

Bespoke Biomarker Combinations for Cancer Survival Prognosis Using Artificial Intelligence on Tumour Transcriptomics

Ricardo Jorge Pais^{1,2*}, Tiago Alexandre Pais¹, Uraquitan Lima Filho¹

¹ Bioenhancer Systems LTD, Manchester, United Kingdom; ² Egas Moniz for Interdisciplinary Research (CiEM), Caparica, Portugal; * Correspondence: RP_ripais@bioenhancersystems.com

BACKGROUND

Accurate cancer survival prognosis remains a major challenge in oncology, with single biomarker approaches often lacking sufficient specificity and sensitivity for clinical application. Recent advances in transcriptomic profiling and artificial intelligence (AI) offer promising new directions to identify combinations of molecular features that better predict patient outcomes. In this study, we applied our AutoML algorithm **O2Pmgen** via the **Digital Phenomics** platform to tumour transcriptomic datasets from The Cancer Genome Atlas (TCGA) for breast, lung, and renal cancers. We further benchmarked our achievements with a well-established autoML approach (**TPOT**) and a single predictive biomarker approach (**BMfinder**).

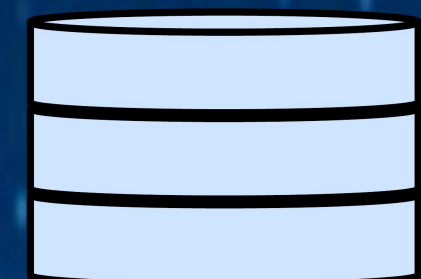
METHODOLOGY/Framework

THE HUMAN PROTEIN ATLAS

https://www.proteinatlas.org/humanproteome/cancer/data#tcga_cancer_samples_rna

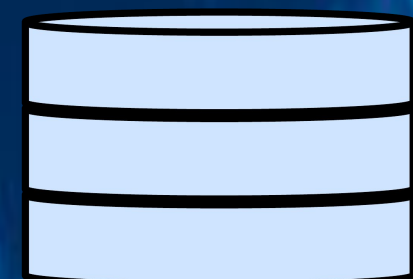
Tumour Transcriptomics Datasets

Breast Cancer



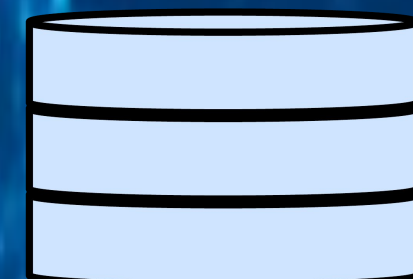
199 good prognosis
40 poor prognosis

Lung Cancer

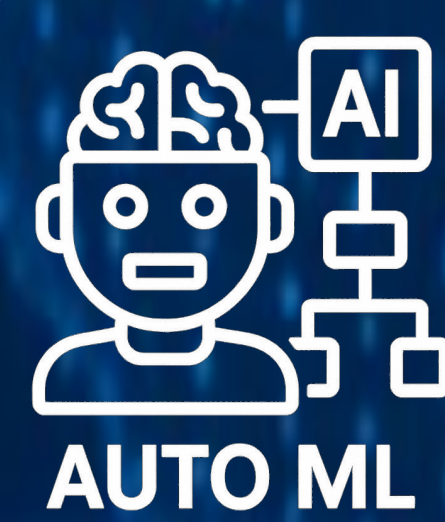


94 good prognosis
210 poor prognosis

Renal Cancer



210 good prognosis
108 poor prognosis



Model
training

Model
validation

Model
selection

Software tools

BMfinder

O2Pmgen

TPOT

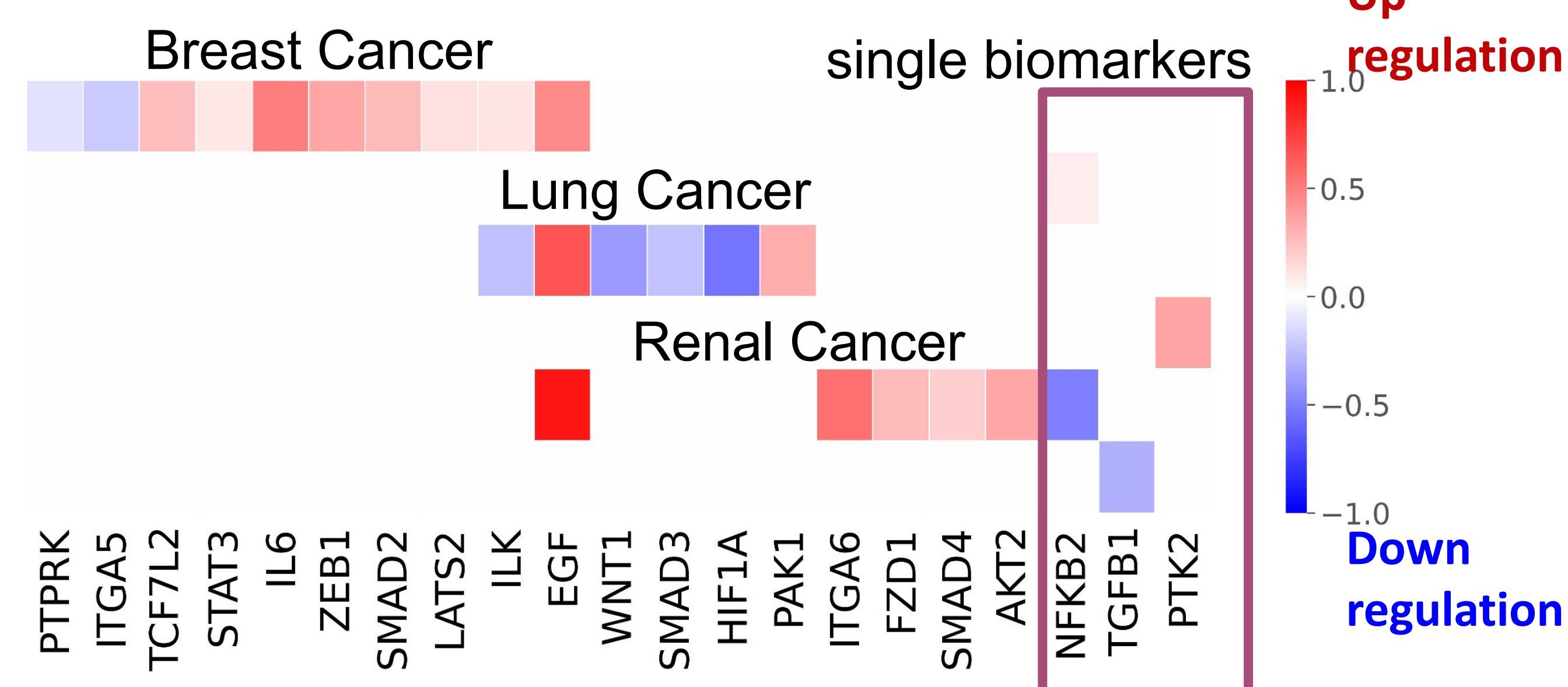
Bespoke
single
predictive
biomarker

Bespoke
predictive
combination of
biomarkers

48 genes
optimised
ML pipeline

RESULTS

biomarker panels for cancer prognostic



Breast cancer prognostic models performance

	Single	EMT genes	Panel
AUC	62%	65%	83%
Sensitivity	84%	95%	62%
Specificity	42%	40%	62%

Lung cancer prognostic models performance

	Single	EMT genes	Small panels
AUC	60%	63%	75%
Sensitivity	71%	79%	81%
Specificity	40%	50%	61%

Renal cancer prognostic models performance

	Single	EMT genes	Small panels
AUC	62%	58%	71%
Sensitivity	73%	82%	81%
Specificity	41%	63%	60%

CONCLUSIONS

- Bespoke predictive models based on small gene panels outcompete single biomarker and large gene panel approaches.
- Predictive models based on small gene panels are solutions for cancer prognosis.

References

- Mann, M.; Kumar, C.; Zeng, W.F.; Strauss, M.T. Artificial Intelligence for Proteomics and Biomarker Discovery. *Cell Syst.* 2021, 12, 759–770.
- Le, T.T.; Fu, W.; Moore, J.H. Scaling Tree-Based Automated Machine Learning to Biomedical Big Data with a Feature Set Selector. *Bioinformatics* 2020, 36, 250–256.
- Filho, U.L.; Pais, T.A.; Pais, R.J. Facilitating "Omics" for Phenotype Classification Using a User-Friendly AI-Driven Platform: Application in Cancer Prognostics. *BioMedInformatics* 2023, 3, 1071–1082.
- Uhlen, M.; Zhang, C.; Lee, S.; Sjöstedt, E.; Fagerberg, L.; Bidkhor, G.; Benfeitas, R.; Arif, M.; Liu, Z.; Edfors, F.; et al. A Pathology Atlas of the Human Cancer Transcriptome. *Science* 2017, 357, eaan2507.

Acknowledgments

This research was fully supported by Bioenhancer Systems. The project also received support from the UK government funded Innovation Navigator Programme (UK Business Growth Hub).

