DeepGRN: Deciphering gene deregulation in cancer development using deep learning

Roland Mathis[1,*], Matteo Manica[1,2,*], María Martínez Rodríguez[1]

[1] IBM Research Zurich, [2] ETH Zurich, [*] Shared first authors

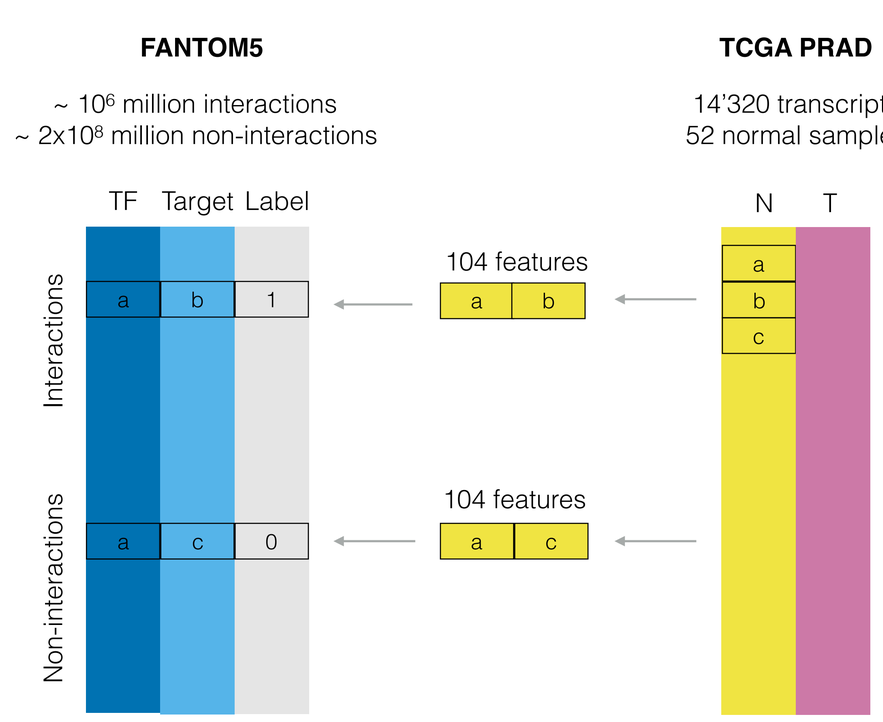
Motivation

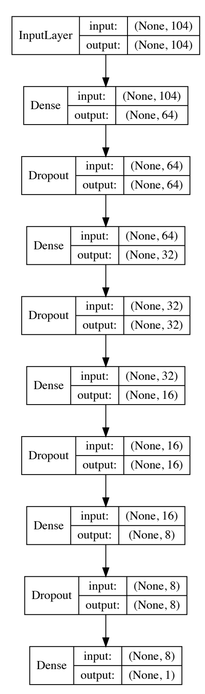
Understanding gene regulatory networks (GRNs) is key towards deciphering gene deregulation in cancer development. We are building on efforts to find tissue-specific and disease-specific gene regulatory networks [1,2]. While large efforts have been devoted to create context specific GRNs for a range of tissues as well as diseases, most currently available cancer GRNs are inferred from unmatched datasets for which only the diseased tissue is available. Our goal is to find disease-specific changes of gene regulation using matched normal and tumor patient data in a cohort-specific fashion.

Training and Prediction

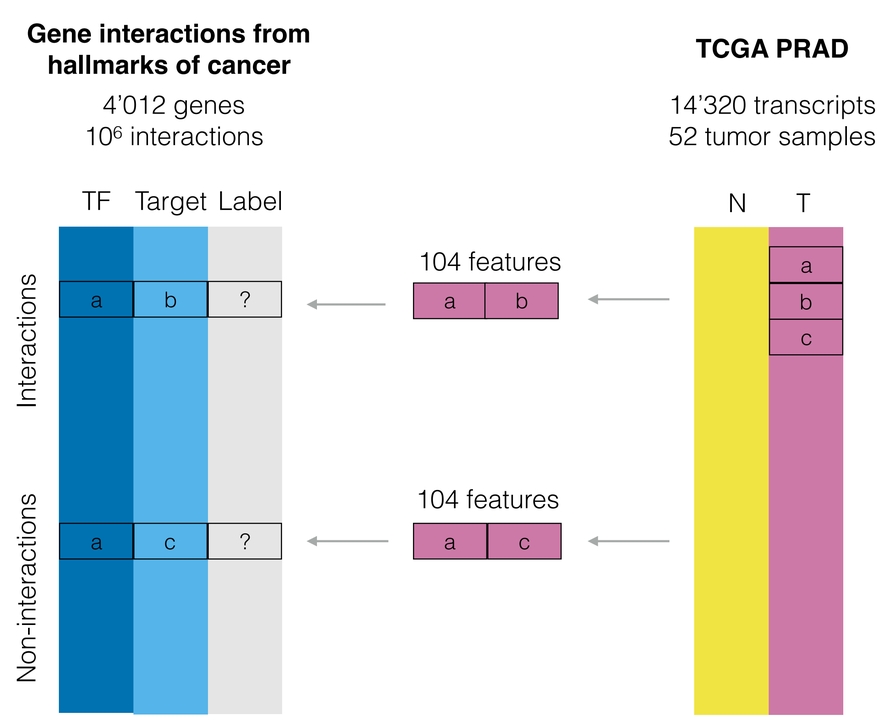
We propose DeepGRN, a deep learning model that enables us to find cohort-specific disease-induced changes in the GRN's of cancer patients by learning the interactions from RNASeq measurements and reported tissue-specific interactions. We apply DeepGRN to the TCGA prostate cancer cohort [3]. For each prostate-specific interaction reported in [1] we use as features the joint RNASeq measurements from the two interacting genes of the patients in a given cohort. We then train a deep learning network on the data from normal patient samples to learn transcription factors to target interactions in disease-free state. Once the model has been trained on normal samples, we predict the GRN and their deregulation for the tumor samples, highlighting differences in the regulation process between normal and tumor samples.

Learning from normal





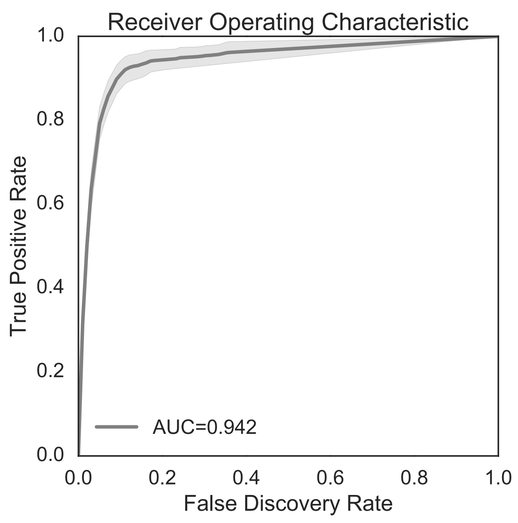
Predicting for tumor

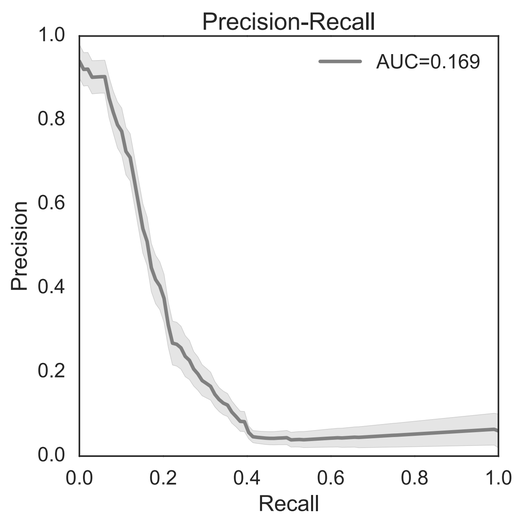


Evaluation and Results

Transcription factors were extracted from FANTOM5 normal network and DeepGRN normal and tumor predictions using Out-Degree Centrality. While DeepGRN can predict transcription factors with a high confidence, target specificity has to be improved. DeepGRN can be used to generate hypotheses for detection of new biological processes relevant for cancer onset and development and puts forward a novel approach to drive drug discovery and suggest targeted therapies.

Evaluation





Results

Top 10 Regulators

FANTOM5 normal
KLF4 MYC IRF1 ETS1 NR3C1 BHLHE40 SMAD3 FOXO3 IRF8 NFKB1

Predicted normal
MYC IRF1 ATF3 GADD45B MARCKS CCND2 KLF4 IRF8 IER3 IGFBP3

Predicted tumor
BTG2 MYC CD44 CRAT GADD45B FAS FOS IDH1 TP53 TAP1

1) FANTOM Consortium and the RIKEN PMI and CLST (DGT), Alistair R. R. Forrest, Hideya Kawaji, Michael Rehli, J. Kenneth Baillie, Michiel J. L. de Hoon, Vanja Haberle, et al. 2014. "A Promoter-Level Mammalian Expression Atlas." Nature 507 (7493): 462–70.
2) Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z. and Bergmann, S., 2016. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nature methods.
3) TCGA Research Network: http://cancergenome.nih.gov
4) Umbehr, M., Kessler, T.M., Sulser, T., Kristiansen, G., Probst, N., Steurer, J. and Bachmann, L.M., 2008. ProCOC: The prostate cancer outcomes cohort study. BMC urology, 8(1), p.9.

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