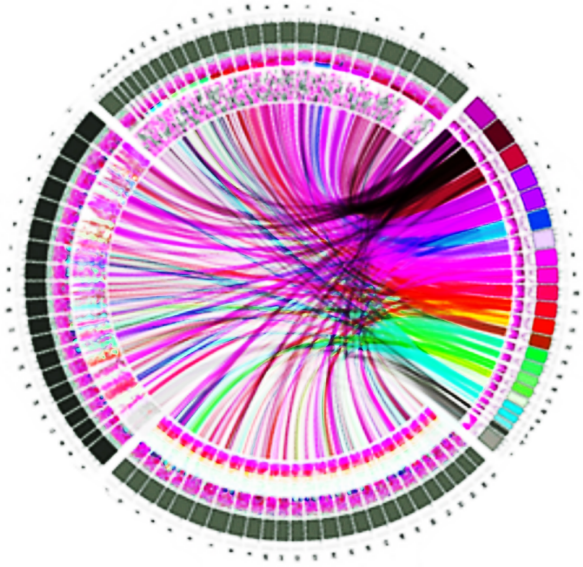
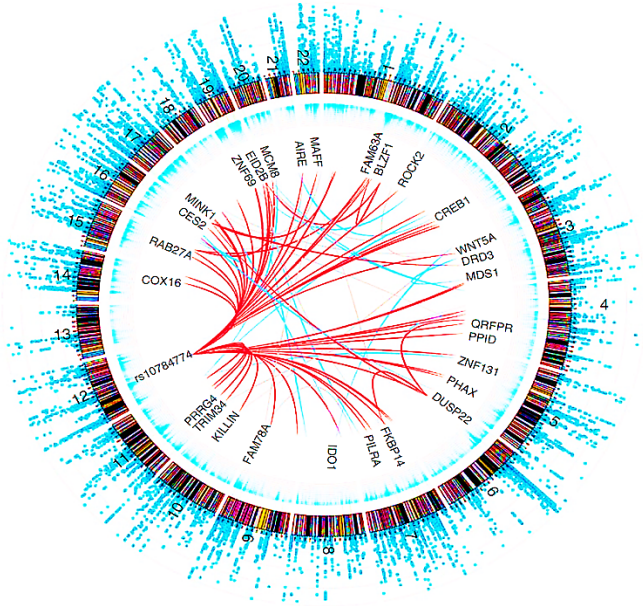
# INTRODUCTION TO SYSTEMS BIOLOGY:
## Bioinformatics 525



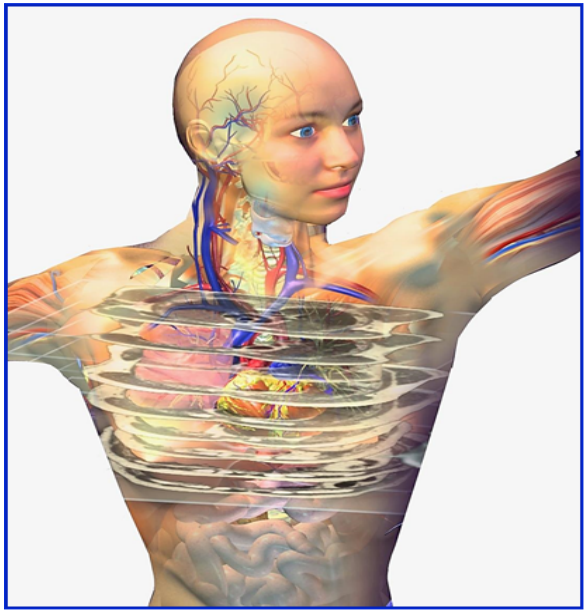


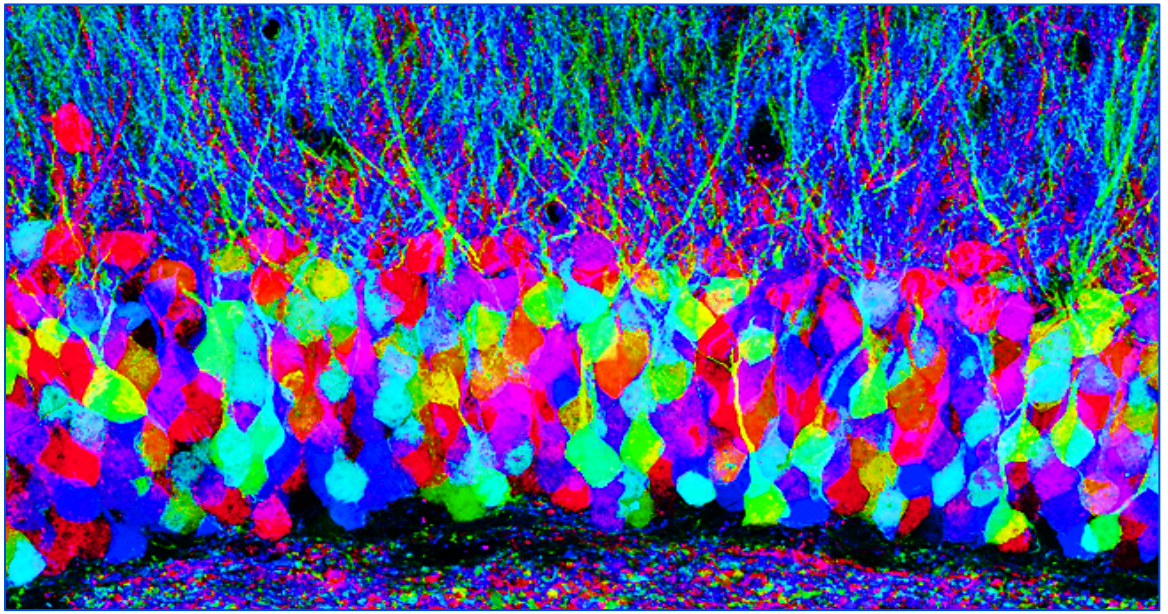

**Gerry Higgins, Ph.D., M.D.**

Adjunct Research Professor, Computational Medicine & Bioinformatics

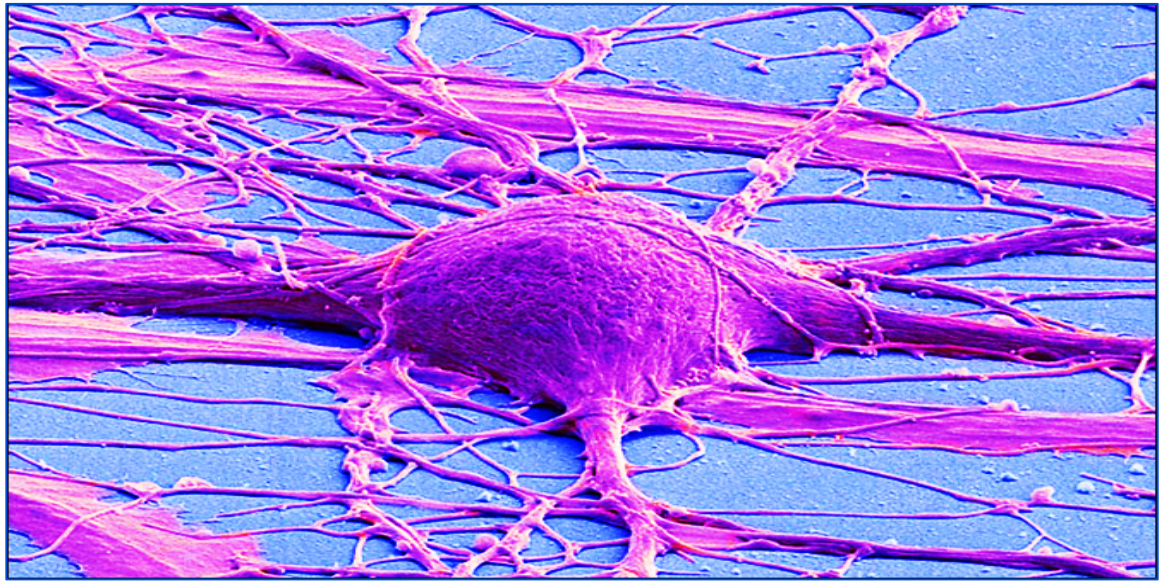
# TOPICS

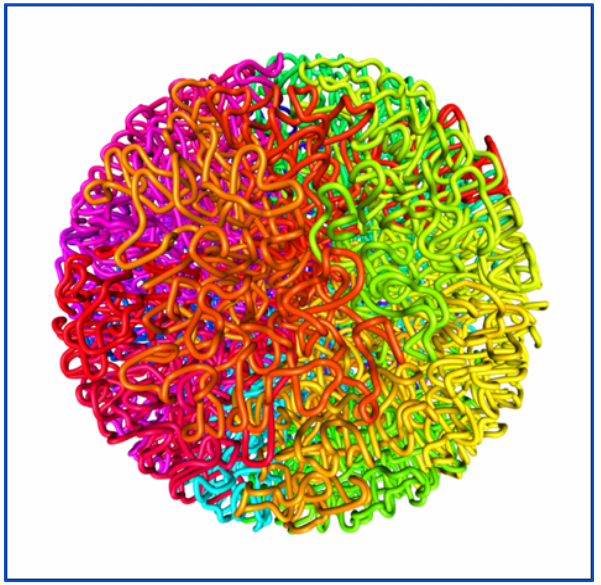
- **→Properties of Biological Systems**
- **→Analytic Methods in Systems Biology**
- **→Networks and Biological Systems**
- **→Case Examples:**
  - →The noncoding genome and transcriptional programs of gene regulation;
  - →Developmental programs of neurogenesis restore function in the damaged human brain;
  - →Systems pharmacology: New approaches for understanding adverse events associated with neuropsychiatric medications.
- **→ Tools, Resources – <u>how can these be applied to your research</u>?**
  - →Gene Set Enrichment Analysis (GSEA), Gene Ontology
  - →Pathway analysis
  - →GTEx Portal
  - →FactorBook, TransFac Pro®





# PROPERTIES OF BIOLOGICAL SYSTEMS





# Properties of Biological Systems

→**They are complex**, consisting of different components that may be similar to each other or not, simple or complex;

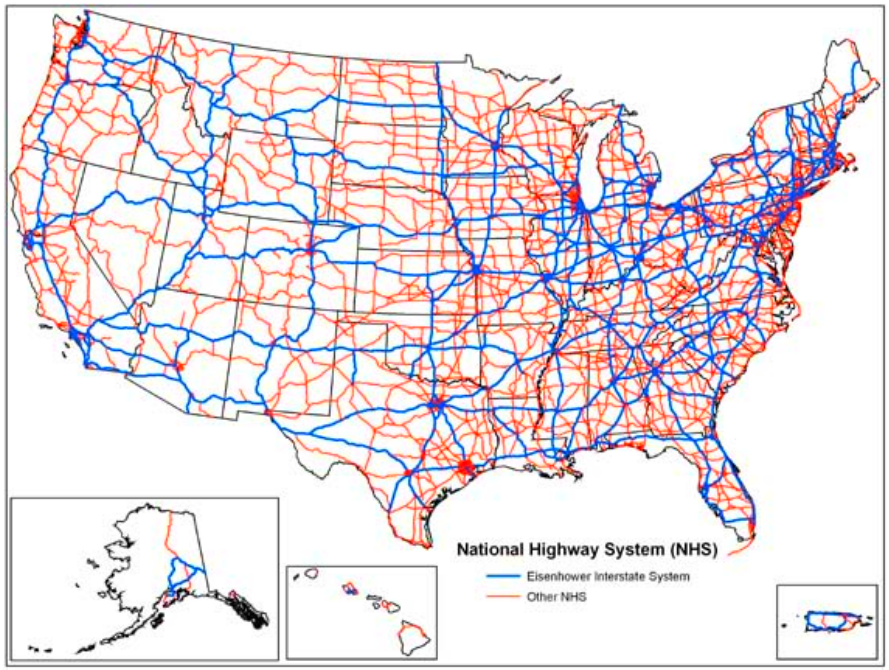
→**They exhibit non-linear dynamics which are not easily understood by humans** – sometimes modeling of biological systems requires computational approaches such as stochastic, multi-dimensional differential equations;

→**They are multi-scalar**, combining atomic, molecular, supramolecular, subcellular, cellular, physiological and behavioral components;

→**They exhibit emergent properties**, whose attributes cannot always be understood through decomposition and the use of reductionist approaches;

→**They become increasingly more complex as they are examined and understood**.

# Size does not equal complexity




U.S. highway network covers >3M miles, but functionality is easy to understand.

Size of web of *Araneus diadematus* is relatively small, but all of its properties cannot be accurately modeled using contemporary computational methods.

# The nonlinear properties of biological systems can often be modeled using more better and simpler linear approaches

→Linear system models: **(1)** Support a large array of mathematical techniques than current nonlinear math, and **(2)** Can often be used to analyze the nonlinearity of biological systems.

---

***EXAMPLE***

**Superposition:** If the Inputs and Outputs of a system can be summed-

$Input1 = Output1$ and $Input2 = Output2$ , then,

$Input1 + Input2 = Output1 + Output2$ .

---

→ Unfortunately, the dynamics of biological systems are nonlinear, *but* we can "flatten out" that part of the nonlinear system that is of interest;

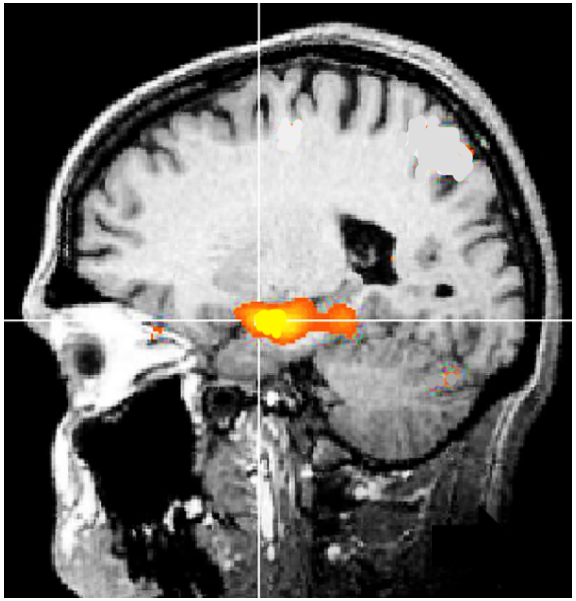
→This is justified by the theorem of Hartman and Grobman, which states that under most conditions, important clues about the behavior of a non-linear system can be deduced using a linear model **[1]**.

**[1]** Guckenheimer J, Holmes P. Nonlinear Oscillations, Dynamical Systems & Bifurcations of Vector Fields. Springer. 1983

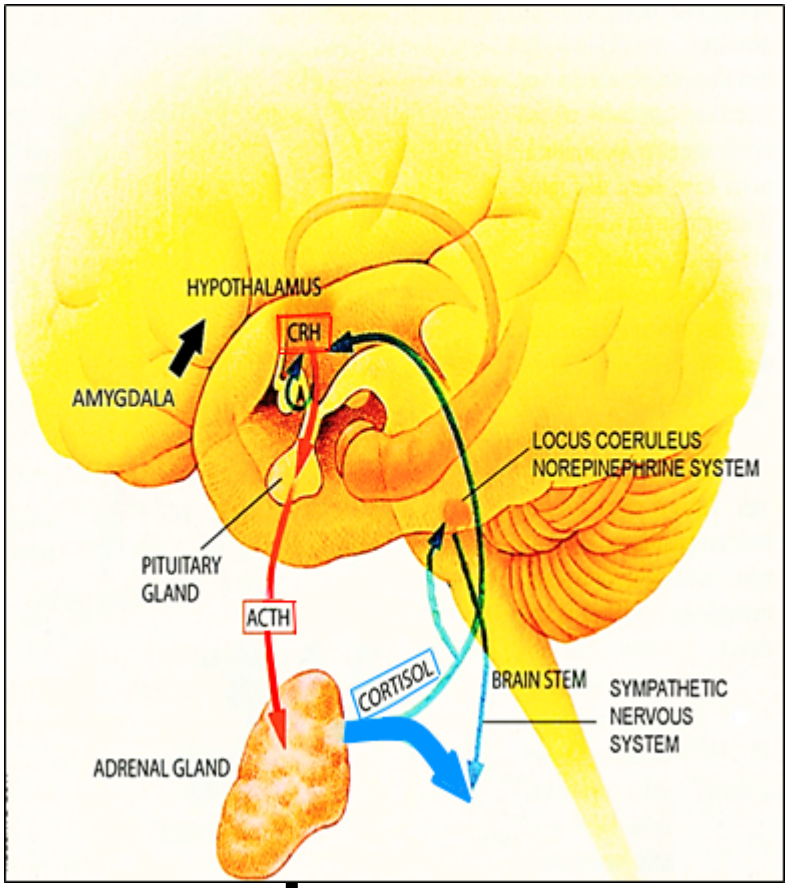
# Example of a Multiscale Biological System – The Stress Response

THREAT

AMYGDALA ACTIVATION

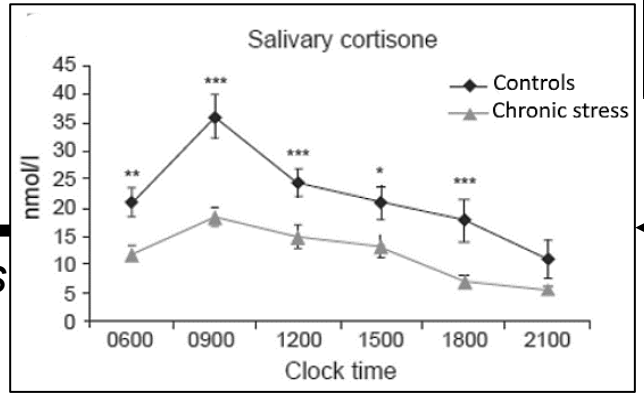


HPA AXIS

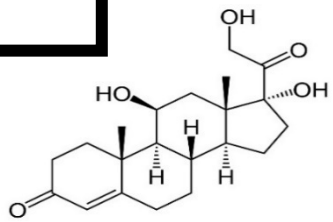



*Emergent properties*

*Temporal dynamics*




*Pathways & molecules*



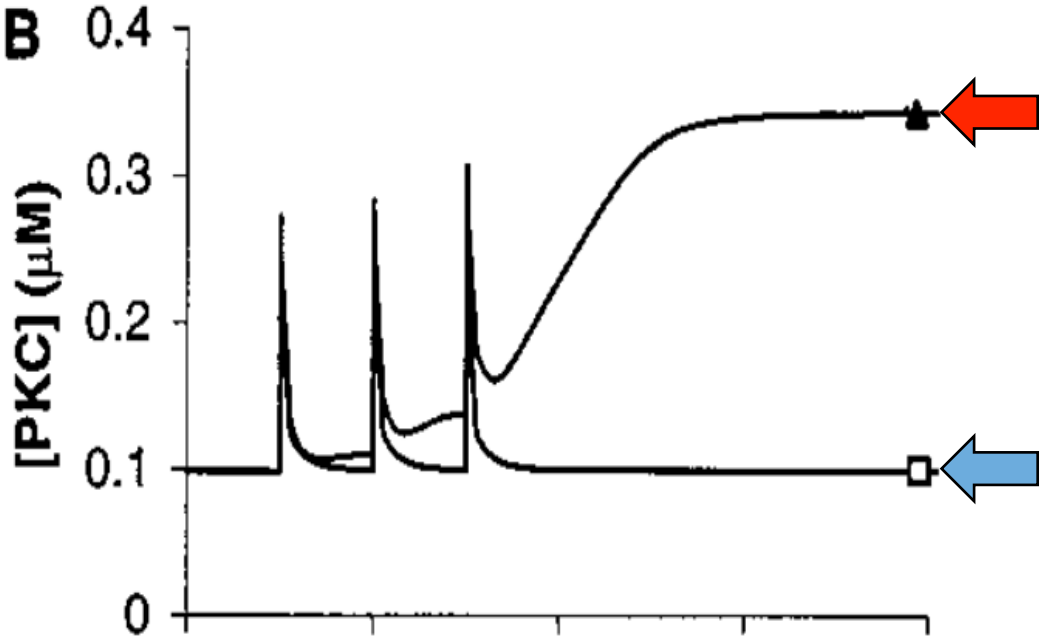

*Decades*

DISEASE

FLAT CORTISOL RESPONSE

CHRONIC CORTISOL

# Example of Emergent Properties of a Biological System – [Protein kinase C] in a glutamate receptor signaling pathway [1]




When feedback is present, successive spikes become larger.

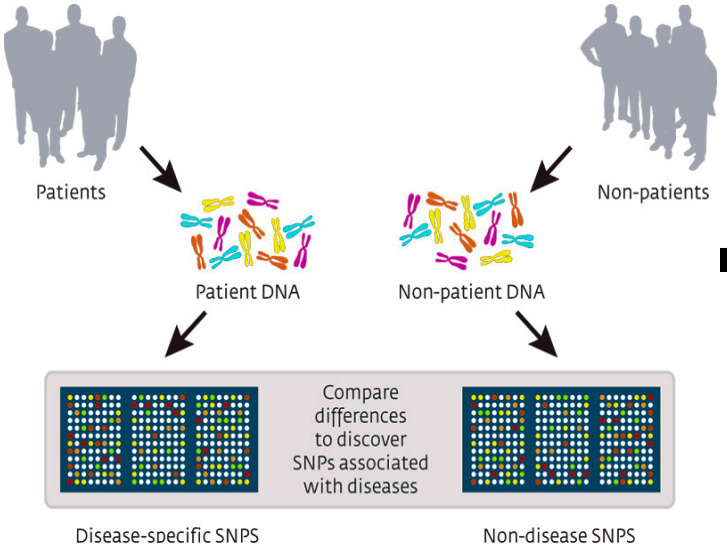
When feedback is absent, successive spikes flat-line.

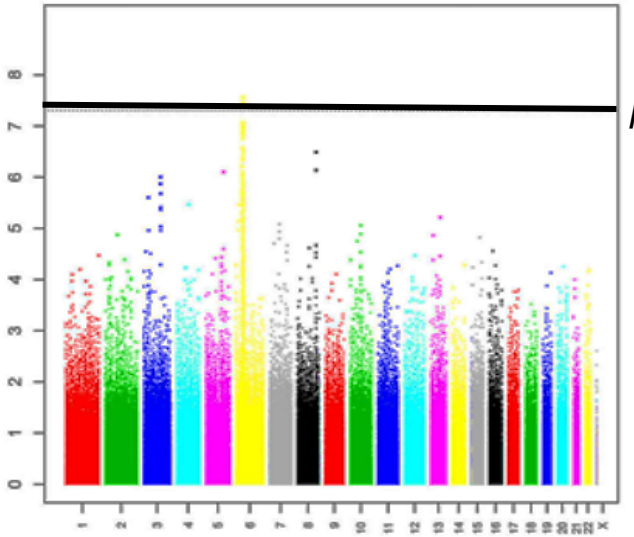
"These properties include the following: (i) Extended signal duration. The coupling of fast responses to the slow responses confers on the system the ability to regulate output for considerable periods after withdrawal of the initial signal. (ii) Activation of feedback loops…These properties of signaling networks raise the possibility that information for "learned behavior" of biological systems may be stored within intracellular biochemical reactions that comprise signaling pathways." **[1]**

**[1]** Bhalla US, Iyengar R. Emergent properties of networks of biological signaling pathways. *Science*. 283, 381-386 (1999).

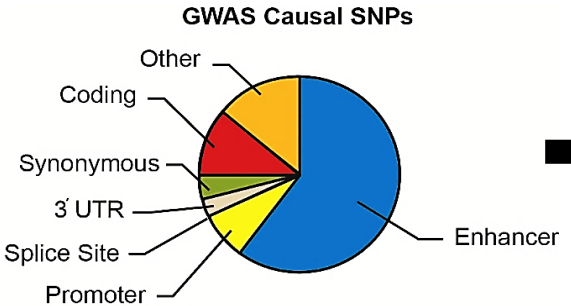
# Biological systems often become increasingly more complex as they are examined and understood – Genomewide association studies (GWAS)

**2006-2011**



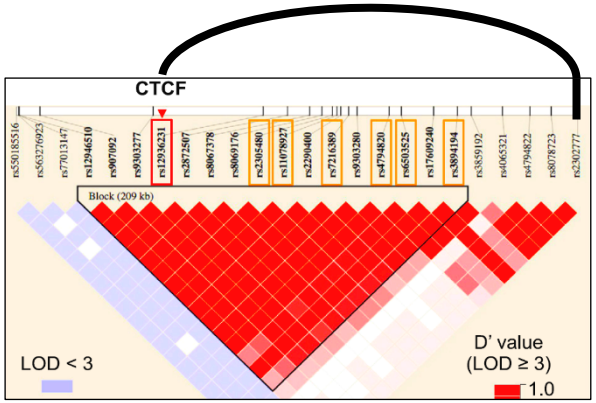

**2011-2016**




→Recognition that many genetic associations involve the regulatory, noncoding genome;

→Change in statistical threshold – e.g., SNPs could be rescued by pathway analysis. But, regulatory SNPs should be in high LD with reported 'lead' SNPs.

**2016 - ?**

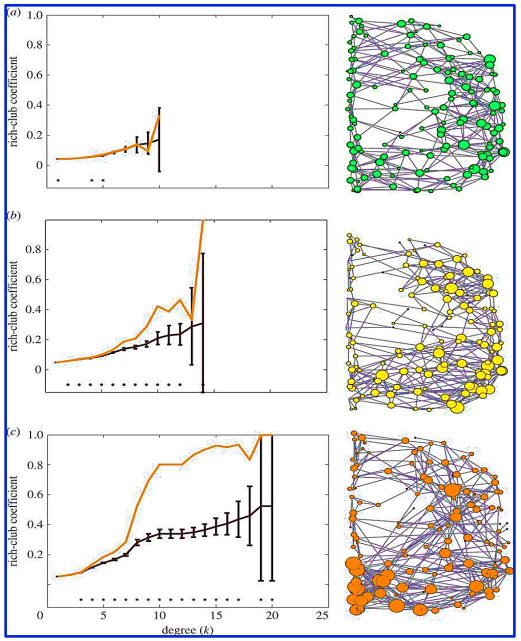


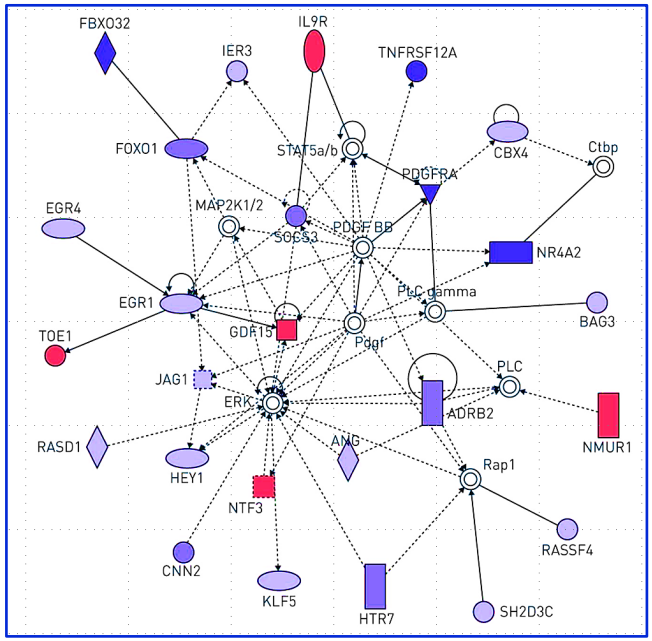

→Demonstration that SNPs connected by CTCF-associated histone loops can provide epigenomic "linkage" **with weak haplotype linkage [1]**;

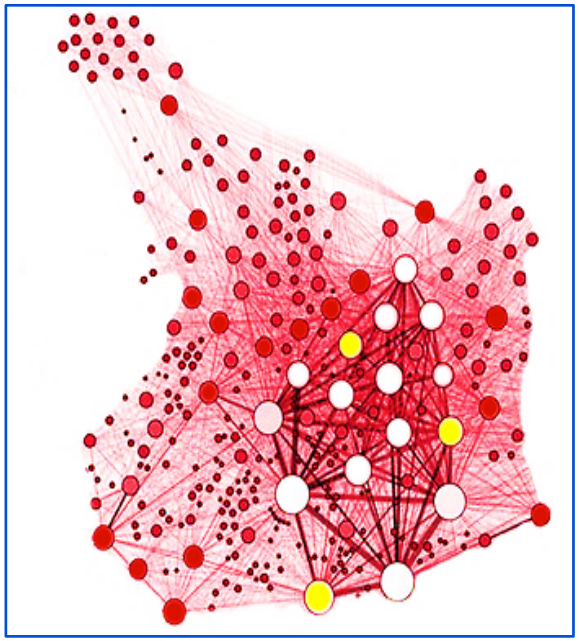
**[1]** Tang Z et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*. 163, 1–17 (2015).

# *More examples* - Properties of Biological Systems

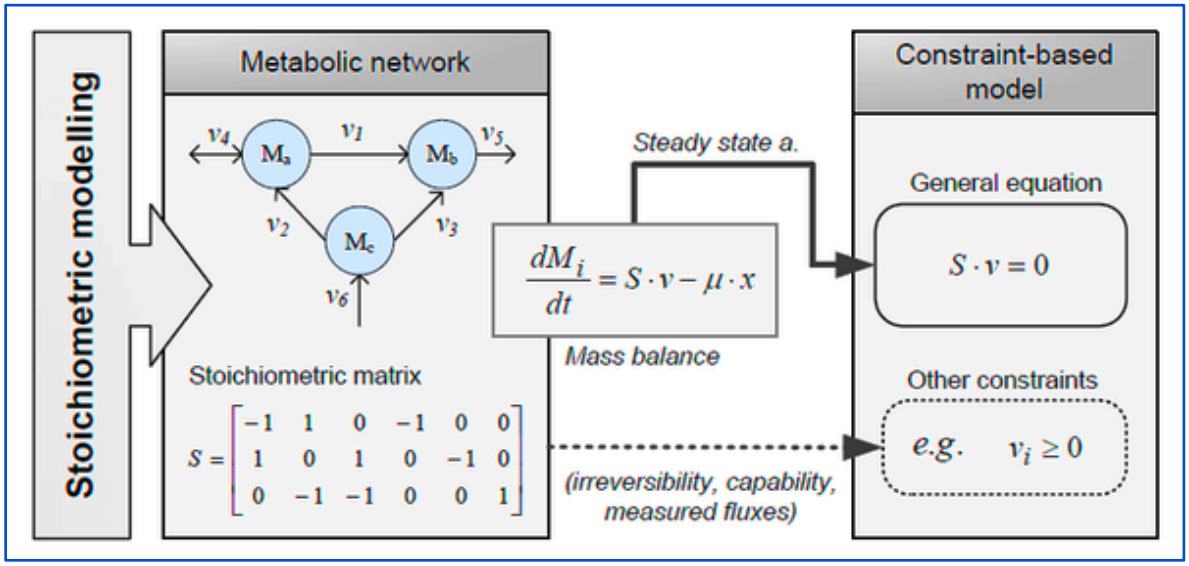
| PROPERTY | EXAMPLE | SCOPE |
|---|---|---|
| Complexity | Human brain | ~90 billion neurons that form 100 trillion synaptic connections |
| Temporal complexity | Ras protein clustering | 12-24nm in diameter; Lifetime of 100 msec |
| Tissue hysteresis | Titan protein in the cardiac wall | Responsible for 320 ± 46 pJ/mm2/ sarcomere |
| Bistable dynamics | On/off control of mitosis | Bistable steady-state response of Cyclin-Dependent Kinase 1 to phosphorylation by Cyclin B1 |
| Emergent property | Human brain | Human-specific **consciousness** |
| Emergent property | Schools of fish | **Self-organizing systems** that have not yielded to reductionist methods |
| Nonlinearity | Bat sonar | Can only be modeled by meta-heurisitic differential equations |

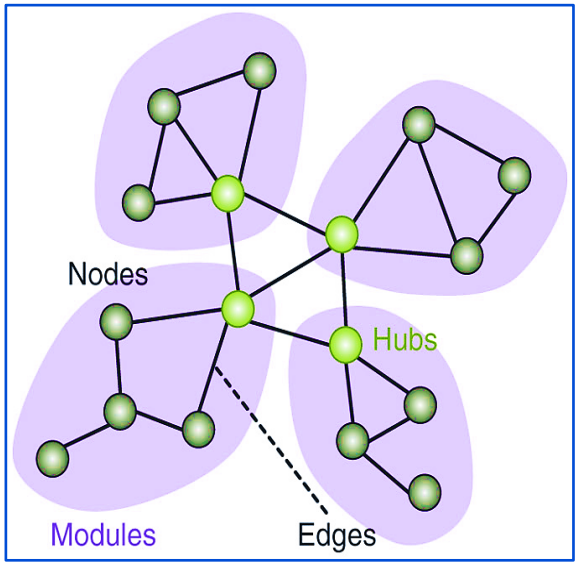


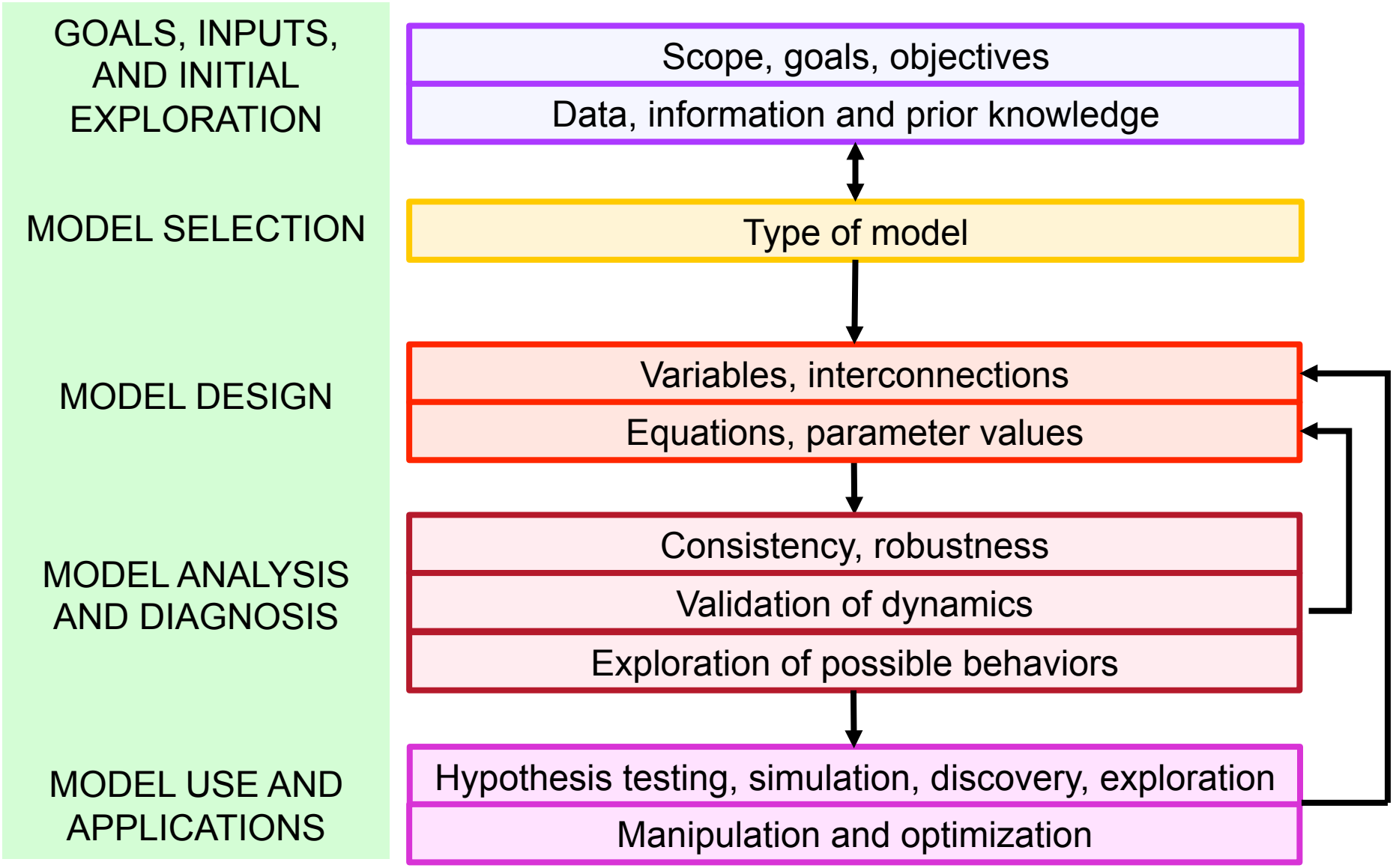



# ANALYTIC METHODS IN SYSTEMS BIOLOGY



# Model Development



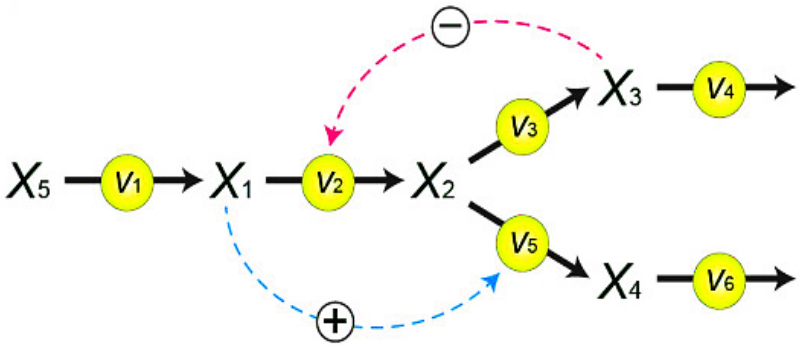

*Modified from*: E.O. Voit. A First Course in Systems Biology. Garland Science, Taylor and Francis Group. New York. ISBN 978-0-8153-4467-4 (2013).

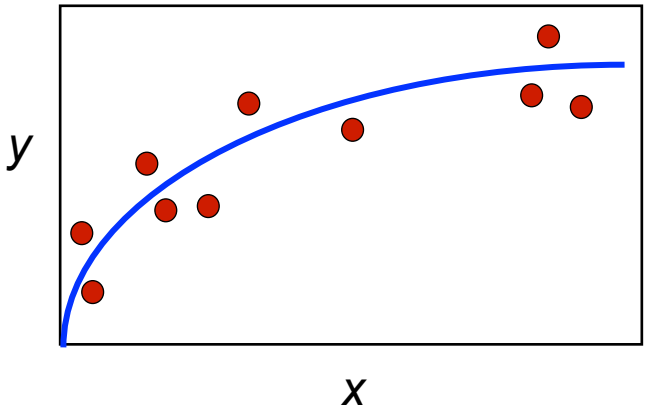
# Model Selection

| DETERMINISTIC | PROBABILISTIC / STOCHASTIC |
|---|---|
| Each element has a planned value in advance. | System elements are not random, but they are random variables drawn from a distribution. |
| Each element has a predecessor and a successor. | Three or more point estimates in the system model can be sampled to determine element duration random variables. |
| The longest path through the network is the critical path. | Critical path through network may change over time. |
| The total duration of the system has a fixed value, thus it is deterministic. | The total duration of the system time is random. |
| The total values of the system outputs is the sum of all outputs. | The total values of the system outputs is a random number. |

# Model Selection: Deterministic

| Type→ | MECHANISTIC | CORRELATIVE |
|---|---|---|
| Objective→ | Attempts to determine causation. | Association provides useful in biomedicine. |
| Attributes / Example → | Complex; Requires comprehensive & accurate data input; Causation hard to prove – changes over time. | Simple linear regression can be used. Can predict a patient's response to a drug when combining clinical and genotype data. |
| Use → | Rich data input; Numerically well defined variables & algorithms. | Sparse but adequate data available. Clinical/ scientific utility, does not prove causation. |

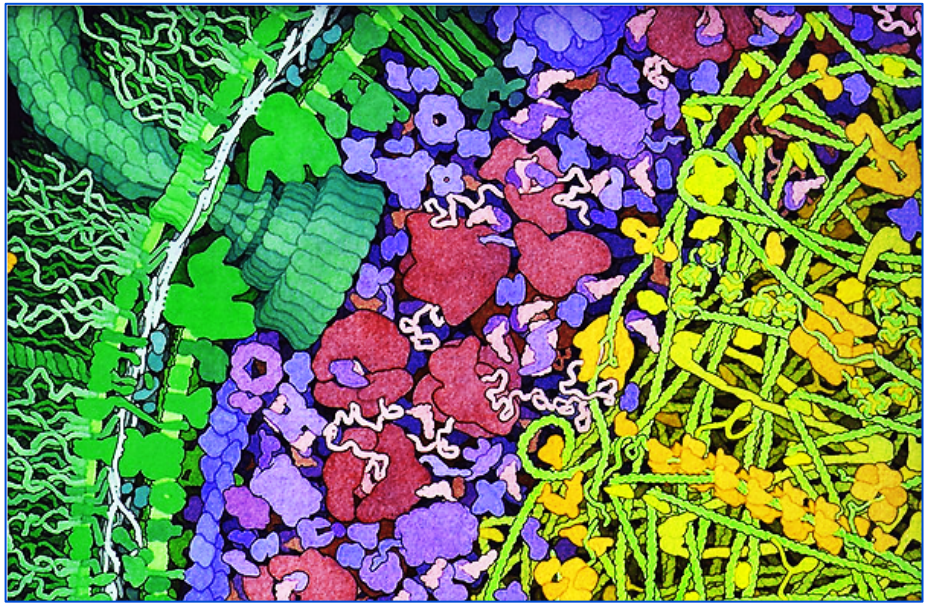


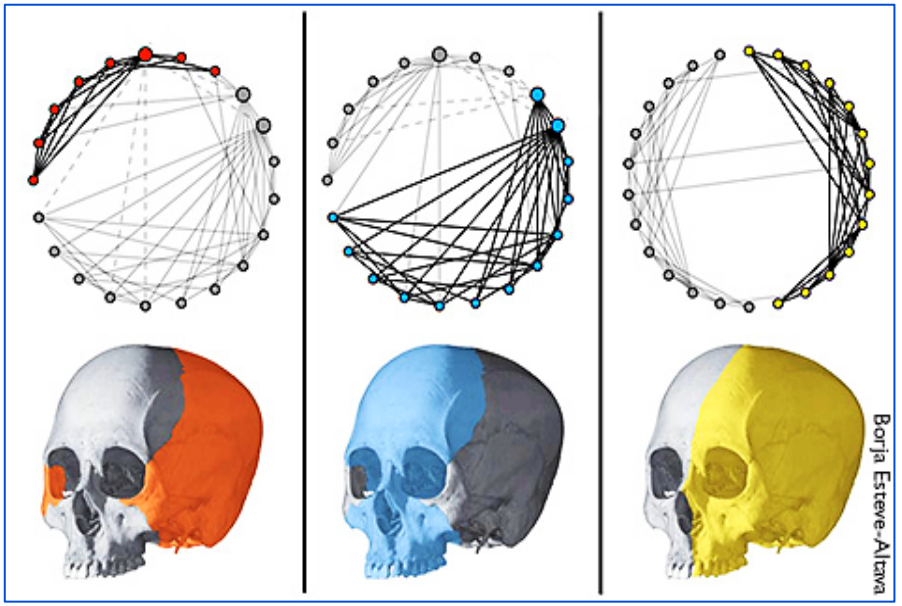




Can be used to explain how positive and negative feedback loops can change the product concentrations of a metabolic pathway.

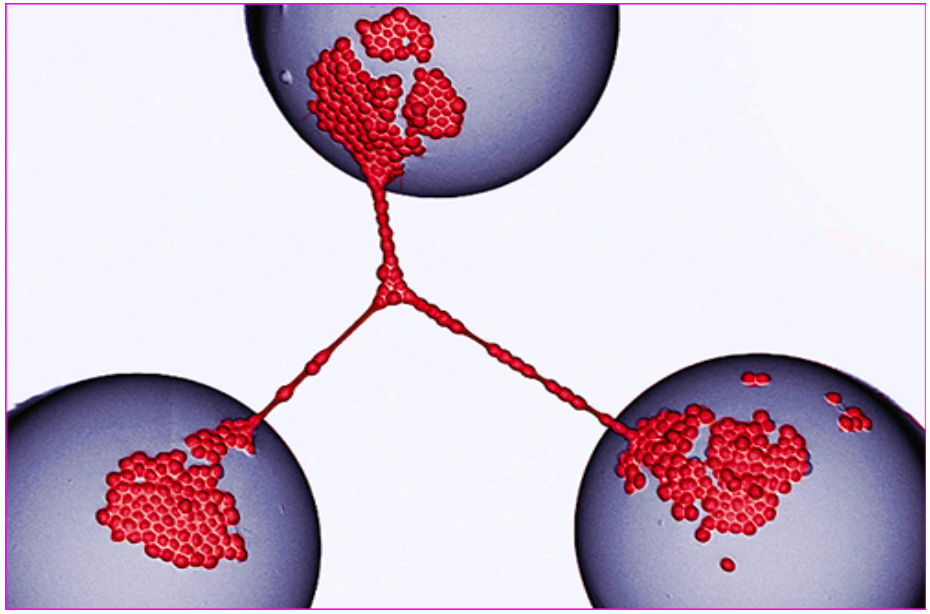
Whether linear or nonlinear, can robustly predict correlation between *x* and *y*, but does not explain connection between *x* and *y*.

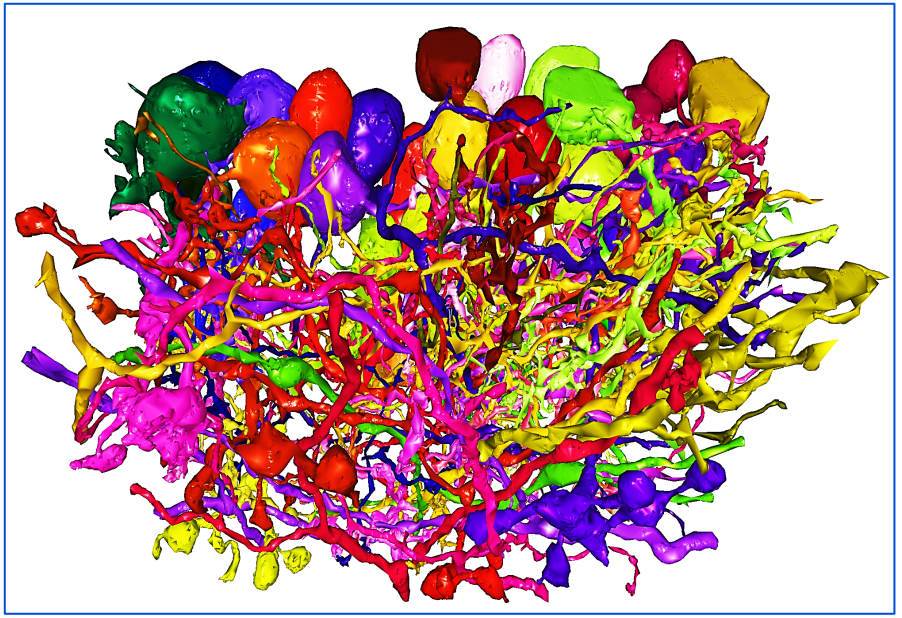
*Modified from*: E.O. Voit. A First Course in Systems Biology. Garland Science, Taylor and Francis Group. New York. ISBN 978-0-8153-4467-4 (2013).





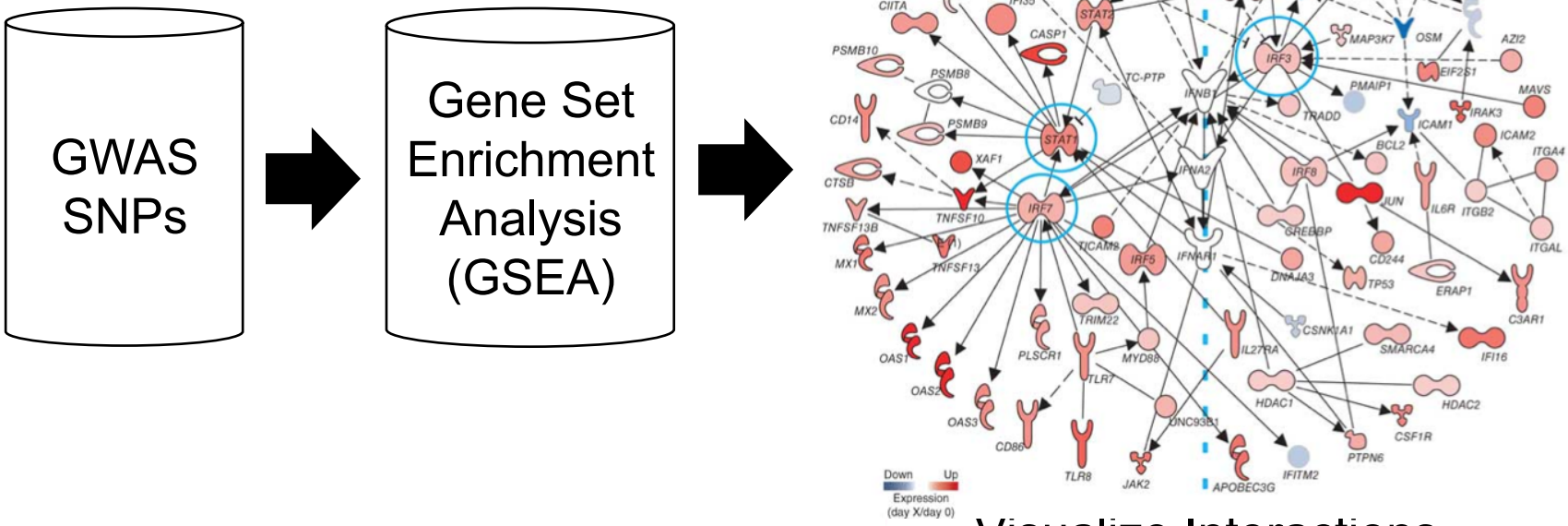

# NETWORK MODELS AND BIOLOGICAL SYSTEMS



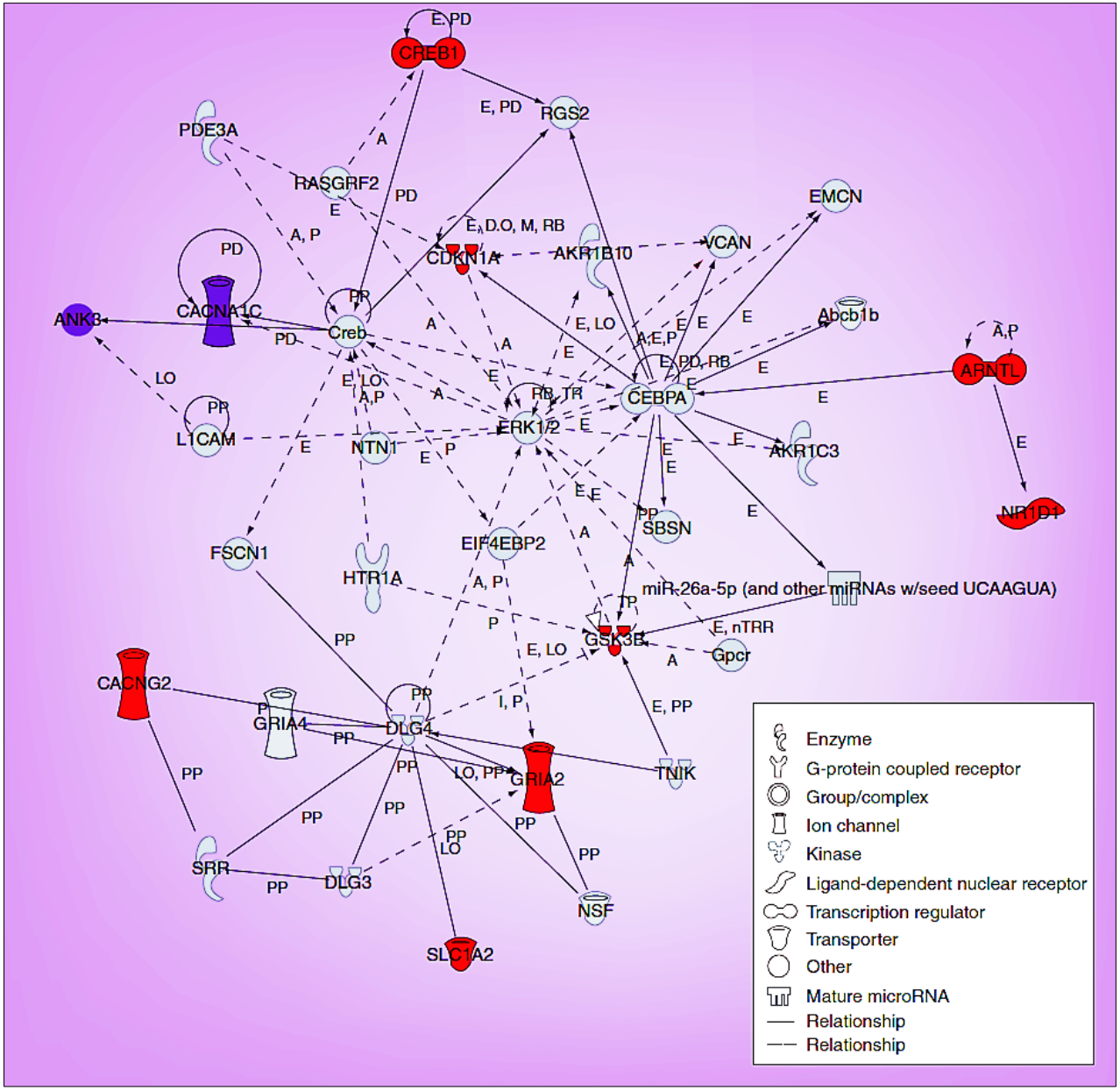


# Graph-based network models

- Commonly used to visualize interactions in a biological system;
- Used for modeling new and unknown interactions;
- Functional annotation of novel molecules using matching and clustering methods;
- Identification of metabolic signaling pathways;
- Detection of the centrality of protein and gene regulatory networks.

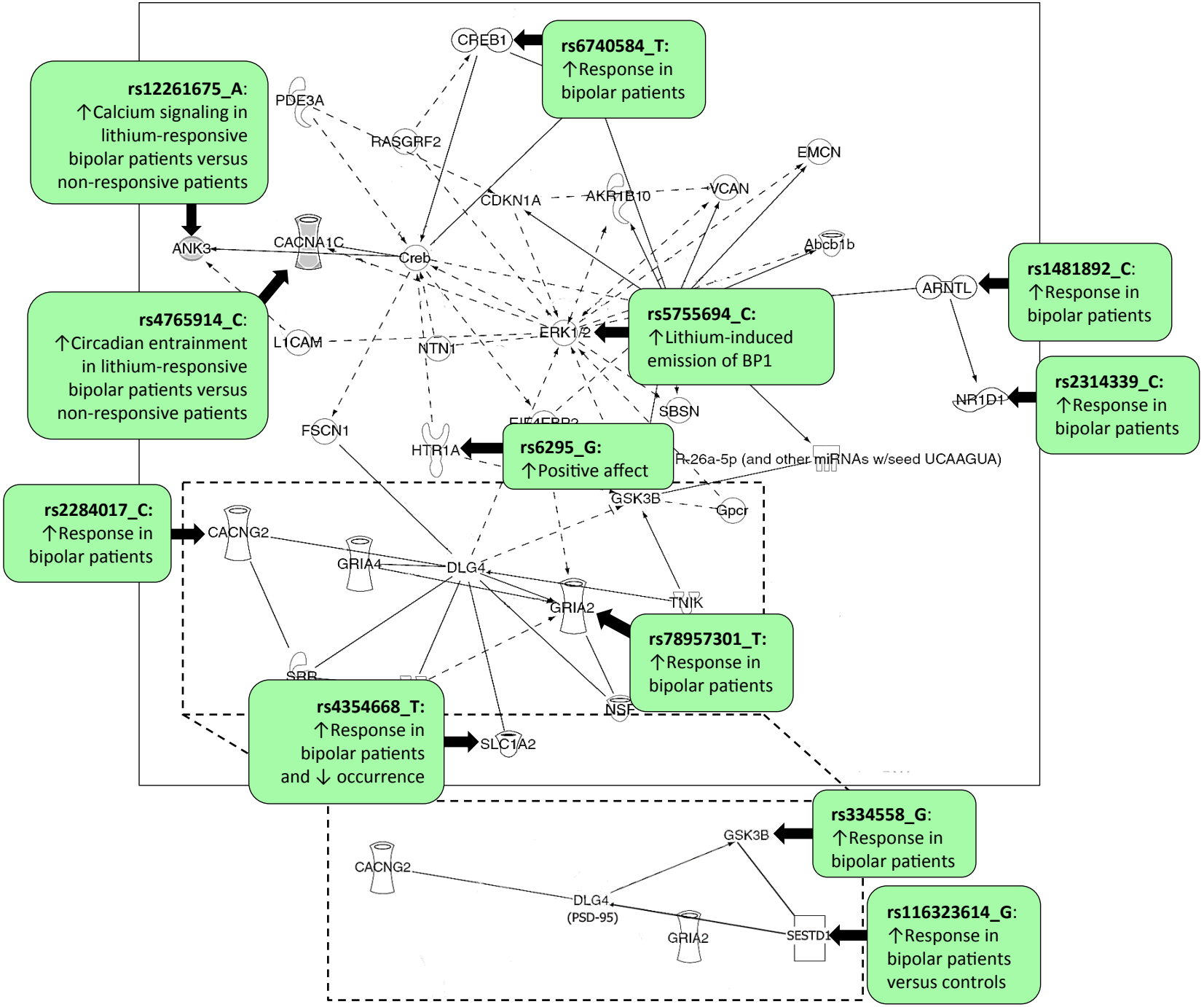


# SNPs that classify lithium response are from a single pathway in brain

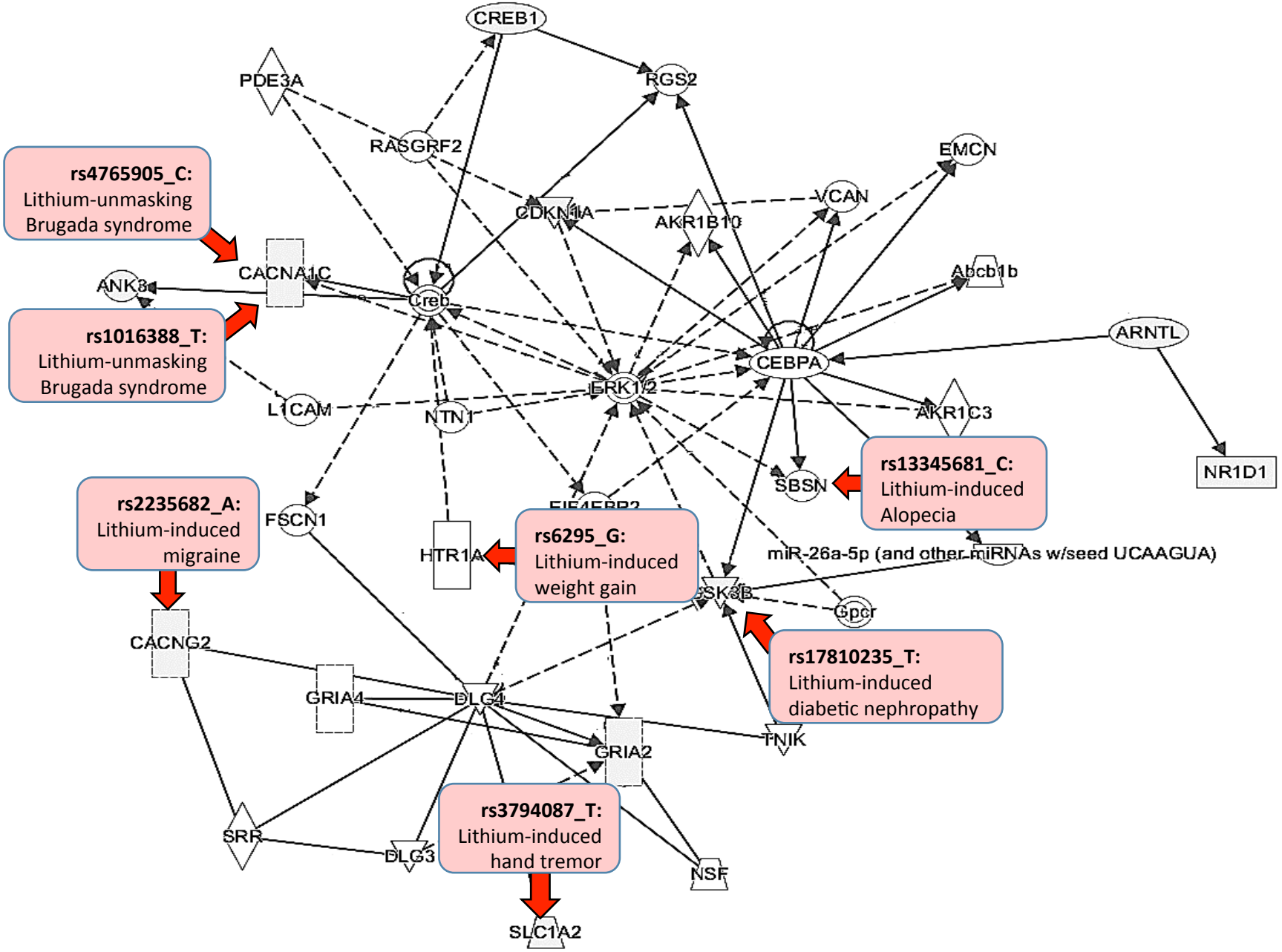


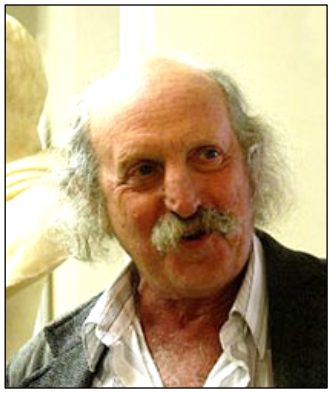

**[1]** Higgins GA, Allyn-Feuer A, Barbour E, Athey BD. A glutamatergic network mediates lithium response in bipolar disorder as defined by epigenome pathway analysis. *Pharmacogenomics.* 16(14), 1547–1563 (2015).

# All significant lithium pharmacogenomic response effects can be traced to this pathway



# SNPs in the lithium response pathway cause lithium-induced adverse events

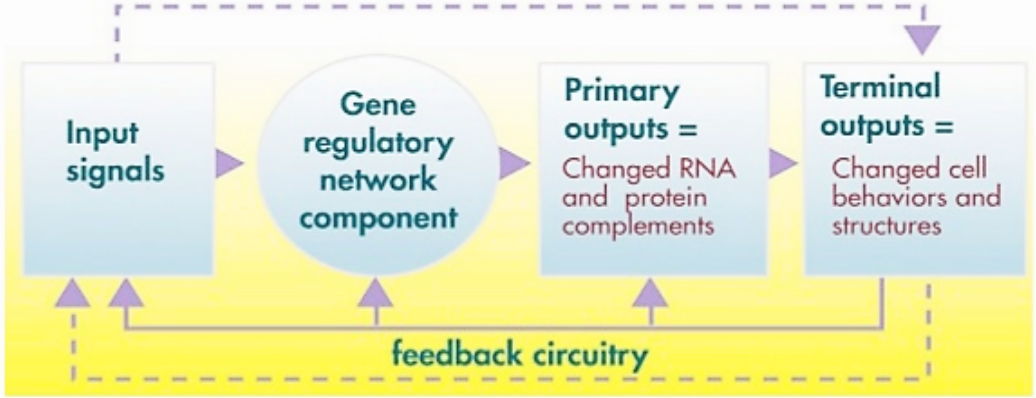


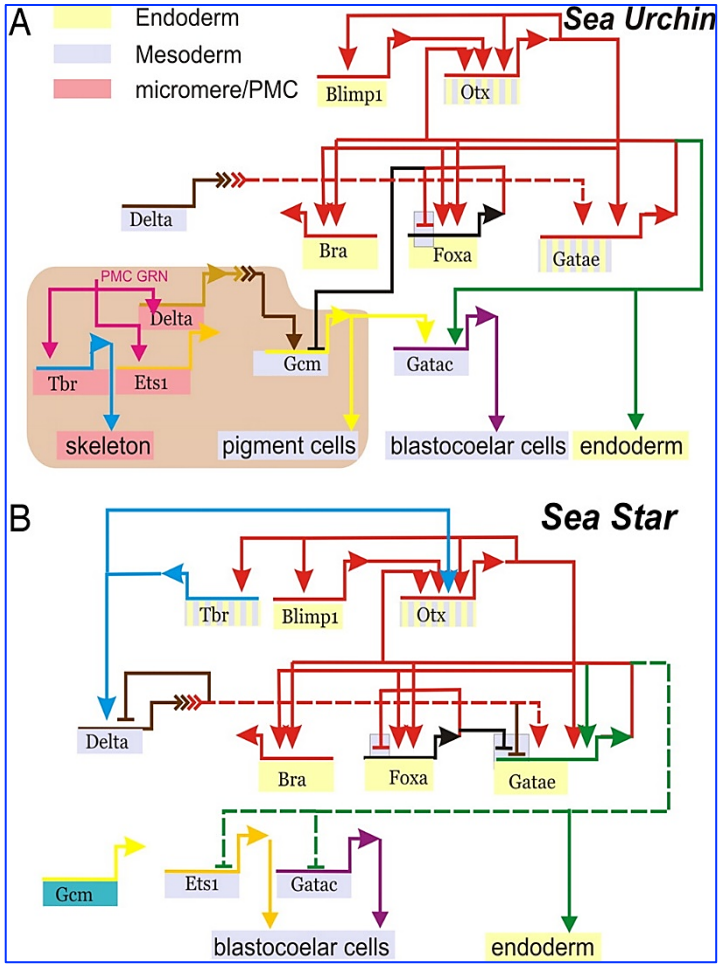


# Network Models: Gene Regulatory Networks (GRNs)

Eric Davidson, Caltech

- In the network
  - Nodes are Genes
  - Input is Transcription Factors (proteins)
  - Output is Gene Expression
  - Arrows show interaction



# Network Models: Weighted Co-Expression Analysis

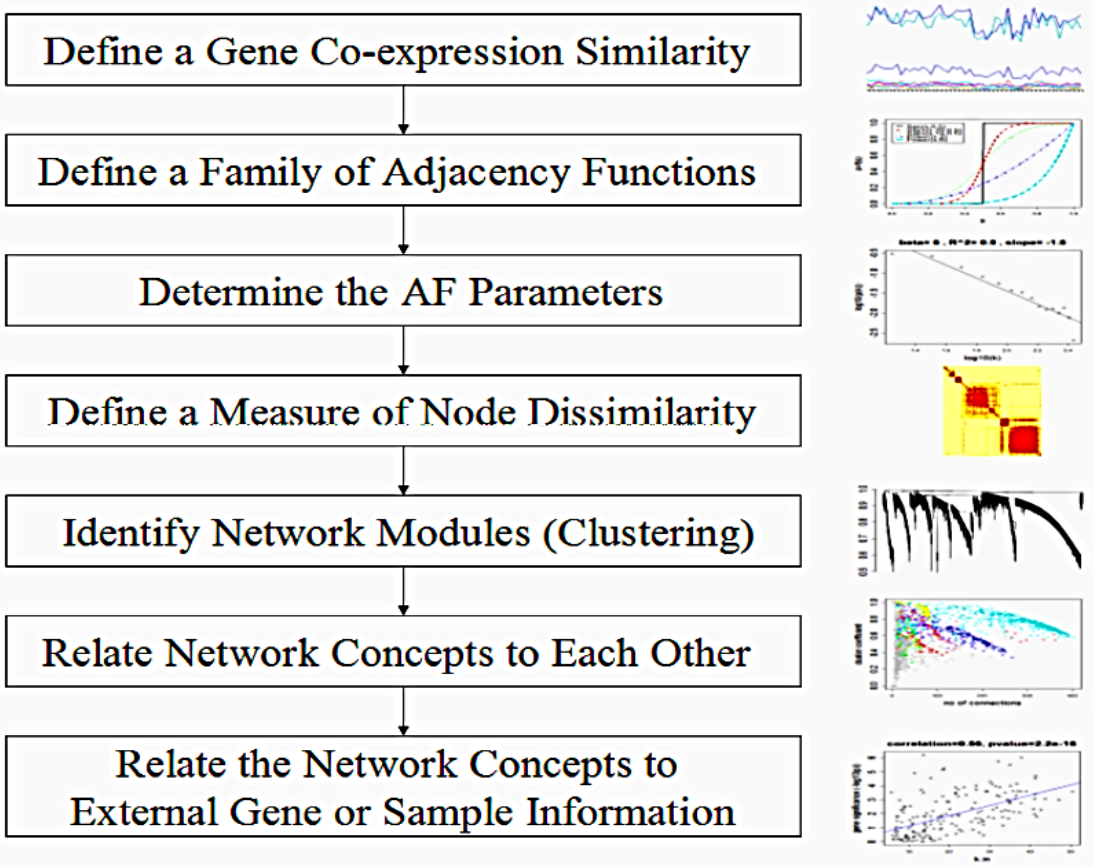



Figure 1: *Flowchart and illustration of gene co-expression network analysis. A typical figure has been placed to the right of each step.*

Zhang, B. and Horvath S. A general framework for weighted gene co-expression network analysis. (2005). *Statistical App. Genet. Mol. Biol.* 4(1), 1-43.

**Table 2.** Module eigengene significance for co-expression modules.

| | | Schizophrenia dataset | | Antipsychotic-free dataset | | |
|---|---|---|---|---|---|---|
| **WGCNA Modules** | **# genes** | **t** | **Adjusted p-value** | **t** | **Adjusted p-value** | **Expressed in brain** |
| Green | 367 | −6.26 | $3.8\times10^{-10}$ | −0.99 | $4.8\times10^{-01}$ | - |
| Magenta | 226 | 5.51 | $3.5\times10^{-08}$ | −0.24 | $9.6\times10^{-01}$ | - |
| **Tan** | **129** | **−4.92** | **$8.8\times10^{-07}$** | **−2.61** | **$4.8\times10^{-02}$** | **61%** |
| Red | 344 | −4.63 | $3.6\times10^{-06}$ | −1.60 | $2.2\times10^{-01}$ | - |
| Turquoise | 789 | 4.37 | $1.3\ x10^{-05}$ | 1.97 | $1.2\times10^{-01}$ | - |
| Yellow | 399 | −3.82 | $1.3\times10^{-04}$ | −1.26 | $3.6\times10^{-01}$ | - |
| Salmon | 121 | 3.02 | $2.5\times10^{-03}$ | 2.51 | $4.8\times10^{-02}$ | 52% |
| Blue | 610 | 2.95 | $3.2\times10^{-03}$ | 2.04 | $1.2\times10^{-01}$ | - |
| Cyan | 115 | 2.87 | $4.1\times10^{-03}$ | −0.13 | $9.6\times10^{-01}$ | - |
| Greenyellow | 197 | 2.59 | $9.7\times10^{-03}$ | −2.51 | $4.8\times10^{-02}$ | - |
| Black | 321 | −2.09 | $3.6\times10^{-02}$ | 0.33 | $9.6\times10^{-01}$ | - |
| Pink | 290 | −2.03 | $4.2\times10^{-02}$ | −0.05 | $9.6\times10^{-01}$ | - |
| Purple | 205 | −1.07 | $2.9\times10^{-01}$ | - | - | - |
| Brown | 447 | −0.24 | $8.1\times10^{-01}$ | - | - | - |

The modules that were found by WGCNA in the first dataset are listed together with the number of genes they contain (shown in the second column). Differences in cases and controls were tested using a linear model with FDR correction. Results for the medicated cases versus controls are presented in column three and four. The modules that were found to be differentially expressed were also tested for significance between cases and controls in the antipsychotic-free set, and results are presented in the fifth and sixth column. The last column indicates the percentage of module content that was also found to be expressed in brain ($\log_2>4$). For all genes in the other modules, this was found to be 45%. For the Tan module, this was significantly higher (Fisher $p=4.3\times10^{-4}$).

doi:10.1371/journal.pone.0039498.t002

De Jong S et al. A gene co-expression network in whole blood of schizophrenia patients Is independent of antipsychotic use and enriched for brain-expressed genes. (2012) *PLoS One*. 7 (6), e39498

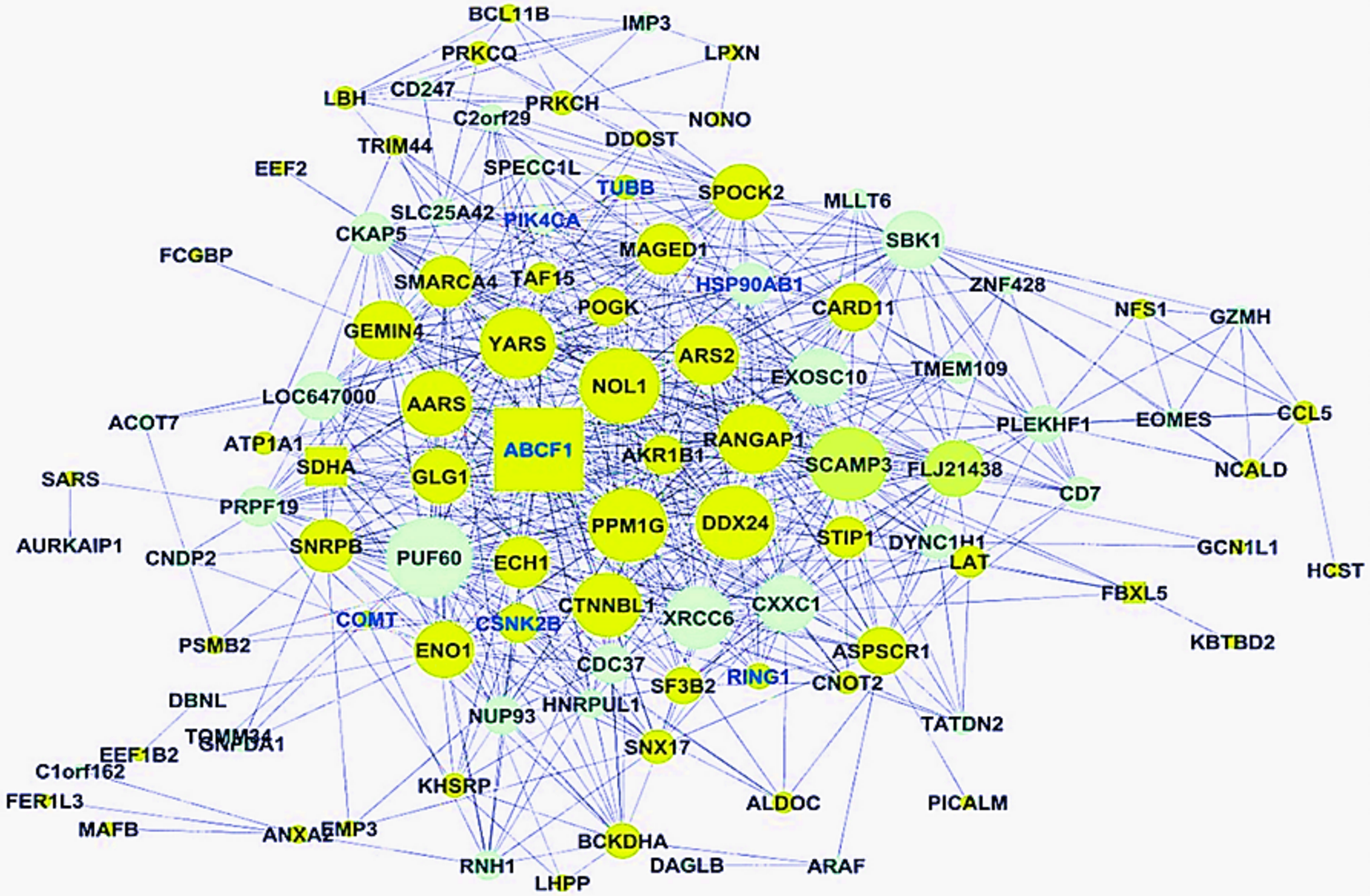
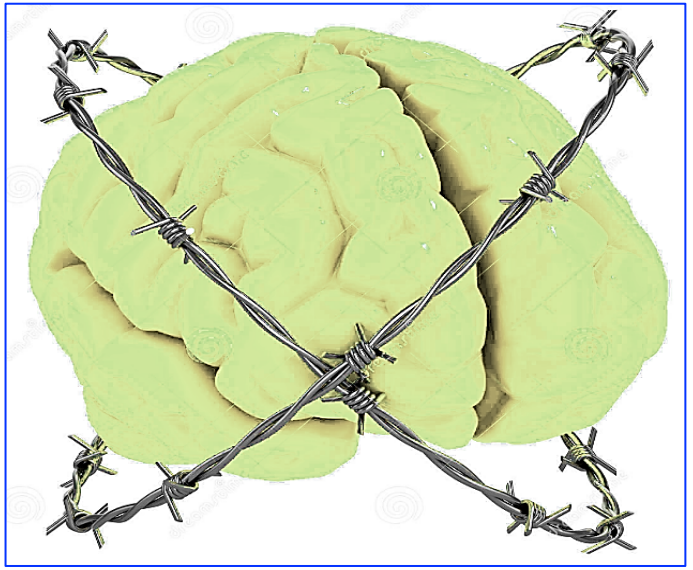
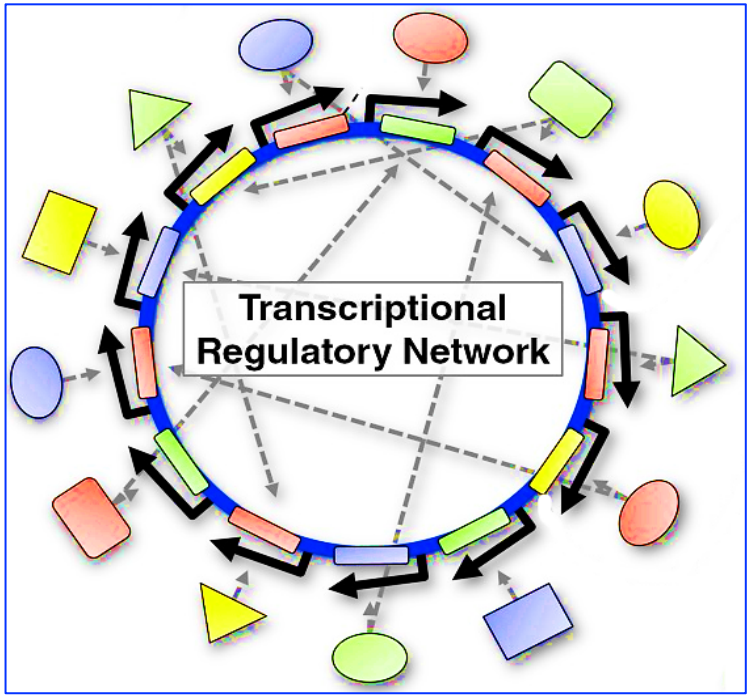


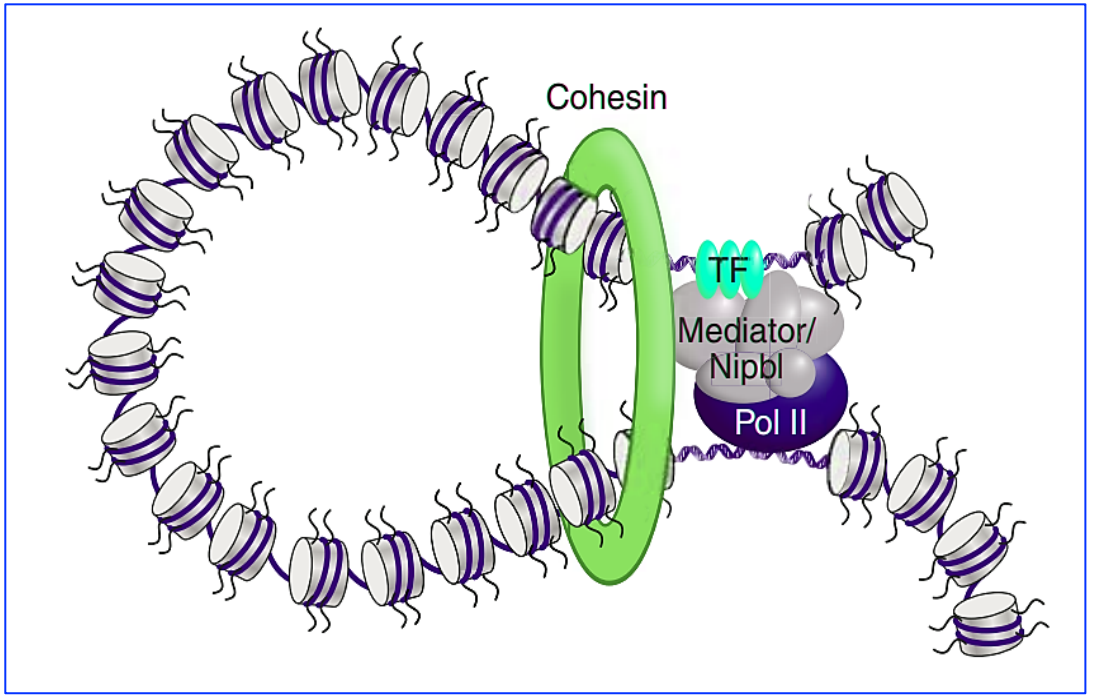

**Figure 2. Visual representation of connections of genes in the Tan schizophrenia module.** This figure shows target genes of the probes in the Tan schizophrenia module with the strongest connections only ($r > 0.64$). Blue-colored nodes represent brain-expressed genes. Square-shape nodes indicate *cis*-regulation. Node size is related to the number of connections of that particular gene; a highly connected gene (i.e. 'hub gene') is therefore larger than genes with fewer connections. Red text indicates genes previously implicated in schizophrenia. Image created using Cytoscape software [69].

doi:10.1371/journal.pone.0039498.g002



# CASE EXAMPLES



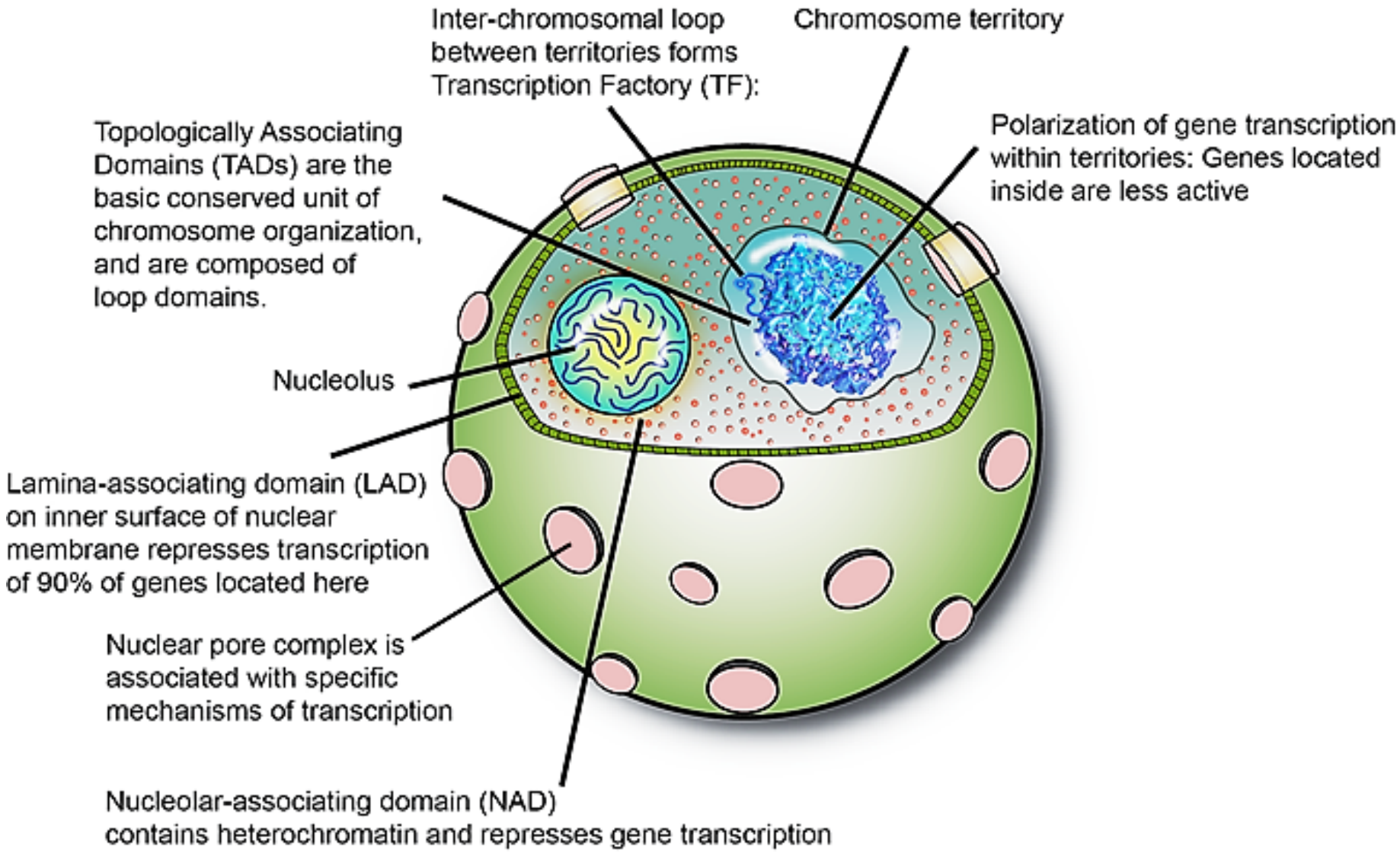
# Case Examples

→The noncoding genome and transcriptional programs of gene regulation;

→Developmental programs of neurogenesis that restore function in the damaged human brain;

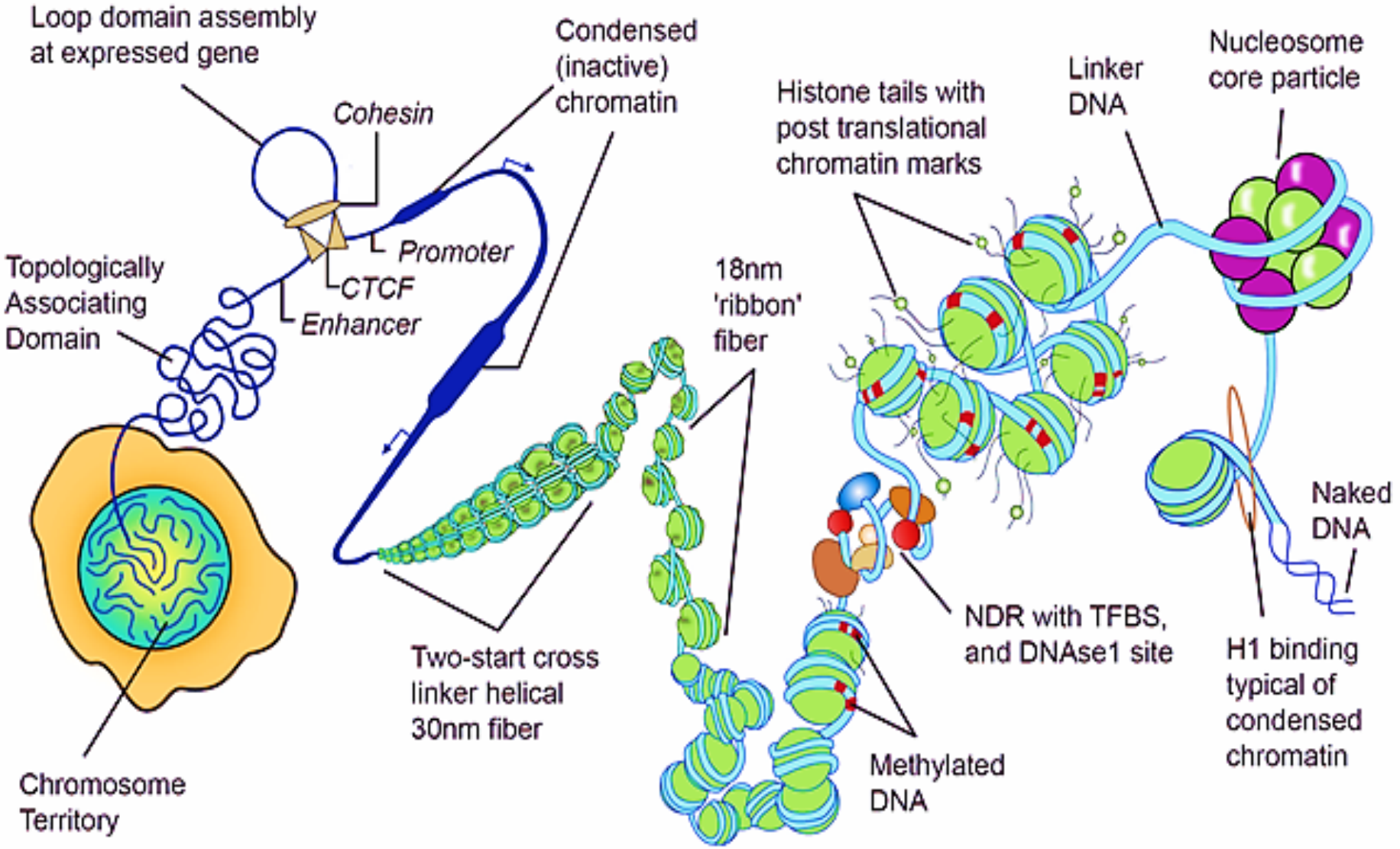
→Systems pharmacology: New approaches for understanding adverse events associated with neuropsychiatric medications.

# Spatial distribution of transcriptional domains in the nucleus




*From:* Higgins, GA et al. The epigenome, 4D nucleome and next-generation neuropsychiatric pharmacogenomics. *Pharmacogenomics.* (2015) 16(14), 1649–1669.

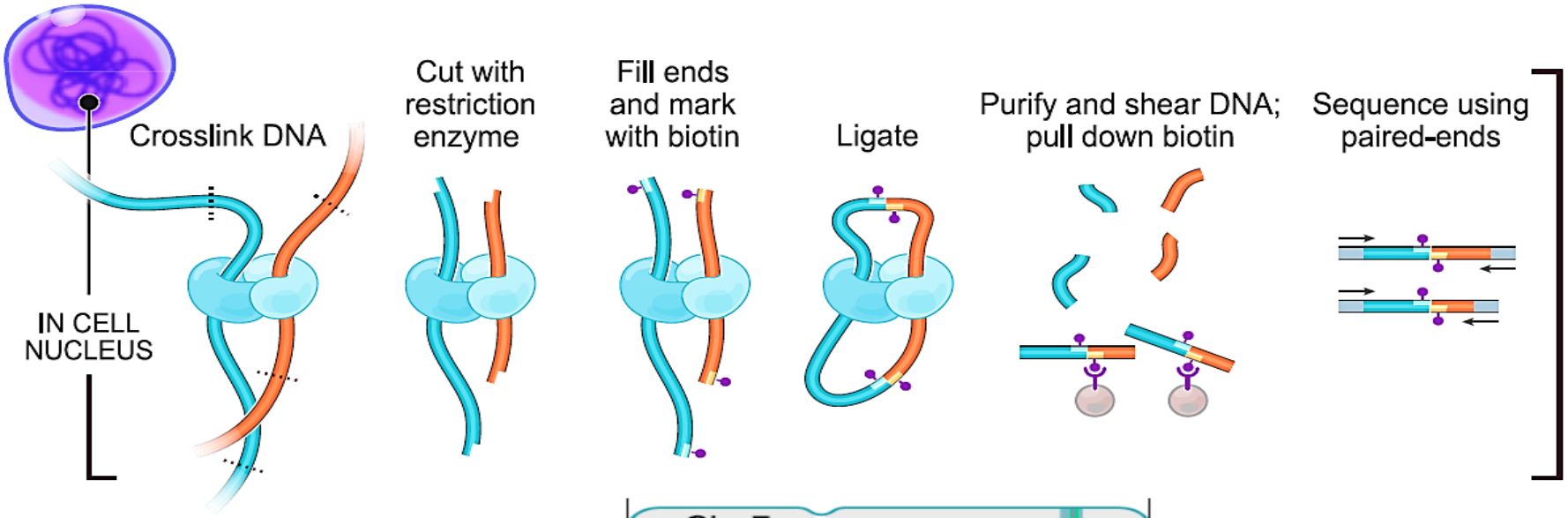
# The epigenome regulates gene transcription




*From:* Higgins, GA et al. The epigenome, 4D nucleome and next-generation neuropsychiatric pharmacogenomics. *Pharmacogenomics.* (2015) 16(14), 1649–1669.

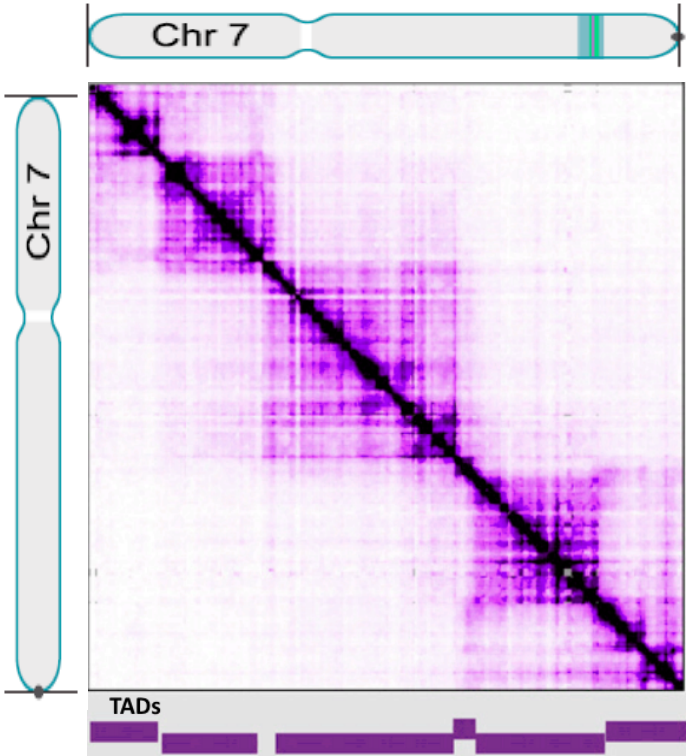
# Hi-C chromatin capture defines the spatial organization of the human genome as topologically-associating domains (TADs)

**A**

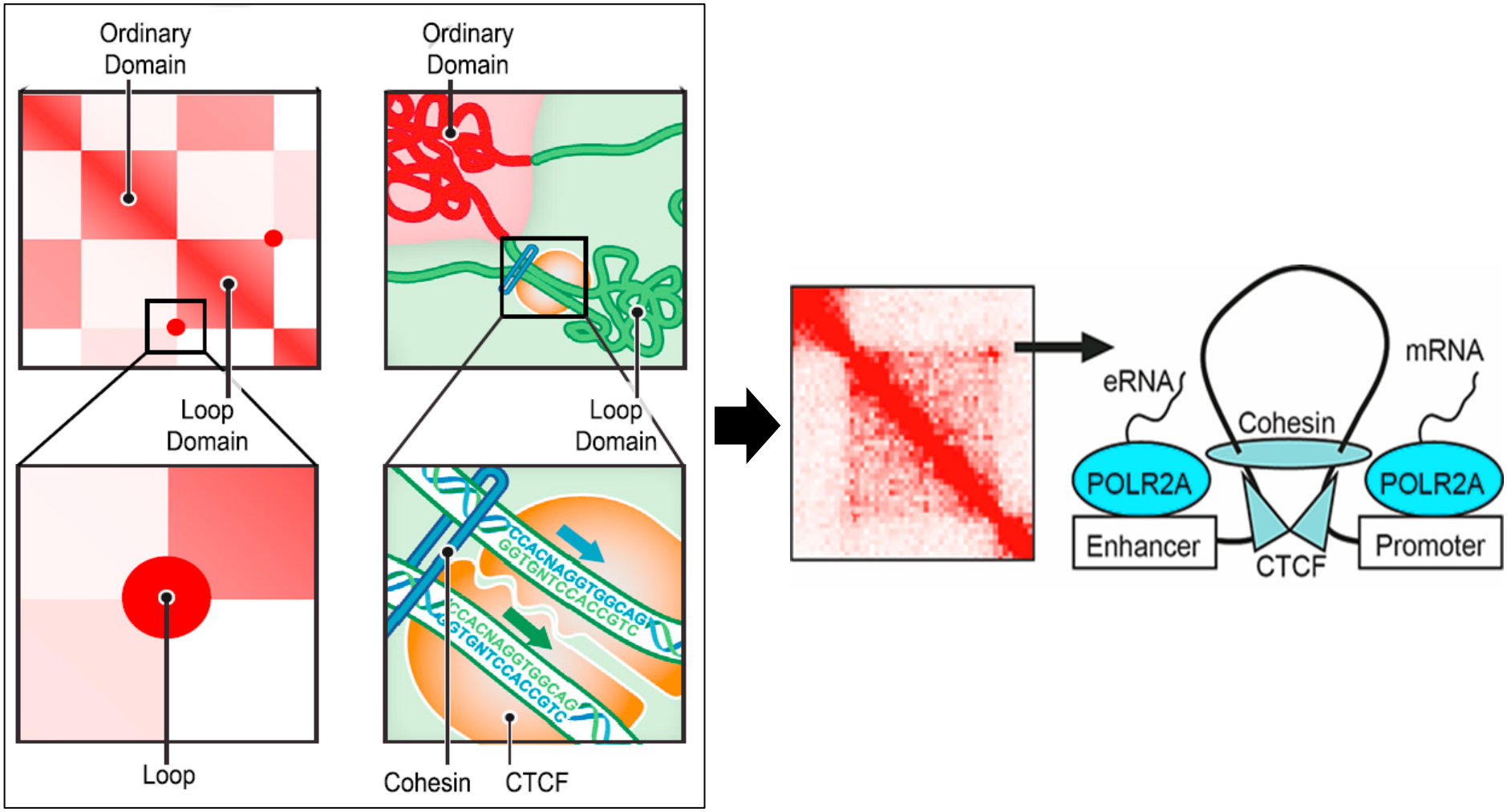



**B**

Plot graph of intra-chromosomal interactions, and you find TADs!



# Topologically-associated domains (TADs) are the basic unit of transcription and contain enhancer-promoter chromatin loops



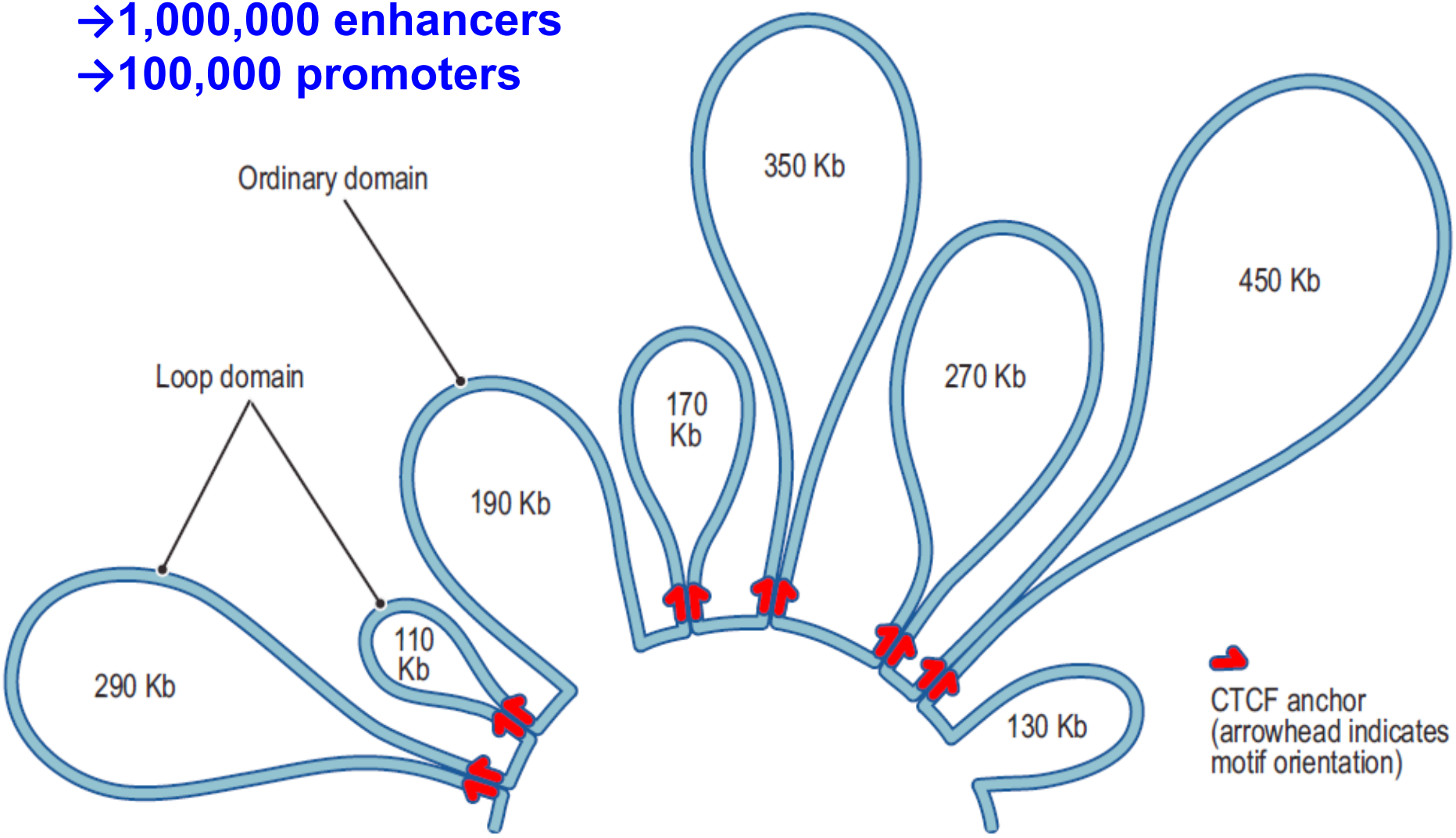

*From:* Higgins, GA et al. The epigenome, 4D nucleome and next-generation neuropsychiatric pharmacogenomics. *Pharmacogenomics.* (2015) 16(14), 1649–1669.

# Control of gene transcription largely takes place in these enhancer-promoter loops

In the human genome, there are approximately:

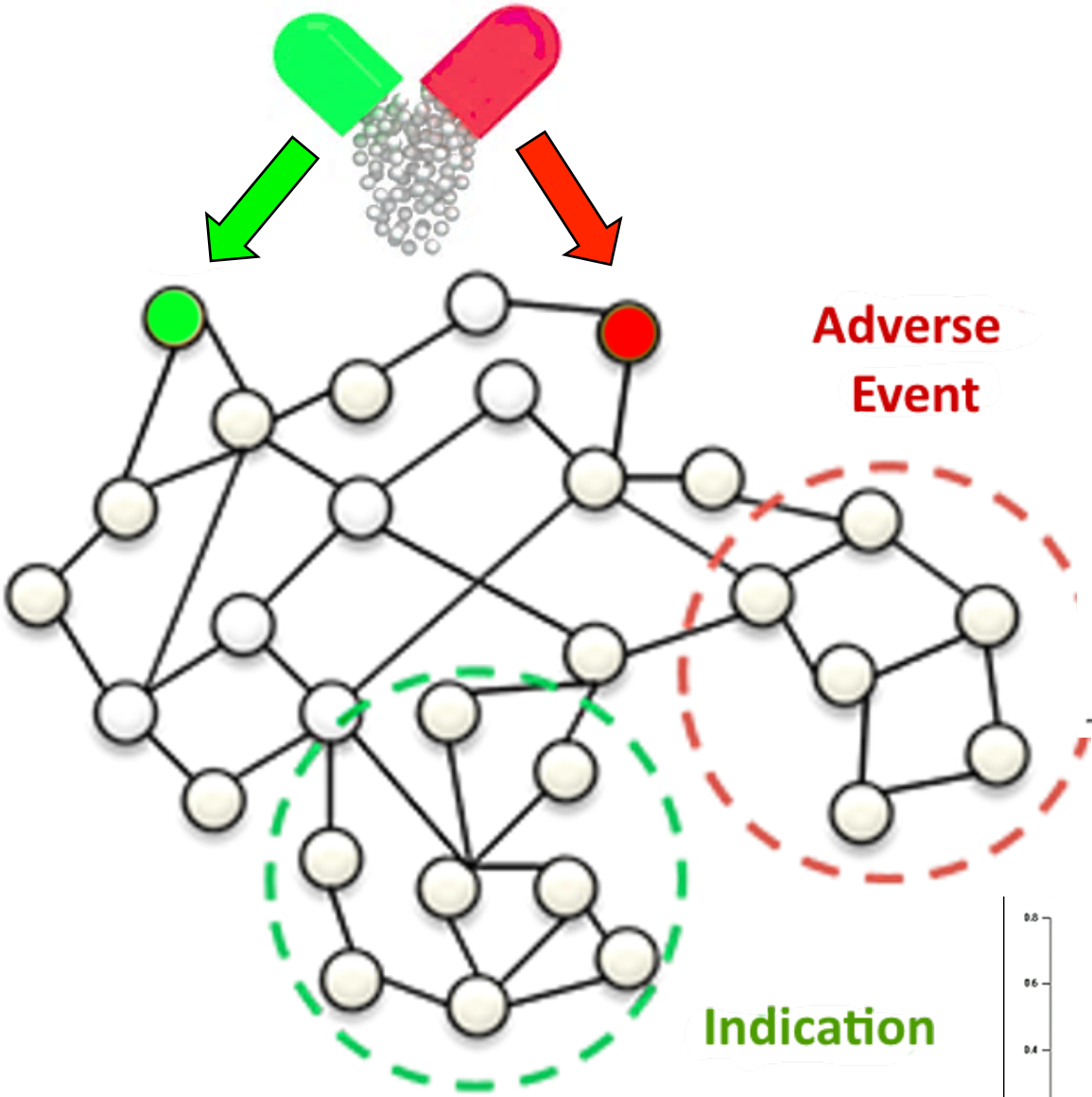
**→1,000,000 enhancers**

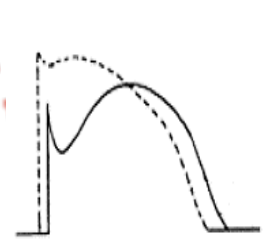
**→100,000 promoters**

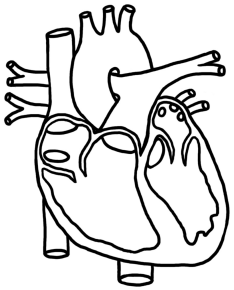


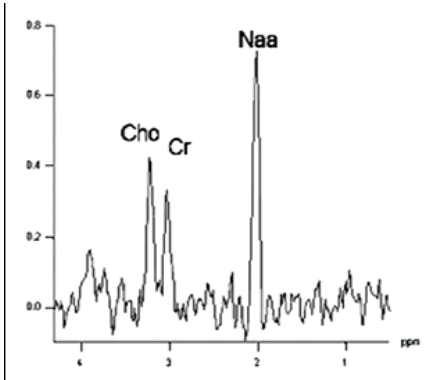

Rao SSP et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* (2014). doi.org/10.1016/j.cell.2014.11.021

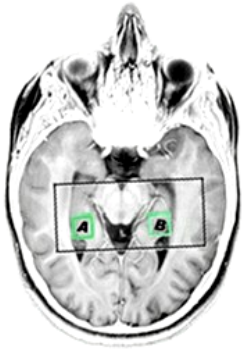
# Pathway Discovery of Adverse Events



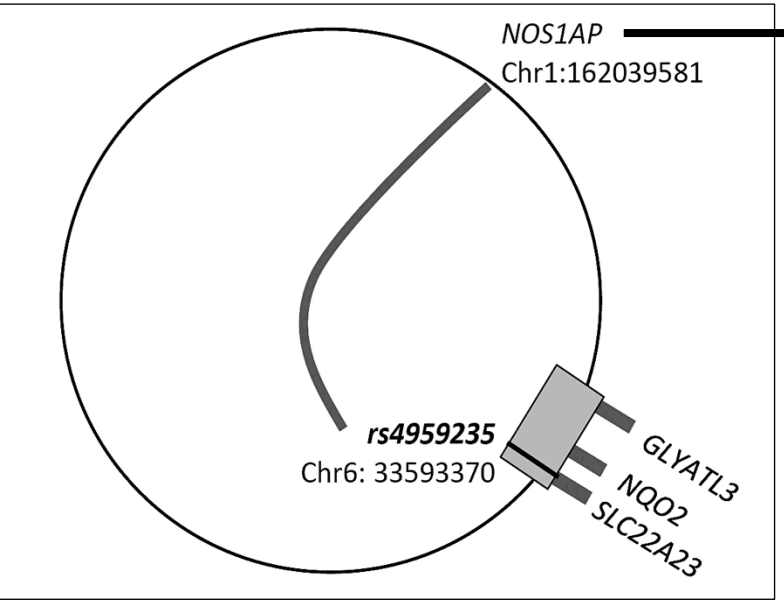







# Spatial mapping using Hi-C data explains basis of QT prolongation during treatment with the antipsychotic drug quetiapine




Aberg K et al. Genome-wide association study of antipsychotic-induced QTc interval prolongation. *Pharmacogenomics.* 12:165–72. (2012).


Aarnoudse AJL et al. Common NOS1AP variants are associated with a prolonged QTc interval in the Rotterdam Study. *Circulation.* 116(1): 10-16 (2007).

Chang KC et al. Nitric oxide synthase 1 adaptor protein, an emerging new genetic marker for QT prolongation and sudden cardiac death. *Acta Cardiol. Sin.* 29:217–225. (2013);

Crotti L et al. NOS1AP is a genetic modifier of the long-QT syndrome. Circulation, 120(17):1657-1663 (2009).

Jamshidi Y et al. Common variation in the NOS1AP gene is associated with drug-induced QT prolongation and ventricular arrhythmia. *J. Amer. Coll. Cardiol.* 60: 841-850. (2012);

Kapoor A et al. QT Interval-International GWAS Consortium: An enhancer polymorphism at the cardiomyocyte intercalated disc protein NOS1AP locus is a major regulator of the QT interval. *Amer. J. Human Genet.* 94: 854-869. (2014);

Jamshidi Y et al. Common variation in the NOS1AP gene is associated with drug-induced QT prolongation and ventricular arrhythmia. *J. Amer. College Cardiol.* 60(9): 841-850 (2012).

Lehtinen AB et al. Association of NOS1AP genetic variants with QT interval duration in families from the Diabetes Heart Study. *Diabetes.* 57(4):1108-1114 (2008).

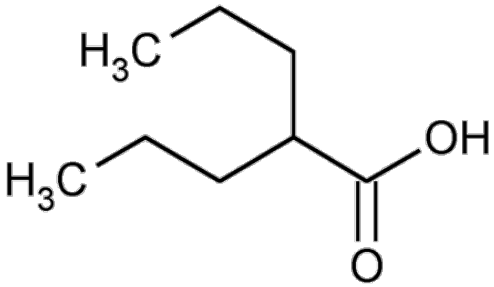
Tomás M, Napolitano C, De Giuli L et al. Priori Polymorphisms in the NOS1AP gene modulate QT interval duration and risk of arrhythmias in the long QT syndrome. *J. Amer. Coll. Cardiol. 55: 2745-2752.* (2010);

Roden DM. Drug-induced prolongation of the QT interval. *New England J. Med.* 350(10), 1013-1022 (2004).

# Valproic Acid Induces the NEUROD1 Transcriptional Program of Neurogenesis Following Brain Injury

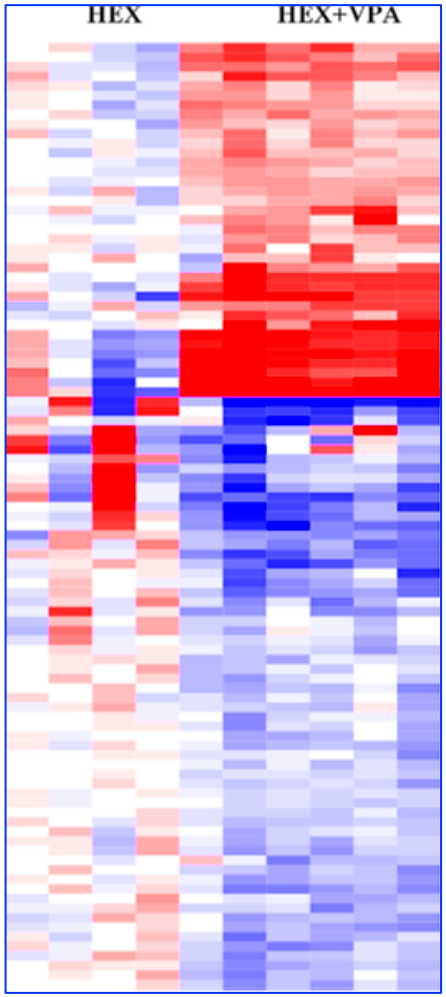
**TRAUMATIC BRAIN INJURY**

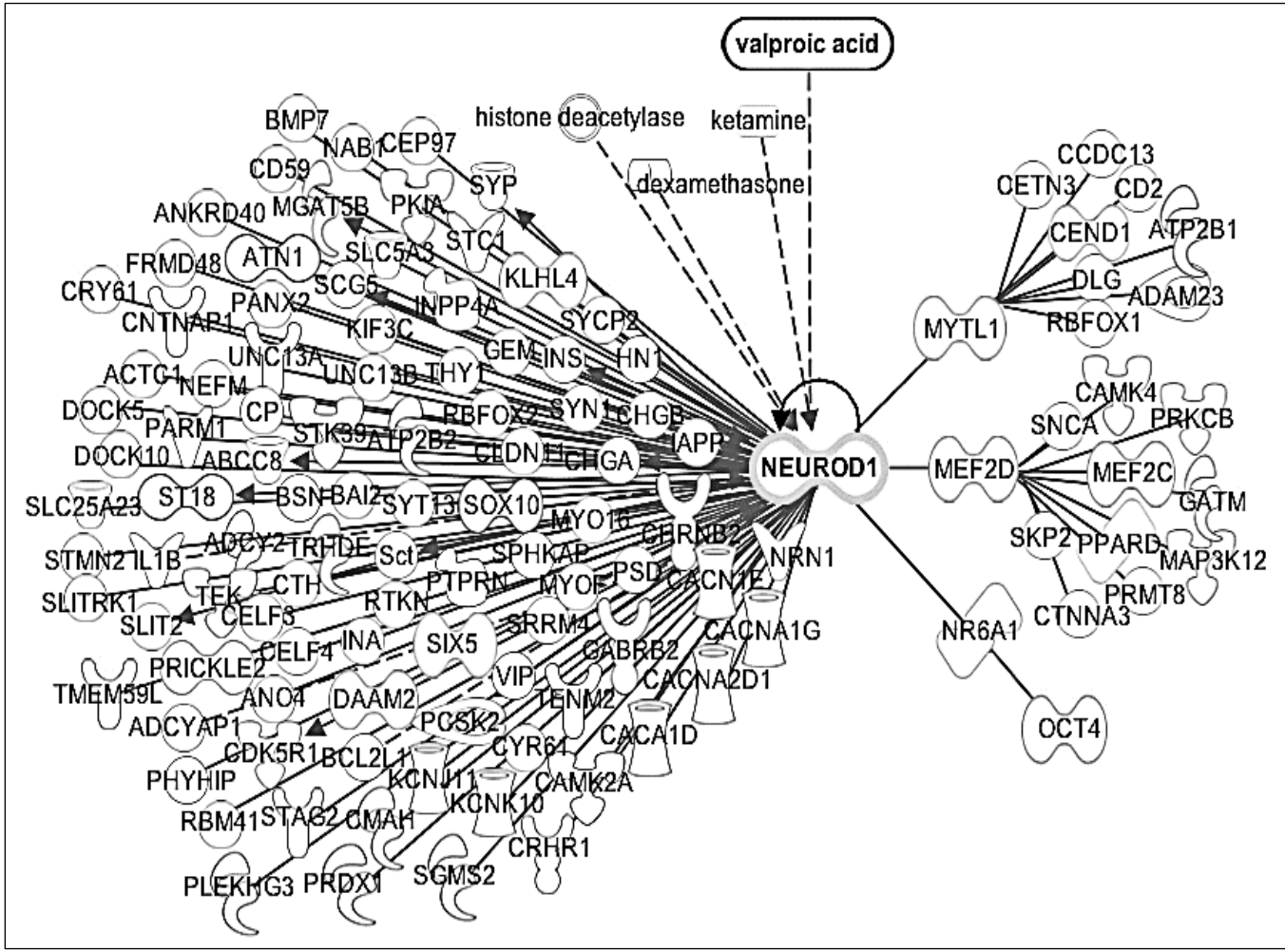
**VALPROIC ACID AND OTHER HDAC INHIBITORS**




**FUNCTIONAL RECOVERY**

# Valproic Acid Induces the NEUROD1 Transcriptional Program of Neurogenesis Following Brain Injury

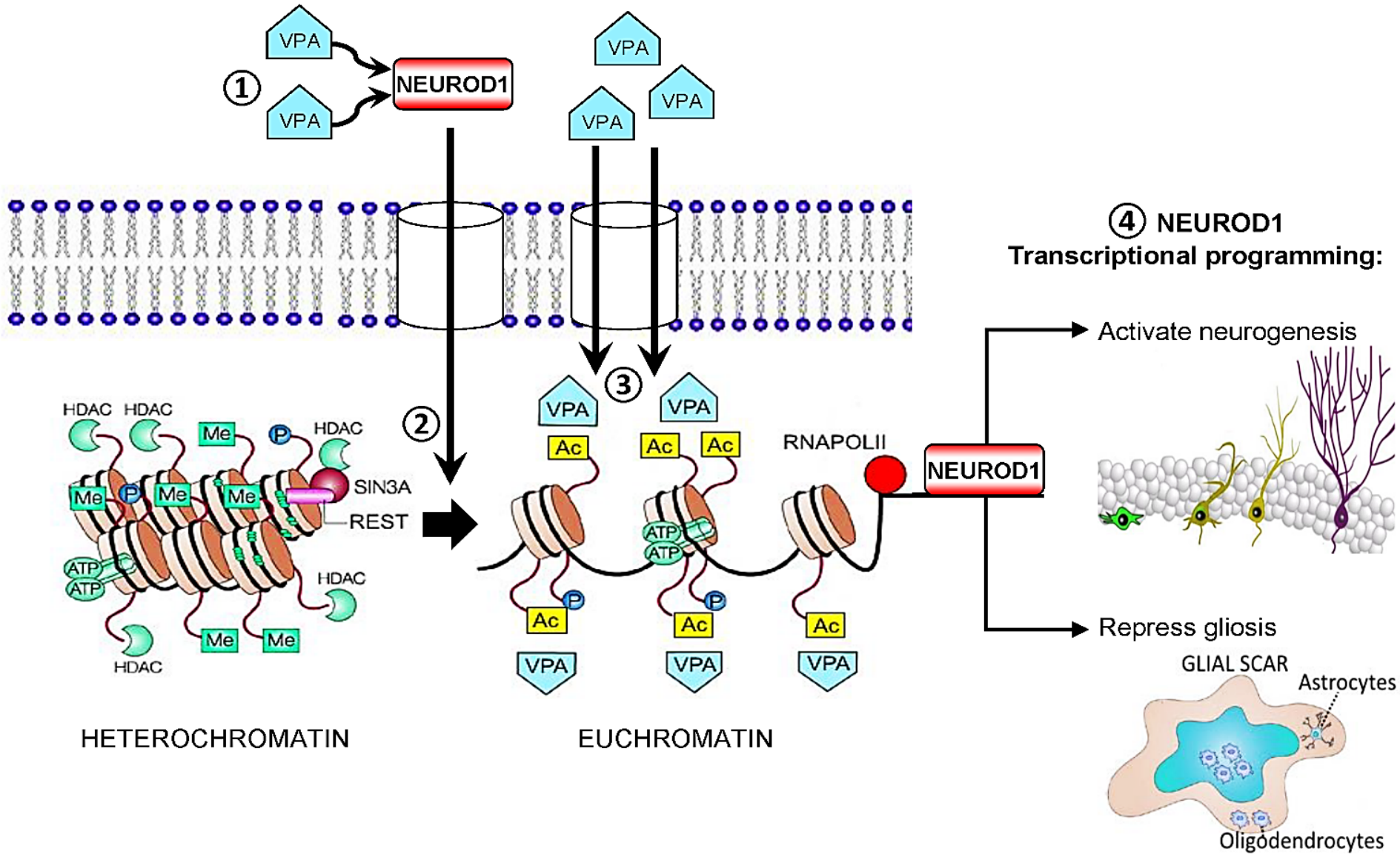



Differential gene expression following VPA therapy in an animal model of TBI

Linked transcriptional programs of neurogenesis, differentiation and neuron survival were identified in the dataset, representing VPA-activation of NEUROD1-mediated pathways.

# Valproic Acid Induces the NEUROD1 Transcriptional Program of Neurogenesis Following Brain Injury




Model of VPA-mediated NEUROD1 programming of transcription in human brain through focused inhibition of HDACs by VPA directed towards a subset of genes.

# TOOLS, RESOURCES

**→Gene Set Enrichment Analysis (GSEA), Gene Ontology**

Amigo 2 Panther Gene Ontology Browser
http://amigo.geneontology.org/amigo

**→Pathway analysis**

Pathway commons: http://www.pathwaycommons.org/about/
Cytoscape: http://www.cytoscape.org/

***Commercial***
Ingenuity Pathway Analysis® (IPA®; Qiagen GmbH)
Pathway Studio® (Elsevier)
MetaCore® (Thomson Reuters)

**→GTEx Portal:** http://www.gtexportal.org/home/

**→Transcription factors:**

FactorBook: http://www.factorbook.org/human/

TransFac Pro® (Biobase International; Qiagen GmbH)

CONTACT INFORMATION (feel free to contact me at any time):

Gerry Higgins, Ph.D., M.D.
gehiggin@med.umich.edu
734-545-2731
Room 2035A Palmer Commons