Journal for ImmunoTherapy of Cancer

Predictive value of preclinical models for CAR-T cell therapy clinical trials: a systematic review and meta-analysis

David Andreu-Sanz,[1] Lisa Gregor [iD],[1] Emanuele Carlini,[1] Daniele Scarcella,[2] Carsten Marr,[2] Sebastian Kobold [iD][1,3,4]

For numbered affiliations see end of article.

Correspondence to
Professor Sebastian Kobold;
sebastian.kobold@med.uni-muenchen.de

ABSTRACT

Background Experimental mouse models are indispensable for the preclinical development of cancer immunotherapies, whereby complex interactions in the tumor microenvironment can be somewhat replicated. Despite the availability of diverse models, their predictive capacity for clinical outcomes remains largely unknown, posing a hurdle in the translation from preclinical to clinical success.

Methods This study systematically reviews and meta-analyzes clinical trials of chimeric antigen receptor (CAR)-T cell monotherapies with their corresponding preclinical studies. Adhering to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, a comprehensive search of PubMed and ClinicalTrials.gov was conducted, identifying 422 clinical trials and 3,157 preclinical studies. From these, 105 clinical trials and 180 preclinical studies, accounting for 44 and 131 distinct CAR constructs, respectively, were included.

Results Patients' responses varied based on the target antigen, expectedly with higher efficacy and toxicity rates in hematological cancers. Preclinical data analysis revealed homogeneous and antigen-independent efficacy rates. Our analysis revealed that only 4% (n=12) of mouse studies used syngeneic models, highlighting their scarcity in research. Three logistic regression models were trained on CAR structures, tumor entities, and experimental settings to predict treatment outcomes. While the logistic regression model accurately predicted clinical outcomes based on clinical or preclinical features (Macro F1 and area under the curve (AUC)>0.8), it failed in predicting preclinical outcomes from preclinical features (Macro F1<0.5, AUC<0.6), indicating that preclinical studies may be influenced by experimental factors not accounted for in the model.

Conclusion These findings underscore the need to better understand the experimental factors enhancing the predictive accuracy of mouse models in preclinical settings.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Mouse models are indispensable for the development of immunotherapies, but their predictive capacity for clinical outcomes remains largely understudied.

⇒ In this meta-analysis, we challenged the relevance of mouse models for cellular therapies and evaluated to what extent preclinical models could predict the actual clinical response in chimeric antigen receptor (CAR)-T cell trials.

WHAT THIS STUDY ADDS

⇒ This study is the largest compiled analysis assessing the predictive value of CAR-T preclinical experiments to date.

⇒ It provides a comprehensive meta-analysis of the clinical efficacy of CAR-T cell monotherapy in hematological and solid malignancies, in parallel with an analysis of the matched preclinical data derived from mouse models.

⇒ Non-immunocompetent models skew efficacy and safety profiles and may not adequately predict clinical outcomes outside of hematology.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Via machine learning models, we identify preclinical features that may be useful to predict clinical outcomes in certain situations.

⇒ Non-immunocompetent mouse models have, if any, limited predictive power for clinical outcomes outside of hematology.

⇒ Future research should focus on developing syngeneic or humanized models with more robust immune systems to better mimic the human tumor microenvironment and improve clinical relevance.

⇒ Finally, our findings highlight the need for better transparency in the reporting of preclinical results, including negative and partial findings.

INTRODUCTION

In cancer immunotherapy, chimeric antigen receptor (CAR)-T cells are engineered with a synthetic receptor targeting tumor antigens, leading to T-cell activation and subsequent cancer cell killing.[1] Approved therapies, including Kymriah, Yescarta and Tecartus, have shown short and long-term success in hematological cancers,[2] where tumor cells are readily available in the bloodstream or bone marrow. Nevertheless, a substantial proportion of patients will still not benefit or only transiently respond,[3–5] while

Table 1 Overview of different mouse models used in preclinical cancer immunology research, including their advantages and disadvantages

	Advantages	Disadvantages
Syngeneic (transplanted tumor cell lines)	▶ Cost efficiency and time-saving model ▶ Possible to study on-target off-tumor effects ▶ Efficacy study concomitantly with native immune system ▶ Possibility of orthotopic implantation	▶ Very limited tumor heterogeneity ▶ Limited modeling of TME ▶ Limited comparison to clinical (human) setting ▶ Lack of human binder-mouse antigen cross-reactivity ▶ Artificial tumor location if implemented subcutaneously
Syngeneic (spontaneous induced cancer development)	▶ Cost-efficient model ▶ More realistic model of tumor development and progression ▶ Possible to study on-target off-tumor effects ▶ Efficacy study concomitantly with native immune system ▶ Potential orthotopic study of tumor model	▶ Variation in spontaneous mutation rates thus limited comparability ▶ Variation in onset and location of tumor development ▶ Time-consuming model due to tumor development ▶ Limited comparison to clinical (human) setting
Transgenic models	▶ Controlled genetic manipulation of the tumor ▶ Close modeling of human tumors ▶ Targeted insights regarding specific genes possible	▶ High maintenance due to advanced gene editing techniques ▶ Time-consuming model due to tumor development ▶ Possibility of unintended phenotypes occurring
Xenograft models	▶ Easy implementation and time-saving model ▶ Allows experimentation with human cancer cell lines and human T cells ▶ Suitable for studying human tumor biology	▶ Use of human cell lines limits the tumor heterogeneity ▶ GvHD ▶ High costs of immunodeficient and immunocompromised mice ▶ Loss of host immune system limits immune response studies
Patient-derived xenograft	▶ Use of patient-specific tumor cells increases clinical relevance ▶ Maintains tumor heterogeneity ▶ Can reflect patient-specific treatment responses (at least with regard to tumor cell-directed effects)	▶ GvHD ▶ Limited access to suitable patient samples ▶ High costs and time-consuming model

GvHD, graft-versus-host disease; TME, tumor microenvironment.

experiencing severe adverse events (AE).[6 7] Conversely, solid tumors possess physical barriers compromising accessibility of therapeutic cells, resulting in lower response rates.[8] Despite promising results for targets like human epidermal growth factor 2 (Her2),[9–11] mesothelin (MSLN),[12 13] and disialoganglioside (GD2),[14 15] clinical efficacy in solid tumors remains limited. We and others have identified access to cancer tissue, limited tumor cell recognition and immune suppression as key resistance mechanisms contributing to clinical inefficacy.[3 16 17]

Overall, translation of preclinical drugs to market authorization stands at a dramatic 1 in 10,000, with a phase I clinical trials' success rate below 5%.[18] Enhancing the predictive capacity of preclinical studies regarding their therapeutic window and future clinical efficacy is essential for increasing the productivity of drug development processes. For CAR-T cells, regulatory agencies mandate preclinical animal testing prior to clinical development[18] and, besides canine or non-human primate models, the majority of preclinical CAR-T cell development relies on mouse models,[19 20] making these a key stepping stone for decision-making during drug development (table 1).

Syngeneic models using spontaneous, induced, or transplantable tumors in mice with C57BL/6 or BALB/c backgrounds allow the study of adoptive transfer alongside a fully competent host immune system.[21 22] Despite their genomic homogeneity, these models face clinical comparability issues due to their artificial tumor microenvironment.[21 22] Subcutaneous implantation simplifies tumor measurements, but lacks the complexity of natural tumor growth. Orthotopic transplantation, on the other hand, may offer a more accurate tumor reflection, but requires training and complex surgical procedures.[23 24]

Human xenograft models consist of transplanting human tumor and immune cells into non-immunocompetent mice such as NOD/SCID/IL2Rγc-KO (NSG) and enable the study of human tumor-immune cell interactions.[25 26] Immunodeficient and immunocompromised mice pose challenges including graft-versus-host disease (GvHD), incomplete human immunity,[22 27 28] and the absence of human stromal cells, which impairs CAR-T cell preservation and limits the study of tumor-supporting stroma in

solid tumor therapies.[22 27 28] Patient-derived xenografts (PDX) replicate tumor cell-intrinsic features into non-immunocompetent mice, but typically have slow-growing tumors and lack human hematopoiesis, complicating immunotherapy evaluation.[29 30]

Despite their disadvantages, preclinical mouse models remain crucial for advancing therapies up to clinical testing. Proactive and comprehensive assessment of potential side effects and toxicities to reliably predict clinical efficacy and safety is imperative.[6 31]

Promising preclinical results, especially in solid tumors, often fail to translate into strong clinical responses. Limited understanding of the accuracy of these results in forecasting clinical responses in patients, coupled with a restricted therapeutic evaluation in a single model, could explain this gap. To gain a better understanding of the performance of preclinical models in CAR-T cell therapy development, we set out to perform a meta-analysis of all available clinical trials investigating CAR-T cell monotherapies together with their preclinical workup in mouse models. Using machine learning models, we sought to predict clinical trial outcomes based on preclinical data and identify potential factors that allow such forecasts.

MATERIALS AND METHODS
Information sources, search strategy, and data collection process

The clinical trial and preclinical records were sourced from PubMed and ClinicalTrials.gov until December 1, 2023, employing specific search criteria (online supplemental methods). These included all publications in English, excluding reviews, systematic reviews, meta-analyses, and retrospective studies. To prevent biases in the assessment, each entry was evaluated in all its aspects, including its fit with regard to inclusion/exclusion criteria and data extraction, by at least two reviewers (DA-S, LG, EC) independently.

Eligibility criteria and selection process
Clinical trial entries were excluded if they fulfilled at least one of the following criteria: non-cancer related, not CAR-T cell therapy, follow-up study, retrospective study, results not available, terminated for non-scientific reasons, not evaluating efficacy, combination therapy, case-report, reported somewhere else, incomplete data. Conversely, preclinical entries were excluded if fulfilling any of the following criteria: not original research article, non-cancer related, clinical study, not CAR-T cell therapy, non-classical CAR setting, target not in clinical setting, not about CAR efficacy, no animal models, follow-up studies, retrospective studies, combination therapy, incomplete data, studying irrelevant variables in the experimental set-up. An additional clarification of the exclusion criteria considered by the investigators as the most interpretable is provided in the online supplemental methods.

Table 2 Summary of all variables collected for clinical trials analyzed

Cancer type	Entity category
	Solid or hematological malignancy
CAR structure	CAR target
	CAR generation
	scFv source
	Transmembrane domain
	Costimulatory domain
	Signaling domain
Additional therapy	Preconditioning therapy
Response	Progressive disease
	Stable disease
	Partial response
	Complete response
Adverse events	GvHD (Grade 1–5)
	CRS (Grade 1–5)
	ICANS (Grade 1–5)
	Treatment-related deaths
	Anemia (Grade 1–5)
	Thrombocytopenia (Grade 1–5)
	Lymphopenia (Grade 1–5)
	Leukopenia (Grade 1–5)
	Neutropenia (Grade 1–5)

CAR, chimeric antigen receptor; CRS, cytokine release syndrome; GvHD, graft-versus-host disease; ICANS, immune effector cell-associated neurotoxicity syndrome; scFv, single-chain variable fragment.

Analysis of clinical trials
For included clinical studies, pre-established variables were collected (table 2): the nature of the tumor entity and target antigen; the full structure of the CAR: single-chain variable fragment (scFv), transmembrane (TM), intracellular domains; the administered chemotherapeutic lymphodepleting regimen; and the number of participants belonging to the efficacy, and the safety assessments of the CAR-T cell therapy of interest. Evaluating the efficacy consisted of dividing the patient population into four categories of overall response, namely (1) progressive disease, (2) stable disease, (3) partial response (PR), and (4) complete response. If not clearly stated, but discernible by the published data, the latest time point of disease assessment was considered. For the safety assessment of each CAR-T cell therapy, all serious and less severe AEs were reported. The severity of all AEs was classified into mild (Grade 1–2) or severe (Grade 3–4), as stated by the investigators. The list of AE included (1) cytokine release syndrome (CRS), (2) immune effector cell-associated neurotoxicity syndrome (ICANS), (3) GvHD, (4) hematotoxicity (distinguished

Table 3 Overview of all variables collected for all preclinical publications analyzed

Mouse model	Disease entity category
	Source of the cancer cell line
	Type of antigen
	Antigen overexpression
	Any further genetic modification
	Injection site of tumor cell line
	Mouse strain
	Mouse type
	Preconditioning therapy
	Rechallenge
CAR structure	CAR target
	CAR generation
	scFv source
	Transmembrane domain
	Costimulatory domain
	Signaling domain
Response	Tumor decrease
	Slower growth
	Tumor clearance
	No benefit
Adverse events	Weight decrease
	CRS
	GvHD
	Treatment-related deaths

CAR, chimeric antigen receptor; CRS, cytokine release syndrome; GvHD, graft-versus-host disease; scFv, single-chain variable fragment.

between anemia, leukopenia, lymphocytopenia, thrombocytopenia and neutropenia), and (5) death.

Analysis of preclinical publications

All approved preclinical studies underwent thorough review and data extraction, with pre-established variables collected. Multiple entries were created for different mouse models or CAR therapies within a single publication (table 3). Variables studied included: the nature of the tumor entity, the origin and potential overexpression of the target antigen; the mouse strain; the route of tumor cell injection; the full structure of the CAR; the use of a preconditioning regimen; and the number of all mice in the study treated with the CAR-T cell therapy of interest. The overall response of each treatment group was determined by the average of all reported responses within that group. The safety assessment encompassed all reported cases of AE, such as (1) relapse, (2) weight loss, (3) CRS, (4) ICANS, (5) GvHD, and (6) death of experimental animals due to therapy side effects. Mouse models were defined as either immunocompetent, immunocompromised or immunodeficient. By immunocompromised, the authors referred to mouse models with partial but significant loss of the host immune system (eg, Balb/c nude, NOG, Rag−/−, SCID), whereas those with a total loss of the host immune system were classified as immunodeficient (eg, NSG). When referring to any mouse model without a fully competent host immune system, the term "non-immunocompetent" was used.

Machine learning-guided analysis

Our analysis used both the preclinical (n=303 experimental entries) and clinical (n=105 clinical trials) CAR-T cell datasets. The response variables were binarized for logistic regression (figure 4): in preclinical data, individual mice were categorized into "No response" (referring to no benefit) versus "Response" (referring to tumor clearance, tumor decrease and slower growth accumulated), while in clinical data, an overall response rate (ORR) cut-off of 0.25 was used (trials with an ORR<0.25: "No response"; those with ORR≥0.25: "Response"). Three logistic regression models were developed: Model A was trained and validated on preclinical data with 14 features; hyperparameters were optimized using grid search with fivefold cross-validation (online supplemental methods). Model B was trained on preclinical data and validated on clinical data, using shared seven features ("Solid or Hematologic tumors", "scFv source", "Target", "CAR generation", "TM domain", "Preconditioning", "Costimulatory domain"). Model C was trained and validated on clinical data with seven features. Performance metrics included area under the curve (AUC) and macro F1 score. Feature importance was assessed through logistic regression coefficients. Datasets were further divided into solid and hematological tumors, with separate models trained and evaluated for each subset. A Lasso linear regression Model D was applied to predict the continuous ORR in clinical data, with feature importance determined by nonzero coefficients (online supplemental methods). Analysis was performed in a conda environment using Python V.3.10.13, scikit-learn V.1.3.2, and seaborn V.0.13.0.

Role of funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

RESULTS

Literature review and data collection

This study sought to bridge the gap between preclinical and clinical data in CAR-T cell therapy, not only by evaluating the clinical trial landscape and its respective preclinical publications, but also by using a logistic regression model to test whether preclinical studies can predict clinical treatment outcome. A systematic review of clinical and preclinical CAR-T cell studies was conducted following Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines[32]
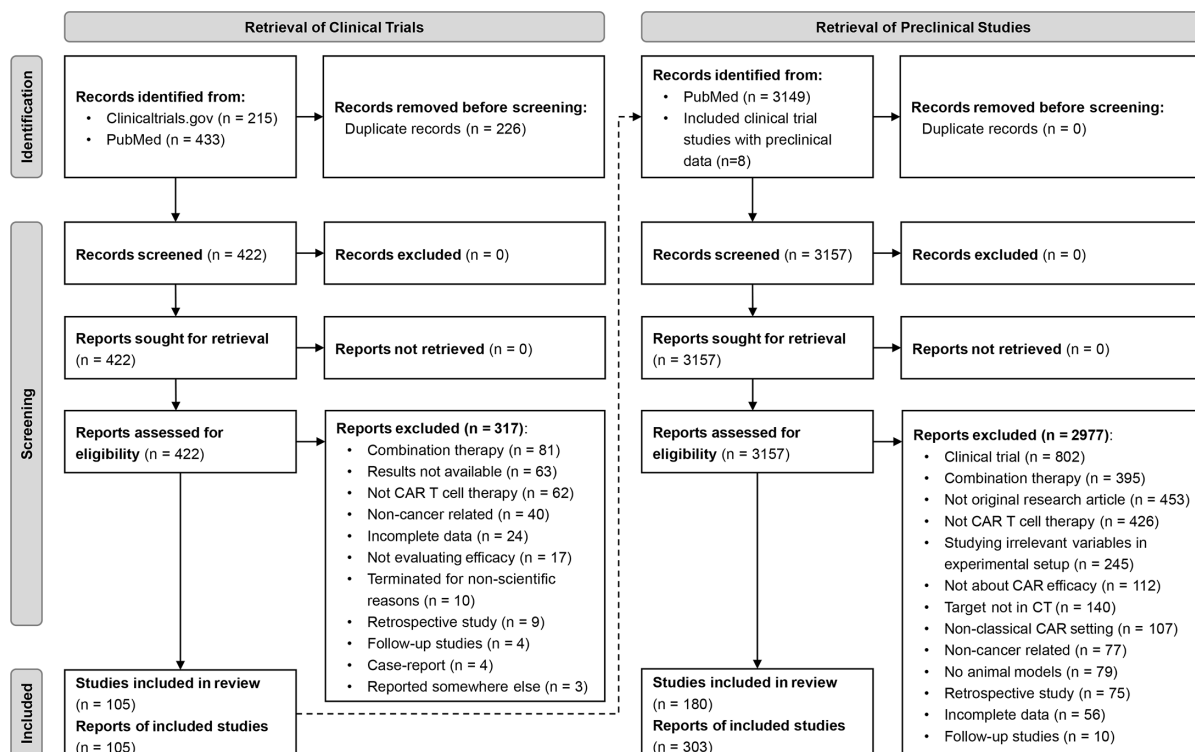
Figure 1 Overview of literature review and data collection process. Literature search performed until December 1, 2023, on Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 led to 105 clinical trials included in our study. Using the target antigens from the included clinical trials (dashed arrow), 303 relevant preclinical studies employing the same target antigen were identified and included in our analysis. CAR, chimeric antigen receptor.

(figure 1). At all times, two researchers independently screened the results, identifying 422 clinical and 3,157 preclinical articles and excluding articles due to combinatorial therapies, mislabeling or lacking data (online supplemental file 1, online supplemental file 3). 105 clinical and 180 preclinical studies met inclusion criteria, with no overlap of patients or animals occurring between studies (online supplemental file 2, online supplemental file 4).

Increased therapeutic efficacy and higher rates of side effects in clinical trials of hematological tumors

Searching PubMed, all clinical trials involving CAR-T cell therapy were retrieved until December 1, 2023, of which 105 clinical trials met inclusion criteria. 86 trials focused on hematological cancers, primarily employing anti-CD19 (n=53) and anti-B-cell maturation antigen (BCMA) (n=16) CAR-T cells (figure 2A). The meta-analysis included 3,312 patients with hematological cancers and 184 with solid cancers (figure 2B, online supplemental figure S1B), with most T-cell products using mouse-derived scFvs, second-generation CAR structures with 4-1BB or CD28 costimulatory domains (91%) (figure 2C). As expected, hematological cancer trials showed higher ORR than those for solid tumors, with only one trial involving anti-GD2 or anti-MSLN CAR-T cells showing notable efficacy, respectively (figure 2D,E). Interestingly, preconditioning therapy, such as chemotherapy

or irradiation, was used in 90% of hematological cancer trials and 37% of all solid cancer trials (online supplemental figure S1A). Expectedly, higher clinical responses in hematological cancers were strongly correlated with an increased rate of side effects, including ICANS and CRS (figure 2F–H). GvHD was rare, mainly occurring in patients treated with mouse scFv CARs (figure 2I). Hematotoxicity symptoms were less common in patients with solid cancer (figure 2J–N), partly explainable by the limited use of preconditioning therapy.

Preclinical mouse studies of CAR-T cell therapy in hematological and solid cancers

Preclinical publications on CAR-T cell therapy targeting antigens from prior clinical trials were retrieved from PubMed by December 1, 2023. Most studies focused on anti-CD19 (n=77), anti-MSLN (n=53), or anti-Her2 CAR-T cells (n=37) (figure 3A). Most B-cell acute lymphoblastic leukemia models (85%) were using anti-CD19 CAR-T cells, while CAR targets for solid tumors covered a comparably wider range of malignancies (figure 3B). The number of mice (hematological n=1,121, solid n=1,311) used for in vivo validation was similar for both tumor types (figure 3C). PRs, including tumor decrease and slower growth, were most commonly reported in non-immunocompetent mouse models, namely in 65.3% of solid tumors and 72.7% of hematological tumor
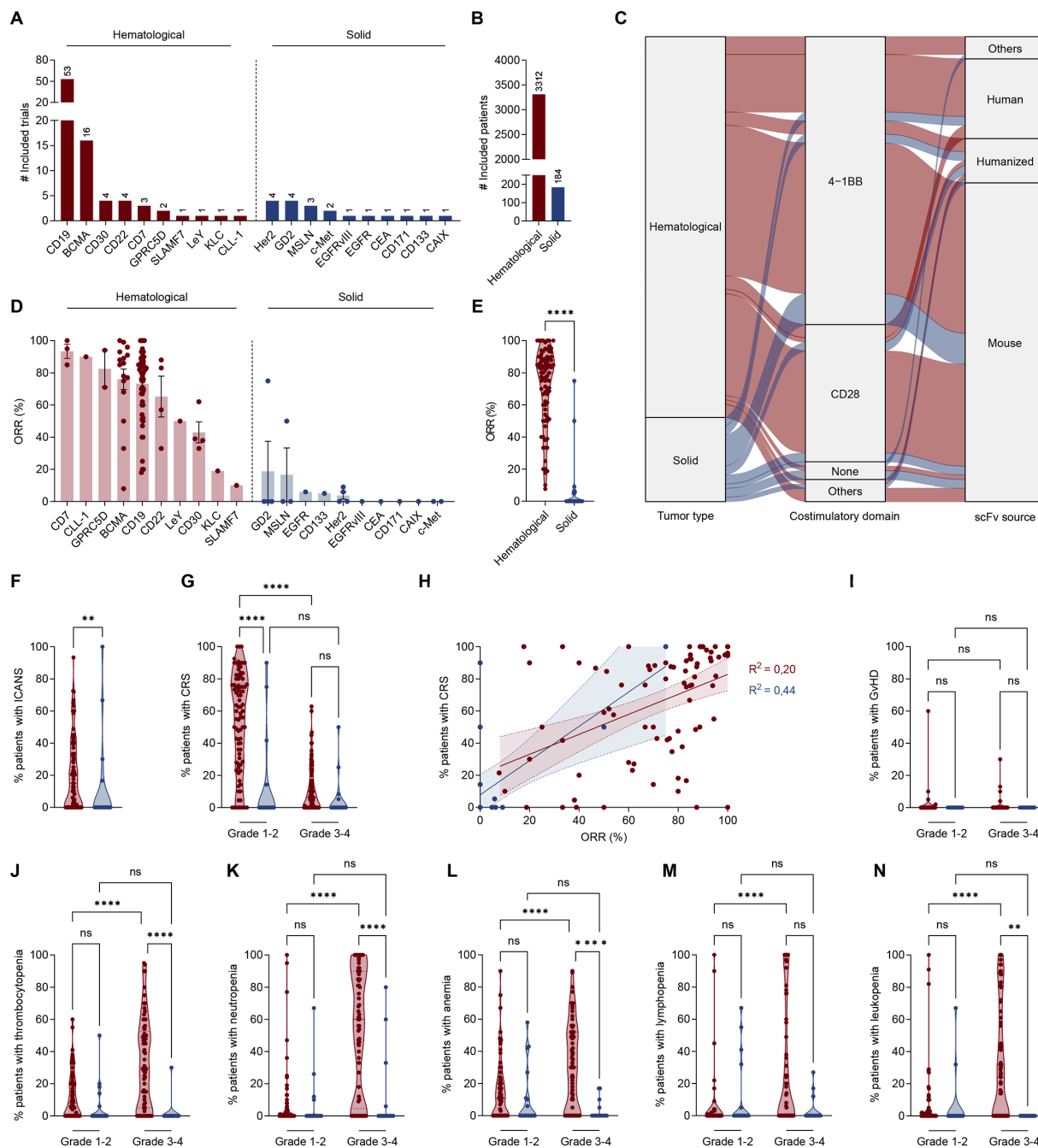
Figure 2 Hematological tumors are associated with higher response rates and toxicity than solid tumors in clinical CAR-T cell trials. (A) Number of clinical trials analyzed for each included target antigen. (B) Total number of included patients for hematological and solid cancer clinical trials. (C) Detailed information regarding costimulatory domain of CAR construct, as well as source of single-chain variable fragment. (D) ORR separated by target antigen. (E) ORR of all clinical trials for hematological and solid tumor entities. (F) Percentage of patients experiencing ICANS in hematological and solid tumor entities. (G) Percentage of patients experiencing CRS in hematological and solid tumor entities. (H) Correlation between ORR and occurrence of side effects in terms of CRS in patients. (I) Fraction of patients experiencing graft-versus-host disease. Distribution of patients experiencing (J) thrombocytopenia, (K) neutropenia, (L) anemia, (M) lymphopenia or (N) leukopenia. CAR, chimeric antigen receptor; CRS, cytokine release syndrome; ICANS, immune effector cell-associated neurotoxicity syndrome; ORR, overall response rate.

entities (figure 3D,E). Tumor clearances were more frequent in the small number of reported immunocompetent models, namely 47.6% for solid tumors and 45.5% for hematological tumors (figure 3F). Expectedly, most CAR-T cells tested (69%) used mouse scFvs with 4-1BB or CD28 as costimulatory domains. CAR-T

cells targeting hematological tumors favored 4-1BB costimulation, while solid tumor CARs leaned toward CD28 (figure 3G). In general, preclinical toxicity was reported in only 4% of the publications studied, with few studies noting issues like weight loss, CRS, ICANS, GvHD, or lethal AE (online supplemental figure
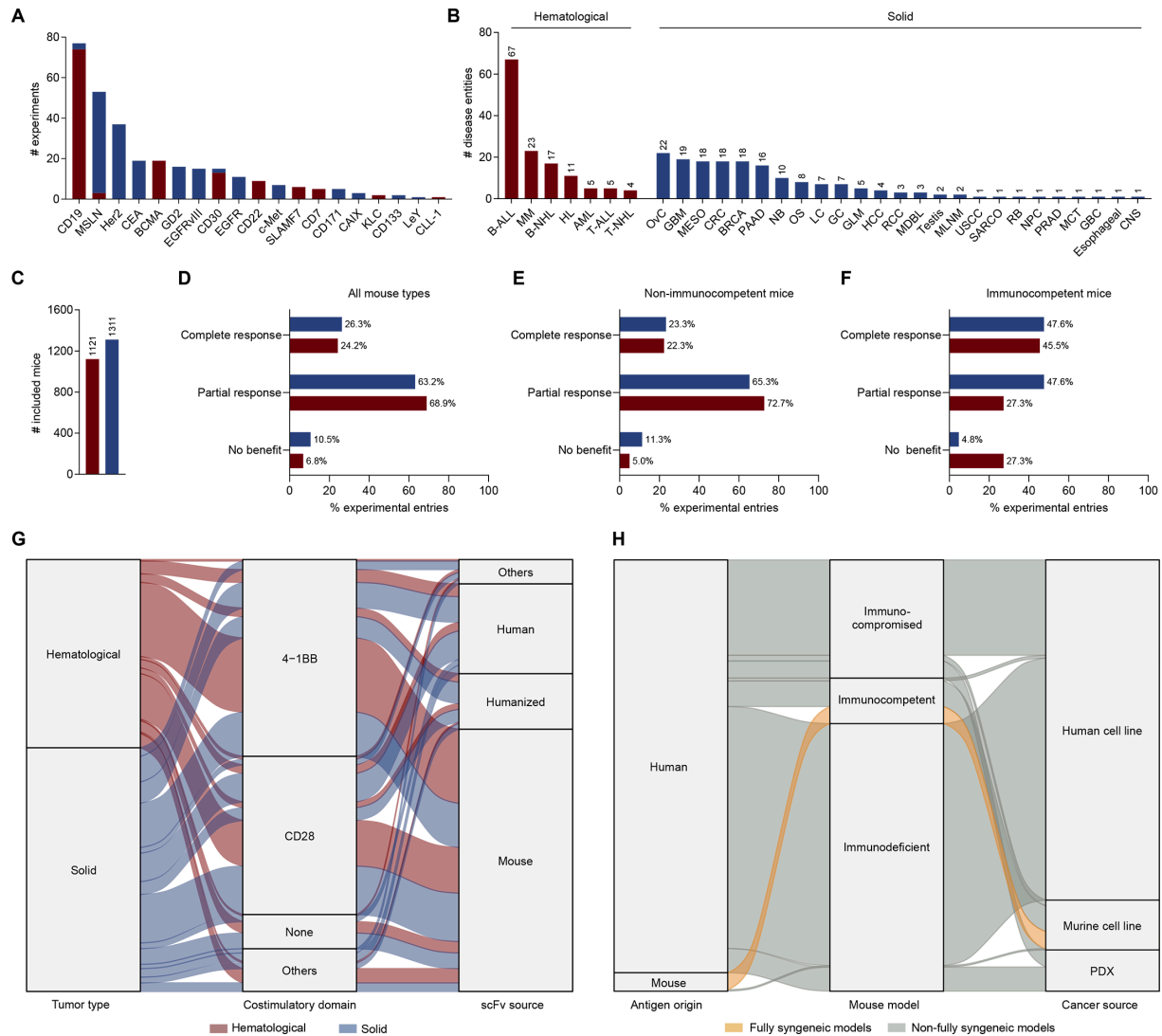
Figure 3 Preclinical mouse studies of CAR-T cell therapy in hematological and solid cancers are primarily of immunodeficient nature. (A) Overall number of preclinical entries per CAR target for either hematological or solid tumors. (B) Overall number of preclinical entries per disease entity for either hematological or solid tumors. (C) Sum of all mice belonging to CAR treatment groups for either category of tumors. (D) Overall responses as for tumor clearance, partial response (tumor decrease and slower growth) or no benefit for all mice, non-immunocompetent mice (E) and immunocompetent mice (F) for either type of tumor. (G) Alluvial plot displaying the different proportions of preclinical studies according to the single-chain variable fragment and costimulatory domain of their CAR molecule for either type of tumor. (F) Alluvial plot highlighting, in yellow, the proportion of preclinical studies employing fully syngeneic mouse models for in vivo CAR-T cell investigation. AML, acute myeloid leukemia; B-ALL, B-cell acute lymphoblastic leukemia; B-NHL, B-cell non-Hodgkin lymphoma; BRCA, breast adenocarcinoma; CAR, chimeric antigen receptor; CNS, central nervous system tumors; CRC, colorectal carcinoma; ESO, esophageal cancer; GBC, gallbladder cancer; GBM, glioblastoma; GC, gastric cancer; GLM, glioma; HCC, hepatocellular carcinoma; HL, Hodgkin's lymphoma; LC, lung cancer; MCT, mast cell tumor; MDBL, medulloblastoma; MESO, mesothelioma; MLNM, melanoma; MM, multiple myeloma; NB, neuroblastoma; NPC, nasopharyngeal cancer; OS, osteosarcoma; OVC, ovarian cancer; PAAD, pancreatic adenocarcinoma; PDX, patient-derived xenografts; PRAD, prostate adenocarcinoma; RB, retinoblastoma; RCC, renal cell carcinoma; SARCO, sarcoma; T-ALL, T-cell acute lymphoblastic leukemia; TC, testicular cancer; T-NHL, T-cell non-Hodgkin's lymphoma; USCC, serous carcinoma of the uterine cervix.

S2A–F). Only a small fraction (n=8) was fully syngeneic, using mouse antigens and tumor cells in immunocompetent mice (figure 3H, online supplemental figure S2G,H). Only 3% of all reported solid tumor experiments and 6.6% of all hematological tumors used PDX models, which predominantly demonstrated PRs (75% for solid tumors and 77% for hematological tumors) (online supplemental figure S2I,J).

This highlights a bias toward an increased use of non-immunocompetent models in preclinical CAR-T cell testing, with easier-to-achieve responses and minimal safety evaluation.

Preclinical and clinical data can predict clinical treatment outcome using a logistic regression model

To evaluate the predictive value of preclinical models in CAR-T cell trials, a comparative machine learning analysis was conducted. Three logistic regression models (A, B, and C) and a linear regression model (Model D) were deployed using different sets of training and testing data (figure 4A): Models A to C were trained on subsets, that is, (1) "all tumors", (2) "hematologic tumors" only, and (3) "solid tumors" only. Due to the small sample size and label imbalance in the clinical solid tumor subset (17 non-responders vs 2 responders), only Model A was trained and tested on the preclinical solid tumor data. When both hematological and solid tumor types ("all tumors") were included, Models B and C achieved higher predictive power than Model A (figure 4B,C), which showed no performance beyond random guessing (Macro F1=0.44±0.05, 95% CI: (0.36 to 0.56)) and AUC=0.52±0.09, 95% CI: (0.34 to 0.66), respectively). To exploit the continuous nature of the ORR in clinical studies, we also trained a regularized linear regression Model D to assess the predictive value of clinical features. Although this model demonstrated poor overall performance (mean R^2=0.51), it correctly predicted a lower ORR for solid tumors (figure 4E), with the largest negative weight assigned to the "Solid" tumor type (figure 4F). "Tumor type" ("Solid" or "Hematological") emerged as the most relevant discriminator in both classification Models B and C (figure 4D) and the regression Model D (figure 4F), indicating that preclinical and clinical data can predict clinical outcome when tumor type information is included. When Model C is trained and tested on the "hematologic tumor" subset, its performance is still beyond random guessing (mean F1=0.51±0.09, 95% CI: (0.36 to 0.67), figure 4B). "TM Domain" is identified as the most predictive feature, followed by "Preconditioning" and the "Costimulatory domain" (figure 4D).

DISCUSSION

The translational value of animal models remains a long-standing concern, encompassing more areas than just the CAR-T cell field.[33 34] Preclinical models often fail to accurately replicate human malignancy and the complexity of the immune system, yielding artificial and unreliable results when translated into the clinic.[35] The question of how accurately animal models can reflect and predict clinical CAR-T cell outcomes remains highly relevant.

To address this question, a comprehensive database of publicly available CAR-T cell data was compiled, including 105 clinical trials with 3,496 patients. As expected, hematological tumors showed the highest clinical responses, as well as higher rates of CAR-associated AE and hematotoxicity compared with solid malignancies, although severe AE have been observed in CAR-treated patients for both tumor types.[36] The majority of the analyzed clinical trials employed anti-CD19 and anti-BCMA CAR-T cells. Besides CD19, BCMA is the only other CAR-T cell target

specifically approved by the Food and Drug Administration (FDA) for the treatment of multiple myeloma.[37 38] Solid tumors accounted for a broader range of targets, accounting for the anticipated heterogeneity of their surface antigen.[39 40] Her2, GD2 and MSLN appeared as the most frequently investigated solid clinical targets. The prevalence of CD19 and BCMA, coupled with the significantly lower numbers of studies for solid tumors, inevitably skewed the overall trend for efficacy and safety. As more trials with novel solid tumor targets are published, a clearer efficacy assessment across different antigens will emerge.

Our preclinical analysis included 303 experiments, with a total of 2,432 mice. CD19, MSLN, and Her2 were the predominant CAR antigens, but CD19 targeting in a B-cell acute lymphoblastic leukemia tumor model represented again a strong bias in the whole analysis. Similarly to the clinical setting, preclinical studies accounted for a small number of solid tumor models with a broad range of different targets. Regardless of the nature of the tumor investigated, the majority of CAR-T cell therapies for blood-borne and solid malignancies were composed of mouse scFv domains and either 4-1BB or CD28 costimulatory domains, in line with the first FDA-approved CAR therapies.[41] These figures are not truly representative of the current pipeline for CAR-T cell therapy development, which reports a growing body of trials (27%) focusing on solid tumors of different origins.[42] Irrespective of the CAR molecule employed, both tumor types reported PRs in mice, with no advantage for blood-borne malignancies.

Three logistic regression models were trained to predict clinical response from preclinical (B) or clinical (C) data, and preclinical response from preclinical data (A). Given the inherent differences between the efficacy readout of clinical and preclinical studies, we binarized the outcomes into responders and non-responders. This underscored a critical distinction: while the reporting of clinical results is highly standardized, preclinical models rely on surrogate endpoints, which may have diverse biological implications. Although necessary for our model, this approach could reduce the granularity of our dataset and overlook subtle patterns in the responses. When predicting clinical response using both clinical and preclinical data, our logistic regression models exhibited superior performance compared with a random classifier. Notably, the predictive strength of these models is predominantly derived from tumor type information. Despite this, the results indicate a degree of concordance between preclinical and clinical studies, even in light of the well-established limitations of animal models in accurately forecasting clinical outcomes in oncology.[43 44] The model predicting preclinical outcomes from preclinical data including solid and hematological tumors (Model A - "All tumors") showed an overall poor performance (Macro F1=0.45±0.05, 95% CI: (0.36 to 0.56), AUC=0.52±0.09, 95% CI: (0.41 to 0.71)), indicating that the available data does not capture the complexity of treatment effects in preclinical experimentation. The most discriminating
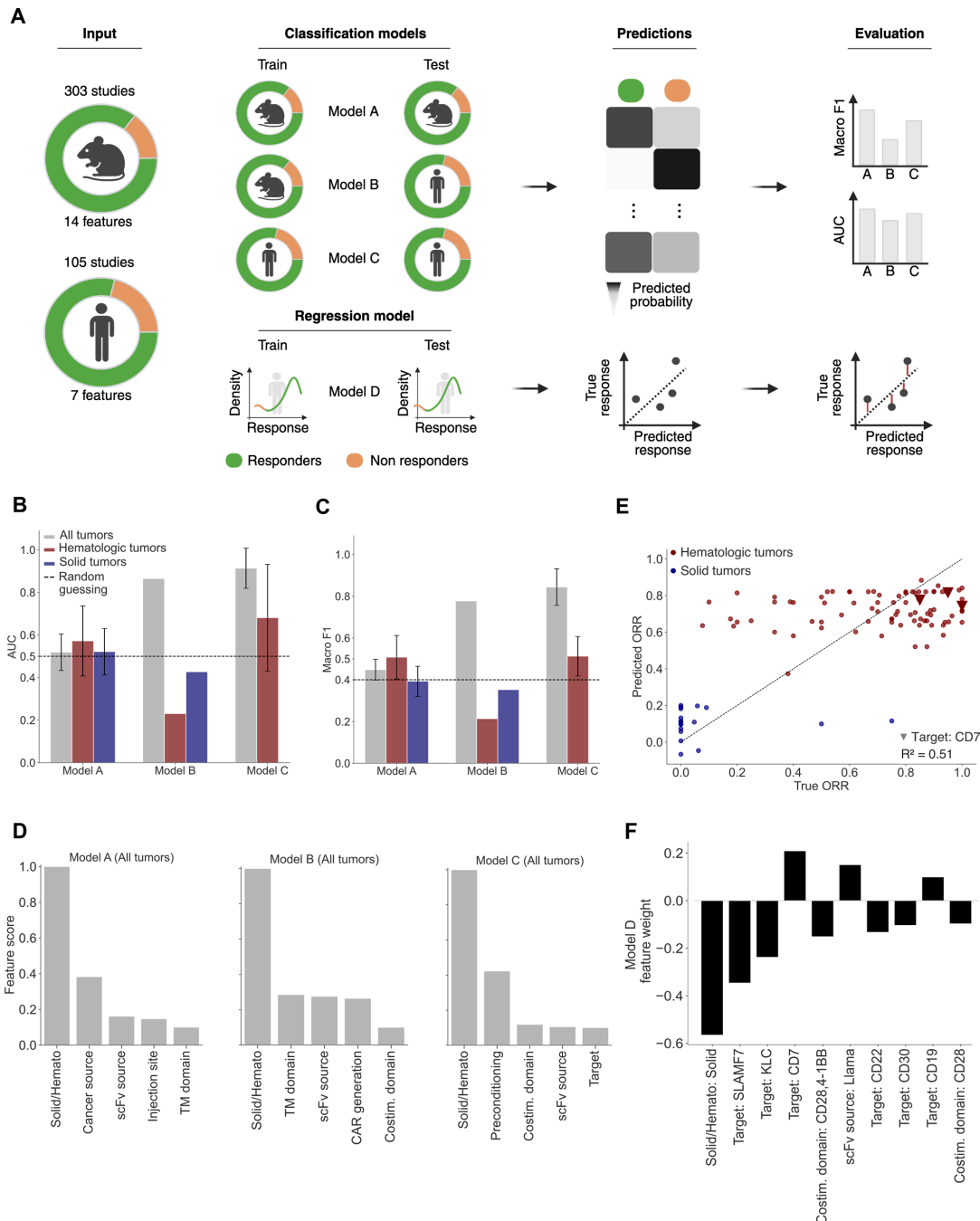
Figure 4 Machine learning analysis of preclinical and clinical datasets identifies tumor type as the most predictive feature across both classification and regression tasks. (A) Schematic outline of the model training and testing strategies: Models A, B, and C are classification models, whereas Model D is a regression model. Models A, C, and D were trained and validated using fivefold cross-validation; specifically, Model A used preclinical data, while Models C and D used clinical data. Model B was trained and validated on preclinical data and subsequently tested on clinical data. (B–C) Performance metrics, including macro F1 score (B) and AUC (C), are reported for Models A, B, and C across the entire dataset ("All tumors"), hematological and solid tumor subsets. Results for the solid tumor subset are only presented for Model A due to the limited size and label imbalance of the clinical solid tumor subset (online supplemental methods). Horizontal, dashed lines indicate the performance of a model using random guessing as a baseline (online supplemental methods). Error bars represent the SD of the scores across the 5*10 validation folds (online supplemental methods) for Models A and C. (D) Feature importance scores from Models A, B and C considering hematologic and solid tumors ("All tumors"). Tumor type (solid vs hematologic) consistently emerges as the most predictive feature for both Models B and C. In Model B, this is followed by "TM domain", whereas in Model C, "Preconditioning" ranks second. All feature importance scores are normalized and expressed on a relative scale. (E) Scatter plots show true versus predicted overall response rates (ORR) across all cross-validation folds for Model D, with a moderate correlation (mean R²=0.51). (F) Feature weights indicate that lower predicted ORRs are mainly linked to solid tumors, which have the most negative weight. AUC, area under the curve; CAR, chimeric antigen receptor; scFv, single-chain variable fragment; TM, transmembrane.

factor for both models predicting clinical responses (B and C) was the cancer's solid or hematological origin, in line with the reported greater efficacy of CAR-T cells in hematological malignancies compared with solid tumors.[45 46] After subsetting the datasets into solid and hematological tumors, the performance of Models B and C dropped considerably, from 0.78 to 0.21 and from 0.84 to 0.51 of Macro F1, respectively. This reduction in performance may be due to the absence of the main discriminating feature, the smaller dataset size, and the increased severity of label imbalance. Consistent with the feature rankings from the classification model, the regression Model D assigned the highest negative weight to the "Solid" tumor type. Given a categorical Lasso regression, where a negative weight indicates a decrease in the target variable and its magnitude reflects the strength of such decrease, our findings confirm that solid tumors are linked to overall poorer response rates.

A number of confounding factors should be taken into consideration for this study. First, preclinical research is biased towards positive data, as negative results are rarely published in prestigious journals, unlike in clinical research.[47] Researchers may also employ mouse models to reinforce their hypotheses by employing control groups in a way that would negatively bias the experimental set-up.[48] It was therefore highly challenging to obtain an accurate and comprehensive understanding of preclinical testing from the currently available data.

Second, preclinical evaluation of CAR-T cells relies predominantly on immunodeficient xenograft models.[20 28] However, tumor xenografts fail to investigate the impact of endogenous immunity on tumor control, reduce tumor heterogeneity, and belittle toxicity against antigen-expressing healthy tissues.[25 30] As a result, the overall rate of toxicity reporting was little to none for both hematological (6.8%) and solid tumors (8.2%), suggesting it to be under-reported or not conducted at all. The higher efficacy and lower safety concerns of immunodeficient mouse models make them the most commonly used in vivo models in the field, despite their high maintenance costs.[25 35] This is clearly reflected in our data, whereby immunodeficient and immunocompromised models displayed PRs on average, in stark contrast to immunocompetent models. The limited information on toxicities prevented any definitive conclusion regarding toxicity profiles in hematological versus solid tumors, highlighting the limitations of preclinical mouse models in accurately predicting clinical toxicity. Researchers possess a number of potentially more suitable animal models for in vivo preclinical evaluation of immunotherapies, namely syngeneic and humanized models.[49] Unfortunately, fully syngeneic mouse models constituted a very small proportion (4%) of overall preclinical records. Such scarcity, combined with insufficient toxicity reporting in preclinical publications, highlighted the strong bias toward immunodeficient models.[28] Considering these low numbers and the phenomenon of positive publication bias, it is challenging to assess whether immunocompetent mice would

be better predictors of clinical response. Bearing in mind the key importance of immunity in immunotherapies, it is tempting to speculate that fully immunocompetent models may better discriminate towards clinical outcome. This hypothesis will require adequate investigation and demonstration.

Third, this work did not account for a number of factors like the CAR transduction vector, the in vitro handling and expansion of cancer and CAR-T cells prior to animal treatment, the differences in dosing regimens, and the timing of in vivo treatment. Furthermore, the strict exclusion criteria here applied led to the removal of CAR therapies involving immune cells other than T lymphocytes and the combination of CAR-T cells with other strategies like antibodies, cytokines, or costimulatory receptors. The use of combination therapies involving CAR-T cells is an ever-growing field of research.[50] In this analysis, we observed a strong interest in exploiting the heterogeneous range of identified solid tumor antigens and can thus anticipate a broader exploration of more sophisticated CAR therapy designs, multiantigen targeting strategies and combinatorial approaches to overcome the challenges posed by solid malignancies.[1 16]

In conclusion, while our machine learning models indicated weak predictive power of clinical responses from preclinical data, the findings underscored the need for more diverse and comprehensive preclinical studies. This, however, should not be seen as a systematic challenge to animal experimentation in this realm, we merely concede that current models have limitations to be considered. Animal models are certainly helpful in attempts to establish if a particular cellular treatment could in principle be useful therapeutically. However, we hereby stress the need to complement and refine current practice to enhance clinical predictive value. In our view, this should include both refined preclinical mouse models and patient-near in vitro models. The integration of more immunocompetent and humanized mouse models, with more standardized testing and reporting guidelines, will allow a more robust analysis of the weight and predictive power of each feature. Such an approach will hopefully lead to more effective and tailored CAR-T cell therapies for both hematological and solid tumors.

Author affiliations
[1]Division of Clinical Pharmacology, LMU Klinikum, Member of the German Center for Lung Research (DZL), Munich, Germany
[2]Institute of AI for Health, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
[3]German Cancer Consortium (DKTK), partner site Munich, a partnership between the DKFZ Heidelberg and the University Hospital of the LMU, Heidelberg, Germany
[4]Einheit für Klinische Pharmakologie (EKLiP), Helmholtz Zentrum München - German Research Center for Environmental Health Neuherberg, Munich, Germany

ORCID iDs
Lisa Gregor http://orcid.org/0000-0002-3300-7167
Sebastian Kobold http://orcid.org/0000-0002-5612-4673

REFERENCES

1. Stoiber S, Cadilha BL, Benmebarek M-R, et al. Limitations in the Design of Chimeric Antigen Receptors for Cancer Therapy. *Cells* 2019;8:472.
2. Amini L, Silbert SK, Maude SL, et al. Preparing for CAR T cell therapy: patient selection, bridging therapies and lymphodepletion. *Nat Rev Clin Oncol* 2022;19:342–55.
3. Lesch S, Benmebarek M-R, Cadilha BL, et al. Determinants of response and resistance to CAR T cell therapy. *Semin Cancer Biol* 2020;65:80–90.
4. Anagnostou T, Riaz IB, Hashmi SK, et al. Anti-CD19 chimeric antigen receptor T-cell therapy in acute lymphocytic leukaemia: a systematic review and meta-analysis. *Lancet Haematol* 2020;7:e816–26.
5. Xiang X, He Q, Ou Y, et al. Efficacy and Safety of CAR-Modified T Cell Therapy in Patients with Relapsed or Refractory Multiple Myeloma: A Meta-Analysis of Prospective Clinical Trials. *Front Pharmacol* 2020;11:544754.
6. Cordas Dos Santos DM, Tix T, Shouval R, et al. A systematic review and meta-analysis of nonrelapse mortality after CAR T cell therapy. *Nat Med* 2024;30:2667–78.
7. Lei W, Xie M, Jiang Q, et al. Treatment-Related Adverse Events of Chimeric Antigen Receptor T-Cell (CAR T) in Clinical Trials: A Systematic Review and Meta-Analysis. *Cancers (Basel)* 2021;13:3912.
8. Umut Ö, Gottschlich A, Endres S, et al. CAR T cell therapy in solid tumors: a short review. *Memo* 2021;14:143–9.
9. Vitanza NA, Johnson AJ, Wilson AL, et al. Locoregional infusion of HER2-specific CAR T cells in children and young adults with recurrent or refractory CNS tumors: an interim analysis. *Nat Med* 2021;27:1544–52.
10. Feng K, Liu Y, Guo Y, et al. Phase I study of chimeric antigen receptor modified T cells in treating HER2-positive advanced biliary tract cancers and pancreatic cancers. *Protein Cell* 2018;9:838–47.
11. Ahmed N, Brawley V, Hegde M, et al. HER2-Specific Chimeric Antigen Receptor-Modified Virus-Specific T Cells for Progressive Glioblastoma: A Phase 1 Dose-Escalation Trial. *JAMA Oncol* 2017;3:1094–101.
12. Haas AR, Tanyi JL, O'Hara MH, et al. Phase I Study of Lentiviral-Transduced Chimeric Antigen Receptor-Modified T Cells Recognizing Mesothelin in Advanced Solid Cancers. *Mol Ther* 2019;27:1919–29.
13. Beatty GL, Haas AR, Maus MV, et al. Mesothelin-specific chimeric antigen receptor mRNA-engineered T cells induce anti-tumor activity in solid malignancies. *Cancer Immunol Res* 2014;2:112–20.
14. Yu L, Huang L, Lin D, et al. GD2-specific chimeric antigen receptor-modified T cells for the treatment of refractory and/or recurrent neuroblastoma in pediatric patients. *J Cancer Res Clin Oncol* 2022;148:2643–52.
15. Majzner RG, Ramakrishna S, Yeom KW, et al. GD2-CAR T cell therapy for H3K27M-mutated diffuse midline gliomas. *Nature New Biol* 2022;603:934–41.
16. Martinez M, Moon EK. CAR T Cells for Solid Tumors: New Strategies for Finding, Infiltrating, and Surviving in the Tumor Microenvironment. *Front Immunol* 2019;10:128.
17. Schmidts A, Maus MV. Making CAR T Cells a Solid Option for Solid Tumors. *Front Immunol* 2018;9:2593.
18. Moreno L, Pearson ADJ. How can attrition rates be reduced in cancer drug discovery? *Expert Opin Drug Discov* 2013;8:363–8.
19. Brown LV, Gaffney EA, Ager A, et al. Quantifying the limits of CAR T-cell delivery in mice and men. *J R Soc Interface* 2021;18:20201013.
20. Duncan BB, Dunbar CE, Ishii K. Applying a clinical lens to animal models of CAR-T cell therapies. *Mol Ther Methods Clin Dev* 2022;27:17–31.
21. Zhong W, Myers JS, Wang F, et al. Comparison of the molecular and cellular phenotypes of common mouse syngeneic models with human tumors. *BMC Genomics* 2020;21:2.
22. Olson B, Li Y, Lin Y, et al. Mouse Models for Cancer Immunotherapy Research. *Cancer Discov* 2018;8:1358–65.
23. Devaud C, Westwood JA, John LB, et al. Tissues in different anatomical sites can sculpt and vary the tumor microenvironment to affect responses to therapy. *Mol Ther* 2014;22:18–27.
24. Bibby MC. Orthotopic models of cancer for preclinical drug evaluation: advantages and disadvantages. *Eur J Cancer* 2004;40:852–7.
25. Chulpanova DS, Kitaeva KV, Rutland CS, et al. Mouse Tumor Models for Advanced Cancer Immunotherapy. *Int J Mol Sci* 2020;21:4118.
26. Jung J. Human tumor xenograft models for preclinical assessment of anticancer drug development. *Toxicol Res* 2014;30:1–5.
27. De La Rochere P, Guil-Luna S, Decaudin D, et al. Humanized Mice for the Study of Immuno-Oncology. *Trends Immunol* 2018;39:748–63.
28. Talmadge JE, Singh RK, Fidler IJ, et al. Murine models to evaluate novel and conventional therapeutic strategies for cancer. *Am J Pathol* 2007;170:793–804.
29. Lai Y, Wei X, Lin S, et al. Current status and perspectives of patient-derived xenograft models in cancer research. *J Hematol Oncol* 2017;10:106.
30. Abdolahi S, Ghazvinian Z, Muhammadnejad S, et al. Patient-derived xenograft (PDX) models, applications and challenges in cancer research. *J Transl Med* 2022;20:206.
31. Kalaitsidou M, Kueberuwa G, Schütt A, et al. CAR T-cell therapy: toxicity and the relevance of preclinical models. *Immunotherapy (Los Angel)* 2015;7:487–97.
32. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev* 2021;10:89.
33. Mukherjee P, Roy S, Ghosh D, et al. Role of animal models in biomedical research: a review. *Lab Anim Res* 2022;38:18.

34 Van Norman GA. Limitations of Animal Studies for Predicting Toxicity in Clinical Trials: Part 2: Potential Alternatives to the Use of Animals in Preclinical Trials. *JACC Basic Transl Sci* 2020;5:387–97.

35 Gould SE, Junttila MR, de Sauvage FJ. Translational value of mouse models in oncology drug development. *Nat Med* 2015;21:431–9.

36 Maus MV, Haas AR, Beatty GL, *et al*. T cells expressing chimeric antigen receptors can cause anaphylaxis in humans. *Cancer Immunol Res* 2013;1:26–31.

37 Raje N, Berdeja J, Lin Y, *et al*. Anti-BCMA CAR T-Cell Therapy bb2121 in Relapsed or Refractory Multiple Myeloma. *N Engl J Med* 2019;380:1726–37.

38 Madduri D, Berdeja JG, Usmani SZ, *et al*. CARTITUDE-1: Phase 1b/2 Study of Ciltacabtagene Autoleucel, a B-Cell Maturation Antigen-Directed Chimeric Antigen Receptor T Cell Therapy, in Relapsed/Refractory Multiple Myeloma. *Blood* 2020;136:22–5.

39 Liu L, Qu Y, Cheng L, *et al*. Engineering chimeric antigen receptor T cells for solid tumour therapy. *Clin Transl Med* 2022;12:e1141.

40 Wagner J, Wickman E, DeRenzo C, *et al*. CAR T Cell Therapy for Solid Tumors: Bright Future or Dark Reality? *Mol Ther* 2020;28:2320–39.

41 Cappell KM, Kochenderfer JN. A comparison of chimeric antigen receptors containing CD28 versus 4-1BB costimulatory domains. *Nat Rev Clin Oncol* 2021;18:715–27.

42 Michaelides S, Obeck H, Kechur D, *et al*. Migratory Engineering of T Cells for Cancer Therapy. *Vaccines (Basel)* 2022;10:1845.

43 Ruggeri BA, Camp F, Miknyoczki S. Animal models of disease: preclinical animal models of cancer and their applications and utility in drug discovery. *Biochem Pharmacol* 2014;87:150–61.

44 de Jong M, Maina T. Of mice and humans: are they the same?--Implications in cancer translational research. *J Nucl Med* 2010;51:501–4.

45 Metzinger MN, Verghese C, Hamouda DM, *et al*. Chimeric Antigen Receptor T-Cell Therapy: Reach to Solid Tumor Experience. *Oncology (Williston Park, NY)* 2019;97:59–74.

46 Bagley SJ, O'Rourke DM. Clinical investigation of CAR T cells for solid tumors: Lessons learned and future directions. *Pharmacol Ther* 2020;205:107419.

47 Fanelli D. Negative results are disappearing from most disciplines and countries. *Scientometrics* 2012;90:891–904.

48 Krebs CE, Herrmann K. Confronting the bias towards animal experimentation (animal methods bias). *Front Drug Discov* 2024;4.

49 Chuprin J, Buettner H, Seedhom MO, *et al*. Humanized mouse models for immuno-oncology research. *Nat Rev Clin Oncol* 2023;20:192–206.

50 Al-Haideri M, Tondok SB, Safa SH, *et al*. CAR-T cell combination therapy: the next revolution in cancer treatment. *Cancer Cell Int* 2022;22:365.