Understanding the molecular basis of human disease by mapping across tissues and organisms

Olga Troyanskaya
Professor
Lewis-Sigler Institute for Integrative Genomics and Department of Computer Science
Princeton University

Deputy Director for Genomics
Simons Center for Data Analysis
From “BIG” genomics data to understanding of biology and human disease

• Gene expression
• Physical PPIs
• Epistatic interactions
• Colocalization
• Sequence
• Protein domains
• Regulatory binding sites

• Using functional genomics data in human to discover disease genes
• Combining functional genomics with quantitative genetics
• Leveraging the power of model organisms on a systems level
Tissue-specificity critical in human disease:
each cell type performs a specialized function

From “Big Data” in functional genomics to
disease-relevant molecular maps

http://www.globalstem.com/image/data/Stem-Neural/gy.png
We need to understand pathways and processes in a **cell- and tissue-specific context**

With Garret Fitzgerald’s and Tilo Grosser’s groups (Penn), Daniel Chasman (Brigham and Women’s Hospital, Harvard), Kara Dolinski (Princeton)
There’s a *flood* of high-throughput genomic data

- Physical & genetic interactions
- TF-binding / miRNA-target sites
- Gene expression profiles
- Noisy & Heterogenous
- Not resolved to cell types & tissues
- Many datasets are not annotated to cell-type/tissue of origin
- Most samples are cell-type/tissue mixtures
- Many cell-types are hard to isolate or not enough to profile

967 datasets with genome-wide measures; ~38,000 conditions from ~14,000 publications

Gene expression profiles

We need to extract tissue-specific signal from data that is not resolved to tissue/cell-type
Integrating human tissue-specific networks

1. 1000s of Genomic datasets

2. Hierarchically integrated tissue-specific knowledge

Machine learning classifier

Body
- Cardio Sys
- Nervous Sys
- Urinary Sys
  - Vascular Sys
  - Brain
  - Kidney
  - Blood Vessel

Brain
- Kidney
- Blood vessel
- Muscle

144 Tissue-specific networks
Tissue networks can predict disease-relevant lineage-specific response

Inflammation in blood vessels is associated with several cardiovascular diseases. Cytokine IL1B is an important mediator of inflammatory response.

Can we use the blood vessel specific network to predict this?

The blood vessel network is the most accurate one in predicting IL1B response in blood vessel cells.

18/20 network neighbors respond to IL1B

With Emanuela Ricciotti and Tilo Grosser
NetWAS: Network-based approach for reprioritizing GWAS results to identify disease genes and potential drug targets

Genome-wide association study (GWAS) is a powerful approach to catalogue trait-associated sequence variants.

- Low frequency mutations
- Small effect sizes
- Epistasis

Can we improve GWAS results using tissue-specific networks?

Greene et al. *Nature Genetics*, 2015
NetWAS: Network-based approach for reprioritizing GWAS results

Top GWAS hits for a disease – e.g. hypertension – are potentially enriched for disease genes

Case study
- Women’s Genome Health Study
- Three endpoints: systolic (SBP), diastolic blood pressure (DBP), and hypertension (HTN) diagnosis
NetWAS: Network-based approach for reprioritizing GWAS results

**Phenotypic**
Hypertension disease genes from OMIM

**Functional**
Blood pressure regulatory genes from GO

**Therapeutic**
Antihypertensive drug-targets from DrugBank
NetWAS: Network-based approach for reprioritizing GWAS results

- **Hypertension (combined)**
  - AUC: 0.75
  - Tissue: heart, liver, gwas

- **Type 2 Diabetes (T2D)**
  - AUC: 0.60
  - Tissue: adipose tissue, pancreas, liver, gwas

- **Alzheimer's Disease (hippocampal vol)**
  - AUC: 0.60
  - Tissue: hippocampus, gwas

- **Obesity (BMI)**
  - AUC: 0.65
  - Tissue: adipose tissue, gwas

- **Systemic Lupus Erythematosus (CRP)**
  - AUC: 0.80
  - Tissue: b lymphocyte, t lymphocyte, gwas

- **Age-related Macular Degeneration (advanced AMD)**
  - AUC: 0.70
  - Tissue: eye, retina, gwas
NetWAS: Network-based approach for reprioritizing GWAS results

GIANT
Genome-scale Integrated Analysis of gene Networks in Tissues

Tissue-specific Interactions
GIANT leverages a tissue-specific gold standard to automatically up-weight datasets relevant to a tissue from a large data compendium of diverse tissues and cell-types. The resulting functional networks accurately capture tissue-specific functional interactions.

Multi-tissue Analysis
Beyond questions pertaining to the role of single genes in single tissues, GIANT also enables examination of changes in gene function across tissues on a broad scale. Users can compare a gene’s functional interaction in different tissues by selecting the relevant tissues in the dropdown menu.

NetWAS Analysis
GIANT can effectively reprioritize functional associations from a genome-wide association study (GWAS) and potentially identify additional disease-associated genes. The approach, named NetWAS, can be applied to any GWAS study, and does not require that the phenotype or disease have any known associated genes.
The challenges of studying human disease or “model systems to the rescue”

Most human diseases are molecularly under-characterized

1. The genes/mechanisms are poorly understood
2. Performing specific genetic experiments is very hard

There are excellent model systems out there

1. Each model organism/system best suited for studying different diseases/processes/aspects
2. Genetics is a lot more tractable
3. There are already genotype-phenotype data available in primates, mouse, zebrafish, worm, fly, yeast (and cell lines)
Which system to use - that is the question!

For a given human disease/phenotype,

1. Which is the best phenotype to test in a given model organism?
2. Which is the best model organism/system?

<table>
<thead>
<tr>
<th>Human disease/phenotype</th>
<th>Zebrafish phenotype</th>
<th>Yeast phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse phenotype</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mapping diseases and phenotypes is challenging:

1. Diseases and phenotypes often poorly understood at the molecular level (e.g. gene-phenotype associations very incomplete)
2. Many-to-many relationships between genetic mutations and observable phenotypes
3. Even semantically similar phenotypes are differently assayed, recorded, and annotated across organisms
Linking human disease to model phenotypes on the molecular level

Disease-associated genes

Phenotype-associated genes

Park et al. *PLOS Comp Biol* 2012
How do we evaluate this approach?

1. Network-based score
2. Overlap-based score
Parkinson's Disease

A synucleinopathy that has a basis in degeneration of the central nervous system that often impairs motor skills, speech, and other functions. Parkinson disease was first described by James Parkinson in 1817. It is the second most common neurodegenerative disorder after Alzheimer disease, affecting approximately 1% of the population over age 50 (Polymeropoulos et al. 1996).

With Kara Dolinski group (Princeton)
Candidate genes for Parkinson's disease in *C. elegans*

- Candidate genes for Parkinson's predicted based on worm dopaminergic neuron network and human GWAS studies
- Example: age-dependent motility defect in BCAT-1 (predominant branched-chain aminotransferase in the nervous system)
- Inhibited by neuronal drug gabapentin (GABA analog)

With Coleen Murphy’s group (Princeton)
Summary: tissue-specific genome-scale view of human biology and disease

A general method can perform tissue-specific data integration into functional networks.

Tissue-specific networks are complementary to quantitative genetics and can disease associations and reprioritize GWAS hits.

Human diseases and pathways can be linked to best model systems through organism-specific functional networks.
Former lab members, now faculty at:
* Curtis Huttenhower (@ Harvard SPH)
* Chad Myers (@ U MN)
* David Hess (@ USantaClara)
* Matthew Hibbs (@JAX)
* Florian Markowetz (@Cancer UK)
* Edo Airoldi (@ Harvard)
* Lars Bongo (@ U Tromso)
* **Casey Greene (@ Dartmouth)**
* Yuanfang Guan (@ Umich AA)
* Maria Chikina (@ Pitt)
* Patrick Bradley (postdoc @Gladstone)
* Chris Park (postdoc @NYGC)
* Ana Pop (@MIT)

Research Scientists:
* **Aaron Wong**
* **Chandra Theesfeld**

Postdoctoral Fellows
* **Arjun Krishnan**
* Benjamin Vandersleus
* Salim Choudhury

Graduate Students
* Qian Zhu
* Dima Gorenshteyn
* Young–Suk Lee
* Vicky Yao
* Max Homilius
* Jonathan Goya
* Jian Zhou
* Ran Zhang
* Ruth Dannenfelser

| IMP | [http://imp.princeton.edu](http://imp.princeton.edu) |
| NanoDissection | [http://nano.princeton.edu](http://nano.princeton.edu) |
| GIANT | [http://girant.princeton.edu](http://girant.princeton.edu) |