Indian Journal of Biochemistry & Biophysics
Vol. 63, February 2026, pp.137-144
DOI:10.56042/ijbb.v63i2.22387

Revolutionizing drug discovery in lung cancer: An artificial intelligence (AI)-assisted framework for identifying target antigens for antibody-drug conjugates

Christanto A[1,2]*, Setyawan UA[1,3] & Chozin IN[1,3]

[1]Department of Pulmonology and Respiratory Medicine, Faculty of Medicine, Brawijaya University, Malang, Indonesia
[2]Brawijaya University Hospital, Malang, Indonesia
[3]Department of Pulmonology and Respiratory Medicine, Saiful Anwar General Hospital, Malang, Indonesia

Received 04 August 2025; *revised 27 August 2025*

Identifying appropriate target antigens continues to be a hindrance in the development of antibody-drug conjugates (ADCs), particularly for lung adenocarcinoma (LUAD). This paper presents a rule-based, Artificial Intelligence (AI)-assisted system that automates the processes of data harmonization, filtering, and prioritization within extensive transcriptome datasets. The TCGA-LUAD (20,530 genes) and GTEx lung (57,233 genes) datasets were harmonized; protein-coding, surface-localized molecules were evaluated for differential expression (Δlog2), housekeeping or essentiality, projected internalization, solubility, and subcellular accessibility. We ranked the candidate molecules by using a composite score, combining normalized Δlog2 and internalization category. We then curated the ranked molecules for their function, relevance to cancer, tissue specificity, and translational feasibility. These processes results in 647 high confidence surface molecule candidates. Several recognized ADC targets (CEACAM5, MET, LRRC15, MUC16) were included in this list, supporting internal validity. Five antigens (PROM2, DSG2, SEZ6L2, CDH3, and CDCP1) met the quantitative thresholds and translational criteria, with commercial antibodies available for testing. Thus, this reproducible, scalable workflow may reduce subjective bias, clarify decision logic, and offer a general template for antigen discovery in the oncology settings. We believe that this combination of scalability, automation, standardization and validation represents a substantial step compared to conventional expert-curated, manually filtered workflows.

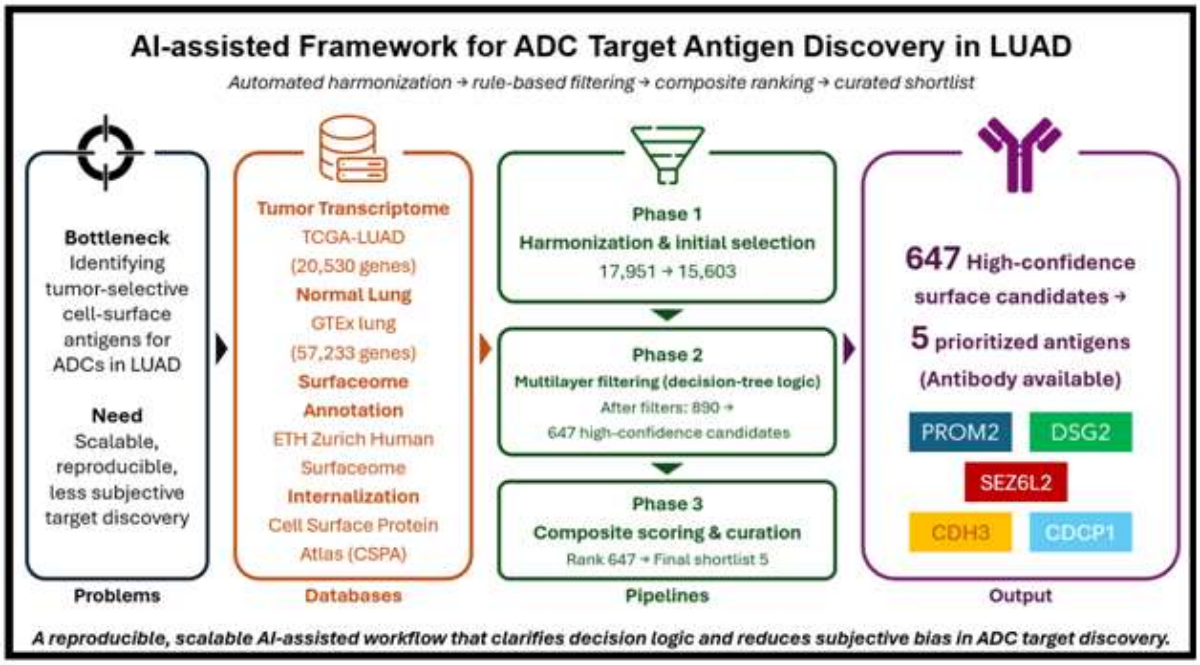
Keywords: Antibody-drug conjugates, Artificial intelligence, Bioinformatics pipeline, Lung adenocarcinoma, Translational oncology

Drug discovery is a time-consuming process, often requires intensive resources and significant effort even with current remarkable developments in precision medicine. Among these processes, finding suitable molecular targets with decent disease specificity, relevance of function, and translational feasibility may prove to be a major bottleneck[1]. Conventional approaches to this target discovery relies mainly on manual, expert-dependent analysis of existing biological data, which is prone to subjectivity, inconsistency, and limited scalability. A systematic, automated method is needed to accelerate and standardize this process across diseases contexts, as the rising complexity of omics-based datasets demands the rise of rapid therapeutic innovation. This is particularly evident in the development of targeted therapies, such as the newly emerging antibody drug conjugates (ADCs), where choosing ideal molecular targets is both a vital and sophisticated process[2,3].

Antibody-drug conjugates (ADC), a new class of drugs combining the selectivity of monoclonal antibodies with the potent cell toxicity of small molecule drugs, represent a new and rapidly emerging class of targeted therapy in oncology. The identification of highly expressed cell surface antigens in tumor tissues that show little expression in normal cells will critically determine the success of ADCs[4,5]. However, identifying perfect target antigens presents major difficulties, especially in malignancies like lung adenocarcinoma (LUAD) where tumor heterogeneity and off-target toxicity risks may complicate the process[6]. Although various ADCs for lung cancer have been developed, there is a significant need of methodical, reproducible methods to identify novel, high-value target antigens.

Conventional methods of target antigen discovery usually consists of expert-driven prioritization, manual literature curation, and sequential data filtering[7]. Although these approaches have produced significant results in therapeutic targets, they are naturally limited by human bias, inconsistency in decision-making, and scaling challenges to the vast complexity of modern

*Correspondence:
E-mail: anthony.chrismd@gmail.com



Graphical abstract

transcriptome and proteomic datasets[8]. Manual processes, with limited reproducibility and transparency, may complicate the process to translate potential targets into clinically active candidates. Thus, the potential of Artificial Intelligence (AI)-based and computational frameworks to automate data integration, standardize filtering criteria, and improve the scalability and reproducibility of this process sparks a growing interest, leading to a rapidly expanding interest of the use of AI in this field.

In this study, we designed and validated an AI-assisted, scalable framework intended for target antigen discovery in ADCs, using lung adenocarcinoma (LUAD) as a case study. Our methodology will demonstrate the efficient discovery and prioritization of candidate target antigens through the integration of automated data preprocessing, biologically-validated rule-based filtering, and AI-assisted decision-making. Our methodology will mitigate significant limitations of traditional methodologies by limiting human participation, hence lowering subjective bias and enhancing repeatability, thereby showcasing its potential for extensive application in next-generation drug development pipelines across diverse domains.

Material and Methods

Study design and AI-assisted discovery framework

We employ a multi-phased, AI-assisted discovery framework aimed at identifying possible cell surface antigens appropriate for ADC production in LUAD. We incorporated extensive transcriptome data, rule-based bioinformatics filtration, and expert curation inside this system. AI-assisted decision making was employed to automate data preprocessing, standardize rule application, and minimize subjective bias, while expert oversight was permitted whenever deemed necessary. Custom Python scripts which facilitated data cleaning, filtering, scoring, and candidate ranking were employed using AI-informed logical hierarchies. Manual interventions were minimized and documented as transparent as possible.

Phase 1: Data acquisition, Preprocessing, and Over expression analysis

We obtained transcriptomic data from TCGA-LUAD HiSeqV2 dataset (LUADs) and the GTEx Lung cohort (normal lung tissues). AI-assisted preprocessing scripts were used to harmonize datasets, eliminate unannotated and duplicate genes, and replace placeholder values. Common genes between both datasets were identified, and non-protein-coding genes were excluded. The remaining genes were cross-referenced with the ETH Zurich Human Surface me Database to identify those encoding predicted surface proteins. For every remaining gene in both normal lung tissues and LUAD, mean $\log_2$ expression values were calculated. Differential expression ($\Delta \log_2$) was calculated, and genes exceeding a defined over expression threshold were kept for further analysis.

Phase 2: Multilayered filtering with AI logic

In this phase, we utilize an AI-assisted decision tree that ran through a series of following biologically relevant filters:

- Homogeneity and Normal Tissue Expression: Low expression genes in normal tissues and low-consistency genes across LUAD samples were excluded to avoid the elimination of biologically viable yet heterogenous or moderately expressed target genes.
- Housekeeping and Essential Genes: Automated scripts were deployed to exclude housekeeping genes and essential, critical biological pathways genes, *i.e*. cell cycle regulation and DNA replication.
- Internalization Potential: Cell Surface Protein Atlas (CSPA) was used to evaluate potential candidate genes' internalization potential, prioritizing high predicted internalization rates proteins.
- Solubility and Subcellular Localization: Based on UniProt annotations and automated text parsing, we excluded genes which encodes proteins with high extracellular solubility or localization incompatible with ADC targeting.

Throughout this step, Uniformity of rules and minimized subjective in decision-making was ensured using AI logic to maintain transparent and reproducible decision pathways.

Phase 3: Candidate prioritization and expert curation

We produce a composite priority score for every gene using AI-automated ranking system which combined internalization predictions and normalized differential expression values ($\Delta \log_2$). This scoring system allocated equal weight to both criteria, aiming to balance the accessibility of predicted target antigens with the magnitude of expression. $\Delta \log_2$ values and internalization predictions were normalized to a scale of 0–1 (High = 1, Medium = 0.5, Low/Unknown = 0). The overall composite score was determined by summing these two values, with a maximum attainable score of 2. The highest-ranked candidates underwent both manual and AI-assisted curation, encompassing gene function, relevance to cancer, tissue specificity, translational potential, and availability of commercial antibodies. This hybrid strategy guaranteed that the final candidate molecule selection process met both quantitative and qualitative scoring criteria.

Among all these processes, AI assisted in the decision-making process by enhances data preprocessing, filtration, and ranking automation, and also to reduce human bias and enhance the reproducibility and adaptability of the process through semi-automated scripting that permitted expert intervention. It also improves scalability for potential application for other datasets and other cancer types, and finally transparency using an encoded decision logic in audible Python scripts. This framework balances computational efficiency with biological insight, offering a reproducible, scalable model for future ADCs target discovery and broader drug development initiatives and pipelines.

Software, Database, and Computational Tools

We use the following tools to preprocess, filter, and score our data: Python version 3.9, withPandas (version 1.5), NumPy (version 1.24), and SciPy (version 1.10) libraries; Surface protein annotations from the ETH Zurich Human Surface me Database (Bausch-Fluck *et al*., 2018; https://wlab.ethz.ch/surfaceome/table_S3_surfaceome.xlsx); Protein coding gene annotations, cross-referenced from UniProt (release 2025_02; https://www.uniprot.org/); and internalization data obtained from the Cell Surface Protein Atlas (CSPA) (Bausch-Fluck *et al*., 2015; https://wlab.ethz.ch/cspa/data/S2_File.xlsx). All custom scripts were developed in Python, and are available upon request for reproducibility purpose for the framework.

Results

Phase 1: Data acquisition, Preprocessing, and Initial candidate selection

We started with transcriptome datasets which include 20,530 genes from the TCGA-LUAD HiSeqV2 dataset and 57,233 genes from the GTEx Lung cohort. Placeholder values in the GTEx dataset were identified and converted toNaN, without excluding any gene. Initial data cleaning eliminated unannotated and duplicate genes, therefore reducing the GTEx dataset to 55,544 genes, while all LUAD genes remained.

17,951 genes common to both datasets were identified using gene intersection and were subsequently filtered for non-protein-coding genes, which results in 16,314 candidates. These genes were further filtered to 16,246 genes surface protein-encoding genes, with the help of ETH Zurich Human Surface me Database. AI-assisted scripts were used to calculate mean $\log_2$ expression levels for each gene over normal lung and LUAD samples. Afterwards, differential expression analysis ($\Delta \log_2$) was conducted, and an overexpression threshold ($\Delta \log_2 > 2$) was applied. This step resulted in 15,603 genes,which exhibited significant over expression in LUADsuitable for further analysis.

Phase 2: Multilayered filtering and candidate refinement

These 15,603 surface-localized, LUAD-over expressed genes first underwent a thorough series of AI-assisted filtering procedures to narrow down the candidate pool. The first step of filtration was the removal of eight housekeeping genes; 331 genes essential to critical biological functions, including mitosis, RNA processing, ribosome biogenesis, and cellular respiration were identified and excluded, which results in 15,264 candidates. Afterwards, internalization potential, a critical determinant of ADC efficacy, was examined by cross-referencing the dataset with the Cell Surface Protein Atlas (CSPA).Proteins with medium and high predicted internalization rate were identified, thereby further reduces the pool to 890 potential genes.

These genes were first subjected via a series of AI-assisted filtering to narrow down the choices. The first phase comprising the elimination of 8 housekeeping genes, and 331 genes important to critical biological operations, including mitosis, RNA processing, ribosome synthesis, and cellular respiration were discovered and removed, resulting in 15,264 candidates. Then, we analyse for internalization potential, which is critical to ADC efficacy, by cross-referencing the dataset with the Cell Surface Protein Atlas (CSPA). Proteins with medium-to-high predicted internalization rate were identified, narrowing the pool to 890 candidate genes.

Afterwards, we delete genes with significant extracellular solubility or shedding risk, decreasing 137 candidates from the pool. Finally, an aggressive final filter was utilized to eliminate genes with subcellular localizations genes likely inaccessible for ADC targeting, including mitochondrion, lysosome, endoplasmic reticulum, Golgi bodies, junctional complex and extracellular matrix proteins. This stage resulted in 647 remaining genes, designated as high-confidence candidates eligible for further processing(Table 1).

Phase 3: Candidate prioritization, biological curation, and Final shortlisting

In this phase, the candidate genes were subjected to prioritizing and biological curation. First, the genes were sorted using an AI-based system, which created a "composite priority score" by combining the normalized differential expression value ($\Delta \log_2$) with estimated internalization potential. This procedure yielded a table of applicants ranked by their suitability as ADC targets. From this list, the top 15 genes were then extracted for extensive biological curation. Each of these genes were appraised under five criteria: gene function, relevance to cancer, tissue specificity, uniqueness (in context of ADV development research), and overall biological feasibility. This process guaranteed that the picked candidates satisfied the requisite standards, combining AI data aggregation with manual expert evaluation to ensure rigorous clinical relevance.

Evaluation of readily-available commercial antibodies for each of top-ranking genes were also done as part of evaluation of translational readiness. Candidates without commercially accessible antibodies were deprioritized to permit further experimental validation(Table 2).

The $\Delta \log_2$ expression and anticipated internalization-based sequencing provide a quantitative basis for prioritizing candidates, while the ultimate selection process was communicated through comprehensive biological curation to guarantee suitable translational relevance and practicality. Notably, several high-ranking genes including CEACAM5, MET, LRRC15, and MUC16 have been either validated previously or are under investigation as ADC targets, which further validates the accuracy of AI-assisted framework.

Table 1 — Summary of Gene Filtering and Candidate Selection Across the AI-Assisted Framework.

Step	Description	Genes Retained
Raw Data Acquisition	GTEx Lung (normal): 57,233; TCGA-LUAD (tumor): 20,530	57,233 / 20,530
Data Cleaning	Removal of unannotated and duplicate genes	55,544 / 20,530
Gene Intersection	Genes common to both datasets	17,951
Protein-Coding Filter	Retained protein-coding genes	16,314
Surface Protein Filter	Based on ETH Zurich Human Surfaceome	16,246
Over expression Thresholding	$\Delta \log_2 > 2$	15,603
Housekeeping Gene Removal	Automated exclusion	15,595
Essential Pathway Exclusion	Vital cellular process genes removal	15,264
Internalization Potential	High or medium predicted internalization	890
Solubility & Localization Filtering	Excluded soluble or inaccessibly localized proteins	647
AI-Assisted Ranking	Based on $\Delta \log_2$ and internalization	647
Manual Biological Curation	Evaluation of top 15 candidates	15
Final Shortlist	Prioritized novel ADC targets	5

Table 2 — Top 15 Ranked Candidate Antigens from the AI-assisted Framework

Rank	Gene	Δ $\log_2$ Expression	Internalization	Antibody Availability	Notes
1	ELFN2	11.18	High	Yes	Excluded: low relevance
2	CEACAM5	10.25	High	Yes	Known ADC target (Tusamitamabravtansine) (*9*)
3	DSG2	9.66	High	Yes	Novel, shortlisted final 5
4	CR2	9.47	High	Yes	Novel candidate
5	LRRC15	9.43	High	Yes	Investigational ADC target (ABBV-085) (*10*)
6	TMEFF1	9.19	High	No	Excluded: low relevance
7	MET	9.16	High	Yes	Known ADC target (Telisotuzumab vedotin) (*11*)
8	PROM2	8.96	High	Yes	Novel, shortlisted final 5
9	SEZ6L2	8.88	High	Yes	Novel, shortlisted final 5
10	CDH3	8.72	High	Yes	Novel, shortlisted final 5
11	FREM2	8.63	High	No	Excluded: low relevance
12	CDH17	8.43	High	Yes	Novel candidate
13	MUC16	8.34	High	Yes	Investigational ADC target (DMUC5754A) (*12*)
14	CNTNAP2	8.28	High	No	Excluded: low relevance
15	CDCP1	8.24	High	Yes	Novel, shortlisted final 5

Table 3 — Summary of Final Shortlisted ADC Target Candidates

Gene	Δ $\log_2$ Expression	Internalization	Antibody Availability	Biological Rationale	Translational Feasibility
PROM2	8.96	High	Yes	Overexpressed in LUAD Membrane localization Associated with cancer proliferation	No prior ADC target Novel candidate Antibody available
DSG2	9.66	High	Yes	Cell adhesion molecule Overexpressed in LUAD Correlates with poor prognosis	No prior ADC target Strong literature support Antibody available
SEZ6L2	8.88	High	Yes	Cell surface protein Overexpressed in LUAD and other tumors	Novel ADC candidate Antibody available
CDH3	8.72	High	Yes	Cadherin family Implicated in cancer invasion and metastasis Overexpressed in LUAD	Investigational ADCs in other cancers Antibody available
CDCP1	8.24	High	Yes	Known cancer-associated antigen Overexpressed in LUAD Prior CAR-T and ADC studies	High translational feasibility Antibody available

From the top 15 candidates, 5 genes were ultimately shortlisted for further analysis, which include PROM2, DSG2, SEZ6L2, CDH3 and CDCP1. These selected genes had a combination of high differential expression, strong predicted internalization, and commercial antibody availability. Beyond numerical ranking, biological relevance played a critical role in candidate selection. For instance, CR2 (CD21) was ranked high in expression, but was excluded due to its primary expression in immune cells (B and T lymphocytes)[13], which increasing off-target toxicity risk. Another highly ranked gene, CDH17, was excluded due to its limited, inconsistent expression in LUAD and low translational feasibility. However, CDCP1, despite being slightly lower ranked, had stable overexpression in LUAD, well-documented literature relevance, proven availability of antibodies, and has been previously investigated as therapy target in several studies[14,15].

This hybrid methodology, integrating AI-assisted quantitative ranking with expert-led biological curation, guaranteed that the final shortlist emphasized both statistical significance and biological feasibility, thereby optimizing the likelihood of successful translational application.

Importantly, there were several genes in the top 15 table that were already validated for investigational ADC targets, including CEACAM5, MET, LRRC15, and MUC16. Their position among the top-ranking genes provides independent support of this framework's robustness and biological relevance. Their inclusion also illustrates the framework's success in identifying both established and novel ADC target antigens, alongside the five unique candidates (PROM2, DSG2, SEZ6L2, CDH3, and CDCP1) obtained from the framework's pipeline approach.

Each of the final choices was analysed, not only for quantitative ranking criteria but also for biological plausibility and translational feasibility (Table 3). The genes PROM2, DSG2, SEZ6L2, and CDH3 are novel, or at least underexplored targets, with solid biological

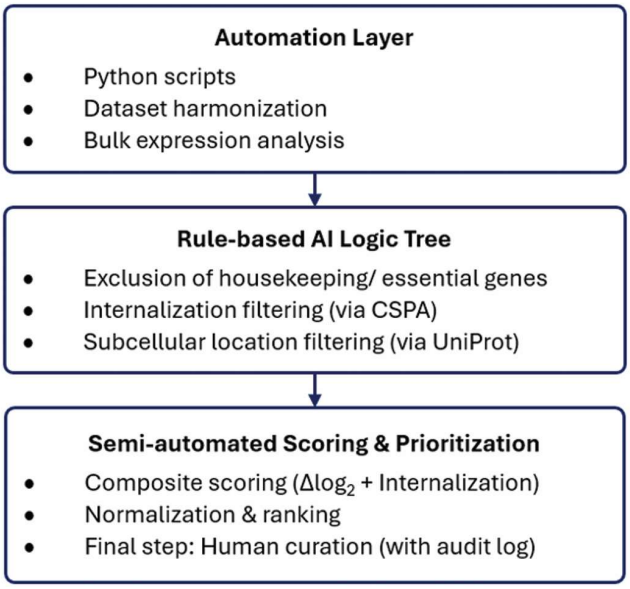


Fig. 1 — Conceptual structure of the AI-assisted antigen discovery framework

rationale and availability of commercial antibodies. These reasons positioned them as attractive candidates for further ADC development. Being previously examined in other types of malignancies, CDCP1 also provides a verified benchmark, underlining the framework's capability to reveal therapeutically relevant targets alongside new discoveries(Table 3).

Discussion

To date, identification and validation of appropriate target antigens is still a major and basic issue in formation of ADCs. This is especially crucial of solid tumors (e.g. LUAD) that exhibit pronounced heterogeneity[4,16]. Traditional methods for doing so are based on manual, human curation, and expert-driven analysis that lack of objectivity bias, scalability as well as reproducibility. We report the successful development and validation of an AI-supported pipeline to simplify, streamline and standardize target discovery in a manner that does not sacrifice biological subtlety thanks to expert curation. Applying this framework to LUAD, we identified five such (novel or relatively under-investigated; potential target candidate) antigens for ADCs and also observed that the framework is robust by identifying multiple known targets of ADCs(Fig. 1).

The main novelty of this framework is its hybrid character, mixing an automated, rule-based filtering with a human-driven flexible manual verification. Automation reduced subjective bias and increased capacity, such that very large-scale data sets (in excess of 57,000 genes) could be analysed across large transcriptomic datasets. The AI-guided decision tree enables consistent use of complex biological filters (expression specificity, essentiality, internalization potential and accessibility). This combination of filters brought the number of potential targets down from more than 15,000 genes to 647 high-confidence targets.

The normalized differential expression-based ranking scheme was combined with an internalization model to objectively rank candidates that also reflected their physiological context, but could be overridden by expert curation. Notably, identification of previously known targets for ADCs (e.g., MET and CEACAM5) supported the accuracy and translational potential of this approach[17–19]. This benchmark validation demonstrates the ability of the framework to discover clinically relevant determinants with innate action ability.

Each of the five shortlisted genes (PROM2, DSG2, SEZ6L2, CDH3, and CDCP1) demonstrating a combination of high differential expression, predicted internalization, and availability of commercial antibody. Notably, while CDCP1 has been utilized in preclinical ADC development and CAR-T studies[20], the remaining four genes were novel or underexplored antigens, each with compelling biological rationales. Their identification underscores the framework's potential to facilitate both rediscovery of previously validated targets and unveiling of novel therapeutic insights. The scalability of this framework brings its potential not only for broad application across diverse cancer types beyond lung cancer, but also other therapeutic modalities involving target discovery beyond ADCs. The framework's modular design allows the incorporation of additional data layers, including proteomic profiles and single cell sequencing, further enhancing the precision and adaptability of the process.

Each one of the five shortlisted genes (PROM2, DSG2, SEZ6L2, CDH3 and CDCP1) exhibited a combination of high differential expression, predicted internalization and commercial antibody availability. It is of interest that though CDCP1 had previously been exploited for preclinical-established ADC and CAR-T studies[20], the other four genes were novel or little-studied antigens with strong biological rationale. This discovery highlights the framework's utility for rediscovery of known drug targets and uncovering of new therapeutic insights. The scalability of this

platform takes its potential not only to other areas beyond lung cancer, but also other therapeutics method where target discovery holds an important aspect. The module-based architecture of the framework also makes it possible to include additional data layers, such as proteomic profiles and single cell sequencing, making it more precise and versatile for future iterations.

Nevertheless, the framework is not without limitations. The potential differences between RNA expression and protein surface presentation based on the reliance of transcriptomic data. While this risk could be negated with cross-referencing of Human Surface me and UniProt databases, a future version should aim to incorporate proteomic validation to more definitively verify antigen accessibility. Additionally, although the framework minimizes the need for manual curation, expert guidance is an integral part of it, particularly towards the end of curation after underlying biological complexity may not be appropriately addressed by purely computational means.

Although our framework does not utilize traditional machine learning (ML) or deep learning (DL) tools, we used the label "AI-assisted" to indicate it incorporates automatic decision-making (implemented by logic trees and enforcing rules straightly), high-throughput filtering data selecting. This is mainly based on the idea of advanced system and symbolic AI that transparent decision-making logic can replace large scale human subjectivity. Even though no predictive ML models were implemented, the framework accomplishes key objectives for AI in biomedical research such as reducing overall bias and bringing reproducibility and scalability of decision across complex, large volume datasets. For future versions, supervised ML models could also be combined with the presented method to improve scoring and detection of non-linear relationships in antigen prioritization.

Conclusion

Our work offers an independent, scalable and biologically plausible AI-based framework for ADC target discovery in LUAD. This framework translates computational rigor into the actual context of biological expertise, providing a plausiblemethodto speed up the discovery of potential targetantigens. The successful application of this pipeline and its validation through the identification of previously known ADC targets establish a foundation for future applications across the oncology drug discovery landscape.

Conflict of interest

All authors declare no conflict of interest.

References

1 MacRae CA, A new phenotypic lexicon for accelerated translation. *Circulation*, 131 (2015) 234.

2 Damelin M, Zhong W, Myers J & Sapra P, Evolving strategies for target selection for antibody-drug conjugates. *Pharm Res*, 32 (2015) 3494.

3 Boni V, Sharma MR & Patnaik A, The resurgence of antibody drug conjugates in cancer therapeutics: novel targets and payloads. *Am Soc Clin Oncol Educ Book*, 40 (2020) e58.

4 Passaro A, Jänne PA & Peters S, Antibody-drug conjugates in lung cancer: recent advances and implementing strategies. *J Clin Oncol*, 41 (2023) 3747.

5 Merle G, Friedlaender A, Desai A & Addeo A, Antibody drug conjugates in lung cancer. *Cancer J*, 28 (2022) 429.

6 Coleman N, Yap TA, Heymach JV, Meric-Bernstam F & Le X, Antibody-drug conjugates in lung cancer: dawn of a new era?*NPJ Precis Oncol*, 7 (2023) 5.

7 Feola S, Chiaro J, Martins B & Cerullo V, Uncovering the tumor antigen landscape: what to know about the discovery process. *Cancers (Basel)*, 12 (2020) 1660.

8 Mirandola L, Marincola F, Rotino G, Figueroa JA, Grizzi F, Bresalier R, Chiriva-Internati M, The quest for the next-generation of tumor targets: discovery and prioritization in the genomics era. In: Tan SL, editor. *Methods in Pharmacology and Toxicology* (Springer US, New York) 2020, 239.

9 Tabernero J, Bedard PL, Bang YJ, Vieito M, Ryu MH, Fagniez N, Chadjaa M, Soufflet C, Masson N, Gazzah A, Tusamitamabravtansine in patients with advanced solid tumors: phase I study of safety, pharmacokinetics, and antitumor activity using alternative dosing regimens. *Cancer Res Commun*, 3 (2023) 1662.

10 Demetri GD, Luke JJ, Hollebecque A, Powderly JD, Spira AI, Subbiah V, Naumovski L, Chen C, Fang H, Lai DW, Yue H, First-in-human phase I study of ABBV-085, an antibody-drug conjugate targeting LRRC15, in sarcomas and other advanced solid tumors. *Clin Cancer Res*, 27 (2021) 3556.

11 Camidge DR, Bar J, Horinouchi H, Goldman J, Moiseenko F, Filippova E, Cicin I, Ciuleanu T, Daaboul N, Liu C, Bradbury P, Telisotuzumab vedotin monotherapy in patients with previously treated c-Met protein-overexpressing advanced nonsquamous EGFR-wildtype non-small cell lung cancer in the phase II LUMINOSITY trial. *J Clin Oncol*, 42 (2024) 3000.

12 Liu JF, Moore KN, Birrer MJ, Berlin S, Matulonis UA, Infante JR, Wolpin B, Poon KA, Firestein R, Xu J, Kahn R, Phase I study of safety and pharmacokinetics of the anti-MUC16 antibody-drug conjugate DMUC5754A in patients with platinum-resistant ovarian cancer or unresectable pancreatic cancer. *Ann Oncol*, 27 (2016) 2124.

13 Smith NA, Coleman CB, Gewurz BE & Rochford R, CD21 (complement receptor 2) is the receptor for Epstein–Barr virus entry into T cells. *J Virol*, 94 (2020) e00428-20.

14 Khan T, Kryza T, Lyons NJ, He Y & Hooper JD, The CDCP1 signaling hub: a target for cancer detection and therapeutic intervention. *Cancer Res*, 81 (2021) 2259.

15 Khan T, Lyons NJ, Gough M, Kwah KK, Cuda TJ, Snell CE, Tse BW, Sokolowski KA, Pearce LA, Adams TE, Rose SE, CUB domain-containing protein 1 (CDCP1) is a rational target for the development of imaging tracers and antibody-drug conjugates for cancer detection and therapy. *Theranostics*, 12 (2022) 6915.

16 Esapa B, Jiang J, Cheung A, Chenoweth A, Thurston DE & Karagiannis SN, Target antigen attributes and their contributions to clinically approved antibody-drug conjugates in haematopoietic and solid cancers. *Cancers (Basel)*, 15 (2023) 1845.

17 Groisberg R, Rosznik J, Meric-Bernstam F & Subbiah V, Comprehensive evaluation of antibody-drug conjugate targets for drug development across tumor types: analysis from TCGA. *Cancer Res*, 78 (2018) 2343.

18 Decary S, Berne PF, Nicolazzi C, Lefebvre AM, Dabdoubi T, Cameron B, Rival P, Devaud C, Prades C, Bouchard H, Cassé A, Preclinical activity of SAR408701: a novel anti-CEACAM5–maytansinoid antibody–drug conjugate for the treatment of CEACAM5-positive epithelial tumors. *Clin Cancer Res*, 26 (2020) 6589.

19 Zanchetta C, De Marchi L, Macerelli M, Pelizzari G, Costa J, Aprile G, Cortiula F, Antibody-drug conjugates in non-small cell lung cancer: state of the art and future perspectives. *Int J Mol Sci*, 26 (2024) 221.

20 Hashimoto SI, Miyamoyo S, Takaiwa A, Nakamura K, Tsukumo Y, Yamashita Y, Antibody drug conjugate using a novel anti-CDCP1 antibody with low hematopoietic stem-cell binding and rapid internalization showed high efficacy for solid cancer. *Cancer Res*, 83 (2023) 438.

21 Bausch-Fluck D, Hofmann A, Bock T, Frei AP, Cerciello F, Jacobs A, Moest H, Omasits U, Gundry RL, Yoon C, Schiess R, A mass spectrometric-derived cell surface protein atlas. *PLoS One*, 10 (2015) e0121314.

22 Bausch-Fluck D, Goldmann U, Müller S, van Oostrum M, Müller M, Schubert OT, Wollscheid B, The in silico human surfaceome. *Proc Natl Acad Sci USA*, 115 (2018) e10988.

23 The GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, 369 (2020) 1318.

24 GDC TCGA Lung Adenocarcinoma (LUAD) cohort. *Xena Browser* [Internet]. 2020; [cited 2025 May 10]. Available from: https://xenabrowser.net/datapages/?dataset=TCGA-LUAD.