Stratification of prostate cancer patients based on molecular interaction profiles

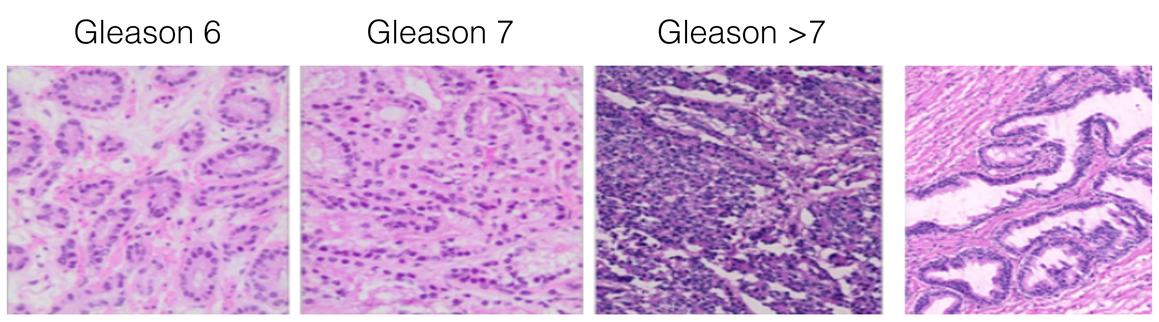
Roland Mathis[1], Matteo Manica[1,2], Maria Rodriguez Martinez[1]

[1] IBM Research Zurich, [2] ETH Zurich

Motivation

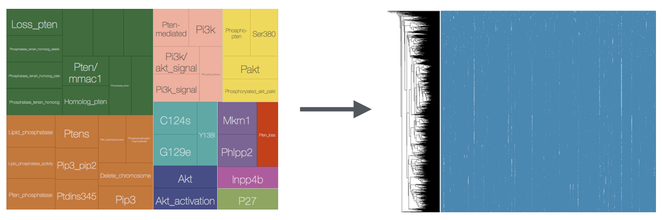
Prostate cancer is a leading cause of cancer death amongst men, however the molecular-level understanding of disease onset and progression are largely unknown. Specifically, stratification of intermediate prostate tumor states based on current markers is difficult. To improve stratification we

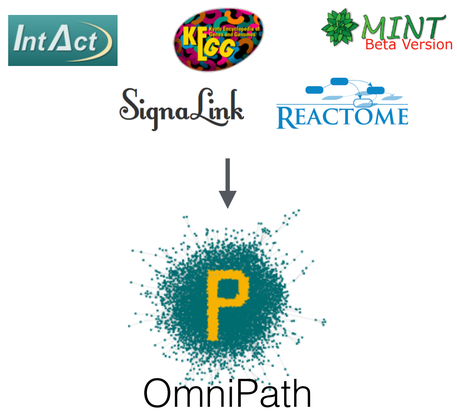
1. integrate data, literature and public databases,
2. extract interaction groups from interactome and cluster patients for each interaction group,
3. train a classifier using the patient profiles to identify marker interactions.

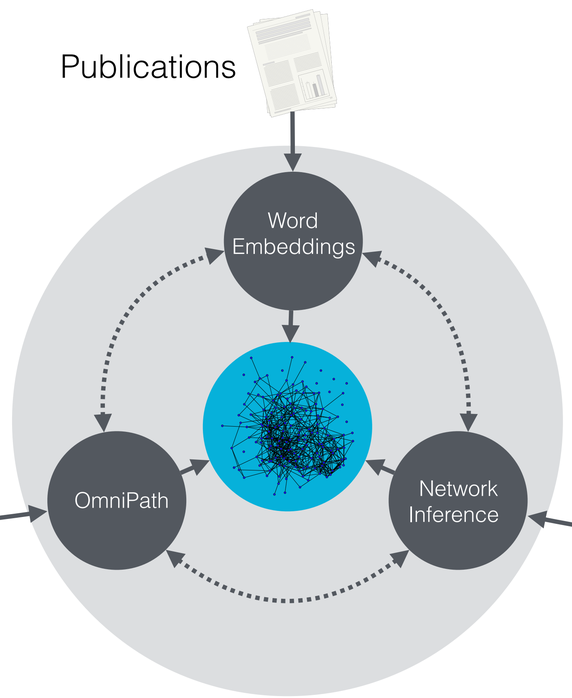


Interactome

Multiple inference methods were applied to a given data-set to construct a data-driven interactome. To enrich the data driven interaction network, separate interactomes were extracted from publications and databases using Omnipath[2)] and word embeddings[3)]. By combining the interaction networks derived from data, public databases and publications a consensus interactome was built.



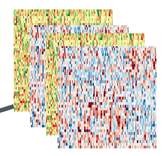




352 prostate tumor samples 152 proteins (RPPA)

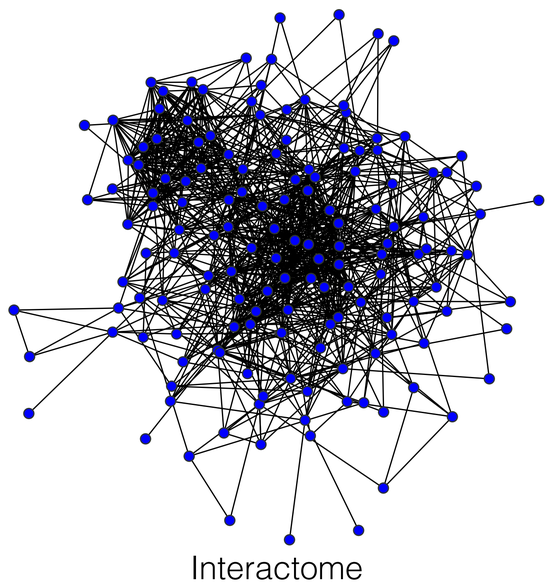
210 prostate tumor samples 3780 proteins (SWATH)

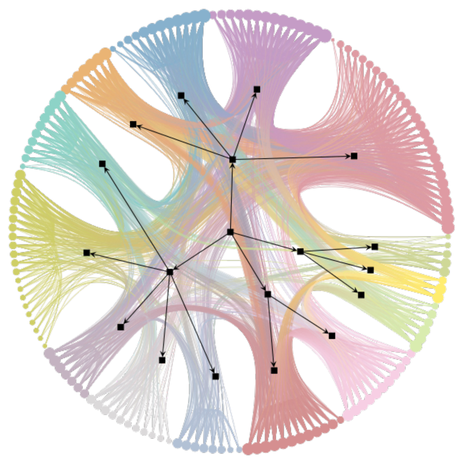
Name	Based on
Pearson	Correlation
Spearman	Ranks correlation
Glasso[4)]	Partial correlations
Aracne[5)]	Mutual information
FunChisq[6)]	Functional depend.



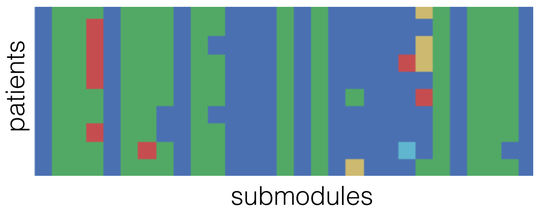
Stratification

From an interaction network we identify relevant interaction modules through graph-theory approaches. For each interaction module we cluster the patients based on molecular states measurements. The patient-specific cluster assignment vectors serve as a personalized interaction signatures and is used to stratify patients.





Cluster submodules using patient data



Stratification of patients based on vectors of submodule cluster-ids

1) Image source: Peter Wild Lab, Zurich University Hospital
2) D Turei, T Korcsmaros and J Saez-Rodriguez (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nature Methods 13(12)
3) Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv Margolin,
4) Adam A., et al. "ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context." BMC bioinformatics 7.Suppl 1 (2006): S7.
5) Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "Sparse inverse covariance estimation with the graphical lasso." Biostatistics 9.3 (2008): 432-441.
6) Yang Zhang and Mingzhou Song. Deciphering interactions in causal networks withoutparametric assumptions.arXiv preprint arXiv:1311.2707, 2013.

This work is part of the PrECISE project. PrECISE combines hypothesis-driven strategies with data-driven analysis in a novel mathematical and computational methodology for the integration of genomic, epigenetic, transcriptomic, proteomic, and clinical data with the goal of risk-stratifying patients and suggesting personalized therapeutic interventions. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 668858.
Project website: www.precise-project.eu