Accelerating antigen-targeting therapy discovery with a scalable pan-cancer bioinformatics platform

Abstract 2982 E. Fox[1], L. Meunier[1], G. Appe[1], A. Behdenna[1], L. Hensen[1], C. Lafond[1], A. Nordor[1], S. Weill[1], C. Marijon[1]. [1]Epigene Labs, Paris, France

AACR American Association for Cancer Research

Introduction

- **Antigen-targeting therapies** have demonstrated significant clinical success in hematologic malignancies, due to their high specificity and therapeutic effectiveness.
- Nearly 90% of clinical trials fail, frequently due to efficacy and safety concerns, underscoring the urgent need to **discover novel, safe, and effective antigen targets**.
- **Data-driven approaches** have been estimated to increase the probability of success of future cancer therapies in clinical development by up to 3x[1].
- We developed a **comprehensive and scalable translational bioinformatics platform** (Fig. 1), which leverages underutilized public omic data to discover and prioritize novel antigen targets across various cancer types.

Conclusion

- Our platform integrates **unbiased data-driven tools with cancer biology insights to streamline antigen target discovery**, from data integration to target selection.
- Scalable to any cancer type or antigen-targeting modality, it offers a robust framework for accelerating oncology drug discovery.

References

1. Nelson M et al. *The support of human genetic evidence for approved drug indications.* Nature Genetics. 2015 Aug;47(8):856-60.
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. Colange M et al. *Differential Expression Analysis with InMoose, the Integrated Multi-Omic Open-Source Environment in Python. [Preprint]. 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!

Methods

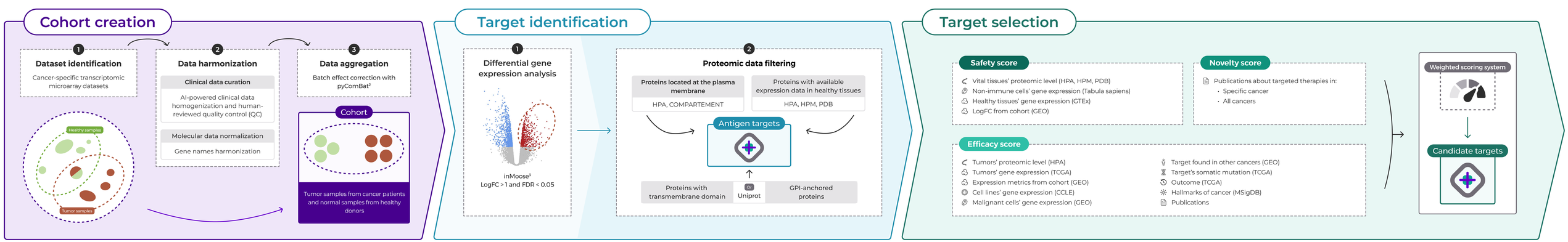

Figure 1: Epigene Labs' antigen target discovery platform
(LogFC = Log Fold Change, FDR = False Discovery Rate, HPA = Human Protein Atlas, HPM = Human Proteome Map, PDB = Proteomics DB, GTEx = Genotype-Tissue Expression, TCGA = The Cancer Genome Atlas , GEO = Genome Expression Omnibus, MSigDB = The Molecular Signatures Database, CCLE = Cancer Cell Line Encyclopedia, FDA = Food and Drug Administration)

Results

Cohort creation

168 Datasets	18 Cancer type	45 Clinical data elements	7,885 Tumor samples	1,569 Healthy samples	20,347 Genes

- Individual GEO datasets average 60 samples, whereas our cohorts (Fig. 2) surpass this across all cancer types, with an average of 556 samples (over a 9-fold increase) —substantially **enhancing statistical power** and improving **population representativeness.**

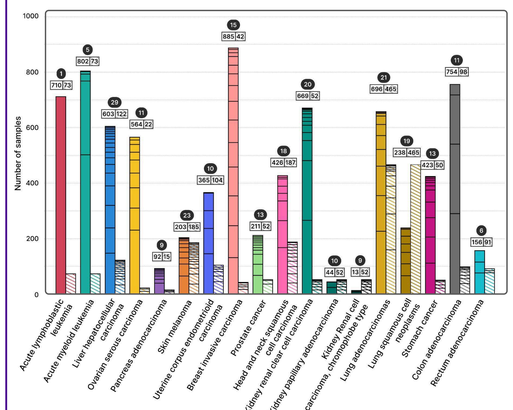

Figure 2: Overview of our harmonized and aggregated indication-specific cohorts.
Numbers of samples and datasets for each indication are indicated at the top of the bars, and portions in each bar correspond to individual datasets used to create the cohort.

Target identification

- On average, **216 potential antigen targets** were identified per indication, ranging from 41 to 556 (Fig. 3).
- A total of 952 targets were shared across 2+ indications, highlighting the **tissue-agnostic potential** of certain antigen targets.
- We successfully identified **8 FDA-approved antigen targets**, along with several investigational antigen targets.

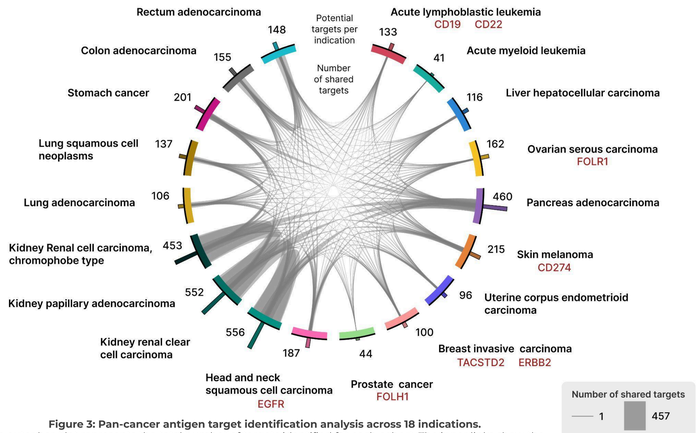

Figure 3: Pan-cancer antigen target identification analysis across 18 indications.
The outer bar plot represents the total number of targets identified for each cohort. The inner links show the number of targets shared between pairs of indications, with FDA-approved targets highlighted in red in the corresponding indications.

Target selection

- **Pan-cancer FDA-approval benchmarks** were established based on all FDA-approved targets for antigen-targeting modalities in relevant indications.
- On average, **nearly 8 candidate targets** per indication met or exceeded FDA-approval benchmarks for **efficacy** and **safety.**
- Around 50% of those also ranked high in **novelty** in every indication.
- Representative results for **breast invasive carcinoma** are shown in Fig. 4.

100 **Identified targets** for **breast invasive carcinoma**
- **31** Targets meeting or exceeding FDA-approved **efficacy** benchmarks
- **7** Highly promising targets with superior predicted **safety**
- **2** Candidate targets with high **novelty**

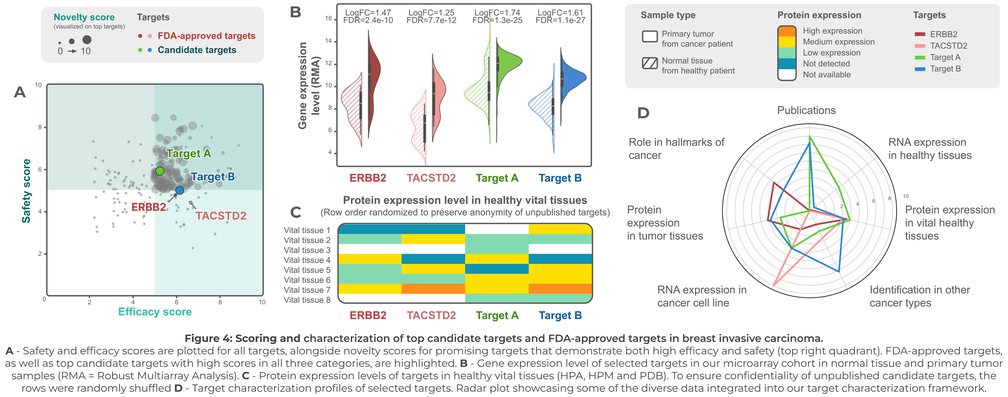

Figure 4: Scoring and characterization of top candidate targets and FDA-approved targets in breast invasive carcinoma.
A - Safety and efficacy scores are plotted for all targets, alongside novelty scores for promising targets that demonstrate both high efficacy and safety (top right quadrant). FDA-approved targets, as well as top candidate targets with high scores in all three categories, are highlighted. **B** - Gene expression level of selected targets in our microarray cohort in normal tissue and primary tumor samples (RMA = Robust Multiarray Analysis). **C** - Protein expression levels of targets in healthy vital tissues (HPA, HPM and PDB). To ensure confidentiality of unpublished candidate targets, the rows were randomly shuffled **D** - Target characterization profiles of selected targets. Radar plot showcasing some of the diverse data integrated into our target characterization framework.

Co-funded by the European Union
European Innovation Council