Integrative *causal* networks for understanding complex human diseases

Jun Zhu, Ph. D.
Professor of Genomics and Genetic Sciences
Icahn Institute of Genomics and Multi-scale Biology
The Tisch Cancer Institute
Icahn Medical School at Mount Sinai
New York, NY

http://research.mssm.edu/integrative-network-biology/
Email: jun.zhu@mssm.edu

@IcahnInstitute

Pittsburg University, October, 2015
Gene sets

Association networks

Probabilistic causal networks

Mechanism based models

Data required to train models

Biological details revealed
Why Bayesian network is chosen for causal modelling?
A framework for building biological causal networks

- High throughput data
- Microarray data
- Proteomic data
- Metabolomic data
- Genomics
- Genetics

probabilistic graphic models

- Medline
- Biocarta/Biopathway
- Biologists

Database

GUI

Hypothesis, test
Association vs Causality

When informed by his doctor of the correlation between fat dogs and their masters, Brian set out immediately to rectify his weight problem.

From Stephen Friend
High-density lipoprotein cholesterol as a predictor of coronary heart disease risk. The PROCAM experience and pathophysiological implications for reverse cholesterol transport

Gerd Assmann, Helmut Schulte, Arnold von Eckardstein, Yadong Huang

Abstract

The incidence of coronary heart disease (CHD) was assessed via the Prospective Cardiovascular Münster (PROCAM) study in 19,698 volunteer subjects aged between 16 and 85 years. An adequate incidence of atherosclerotic CHD was only found in male subjects greater than 40 years of age. The analysis and subsequent 6 year follow-up period was, therefore, confined to 4559 male participants aged 40–64 years. In the follow-up period, 186 study participants developed atherosclerotic CHD (134 definite non-fatal myocardial infarctions (Mls) and 52 definite atherosclerotic CHD deaths including 21 sudden cardiac deaths and 31 fatal Mls). Univariate analysis revealed a significant association between the incidence of atherosclerotic CHD and high-density lipoprotein cholesterol ($P < 0.001$), which remained after adjustment for other risk factors.
### Table 3 | Description of included studies of cholesteryl ester transfer protein (CETP) inhibitors

<table>
<thead>
<tr>
<th>Reference</th>
<th>Trial drugs and dose</th>
<th>Control</th>
<th>Follow-up (months)</th>
<th>No enrolled (No intervention, No control)</th>
<th>Statin use (%)</th>
<th>Men (No intervention, No control)</th>
<th>Mean (SD) age (years) (intervention, control)</th>
<th>White ethnicity (%)</th>
<th>Increase in HDL from baseline in active arm (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dal-OUTCOMES&lt;sup&gt;33&lt;/sup&gt; 2012</td>
<td>Dalcetrapib 600 mg daily</td>
<td>Placebo</td>
<td>31</td>
<td>15 871 (7938, 7933)</td>
<td>97</td>
<td>6365, 6436</td>
<td>60.3 (9.1), 60.1 (9.1)</td>
<td>88</td>
<td>40</td>
</tr>
<tr>
<td>Dal-PLAQUE&lt;sup&gt;34&lt;/sup&gt; 2011</td>
<td>Dalcetrapib 600 mg daily</td>
<td>Placebo</td>
<td>24</td>
<td>130 (64, 66)</td>
<td>87</td>
<td>51, 55</td>
<td>62.6 (8.2), 64.6 (7.8)</td>
<td>92</td>
<td>31</td>
</tr>
<tr>
<td>Dal-VESSEL&lt;sup&gt;35&lt;/sup&gt; 2012</td>
<td>Dalcetrapib 600 mg daily</td>
<td>Placebo</td>
<td>8</td>
<td>476 (239, 237)</td>
<td>95</td>
<td>211, 211</td>
<td>62.3 (7.0), 61.9 (7.92)</td>
<td>NR</td>
<td>31</td>
</tr>
<tr>
<td>Define&lt;sup&gt;36&lt;/sup&gt; 2010</td>
<td>Anacetrapib 100 mg daily</td>
<td>Placebo</td>
<td>18</td>
<td>1623 (811, 812)</td>
<td>99</td>
<td>629, 618</td>
<td>62.5 (8.7), 62.9 (9.0)</td>
<td>83</td>
<td>138</td>
</tr>
<tr>
<td>Illuminate&lt;sup&gt;37&lt;/sup&gt; 2007</td>
<td>Torcetrapib 60 mg daily</td>
<td>Placebo</td>
<td>18</td>
<td>15 054 (7528, 7526)</td>
<td>100</td>
<td>5854, 5861</td>
<td>61.3 (7.6), 61.3 (7.6)</td>
<td>93</td>
<td>72</td>
</tr>
<tr>
<td>Illustrate&lt;sup&gt;38&lt;/sup&gt; 2007</td>
<td>Torcetrapib 60 mg daily</td>
<td>Placebo</td>
<td>24</td>
<td>1188 (591, 597)</td>
<td>100</td>
<td>416, 421</td>
<td>56.9 (9.1), 57 (9.2)</td>
<td>NR</td>
<td>61</td>
</tr>
<tr>
<td>Radiance 1&lt;sup&gt;39&lt;/sup&gt; 2007</td>
<td>Torcetrapib 60 mg daily</td>
<td>Placebo</td>
<td>24</td>
<td>850 (423, 427)</td>
<td>100</td>
<td>214, 232</td>
<td>46.8 (12.0), 45.2 (12.9)</td>
<td>NR</td>
<td>52</td>
</tr>
<tr>
<td>Radiance 2&lt;sup&gt;40&lt;/sup&gt; 2007</td>
<td>Torcetrapib 60 mg daily</td>
<td>Placebo</td>
<td>20</td>
<td>752 (377, 375)</td>
<td>100</td>
<td>237, 245</td>
<td>57.9 (8.1), 56.5 (8.2)</td>
<td>NR</td>
<td>63</td>
</tr>
</tbody>
</table>

HDL=high density lipoprotein; NR=not reported.
Fig 4 The statin revolution: without background statin treatment, fibrates and niacin were found to reduce non-fatal myocardial infarction.

<table>
<thead>
<tr>
<th>Non-fatal myocardial infarction</th>
<th>No of events/total</th>
<th>Odds ratio M-H, random (95% CI)</th>
<th>Odds ratio M-H, random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td><strong>Niacin</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No background statin</td>
<td>136/1659</td>
<td>394/3332</td>
<td>0.67 (0.54 to 0.82)</td>
</tr>
<tr>
<td>Background statin</td>
<td>509/15371</td>
<td>527/14939</td>
<td>0.94 (0.83 to 1.06)</td>
</tr>
<tr>
<td>Test for heterogeneity: $I^2$=87%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fibrate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No background statin</td>
<td>773/14236</td>
<td>1181/15896</td>
<td>0.72 (0.65 to 0.79)</td>
</tr>
<tr>
<td>Background statin</td>
<td>173/2765</td>
<td>186/2753</td>
<td>0.92 (0.74 to 1.14)</td>
</tr>
<tr>
<td>Test for heterogeneity: $I^2$=78%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CETP inhibitor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background statin</td>
<td>582/18003</td>
<td>553/18008</td>
<td>1.05 (0.93 to 1.18)</td>
</tr>
</tbody>
</table>

Daniel Keene et al. BMJ 2014;349:bmj.g4379
The cost of developing a prescription drug that gains market approval
A simple biological question: are there causal/reactive relationships?
A Bayesian network approach:

Best model
A Bayesian network approach:

Best models

Markov Equivalent models
A Bayesian network ≠ a causal structure

Markov Equivalent models

\[ B \perp C \mid A \]
Bayesian network: how to break Markov equivalent?

Animal model: mouse F2 intercrosses
Causal inference: genetics

Perturbations with a causal anchor
--Natural variation in a segregating population provides the same type of causal anchor

DNA Supporting Gene X

Variation in DNA leads to variation in mRNA

Variation in mRNA leads to variation in protein, which in turn can lead to disease

Central Dogma of Biology

Schadt et al. Nature Genetics (2005)
A Bayesian network approach:

Best models

Markov Equivalent models
A Bayesian network approach:

Genotype

Cis regulation

Genotype

Cis regulation

Genotype

Cis regulation

Not Markov Equivalent models
Cis regulation

Genotype

Cis regulation

Not Markov Equivalent models
Structure priors based on causality

- Estimate confidence of causality
  - Bootstrap samples for 200 times
  - Factions of causal, reactive, independent calls

- The pair is independent
  \[
  p(X_a \rightarrow X_b) = 1 - \frac{\sum_{i \in \mathcal{P}E} p(X_a \perp X_b | l_i)}{\sum_{i \in \mathcal{P}E} 1}
  \]

- The pair is causa/reactive
  \[
  p(X_a \rightarrow X_b) = \frac{2 \sum_i p(X_a \rightarrow X_b | l_i)}{\sum_i p(X_a \rightarrow X_b | l_i) + p(X_b \rightarrow X_a | l_i)}
  \]

Zhu et al., PLoS CompBio, 2007
Bayesian Network: a simulation study

Simulation of data with network and genetics constraints

- Genetic map
- Rqtl
- QTL models
- Rcross
- genotypes
- Trait values
- Zmap
- QTLs

QTLCartographer Suite

Bayesian network (a network structure, Conditional probability density functions)

Values of head nodes

Values of all nodes

Network reconstruction program

Zhu et al., PLoS CompBio, 2007
Bayesian network: Genetics information is critical when sample size is small

Largest improvement in recall occurs with smaller sample sizes

Zhu et al., PLoS CompBio, 2007
A framework for data integration

Microarray data
Proteomic data
Metabolomic data
Genomics
Genetics

High throughput data

knowledge

Medline
Biocarta/Biopathway
Biologists

Database

GUI

probabilistic graphic models

Hypothesis, test
Bayesian network: PPI

Bayesian network: PPI

4-clique

3-clique

Clique community (partial clique)

Bayesian network: Transcription Factors

Introducing *scale-free priors* for TF or protein complex

\[ p(T \rightarrow g) \propto w(T) \]

\[ w(T) = \log\left( \sum_{g_i \in R} \left| r(T, g_i) \right| > r_{cutoff} \right) \]

Yeast segregants

Synthetic complete medium Logarithm growth

Gene expression

Genotypes

Public databases

Protein-protein interactions

Transcription factor binding sites

Protein Metabolite interactions

Bayesian network

Integration improves network qualities

<table>
<thead>
<tr>
<th>BN</th>
<th>KO data</th>
<th>GO terms</th>
<th>TF data</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o any priors</td>
<td>125</td>
<td>55</td>
<td>26</td>
</tr>
<tr>
<td>w/ genetics priors</td>
<td>139</td>
<td>59</td>
<td>34</td>
</tr>
<tr>
<td>w/ genetics, TF and PPI priors</td>
<td>152</td>
<td>66</td>
<td>52</td>
</tr>
</tbody>
</table>

Prospective validation is the gold standard

ILV6 gives rise to large expression signature
- ILV6 KO sig enriched (p~10E-52)
- GCN4 upregulated in ILV6 KO $\rightarrow$ large signature

LEU2 KO gives rise to small expression signature
- LEU2 KO sig enriched (p~10E-18)
- GCN4 downregulated in LEU2 KO $\rightarrow$ small signature

How does LEU2 affect LEU3 activity?

LEU3 binding sites

LEU2

mRNA expression

Surrogate marker for Leu3p activity
A framework for building causal networks

probabilistic graphic models

knowledge

High throughput data

Microarray data

Proteomic data

Metabolomic data

Genomics

Genetics

Medline

Biocarta/Biopathway

Biologists

Database

Hypothesis, test

GUI
Yeast segregants

Synthetic complete medium
Logarithm growth

Gene expression

metabolites

Public databases

Protein-protein interactions

Transcription factor binding sites

Protein Metabolite interactions

Bayesian network

Metabolite abundance is under genetic control

KEGG biochemical pathways

\[ p(m \rightarrow e) \propto e^{-\lambda d_{m,e}} \]

LEU2 mRNA is causal to 2-isopropylmalate

LEU3 binding site

With metabolomic data

LEU3 regulation

- The activity of Leu3p is positively regulated by alpha-isopropylmalate (IPM), the product of the first step in leucine biosynthesis

- The degree of activation by Leu3p is Leu3p concentration dependent, and it has been shown that LEU3 gene expression is regulated by general amino acid control, which is mediated by the GCN4 transcription factor
2-isopropylmalate: mechanism of causal regulator LEU2

LEU2 genotype → LEU2 activity → 2-isopropylmalate

Transcriptional response for genes with LEU3 binding sites

LEU3 activity
**Networks facilitate direct identification of genes that are causal for disease**

*Yang et al, Nature Genetics (2009)*

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Variance of OFPM explained by gene expression*</th>
<th>Mouse model</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zfp90</td>
<td>Zinc finger protein 90</td>
<td>68%</td>
<td>tg</td>
<td>Constructed using BAC transgenics</td>
</tr>
<tr>
<td>Gas7</td>
<td>Growth arrest specific 7</td>
<td>68%</td>
<td>tg</td>
<td>Constructed using BAC transgenics</td>
</tr>
<tr>
<td>Gpx3</td>
<td>Glutathione peroxidase 3</td>
<td>61%</td>
<td>tg</td>
<td>Provided by Prof. Oleg Mirochnitchenko (University of Medicine and Dentistry at New Jersey, NJ) [12]</td>
</tr>
<tr>
<td>Lactb</td>
<td>Lactamase beta</td>
<td>52%</td>
<td>tg</td>
<td>Constructed using BAC transgenics</td>
</tr>
<tr>
<td>Me1</td>
<td>Malic enzyme 1</td>
<td>52%</td>
<td>ko</td>
<td>Naturally occurring KO</td>
</tr>
<tr>
<td>Gyk</td>
<td>Glycerol kinase</td>
<td>46%</td>
<td>ko</td>
<td>Provided by Dr. Katrina Dipple (UCLA) [13]</td>
</tr>
<tr>
<td>Lpl</td>
<td>Lipoprotein lipase</td>
<td>46%</td>
<td>ko</td>
<td>Provided by Dr. Ira Goldberg (Columbia University, NY) [11]</td>
</tr>
<tr>
<td>C3ar1</td>
<td>Complement component 3a receptor 1</td>
<td>46%</td>
<td>ko</td>
<td>Purchased from Deltagen, CA</td>
</tr>
<tr>
<td>Tgfbr2</td>
<td>Transforming growth factor beta receptor 2</td>
<td>39%</td>
<td>ko</td>
<td>Purchased from Deltagen, CA</td>
</tr>
</tbody>
</table>

*Schadt et al. Nature Genetics (2005)*
Multiple genes in a network causing diseases!

Guilt by association

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium*

Robust associations of four from genome-wide analyses

John A Todd¹, Neil M Walker¹,⁹, Jason D Cooper¹,⁹, Del

<table>
<thead>
<tr>
<th>Disease/Trait</th>
<th>GWAS_snp</th>
<th>Reported Gene(s)</th>
<th>eSNP_gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triglycerides</td>
<td>rs10889353</td>
<td>ANGPTL3</td>
<td>ANGPTL3</td>
</tr>
<tr>
<td>Testicular germ cell tumor</td>
<td>rs710138</td>
<td>BAK1</td>
<td>BAK1</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>rs4444235</td>
<td>BMP4</td>
<td>BMP4</td>
</tr>
<tr>
<td>Serum IgE levels</td>
<td>rs2251746</td>
<td>FCER1A</td>
<td>FCER1A</td>
</tr>
<tr>
<td>Autoimmune diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>rs9271366</td>
<td>HLA-DRB1</td>
<td>HLA-DRB1</td>
</tr>
<tr>
<td>Plasma triglycerides</td>
<td>rs6457617</td>
<td>MHC</td>
<td>HLA-DQA1</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>rs9272346</td>
<td>MHC</td>
<td>HLA-DQB1</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>rs3197999</td>
<td>MST1</td>
<td>MST1</td>
</tr>
<tr>
<td>Body mass index</td>
<td>rs10838738</td>
<td>MTCH2</td>
<td>MTCH2</td>
</tr>
<tr>
<td>HDL cholesterol, Triglycerides</td>
<td>rs7679</td>
<td>PLTP</td>
<td>PLTP</td>
</tr>
<tr>
<td>Glioma</td>
<td>rs6010620</td>
<td>RTE1</td>
<td>RTE1</td>
</tr>
<tr>
<td>Attention deficit hyperactivity disorder</td>
<td>rs260461</td>
<td>ZNF544</td>
<td>ZNF544</td>
</tr>
<tr>
<td>Coronary disease, LDL</td>
<td>rs599839</td>
<td>CELSR2, PSRC1, MYBPHL</td>
<td>SOR1, PSRC1, CELSR2</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>rs8034191</td>
<td>CHRNA3, CHRNA5, PSMA4, LOC123688</td>
<td>PSMA4</td>
</tr>
<tr>
<td>Weight</td>
<td>rs2844479</td>
<td>AIF1, NCR3</td>
<td>BAT3</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>rs7804356</td>
<td>Intergenic</td>
<td>SKAP2</td>
</tr>
<tr>
<td>Testicular germ cell tumor</td>
<td>rs4699052</td>
<td>Intergenic</td>
<td>LOC56898</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>rs6596075</td>
<td>Intergenic</td>
<td>SLCL2A5</td>
</tr>
<tr>
<td>Cognitive test performance</td>
<td>rs2832077</td>
<td>Intergenic</td>
<td>CCT8</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>rs10798269</td>
<td>Intergenic</td>
<td>C1orf9</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>rs4779584</td>
<td>Intergenic</td>
<td>GREM1</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>rs1701704</td>
<td>RAB5B, SUOX, IKZF4, ERBB3, CDK2, RPS26</td>
<td></td>
</tr>
<tr>
<td>QT interval</td>
<td>rs4725982</td>
<td>KCNH2</td>
<td>IAN4L1</td>
</tr>
<tr>
<td>Intracranial aneurysm</td>
<td>rs700651</td>
<td>BOLL, PLCL1</td>
<td>PLCL1</td>
</tr>
<tr>
<td>Body mass index</td>
<td>rs7498665</td>
<td>SH2B1</td>
<td>EIF3C</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>rs881375</td>
<td>TRAF1, C5</td>
<td>TRAF1</td>
</tr>
<tr>
<td>Attention deficit hyperactivity disorder</td>
<td>rs2842643</td>
<td>TFEB</td>
<td>UNC5CL</td>
</tr>
</tbody>
</table>
Our Data Supports SORT1 as the Strongest Candidate Gene

<table>
<thead>
<tr>
<th>eSNP</th>
<th>Genes associated</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs599839</td>
<td>SORT1</td>
<td>1.52E-56</td>
</tr>
<tr>
<td>rs599839</td>
<td>PSRC1</td>
<td>2.17E-53</td>
</tr>
<tr>
<td>rs599839</td>
<td>CELSR2</td>
<td>4.31E-23</td>
</tr>
</tbody>
</table>

SORT1 subnetwork is enriched for chemokine, angiogenesis, insulin signaling genes

Schadt et al., PLoS Biol., 2008; 6:e107
From noncoding variant to phenotype via \textit{SORT1} at the 1p13 cholesterol locus

Kiran Musunuru$^{1,2,3}$*, Alanna Strong$^{4}$*, Maria Frank-Kamenetsky$^{5}$, Noemi E. Lee$^{1}$, Tim Ahfeldt$^{1,6}$, Katherine V. Sachs$^{4}$, Xiaoyu Li$^{4}$, Hui Li$^{3}$, Nicolas Kuperwasser$^{1}$, Vera M. Ruda$^{1}$, James P. Pirruccello$^{1,2}$, Brian Muchmore$^{7}$, Ludmila Prokunina-Olsson$^{7}$, Jennifer L. Hall$^{2,8}$, Eric E. Schadt$^{9}$, Carlos R. Morales$^{10}$, Sissel Lund-Katz$^{11}$, Michael C. Phillips$^{11}$, Jamie Wong$^{5}$, William Cantley$^{5}$, Timothy Racie$^{5}$, Kenechi G. Ejebe$^{1,2}$, Marju Orho-Melander$^{12}$, Olle Melander$^{12}$, Victor Koteliantsky$^{5}$, Kevin Fitzgerald$^{5}$, Ronald M. Krauss$^{13}$, Chad A. Cowan$^{1,2}$, Sekar Kathiresan$^{1,2}$* & Daniel J. Rader$^{4}$*

**Overexpression**  

\begin{align*}
\text{mg} \cdot \text{dl}^{-1} & \\
\text{Baseline} & \quad 2 \text{ weeks} \\
\text{Total cholesterol (Mira)} & \quad 6 \text{ weeks}
\end{align*}

\begin{align*}
\text{Normalized level} & \\
\text{Baseline} & \quad 2 \text{ weeks}
\end{align*}

\begin{align*}
\text{P} & = 4 \times 10^{-5} \\
\text{P} & = 3 \times 10^{-4}
\end{align*}

**Knockdown**

\begin{align*}
\text{mg} \cdot \text{dl}^{-1} & \\
\text{Baseline} & \quad 2 \text{ weeks} \\
\text{Total cholesterol (Mira)} & \quad \text{LDL-C} (\text{pooled FPLC})
\end{align*}

\begin{align*}
\text{Control} & \quad \text{Sort1 siRNA} \\
\text{P} & = 0.03
\end{align*}
Why it is so hard to model biological systems?

- The more we learn, the more complicated it becomes!

**Epigenetic regulation**: heritable changes in gene function that cannot be explained by changes in DNA sequence
- DNA methylation
- Chromatin structure

**Junk DNA?**

**Post transcriptional regulation**
- Splicing (1981)
- RNA editing (1986)
- miRNA mediated regulation (1993)

**Post translational regulation**
- Phosphorylation
- Glycosylation
- Acetylation

It is not one gene to one protein anymore!
Integrating omics data
Complex diseases: observations to models
Integrating omics data into Bayesian network models

- Genetics
  - Zhu et al, Cytogenet Genome Res, 2004
  - Schadt et al, Nature Genetics, 2005

- Proteomics and Genomics
  - Zhu et al, Nature Genetics, 2008

- CNV

- Metabolomics

- Methylation data
  - Yoo et al, PLoS Genetics, 2015
Complex traits

Genetic background

time
Integration of time dimension data into Bayesian network

Genetics

Time

Causal inference
Constructing causal networks

Dynamic Bayesian network

Granger causality

Zhu et al., PLoS CompBio, 2010
Human blood expression profile time series

(1) Remove individual scale variation reference time point 0

(2) Combine short time series into long ones
Integration of genetics, genomics and temporal data using Dynamic Bayesian network

Zhu*, Chen* et al., PLoS CompBio, 2010
Integration of time dimension data into Bayesian network
Time series of drug response in a yeast F2 cross

95 segregants

Add 28µl of 0.25 mM rapamycin → Incubate in shaking water bath at 30°C.

Wash each cell pellet with 1 ml cold sterile DW, flash freeze it and store it at -80°C.

Sample each culture at 10 min intervals after rapamycin addition for 70 min as described for T₀.

Yeung et al., PNAS, 2011
Multi-polynomial Temporal Genetic Association

- Assume that for each genotype, time-series gene expression levels follow a multivariate normal density (Wu et al.)
- The mean vector is modeled by the cubic polynomial curve, $m$ is the number of time points
  \[
  \tilde{g}_j = [g_j(t)]_{1 \times m} = [\beta_{0j} + \beta_{1j}t + \beta_{2j}t^2 + \beta_{3j}t^3]_{1 \times m}
  \]
- The covariance matrix is assumed to be the same for both genotypes and modeled using AR(1) repeated measurement errors (Daviddian et al; Verbeke et al)

\[
\Sigma = \sigma_e^2 \begin{bmatrix}
1 & \rho & \ldots & \rho^{m-1} \\
\rho & 1 & \ldots & \rho^{m-2} \\
\vdots & \vdots & \ddots & \vdots \\
\rho^{m-1} & \rho^{m-2} & \ldots & 1
\end{bmatrix}
\]
MPTGA

- Density for time-series data, \( m=6 \)
  \[
f_j(y) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \exp[-(\vec{y} - \vec{g}_j)^T \Sigma^{-1} (\vec{y} - \vec{g}_j) / 2]
\]
- Joint likelihood for N-95 segregants, \( \Theta = (\beta_{0j}, \beta_{1j}, \beta_{2j}, \beta_{3j}, \rho, \sigma_e^2) \)
  \[
  L(\Theta) = \prod_{i=1}^{N} \left[ \delta_{i0} f_0(\vec{y}_i) + \delta_{i1} f_1(\vec{y}_i) \right]
\]
- Maximum likelihood estimate (MLE)
- Likelihood ratio test

\[
H_0 : \beta_{00} = \beta_{01}, \beta_{10} = \beta_{11}, \beta_{20} = \beta_{21}, \beta_{30} = \beta_{31}
\]
\[
H_1 : \text{at least one of the equalities does not hold}
\]

Unpublished results
Simulation Study

- Simulate time series data from multivariate normal distribution, with mean vector modeled by various patterns that are similar to the observed experimental results.
- Data drawn either from a single model or two separate models
- 10,000 groups of time series data were simulated, in which each group is composed of 100 6-point time series data
- Performance comparison
  - ROC curve
  - simulation under different auto-regressive coefficient $\rho$

Unpublished results
ROC curves

- ROC curve comparing performance among three different time-dependent genetic association methods

Unpublished results
Mathematical Models
eQTL effect

M1: \( X_{i,t} = \delta_{i0}(\alpha_{00} + \alpha_{10}X_{i,t-1}) + \delta_{i1}(\alpha_{01} + \alpha_{11}X_{i,t-1}) + \epsilon_i \)

\[ Y_{i,t} = \beta_0 + \beta_1Y_{i,t-1} + \beta_2X_{i,t-1} + \mu_i \]

Auto-regressive term Causal effect

M2: \( Y_{i,t} = \delta_{i0}(\beta_{00} + \beta_{10}Y_{i,t-1}) + \delta_{i1}(\beta_{01} + \beta_{11}Y_{i,t-1}) + \mu_i \)

\[ X_{i,t} = \alpha_0 + \alpha_1X_{i,t-1} + \alpha_2Y_{i,t-1} + \epsilon_i \]

M3: \( X_{i,t} = \delta_{i0}(\alpha_{00} + \alpha_{10}X_{i,t-1}) + \delta_{i1}(\alpha_{01} + \alpha_{11}X_{i,t-1}) + \epsilon_i \)

\[ Y_{i,t} = \delta_{i0}(\beta_{00} + \beta_{10}Y_{i,t-1}) + \delta_{i1}(\beta_{01} + \beta_{11}Y_{i,t-1}) + \beta_2X_{i,t-1} + \mu_i \]

M4: \( Y_{i,t} = \delta_{i0}(\beta_{00} + \beta_{10}Y_{i,t-1}) + \delta_{i1}(\beta_{01} + \beta_{11}Y_{i,t-1}) + \mu_i \)

\[ X_{i,t} = \delta_{i0}(\alpha_{00} + \alpha_{10}X_{i,t-1}) + \delta_{i1}(\alpha_{01} + \alpha_{11}X_{i,t-1}) + \alpha_2Y_{i,t-1} + \epsilon_i \]

M5: \( X_{i,t} = \delta_{i0}(\alpha_{00} + \alpha_{10}X_{i,t-1}) + \delta_{i1}(\alpha_{01} + \alpha_{11}X_{i,t-1}) + \epsilon_i \)

\[ Y_{i,t} = \delta_{i0}(\beta_{00} + \beta_{10}Y_{i,t-1}) + \delta_{i1}(\beta_{01} + \beta_{11}Y_{i,t-1}) + \mu_i \]

Model Selection: BIC or AIC

Unpublished results
# eQTL and CI

<table>
<thead>
<tr>
<th></th>
<th>Static (T0)</th>
<th>MPTGA</th>
<th>Union</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td># eQTL identified</td>
<td>2698</td>
<td>3078</td>
<td>5336</td>
<td>2534</td>
</tr>
<tr>
<td>Avg. 1-LOD drop CI (kb)</td>
<td>41.44</td>
<td>16.91</td>
<td>46.99</td>
<td>7.37</td>
</tr>
</tbody>
</table>

*Unpublished results*
Histogram plot

- Number of linkages plotted against genome location by static approach and three time-dependent approaches.
### Hotspots identified and enrichment analysis of Rapamycin respond signature

<table>
<thead>
<tr>
<th>HotSpot</th>
<th>Chr</th>
<th>Pos</th>
<th>Size</th>
<th>Overlap with Rapamycin Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T0(&gt;=51)</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10,000</td>
<td>63</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>230,000</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>390,000</td>
<td>53</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>210,000</td>
<td>82</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>550,000</td>
<td>309</td>
<td>737</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>90,000</td>
<td>95</td>
<td>63</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>430,000</td>
<td>59</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>370,000</td>
<td>60</td>
<td>84</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>110,000</td>
<td>101</td>
<td>151</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>670,000</td>
<td>425</td>
<td>215</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>50,000</td>
<td>260</td>
<td>331</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>450,000</td>
<td>1030</td>
<td>906</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>490,000</td>
<td>751</td>
<td>722</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>110,000</td>
<td>557</td>
<td>709</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>170,000</td>
<td>450</td>
<td>1218</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>510,000</td>
<td>254</td>
<td>425</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>50000</td>
<td>45</td>
<td>56</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>90,000</td>
<td>50</td>
<td>92</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>190,000</td>
<td>15</td>
<td>207</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>150,000</td>
<td>37</td>
<td>72</td>
</tr>
<tr>
<td>21</td>
<td>9</td>
<td>70,000</td>
<td>34</td>
<td>84</td>
</tr>
<tr>
<td>22</td>
<td>9</td>
<td>250,000</td>
<td>50</td>
<td>43</td>
</tr>
<tr>
<td>23</td>
<td>12</td>
<td>30,000</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>24</td>
<td>13</td>
<td>790,000</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>970,000</td>
<td>23</td>
<td>87</td>
</tr>
<tr>
<td>26</td>
<td>8</td>
<td>130,000</td>
<td>33</td>
<td>73</td>
</tr>
</tbody>
</table>
GO term enrichment analysis for eQTL hot spots identified by MPTGA

<table>
<thead>
<tr>
<th>HotSpot</th>
<th>Chr</th>
<th>Pos</th>
<th>GO-Term</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10,000</td>
<td>cytosolic large ribosomal subunit (sensu Eukaryota)</td>
<td>2.31E-06</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>230,000</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>390,000</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>210,000</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>550,000</td>
<td>structural constituent of ribosome</td>
<td>3.25E-43</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>90,000</td>
<td>branched chain family amino acid biosynthesis</td>
<td>2.89E-09</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>430,000</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>370,000</td>
<td>protein folding</td>
<td>1.12E-14</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>110,000</td>
<td>mating projection tip</td>
<td>8.40E-08</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>670,000</td>
<td>ergosterol biosynthesis</td>
<td>6.29E-25</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>50,000</td>
<td>phosphate transport</td>
<td>2.15E-07</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>450,000</td>
<td>endopeptidase activity</td>
<td>5.55E-16</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>490,000</td>
<td>endopeptidase activity</td>
<td>5.66E-20</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>110,000</td>
<td>structural constituent of ribosome</td>
<td>2.48E-41</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>170,000</td>
<td>structural constituent of ribosome</td>
<td>1.45E-35</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>510,000</td>
<td>structural constituent of ribosome</td>
<td>1.55E-54</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>50000</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>90,000</td>
<td>structural constituent of ribosome</td>
<td>3.21E-28</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>190,000</td>
<td>structural constituent of ribosome</td>
<td>1.67E-102</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>150,000</td>
<td>structural constituent of ribosome</td>
<td>5.58E-36</td>
</tr>
<tr>
<td>21</td>
<td>9</td>
<td>70,000</td>
<td>structural constituent of ribosome</td>
<td>5.50E-37</td>
</tr>
<tr>
<td>22</td>
<td>9</td>
<td>250,000</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>12</td>
<td>30,000</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>13</td>
<td>790,000</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>970,000</td>
<td>structural constituent of ribosome</td>
<td>1.06E-51</td>
</tr>
<tr>
<td>26</td>
<td>8</td>
<td>130,000</td>
<td>mating projection tip</td>
<td>1.10E-07</td>
</tr>
</tbody>
</table>

- A key mechanism of cell controlling growth is to regulate ribosome biogenesis.
- Ribosomal protein gene expression is regulated by mTOR, target of rapamycin.
- All eQTL hot spots identified by MPTGA method and enriched for the rapamycin respond signature are enriched for GO term structural constituent of ribosome.
Causal regulator identification

- Causal regulators identification by time-dependent genetic causality test, and comparison with previous findings in Yvert et al and Zhu et al. Genes in red bold font are overlapping with previous findings.

<table>
<thead>
<tr>
<th>HotSpot</th>
<th>Chr</th>
<th>Pos</th>
<th>Yvert et al.</th>
<th>BN Full</th>
<th>TGCT(BIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10,000</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>230,000</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>390,000</td>
<td>N/A</td>
<td>N/A</td>
<td>YR016W</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>210,000</td>
<td>METALPHA1</td>
<td>METALPHA1</td>
<td>METALPHA1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>550,000</td>
<td>AMN1, MAK5</td>
<td>TBS1, TOS1, ARA1, CSH1, SUP45, CNS1, AMN1</td>
<td>SEC66, ICS2, RPBS, UBS1</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>90,000</td>
<td>LEU2</td>
<td>LEU2, ILV6, NFS1, CIT2, METALPHA1</td>
<td>LEU2</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>430,000</td>
<td>N/A</td>
<td>N/A</td>
<td>AVT6</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>370,000</td>
<td>N/A</td>
<td>N/A</td>
<td>MRH4, YGL081W, PUS2</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>110,000</td>
<td>N/A</td>
<td>N/A</td>
<td>ERG11, SHU1</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>670,000</td>
<td>HAP1</td>
<td>HAP1</td>
<td>HAP1</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>50,000</td>
<td>None</td>
<td>None</td>
<td>MDM1</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>450,000</td>
<td>N/A</td>
<td>N/A</td>
<td>PHO23, TOP2, MKT1, YPT53</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>490,000</td>
<td>None</td>
<td>SAL1, TOP2</td>
<td>TOP2, YDJ1, MKT1</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>110,000</td>
<td>N/A</td>
<td>N/A</td>
<td>MKT1</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>170,000</td>
<td>None</td>
<td>None</td>
<td>PHM7, YOL073C, ATP19, PHM7, MSH2, HAL9</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>510,000</td>
<td>N/A</td>
<td>N/A</td>
<td>YPL038W-A, MET12, MET31, RMI1, YPL039W</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>50000</td>
<td>N/A</td>
<td>N/A</td>
<td>PUS2, ASF2</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>90,000</td>
<td>N/A</td>
<td>N/A</td>
<td>YDL203C, ASF2, PRR2, RTN2</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>190,000</td>
<td>N/A</td>
<td>N/A</td>
<td>ISC1, SMB1</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>150,000</td>
<td>N/A</td>
<td>N/A</td>
<td>UBS1, YLC1353W</td>
</tr>
<tr>
<td>21</td>
<td>9</td>
<td>70,000</td>
<td>N/A</td>
<td>N/A</td>
<td>RRD1, ECM37, YL130W</td>
</tr>
<tr>
<td>22</td>
<td>9</td>
<td>250,000</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>23</td>
<td>12</td>
<td>30,000</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>24</td>
<td>13</td>
<td>790,000</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>970,000</td>
<td>N/A</td>
<td>N/A</td>
<td>CPD1, YHB1, SOL4</td>
</tr>
<tr>
<td>28</td>
<td>8</td>
<td>130,000</td>
<td>GPA1</td>
<td>GPA1</td>
<td>ERG11, GPA1</td>
</tr>
</tbody>
</table>
**RRD1** as a causal gene for the eQTL hot spot at chrIX:70,000

RRD1 eQTL based on MPTGA

RRD1 ko vs wildtype
- 65 genes differentially expressed
- 3 of them overlap with genes in eQTL hot spot chrIX:70,000
- 21 of them overlap with genes in eQTL hot spot chrXV:170,000 (5.7 fold enrichment, p-value=1.9e-11)

Lin et al., (unpublished)
RRD1 as a causal gene for the eQTL hot spot at chrIX:70,000

RRD1 ko vs wildtype both treated with rapamycin
- 584 genes differentially expressed
- 52 of them overlap with genes in eQTL hot spot chrIX:70,000 (6.1 fold enrichment, p=2.0e-31)
- 35 of them overlap with genes in eQTL hot spot chrXV:170,000 (2.5 fold enrichment, p-value=5.6e-79)

Lin et al., (unpublished)
RRD1 as a causal gene for the eQTL hot spot at chrIX:70,000

Lin et al., (unpublished)
Lin et al., (unpublished)
## Acknowledgements

**Zhu lab**  
Seungyeul Yoo  
Eunjee Lee  
Li Wang  
Luan Lin  
Quan Long

**Mount Sinai Genomics Institute**  
Eric Schadt  
Bin Zhang  
Zhidong Tu  
Charles Powell  
Patrizia Casaccia

**Supported by:**

- Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai
- Janssen
- Canary Foundation
- Prostate Cancer Foundation
- NIH
- NCI

**Boston University**  
Avrum Spira  
Joshua Campbell

**U Washington**  
Roger Baumgarner

**Berkeley**  
Rachel Brem

**Princeton**  
Lenoid Kruglyak