

Pan-cancer discovery and characterization of clinically informative lncRNAs

Keren Isaev^{1,2}, Jüri Reimand^{1,2}

¹Department of Medical Biophysics, University of Toronto

²Ontario Institute for Cancer Research

Background

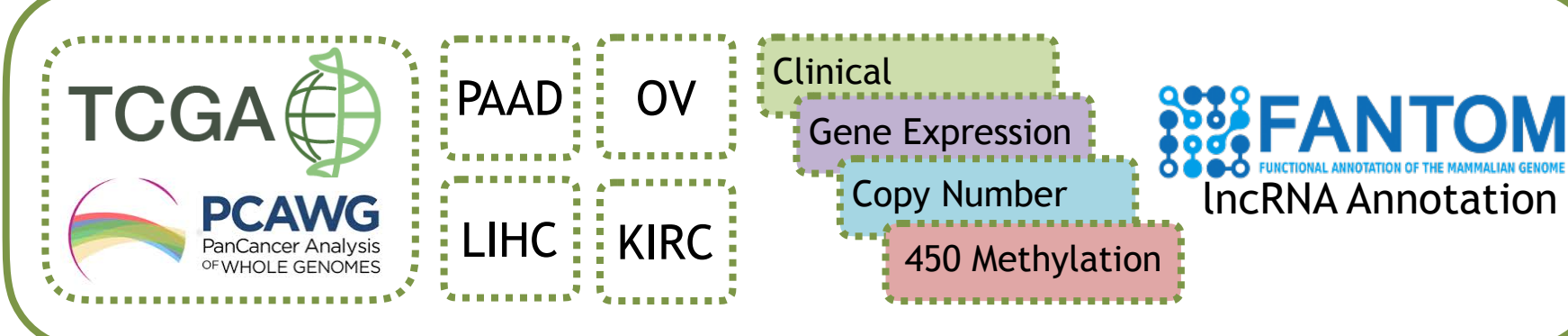
It is becoming increasingly evident that the genome is pervasively transcribed into non-coding RNA including long non-coding RNAs (~20,000 estimated to be functional by FANTOM). Several lncRNAs including *HOTAIR*, *MALAT1* and *SCHALP1* have demonstrated a relationship between gene expression and patient survival outcomes or metastasis in several cancers. However, the majority of lncRNAs remain both functionally and clinically undefined while presenting a diverse array of translational opportunities due to their tissue specific expression and potential to regulate cancer pathways. Thus, the purpose of this project is to define a subset of high-confidence lncRNAs for further study and potential translation for diagnostic and prognostic value.

Methods and workflow

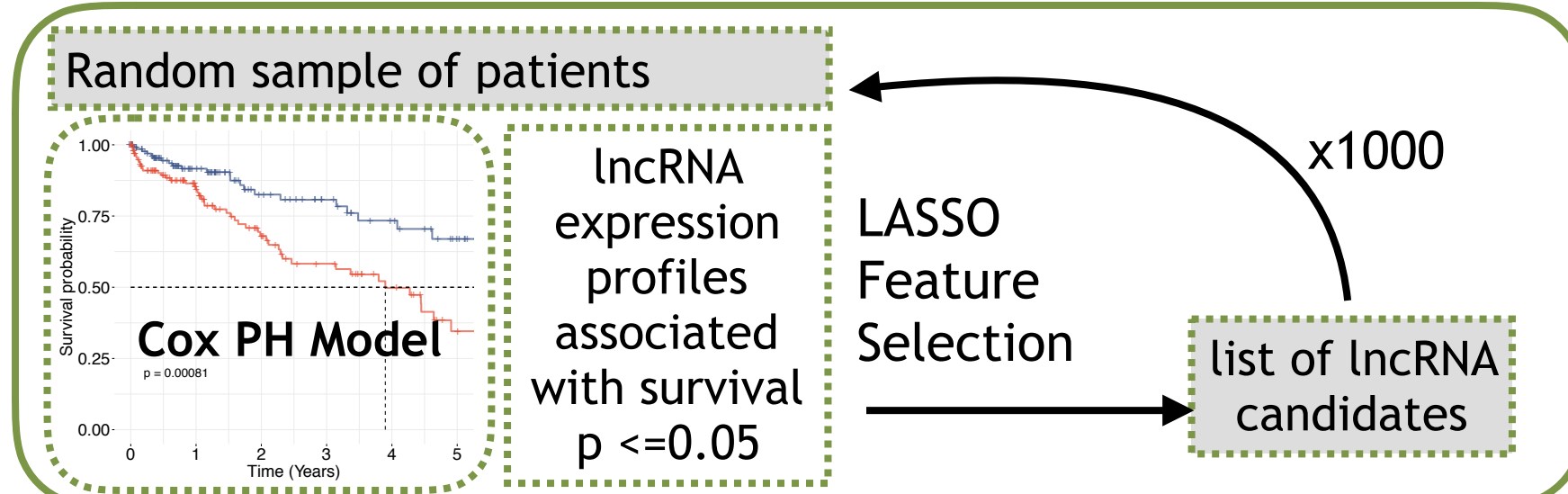
Aims:

1. Identify a set of high confidence prognostic lncRNA expression profiles
2. Elucidate the underlying genetic or epigenetic differences between the lncRNA annotated risk groups
3. If function is unknown, predict its PCG interactions and associated biological pathways

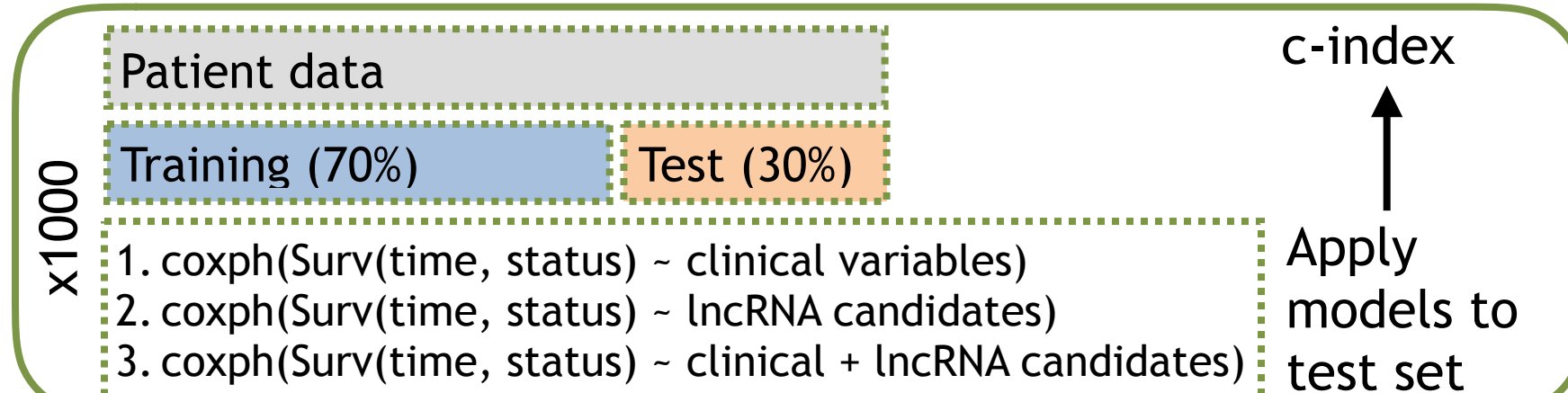
1. -omics datasets across 4 cancer types



2. Sample 70% of patients for feature selection



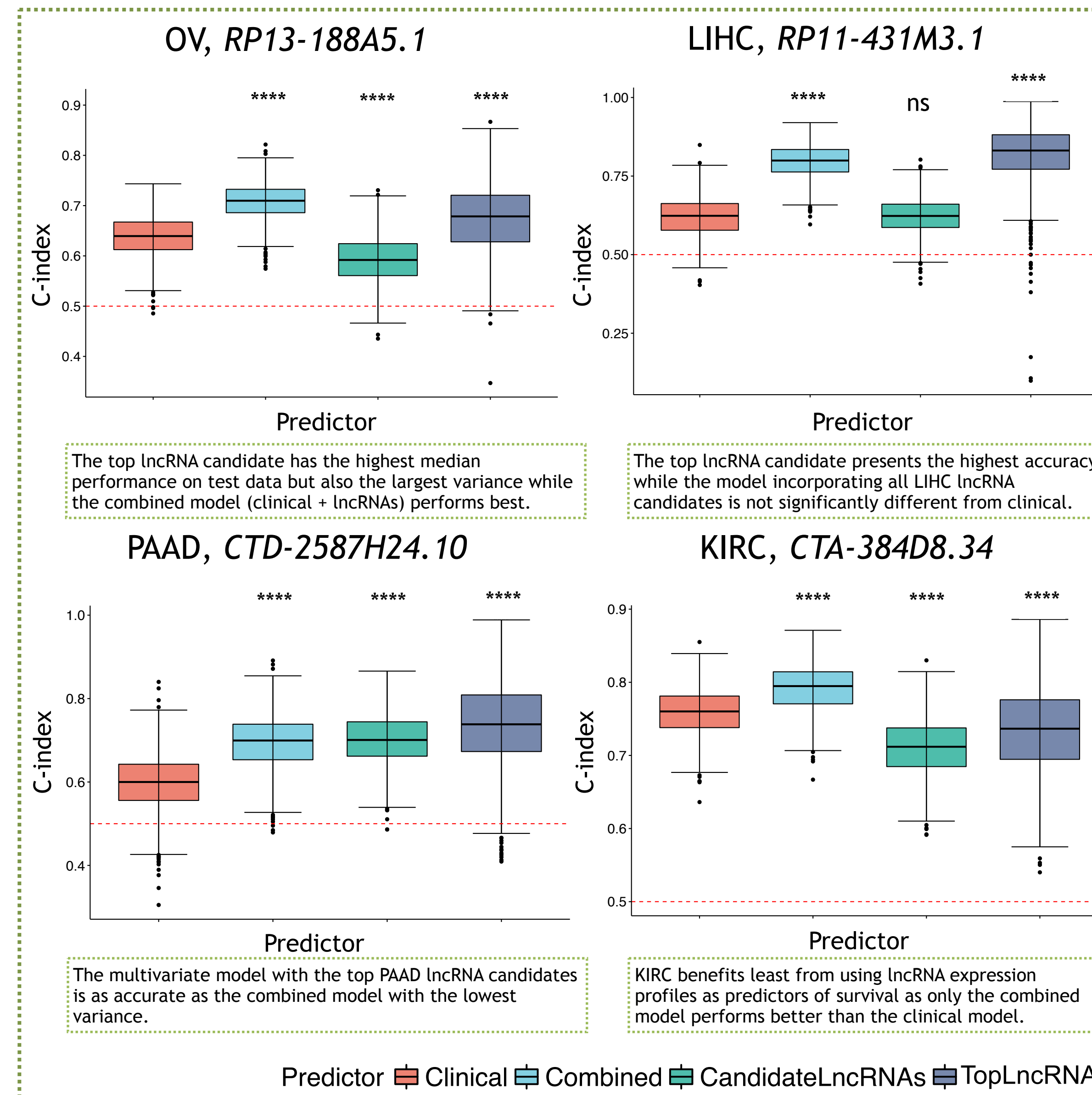
3. Final list of lncRNA candidates consists of those chosen in at least 40% of the random splits



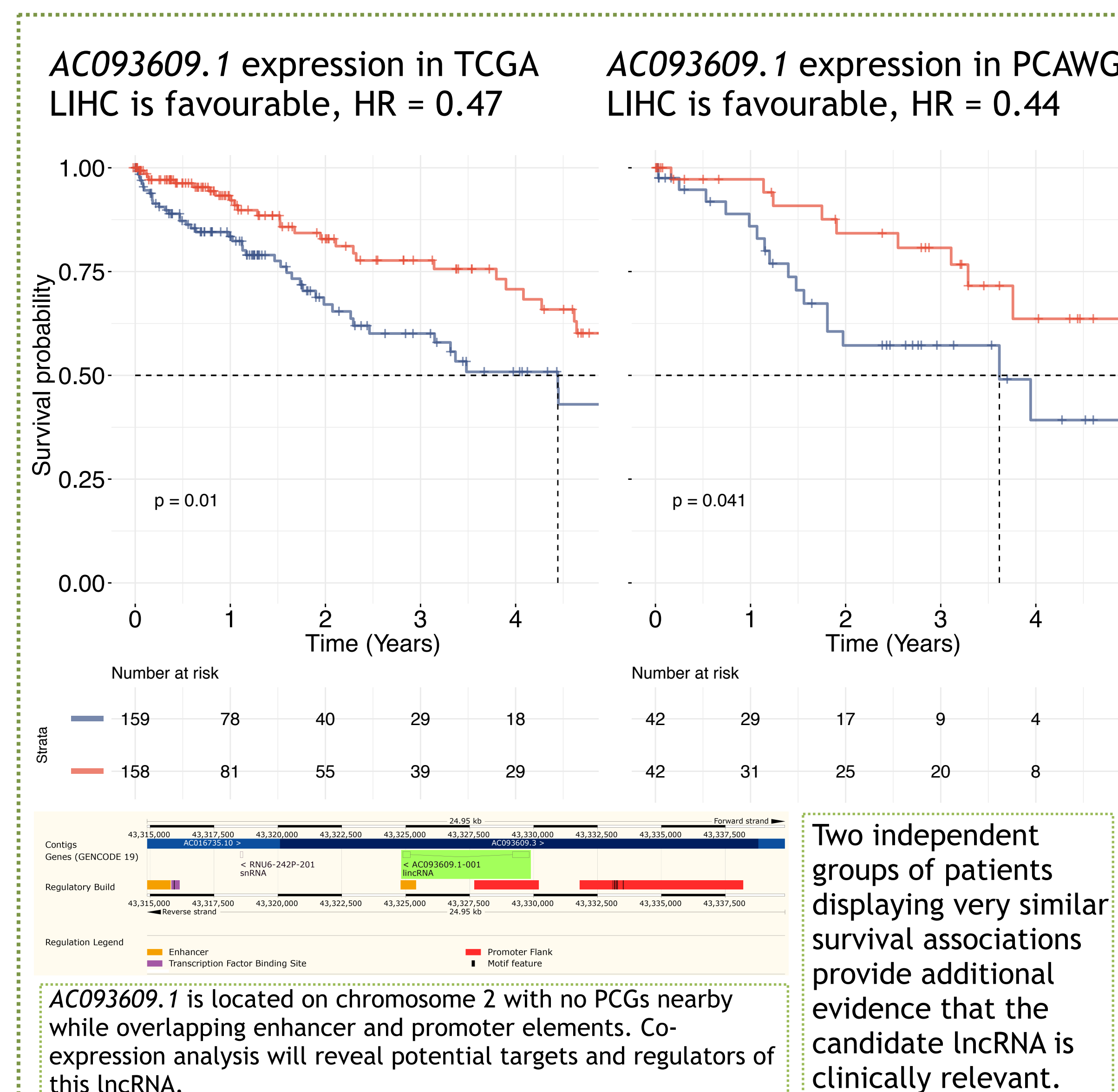
4. Identify genetic or epigenetic aberrations driving expression profiles and predict functional associations

1. lm(lncRNA expression ~ promoter methylation)
2. lm(lncRNA expression ~ segment mean)
3. glmnet(lncRNA expression (high/low) ~ PCGs, family = "binomial")

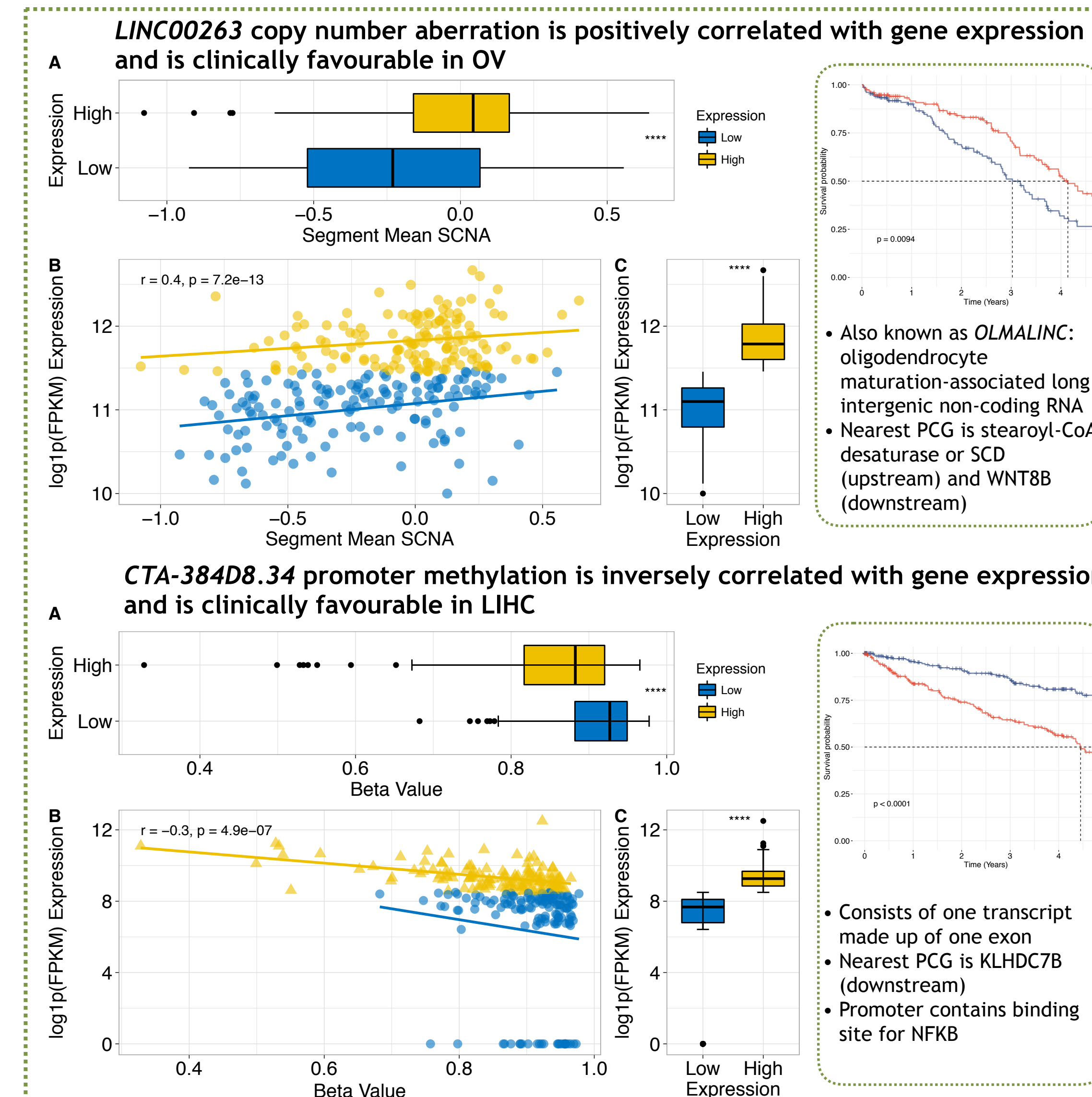
Top lncRNA candidates outperform clinical variables in predicting survival in LIHC, OV and PAAD



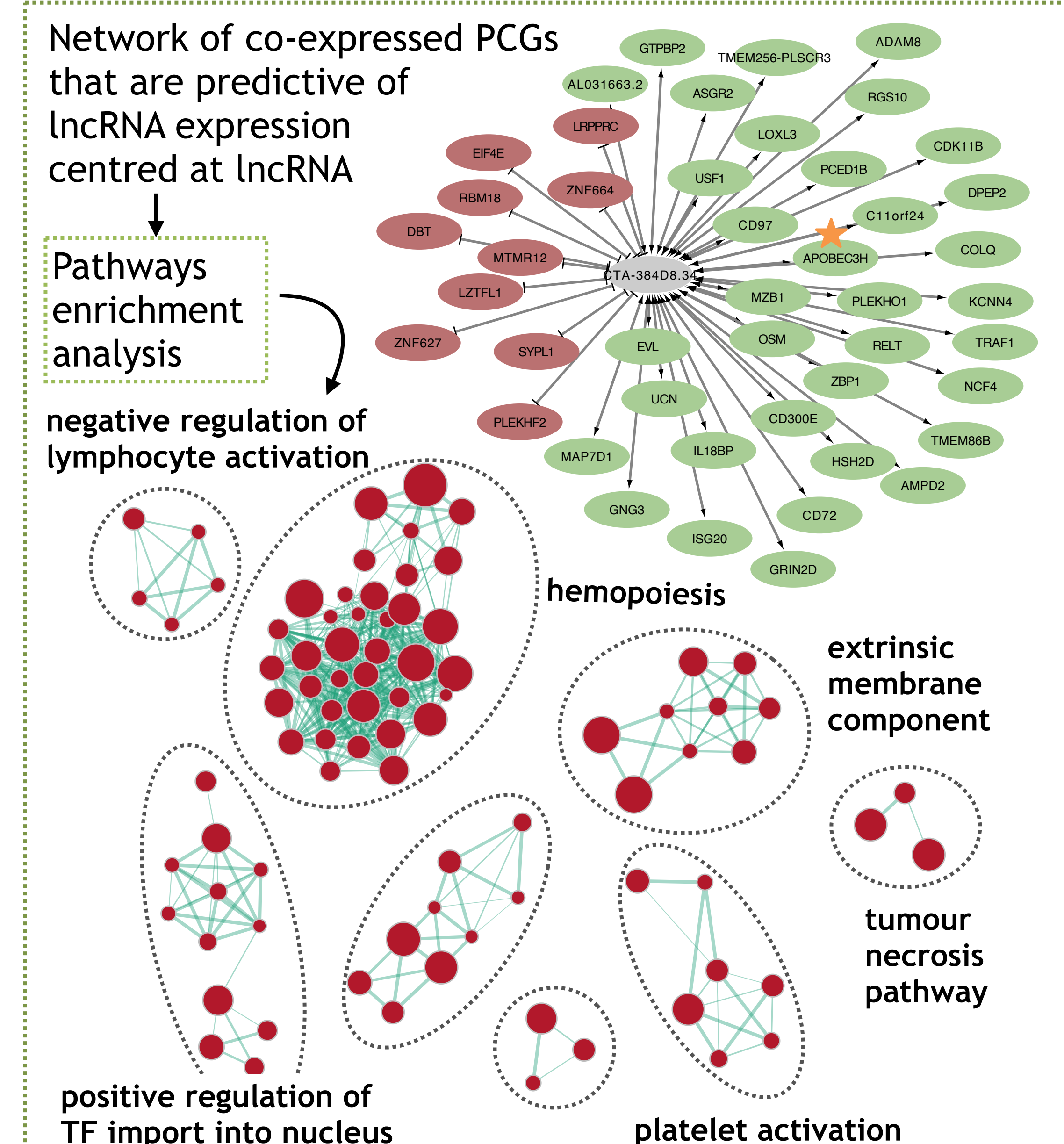
AC093609.1's prognostic expression is validated in an external dataset PCAWG



Copy number aberrations and methylation influence the expression of candidate lncRNAs



CTA-384D8.34 is associated with immune response in KIRC

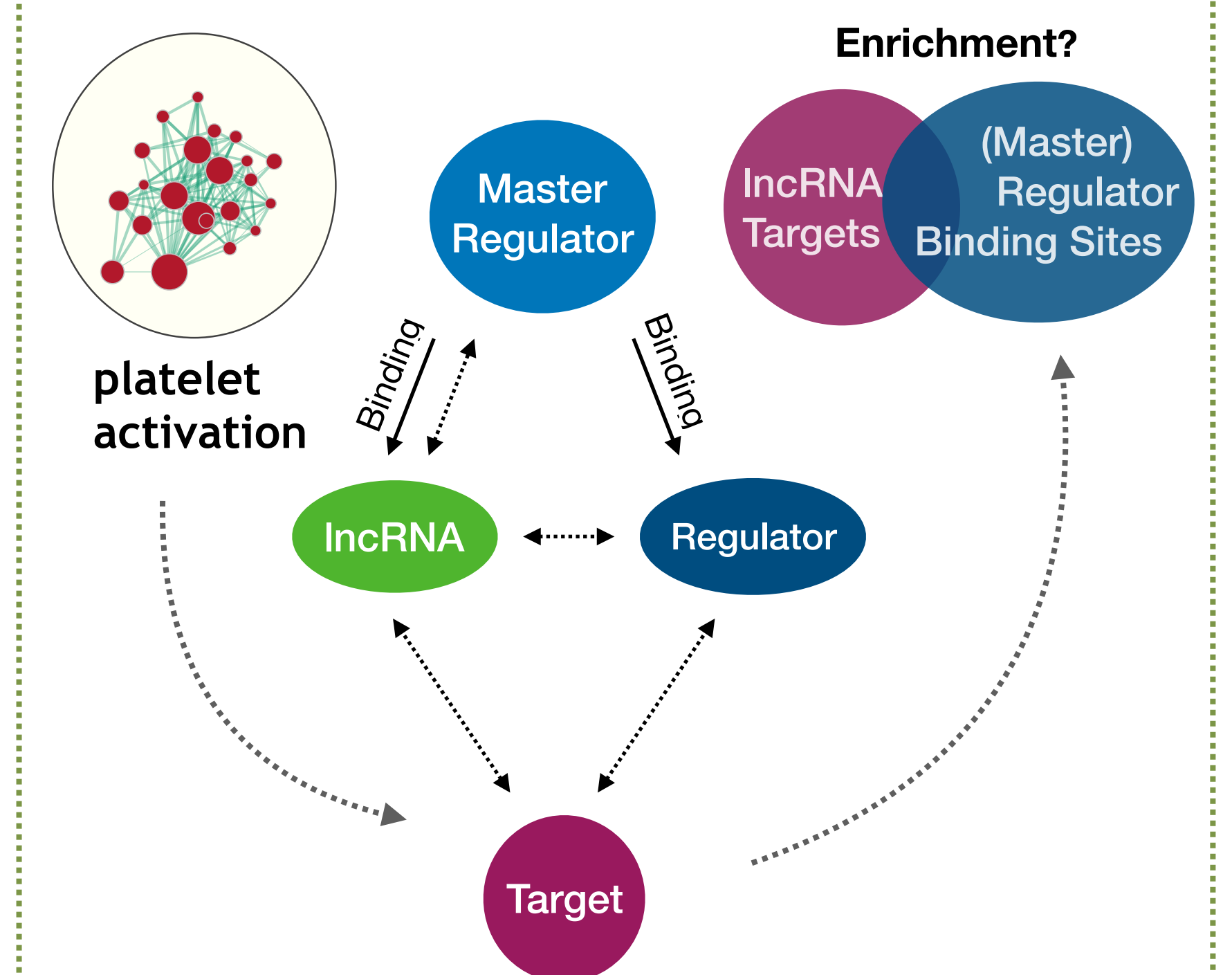


Future directions and applications

1. Consider additional clinical variables including drug response, metastasis and recurrence
- Combine best performing lncRNA molecular profiles into multivariate model
- Compare performance of lncRNA predictors with other genetic factors such as the presence of known cancer driver mutations

2. What is the biological mechanism driving the differences in expression between the different risk groups?

- Integrate transcription factor binding data



Summary

There is a large group of lncRNAs whose function is mostly unknown. We can focus on a smaller subset that promises more clinical relevance and translation. This sort of filtering can allow us to narrow the list of 1000s of genes to those that are most likely improving patient diagnosis and prognosis.

~ 20,000 lncRNAs

Filtering

lncRNAs associated with survival

Prognostic biomarker

Functional associations

Citations and Acknowledgements

TCGA, PCAWG, ...

Medical Biophysics
UNIVERSITY OF TORONTO

Ontario

Funding for the Ontario Institute for Cancer Research is provided by the Government of Ontario.