Mining human genomes for genetic interactions underlying disease

Chad Myers

Department of Computer Science and Engineering
Program in Bioinformatics and Computational Biology

University of Minnesota- Twin Cities
Talk Overview

- Crash course in molecular biology
- The big questions in computational/systems biology
- An example problem in network biology from my lab
  - Understanding how genes interact in yeast and humans
- Sampling of other problems we’re working on
A crash course in molecular biology

- DNA: Deoxyribonucleic Acid
- Instructions, written in a 4-letter alphabet
  - Adenine, Thiamine, Cytosine, Guanine
A crash course in molecular biology

Information storage (DNA) → cellular machinery (proteins)

(A)lmost all organisms have this basic process in common

(adapted from Charles Mallery, http://fig.cox.miami.edu/~cmallery/150/gene/sf13x1.jpg)
The genome sequencing revolution

http://www.genome.gov/sequencingcosts/
Revolutionizing genome sequencing technology

1984
- 1st public discussion of Human genome sequencing project

1995
- 1st bacterial genome sequenced

1996
- Human genome sequencing begins

1997
- Yeast genome sequenced

1998
- C. elegans (roundworm)

2000
- Mouse

2002
- Human, fly finished

2018
- ~40,000 species/virus genomes sequenced!
What’s left to do?

With rapid sequencing technology, and many complete genomes sequenced, are we done? **NO!**

- the genome sequence is really just a “parts list”
- understanding the cell requires learning what each part does (e.g. which other parts it interacts with, which function(s) it carries out)
The big questions in computational/systems biology

• What are all of the parts? What is their function? When are they used?

• How do the various parts interact to carry out biological processes/form systems?

• How to genomes vary (evolve) across larger time scales?

• How do genomes vary across individuals? Which variants relate to which traits (e.g. disease)?

Computer science enables the investigation of all of these!
General comments about research in computational biology

• Highly collaborative

• Many problems related to managing and interpreting large, multidimensional datasets (machine learning, data mining, statistics are the basis of many approaches)

• Work is tied to or motivated by biological/biomedical questions, can work on a variety of them
My lab’s focuses

Gene expression variation in maize

Regulatory networks in C. elegans

Network analysis & algorithms

Genetic interactions in Arabidopsis

Targeted cancer therapeutics

Genetic interactions in yeast

Chemical genomics in yeast
Summary of computational techniques we apply/develop

• Data mining
  – Unsupervised clustering approaches
  – Association rule mining
  – PCA

• Machine/statistical learning methods
  – Decision trees/forests, SVMs, linear/logistic regression w/ regularization (e.g. lasso/elastic net), Bayesian networks

• Graph algorithms
  – Analysis/measurement of network properties, mining for local structures
The genome sequencing revolution: individual human genomes

Data from:
Human disease/trait-associations
2006-2018
May 2018: ~69,000 SNP-trait associations
> 5000 studies

Source: NHGRI-EBI Catalog
Genomics gets personal

scientist

physician

My genome

Variants:

- 3.4 million bases (1/1000 bases are unique)
- 10,231 variants affect a protein

“Clinically significant” findings:

<table>
<thead>
<tr>
<th>Variant</th>
<th>Interpretation</th>
<th>Associated Condition</th>
<th>Mode of Inheritance</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMPD1 c.133C&gt;T (p.Gln45*)</td>
<td>Pathogenic</td>
<td>Myoadenylate Deaminase Deficiency</td>
<td>Autosomal Recessive</td>
</tr>
</tbody>
</table>

Image credit: Jason Bobe, http://thepersonalgenome.com/category/zeitgeist/
Genomics gets personal

<table>
<thead>
<tr>
<th>Name</th>
<th>Confidence</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol Flush Reaction</td>
<td>★★★★★</td>
<td>Does Not Flush</td>
</tr>
<tr>
<td>Bitter Taste Perception</td>
<td>★★★★★</td>
<td>Can Taste</td>
</tr>
<tr>
<td>Earwax Type</td>
<td>★★★★★</td>
<td>Wet</td>
</tr>
<tr>
<td>Eye Color</td>
<td>★★★★★</td>
<td>Likely Blue</td>
</tr>
<tr>
<td>Hair Curl</td>
<td>★★★★★</td>
<td>Slightly Curlier Hair on Average</td>
</tr>
<tr>
<td>Lactose Intolerance</td>
<td>★★★★★</td>
<td>Likely Tolerant</td>
</tr>
<tr>
<td>Malaria Resistance (Duffy Antigen)</td>
<td>★★★★★</td>
<td>Not Resistant</td>
</tr>
<tr>
<td>Male Pattern Baldness</td>
<td>★★★★★</td>
<td>Increased Odds</td>
</tr>
<tr>
<td>Muscle Performance</td>
<td>★★★★★</td>
<td>Likely Sprinter</td>
</tr>
<tr>
<td>Non-ABO Blood Groups</td>
<td>★★★★★</td>
<td>See Report</td>
</tr>
<tr>
<td>Norovirus Resistance</td>
<td>★★★★★</td>
<td>Not Resistant</td>
</tr>
<tr>
<td>Resistance to HIV/AIDS</td>
<td>★★★★★</td>
<td>Not Resistant</td>
</tr>
<tr>
<td>Smoking Behavior</td>
<td>★★★★★</td>
<td>If a Smoker, Likely to Smoke More</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#SNP ID</th>
<th>chromosome</th>
<th>position</th>
<th>genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4477212</td>
<td>1</td>
<td>72017</td>
<td>AA</td>
</tr>
<tr>
<td>rs3094315</td>
<td>1</td>
<td>74265</td>
<td>AG</td>
</tr>
<tr>
<td>rs1313972</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs12124819</td>
<td>1</td>
<td>742429</td>
<td>AG</td>
</tr>
<tr>
<td>rs66681049</td>
<td>1</td>
<td>976643</td>
<td>AA</td>
</tr>
<tr>
<td>rs49703838</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs4475691</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs7537756</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs1110052</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs2272756</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs3748597</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs13303106</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs28413673</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs13303118</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs66665000</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs2341362</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs1897703</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs9679475</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs59401377</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs3128117</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs2461126</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs341365</td>
<td>1</td>
<td>742584</td>
<td>AG</td>
</tr>
<tr>
<td>rs271088</td>
<td>1</td>
<td>949705</td>
<td>CC</td>
</tr>
<tr>
<td>rs3128117</td>
<td>1</td>
<td>952073</td>
<td>AA</td>
</tr>
<tr>
<td>rs271087</td>
<td>1</td>
<td>967643</td>
<td>TT</td>
</tr>
<tr>
<td>rs2461136</td>
<td>1</td>
<td>980280</td>
<td>AA</td>
</tr>
<tr>
<td>rs2468991</td>
<td>1</td>
<td>984254</td>
<td>TT</td>
</tr>
<tr>
<td>rs49703838</td>
<td>1</td>
<td>982858</td>
<td>AG</td>
</tr>
<tr>
<td>rs3766192</td>
<td>1</td>
<td>100706</td>
<td>AA</td>
</tr>
<tr>
<td>rs3766191</td>
<td>1</td>
<td>1007450</td>
<td>AA</td>
</tr>
<tr>
<td>rs9442972</td>
<td>1</td>
<td>1011220</td>
<td>AA</td>
</tr>
<tr>
<td>rs10907177</td>
<td>1</td>
<td>1011278</td>
<td>AA</td>
</tr>
<tr>
<td>rs3737722</td>
<td>1</td>
<td>1011446</td>
<td>AC</td>
</tr>
</tbody>
</table>
“I should have been a sprinter??

**Alpha-actinin-3 (ACTN3)**

(Actin-binding protein expressed only in fast twitch fibers)

<table>
<thead>
<tr>
<th>Who</th>
<th>Genotype</th>
<th>What It Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mom</td>
<td>CC</td>
<td>Two working copies of alpha-actinin-3 in fast-twitch muscle fiber. <em>Many</em> world-class sprinters and some endurance athletes have this genotype.</td>
</tr>
<tr>
<td>Me</td>
<td>CT</td>
<td>One working copy of alpha-actinin-3 in fast-twitch muscle fiber. <em>Many</em> world-class sprinters and some endurance athletes have this genotype.</td>
</tr>
<tr>
<td>Dad</td>
<td>TT</td>
<td>No working copies of alpha-actinin-3 in fast-twitch muscle fiber. <em>Few</em> world-class sprinters have this genotype, but many world-class endurance athletes do.</td>
</tr>
</tbody>
</table>
Most variants result in relatively minor increase in disease risk

### Elevated Risk

<table>
<thead>
<tr>
<th>Name</th>
<th>Confidence</th>
<th>Your Risk</th>
<th>Avg. Risk</th>
<th>Compared to Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gallstones</td>
<td>★★★★★</td>
<td>11.1%</td>
<td>7.0%</td>
<td>1.66x</td>
</tr>
<tr>
<td>Age-related Macular Degeneration</td>
<td>★★★★★</td>
<td>9.5%</td>
<td>7.0%</td>
<td>1.35x</td>
</tr>
<tr>
<td>Restless Legs Syndrome</td>
<td>★★★★★</td>
<td>2.5%</td>
<td>2.0%</td>
<td>1.26x</td>
</tr>
<tr>
<td>Esophageal Squamous Cell Carcinoma (ESCC)</td>
<td>★★★★★</td>
<td>0.4%</td>
<td>0.4%</td>
<td>1.21x</td>
</tr>
<tr>
<td>Stomach Cancer (Gastric Cardia Adenocarcinoma)</td>
<td>★★★★★</td>
<td>0.3%</td>
<td>0.2%</td>
<td>1.22x</td>
</tr>
<tr>
<td>Scleroderma (Limited Cutaneous Type)</td>
<td>★★★★★</td>
<td>0.06%</td>
<td>0.07%</td>
<td>1.24x</td>
</tr>
</tbody>
</table>

### Decreased Risk

<table>
<thead>
<tr>
<th>Name</th>
<th>Confidence</th>
<th>Your Risk</th>
<th>Avg. Risk</th>
<th>Compared to Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer's Disease</td>
<td>★★★★★</td>
<td>4.9%</td>
<td>7.2%</td>
<td>0.69x</td>
</tr>
<tr>
<td>Melanoma</td>
<td>★★★★★</td>
<td>2.2%</td>
<td>2.9%</td>
<td>0.75x</td>
</tr>
<tr>
<td>Rheumatoid Arthritis</td>
<td>★★★★★</td>
<td>1.0%</td>
<td>2.4%</td>
<td>0.43x</td>
</tr>
<tr>
<td>Ulcerative Colitis</td>
<td>★★★★★</td>
<td>0.6%</td>
<td>0.8%</td>
<td>0.72x</td>
</tr>
<tr>
<td>Exfoliation Glaucoma</td>
<td>★★★★★</td>
<td>0.2%</td>
<td>0.7%</td>
<td>0.22x</td>
</tr>
<tr>
<td>Type 1 Diabetes</td>
<td>★★★★★</td>
<td>0.2%</td>
<td>1.0%</td>
<td>0.15x</td>
</tr>
<tr>
<td>Crohn's Disease</td>
<td>★★★★★</td>
<td>0.09%</td>
<td>0.53%</td>
<td>0.18x</td>
</tr>
<tr>
<td>Celiac Disease</td>
<td>★★★★★</td>
<td>0.07%</td>
<td>0.12%</td>
<td>0.58x</td>
</tr>
</tbody>
</table>
“Missing” heritability

• Variation in height is ~80% heritable
• known genetic loci (~40) only explain ~5% of heritable variance

<table>
<thead>
<tr>
<th>Disease</th>
<th>Number of loci</th>
<th>Proportion explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age-related macular degeneration</td>
<td>5</td>
<td>50%</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>32</td>
<td>20%</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>6</td>
<td>15%</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>18</td>
<td>6%</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>7</td>
<td>5.2%</td>
</tr>
<tr>
<td>Height</td>
<td>40</td>
<td>5%</td>
</tr>
<tr>
<td>Early onset myocardial infarction</td>
<td>9</td>
<td>2.8%</td>
</tr>
<tr>
<td>Fasting glucose</td>
<td>4</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

Manolio et al. Nature 461, 747-753 (8 October 2009)

http://www.dailymail.co.uk/health/article-1316471/Genes-decide-tall-short.html

Maybe combinations of genetic variants are the cause?
Detecting genetic interactions: a hopeless cause?

Exhaustive pairwise tests:
~500k-1 million loci
~0.5 trillion tests

Current paradigm:
Test individual variants for association

Statistical power issue!
A simpler organism, powerful genetic tools

Baker’s yeast (*Saccharomyces cerevisiae*)

How harmful are single gene deletions to a yeast cell?

Fitness distribution of all 6000 yeast deletion mutants

(Costanzo et al., 2010)
One interesting outcome of combining gene deletions

```
A
B
Wild-type
```
```
aΔ
X
B
Viable
```
```
A
X
bΔ
Viable
```
```
aΔ
X
bΔ
Lethal
```

“synthetic lethality”

Or, more generally, “genetic interaction”
A quantitative definition of genetic interaction

Fitness

wt
a
b
ab

"Neutral"
Expected Result
Multiplicative Model

0.25
0.5
0.5
1

(ab = a x b)
Types of genetic interaction: Negative

```
<table>
<thead>
<tr>
<th>Fitness</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>ab</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>
```

“Negative”
- Synthetic Lethal
- Expected Result

“Neutral”
- Parallel pathways

Example: Parallel pathways

Essential function

A

B
Types of genetic interaction: Positive

Fitness

- **wt**: 1
- **a**: 0.5
- **b**: 0.5
- **ab**: 0.5

“Neutral” Expected Result

“Positive” e.g. two genes whose products are not functionally independent
Types of genetic interaction: Positive

Fitness

- **wt**: 1
- **a**: 0.3
- **b**: 0.8
- **ab**: 0.8

0.24 = “Neutral” Expected Result

“Positive” (Suppression)
e.g. Mutant b suppresses growth defect of mutant a
Summary: quantitative definition of genetic interaction

\[ \varepsilon = f_{ij \text{observed}} - f_{ij \text{expected}} = f_{ij \text{observed}} - f_i f_j \]
The problem with combinatorial perturbation

Consider yeast, \( \sim 6000 \) genes

- \( \sim 18 \) million pairs
- \( \sim 10^{10} \) triples

Need automated, high throughput approach!
Robots to the rescue!


Charlie Boone, Brenda Andrews (U. Toronto)
From raw data to reliable, quantitative interactions

Synthetic Genetic Array

Colony growth model

(1536 colonies, 308 different double mutants)

$$
\epsilon = f_{ij \text{observed}} - f_{ij \text{expected}} = f_{ij \text{observed}} - f_i f_j
$$

Baryshnikova et al. Nature Methods 2010

Anastasia Baryshnikova, Ben VanderSluis, Elizabeth Koch
Automated yeast genetics


Charlie Boone, Brenda Andrews (U. Toronto)
From raw data to reliable, quantitative genetic interactions

Synthetic Genetic Array

(1536 colonies, 308 different double mutants)

Colony growth model

\[ \varepsilon = f_{ij_{\text{observed}}} - f_{ij_{\text{expected}}} = f_{ij_{\text{observed}}} - f_i f_j \]

Baryshnikova et al. Nature Methods 2010
A near complete genetic interaction map in yeast

~23 million double mutants screened for genetic interactions (~90% of possible gene pairs)

~1 million genetic interactions discovered
- 550,000 negative interactions
- 350,000 positive interactions

Global distribution of double mutant phenotypes

- Approximately 1 million genetic interactions
  - ~550,000 negative interactions (3%)
  - ~350,000 positive interactions (2%)

Genetic interactions are rare, but several-fold more dense than physical networks.

\[ \mathcal{E} = f_{ij, \text{observed}} - f_i f_j \]
Interaction profiles are predictive of gene function

Constructing a genetic interaction similarity network

Edge-weighted spring-embedded layout
Similar genetic interaction profiles reveal global functional map

(connection → similar interaction profile)
DNA replication

Similar genetic interaction profiles reveal global functional map
Similar genetic interaction profiles reveal global functional map.
Similar genetic interaction profiles reveal a global functional map.
Similar genetic interaction profiles reveal global functional map
Similar genetic interaction profiles reveal global functional map

- RNA splicing
- RNA decay
- Nuclear pore
- Vacuole sorting
- Actin cytoskeleton
- Budding, cytokinesis
- Spindle assembly, Cohesion and Kinetochore
- DNA replication
- DNA recomb. and repair
- Dynein-dynactin
- Tubulin folding
- Chromatin remodeling
- Proteosome
- APC complex
- RNA Pol II
Major lesson: Genetic interactions are highly structured
<table>
<thead>
<tr>
<th>TID</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{Bread, Milk}</td>
</tr>
<tr>
<td>2</td>
<td>{Bread, Diapers, Beer, Eggs}</td>
</tr>
<tr>
<td>3</td>
<td>{Milk, Diapers, Beer, Cola}</td>
</tr>
<tr>
<td>4</td>
<td>{Bread, Milk, Diapers, Beer}</td>
</tr>
<tr>
<td>5</td>
<td>{Bread, Milk, Diapers, Cola}</td>
</tr>
</tbody>
</table>
## Crash course in association rule mining

<table>
<thead>
<tr>
<th></th>
<th>Bread</th>
<th>Milk</th>
<th>Diapers</th>
<th>Beer</th>
<th>Eggs</th>
<th>Cola</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Enabling exhaustive combinatorial pattern search: smart pruning of patterns

Enabling exhaustive combinatorial pattern search: smart pruning of patterns

Exhaustive mining for local network structures

Results (negative interactions):
Real network: 10,459 blocks
Randomized networks: ~20 blocks

(>= 3 x 3 interaction blocks)

Bellay et al., *Genome Research* 2011
Is the yeast genetic interaction map relevant for helping find interactions in human populations?
Major lesson: Genetic interactions are highly structured

Estimated frequency: ~25%

~75%

Bellay et al., Genome Research 2011
BridGE: a method for detecting genetic interactions from GWAS data

Human genotype data (~1000 diseased/control individuals)

Pathways/gene relationships

BridGE algorithm

Pathway 1

Pathway 2

Disease-associated genetic interactions (pathway-level)

SNP-level associations

Wang et al. PLoS Genetics 2017
Fang, Wang et al. bioRxiv 2017
BridGE: a method for detecting genetic interactions from GWAS data

(1) Map SNPs to established pathways/gene sets (KEGG, Biocarta, Reactome)

(2) Construct network from individual pairwise SNP interaction tests (noisy)

(3) Test pathway-pathway combinations for enrichment of noisy SNP-SNP interactions

(4) Estimate significance through SNP and sample permutations

Wen Wang
Gang Fang
Vipin Kumar
BridGE: network motifs we hope to detect
Example interaction (Parkinson’s disease)

Parkinson’s disease GWAS
Simón-Sánchez et al., Nature Genetics, 2009

Wen Wang
Gang Fang
Vipin Kumar
Example interaction (Parkinson’s disease)

2 interactions discovered at FDR < 0.05
1 replicates in independent cohort

Wen Wang
Gang Fang
Vipin Kumar
Evidence for many genetic interactions in Parkinson’s disease

FDR < 0.25: 23 interactions
Evidence for many genetic interactions in Parkinson’s disease

FDR < 0.25: 23 interactions

Pathways that involve single associated SNP in dbGaP ($p < 10^{-5}$)
Evidence for genetic interactions in many diseases

We’ve discovered interactions in 7 of the 8 diseases we’ve analyzed to date

- Parkinson’s Disease
- Schizophrenia
- Bipolar Disorder
- Breast Cancer
- Prostate Cancer
- Type II Diabetes
- Hypertension

- 14 cohorts examined, representing 8 different diseases
- Significant interactions in 12/14 cohorts, representing 7 diseases
- ~25% of pathways identified were implicated using traditional GWAS methods
Evidence for genetic interactions in many diseases

Wen Wang
What’s next?  Challenges and opportunities

• Causal genetic mechanism $\rightarrow$ therapeutics? (there are many more “good targets” than are currently targeted for most diseases)

• Convergence of population genetics and reverse genetic (i.e. genome editing) approaches in human cells
  • we can now introduce/phenotype individual variants and combinations in cell-based models
What’s next? Challenges and opportunities

GWAS significant loci (2005-present)

Targets of drugs (1940-present)
Sampling of other projects in the lab

- Higher order interactions (triple mutants)
- Interactions in other conditions (environments)
- Chemical genetic approaches for large-scale characterization of chemicals’ effects on cells
- CRISPR-enabled genetic interaction screens in human cells
Chemical genetic profiling

Deletion Mutants

Drug

\[\text{\textcolor{red}{\square}} = \text{Sensitivity}\]

Gaiever/Nislow:
Hillenmeyer et al. 2008
Lee et al. 2014
Combining gene-gene interactions and chemical-gene interactions

Key idea: Drug mimics target’s genetic interaction profile
Barcode sequencing to generate chemical-genetic interaction profiles

Barcoded Deletion Collection

ACAG... TAGG... GATA... CCAG...

Barcodes

Drug_A Drug_B Drug_C

Multiplex via PCR

(Identifies the strain) (Identifies the condition)
Barcode Multiplex Tag

Deletion Mutants

Drug

768 compound conditions per HiSeq lane

Barseq: Smith et al. Genome Research 2009
## Compound libraries screened

<table>
<thead>
<tr>
<th>Library Name</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIKEN Natural Product Depository</td>
<td>9849</td>
</tr>
<tr>
<td>NCI Structural Diversity Set</td>
<td>1596</td>
</tr>
<tr>
<td>NCI Mechanistic Diversity Set</td>
<td>909</td>
</tr>
<tr>
<td>NCI Oncology</td>
<td>113</td>
</tr>
<tr>
<td>NCI Natural Products</td>
<td>117</td>
</tr>
<tr>
<td>NIH Clinical Collection</td>
<td>816</td>
</tr>
<tr>
<td>GSK Published Kinase Inhibitors</td>
<td>384</td>
</tr>
<tr>
<td>CNCL Compound Library</td>
<td>4480</td>
</tr>
</tbody>
</table>

**Total: > 18,000 compounds screened**
8419 compounds
289 genes
Genes \rightarrow Compounds
8419 compounds
289 genes

Genes → Compounds →

Glycosylation  Unfolded protein response

Compounds →

TUB3  CIN1

SMI1  BEM2
UME1  SSD1
ABP1

RAD5  RAD55  MMS4
PDA1  SHE9

SWC5  LSM1  LSM6  PAT1

RPN4  UBX4

ALG5  ALG6  ALG8
HAC1  IRE1

DFG5  FLC2

GCS1  PAR32  GYP1  GUP1  ROM1
New molecular probes span a diversity of biological processes

CRISPR-enabled genetic interaction screens in human cells
CRISPR enables genetic interaction screens in human cell lines

(70k guides targeting 18k genes)

Knockout query gene  Genome-wide lentiviral KO-library  ~18 doublings

mutant A  double mutant pool AX, AY, AZ, ...

mutant B  double mutant pool BX, BY, BZ, ...

wild type  single mutant pool X, Y, Z, ...

A & Y synthetic lethal

B & Z synthetic lethal

(collaboration with Jason Moffat Lab, U. Toronto)
Take-home messages

- Genetic interactions are a powerful approach to functional characterization from global structure down to the level of specific pathways.

- Fundamental questions remain about the interpretation of genetic interactions.

- Many computational challenges remain in obtaining/interpreting genetic interaction networks.

- Insights from model organisms are directly relevant for understanding/treating of human disease—computational scientists help build these “bridges” across species.
Some parting thoughts about inter-disciplinary science

- Many exciting problems lie in between the boundaries, most of the barriers relate to communication
- Productive relationships take years to build
- Your selfish interests need to be aligned to be productive
- There is more than 100% credit
Acknowledgements

Wen Wang

Gang Fang

Vipin Kumar’s lab (UofM), now: Asst. Prof. at Mt. Sinai SoM
Acknowledgements

University of Minnesota

Wen Wang
Elizabeth Koch
Benjamin VanderSluis
Carles Pons
Raamesh Deshpande
Scott Simpkins
Justin Nelson
Hamid Safizadeh
Jeremy Bellay
Yungil Kim
Tian Xia
Roman Briskine
Rob Schaefer

Funding
NSF (DBI 0953881, MCB 0918908, MCB 1818293)
NIH (R01HG005084, R01HG005853, R01GM104975)
UMN BICB program
UMN Supercomputing Institute
MN Partnership for Biomedical Genomics
Canadian Institute for Advanced Research

Princeton
Anastasia Baryshnikova (now Calico)

Toronto
Charlie Boone
Brenda Andrews
Michael Costanzo
Sheena Li (RIKEN)
Jeff Piotrowski
Jacqui Barber (RIKEN)
Kerry Andrusiak
Elena Kuzmin

University of Minnesota

Vipin Kumar
Gang Fang
Gowtham Atluri
Gaurav Pandey
Michael Steinbach

Nathan Pankratz
Carol Lange