

# A SYSTEMS STUDY OF AGING AND AGE-RELATED PATHOLOGIES

A. Grama   S. Mohammadi

Department of Computer Science  
Purdue University

**CSol Big Data Workshop, 2013**

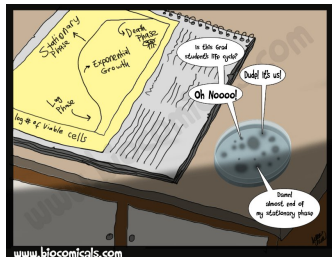
# OUTLINE

- 1 PHASE 1: CONSTRUCTING A MAP OF AGING PATHWAYS IN YEAST
  - Overview
  - Materials and Methods
    - Datasets
    - Tracing Information Flow
  - Results and Discussion
- 2 PHASE 2: PROJECTING YEAST AGING PATHWAYS TO HUMAN TISSUES
  - Motivation
  - Datasets
  - Results

# OUTLINE

- 1 PHASE 1: CONSTRUCTING A MAP OF AGING PATHWAYS IN YEAST
  - Overview
  - Materials and Methods
    - Datasets
    - Tracing Information Flow
  - Results and Discussion
- 2 PHASE 2: PROJECTING YEAST AGING PATHWAYS TO HUMAN TISSUES
  - Motivation
  - Datasets
  - Results

# YEAST AGING



Courtesy of Alper Uzan, PhD.

- Yeast as a model organism for aging research:
  - ✓ Rapid growth
  - ✓ Ease of manipulation
- **Replicative life-span (RLS):** the number of buds a mother cell can produce before senescence occurs
- **Chronological life-span (CLS):** duration of viability after entering the stationary-phase

# OUTLINE

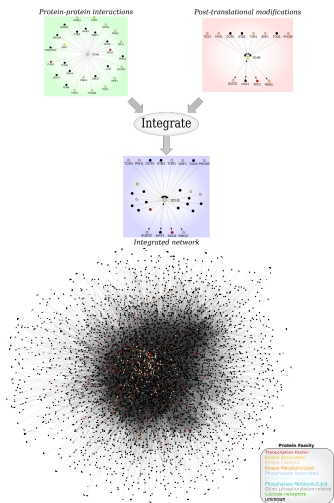
## 1 PHASE 1: CONSTRUCTING A MAP OF AGING PATHWAYS IN YEAST

- Overview
- Materials and Methods
  - Datasets
  - Tracing Information Flow
- Results and Discussion

## 2 PHASE 2: PROJECTING YEAST AGING PATHWAYS TO HUMAN TISSUES

- Motivation
- Datasets
- Results

# YEAST INTERACTOME

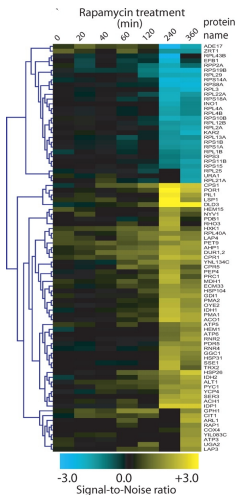


- **Mixed network:** Contains both directed (biochemical activities) and undirected (protein-protein interactions) edges
- 103,619 (63,395 non-redundant) physical interactions among 5,691 proteins.
- 5,791 (5,443 non-redundant) biochemical activities (mostly phosphorylation events) among 2,002 kinase-substrate pairs.

# TRANSCRIPTIONAL REGULATORY NETWORK (TRN) OF YEAST

- Directed graph
- Downloaded from the Yeast Search for Transcriptional Regulators And Consensus Tracking (YEASTRACT)
- Consists of 48,082 interactions between 183 transcription factors (TF) and 6,403 target genes (TG).

# RAPAMYCIN-TREATMENT DATASET



Adopted from Fournier et al.,  
2010

- **Rapamycin:** A lipophilic macrolide that directly binds to and inhibits TOR *in vivo*
- Temporal analysis of gene expression changes for 6,000 ORFs in Baker's yeast, *Saccharomyces cerevisiae*, over 6h of rapamycin treatment.
- 366 repressed and 291 induced genes at a minimum threshold of 2-fold change.



# DIRECTIONAL INFORMATION FLOW

## RANDOM WALK

### DEFINITION

**Random walk** on a graph  $G$ , initiated from vertex  $v$ , is the sequence of transitions among vertices, starting from  $v$ . At each step, the random walker randomly chooses the next vertex from among the neighbors of the current node.

It is a Markov chain with the transition matrix  $P$ , where  $p_{ij} = \text{Prob}(S_{n+1} = v_i | S_n = v_j)$  and random variable  $S_n$  represents the state of the random walk at the time step  $n$ .

# DIRECTIONAL INFORMATION FLOW

## RANDOM WALK WITH RESTART

### DEFINITION

**Random walk with restart (RWR)** is a modified Markov chain in which, at each step, a random walker has the choice of either continuing along its path, with probability  $\alpha$ , or jump (teleport) back to the initial vertex, with probability  $1 - \alpha$ .

The transition matrix of the modified chain,  $M$ , can be computed as  $M = \alpha P + (1 - \alpha) \mathbf{e}_v \mathbf{1}^T$ , where  $\mathbf{e}_v$  is a stochastic vector of size  $n$  having zeros everywhere, except at index  $v$ , and  $\mathbf{1}$  is a vector of all ones.

# DIRECTIONAL INFORMATION FLOW

## STATIONARY DISTRIBUTION

The portion of time spent on each node in an infinite random walk with restart initiated at node  $v$ , with parameter  $\alpha$ .

### DEFINITION

**Stationary distribution** of the modified chain

$$\begin{aligned}\pi_v(\alpha) &= M\pi_v(\alpha) \\ &= (\alpha P + (1 - \alpha)\mathbf{e}_v\mathbf{1}^T)\pi_v(\alpha)\end{aligned}$$

Enforcing a unit norm on the dominant eigenvector to ensure its stochastic property,  $\|\pi_v(\alpha)\|_1 = \mathbf{1}^T\pi_v = 1$ , we will have:

# DIRECTIONAL INFORMATION FLOW

## STATIONARY DISTRIBUTION—CONTINUE

### DEFINITION

**Iterative form** of the information flow process:

$$\pi_v(\alpha) = \alpha P \pi_v(\alpha) + (1 - \alpha) \mathbf{e}_v,$$

### DEFINITION

**Explicit (direct) formulation** of the information flow process:

$$\pi_v(\alpha) = \underbrace{(1 - \alpha)(I - \alpha P)^{-1}}_Q \mathbf{e}_v,$$

# DIRECTIONAL INFORMATION FLOW

## INTERPRETATION

### DEFINITION

Expansion using the Neumann series:

$$\pi_v(\alpha) = (1 - \alpha) \sum_{i=0}^{\infty} (\alpha P)^i \mathbf{e}_v$$

Thus,  $\pi_v(\alpha)$  is a function of:

- Distance to source node ( $v$ )
- Multiplicity of paths

# OUTLINE

## 1 PHASE 1: CONSTRUCTING A MAP OF AGING PATHWAYS IN YEAST

- Overview
- Materials and Methods
  - Datasets
  - Tracing Information Flow
- Results and Discussion

## 2 PHASE 2: PROJECTING YEAST AGING PATHWAYS TO HUMAN TISSUES

- Motivation
- Datasets
- Results

## EXPERIMENTAL SETTINGS

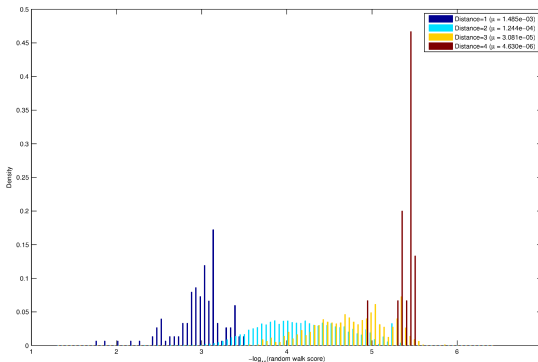
We set the **preference vector** as:

$$e_S(i) = \begin{cases} \frac{1}{|S|} & \text{if } v_i \in S, \\ 0 & \text{O.W.} \end{cases}$$

for  $S$  being the subset of vertices in the yeast interactome corresponding to members of the TORC1 protein complex. The diameter of the network is computed to be 6 and  **$\alpha$  parameter** is set to  $\frac{d}{d+1} = \frac{6}{7} \sim 0.85$  accordingly to give all nodes a fair chance of being visited.

# DISTRIBUTION OF INFORMATION FLOW SCORES

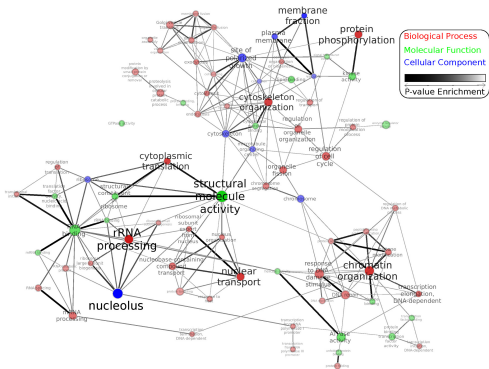
Distribution of information flow scores across nodes with similar distance from members of TORC1 are color coded accordingly. The  $\mu$  parameter is the average of information flow scores for nodes under each distribution.





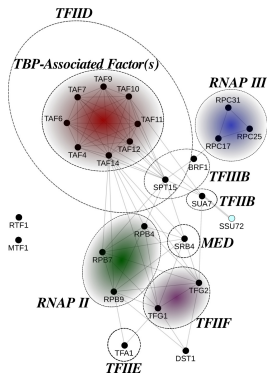
# ENRICHMENT MAP OF YEAST GOSLIM TERMS

Enriched terms are identified by mHG p-value, computed for the ranked-list of genes based on their information flow scores. Each node represents a significant GO term and edges represent the overlap between genesets of GO terms. Terms in different branches of GO are color-coded with red, green, and blue. Color intensity of each node represents the significance of its p-value, while the node size illustrates the size of its geneset. Thickness of edges is related to the extent of overlap among genesets.



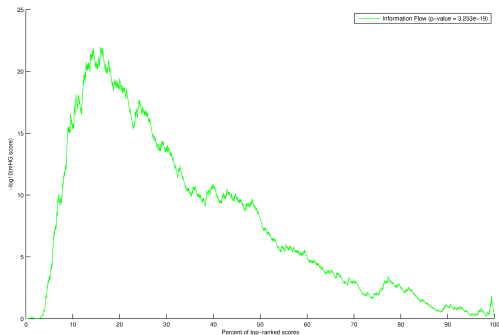
# TOR-DEPENDENT CONTROL OF TRANSCRIPTION INITIATION

Induced subgraph in the yeast interactome, constructed from the top-ranked genes in the information flow analysis that are annotated with the transcription initiation GO term. Different functional subunits are marked and color-coded appropriately.



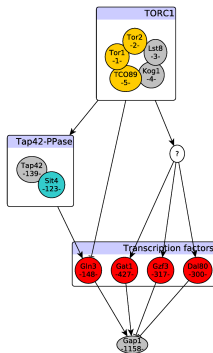
# ENRICHMENT PLOT FOR RAPAMYCIN-TREATMENT DATASET

Enrichment score as a function of the score percentage. Computations are based on the set of differentially expressed genes in response to Rapamycin treatment. The peak of plot occurs at around top 15% of scores, resulting in the minimum hypergeometric (mHG) score of  $\sim 1e - 22$ . The exact p-value for this score is computed, using dynamic programming, to be  $3.3e - 19$ .



# TORC1-DEPENDENT REGULATION OF GAP1

The schematic diagram is based on literature evidence for the interactions. Each node in the signaling pathway is annotated with the rank of its information flow score from TORC1. Ranking of nodes based on their information flow scores respect our prior knowledge on the structure of this pathway.



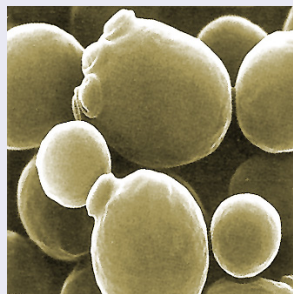
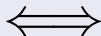
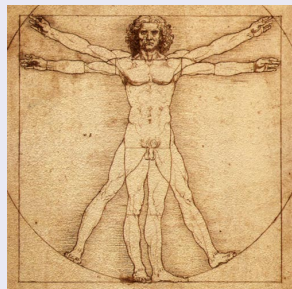
# OUTLINE

- 1 PHASE 1: CONSTRUCTING A MAP OF AGING PATHWAYS IN YEAST
  - Overview
  - Materials and Methods
    - Datasets
    - Tracing Information Flow
  - Results and Discussion
- 2 PHASE 2: PROJECTING YEAST AGING PATHWAYS TO HUMAN TISSUES
  - Motivation
  - Datasets
  - Results

# COMPARATIVE NETWORK ANALYSIS

## TRADITIONAL APPROACH

### BASIC IDEA

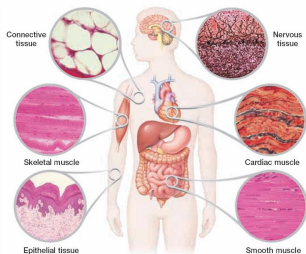


To project functional pathways from a well-studied organism, such as yeast, back to a higher-order organism, such as humans.

# NEW OUTLOOK

## TISSUE-SPECIFIC ANALYSIS

Different human cell types inherit a similar genetic code, but exhibit unique characteristics and functions.



- How does regulation of different genes contribute to functional differences in human tissues?
- How does a uni-cellular organism, such as yeast, contribute to the biological understanding of higher-order, multi-cellular organisms, such as humans?

# OUTLINE

- 1 PHASE 1: CONSTRUCTING A MAP OF AGING PATHWAYS IN YEAST
  - Overview
  - Materials and Methods
    - Datasets
    - Tracing Information Flow
  - Results and Discussion
- 2 PHASE 2: PROJECTING YEAST AGING PATHWAYS TO HUMAN TISSUES
  - Motivation
  - Datasets
  - Results



# TISSUE-SPECIFIC GENE EXPRESSION

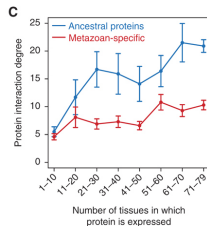
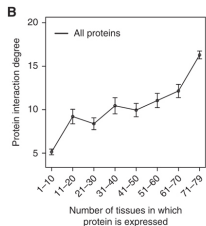
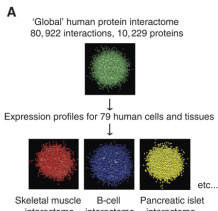
The GNF Gene Atlas dataset:

The screenshot shows the BioGPS website. At the top, there is a navigation bar with 'BioGPS' and a tagline: 'A free extensible and customizable gene annotation portal, a complete resource for learning about gene and protein function.' Below this, there are two main sections: 'Simple to use' and 'News and Musings'. The 'Simple to use' section includes a search bar and a list of steps: 1. Search, 2. View, 3. Browse, 4. Build. The 'News and Musings' section contains a news item about the ISMB 2012 recap and a data chart plugin version 2.0. There is also a 'User Love' section with a quote from Robert.

- 79 different tissues
- 44,775 human transcripts
- Platforms:
  1. Affymetrix HG-U133A.
  2. Custom GNF1H array.

# TISSUE-SPECIFIC INTERACTOMES

- Vertex-induced sub-graphs of the human interactome
- Based on the GNF Gene Atlas dataset
  - ⇒ A gene is considered as present in a tissue, if its normalized expression level is  $> 200$  (average difference between match-mismatch pairs).



Adopted from Bossi et al., 2009

## SEQUENCE SIMILARITY OF PROTEIN PAIRS

- Protein sequences are downloaded from Ensembl database, release 69.
- Reference genomes:
  - ▷ **Human:** GRCh37
  - ▷ **Yeast:** EF4
- Number of sequences:
  - ▷ **Human:** 101,075
  - ▷ **Yeast:** 6,692
- Low-complexity regions are masked using **pseg**
- Smith-Waterman algorithm is used to compute local alignments.

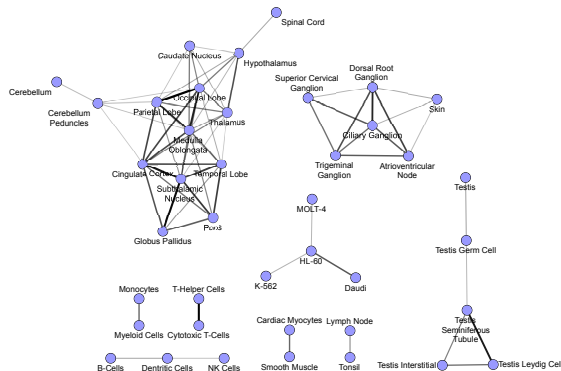
# OUTLINE

- 1 PHASE 1: CONSTRUCTING A MAP OF AGING PATHWAYS IN YEAST
  - Overview
  - Materials and Methods
    - Datasets
    - Tracing Information Flow
  - Results and Discussion
- 2 PHASE 2: PROJECTING YEAST AGING PATHWAYS TO HUMAN TISSUES
  - Motivation
  - Datasets
  - Results



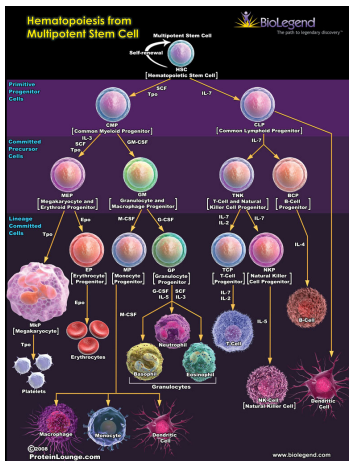
# TISSUE-TISSUE SIMILARITY NETWORK (TTSN)

- First of its kind
- Based on similarity of expression signatures
- Differentiates between similar and dissimilar tissues

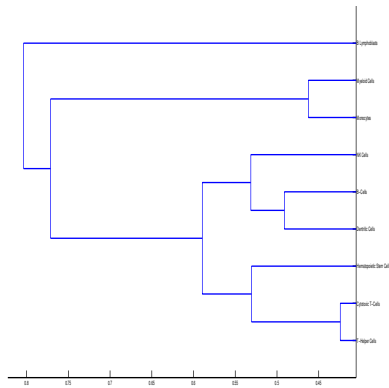


*TTSN with stringent threshold (1.96 < Z-score)*

# RECONSTRUCTING THE DIFFERENTIATION TREE OF IMMUNE CELLS



Adopted from BioLegend



# SIMILARITY OF DIFFERENT TISSUES TO YEAST

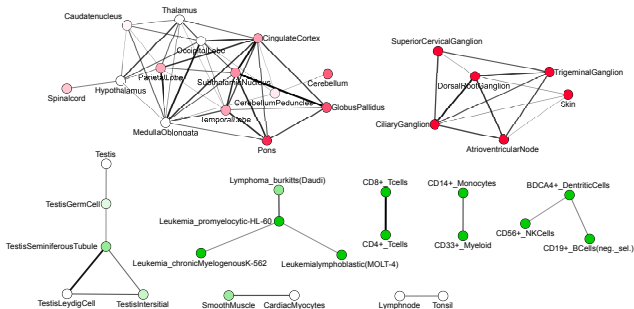
## SUMMARY OF METHODS

- Align each tissue-specific network to yeast using *Belief Propagation (BP)* algorithm.
- For each alignment, compute:
  - Number of conserved edges.
  - Sequence similarity of the aligned proteins.
- Create an ensemble of random, *pseudo tissues* seeded around housekeeping proteins (proteins that are expressed in all tissues).
- Align random tissues with yeast and compute p-values of original alignment.



# SIMILARITY OF DIFFERENT TISSUES TO YEAST

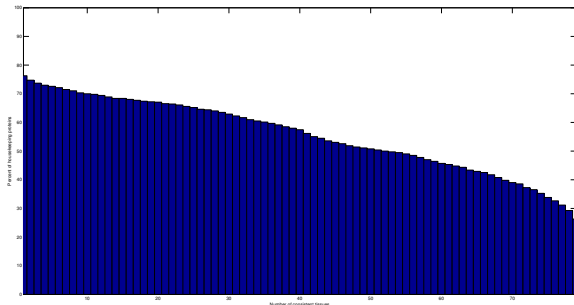
## PRELIMINARY P-VALUE RESULTS



- Green nodes show tissues with significant similarity to yeast, while red nodes show dissimilar tissues.
- Similar tissues tend to have consistent p-values.

# SIMILARITY OF DIFFERENT TISSUES TO YEAST

## ALIGNMENT CONSISTENCY OF HOUSEKEEPING PROTEINS



- Approximately 75% of housekeeping proteins are aligned with yeast proteins.
- Among aligned HK proteins, 25% of them consistently aligned across all 79 tissues.

## SUMMARY

- Aging is the primary risk factor for a number of human diseases.
- Emerging evidence supports the hypothesis that large classes of age-related pathologies share their underlying biology
- Constructing a comprehensive map of aging pathways is a critical step towards deciphering key lifespan mediators, their crosstalk, and systems-level organization.
- Tissue-specificity analysis is needed to precisely model aging in both proliferating and post-mitotic cells.
  
- Future works:
  - Construction of comprehensive, yet reliable tissue-specific networks by integrating various available high-throughput datasets.
  - Devising a *biased* information flow method for targeting specific subsets of TOR effectors.

# FOR FURTHER READING I



SC. Johnson et al.

mTOR is a key modulator of ageing and age-related disease

*Nature*, 493(7432):338–45, 2013.



RM. Naylor et al.

Senescent cells: a novel therapeutic target for aging and age-related diseases

*clinical pharmacology and therapeutics*, 93:105–116, 2013.



J Campisi et al.

Cellular senescence: a link between cancer and age-related degenerative disease?

*Seminars in cancer biology*, 21: 354–359, 2012.