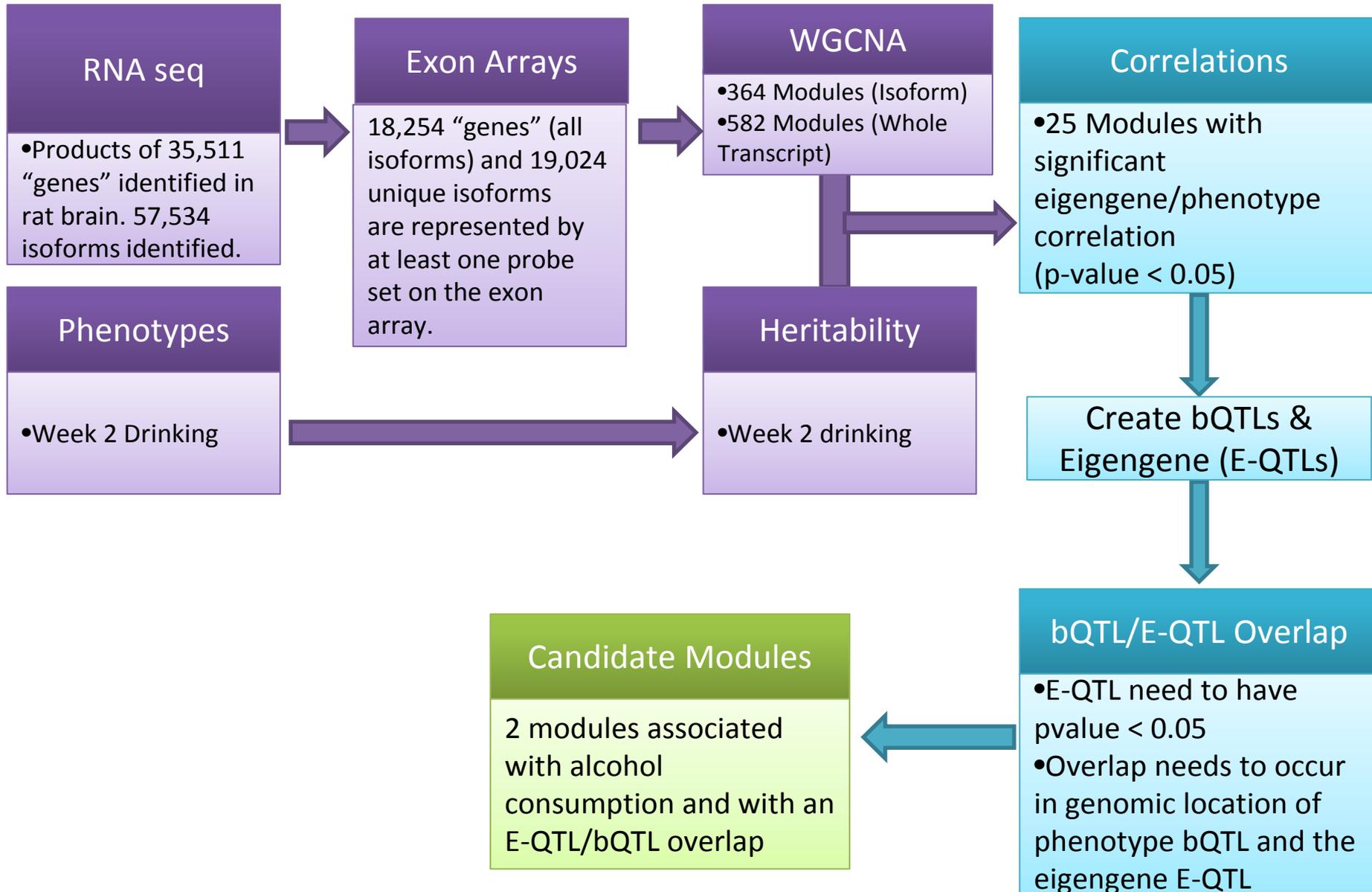
Targeted a region of the genome that harbors a known polymorphism between the parental inbred strains of HXB/BXH panel (approx. 100K probes)

Remove **probe sets** if less than 3 probes remain (approx. 200K)

2. Use RNA-Seq data to identify isoforms

3. Use RNA-Seq data to identify transcripts from unannotated regions using probes directed at “predicted” gene products

Flow Chart for Data Analysis to Discern Candidate Modules Predisposing to High or Low Alcohol Consumption

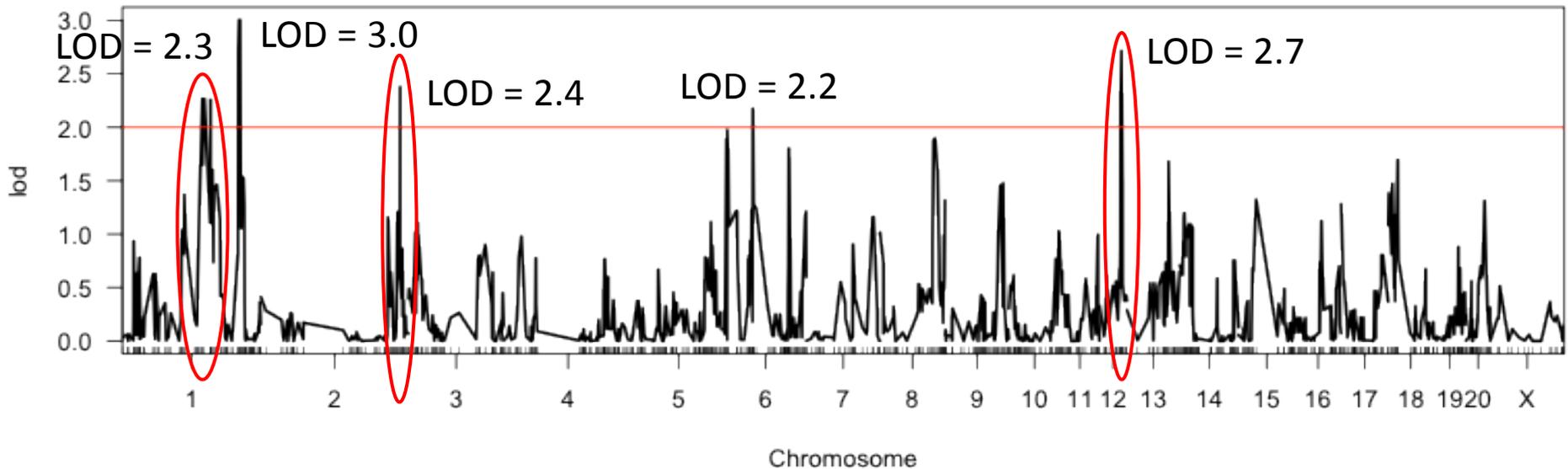


Indianred4 and Aquamarine1

Modules Have the Best Correlation of Eigengene with Drinking in the HXB/BXH Panel and E-QTL Overlaps the b-QTL for Alcohol Consumption

		Indianred4	Aquamarine1 (isoform-specific)
Number of Transcripts in Module		14	8
Proportion of Variance in Module Explained By Eigengene		0.59	0.61
Hub Gene	Gene Symbol	Brain.13.100	Tmem116
	Position	Chr12:40.9 Mb	Chr12:42.4 Mb
Module Eigengene QTL	Location [chromosome:Mb (95% confidence interval)]	Chr12:41.0-45.7 Chr2:266.4-267.6 Chr1:163.3-164.4 and 167.9-178.3	Chr12:42.4 (40.7-44.7)
Correlation with Drinking	Correlation Coefficient	-0.59	0.33
	P-value	0.005	0.145

bQTL Analysis for Alcohol Consumption in the HXB/BXH RI Rat Panel (A GWAS Analysis) and Overlap in Eigengene QTL and bQTL



b-QTLs for alcohol consumption (Mb):

Chr 1: 167-203

Chr 1: 235-259

Chr 2: 256-276

Chr 6: 43-63

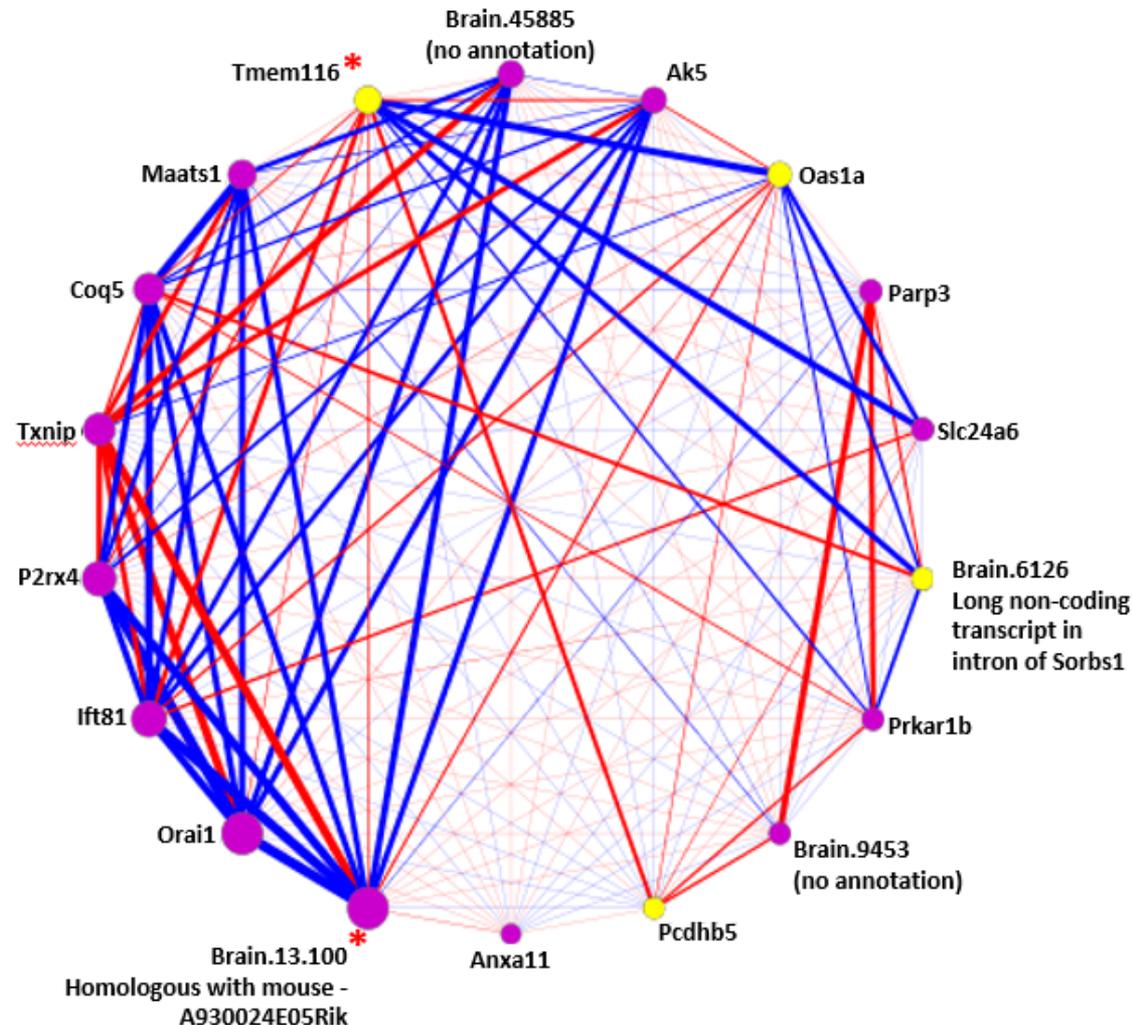
Chr 12: 31-52

E-QTL for [aquamarine1](#) module = Chr12:42.4 Mb
(from isoform-specific probes)

Circled b-QTLs are overlapped by
[Indianred4](#) E-QTLs

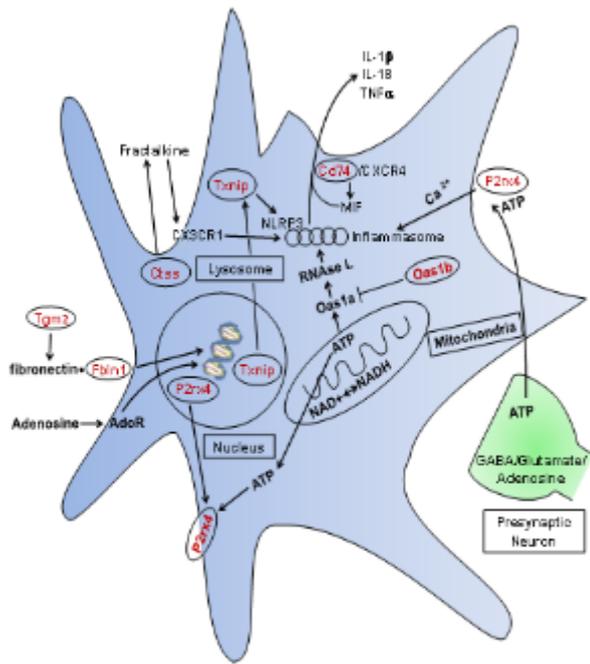
Connectivity Within and Between Members of the Indianred4 and Aquamarine1 Modules

- Edge thickness is weighted based on magnitude of correlation between nodes.
- Blue edges represent a positive correlation between nodes.
- Red edges represent a negative correlation between nodes.
- Node size is weighted based on connectivity within module.
- Yellow nodes were included in the module at the transcript and isoform level.



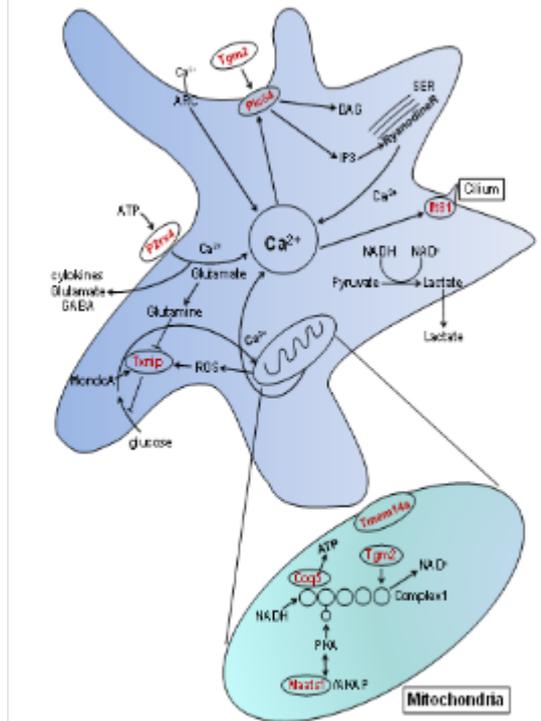
Biological Context from Pathway

Astrocyte / Microglia



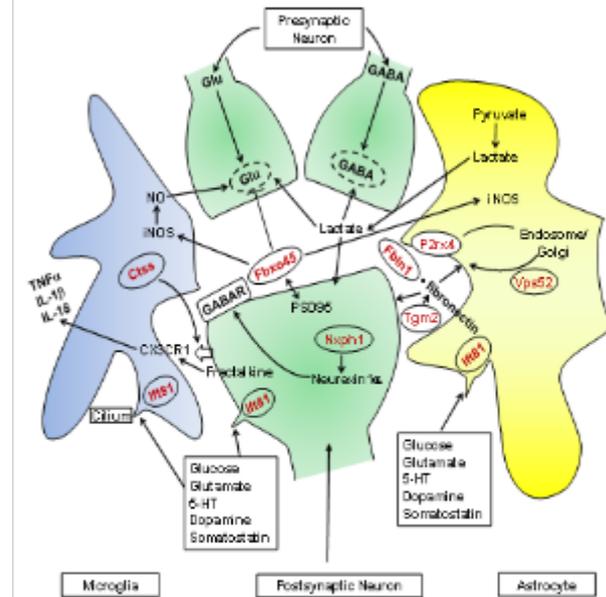
Inflammation / Immune Response

Astrocyte / Microglia



Energy / Ca²⁺ Homeostasis / Redox

Astrocyte / Microglia / Neurons



Glial/Neuronal Communication

Conclusions

- Living organisms are integrated networks of specialized function
- Graph theory is an approach to analyze and visualize interacting networks
- The graph theory analysis also allows one to analyze how a perturbation at one node of a network can perturb nodal relationships throughout a network
- Network analysis can also provide valuable information about an organisms predisposition to pathological consequences of a perturbation at any point in a network, OR a site (target) for intervention
- Network analysis can be applied at multiple levels of biologic function (molecular, electrophysiologic, imaging, etc.) and integrated between levels
- In all cases, one needs a proper population structure to apply network analysis
- **The Rat Hybrid Diversity Panel Provides an Excellent Opportunity for a Novel Approach to Apply Network Analysis to Understand Toxicology and Pathology**

http://phenogen.ucdenver.edu

The screenshot shows a web browser window with the URL phenogen.ucdenver.edu/PhenoGen/. The page header features the title "PhenoGen Informatics" and the tagline "The site for quantitative genetics of the transcriptome." Below the header is a navigation menu with the following items: Overview, Genome / Transcriptome Data Browser, Available Data Downloads, Microarray Analysis Tools, Gene List Analysis Tools, QTL Tools, About, Help, and Login/Register. The main content area includes a welcome message and a central graph with several nodes: Gene List Analysis, Pathway Analysis, Statistics / Expression Values, Exon Expression Correlations, and Microarray Analysis. A text box on the left explains that green nodes are accessible without login, while blue nodes require login. A "Compare/Share" section is visible at the bottom right, showing a "Demo/Screen Shots" interface with various comparison options.

PhenoGen Informatics
The site for quantitative genetics of the transcriptome.

Overview | Genome / Transcriptome Data Browser | Available Data Downloads | Microarray Analysis Tools | Gene List Analysis Tools | QTL Tools | About | Help | Login/Register

Welcome to PhenoGen Informatics
The site for quantitative genetics of the transcriptome.

Hover over or click on nodes in the graph below to see the tools/data available on the site.
Green no login required.
Blue sections require a login.
[Pause](#)

Pathway Analysis
Statistics / Expression Values
Exon Expression Correlations
Gene List Analysis
Microarray Analysis

Compare/Share
Demo/Screen Shots