A SYSTEMS APPROACH TO DISSECTING THE TISSUE-SPECIFIC ARCHITECTURE OF CELLULAR NETWORKS

S. Mohammadi, B. Saberidokht, S. Subramaniam, A. Grama

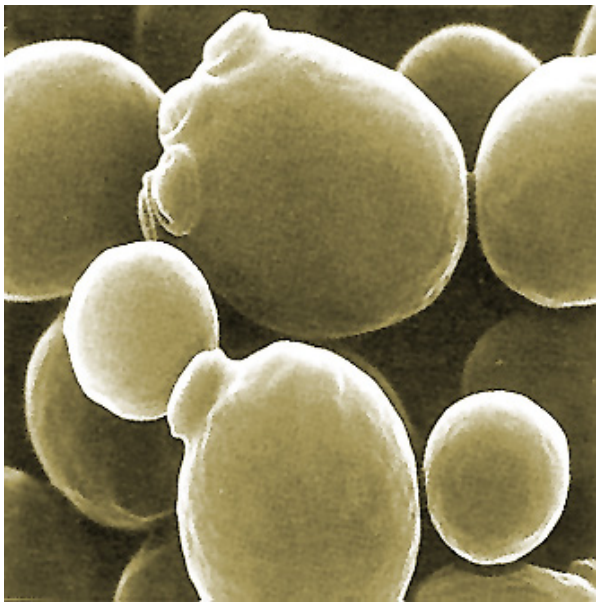
Department of Computer Science, Purdue University
Department of Bioengineering, University of California, San Diego

Outline

1. Motivation
2. Materials and methods
   - Datasets
   - Algorithmic contributions
3. Results and discussion
   - Core alignment graph of housekeeping genes
   - Computing similarity of human tissues with yeast
   - Consistency of alignment p-values
   - Partitioning tissue-selective genes
   - Conservation and functional role of tissue selective genes
   - Tissue-selective genes predict tissue-specific pathologies

Yeast as a pioneering model organism

Simple yet powerful



"... yeast has graduated from a position as the premier model for eukaryotic cell biology to become the pioneer organism that facilitated the establishment of entirely new fields of study called *functional genomics* and *systems biology*." – D. Botstein and G. Fink (2011).

Yeast as a pioneering model organism

Why yeast?

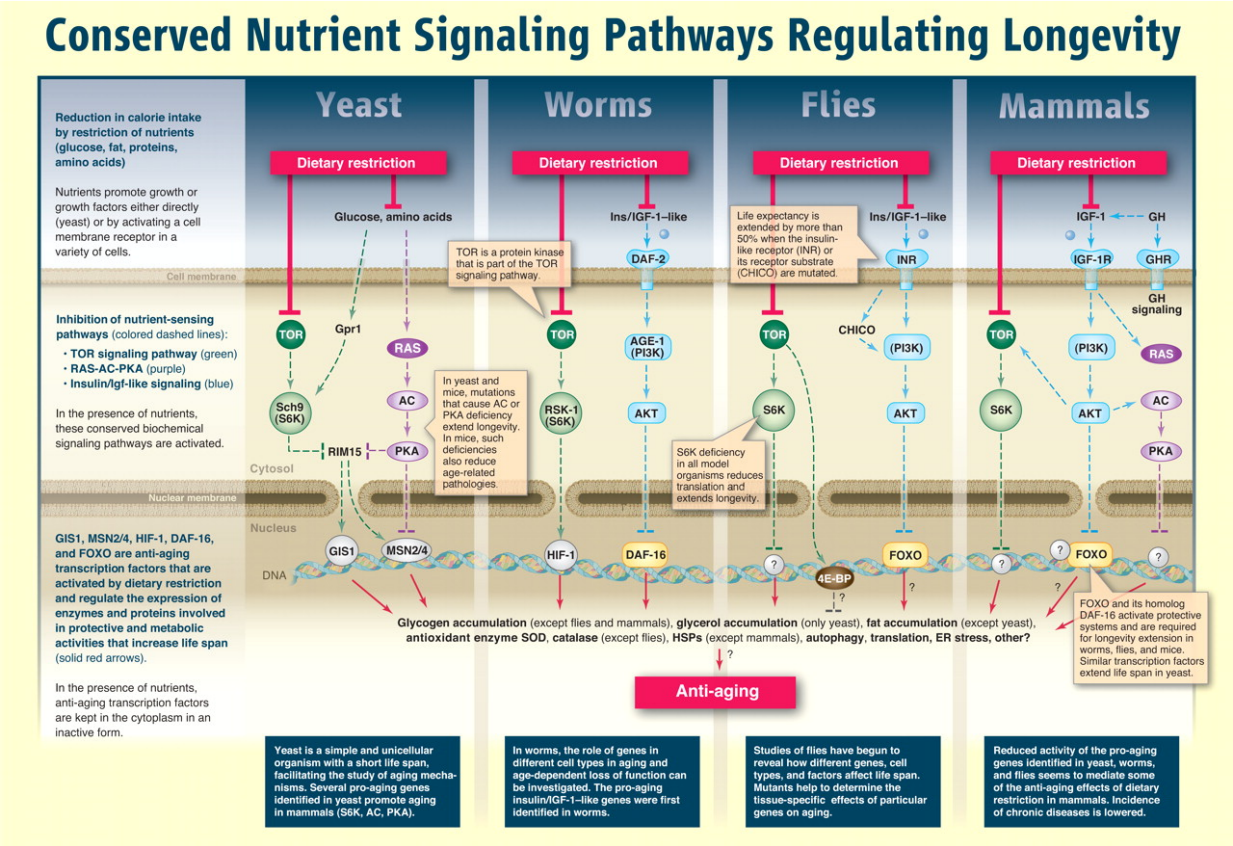
- Rapid growth and ease of manipulation
- Mature genetic and molecular toolbox, including deletion mutants, over-expression libraries, and green fluorescent protein (GFP)-tagged yeast strains
- Multitude of high-throughput datasets, ranging from genetic arrays, transcriptome, proteome, and metabolome profiles
- Saccharomyces Genome Database (SGD)

Conserved pathways between yeast and higher-order organisms

Many of the underlying functionalities and associated machineries are shared with higher eukaryotes:

- Cell cycle
- Programmed cell death
- Protein folding, quality control, and degradation
- Signaling pathways, such as MAPK, TOR, and insulin/IGF-I
- Aging and CR-mediated pathways
  - **Chronological:** amount of time cells survive in post-mitotic state
  - **Replicative:** number of times a cell can divide before senescence occurs.

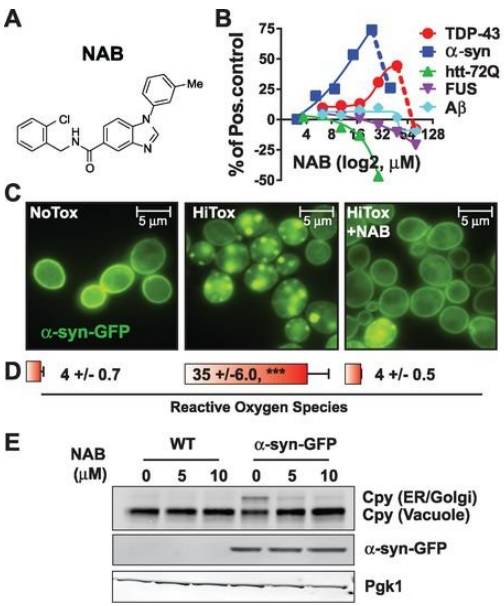
Conserved pathways
Continued



Fontana *et al*, Science (2010)

YEAST AS A MODEL ORGANISM FOR HUMAN DISEASE

RECENT SUCCESS STORIES

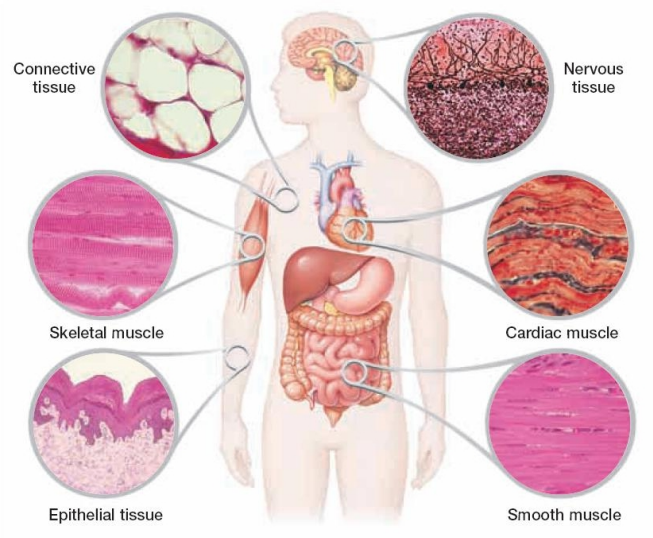


Adopted from D. Tardiff *et al.* (2013)

- Heterologous expression of disease gene(s)
- Yeast as an unbiased phenotypic screen
- N-aryl benzimidazole (NAB) strongly protects cells from α-synuclein toxicity in the humanized yeast model
- Validated this discovery using iPS cell from Parkinson's patients with α-Syn mutation

Problem statement

For which tissues is yeast a good model organism?



Different human tissues, while inheriting a similar genetic code, exhibit unique anatomical and physiological properties.

What are the shared/ missing functional components in yeast, compared to human tissues?

Outline

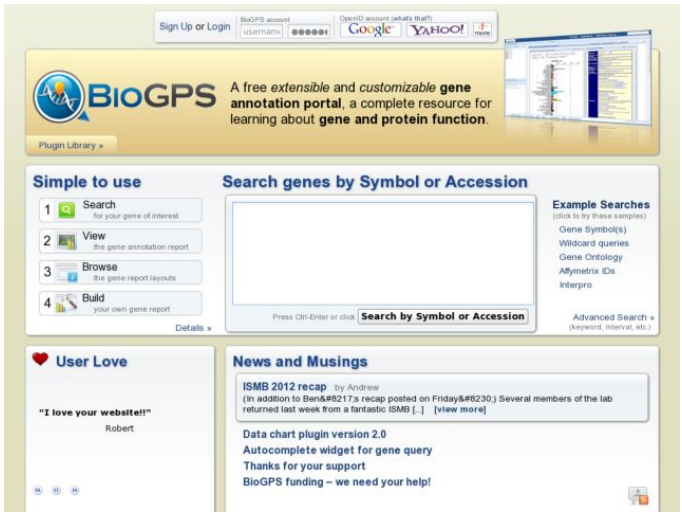
1 Motivation

2 Materials and methods
- Datasets
- Algorithmic contributions

3 Results and discussion
- Core alignment graph of housekeeping genes
- Computing similarity of human tissues with yeast
- Consistency of alignment p-values
- Partitioning tissue-selective genes
- Conservation and functional role of tissue selective genes
- Tissue-selective genes predict tissue-specific pathologies

TISSUE-SPECIFIC GENE EXPRESSION

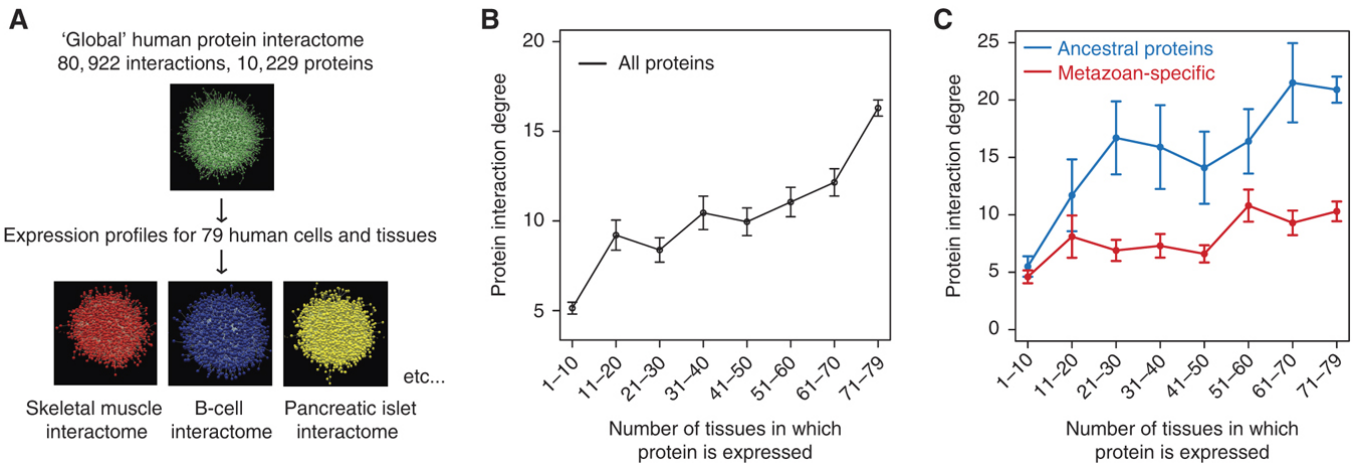
The GNF Gene Atlas dataset:



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

TISSUE-SPECIFIC INTERACTOMES

- Vertex-induced subgraphs of the global human interactome
- Based on the GNF Gene Atlas dataset
  - $\Rightarrow$ 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 protein sequences:
  - **Human**: 101,075
  - **Yeast**: 6,692
- Low-complexity regions are masked using **pseg**
- Smith-Waterman algorithm is used to compute local sequence alignments.

Outline

1 Motivation

2 Materials and methods
- Datasets
- Algorithmic contributions

3 Results and discussion
- Core alignment graph of housekeeping genes
- Computing similarity of human tissues with yeast
- Consistency of alignment p-values
- Partitioning tissue-selective genes
- Conservation and functional role of tissue selective genes
- Tissue-selective genes predict tissue-specific pathologies

SPARSE NETWORK ALIGNMENT

Integer Quadratic Program– Approximated using Belief Propagation:

$$\max_{\boldsymbol{x}} \quad (\alpha \boldsymbol{w}^T \boldsymbol{x} + \frac{\beta}{2} \boldsymbol{x}^{\mathsf{T}} \boldsymbol{S} \boldsymbol{x})$$

$$\text{Subject to:} \quad \begin{cases} \mathcal{C}\boldsymbol{x} \leq \mathbf{1}_{n_G * n_H} & \text{Matching constraints;} \\ x_{ii'} \in \{0, 1\}, & \text{Integer constraint.} \end{cases}$$

- $\boldsymbol{x}$: Matching vector
- L: Bipartite graph of similarities between pair of proteins in input networks
- $\boldsymbol{w}$: Edge-weights in the graph L (based on sequence similarities)
- $\boldsymbol{S}$: Matrix encoding conserved edges in the product graph ($G \otimes H$)
- $\mathcal{C}$: Incidence matrix of graph L

Random model for tissue-specific networks

Definition

- Global human interactome: All potential interactions between human proteins, represented by graph $G = (V_G, E_G)$
- Tissue-specific network(s): Vertex-induced subgraph(s) of the Global human interactome, represented by $G_T = (V_T, E_T)$ with $n_T = |V_T|$, $V_T \subset V_G$, and $E_T \subset E_G$
- Universal genes: Ubiquitously expressed subset of human genes corresponding to houskeeping functions, represented by $V_U \subset V_G$, and $n_U = |V_U|$
- Random tissue-specific network(s): Vertex-induced subgraphs of G, constructed from $V_{\mathcal{R}} = V_U \cup V_S$, with V_S being random set of vertices of size $n_T - n_U$ selected from $V_G \setminus V_U$

Significance of network alignment(s)

Definition

- Original alignment: $\mathcal{W} = \boldsymbol{w}^T \boldsymbol{x}$, $\mathcal{O} = \frac{1}{2} \boldsymbol{x}^T \boldsymbol{S} \boldsymbol{x}$
- Monte-Carlo simulation: Let $\mathcal{W}_{\mathcal{R}}$ and $\mathcal{O}_{\mathcal{R}}$ be the random vectors representing the weight and overlap of aligning $k_{\mathcal{R}}$ random tissue-specific networks with yeast
- Positive/Negative cases: k_P is the number of random cases with both $\mathcal{W}_{\mathcal{R}} \leq \mathcal{W}$ and . $\mathcal{O}_{\mathcal{R}} \leq \mathcal{O}$. k_N is defined as the size of complement set.
- p-value bounds:

$$\delta_{\mathcal{R}} = \frac{k_P}{k_{\mathcal{R}}} \leq \text{alignment p-value} \leq 1 - \frac{k_N}{k_{\mathcal{R}}} = \Delta_{\mathcal{R}}$$

- Alignment p-value:

$$p - value = Prob(\alpha * \mathcal{O} + \beta * \mathcal{W} \leq \mathcal{O}\mathcal{W}_{\mathcal{R}})$$

Partitioning human genes based on their expression selectivity

Definition

Selectivity *p*-value– Given a cluster of homogenous tissues:

$$\begin{aligned} p\text{-value}(X = c_n) &= Prob(c_n \leq X) \\ &= HGT(c_n|N, n, c_N) \\ &= \sum_{x=c_n}^{min(c_N,n)} \frac{C(c_N, x)C(N - c_N, n - x)}{C(N, n)} \end{aligned}$$

N: total number of tissues, n: number of tissues in the cluster, c_N: number of tissues in which a given gene is expressed, c_n: number of tissue in the cluster that the given gene is expressed.

HUMAN-SPECIFIC OR CONSERVED?

DEFINITION

Classification of human tissue-selective genes:

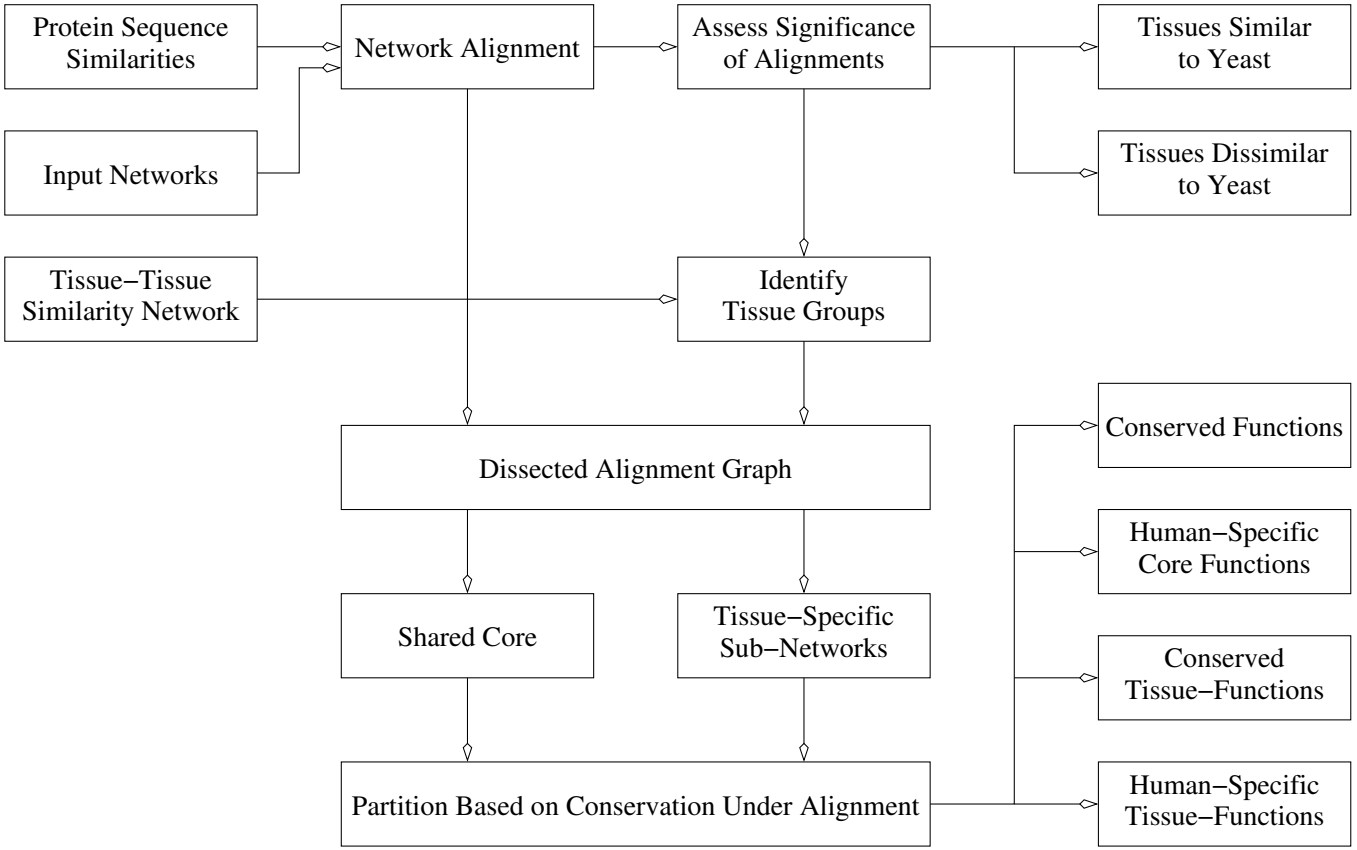
- Conserved: Subset of tissue-selective genes that are consistently aligned in the "majority" of aligned tissues in the given group
- Human-specific: Subset of tissue-selective genes that are consistently unaligned in the "majority" of tissues in the given group
- Unclassified: None of the above

DEFINITION

Majority voting:

- Alignment consistency table: Yeast partner of each tissue-selective gene in the given cluster of tissues
- Consensus rate: Minimum percentage of tissues (columns) in each row of the alignment consistency table that have to agree to make a decision about conserved/human-specificity

SUMMARY


Protein Sequence Similarities
Input Networks
Tissue–Tissue Similarity Network
Network Alignment
Assess Significance of Alignments
Identify Tissue Groups
Dissected Alignment Graph
Shared Core
Tissue–Specific Sub–Networks
Partition Based on Conservation Under Alignment
Tissues Similar to Yeast
Tissues Dissimilar to Yeast
Conserved Functions
Human–Specific Core Functions
Conserved Tissue–Functions
Human–Specific Tissue–Functions
Input
Processing
Output

Outline

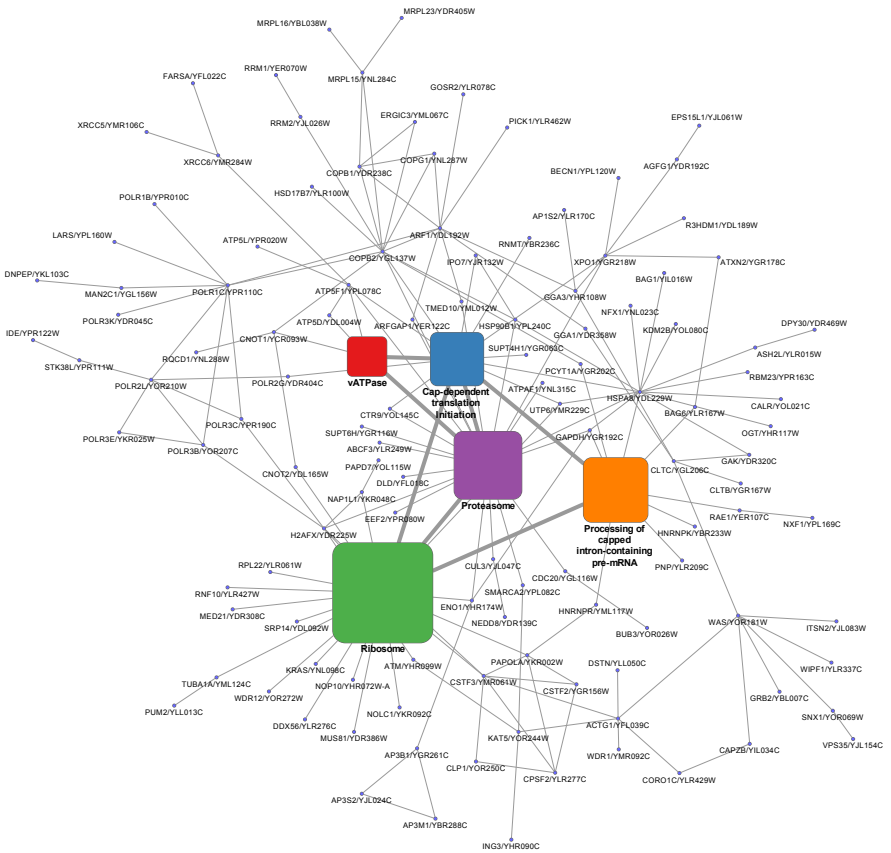
1 Motivation

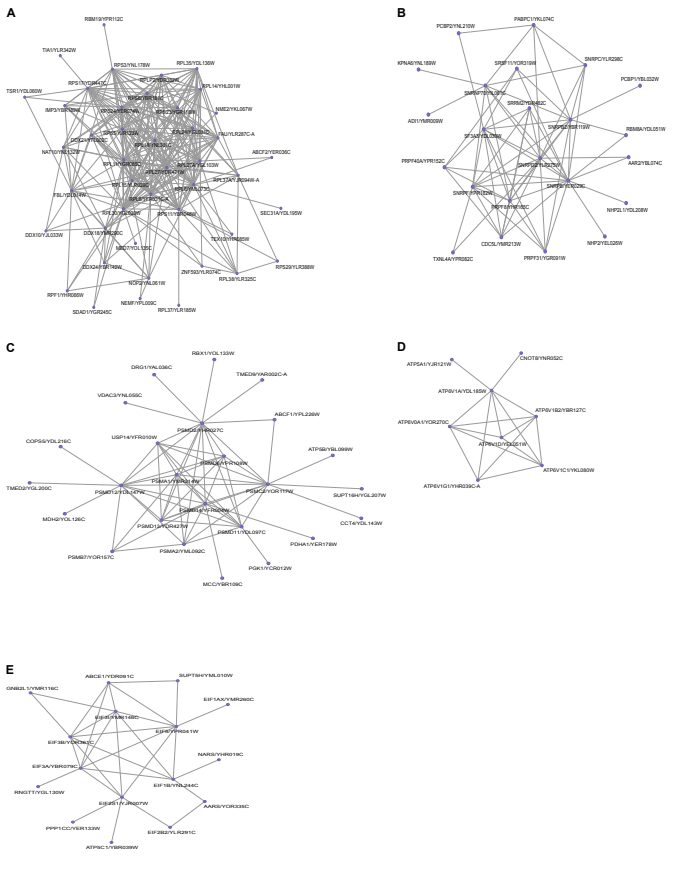
2 Materials and methods
- Datasets
- Algorithmic contributions

3 Results and discussion
- Core alignment graph of housekeeping genes
- Computing similarity of human tissues with yeast
- Consistency of alignment p-values
- Partitioning tissue-selective genes
- Conservation and functional role of tissue selective genes
- Tissue-selective genes predict tissue-specific pathologies

Core alignment graph of housekeeping genes
Computing similarity of human tissues with yeast
Consistency of alignment p-values
Partitioning tissue-selective genes
Conservation and functional role of tissue selective genes
Tissue-selective genes predict tissue-specific pathologies

CORE GENES– THE MOST CONSERVED SUBSET OF HOUSEKEEPING GENES





Functional enrichment of HK genes

Core subset

- Ribosome biogenesis
- Translation
- Protein targeting
- RNA splicing
- mRNA surveillance

Functional enrichment of HK genes

Human-specific subset

- Anatomical structure development
- Paracrine signaling
- NADH dehydrogenase (mitochondrial Complex I)

Outline

1 Motivation

2 Materials and methods
- Datasets
- Algorithmic contributions

3 Results and discussion
- Core alignment graph of housekeeping genes
- Computing similarity of human tissues with yeast
- Consistency of alignment p-values
- Partitioning tissue-selective genes
- Conservation and functional role of tissue selective genes
- Tissue-selective genes predict tissue-specific pathologies

THE MOST SIMILAR TISSUES TO YEAST

Name	pval lower bound	overall pval	pval upper bound	confidence
Myeloid Cells	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
Monocytes	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
Dentritic Cells	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
NK Cells	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
T-Helper Cells	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
Cytotoxic T-Cells	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
B-Cells	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
Endothelial	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
Hematopoietic Stem Cells	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
MOLT-4	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
B Lymphoblasts	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
HL-60	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
K-562	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
Early Erythroid	< 1.00e-04	< 1.00e-04	< 1.00e-04	1
Bronchial Epithelial Cells	< 1.00e-04	< 1.00e-04	0.0002	0.9998
Colorectal Adenocarcinoma	< 1.00e-04	< 1.00e-04	0.0004	0.9996
Daudi	< 1.00e-04	< 1.00e-04	0.0009	0.9991
Testis Seminiferous Tubule	< 1.00e-04	< 1.00e-04	0.0012	0.9988
Smooth Muscle	< 1.00e-04	< 1.00e-04	0.0016	0.9984
Blood (Whole)	< 1.00e-04	< 1.00e-04	0.0053	0.9947
Thymus	< 1.00e-04	0.0001	0.0062	0.9938
Testis Interstitial	< 1.00e-04	0.0004	0.0086	0.9914

THE LEAST SIMILAR TISSUES TO YEAST

Name	pval lower bound	overall pval	pval upper bound	confidence
Trigeminal Ganglion	0.9947	0.9994	1	0.9947
Superior Cervical Ganglion	0.9847	0.9991	1	0.9847
Ciliary Ganglion	0.9407	0.9813	0.9964	0.9443
Atrioventricular Node	0.8746	0.9792	0.9921	0.8825
Skin	0.8355	0.9297	0.9809	0.8546
Heart	0.7934	0.9585	0.9815	0.8119
Appendix	0.7596	0.9371	0.973	0.7866
Dorsal Root Ganglion	0.7065	0.933	0.9717	0.7348
Skeletal Muscle	0.3994	0.5902	0.7866	0.6128
Uterus Corpus	0.233	0.7736	0.8769	0.3561
Lung	0.0771	0.3853	0.5544	0.5227
Pons	0.0674	0.5201	0.6983	0.3691
Salivary Gland	0.0639	0.3449	0.5173	0.5466
Liver	0.0600	0.6857	0.8519	0.2081
Ovary	0.0388	0.2735	0.4481	0.5907
Trachea	0.0259	0.2376	0.4146	0.6113
Globus Pallidus	0.0206	0.2471	0.4336	0.587
Cerebellum	0.0127	0.1950	0.3783	0.6344

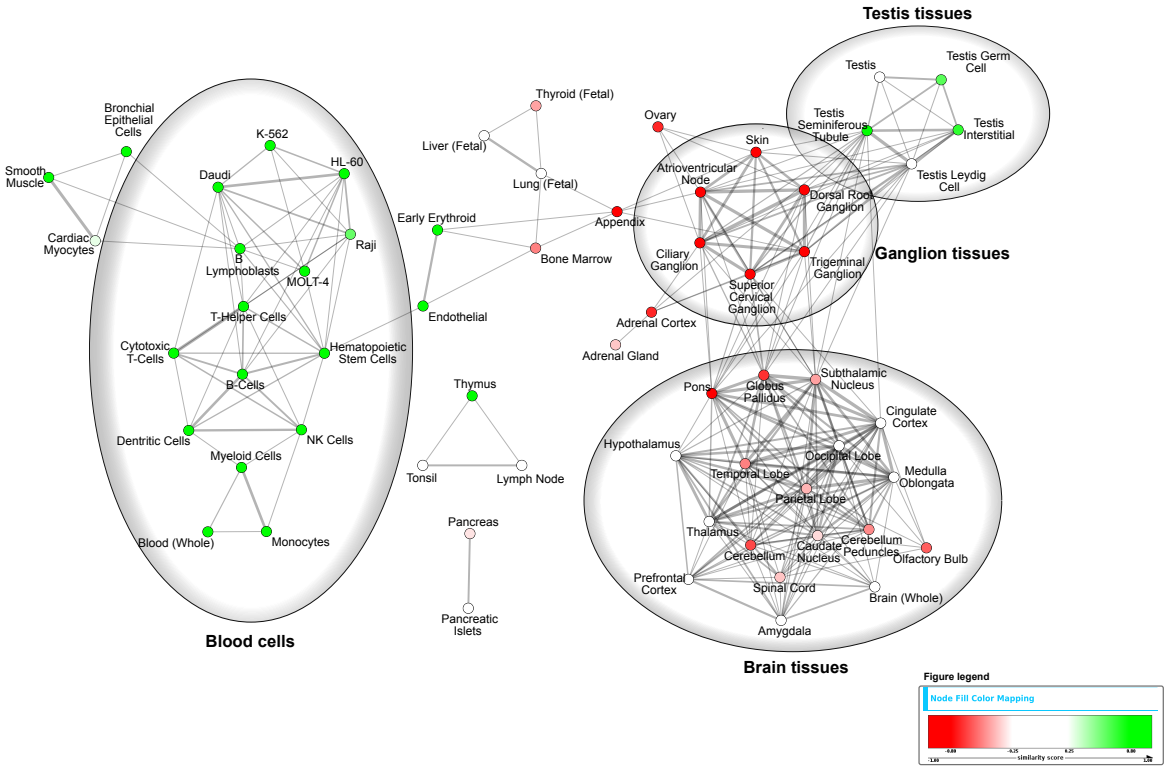
Outline

1 Motivation

2 Materials and methods
- Datasets
- Algorithmic contributions

3 Results and discussion
- Core alignment graph of housekeeping genes
- Computing similarity of human tissues with yeast
- Consistency of alignment p-values
- Partitioning tissue-selective genes
- Conservation and functional role of tissue selective genes
- Tissue-selective genes predict tissue-specific pathologies

TISSUE-TISSUE SIMILARITY NETWORK



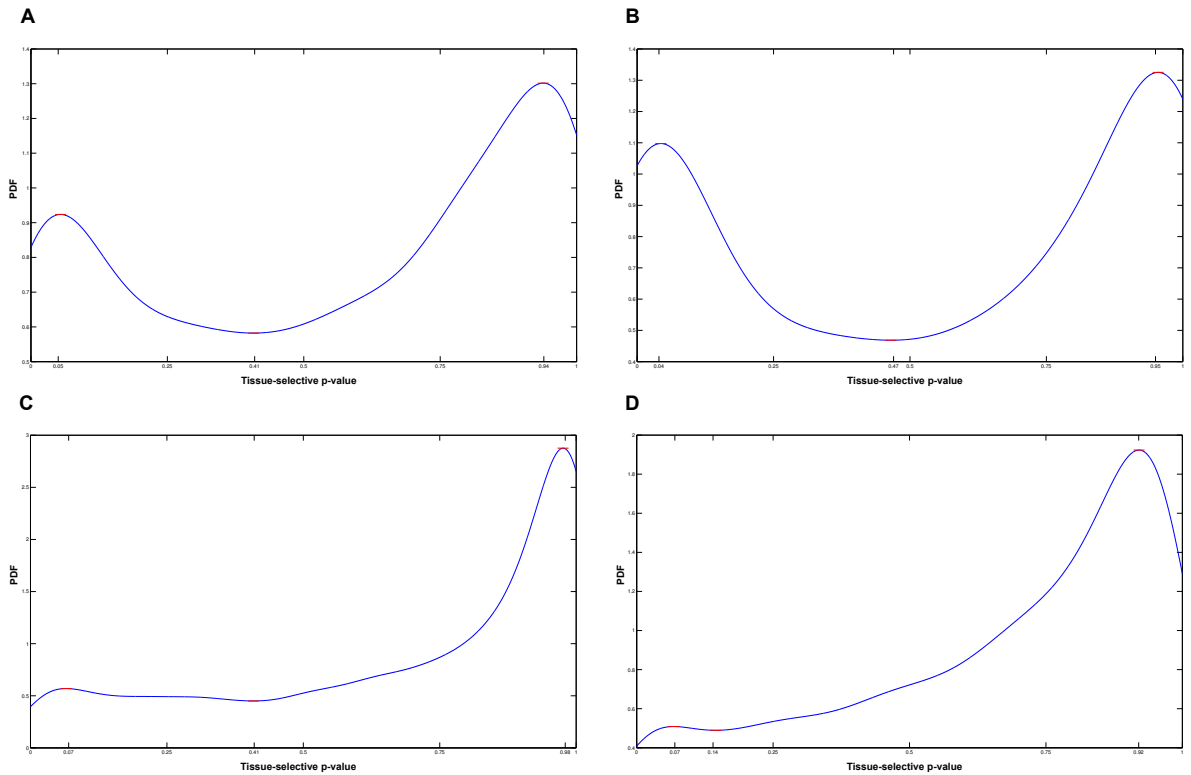
Outline

1 Motivation

2 Materials and methods
- Datasets
- Algorithmic contributions

3 Results and discussion
- Core alignment graph of housekeeping genes
- Computing similarity of human tissues with yeast
- Consistency of alignment p-values
- Partitioning tissue-selective genes
- Conservation and functional role of tissue selective genes
- Tissue-selective genes predict tissue-specific pathologies

TISSUE-SELECTIVITY DENSITY FUNCTION

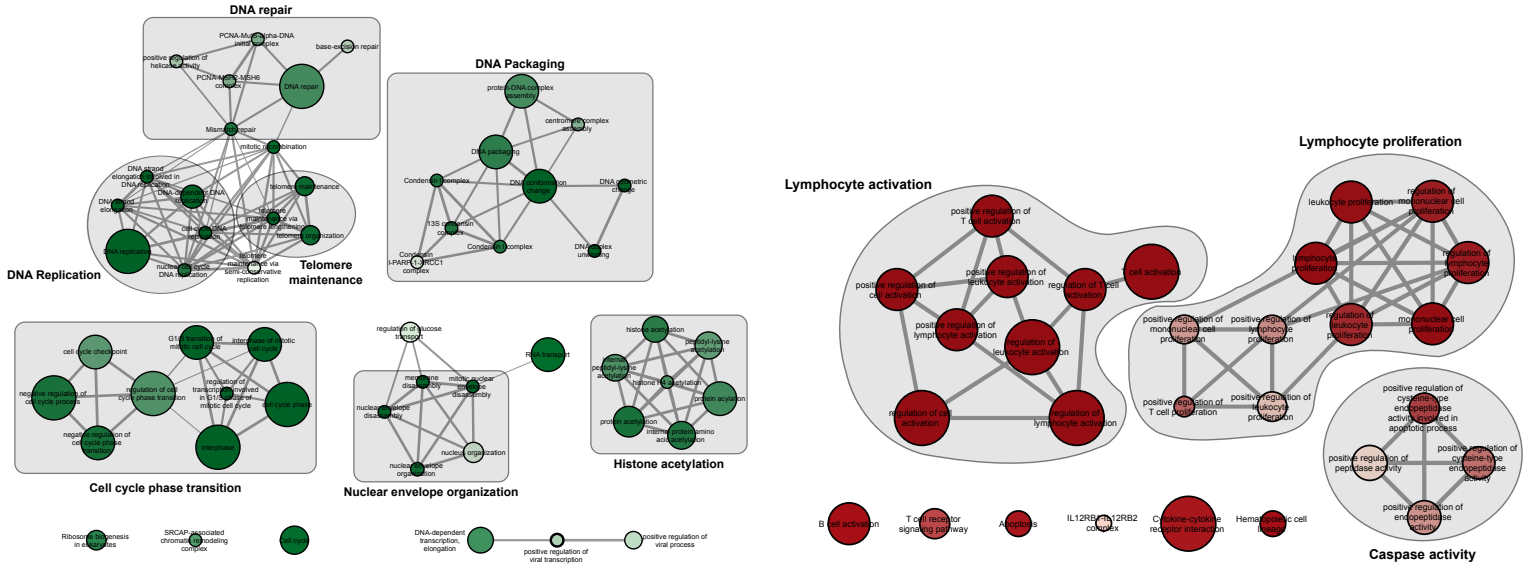


(**A**) Brain tissues, (**B**) Blood cells, (**C**) Ganglion tissues, (**D**) Testis tissues

Outline

1. Motivation
2. Materials and methods
   - Datasets
   - Algorithmic contributions
3. Results and discussion
   - Core alignment graph of housekeeping genes
   - Computing similarity of human tissues with yeast
   - Consistency of alignment p-values
   - Partitioning tissue-selective genes
   - Conservation and functional role of tissue selective genes
   - Tissue-selective genes predict tissue-specific pathologies

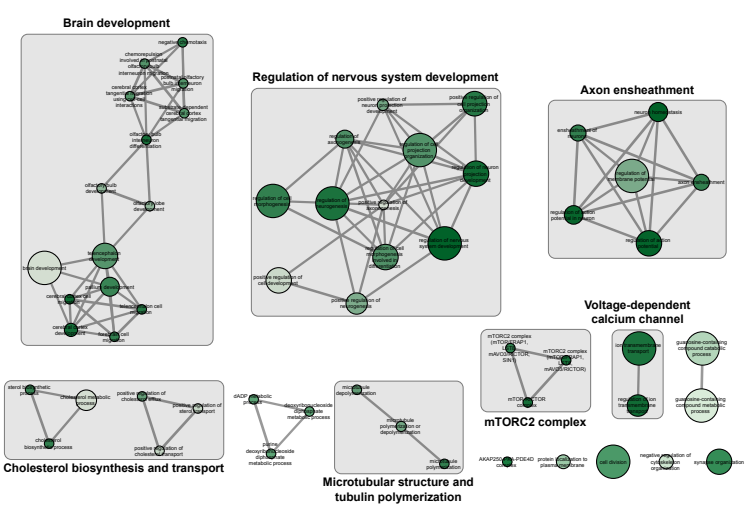
BLOOD CELLS



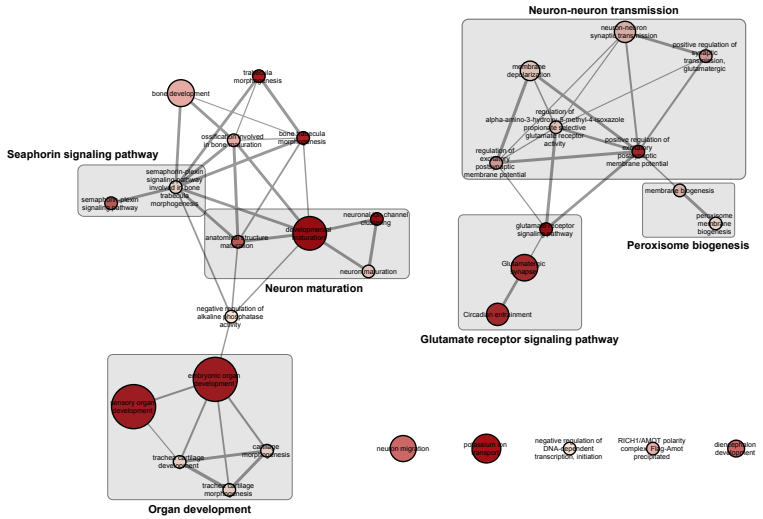
(a) Conserved (b) Human-specific

FIGURE : **Enrichment map of unique blood-selective functions.**

BRAIN TISSUES



(a) Conserved



(b) Human-specific

FIGURE : **Enrichment map of unique brain-selective functions.**

Outline

1 Motivation

2 Materials and methods
- Datasets
- Algorithmic contributions

3 Results and discussion
- Core alignment graph of housekeeping genes
- Computing similarity of human tissues with yeast
- Consistency of alignment p-values
- Partitioning tissue-selective genes
- Conservation and functional role of tissue selective genes
- Tissue-selective genes predict tissue-specific pathologies

Enriched disease classes

	Conserved genes		Human-specific genes	
	Disease class	*p*-value	Disease class	*p*-value
Blood cells	Cancer	$2.85 * 10^{-3}$	Immune	$1.88 * 10^{-9}$
			Infection	$1.00 * 10^{-2}$
Brain tissues	Psych	$3.59 * 10^{-4}$	Psych	$5.70 * 10^{-8}$
	Chemdependency	$2.60 * 10^{-3}$	Neurological	$2.97 * 10^{-2}$
	Pharmacogenomic	$9.74 * 10^{-2}$		

Comparative analysis of brain-specific pathologies

Disorder	Conserved genes	Human-specific genes
schizophrenia	0.008573	8.4905E-06
autism	0.048288	0.00077448
dementia	0.0014356	-
schizophrenia; schizoaffective disorder; bipolar disorder	-	0.0021433
myocardial infarct; cholesterol, HDL; triglycerides; atherosclerosis, coronary; macular degeneration; colorectal cancer	0.0051617	-
epilepsy	0.071562	0.0064716
seizures	-	0.020381
bipolar disorder	0.048288	0.022016
attention deficit disorder conduct disorder oppositional defiant disorder	0.032444	0.023865