

“Good Enough Solutions” and the Genetics of Complex Disease

Weiss JN, Karma A, MacLellan WR, Deng M, Rau CD, Rees CM, Wang J, Wisniewski, Eskin E, Horvath S, Qu Z, Wang Y, and Lusis AJ (2012). Circ Res 111:493-504

Laura Saba, PhD

Research Assistant Professor

Department of Pharmaceutical Sciences

Skaggs School of Pharmacy and Pharmaceutical Sciences

University of Colorado Anschutz Medical Campus

Disclosures

- Journal Club
 - Graphics and some text have been copied directly
- My biases
 - Inbred Rodent Panels
 - Currently part of an R24 that makes RNA expression of several panels publically available and is collecting RNA-Seq data on Hybrid Rat Diversity Panel
 - <http://phenogen.ucdenver.edu>
 - <https://github.com/TabakoffLab/NextGenSequencing>
 - Over ten years working with these types of panels
 - Weighted Gene Co-expression Network Analysis (WGCNA)
 - Did a short training with Steve Horvath at UCLA
 - Several publications using this technique and WGCNA modules are also implemented on PhenoGen

Problems with the single SNP approach

- Modest increased risk (average = 1.3-fold)
- Explains small fraction of genetic component (generally less than 20%)
- Difficult to elucidate biological mechanism
- Most common diseases are polygenic (hundreds or even thousands of genes involved) and are caused by modest effects of multiple genes interacting with the environment

Business Failure and Genetic Disease Analogy

Genetic Disease

- Bottom-up approach for identifying cause of disease
 - GWAS – examine all SNPs for association with disease
- Modules
 - Biological entities relate to the physiological regulation of metabolism or cell cycle, etc.
- Multifactorial reasons for disease
 - Complex polygenic disease

Business Failure

- Bottom-up approach for identifying cause of failure
 - Examine all personnel files to identify single rogue employee
- Modules
 - Employees within a certain unit, e.g., production, distribution, marketing, etc.
- Multifactorial reason for failure
 - E.g., underperforming department or lack of integration between departments

Top-Down Approach in Business

1. Group (cluster) employees into divisions (modules)
2. Relate division to performance measures
3. Identify underperforming employees within poorly performing divisions

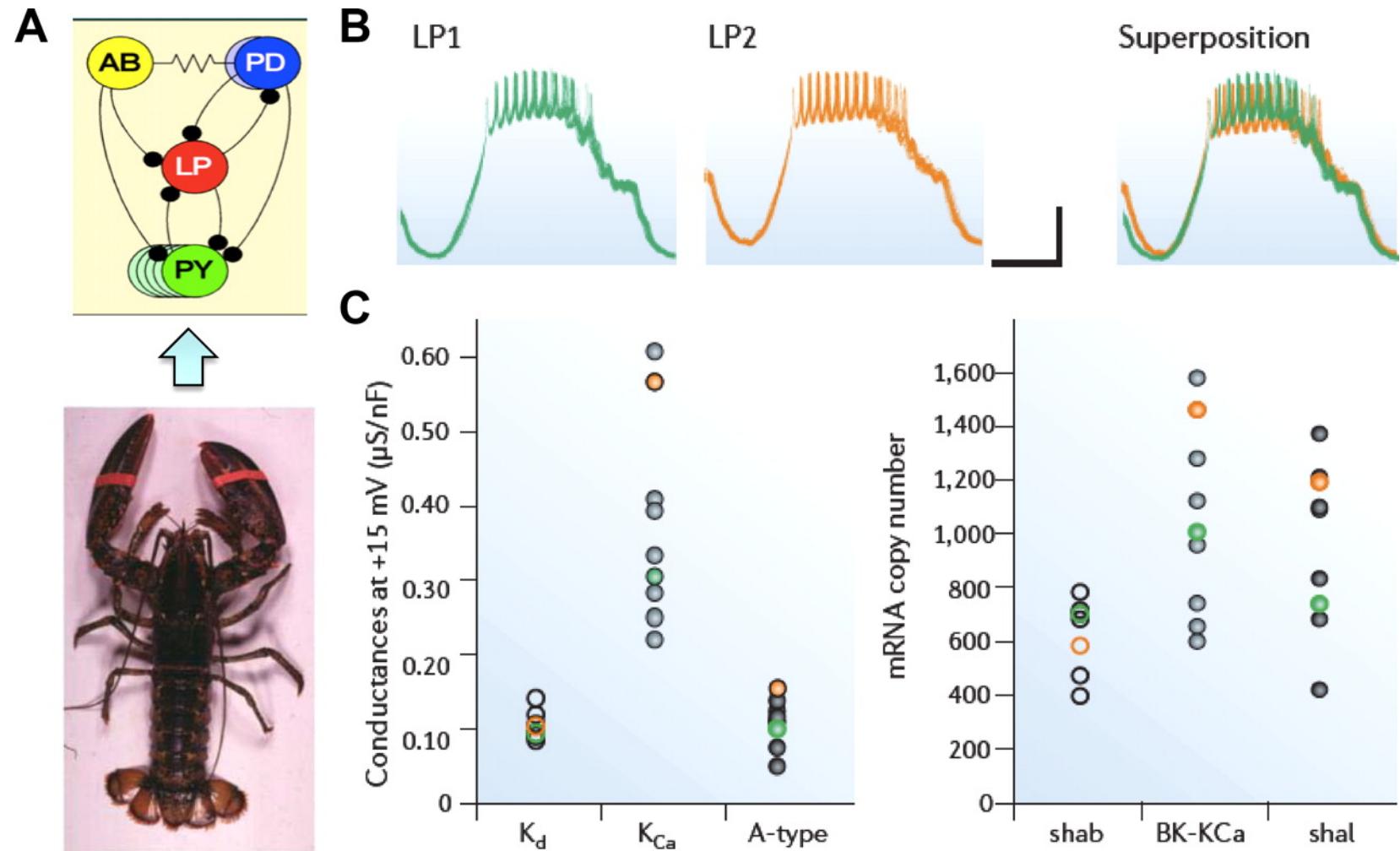
“Good Enough Solutions”



http://evolution.berkeley.edu/evolibrary/article/mantisshrimp_01

- Concept from evolutionary biology
- “in complex systems, many different combinations of the system’s parameters can produce a nearly identical output”
- a wide range of individual gene expression patterns may all be perfectly adequate for normal function, but have different abilities to adapt to an environmental stress (e.g., differential disease susceptibility and drug reactions)

“Good enough solutions” in the somato-gastric ganglia of lobsters.



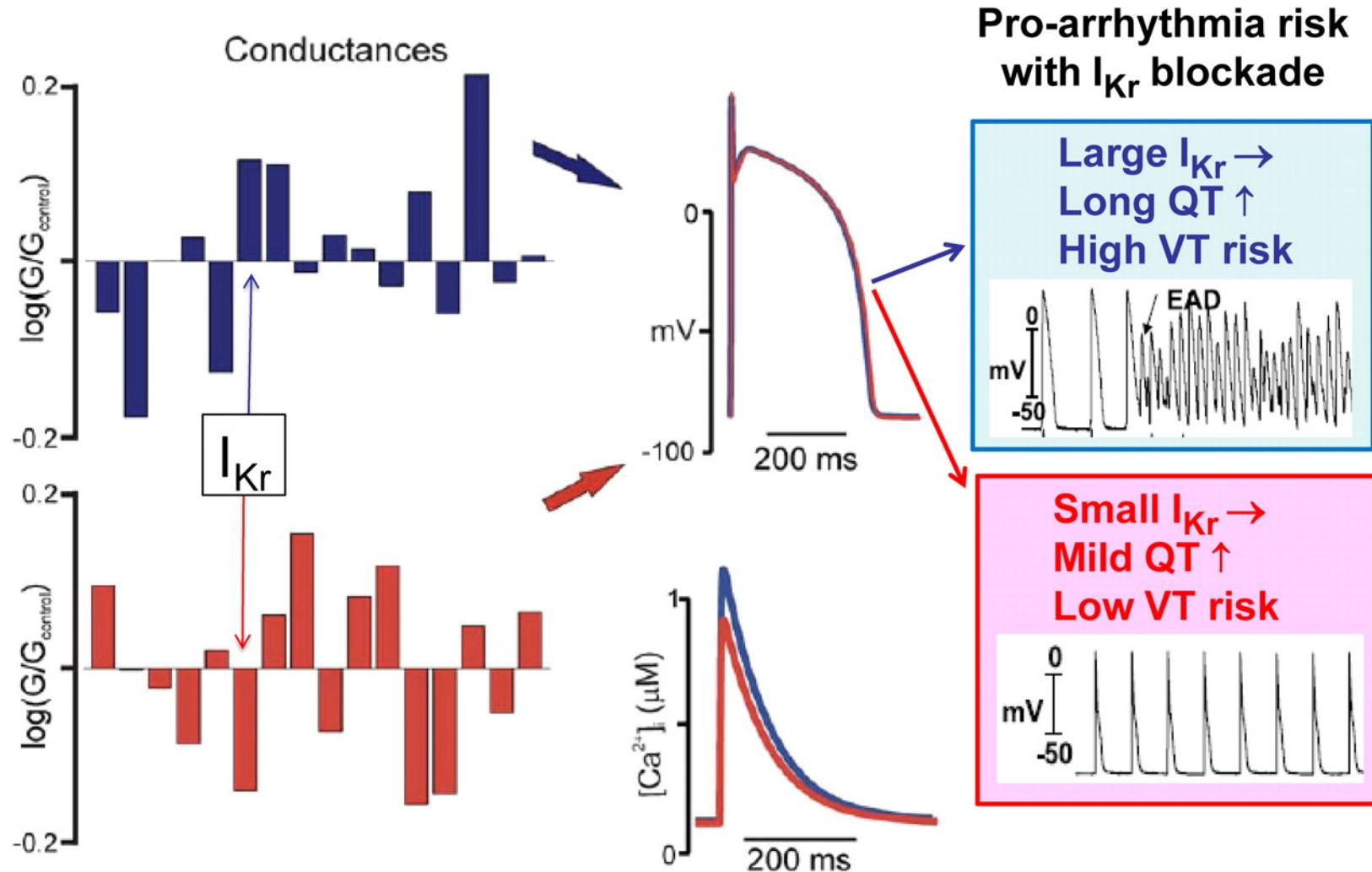
James N. Weiss et al. Circulation Research.

2012;111:493-504



Copyright © American Heart Association, Inc. All rights reserved.

Two sets of ionic conductances [$\log(G/G_{\text{control}})$] in a human ventricular action potential model that produce nearly identical action potentials and Ca transients (middle red and blue traces, respectively), for example, representing 2 “good enough solutions” for normal cardiac function.



James N. Weiss et al. Circulation Research.

2012;111:493-504



Copyright © American Heart Association, Inc. All rights reserved.

Why Networks?

- Complex polygenic trait \neq one gene
- Pathways add biological context to unannotated or under annotated genes.
- Fits the “Good Enough Solutions” idea.

How Can We Describe Genetic Networks?

- Nodes (genes) and Edges (interactions between genes)
- Selective pressure leads to a scale-free network
- One systems genetic approach is to use RNA expression levels to describe the network.

Selective Pressure and Scale-Free Networks

- Random Network
 - Majority of nodes have the average number of connections
 - Few nodes have many more or many less than average
- Growth of the Network
 - Random addition of new nodes and new links
 - Old nodes have more links than new nodes (due to time in network)
 - “the strong get stronger”, i.e., preferential attachment



Scale-Free Network

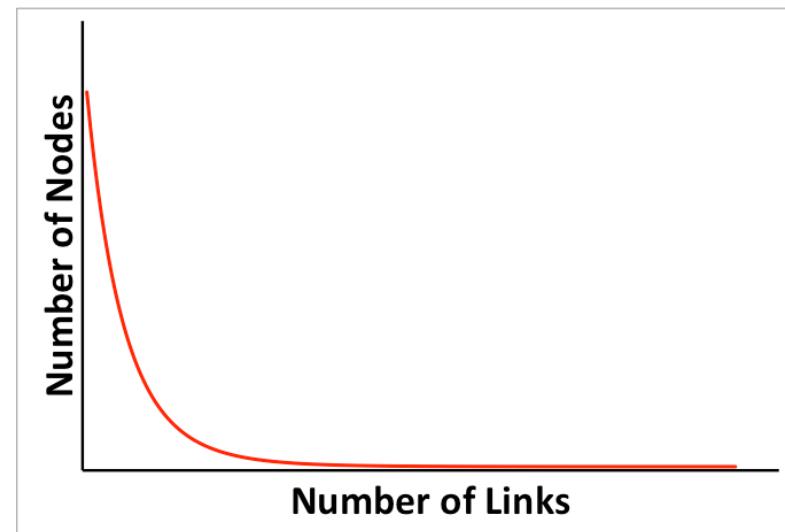
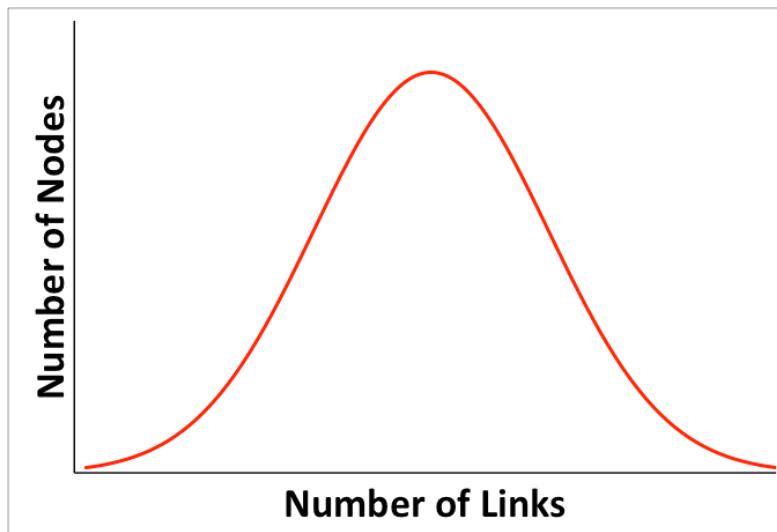
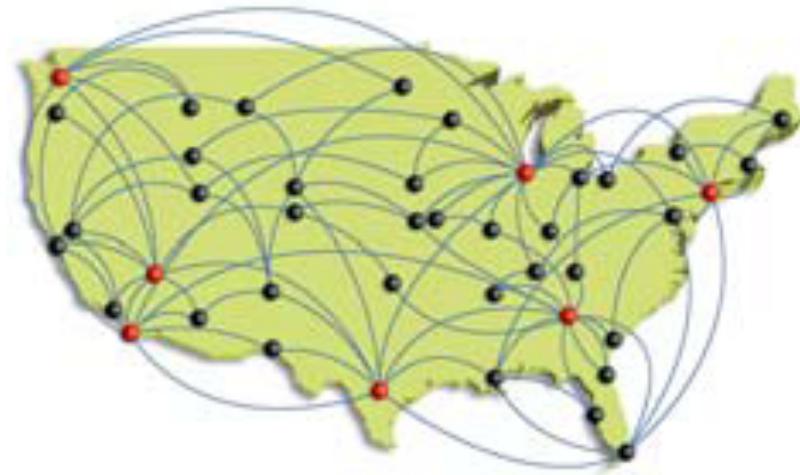
Number of links per node follows a power law distribution

Scale-Free Networks

Random Network



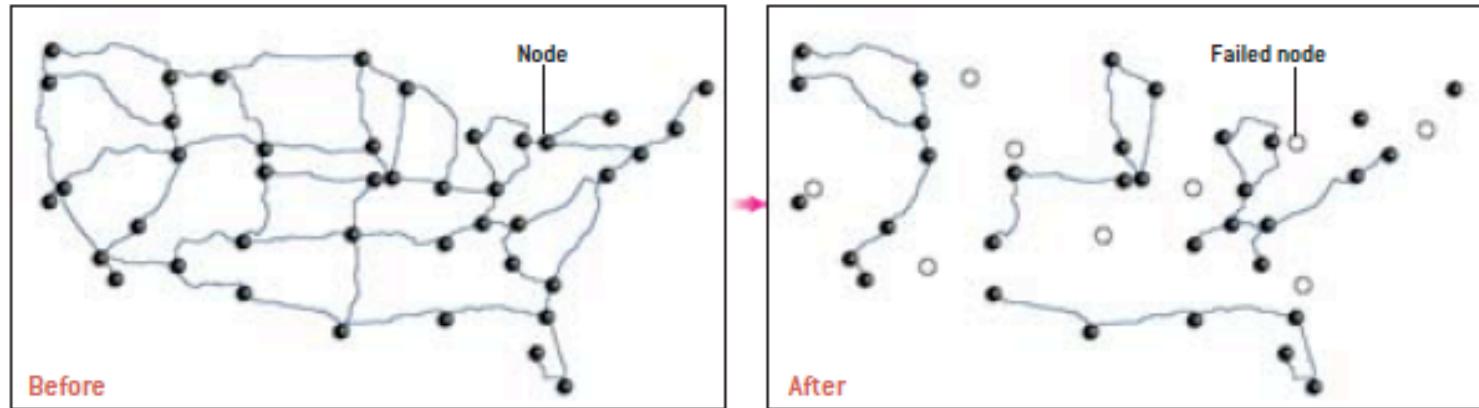
Scale-Free Network



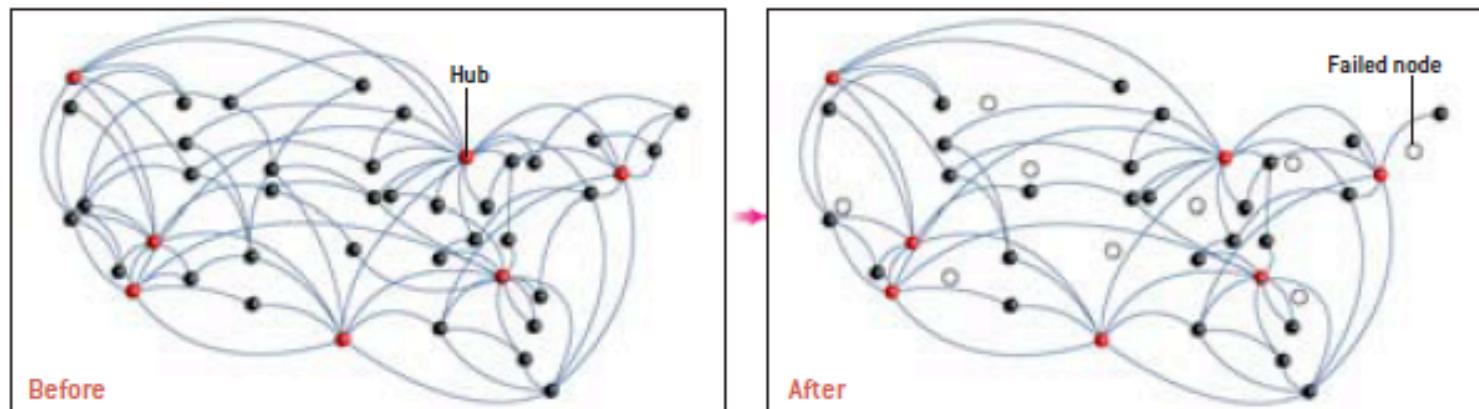
Advantage of Scale-Free Networks

- Robustness
 - Random failure of a node does not bring the whole system down.
 - “In a typical scale-free network, for example, up to 80% of nodes and links can be randomly destroyed before the network fails catastrophically.”
- Efficiency and Adaptability
 - “small-world property”
 - The ability to access any individual node from a multitude of alternative pathways makes scale-free networks inherently adaptable to changing environmental conditions

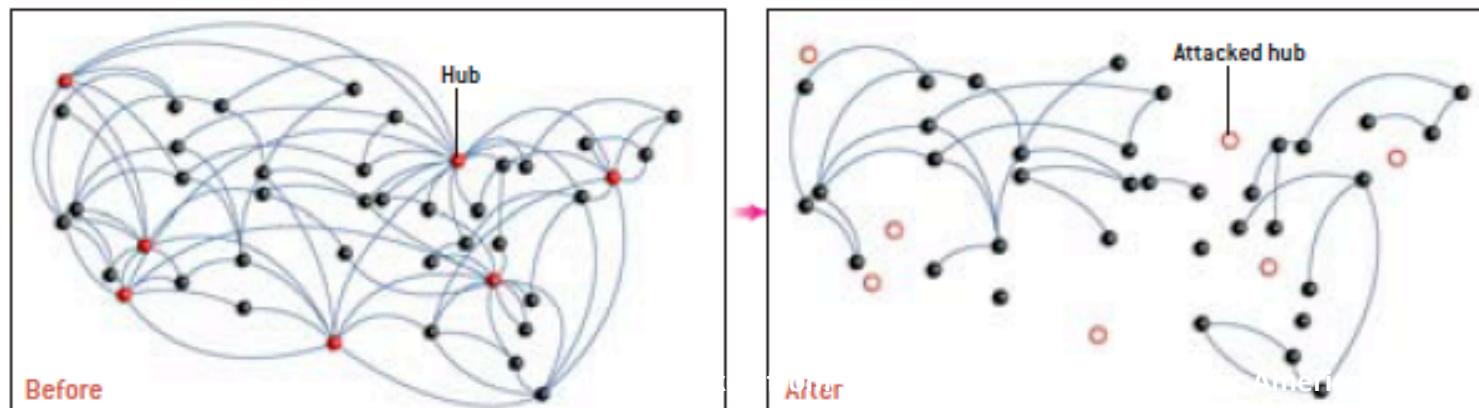
Random Network, Accidental Node Failure



Scale-Free Network, Accidental Node Failure



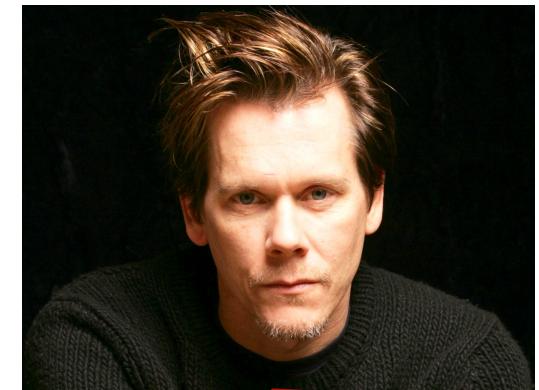
Scale-Free Network, Attack on Hubs



6 degrees of Kevin Bacon



Faster (2010)

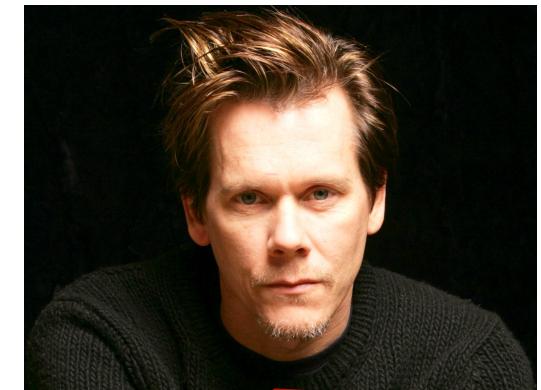


Jayne
Mansfield's Car
(2012)

6 degrees of Kevin Bacon

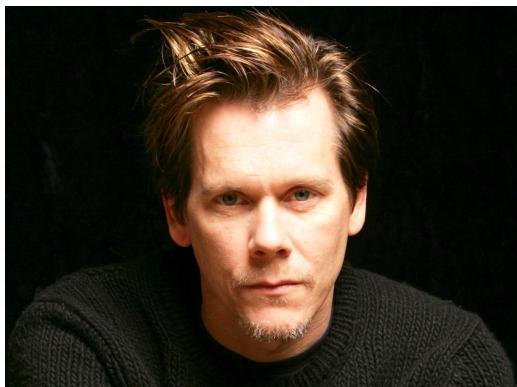
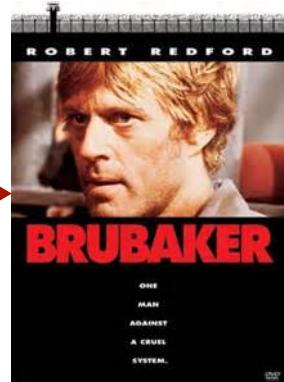


Get Smart
(2008)



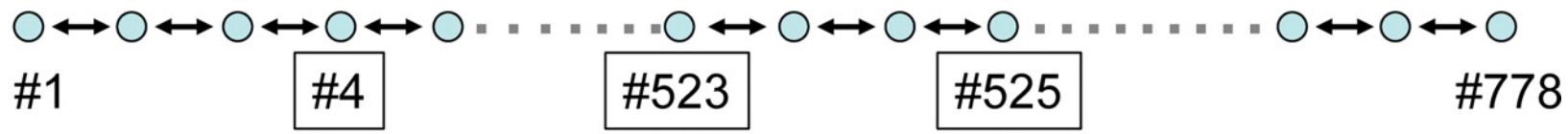
Wild Things
(1998)

6 degrees of separation

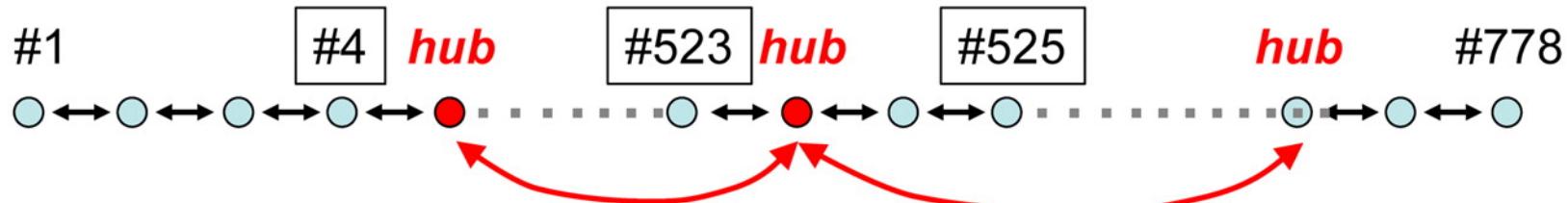


Small-world properties in metabolic networks.

A Linear array



B Network with hubs

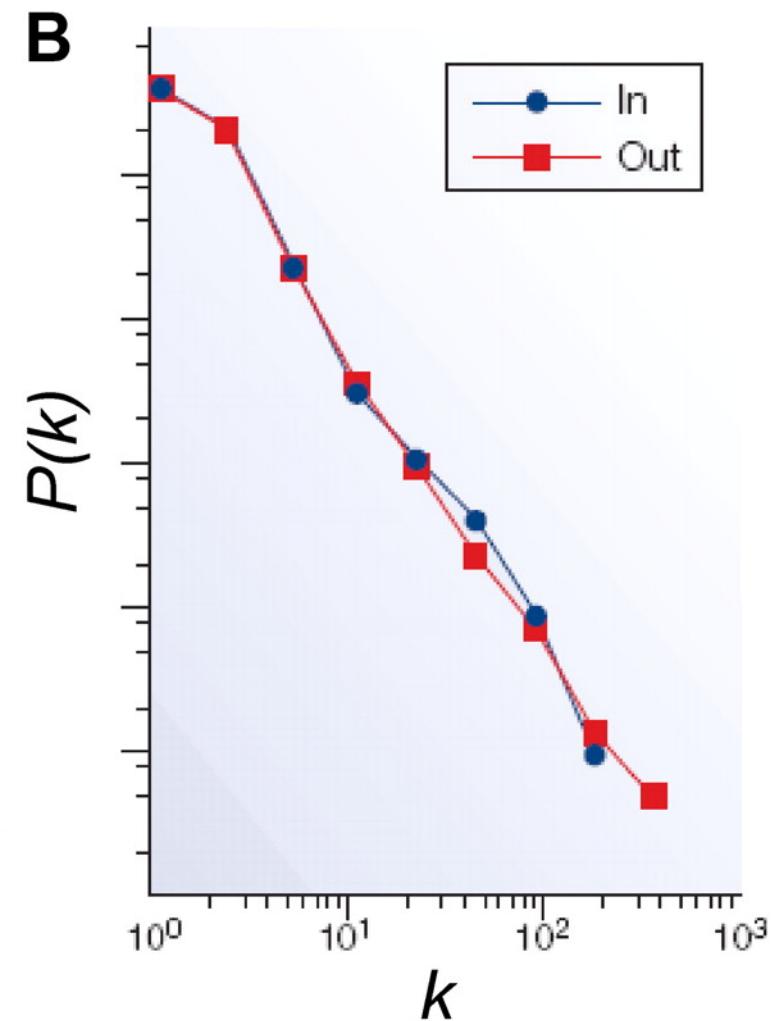
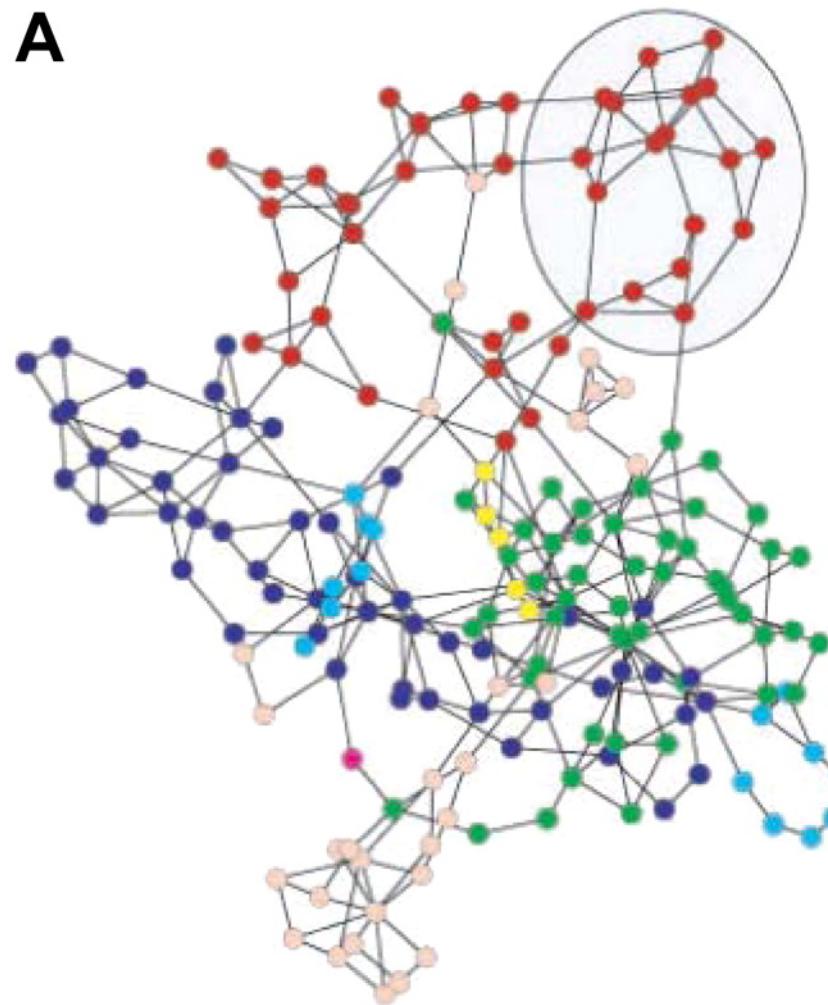


James N. Weiss et al. Circulation Research.
2012;111:493-504



Copyright © American Heart Association, Inc. All rights reserved.

A, The E. coli metabolic network, displayed as a topological overlap map, in which metabolites are nodes (circles) and enzymes converting one metabolite to another are links (lines).



James N. Weiss et al. Circulation Research.

2012;111:493-504



Copyright © American Heart Association, Inc. All rights reserved.

How to Construct a Gene Network

Assumption: if genes belong to the same module, they will be co-regulated by the same factors, i.e., their expression levels should correlate

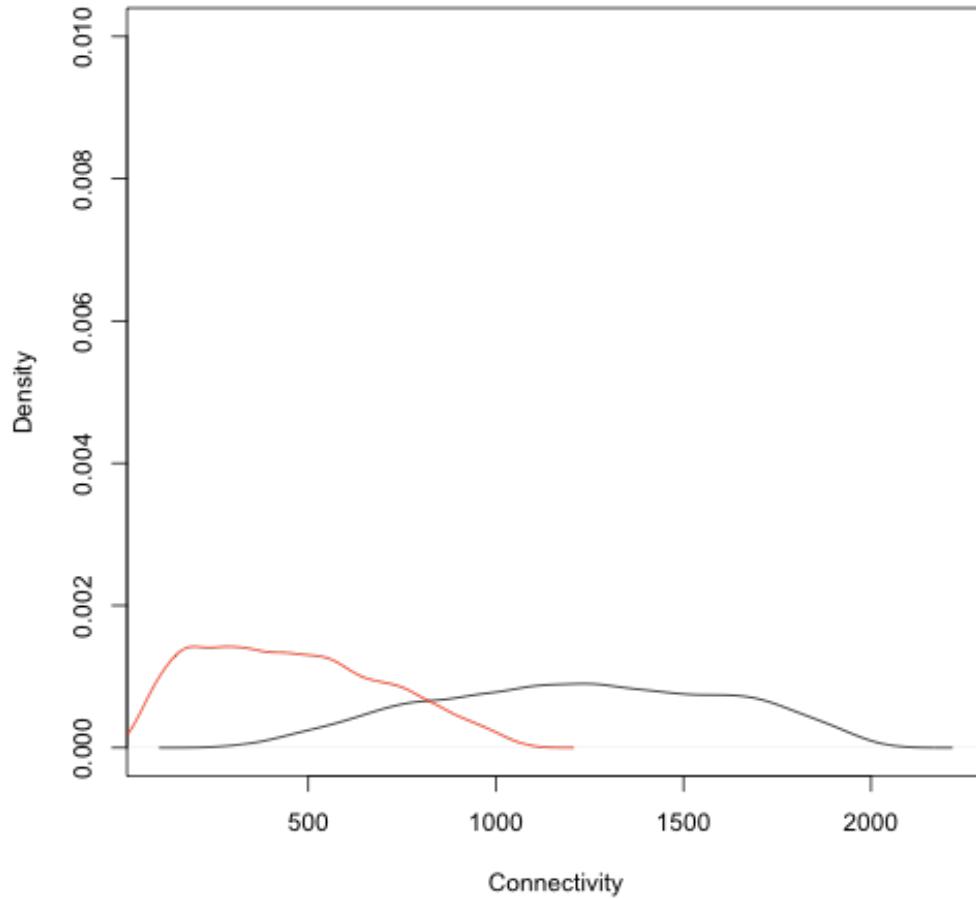
1. Quantify expression in a genetically diverse population
2. Analyze the data using a network algorithm, e.g., weighted gene co-expression network analysis (WGCNA)

WGCNA

- Step 1: calculate all pairwise correlation coefficients
- Step 2: create adjacency matrix using soft threshold to model a scale-free network
 - Raise the correlation coefficient to the β power
 - Choose smallest β that generates a scale-free network.

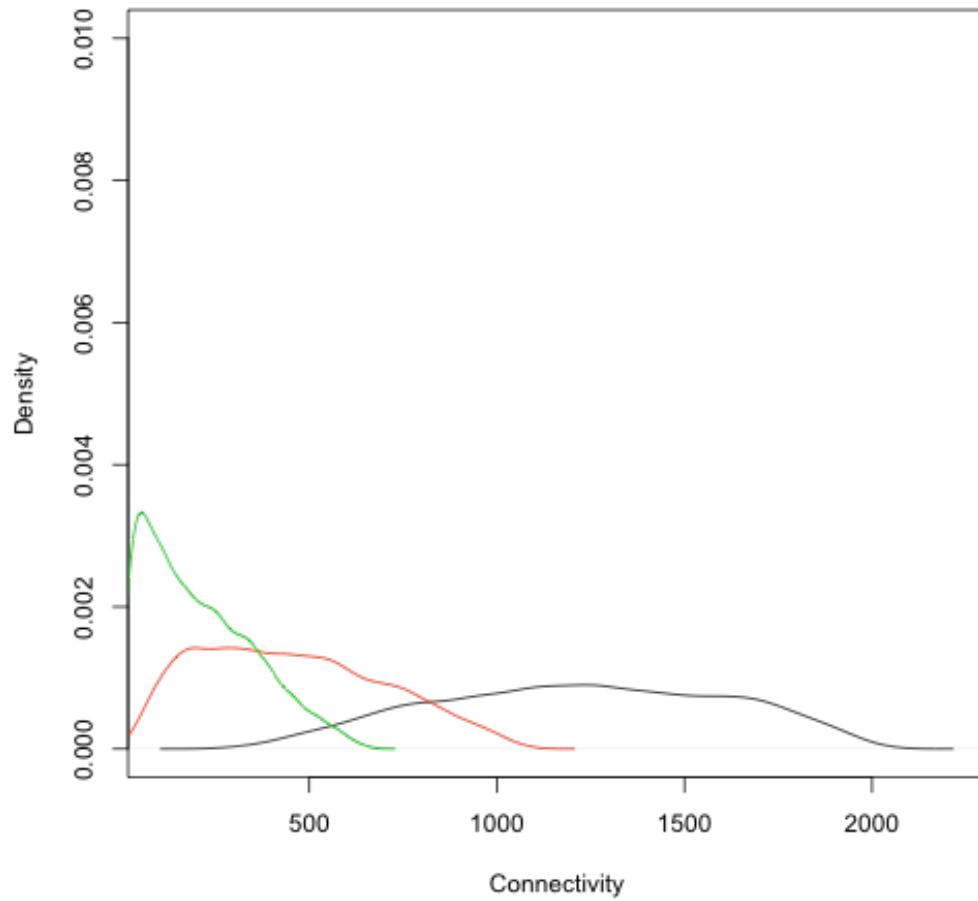
Scale-Free Network

$$\text{Connectivity}_i = \sum_{j \neq i} |\rho_{ij}^2|$$



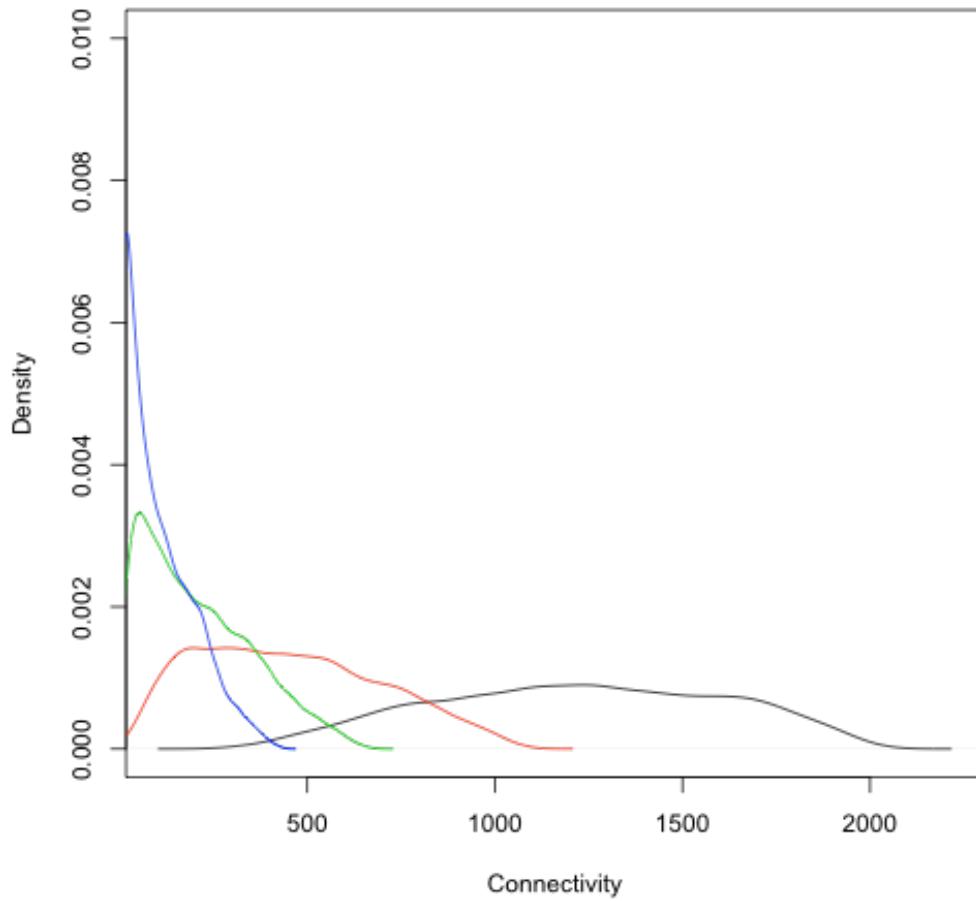
Scale-Free Network

$$\text{Connectivity}_i = \sum_{j \neq i} |\rho^3_{ij}|$$



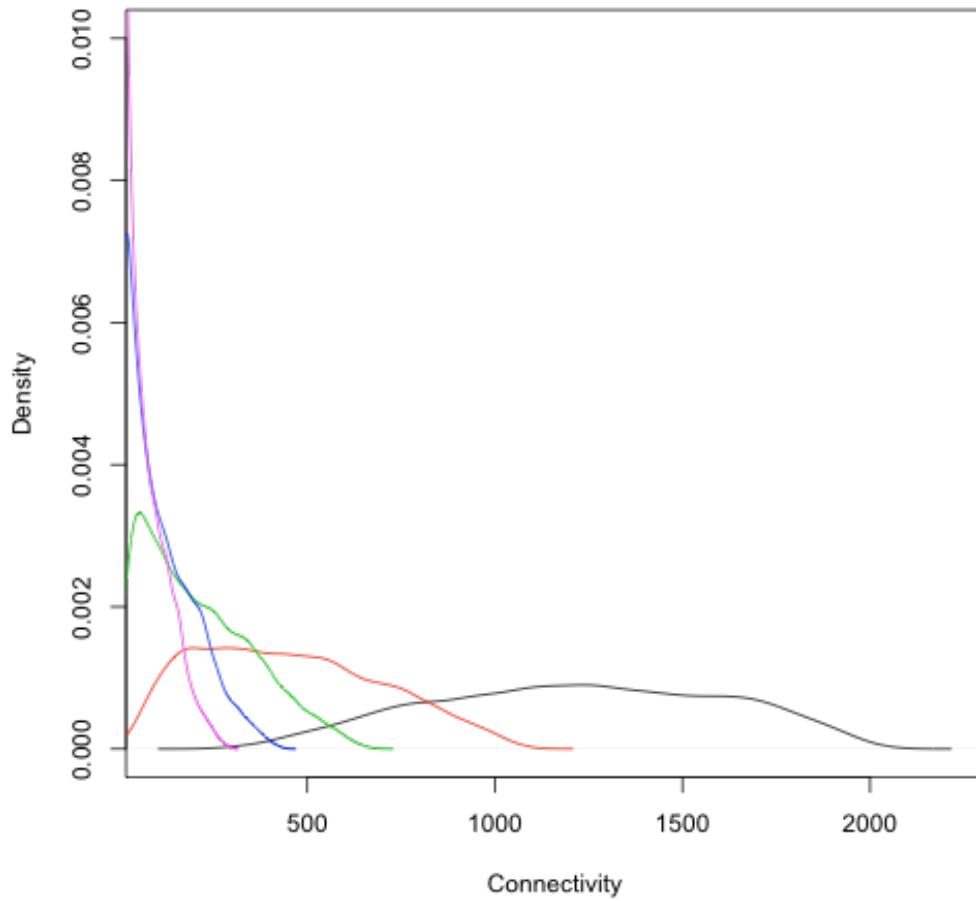
Scale-Free Network

$$\text{Connectivity}_i = \sum_{j \neq i} |\rho_{ij}^4|$$



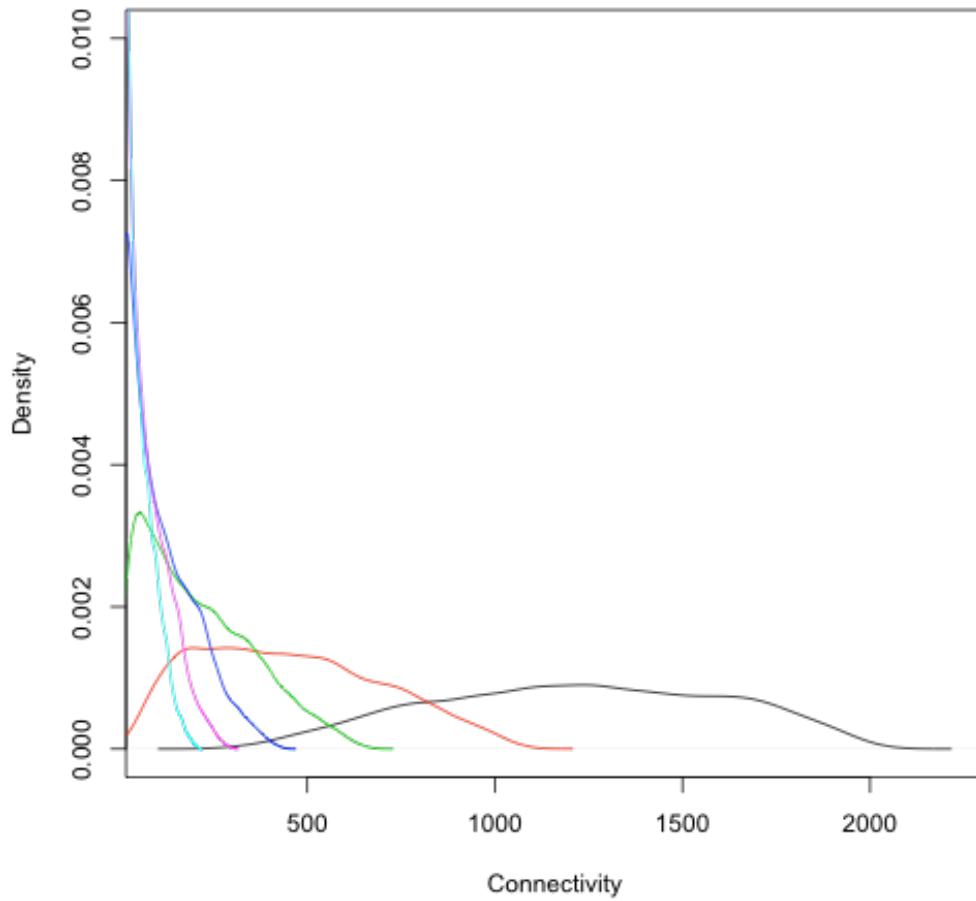
Scale-Free Network

$$\text{Connectivity}_i = \sum_{j \neq i} |\rho^5_{ij}|$$



Scale-Free Network

$$\text{Connectivity}_i = \sum_{j \neq i} |\rho^6_{ij}|$$



WGCNA

- **Step 3: calculate topological overlap**
 - “Are their friends friends?”
 - More robust
- **Step 4: define modules using hierarchical clustering and dynamic tree cutting**
- **Step 5: calculate eigengene for each module**
 - First principal component
 - Used to summarize expression across all genes within a module

Characterization of WGCNA Modules

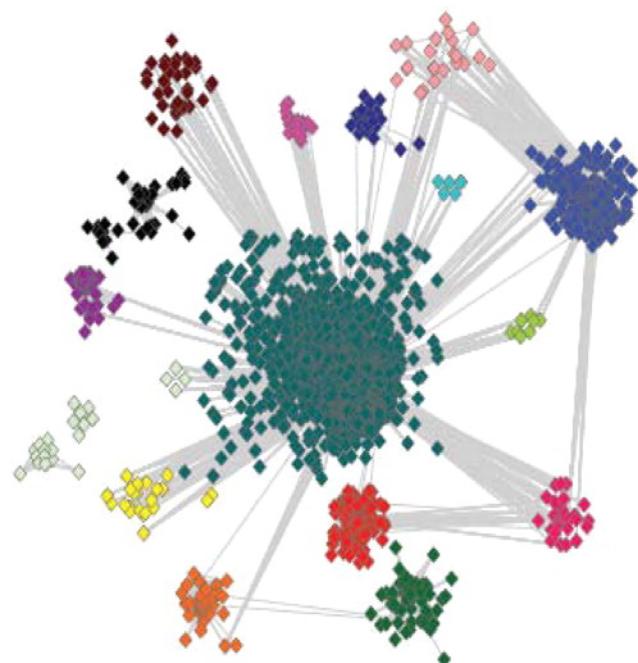
- Gene Ontology and Pathway Enrichment
- Cell-type specific expression
- Correlation of eigengene with physiological or behavioral phenotype
- Linkage disequilibrium
- Source of expression control (e.g., transcription factors, microRNA, common eQTL)

Models to Construct and Validate Gene Networks

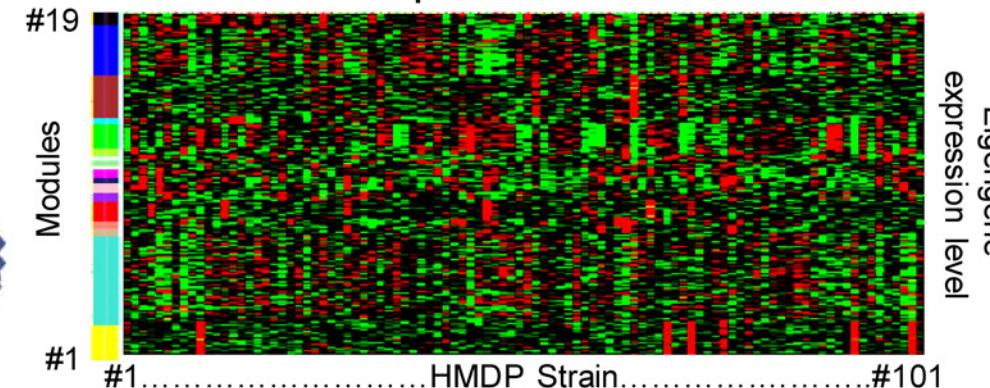
- Inbred Rodents
 - “close” to humans
 - Tissue-specific RNA available
 - Can accumulate data since animals are inbred
- Hybrid Mouse Diversity Panel (HMDP)
 - 100 (or more) recombinant inbred and common inbred mouse strains
 - All have been sequenced or densely genotyped
 - Commercially available
 - Biological replicates allow for heritability calculations
 - Requires the use of mixed models to handle population structure

A, Topological overlay map of the cardiac gene network from the HMDP. Colors indicate nodes belonging to the same gene module.

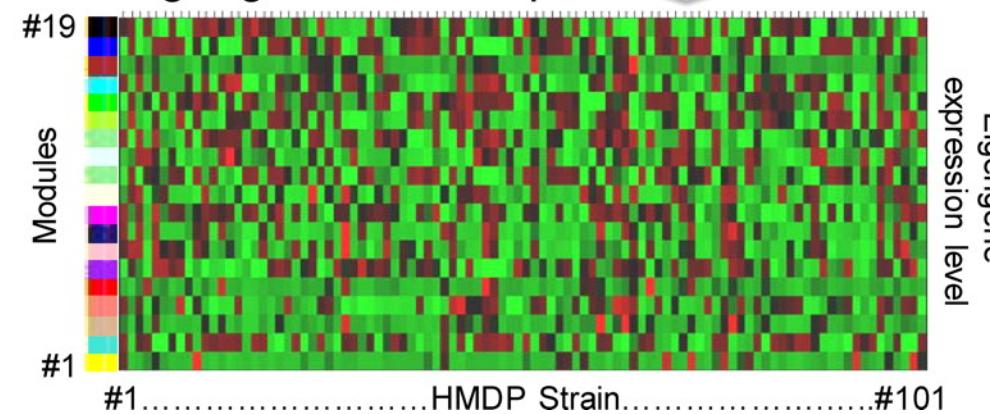
A Topological overlay map



B Gene heat map



C Eigengene heat map



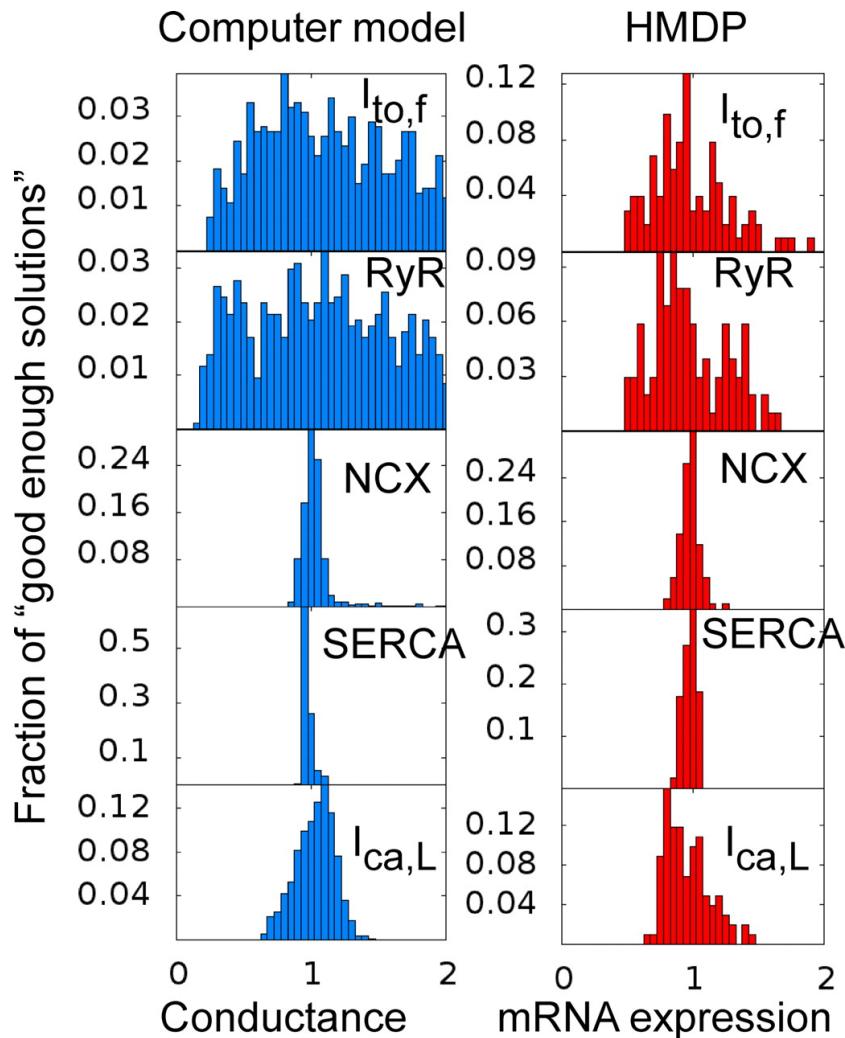
**each HMDP strain represents a different “good enough solution” for survival in captivity under normal conditions

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



Copyright © American Heart Association, Inc. All rights reserved.

Correspondence between gene expression and “good enough solution” predictions.



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



Copyright © American Heart Association, Inc. All rights reserved.

Gene Module Association Study (GMAS)

observations/assumptions

1. Applying WGCNA to the HMDP allows genes to be grouped into an approximately scale-free network of gene modules
2. The aggregate behavior of the genes in a given module can be represented by the module's eigengene (first principal component)
3. The expression patterns of eigengenes are very similar among individuals within a strain but vary considerably between strains (different “good enough solutions”)
4. Exposure of the different strains to an environmental stressor produces different results.
5. If the eigengene patterns do represent different “good enough solutions” then the eigengene should predict the phenotypic response to the stressor.

Top-Down Strategy using GMAS

1. Expose HMDP strains to an environmental stressor
2. Quantitate their response to the stressor
3. Correlate the eigengene expression patterns of the HMDP strains with their quantitative response to the stressor

**Because we have dramatically reduced the number of associations by looking at modules rather than individual genes, we can also ‘afford’ to look at interaction effects of multiple eigengenes on the response.

Caveats and Challenges

1. How do we validate assumptions?
 - E.g., Are modules truly biologically meaningful?
2. How should we decide which genes to include in a module?
3. How should we infer function of modules?
4. Are eigengenes the “best” summary measure?
5. Are the assumptions of linearity realistic?
6. Is gene (RNA) expression the best/only measure to use?
7. How valid is the assertion that genes don’t jump to different modules in response to a stressor? Does the expression of the entire module change?

Clinical Implications

- Possible to identify commonalities between human/mouse modules to allow for translational studies
- Identify novel therapies if they alter eigengene expression (i.e., altered disease susceptibility)
- Provide a synergy between GWAS and GMAS by providing alternative therapeutic targets to genes identified in GWAS

Acknowledgements

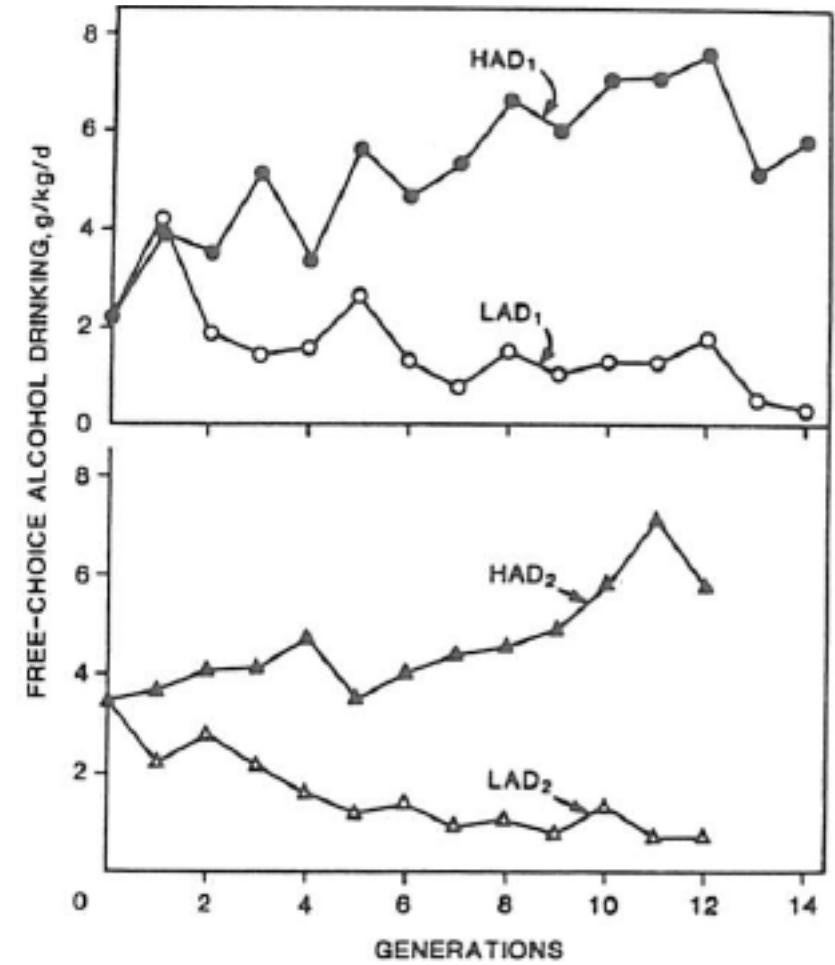
- Supported by:
 - NIAAA (AA013162, AA013162-08S1, AA016663)
 - INIA-West Pilot Grant
 - National Foundation for Prevention of Chemical Dependency Disease (NFPCCD Career Development Award)
 - Banbury Fund
- Special thanks to:
 - Dr. Boris Tabakoff, UCD
 - Dr. Stephen Flink
 - Spencer Mahaffey, ME
 - Lauren Vanderlinden, MS
 - Yinni Yu, MS
 - Dr. Paula Hoffman, UCD
 - Seija Tillanen
 - Dr. Katerina Kechris, CSPH

Example

Selective Breeding

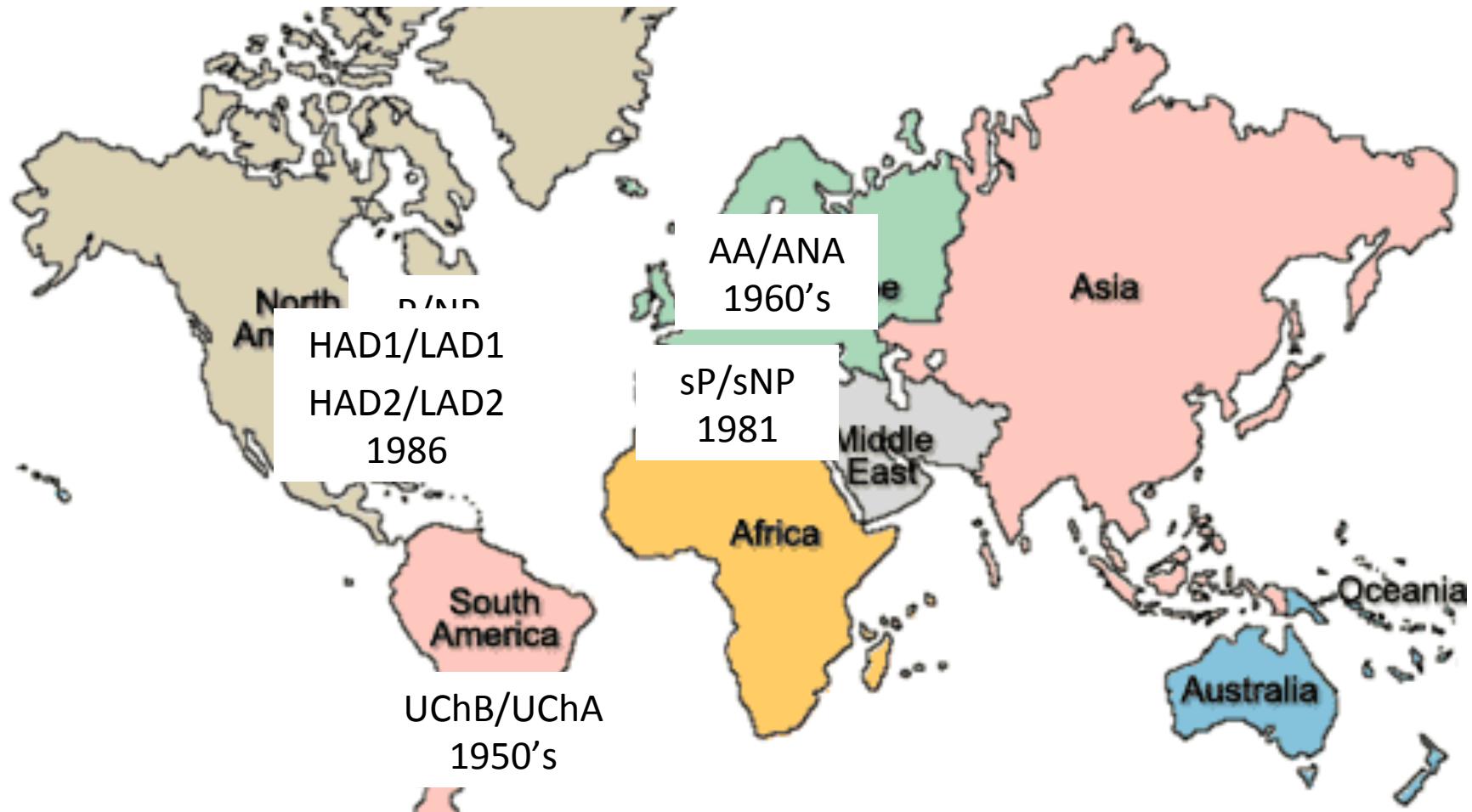


"HERE COMES THE PIMP."

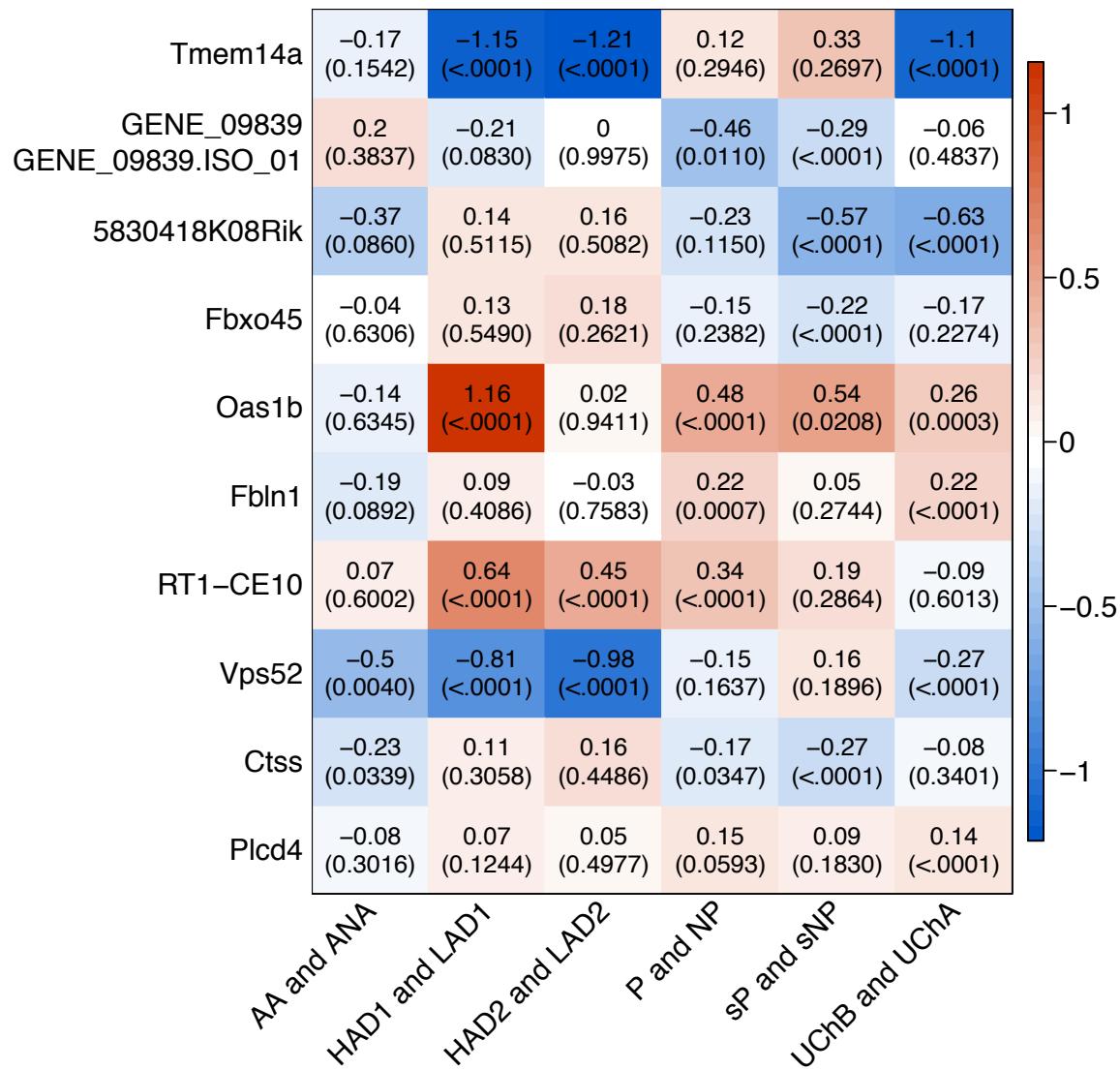


Li T-K, Lumeng L, and Doolittle DP (1993). Selective breeding for alcohol preference and associated responses. *Behavior Genetics* 23(2):163-170

Origin of Six Pairs of Selected Lines



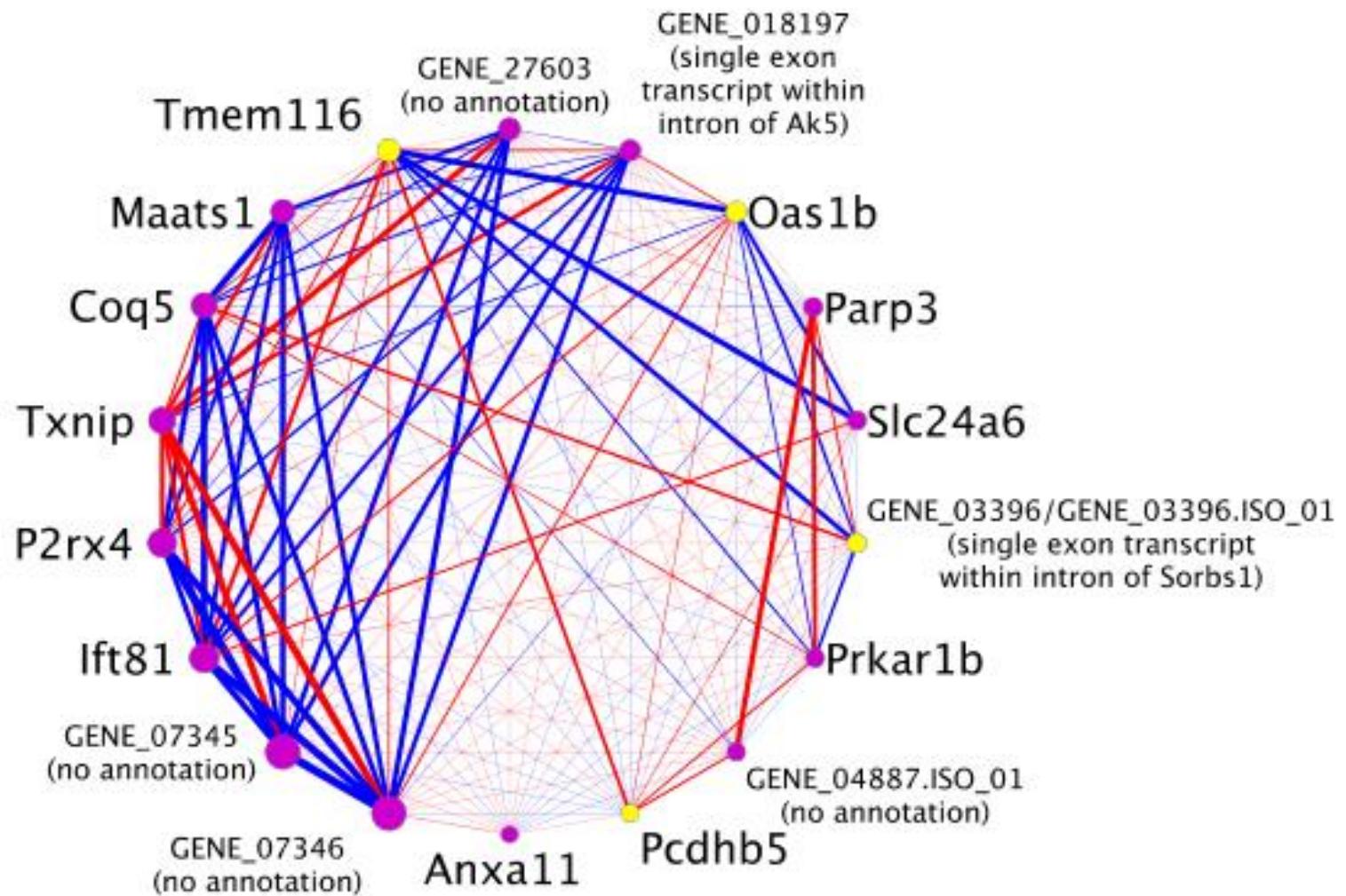
Genes/Isoforms differentially expressed between high alcohol consuming and low alcohol consuming selected lines of rats



Candidate Modules

1. Significant eQTL for the co-expression module within a QTL for alcohol consumption (**DNA → RNA**)
2. Module expression correlated with alcohol consumption and/or genes within the coexpression module tend to be differentially expressed in the selected lines (**RNA → phenotype**)

Transcriptional Pathway Associated With Alcohol Consumption



Biological Context from Pathway

