

# **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**

# 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)
- 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 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 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)

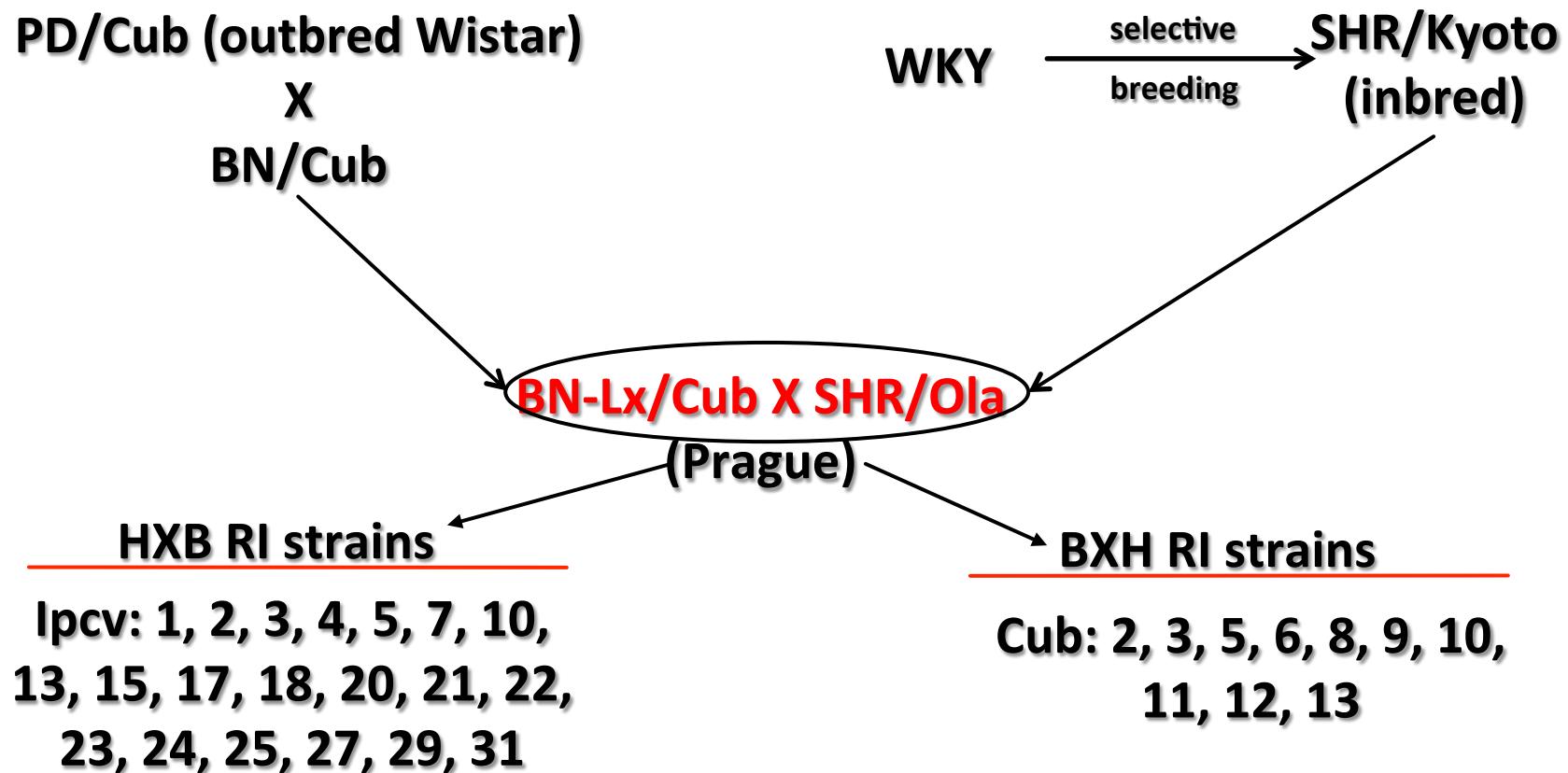
Total 60 strains

Mapping power adequate

Correlational power excellent

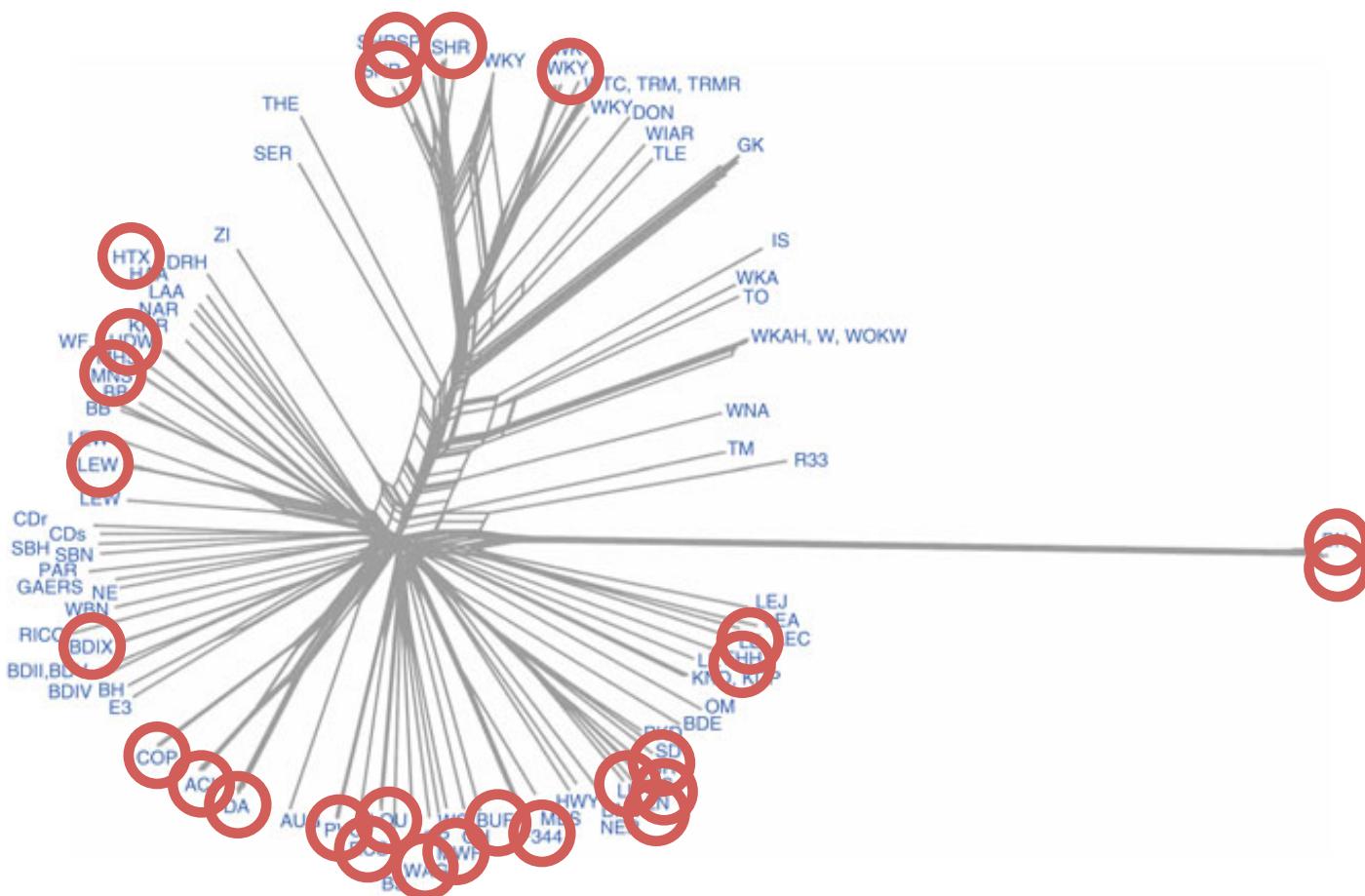
Power to detect transcript variation (advantage using both inbred and recombinant)

# Origin of HXB/BXH RI Panel



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

# Genetic Relationship Between Inbred Strains



SNP and haplotype mapping for genetic analysis in the rat. The STAR Consortium\*. Nature Genetics 40, 560-566 (2008)  
PD, M520/N, and MR/N also added for a total of 30

# Why Networks?

- “The human **brain** can be conceptualized as a **complex hierarchical network**... Spatio-temporal activity over... neurons intimately linked structure and function. Brain networks can be examined at any level of their hierarchy...” Power, J.D. et al., *Neuron* 67:735 (2010).
- Psychiatric disorders (**complex traits**) are “**unlikely a disease of a single gene, brain region or neurotransmitter system**. Rather, the syndrome is conceptualized as a systems disorder with a depressive (psychiatric) episode viewed as the net effect of failed network regulation.” Mayberg, H.S. *Biol Psychiatry* 61:729 (2007).
- “The **brain’s structural and functional systems have features of complex networks** – such as small-world topology, highly connected hubs and modularity – both at the whole brain level of human neuroimaging and... in non-human animals.” Bullmore, E. and Sporns, O. *Nature Reviews* 10:186 (2009).
- It has been proposed that the brain is organized into **several canonical functional networks**, i.e.:
  - ❖ The default mode network (DMN) – the resting state network
  - ❖ The dorsal attention network (DAN)
  - ❖ The executive control network (ECN)
  - ❖ The salience network (SN)

Raichle, M.E. and Snyder, A.Z., *Neuroimage* 37:1083 (2007)

Fox, M.D. and Raichle, M.E., *Nat Rev Neurosci* 8:700 (2007)

- **Resting-State connectivity can be a predictor of response** to chemical (drug) and other environmental perturbations. Crowther, A., et al., *Neuropsychopharmacol* \_\_:1 (2015).
- “Functional connectivity in fMRI is defined as **the temporal coherence, or statistical dependence**, between measurements of activity...” e.g.: transcription Power, J.D. et al., *Neuron* 67:735 (2010).
- We posit that “**Resting-State**” **transcriptional networks across brain provide power for understanding predisposition** to disease, etiology of pathology and prediction of response to medications or toxins “Transcriptional Connectome”.

# Why Study the RNA Dimension

Transcriptome links DNA and complex traits/diseases

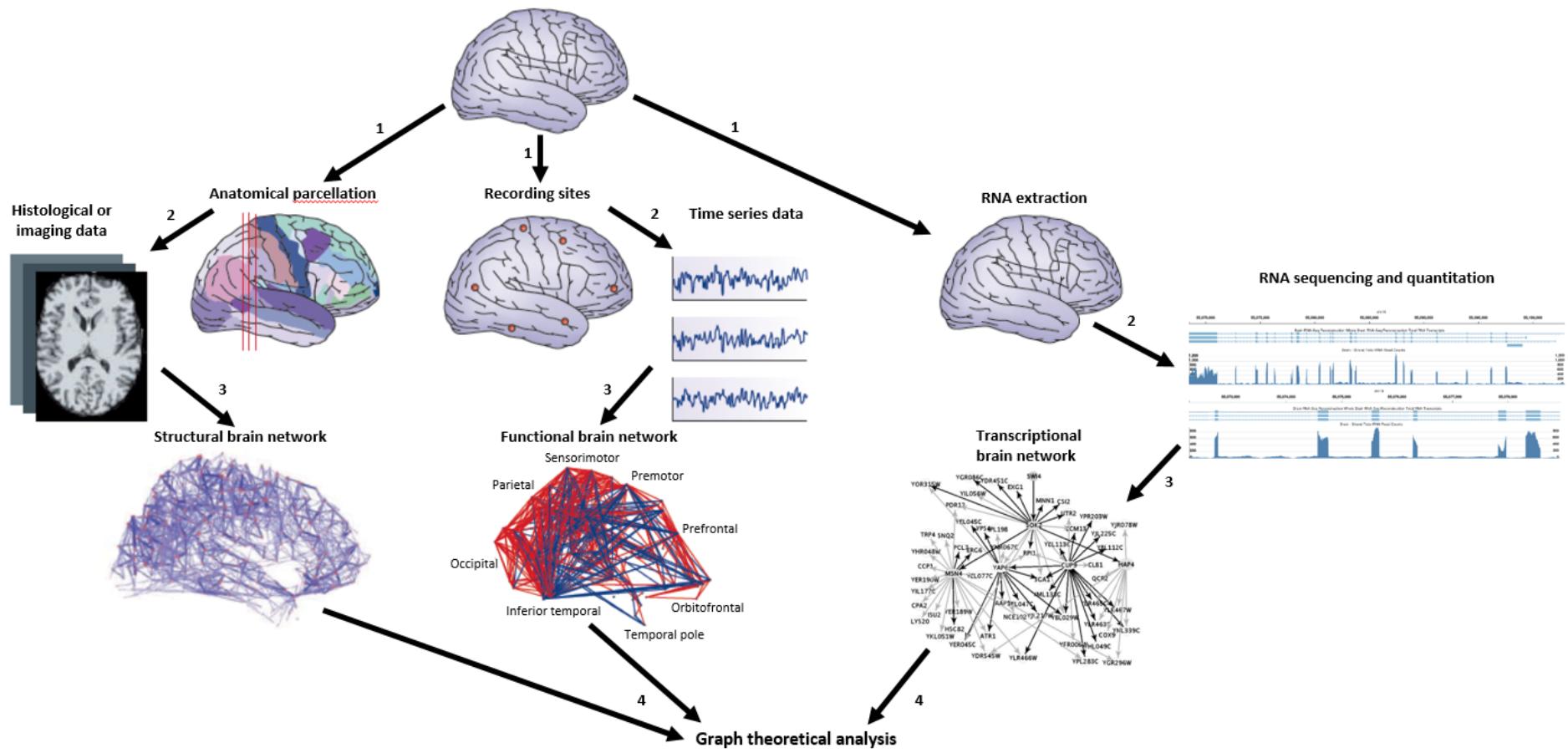
- A. First quantitative link between DNA sequence and phenotype (an endophenotypes).
- B. Transcriptome information addresses part of the GWAS Gap: how does identified DNA polymorphic locus contribute to disease?
- C. First step where DNA sequence and environment interact.
- D. Implementation of graph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

# Goal

## 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).
  - Completed exon array analysis for all 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

# Brain Transcriptional Connectome



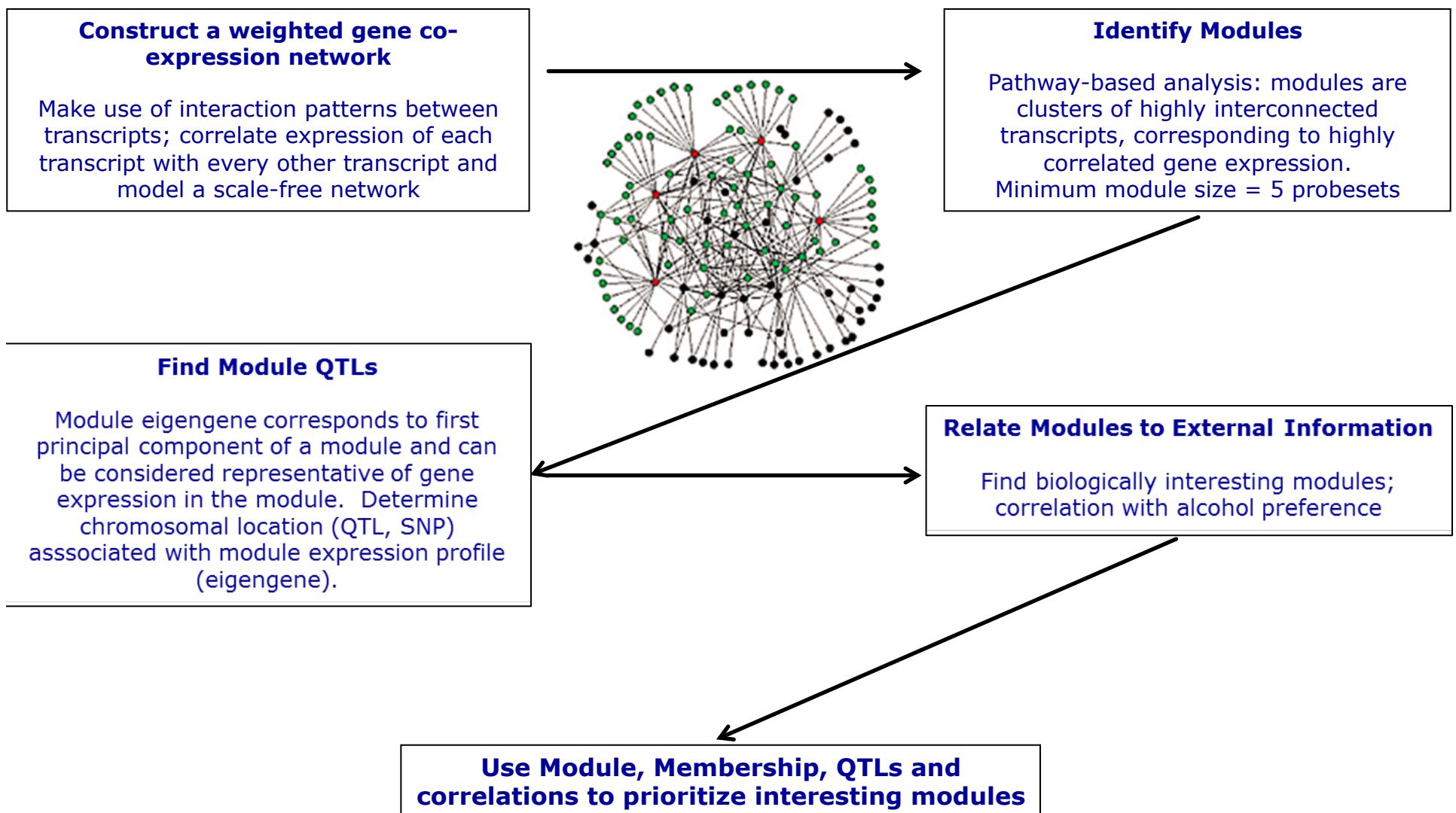
# Transcriptome Reconstructions

## Data from RNA-Seq on “Total” RNA

- Brain
  - 39,988 transcript isoforms (32,908 genes)
    - 9,674 Ensembl annotated transcripts
    - 30,314 novel transcripts (including 5,402 novel isoforms of Ensembl genes)
- Liver
  - 18,833 transcript isoforms (15,176 genes)
    - 6,745 Ensembl annotated transcripts
    - 9,988 novel transcripts (including 2,715 novel isoforms of Ensembl genes)
- Heart
  - 39,368 transcript isoforms (34,773 genes)
    - 7,478 Ensembl annotated transcripts
    - 31,890 novel transcripts (including 3,363 novel isoforms of Ensembl genes)

A significant amount of annotation left to complete

# Weighted Gene Coexpression Network Analysis



# 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
  - 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 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. Aberrant 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,<sup>1</sup> G. R. Reddy,<sup>1</sup> K. S. Murthy,<sup>2</sup> J. Johnson,<sup>1</sup> K. Sajwan,<sup>1</sup> and D. Desaiah<sup>3</sup>

<sup>1</sup>Savannah State University, Savannah, Georgia, USA

<sup>2</sup>Medical College of Virginia, Richmond, Virginia, USA

<sup>3</sup>University of Mississippi Medical Center, Jackson, Mississippi, USA

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



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 Selvin-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

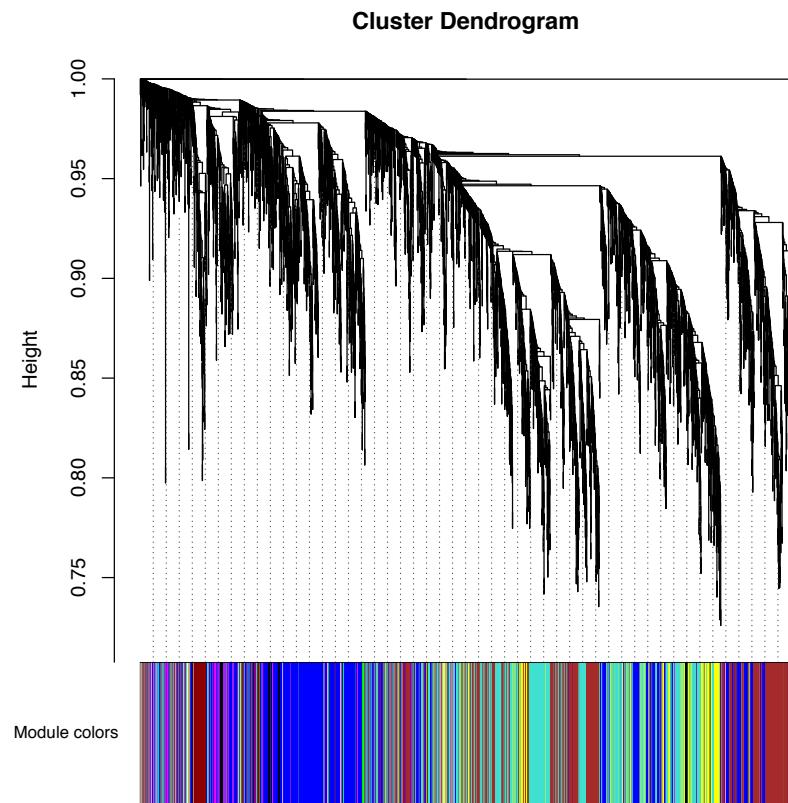
But how does this lead to the phenotype?

# Description of Data for Network Construction

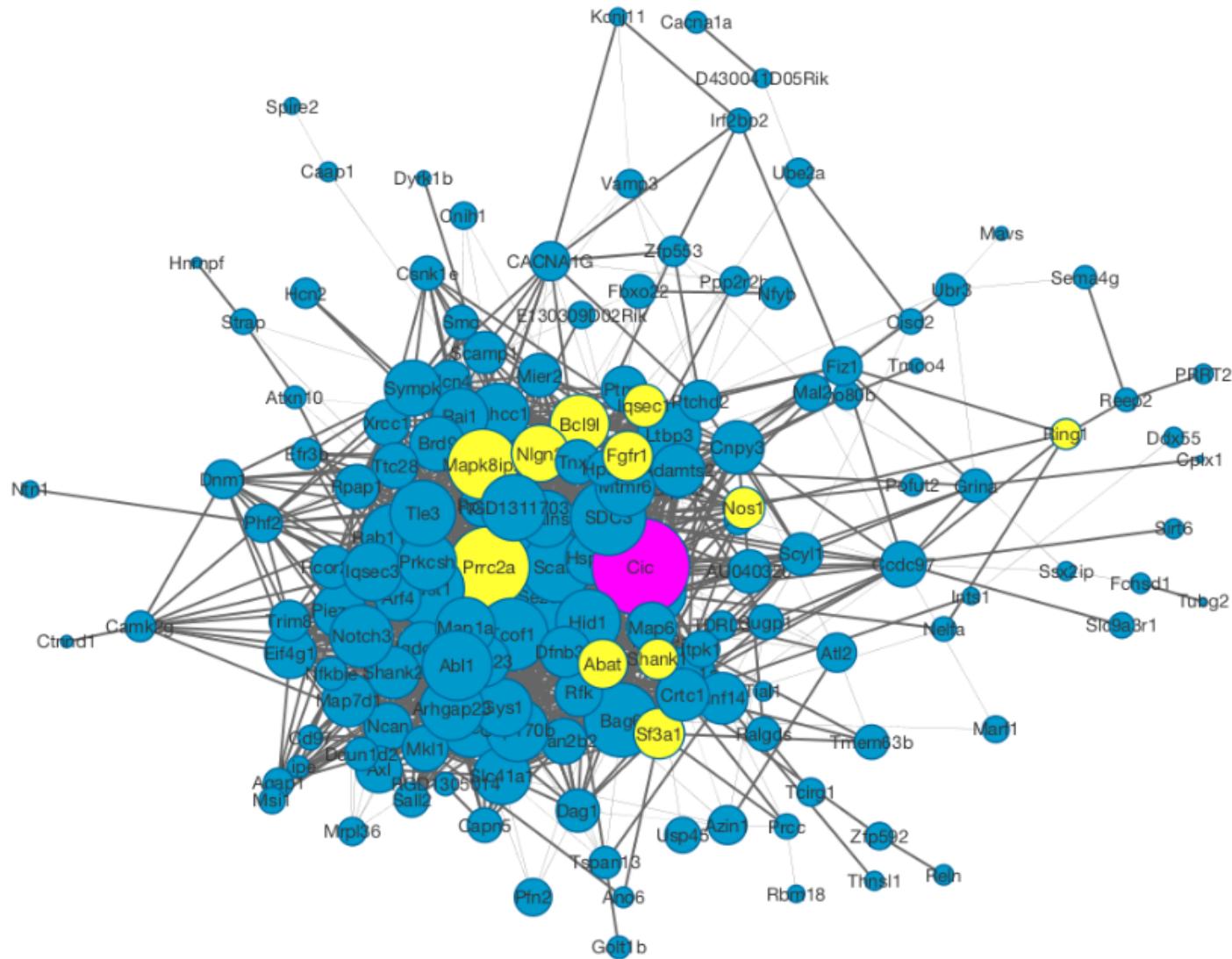
- Gene Expression Analysis on 29 HXB/BXH Recombinant Inbred Strains and Parental Strains
  - 1 to 2 biological replicates per strain for RNA-Seq
  - Included parental strains in data processing (e.g., normalization), but not in WGCNA due to population structure
- 17,293 Ensembl genes (transcripts used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



# Network Module With the Transcript for Nitric Oxide Synthase (nNOS)



# 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 75<sup>th</sup> in intramodular connectivity

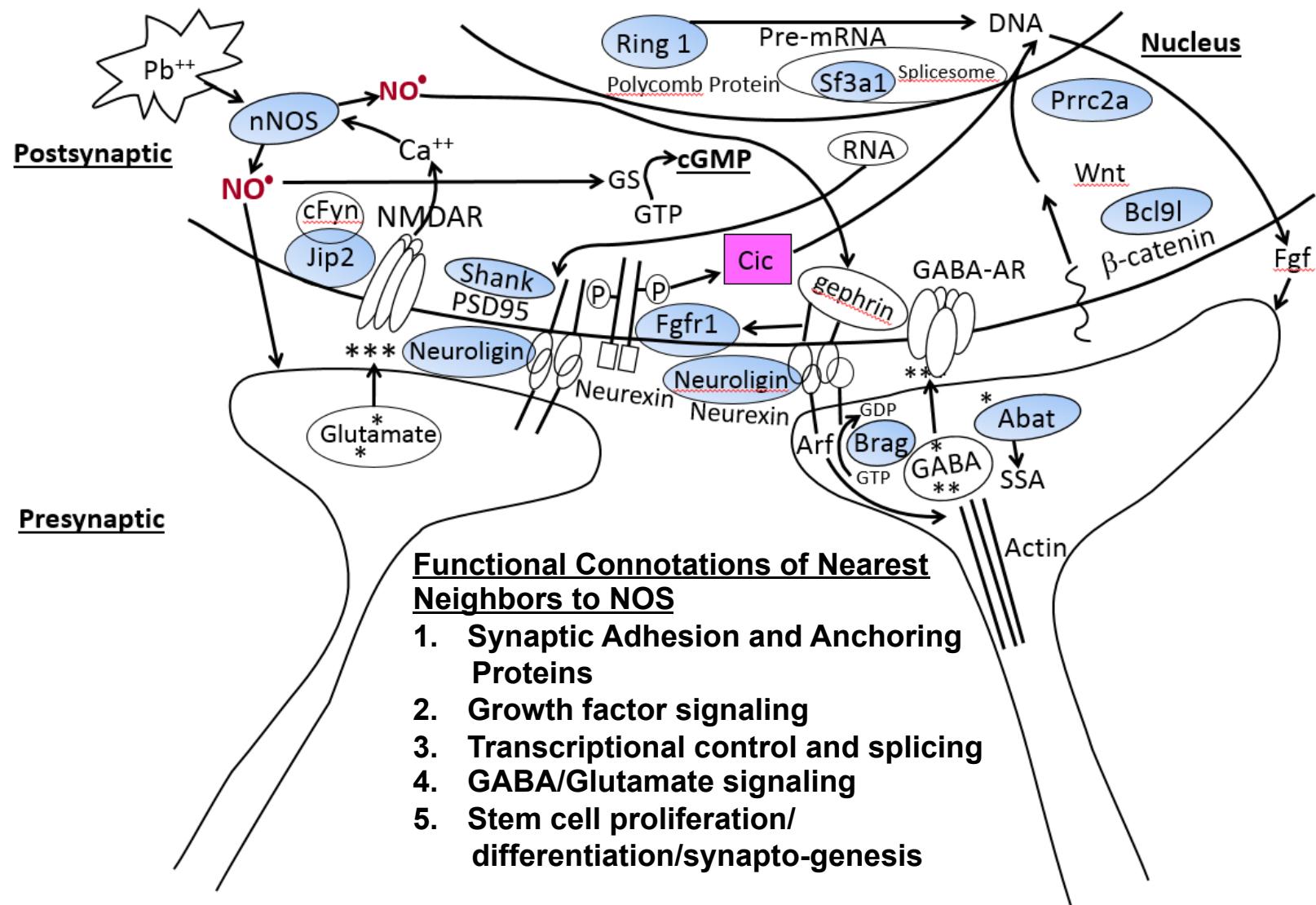
# Functional Enrichment in Pink Module

- 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



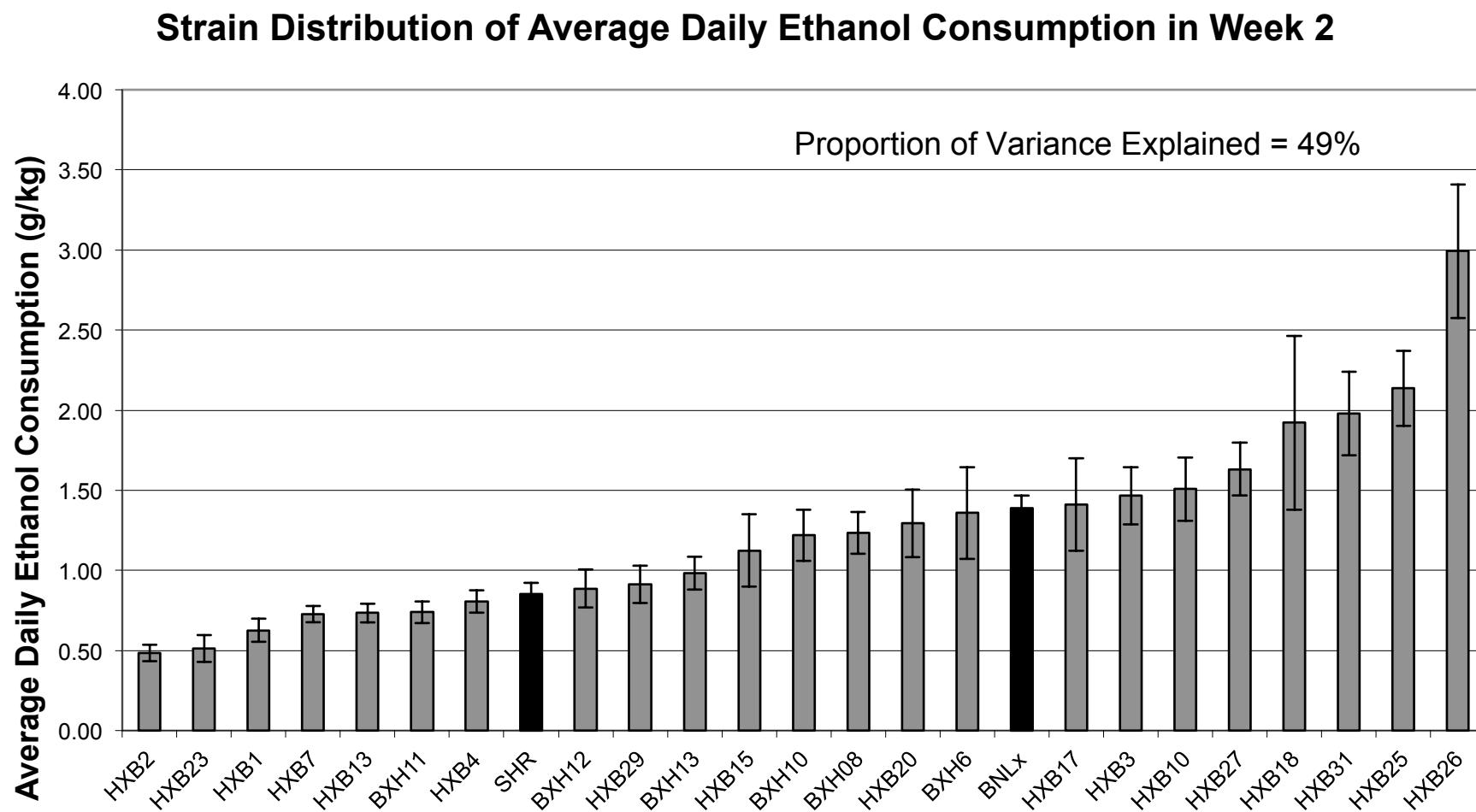
# Predisposition to Phenotype (Susceptibility) Genetics and Alcohol Consumption

- You can't be an alcoholic if you have never tried alcohol, i.e., it is an **etiological 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.



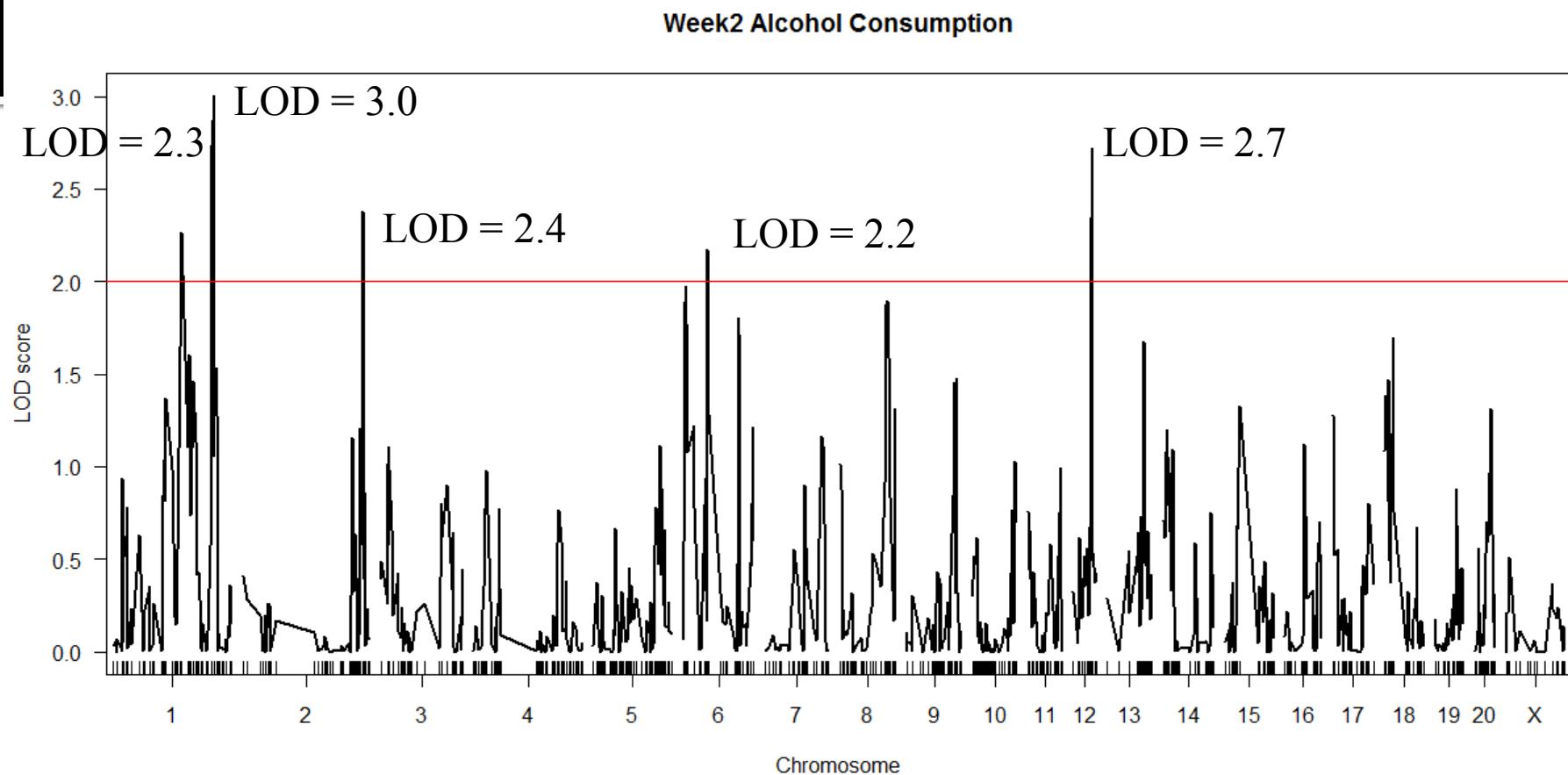
<http://less-social.com/the-power-of-community-migration/> (Budapest) page 4

# Distribution of Alcohol Consumption Across 25 HXB/BXH RI Strains



Copied From Tabakoff B, Saba L et al 2009. BMC Biol. 7:170

# 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

Chr 12: 31 – 53 Mb

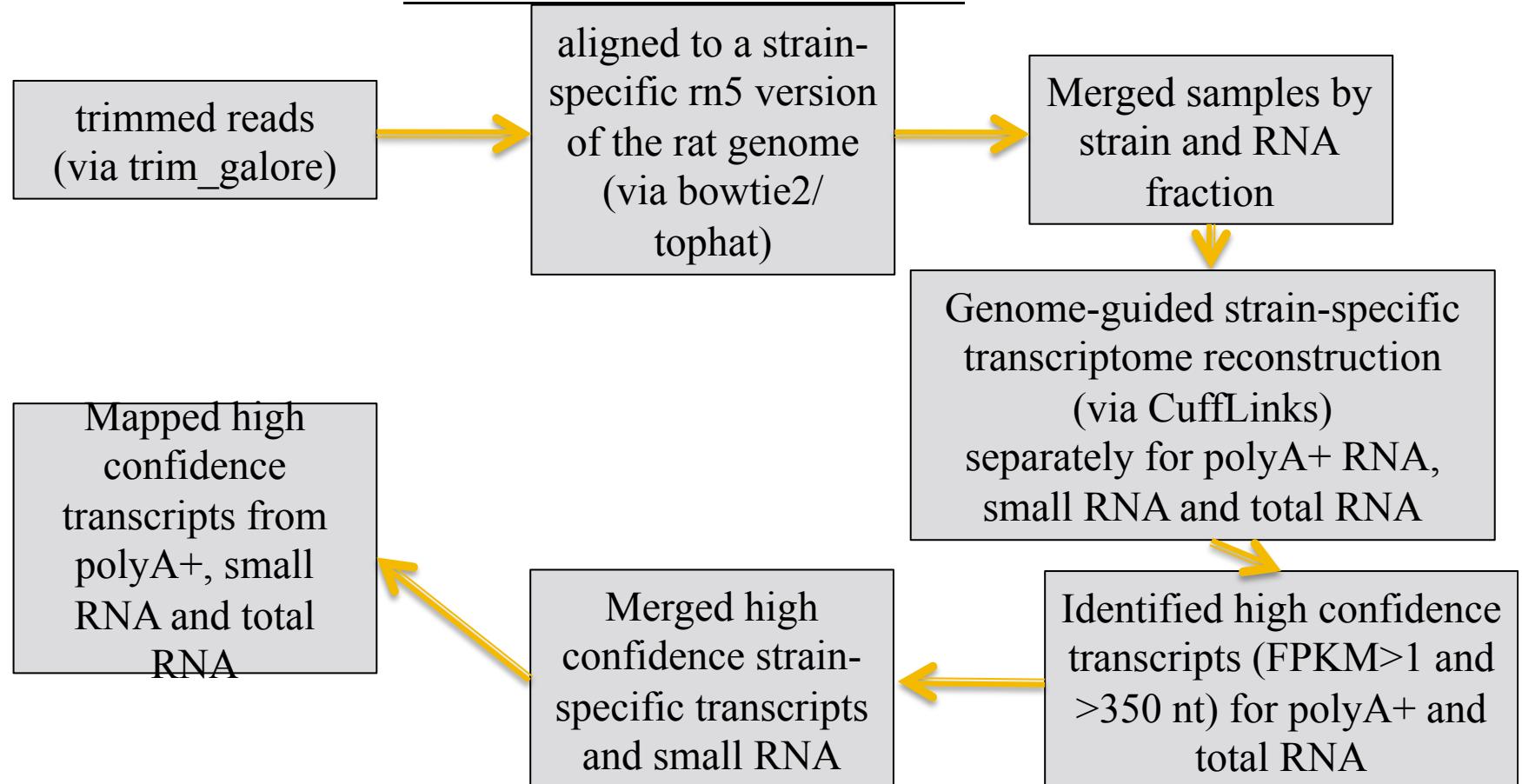
# **RNA Expression Estimates**

# Alcohol Naïve Brain RNA Expression

- Selected Lines – 60 Affymetrix Rat Exon 1.0 ST Arrays
  - 6 pairs of selected bred lines
  - One high drinking line, one low drinking line
  - 5 samples per line
  - 1 array per brain sample
- RI Panel - 84 Affymetrix Rat Exon 1.0 ST Arrays
  - 21 RI strains
  - 4 samples per strain
  - 1 array per sample
- RNA-Seq data collected on brain tissue from the RI strain progenitors (BN-LX/Cub and SHR/Ola)

# Preprocessing of RNA-Seq Data

## SHR/OlaPrin and BN-Lx/CubPrin Brain RNA from Ribosome Depleted PolyA+ RNA fraction and total RNA fraction



FPKM = fragments per kilobase per million reads 30

# **Quantify Transcription Levels of Genes and Individual Isoforms Expressed in Rat Brain From Exon Array**

## 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:

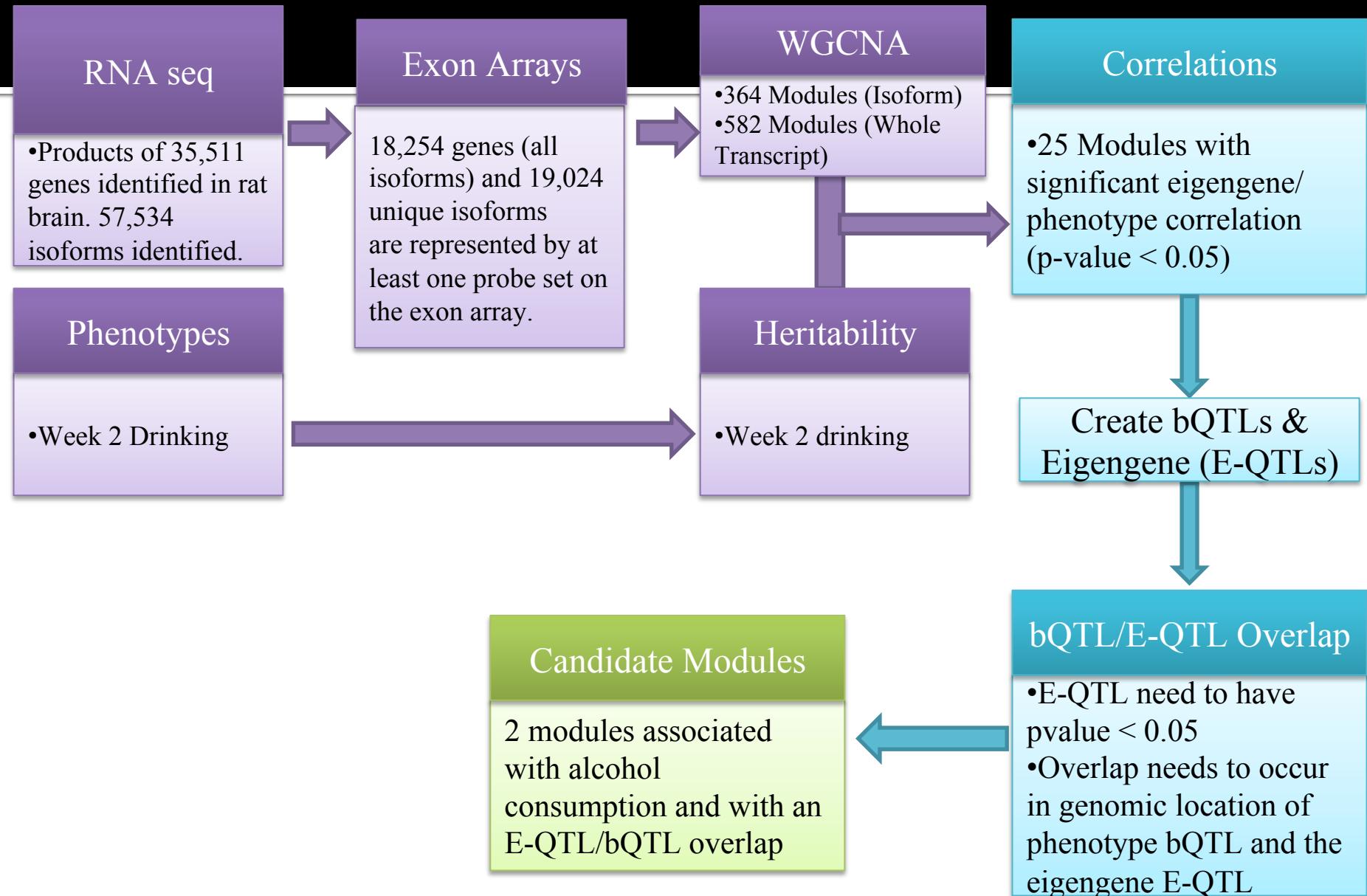
Aligned to the rat genome (rn5) more than once or did not align to the rat genome at all (approx. 500K probes)

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

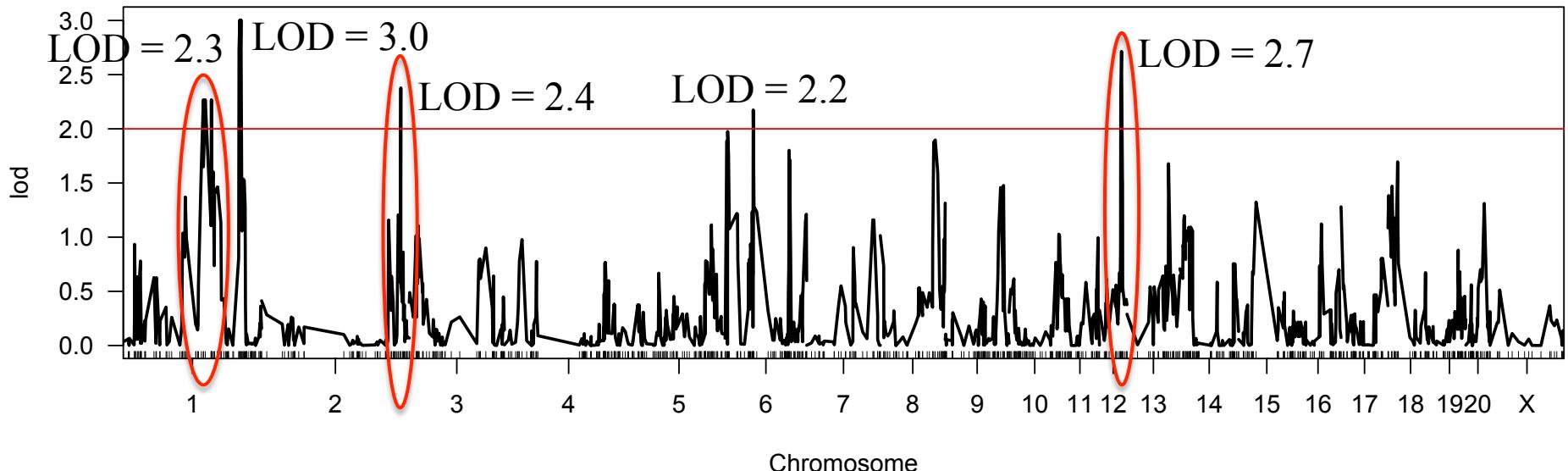
# Flow Chart for Data Analysis to Discern Candidate Modules Predisposing to High or Low Alcohol Consumption



# 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)
	Empirical Genome-wide P-value	0.033	0.027
Correlation with Drinking	Correlation Coefficient	-0.59	0.33
	P-value	0.005	0.045

# 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

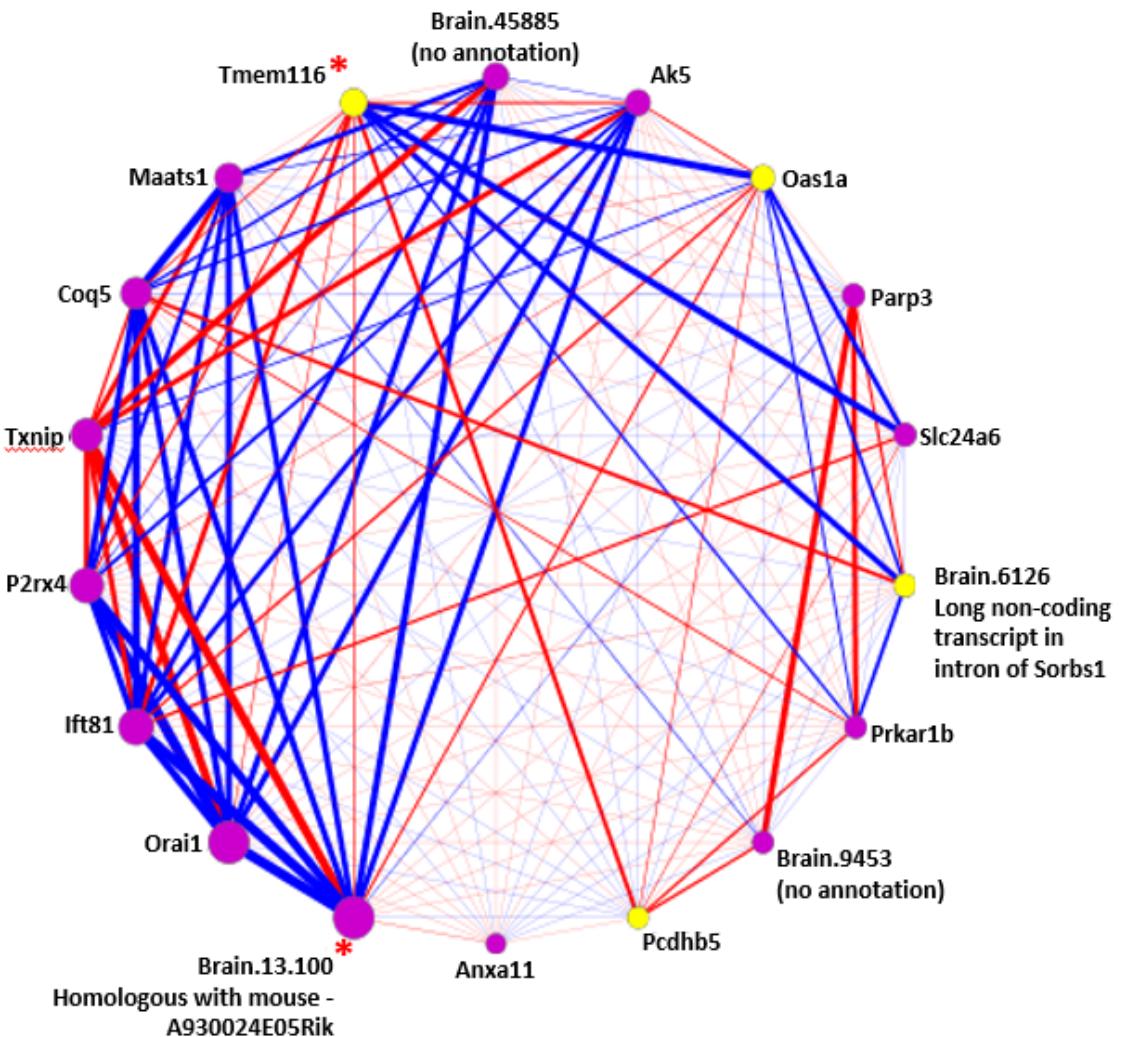
Chr 12: 31-52

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

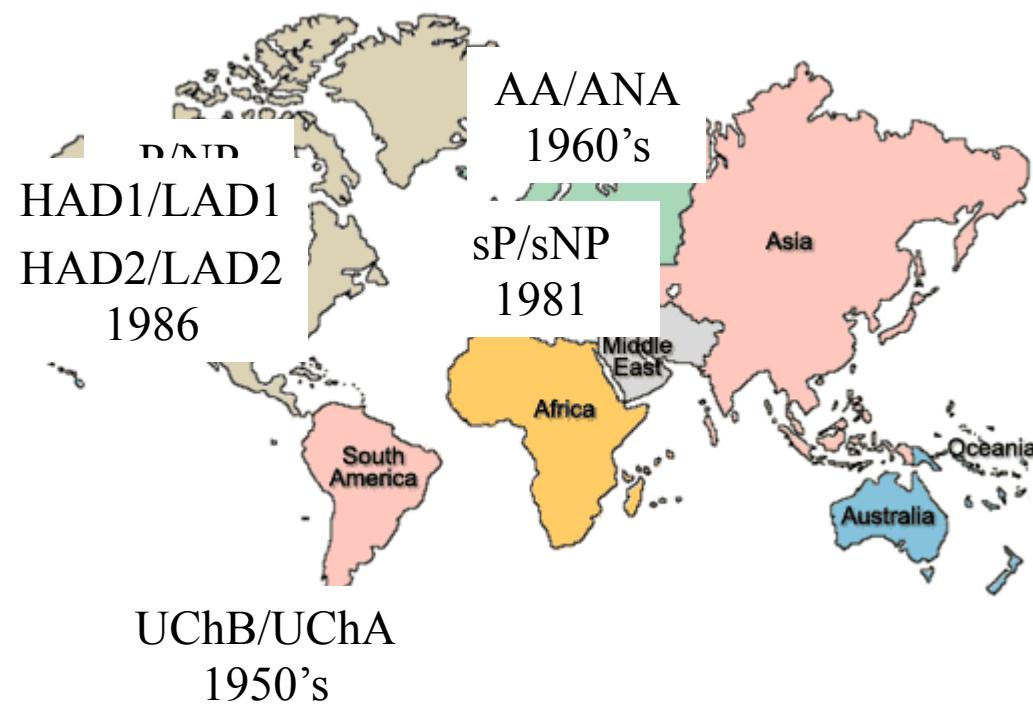
Circled b-QTLs are overlapped by  
Indianred4 E-QTLs

# Connectivity Within and Between Members of the Indianred4 and Aquamarine1 Modules

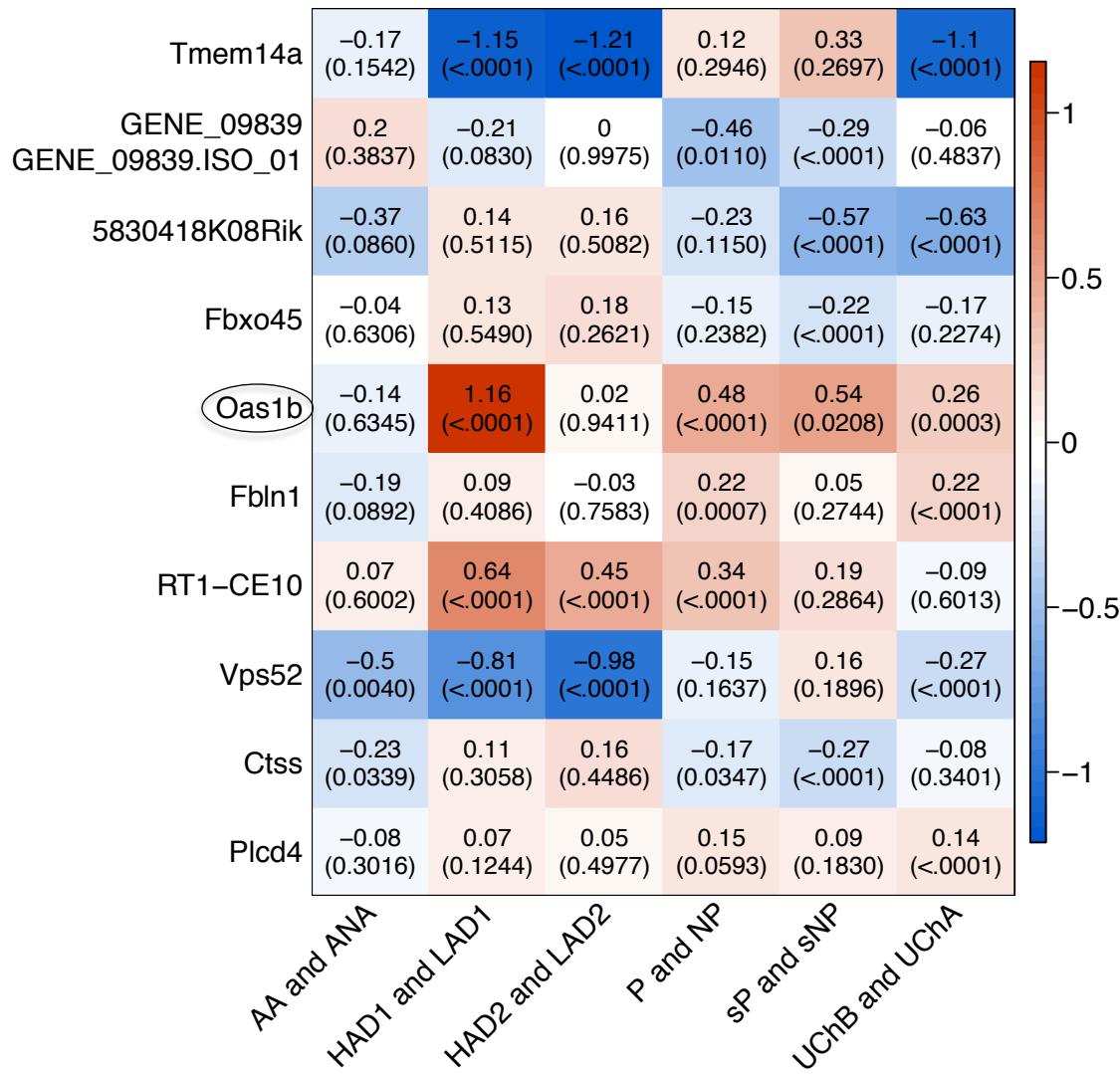
- Edge thickness is weighted based on magnitude of correlation between nodes.
- Blue edges represent a positive correlation between nodes.
- Red edges represent a negative correlation between nodes.
- Node size is weighted based on connectivity within module.
- Yellow nodes were included in the module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



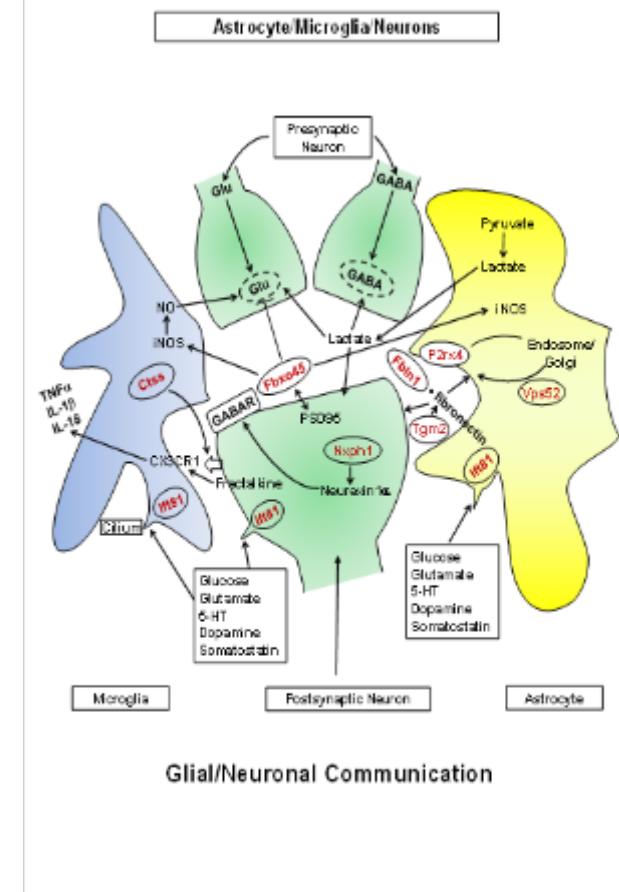
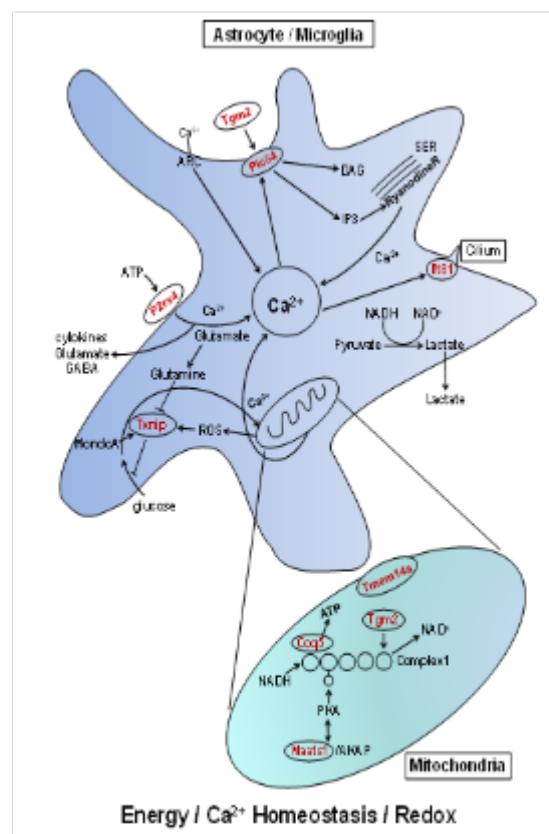
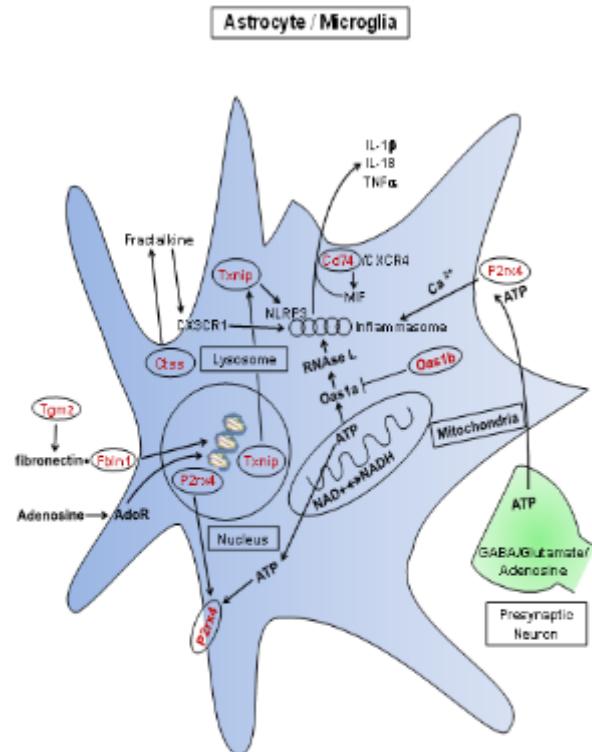
# Origin of Six Pairs of Selected Lines



# Genes/Isoforms differentially expressed between high alcohol consuming and low alcohol consuming selected lines of rats



# Biological Context from Pathway



# Conclusions

1. Approximately 34% of the rat genome is transcribed in brain.
2. The measure of abundance of transcripts across multiple genetically defined strains of animals can be used to organize the expressed transcripts into modules which can indicate functional (biologic) relationships within a module.
3. Module membership of differentially expressed genes (particularly if they are highly connected within a module) can indicate the pathway (system) that can be affected by the differentially expressed gene.
4. Our data indicate that two modules involved in 1) cell calcium homeostasis, 2) immune response regulation, 3) communication between microglia, astrocytes and neurons are of significant importance for distinguishing high alcohol consuming animals from low alcohol consuming animals.
5. These data open novel avenues for medication development (drugs that target neuroinflammation) to treat excessive alcohol consumption in humans.

# <http://phenogen.ucdenver.edu>

The screenshot shows a web browser window for the PhenoGen Informatics site. The URL in the address bar is [phenogen.ucdenver.edu/PhenoGen/](http://phenogen.ucdenver.edu/PhenoGen/). The page features a dark blue header with the title "PhenoGen Informatics" and a subtitle "The site for quantitative genetics of the transcriptome." Below the header is a navigation menu with green and blue buttons: 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 has a white background. It displays a welcome message: "Welcome to PhenoGen Informatics" and "The site for quantitative genetics of the transcriptome." A callout box on the left side provides instructions: "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](#)".

A central feature is a network graph with four main nodes: "Gene List Analysis" (green), "Microarray Analysis" (blue), "Pathway Analysis" (blue), and "Statistics / Expression Values" (blue). Lines connect "Gene List Analysis" to the other three nodes. To the right of the graph is a "Compare/Share" section titled "Demo/Screen Shots". This section includes tabs for "List", "Annotations", "Location(GPL)", "Literature", "Promoter", "Homologs", "Analyses", "Pathways", "Expression Values", "Event Correlation", "Save All", and "Compare". It also shows a comparison between "Example Pathway List 1" and "Example Pathway List 2".

# Working to Complete Data Acquisition

Strains	Tissue	Sex	Number of Biological Replicates Per Strain/Sex	Number of Paired-End Reads (rRNA-depleted total RNA)	Number of Single- End Reads (small RNA)
30 Classic Inbred Strains	brain	Male/female	4	36 trillion	11.0 billion
30 RI Strains	brain	Male/female	4	36 trillion	11.0 billion
30 Classic Inbred Strains	liver	Male/female	4	36 trillion	11.0 billion
30 RI Strains	liver	Male/female	4	36 trillion	11.0 billion

# 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

# 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](http://phenogen.ucdenver.edu) for data and details**

# The RNA Dimension (the true intermediate phenotype)

