Understanding Protein Function on a Genome-scale through the Analysis of Molecular Networks

Mark B Gerstein
Yale

slides at Lectures.GersteinLab.org

(See Last Slide for References & More Info.)
The problem: Grappling with Function on a Genome Scale?

- 250 of ~530 originally characterized on chr. 22 [Dunham et al. Nature (1999)]
- >25K Proteins in Entire Human Genome (with alt. splicing)
Traditional single molecule way to integrate evidence & describe function

Descriptive Name: Elongation Factor 2

Summary sentence describing function:
This protein promotes the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome.

Lots of references to papers
Some obvious issues in scaling single molecule definition to a genomic scale

- Fundamental complexities
  - Often >2 proteins/function
  - Multi-functionality:
    2 functions/protein
  - Role Conflation:
    molecular, cellular, phenotypic
Some obvious issues in scaling single molecule definition to a genomic scale

- Fundamental complexities
  ◦ Often >2 proteins/function
  ◦ Multi-functionality:
    2 functions/protein
  ◦ Role Conflation:
    molecular, cellular, phenotypic
- Fun terms… but do they scale?....
  ◦ Starry night (P Adler, ’94)
An Ontology of Naming Pathologies

Single

M

Explicit meaning

M-scientific

SEMA5A

Not "funny"; usually acronym or concatenation of long descriptive scientific name

M-literal

drop dead

Inherent meaning of words is sufficient to describe gene function in some way; no cultural knowledge is required

M-embed

Clever reference or allusion. Cultural savvy or other knowledge required to make sense

Literary

malvolio

Acronym

LOV

Historical

yuri

Pop culture

tribbles

P

No explicit meaning

~M

~M-outside

kuzbanian

Some outside, non-obvious reason for name

~M-irrel

ring

Irrelevant acronym; not tied to gene function

~M-nr

yippee

Silly or funny names. No relevance to underlying gene function

Multi

T

Transferred naming system

T-relation

kryptonite and superman

Naming ceases to make sense if names are shuffled among genes

T-norelation

arleekin valiet tungus...

Names could be shuffled among genes with no loss of meaning

P

Problematic relationships

P-clash

PKD1 and lov-1

Analogous genes with very different names

P-confusion

MT-1

Many genes with same name, or many names for one gene

P-defunct

BAF45 and BAF47

Gene named to reflect information later shown to be inaccurate or untrue

[Seringhaus et al. GenomeBiology (2008)]
Gene Name Skew

[Seringhaus et al. GenomeBiology (2008)]
Hierarchies & DAGs of controlled-vocab terms but still have issues...

MIPS (Mewes et al.)

GO (Ashburner et al.)

[Seringhaus & Gerstein, Am. Sci. '08]
Towards Developing Standardized Descriptions of Function

• Subjecting each gene to standardized expt. and cataloging effect
  ◊ KOs of each gene in a variety of std. conditions => phenotypes
  ◊ Std. binding expts for each gene (e.g. prot. chip)

• Function as a vector

<table>
<thead>
<tr>
<th></th>
<th>DNA</th>
<th>RNA</th>
<th>ATP</th>
<th>Metal</th>
<th>CoA</th>
<th>NAD</th>
<th>...</th>
<th>G protein</th>
<th>CDC28</th>
<th>Calmodulin</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein 1</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>protein 2</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>protein 3</td>
<td>1.0</td>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>protein 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>...</td>
</tr>
<tr>
<td>protein 5</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>protein 6</td>
<td>0.9</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>protein 7</td>
<td>0</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Interaction Vectors [Lan et al, IEEE 90:1848]
Networks (Old & New)

**Fringe:** Vital in boundary formation in developing fly wing.
**Numb:** mutations impair sensory organs in flies
**Notch:** with defects, flies develop notches in wings

**Itch:** linked to itchy skin in mice

Classical KEGG pathway

Same Genes in High-throughput Network

[Seringhaus & Gerstein, Am. Sci. '08]
Networks occupy a midway point in terms of level of understanding.

1D: Complete Genetic Parts list

~2D: Bio-molecular Network Wiring Diagram

3D: Detailed structural understanding of cellular machinery


[Fleischmann et al., Science, 269:496]
Networks as a universal language

- Internet [Burch & Cheswick]
- Food Web
- Electronic Circuit
- Disease Spread [Krebs]
- Neural Network [Cajal]
- Protein Interactions [Barabasi]
- Social Network
Using the position in networks to describe function

Guilt by association

Finding the causal regulator (the "Blame Game")

[NY Times, 2-Oct-05, 9-Dec-08]
Network pathology & pharmacology

Breast Cancer
Alzheimer’s Disease
Parkinson’s Disease
Multiple Sclerosis

Interactome networks

[Adapted from H Yu]
Outline: Molecular Networks

• Why Networks?

• Central Points in Networks
  ◊ Hubs & Bottlenecks
    (yeast ppi & reg. net)
  ◊ Tops of Hierarchies
    (yeast reg. net)
  ◊ Identified by score
    (human miRNA-targ. net)

• Dynamics of Networks
  ◊ Across environments
    (in prokaryote metab. pathways)

• Protein Networks & Variation
  (human ppi & miRNA-targ. net)
Different Types of Molecular Networks

- Protein-protein Interaction networks
- TF-target-gene Regulatory networks
- Metabolic pathway networks
- miRNA-target networks

Finding Central Points in Networks: Hubs & Bottlenecks

Where are key points in networks? How do we locate them?
Global topological measures

Indicate the gross topological structure of the network

Interaction and expression networks are *undirected*

[Barabasi]
Global topological measures for directed networks

Regulatory and metabolic networks are directed
Scale-free networks

Power-law distribution

\[ P(k) \sim k^{-\gamma} \]

*Hubs* dictate the structure of the network

[Barabasi]
Hubs tend to be Essential

Integrate gene essentiality data with protein interaction network. Perhaps hubs represent vulnerable points?

[Lauffenburger, Barabasi]
Relationships extends to "Marginal Essentiality"

Marginal essentiality measures relative importance of each gene (e.g. in growth-rate and condition-specific essentiality experiments) and scales continuously with "hubbiness"

[Yue et al., 2003, TIG]
Another measure of Centrality: Betweenness centrality

Betweenness of a node is the number of shortest paths of pairs of vertices that run through it -- a measure of information flow.


Betweenness centrality -- Bottlenecks

Proteins with high betweenness are defined as *Bottlenecks* (top 20%), in analogy to the traffic system.
Bottlenecks & Hubs

Hub-bottleneck node
Non-hub-bottleneck node
Hub-non-bottleneck node
Non-hub-non-bottleneck node

[Yu et al., PLOS CB (2007)]
Bottlenecks are what matters in regulatory networks

\[ P < 10^{-20} \]

\[ P < 10^{-4} \]

Finding Central Points in Networks #2: Tops of the Hierarchy

Where are key points networks? How do we locate them?
Determination of "Level" in Regulatory Network Hierarchy with Breadth-first Search

I. Example network with all 4 motifs

II. Finding terminal nodes (Red)

III. Finding mid-level nodes (Green)

IV. Finding top-most nodes (Blue)

Level 3
Level 2
Level 1

[Yu et al., PNAS (2006)]
Regulatory Networks have similar hierarchical structures

[S. cerevisiae

E. coli

Yu et al., Proc Natl Acad Sci U S A (2006)]
Example of Path Through Regulatory Network

Expression of MOT3 is activated by heme and oxygen. Mot3 in turn activates the expression of NOT5 and GCN4, mid-level hubs. GCN4 activates two specific bottom-level TFs, Put3 and Uga3, which trigger the expression of enzymes in proline and nitrogen utilization.

[Yu et al., PNAS (2006)]
Yeast Regulatory Hierarchy: the Middle-managers Rule

A. Regulatory hierarchy in *S. cerevisiae*

[Yu et al., PNAS (2006)]
Yeast Network Similar in Structure to Government Hierarchy with Respect to Middle-managers

B. Governmental hierarchy of a representative city (Macao)
Characteristics of Regulatory Hierarchy: Middle Managers are Information Flow Bottlenecks

Average betweenness at each level

[Yu et al., PNAS (2006)]
Characteristics of Regulatory Hierarchy: The Paradox of Influence and Essentiality

[Yu et al., PNAS (2006)]
Finding Central Points in Networks #3: Points of Maximal Regulatory Effect
• How much does a regulator influence its targets?
• For miRNA-target networks easy to calculate, as all influence is down-regulation
  ◊ target prediction via: TargetScan, PITA, PicTar, miRanda, …
• Look at down-reg. genes in a sample & compare with targets of a specific micro-RNA
  ◊ more down-reg genes => stronger regulatory effect

**RE-score: Another way to identify "important" network nodes**

**Formula:**

\[ \text{RE score} = R_n - R_t \]

Cheng et al., Genome Biology, 2009
Application of RE-score to measure changing miRNA effect in different conditions (ER- and ER+ breast cancer)

Cheng et al., Genome Biology, 2009
RE-score can be used to classify cancers

(1) RE-score profile for different miRNA in 1 cancer sample.
(2) Tabulate over many different breast cancer samples.

(3) Clustering based on RE score divides samples into 2 main types of cancer.

(4) Clustering better than based on individual gene expression levels.

Cheng et al., Genome Biology, 2009
Network Dynamics: Environments
How do molecular networks change across environments? What pathways are used more? Used as a biosensor?
What is metagenomics?

**Genomics Approach**

1. Culture Microbes
2. Extract DNA
3. Sequence: ATCGTATA CGCGAAG ACGTCTGA AGTGCTGCT
4. Assemble and Annotate

**Problem:** Estimated that less than 1% can be cultured in the lab.

**Metagenomics Approach**

1. Collect Sample
2. Extract DNA
3. Sequence: ATCGTATA CGCGAAG ACGTCTGA AGTGCTGCT
4. Partially Assemble and Annotate

**Problem:** Lose information about which gene belongs to which microbe.
Global Ocean Survey Statistics (GOS)

- 6.25 GB of data
- 7.7M Reads
- 1 million CPU hours to process

Rusch, et al., PLOS Biology 2007
Expressing data as matrices indexed by site, env. var., and pathway usage

[Rusch et. al., (2007) PLOS Biology; Gianoulis et al., PNAS (in press, 2009)]
Simple Relationships: Pairwise Correlations
 Canonical Correlation Analysis: Simultaneous weighting

\[
\text{UPI} = a \cdot \text{GRE} + b \cdot \text{GPA} + c
\]

\[
\text{GPI} = a' \cdot \text{GRE} + b' \cdot \text{PowerPoint} + c'
\]

[ Gianoulis et al., PNAS (in press, 2009) ]
Canonical Correlation Analysis: Simultaneous weighting

<table>
<thead>
<tr>
<th>Score</th>
<th># of papers published</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRE</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Undergraduate Performance Index (UPI)</th>
<th>Graduate School Performance Index (GPI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRE</td>
<td>GPA</td>
</tr>
</tbody>
</table>

Environmental Features
- Temp
- Chlorophyll

Metabolic Pathways
- Photosynthesis
- Lipid Metabolism

[ Gianoulis et al., PNAS (in press, 2009) ]
The goal of this technique is to interpret cross-variance matrices. We do this by defining a change of basis.

Given \( X = \{x_1, x_2, \ldots, x_n\} \) and \( Y = \{y_1, y_2, \ldots, y_m\} \)

\[
C = \frac{\sum_X \sum_{X,Y} \sum_{Y,X}}{\sqrt{a' \sum_{11} a \sqrt{b' \sum_{22} b}}}
\]

\[
\max \ Corr(U, V) = \frac{a' \sum_{12} b}{\sqrt{a' \sum_{11} a \sqrt{b' \sum_{22} b}}}
\]

[ Gianoulis et al., PNAS (in press, 2009) ]
**DPM: Discriminative Partition Matching**

Environment

Cluster (Partition)

Test

<table>
<thead>
<tr>
<th>Functional class</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>InfoStorage &amp; Processing</td>
<td>.07</td>
</tr>
<tr>
<td>Cellular Process</td>
<td>.08</td>
</tr>
<tr>
<td>Metabolism</td>
<td>4x10-14</td>
</tr>
</tbody>
</table>

[ Gianoulis et al., PNAS (in press, 2009) ]
Strength of Pathway co-variation with environment

[ Gianoulis et al., PNAS (in press, 2009) ]
Conclusion #1: energy conversion strategy, temp and depth

[ Gianoulis et al., PNAS (in press, 2009) ]
Conclusion #2: Outer Membrane components vary the environment

[ Gianoulis et al., PNAS (in press, 2009) ]
Conclusion #3: Covariation of AA biosynthesis and Import

Why is their fluctuation in amino acid metabolism? Is there a feature(s) that underlies those that are environmentally-variant as opposed to those which are not?

[ Gianoulis et al., PNAS (in press, 2009) ]
Biosensors: Beyond Canaries in a Coal Mine

[ Gianoulis et al., PNAS (in press, 2009) ]
Networks & Variation

Which parts of the network vary most in sequence?
Which are under selection, either positive or negative?
METHODOLOGY: MAP SNP AND CNV DATA ONTO ENSEMBL GENES, AND THEN MAP ENSEMBL GENES TO THE KNOWN INTERACTOME

ILLUSTRATIVE

Hapmap/Perlegen

SNPs

Ensembl Genes

ENSG000XXXX: rsSNP00XXX
CNV_XXX
DN/DS XXXX
Recombination rate

Interactome

~30000 interactions from HPRD and Y2H screens

Database of Genomic Variants

CNVs + SDs

Result

• Dataset of network position / parameters (e.g. degree centrality or betweenness centrality) in relationship to SNPs, CNV’s, recombination rates and positive selection tests


Source: PMK
ADAPTIVE EVOLUTION CAN BE SEEN ON TWO DIFFERENT LEVELS

Intra-species variation

- Single-basepair
  - Positive Selection

Single-Nucleotide Polymorphisms

- Fixed Differences
  - Positive Selection

Structural variation

- Copy Number Variants
  - Positive Selection

Segmental Duplications

Fixed mutations (differences to other species)

Source: PMK
POSITIVE SELECTION LARGELY TAKES PLACE AT THE NETWORK PERIPHERY

CENTRAL PROTEINS ARE LESS LIKELY TO BE UNDER POSITIVE SELECTION

Peripheral genes are likely to undergo positive selection, whereas hubs aren’t. This is likely due to the following reasons:

- Hubs have stronger structural constraints, the network periphery doesn’t.
- Most recently evolved functions (e.g. “environmental interaction genes” such as sensory perception genes etc.) would probably lie in the network periphery.

Effect is independent of any bias due to gene expression differences.

* With a probability of over 80% to be positively selected as determined by Ka/Ks. Other tests of positive selection (McDonald Kreitmann and LDD) corroborate this result.

CENTRAL NODES ARE LESS LIKELY TO LIE INSIDE OF SDs

This result also confirms our initial hypothesis – peripheral nodes tend to lie in regions rich in SDs.

Since segmental duplications are a different mechanism of ongoing evolution, the less constrained peripheral proteins are enriched in them.

Note that despite the small size of our dataset for known SD’s we get significant correlations. It is to be expected that the correlations will get clearer as more data emerges.

* Specifically, a number of the SDs are likely not fixed, but rather common CNVs in the reference genome.

Another explanation: THE NETWORK PERIPHERY CORRESPONDS TO THE CELLULAR PERIPHERY

Source: Gandhi et al. (Nature Genetics 2006), Kim et al. PNAS (2007)
IS RELAXED CONSTRAINT OR ADAPTIVE EVOLUTION THE REASON FOR THE PREVALENCE OF BOTH SELECTED GENES AND SDs AT THE NETWORK PERIPHERY?

<table>
<thead>
<tr>
<th><strong>Relaxed Constraint</strong></th>
<th><strong>Adaptive Evolution</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inter-Species Variation (Fixed differences)</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intra-Species Variation (Polymorphisms)</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Kim et al. PNAS (2007)
SOME, BUT NOT ALL OF THE SINGLE-BASEPAIR SELECTION AT THE PERIPHERY IS DUE TO RELAXED CONSTRAINT

Inter vs. Intra-Species Variation in Networks

<table>
<thead>
<tr>
<th></th>
<th>Inter-Species (Fixed differences)</th>
<th>Intra-Species (SNPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betweenness Centrality ($\times 10^4$)</td>
<td>2.71</td>
<td>4.08</td>
</tr>
<tr>
<td>Genes with $dN/dS &gt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genes with $dN/dS \leq 1$</td>
<td>4.37</td>
<td>4.35</td>
</tr>
<tr>
<td>p &lt;= 0.01</td>
<td>p &lt; 0.05</td>
<td></td>
</tr>
</tbody>
</table>

Reasoning

- There is a difference in **variability** (in terms of SNPs) between the network periphery and the center
- However, this difference is much smaller than the difference in **selection**
- This most likely means, that part of the effect we're seeing is due to relaxed constraint (and higher variability)
- But, not the entire effect*

*But it’s hard to quantify
Source: Kim et al. (2007) PNAS
Similar Results for Large-scale Genomic Changes (CNVs and SDs)

- There is a small difference in variability (in terms of CNVs) between the network periphery and the center.

- But, there is a (as shown before) marked difference in fixed (and hence, presumably, selected) SDs at the network periphery and center.

Source: Kim et al. (2007) PNAS
Networks & Variation 2

Variation in the miRNA network
Analyze Regulation in microRNA-target Network

• Relationship between target in degree (number of micro-RNAs that regulate gene) & evolutionary rate of gene?
  ◊ In deg. related 3' UTR size

• Expectation: more regulation, more constraint
Relationship between microRNA regulation and protein evolution

Important genes are regulated more intensively regulated by the microRNAs

[Cheng et al., BMC Genomics, 2009 (in press)]
MicroRNA regulation: a two-way strategy

For non-housekeeping genes, functionally critical genes are intensively regulated by miRNAs and prefer long 3'UTR.

housekeeping genes, however conserved, are selected to have shorter 3'UTRs to avoid miRNA regulation.

[Cheng et al., BMC Genomics, 2009 (in press)]
Analogies show it reasonable for more variable part of network to be periphery

• **Computer Networks**
  – Servers in center have much depending on them; thus, can't be frequently updated & patched
  – Servers on periphery often attacked and so need frequent patches

• **Social Networks**
  – Individuals at center under more constraint (to conform), whereas those at periphery have more freedom to experiment
Speculation: Why more tightly regulated gene might have less variation

Example: MAP Kinase signaling pathway

- ODE model with Michaelis-Menten kinetics
- parameters fit to time series data of protein activities in response to EGF and NGF from rat PC12 cell line

In sensitivity analysis, stiff parameters cluster around Ras and Raf.

Population study in fruit flies:
- allele variation based on PCR of pathway genes

Ras and Raf have less allele variation than other proteins in the network.

Outline: Molecular Networks

• Why Networks?

• Central Points in Networks
  ◊ Hubs & Bottlenecks
    (yeast ppi & reg. net)
  ◊ Tops of Hierarchies
    (yeast reg. net)
  ◊ Identified by score
    (human miRNA-targ. net)

• Dynamics of Networks
  ◊ Across environments
    (in prokaryote metab. pathways)

• Protein Networks & Variation
  (human ppi & miRNA-targ. net)
Conclusions:
Analysis of Network Structure

- Centrality Measures in Protein Network
  ◊ Hubs & Bottlenecks
  ◊ Importance of later in regulatory networks

- Regulatory Network Hierarchies
  ◊ Middle managers dominate, sitting at info. flow bottlenecks
  ◊ Paradox of influence and essentiality
  ◊ Topmost proteins sit at center of interaction network
Conclusions: Points of Network Centrality

• RE-score measures degree of (down) regulation of targets vs. non-targets
• Application to miRNA network
• Different RE-score of miRNAs can be used in cancer classification
Conclusions: Networks Dynamics across Environments

- Developed and adapted techniques to connect quantitative features of environment to metabolism.
- Applied to available aquatic datasets, we identified footprints that were predictive of their environment (potentially could be used as biosensor).
- Strong correlation exists between a community's energy conversion strategies and its environmental parameters (e.g. temperature and chlorophyll).
- Suggest that limiting amounts of cofactor can (partially) explain increased import of amino acids in nutrient-limited conditions.
Conclusions: Connecting Networks & Variation

• We find ongoing evolution (positive selection) at the network periphery.
  ◊ This trend is present on two levels:
    • On a sequence level, it can be seen as positive selection of peripheral nodes
    • On a structural level, it can be seen as the pattern of SDs that display significantly higher allele frequencies in non-central genes
  ◊ 2 possible mechanisms for this: adaptive evolution at cellular periphery & relaxation of structural constraints at the network periphery
    • We show that the latter can only explain part of the increased variability
Conclusions: Connecting Networks & Variation 2

- More highly regulated genes are under more constraint in miRNA-target networks
- Exception for housekeeping genes
- Speculation as to why variation at periphery is quite reasonable
TopNet – an automated web tool

- Normal website + Downloaded code (JAVA)
- Web service (SOAP) with Cytoscape plugin

[Yu et al., NAR (2004); Yip et al. Bioinfo. (2006);
Similar tools include Cytoscape.org, Idekar, Sander et al]

Yale Network Analyzer

Topology of Networks
Acknowledgements

H Yu
P Kim
K Yip
T Gianoulis
C Cheng

A Paccanaro
P Alves
T Emonet
P Cayting
M Seringhaus
Y Xia
J Korbel
A Sboner
P Patel
P Bork
J Raes
E Franzosa
M Snyder
N Bhardwaj
R Alexander

Networks.GersteinLab.org
Job opportunities currently for postdocs & students
More Information on this Talk

**TITLE:** Understanding Protein Function on a Genome-scale through the Analysis of Molecular Networks

**SUBJECT:** Networks

**DESCRIPTION:**
Network Biology: Understanding metabolic and protein interactions, VIB workshop on the future of proteome research, Ghent, Belgium; 2009.10.08, 9:30-10:10; [I:VIB] (Medium networks talk, shortened from [I:MBINETS].)

(PPT works on mac & PC and has many photos. Paper references in the talk were mostly from Papers.GersteinLab.org. The above topic list can be easily cross-referenced against this website. Each topic abbrev. which is starred is actually a papers “ID” on the site. For instance, the topic **pubnet** can be looked up at [http://papers.gersteinlab.org/papers/pubnet](http://papers.gersteinlab.org/papers/pubnet)

**PERMISSIONS:** This Presentation is copyright Mark Gerstein, Yale University, 2008. Please read permissions statement at [http://www.gersteinlab.org/misc/permissions.html](http://www.gersteinlab.org/misc/permissions.html). Feel free to use images in the talk with PROPER acknowledgement (via citation to relevant papers or link to gersteinlab.org).

**PHOTOS & IMAGES:** For thoughts on the source and permissions of many of the photos and clipped images in this presentation see [http://streams.gerstein.info](http://streams.gerstein.info). In particular, many of the images have particular EXIF tags, such as **kwpotppt**, that can be easily queried from flickr, viz: [http://www.flickr.com/photos/mbgmbg/tags/kwpotppt](http://www.flickr.com/photos/mbgmbg/tags/kwpotppt).