Data-Driven Discovery of Novel Antigen Targets: A Scalable Bioinformatics Pipeline

L. Meunier[1], G. Appé[1], M. Colange[1], E. Fox[1], L. Hensen[1], C. Marijon[1], A. Nordor[1], S. Weill[1], and A. Behdenna[1]. [1]Epigene Labs, Paris, France

JOBIM 2025 08>11 JUIL Bordeaux

Introduction

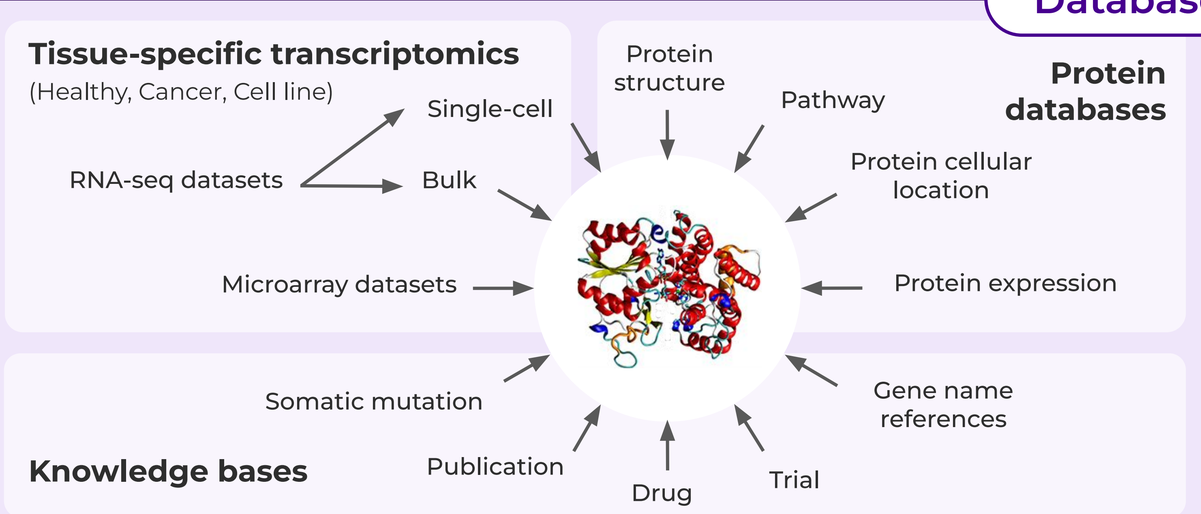
- Antigen-targeting therapies hold promise but face ~90% trial failure, mainly due to low efficacy or toxicity.
- Early, data-driven target selection can triple oncology approval rates by improving selection and reducing failures[1].
- We present a scalable platform that mines public omics data to identify and rank novel antigen targets across cancers.

Conclusion

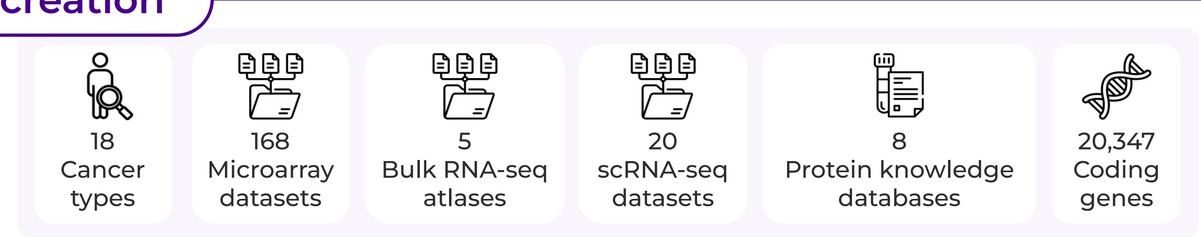
- Applied to 18 cancer cohorts, our pipeline identified ~216 candidate antigens per indication (range: 41-556), including 8 FDA-approved targets.
- On average, 8 candidates per cancer type outperformed FDA-approved benchmarks in efficacy and safety, while this approach also flagged potential on-target off-tumor toxicities.
- These results highlight the pipeline's potential to accelerate discovery of novel, clinically relevant oncology targets.

Methods

Database creation



Results



- All data sources were curated using an AI-assisted, human-supervised pipeline and mapped to a proprietary ontology for consistent annotation.
- Gene names were standardized across sources using a common reference.
- Microarray datasets were normalized, and assembled into 18 indication-specific disease atlases comprising 7,885 tumor and 1,569 healthy samples, using batch-correction[2] to preserve biological signal while removing technical artifact.

Target identification

1 Identifying cancer-specific targets via differential expression analysis

On microarray cancer specific atlas - healthy samples vs tumor samples
inMoose[3] LogFC > 1 and FDR < 0.05

2 Selecting cell surface proteins using proteomic filters

Proteins located at the plasma membrane	Proteins with available expression data in healthy tissues
HPA, COMPARTEMENT	HPA, HPM, PDB

Antigen targets

Proteins with transmembrane domain	Or Uniprot	GPI-anchored proteins

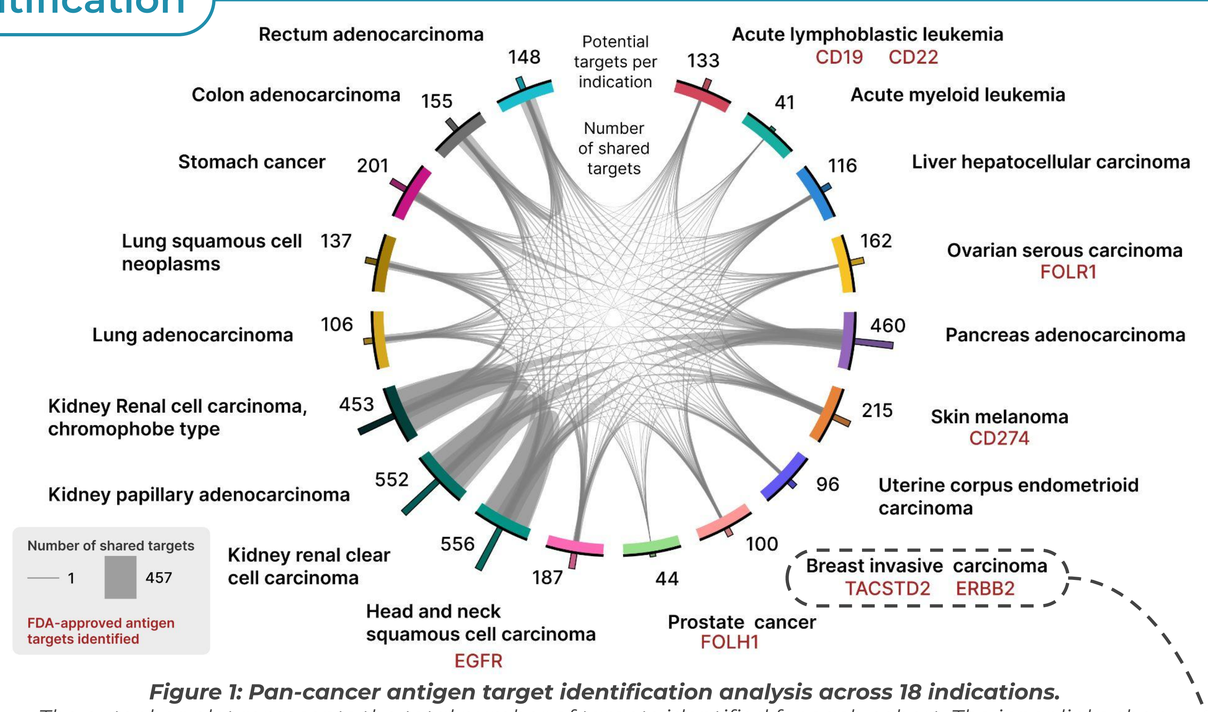


Figure 1: Pan-cancer antigen target identification analysis across 18 indications.
The outer bar plot represents the total number of targets identified for each cohort. The inner links show the number of targets shared between pairs of cancer type, with FDA-approved targets highlighted in red.

Target selection

Efficacy score

Protein expression in tumor tissues	HPA
RNA expression in tumor tissues	TCGA
RNA expression in tumor tissues	Epigene Labs
RNA expression in cancer cell lines	CCLE
RNA expression in malignant cells	GEO
Identification in other cancer types	Epigene Labs
Mutation frequency in tumor tissues	TCGA
Correlation with patient outcome	TCGA
Role in hallmarks of cancer	MSigDB
Publications	PubMed

Safety score

Protein expression in vital healthy tissues	HPA, HPM, PDB
Protein expression in non-vital healthy tissues	HPA, HPM, PDB
RNA expression in non-immune cells	Tabula Sapiens
RNA expression in healthy tissues	GTEx
LogFC between healthy/tumor tissues	Epigene Labs

Novelty score

Publications about targeted therapies in all cancers	PubMed
Publications about targeted therapies in specific cancer	PubMed

Weighted scoring system → Candidate targets

- We develop a weighted scoring system based on efficacy, safety, and novelty metrics, trained and calibrated using FDA-approved targets.
- Our score can also predict potential on-target off-tumor toxicities.

- On average, 8 candidates per indication exceeded our FDA-approved target benchmarks for efficacy and safety.

100 **Identified targets** for **breast invasive carcinoma**

31 Targets meeting or exceeding FDA-approved **efficacy** benchmarks

7 Highly promising targets with superior **safety** indicators

2 Candidate targets with high **novelty**

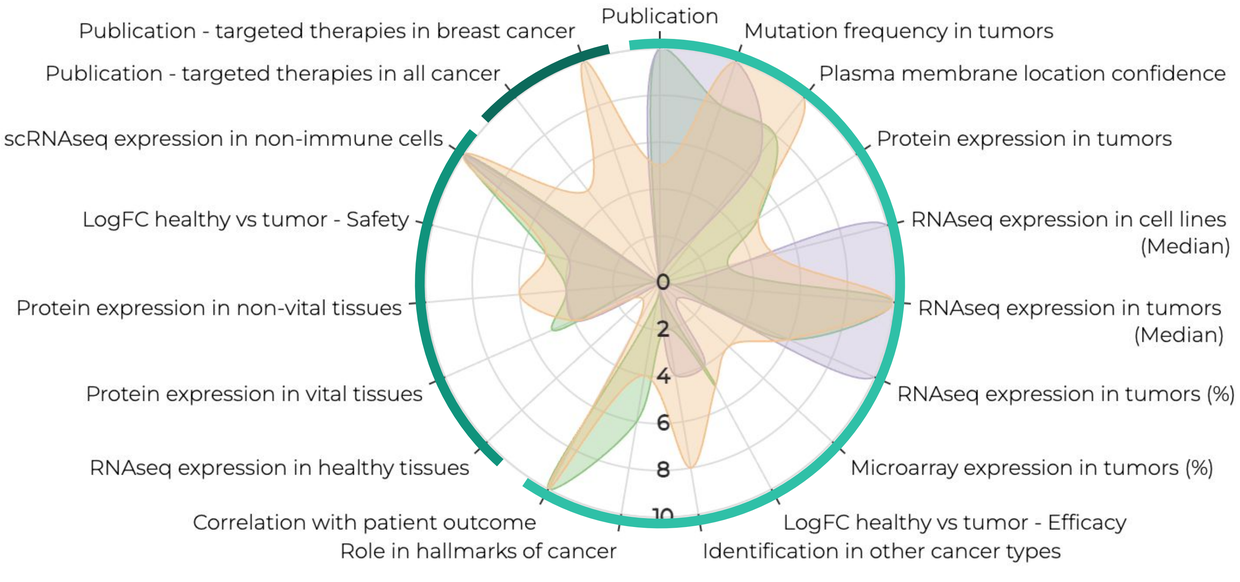


Figure 2: Scoring of one top candidate target and FDA-approved targets in breast invasive carcinoma. Radar plot showcasing the data integrated into our framework to compute efficacy, safety and novelty score.

References

1. Minikel, E.V., Painter, J.L., Dong, C.C. et al. Refining the impact of genetic evidence on clinical success. Nature 629, 624–629 (2024).
2. Behdenna A *et al. pyComBat, a Python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods*. BMC Bioinformatics. 2023 Dec 7; 24, 459.
3. Differential Expression Analysis with InMoose, the Integrated Multi-Omic Open-Source Environment in Python. BMC Bioinformatics. Accepted for publication, 2025. *https://doi.org/10.1101/2024.11.14.623578, 2024.*

Contact

Akpéli Nordor, PharmD, PhD
(akpeli@epigenelabs.com).
The authors have no conflict of interest to declare.

CHECK US OUT!