

Age in gene regulatory networks

- Do genes of the same age tend to interact more?

Student: **Yuliya Burankova**

Supervisor: **Erik Zhivkopoulos,**
SciLifeLab, Stockholm University



Background

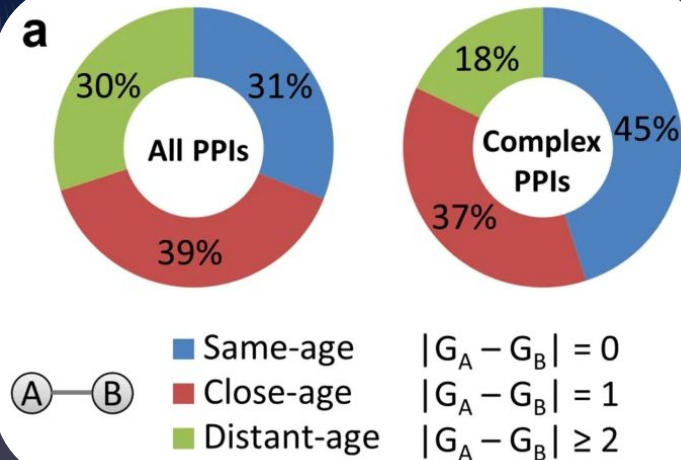
For protein-protein interaction (PPI) networks in yeast and human, it was shown that proteins of the same age tend to interact more ¹.

Goal

The goal of this project is

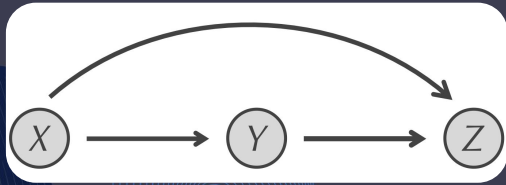
to explore if gene interaction preference for genes with similar age holds in gene regulatory networks (GRNs),

in particular in those that describe direct regulatory interaction (transcription factor-target gene).





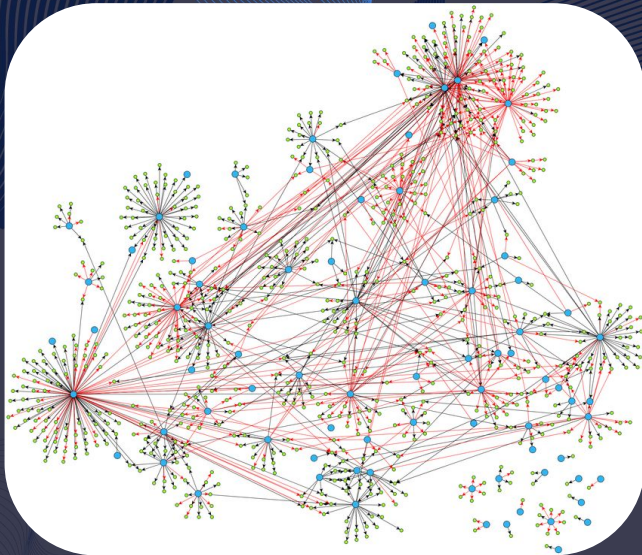
Gene regulatory networks



GRNs are composed of:

- **nodes** are the genes and their regulators
- **edges** are the regulatory relationships between the nodes

It is usually represented mathematically as an **oriented graph**. GRNs models regulatory interactions (indirect, not physical!).



Data (adjacency lists):

- Yeast GRN — full transcriptional regulatory network ^{2,3} (Tnet)
- Mouse GRN — manually curated database ⁴ (TRRUST v2)
- Human GRN — manually curated database ⁴ (TRRUST v2)

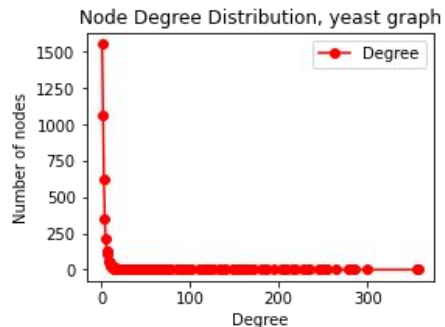
Data contain the list of links between transcription factors (TF) and corresponding target genes (TG). All edges have been experimentally confirmed earlier.



1. GRNs structure

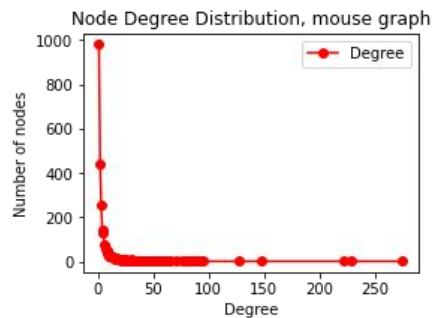
Yeast

- Total genes number: 4 441.
- Interactions number: 12 873.
- Contains only transcription factors and their targets.
- Average number of interaction for nodes: 2.8987.



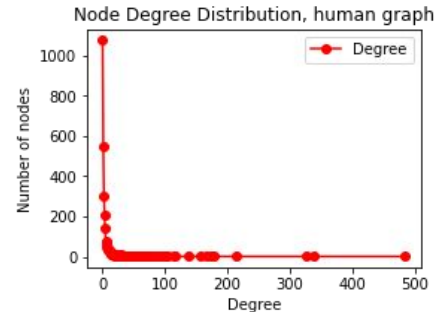
Mouse

- Total genes number: 2 456.
- Interactions number: 7 057.
- Contains transcription factors, their targets and mode of regulation.
- Average number of interaction for nodes: 2.6425.



Human

- Total genes number: 2 862.
- Interactions number: 8 427.
- Contains transcription factors, their targets and mode of regulation.
- Average number of interaction for nodes: 2.9444.



2. Protein into gene age classes

Protein age classes ⁵ (downloaded from [github](#)) were translated into gene age classes using protein-gene name matching from [YeastGenome](#) and [UNIPROT](#) databases.

Yeast

- age classes are available for 5 000 proteins.
- After mapping:

	Age class	Genes number
0	Ascomycota	111
1	Cellular_organisms	372
2	Dikarya	241
3	Euk+Bac	505
4	Euk_Archaea	104
5	Eukaryota	1364
6	Opisthokonta	186
7	Saccharomyceta	554
Total	Total	3437

Mouse

- age classes are available for 19 950 proteins.
- After mapping:

	Age class	Genes number
0	Cellular_organisms	85
1	Euk+Bac	153
2	Euk_Archaea	20
3	Eukaryota	436
4	Eumetazoa	878
5	Mammalia	152
6	Opisthokonta	130
7	Vertebrata	433
Total	Total	2287

Human

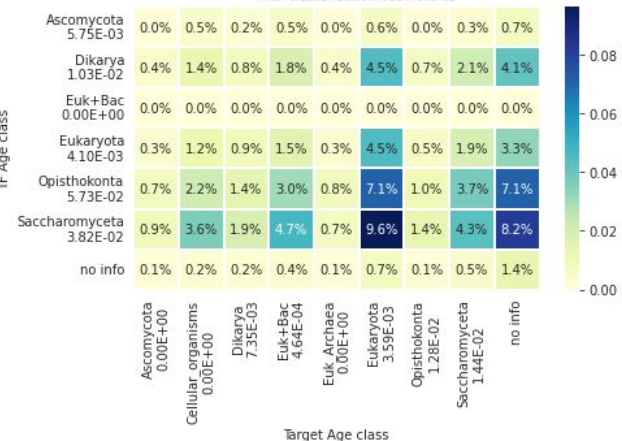
- age classes are available for 20 331 proteins.
- After mapping:

	Age class	Genes number
0	Cellular_organisms	145
1	Euk+Bac	230
2	Euk_Archaea	27
3	Eukaryota	650
4	Eumetazoa	891
5	Mammalia	212
6	Opisthokonta	167
7	Vertebrata	533
Total	Total	2855

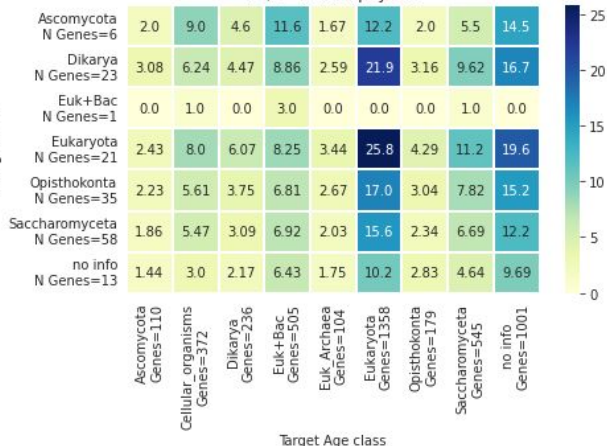


2.1. Results for Yeast GRN

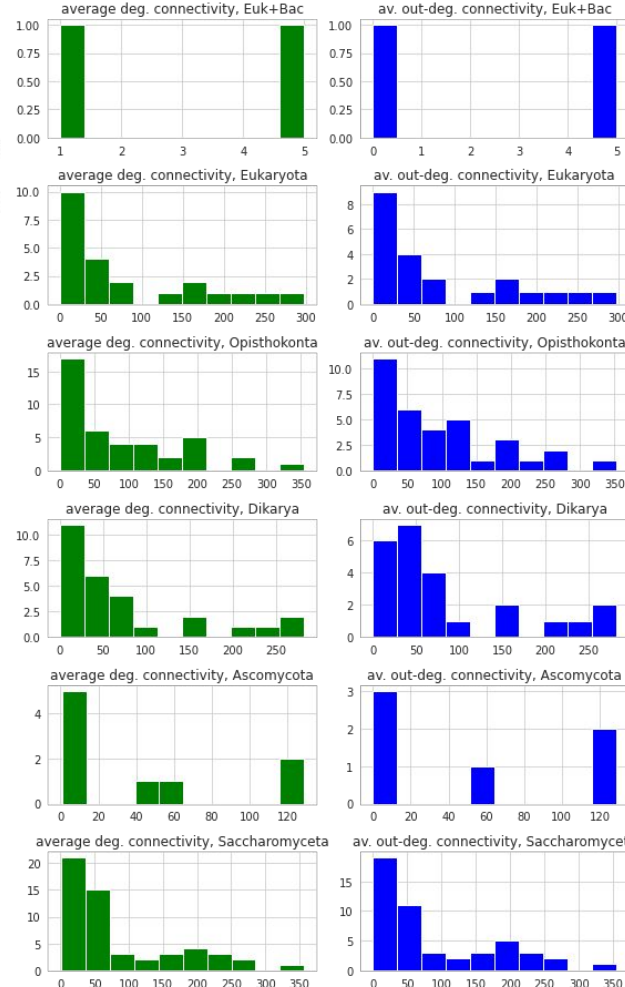
Age classes heatmap, yeast with clusterization coefficients



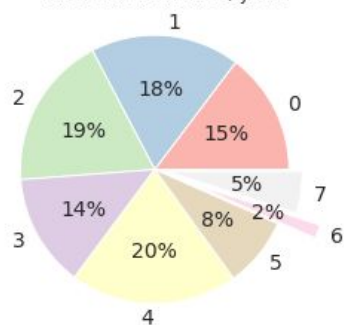
TG/TF heatmap, yeast



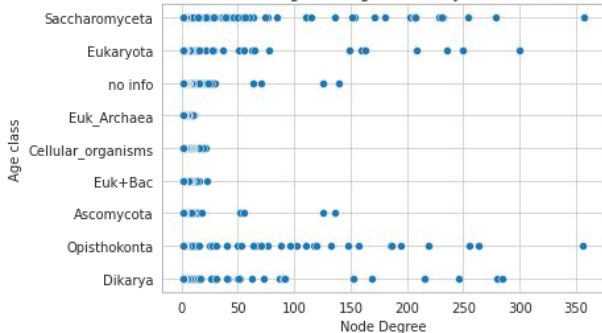
Connectivity, Yeast GRN



Interaction distance, yeast



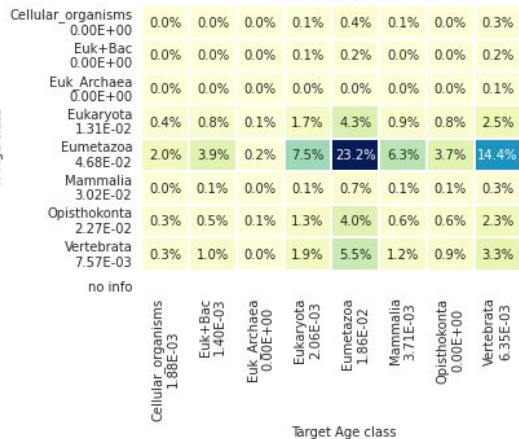
Node Degree in age classes, yeasts GRN





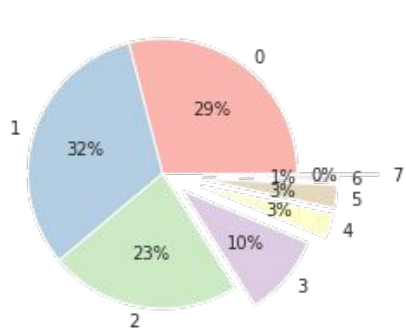
2.2. Results for Mouse GRN

Age classes heatmap, mouse
with clusterization coefficients

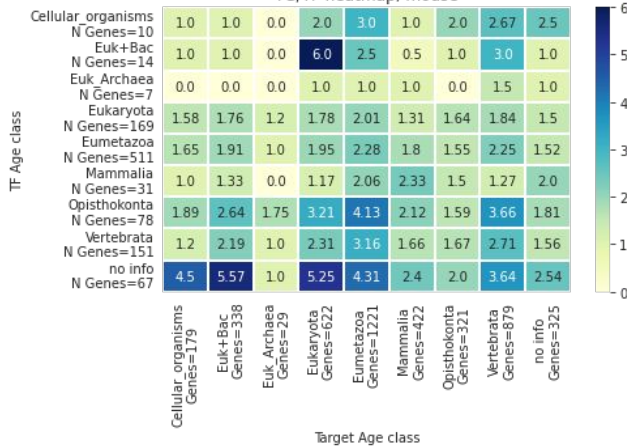


Target Age class

Interaction distance, mouse

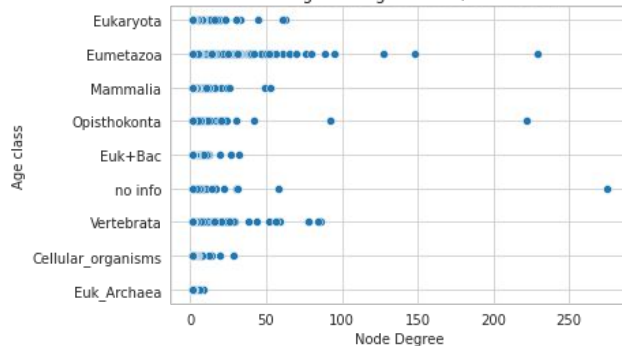


TG/TF heatmap, mouse

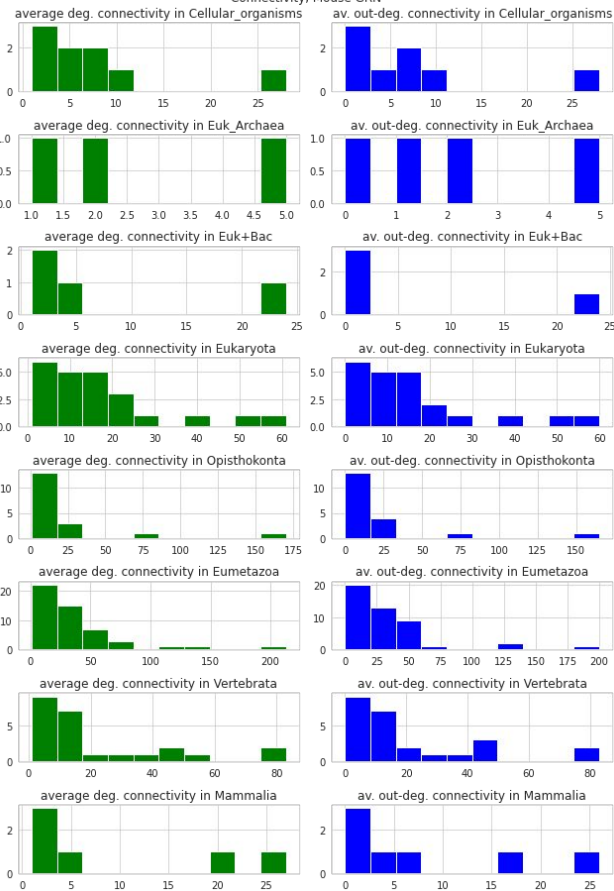


Target Age class

Node Degree in age classes, mouse GRN



Connectivity, Mouse GRN



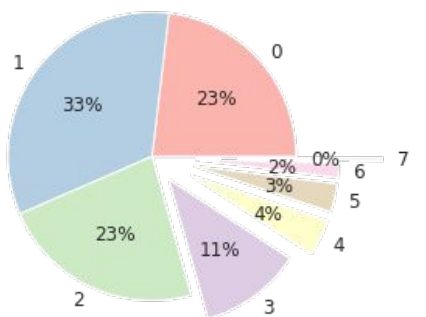


2.3. Results for Human GRN

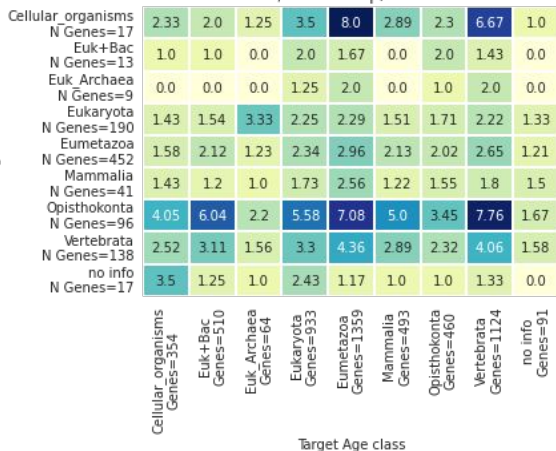
Age classes heatmap, human
with clusterization coefficients



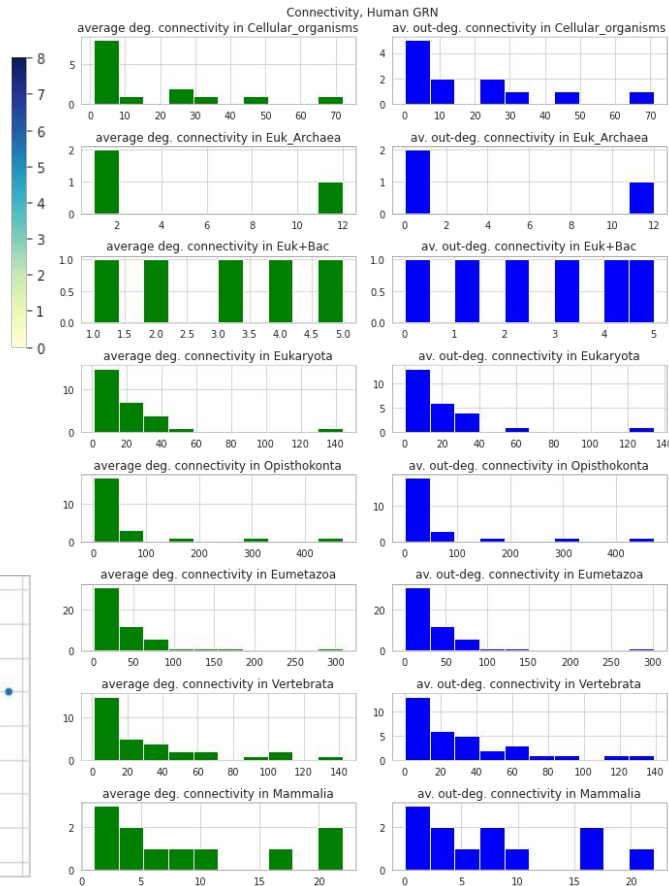
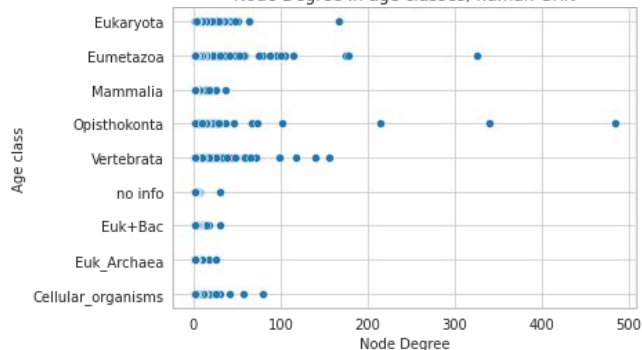
Interaction distance, human



TG/TF heatmap, human



Node Degree in age classes, human GRN



3. GenOrigin age classes

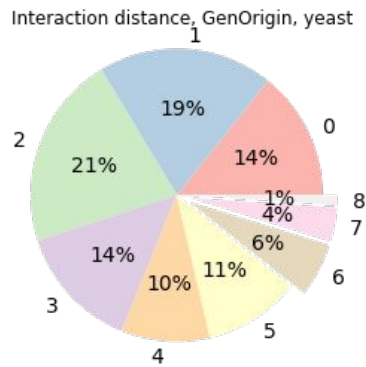
The gene ages calculated from the protein ages gave different results for human and mouse and yeast GRNs. We decided to use the gene ages from the [GenOrigin](#)⁶ database for Yeast GRN parameters calculation.

Yeast

- age classes are available for 6 600 genes.
- After mapping:

Age class	Genes number
0	> Eukaryota 332
1	Ascomycota 950
2	Dikarya 213
3	Eukaryota 539
4	Fungi 463
5	Opisthokonta 81
6	Saccharomyces 688
7	Saccharomyceta 445
8	Saccharomycetaceae 474
Total	Total 4185

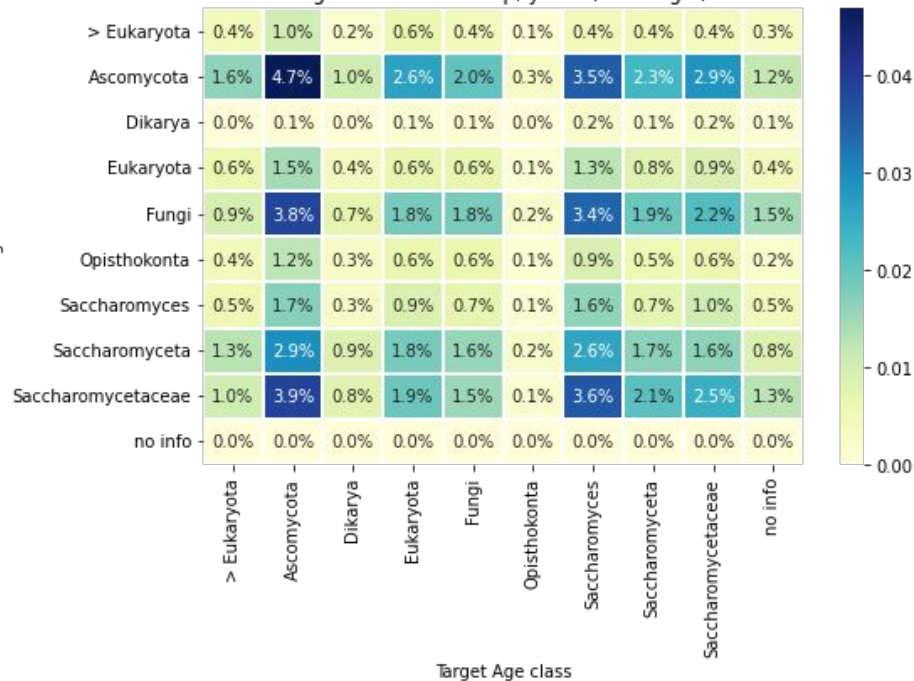
We used the GenOrigin phylogenetic tree to convert a numerical age into an age class.



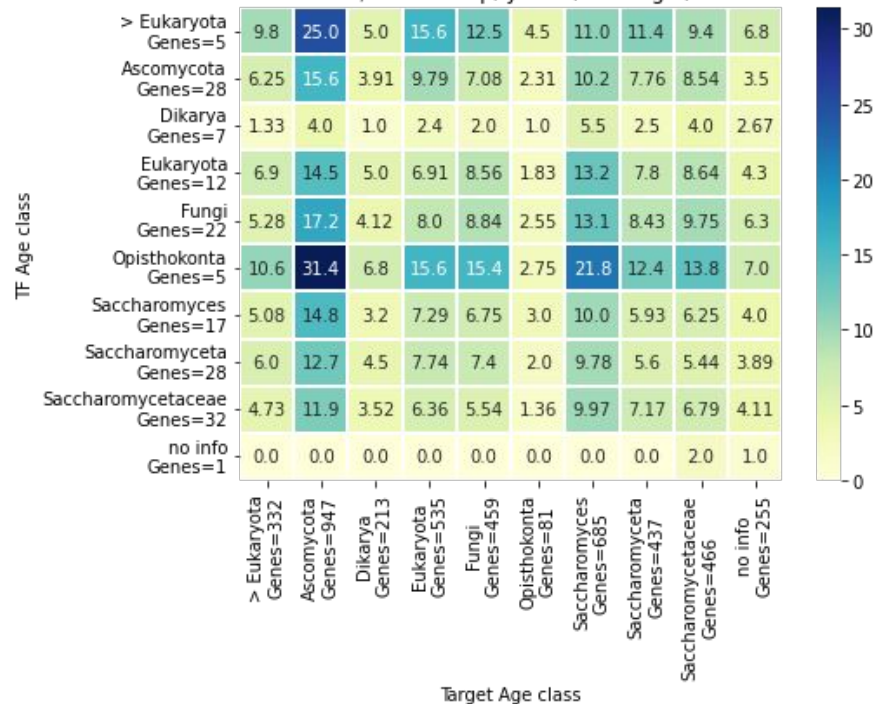


Heatmaps:

Age classes heatmap, yeast (GenOrigin)



TG/TF heatmap, yeast (GenOrigin)



4. Phylostratigraphy

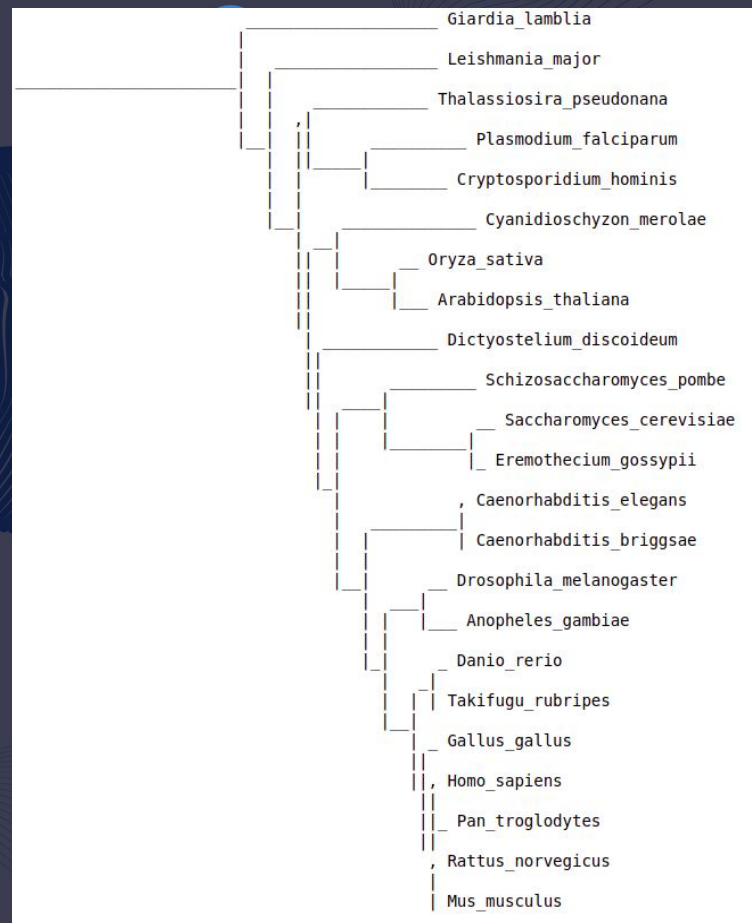
We used a phylostratigraphy approach ⁷ to determine the yeast genes age in GRN.

BLASTx:

- iTOL tree ⁸ (leave only Eukaryotes, for truncate the swiss db);
- truncated Swiss-prot database (94 268 sequences);
- e-value 10^{-3} .

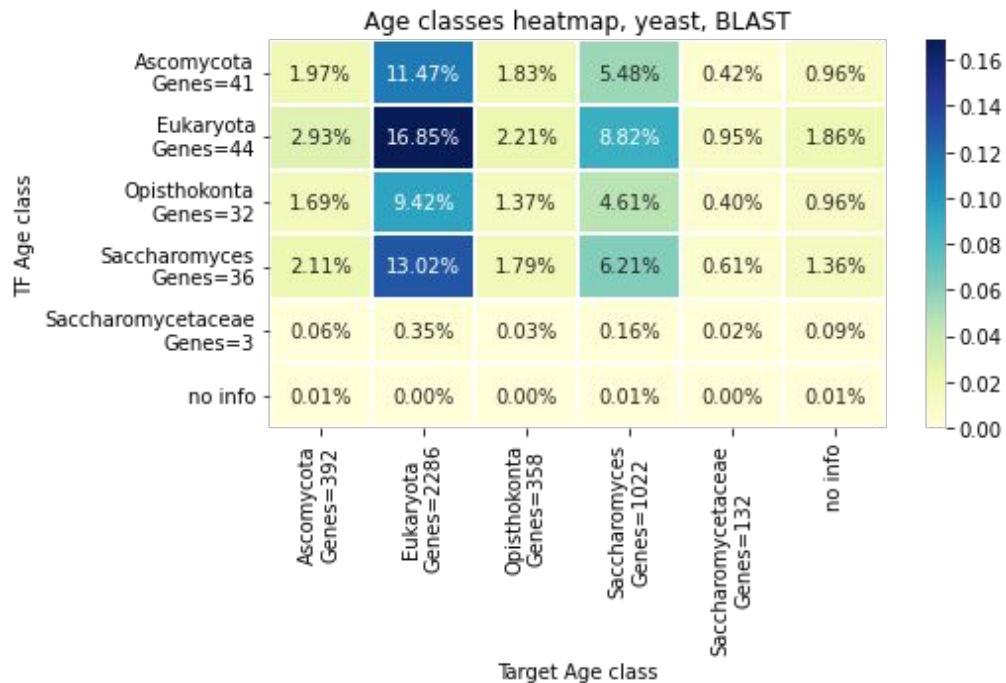
The database
class structure:

	Age class	Seq number
0	Ascomycota	5106
1	Eukaryota	25332
2	Opisthokonta	55622
3	Saccharomyces	7281
4	Saccharomycetaceae	927
Total	Total	94268

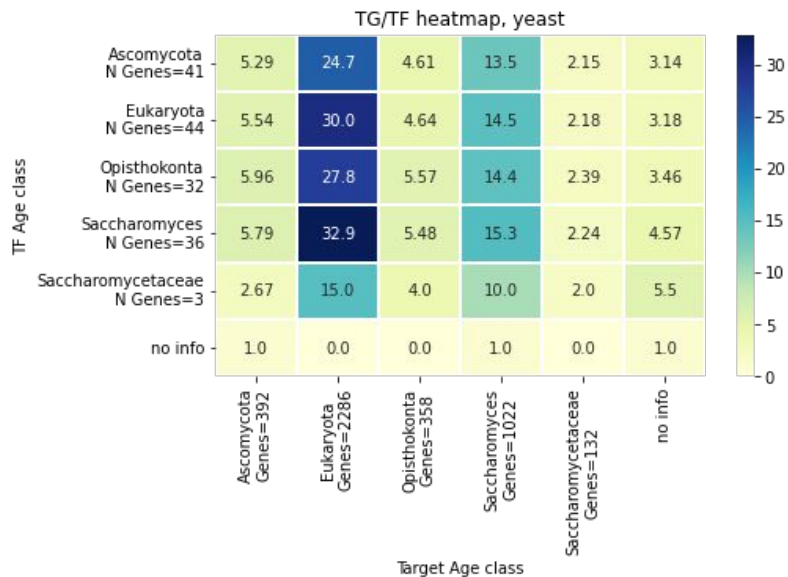
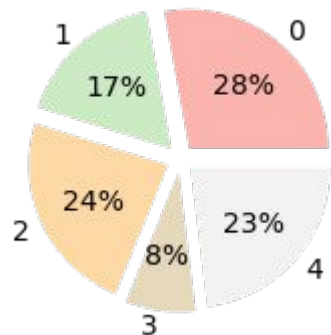




4.1. Phylostratigraphy results for Yeast GRN



Interaction distance (BLAST), yeast



Conclusion

Preliminary analysis shows a more complex age structure of the yeast gene regulatory network than in mice or human GRNs.

We plan to blast GRN genes to a fine-grained tree with more uniform representation of nodes across gene classes.

All code and data is available on GitHub page:
github.com/Freddsle/age_patterns





References

1. Chen, C-Y., et al. "**Dissecting the human protein-protein interaction network via phylogenetic decomposition.**" Scientific reports 4.1 (2014): 1-10. [DOI](#).
2. Balaji, S., et al. "**Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast.**" Journal of molecular biology 360.1 (2006): 213-227. [DOI](#).
3. Vavulov, O., Zhivkopoulos, E. "**Building gene regulatory networks with structural properties.**" Bioinformatic institute (2021). [LINK](#).
4. Han, H., et al. "**TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions.**" Nucleic acids research 46.D1 (2018): D380-D386. [DOI](#).
5. Liebeskind, B.J., McWhite, C.D., and Marcotte, E.M.. "**Towards consensus gene ages.**" Genome biology and evolution 8.6 (2016): 1812-1823. [DOI](#).
6. Tong, Y.-B., et al. "**GenOrigin: A comprehensive protein-coding gene origination database on the evolutionary timescale of life.**" Journal of Genetics and Genomics (2021). [DOI](#).
7. Domazet-Lošo, T., Brajković J., and Diethard Tautz. "**A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages.**" Trends in Genetics 23.11 (2007): 533-539. [DOI](#).
8. [iTOL tree](#).