ARTICLE

Check for updates

The target atlas for antibody-drug conjugates across solid cancers

Jiacheng Fang [ID][1,2], Lei Guo[1 ✉], Yanhao Zhang[3], Qing Guo[4], Ming Wang[5 ✉] and Xiaoxiao Wang[2 ✉]

Antibody-Drug Conjugates (ADCs) represent a rapidly advancing category of oncology therapeutics, spanning the targeted therapy for both hematologic malignancies and solid cancers. A crucial aspect of ADC research involves the identification of optimal surface antigens that can effectively differentiate target cells from most mammalian cell types. Herein, we have devised an algorithm and compiled an extensive dataset annotating cell membrane proteins. This dataset is derived from comprehensive transcriptomic, proteomic, and genomic data encompassing 19 types of solid cancer as well as normal tissues. The aim is to uncover potential therapeutic surface antigens for precise ADC targeting. The resulting target landscape comprises 165 combinations of targets and indications, along with 75 candidates of cell surface proteins. Notably, 35 of these candidates possess characteristics suitable for ADC targeting, and have not been previously reported in ADC research and development. Additionally, we have identified a total of 159 ADCs from a pool of 760 clinical trials. Of these, 72 ADCs are presently undergoing interventional evaluation for a variety of solid cancer types, targeting 36 unique antigens. We conducted an analysis of their expression in normal tissues using this comprehensive annotation dataset, revealing a diverse range of profiles for the current ADC targets. Moreover, we emphasize that the biological impacts of target antigens have the potential to enhance their clinical effectiveness. We propose a comprehensive assessment of the drugability of target antigens, considering multiple facets. This study represents a thorough exploration of pan-cancer ADC targets over the past two decades, underscoring the potential of a comprehensive solid cancer target atlas to broaden the scope of ADC therapies.

INTRODUCTION

Cancer ranks as the primary or secondary cause of premature death (occurring between ages 30–69) in 134 out of 183 countries. In an additional 45 countries, it holds the position of the third or fourth leading cause [1]. The section of Cancer Surveillance (CSU) of International Agency for Research on Cancer (IARC) projects that in 2020, there will be a total of 19.3 million new cancer cases and 10.0 million cancer-related deaths worldwide, encompassing both sexes and all age groups (https://gco.iarc.fr/today/home). Out of these cases and deaths, 18.0 million (93.37%) and 9.3 million (92.85%) respectively will be attributed to solid tumors. An extensive assessment of oncology reveals that a majority of cancer drugs authorized by the European Medicines Agency between 2009–13 were introduced to the market without achieving a significant enhancement in the overall quality or quantity of patients' lives [2]. This underscores the critical necessity for advancing the therapeutic effectiveness of drugs targeting solid tumors.

ADCs are specialized compounds composed of recombinant monoclonal antibodies (mAbs) chemically linked to potent cytotoxic agents, often referred to as payloads or warheads. They are engineered to utilize the highly specific targeting abilities of mAbs to deliver and release these potent chemotherapeutic agents to tumor cells, as well as the neighboring stromal cells and cancer stem cells [3, 4]. Currently, fifteen ADCs have received marketing approval, and there are 760 clinical studies registered in ClinicalTrials.gov (Supplementary Tables 1, 2). Of these, more than 400 trials are actively underway at various stages, with 227 trials focused on different types of solid tumors (Supplementary Table 2). The first ADC, gemtuzumab ozogamicin, gained regulatory approval from the US Food and Drug Administration (FDA) in 2000. However, in the subsequent ten years, no ADCs received global approval. In the past three years, from 2019 to 2022, a total of ten ADCs have been approved, with seven of them authorized for the treatment of solid cancers (Supplementary Table 1). Trastuzumab deruxtecan (DS-8201), a monotherapy targeting HER2, has demonstrated significant effectiveness in pretreated patients with HER2-positive metastatic breast cancer, achieving a confirmed overall response rate of 60.9% and a disease-control rate of 97.3% [5]. ADCs have emerged as a prominent area for drug innovation and development.

An optimal ADC should encompass three essential components [6]: 1) a mAb with high specificity for an antigen, whether homogeneous or heterogeneous, that is overexpressed in

[1]Interdisciplinary Institute of Medical Engineering, Fuzhou University, Fuzhou, Fujian 350108, China. [2]State Key Laboratory of Environmental and Biological Analysis, Department of Chemistry, Hong Kong Baptist University, RRS812 Kowloon Tong, Hong Kong SAR, China. [3]School of Ecology and Environment, Zhengzhou University, Zhengzhou, Henan 450001, China. [4]Department of Chemistry, Hong Kong Baptist University, RRS812 Kowloon Tong, Hong Kong SAR, China. [5]College of Food Science & Engineering, Northwest University, 229 Taibai North Road, Xi'an, Shaanxi 710069, China. ✉email: gankLei@fzu.edu.cn; wangming@nwu.edu.cn; wangxiaoxiao_523@163.com

tumors; 2) a linker that maintains stability in the bloodstream but can readily cleave at target sites; 3) a warhead with high sensitivity for specific indications. The antibody, linker, and payload can be fine-tuned using various approaches. For instance, site-specific conjugation technology can enhance the safety and efficacy of ADCs, with or without antibody engineering [7–11]. Off-target payload exposure can be minimized by bolstering the stability of conjugation and linker [12–14]. The efficiency of ADC uptake and processing can be augmented through the use of bispecific or biparatopic antibodies [15, 16]. Exploring payloads with novel anti-tumor mechanisms addresses the challenge of drug resistance posed by the warhead component [3, 17]. However, the intrinsic attributes of ADC target proteins, apart from these three, generally remain constant. Their gene copy number variation, endocytosis rate, trafficking route, absolute protein levels, and differential expression patterns cannot be optimized. Moreover, since most antigens are tumor-associated rather than tumor-specific, on-target off-tumor toxicity becomes inevitable. For example, BR96-doxorubicin treatment induced hemorrhagic gastritis in patients due to the unforeseen presence of Lewis(y) antigen on gastric mucosal cells [18]. Treatment with CD44v6-directed Bivatuzumab mertansine and BAY794620 targeting CA9 antigen resulted in fatal epidermal necrolysis and gastrointestinal toxicity, respectively, due to the expression of targets in corresponding tissues [17, 19]. Therefore, selecting a promising ADC target must meet numerous criteria to enable effective tumor eradication while preventing intolerable damage to normal tissues.

Crucially, an optimal ADC target antigen should exhibit exceptionally high density in target cells across a significant proportion of patients, regardless of the pathologic stage. It should also demonstrate limited expression in normal tissues, particularly in vital tissues, hematopoietic stem cells (HSCs), and multipotential progenitors (MPPs). Additionally, it is worth noting that elevated levels of a target antigen on various types of blood cells could influence drug exposure, potentially impeding the aggregation of ADCs at the target site. Once these criteria and considerations are met, each candidate should demonstrate a substantial ratio of differential expression in tumors compared to both adjacent normal tissue and other normal tissues. Beyond protein expression profiles, the confirmed biological functions of the target antigen in tumorigenesis are known to be additional factors that can significantly impact the efficacy of ADCs [20]. Thus, it is imperative to thoroughly assess the druggability of candidates based on comprehensive antigen annotation sources, as well as specialized analytical tools tailored to identify potential ADC target antigens.

The objective of this study is to establish a strategy for ADC target discovery by integrating data from transcriptome, proteome, and genome of both solid tumor and normal cell populations, and compiling a comprehensive dataset of cell surface antigen annotations to identify potential therapeutic targets. Concurrently, we also provide an overview of the ADC interventional studies documented in ClinicalTrials.gov, so as to furnish the academic and pharmaceutical sectors with an all-encompassing ADC target atlas spanning various types of solid cancers.

METHODS
Clinical trials search strategy
To identify the most comprehensive target atlas from clinical trials, ClinicalTrials.gov and PubMed were searched on 31 December 2022 with the following search terms 'antibody-drug conjugate', 'cancer', 'tumour', and 'oncology' in various combinations. Just interventional studies were included, and the early phase I trials were grouped with phase I trials. In addition, abstracts and posters from ASCO 2022, ECCO 2022 and ESMO 2022 congresses were included for ADC searching with the terms 'antibody-drug conjugate' or 'ADC'.

Data acquisition of the normal tissue transcriptome and proteome
Expression profiles for human tissue proteins based on IHC were retrieved from the Human Protein Atlas (HPA) under the entry of 'Normal tissue data' (https://www.proteinatlas.org/about/download, normal_tissue.tsv.zip). Proteomic sequencing data based on LC-MS/MS was retrieved from the Human Proteome Map (HPM) (http://www.humanproteomemap.org/, HPM_normal protein_level_expression_matrix_Kim_et_al_052914 - LC-MSMS.xlsx). Consensus transcript expression profiling integrated from the HPA, GTEx and FANTOM5 was retrieved from the HPA under the entry of 'RNA consensus tissue gene data' (https://www.proteinatlas.org/about/download, rna_tissue_consensus.tsv.zip).

Retrieval and compilation of protein subcellular location data
Subcellular localization of proteins was retrieved from the HPA under the entry of 'Subcellular location data' (https://www.proteinatlas.org/about/download, subcellular_location.tsv.zip), and the knowledge channel of COMPARTMENTS (https://compartments.jensenlab.org/Downloads, human_compartment_knowledge_full.tsv). The membrane protein annotation dataset (includes 6176 entries) was compiled by extracting all cell surface membrane protein information first through the R language (version 4.0.3), and then merging the above information.

Rearrangement and mapping of the human tissues
Since the human tissue nomenclature differs among source repositories, each data set was mapped to a set of consensus tissue labels. In cases when mapping multiple tissues in one repository to a single tissue label in another source, the maximum expression value was selected. For example, the caudate, cerebellum, cerebral cortex, choroid plexus, dorsal raphe, hippocampus, hypothalamus, pituitary gland, and substantia nigra were collapsed into a single tissue category, "brain". The cervix, uterine, endometrium, ovary, fallopian tube, vagina, epididymis, seminal vesicle, testis, and prostate were collapsed into internal genitalia. In addition, the adult adrenal, adult colon, adult esophagus, adult frontal cortex, adult gallbladder, adult heart, adult kidney, adult liver, adult lung, adult ovary, adult pancreas, adult prostate, adult rectum, adult retina, adult testis, and adult urinary bladder in the HPM were mapped to the adrenal gland, colon, esophagus, brain, gallbladder, heart muscle, kidney, liver, lung, internal genitalia, pancreas, internal genitalia, rectum, eye, internal genitalia, and urinary bladder, respectively. To maintain consistency, fetal tissues were also discarded, resulting in 32 unique tissue categories.

Normal tissue expression and binning
In the interest of facilitating target screening, the expression values were classified into five categories, including 'High', 'Medium', 'Low', 'Not Detected', and 'Not Available'. To accomplish the binning, we first perform $\log_{10}$ conversion on HPM dataset, and then temporarily correct it for the purpose of abundance distribution estimation. In order to best fit the normal curves to the observed distributions, we applied the Broyden-Fletcher-Goldfarb-Shanno algorithm [21], and subsequently obtained the peak maximum and standard deviation measure. Expression values in the range of one standard deviation above the peak maximum were recorded as 'Medium', and expression values above this threshold were recorded as 'High'. Similarly, expression values in the range of one standard deviation below the peak maximum were recorded as 'Low', and for those falling below one standard deviation were recorded as 'Not Detected'. Proteins without expression values were recorded to be of 'Not Available' abundance. The natural format of the expression profile of human tissue proteins obtained from IHC is the five aforementioned categories, so there is no adjustment. While for the RNA consensus expression profiling integrated from the HPA, GTEx and FANTOM5, the consensus normalized expression (NX) values between 20 and 40 were recorded as 'Medium', and the NX values above this threshold were recorded as 'High'. Similarly, the NX values in the range of 1–20 were recorded as 'Low', and for those falling below 1 were recorded as 'Not Detected'. The NX values in other cases were recorded as 'Not Available'.

Differential expression analysis of genes between tumour and its adjacent normal tissue
We gathered the uniformly processed TCGA and GTEx RNA-sequencing data from the RNAseqDB (https://github.com/mskcc/RNAseqDB). All together there were 9,109 high-quality samples covering 14 normal tissues and 19 types of solid cancer. The DESeq2 package [22] was used to identify HUGO genes that are differentially expressed between tumours

and their adjacent normal tissues. By setting a threshold of Benjamini-Hochberg adjusted p-values of 0.01 and log$_2$FoldChange of 1.0, those HUGO genes that were significantly upregulated in the tumours were retained.

Differential expression analysis of genes between tumour and other normal tissues

The read count data of RNA-sequencing gathered from the RNAseqDB was first transformed to TPM (transcripts per million) format that can be directly used to compare gene expression. To transform read counts into TPM format, we need to normalize for gene length, and then normalize for gene depth, in that order. For the gene length normalizing step, we fist calculated gene length from GTF files (GDC.h38 GENCODE v22 GTF (used in RNA-Seq alignment and by HTSeq)) by counting the longest transcript of each gene (sum of exons) or the sum of all exons, then divided each count by the length of its respective gene. For the gene depth normalizing step, we performed in the order as follows: 1) sum all counts within each sample column; 2) divide each column sum by the desired depth (1,000,000) to yield scaling factors; 3) divide each sample count within a column by its respective scaling factor.

The TPM values of each gene in its paired indication and normal tissues were used as input data for differential expression analysis. The non-parametric Mann–Whitney U test analysis was applied to calculate the differential expression ratios. The differential expression patterns of target genes between their paired indication and normal tissues were visualized via the OmicCircos package [23].

Tumour tissue transcriptome, genome, and phenotype compilation

To integrate transcriptome and genome and phenotype information of the gene set of interest. We downloaded the phenotype data of TCGA patients with various solid cancers from UCSC Xena (https://xenabrowser.net/datapages/), and then extracted and organized the information about gender, neoplasm histologic grade, pathologic stage, and TNM staging. The non-silent mutations (SNP and INDEL) for each gene in individual cancer type were determined through mining the MC3 ("Multi-Center Mutation Calling in Multiple Cancers") TCGA MAF (mutation annotation format) file. The gene-level transcription estimates (in log$_2$(x + 1) transformed RSEM normalized count format) were transformed to TPM format that can be directly used to compare gene expression. Thereafter, we compiled a comprehensive dataset via integrating the aforementioned tumour tissue transcriptome and genome information and the organized phenotype data. The heterogeneous expression pattern analysis was performed using the ggstatsplot package in batch mode (https://github.com/IndrajeetPatil/ggstatsplot).

Annotation of functionally relevant mutation

OncoKB (https://www.oncokb.org/) contains annotation information about the impact and therapeutic significance of 5616 specific alterations in 682 cancer genes. It combines multiple resources, including FDA, NCCN (National Comprehensive Cancer Network) and other guidelines, ClinicalTrials.gov and scientific literatures. We utilized the annotation information about oncogenic and clinically actionable alterations from the OncoKB and discarded somatic mutations that were labeled as likely oncogenic or predicted oncogenic, resulting in the set of driver mutations, which were not contaminated by passenger mutations. The collected information was applied to analyze the therapeutic significance of potential target genes that altered in a large proportion of patients.

Predicting overexpression rates of target antigens

We applied the method called functional genomic mRNA profiling to predict overexpression rates of target antigens [24]. Typically, principal component analysis (PCA) was used to analyze the mRNA transcriptome and n eigenvalues and n corresponding eigenvalues (transcriptional components) were subsequently obtained. We identified these subsets of transcriptional components that describe non-genetic regulatory factors (physiological and experimental factors) and used them as covariates to in multiple linear regression to correct the original gene expression data (the so-called functional genomic mRNA profiles that capture the effects of genomic alterations on gene expression levels). The overexpression percentages of each target antigens in samples per cancer type were determined based on the threshold that was defined in the set of

functional genomic mRNA profiles of normal tissues by calculating the 97.5th percentile for the functional genomic mRNA signal.

Gene set enrichment score analysis

We first downloaded the pan-cancer phenotype data and expression matrix from XENA (https://xenabrowser.net/datapages/?dataset=GDC-PANCAN.basic_phenotype.tsv&host=https%3A%2F%2Fgdc.xenahubs.net&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443), GDC-PANCAN.basic_phenotype.tsv) and the TCGA PanCanAtlas (https://gdc.cancer.gov/about-data/publications/pancanatlas, EBPlusPlusAdjust-PANCAN_IlluminaHiSeq_RNASeqV2.geneExp.tsv), respectively. We then extracted the paired samples and their corresponding expression matrix according to the rules of TCGA sample barcode (https://docs.gdc.cancer.gov/Encyclopedia/pages/TCGA_Barcode/). After which the enrichment score of gene set of interest was calculated based on the ssGSEA [25], and the z-score was transformed to evaluate the expression similarities and differences of gene set in TCGA pan-cancer.

Statistical analysis

All statistical analyses described above within the context of individual analyses in the Methods section were carried out using R statistical environment. The non-parametric Mann–Whitney U test and non-parametric Kruskal–Wallis one-way ANOVA were carried out for the analyses of two groups and more than two groups, respectively.

RESULTS
Assembling a comprehensive dataset and designing an algorithm to identify ADC targets

We assembled a comprehensive annotation dataset and designed an algorithm to discover surface molecules differentially expressed between tumour tissues and normal tissues. We used RNA-sequencing data to discover differentially expressed therapeutic targets, while the best-in-class RNA-seq data processing pipelines were shown to produce consistent expression estimates for just 88% of protein-coding gene [26]. A pipeline was successfully developed to unify RNA-sequencing data of TCGA and GTEx by minimizing differences between matching tissues from TCGA and normal tissues from GTEx, and by processing data sets uniformly without the inter-project normalization step [27]. We performed differential analysis for 20,242 HUGO genes across 19 solid cancer types based on the sequencing data processed from this pipeline. By setting a threshold of Benjamini-Hochberg adjusted p-values of 0.01 and log$_2$FoldChange of 1.0, we kept those HUGO genes significantly overexpressed in tumour cell populations. To search for potential extracellular membrane proteins from the overexpression lists, we compiled a membrane protein annotation dataset (6176 entries) from three databases (i.e., Uniprots, HPA, and Compartments) that contain protein subcellular location information, and then merged it with the above overexpression lists. To exclude molecules highly expressed in normal tissues, we used the mRNA expression dataset integrated from the GTEx, HPA, and FANTOM5, and supplemented the dataset with proteome data from the HPA and HPM. We excluded molecules exhibiting high expression in the vital tissues (i.e., heart, liver, spleen, lung, kidney), as well as molecules exhibiting high expression in more than three normal tissues except for the blood, bone marrow, brain, and placenta. Thereafter, we used the publicly available resource BloodSpot [28] and found those surface molecules that are restrictedly expressed in the bone marrow Lin$^-$ CD34$^+$ CD38$^-$ CD90$^+$ CD45RA$^-$ HSCs and Lin$^-$ CD34$^+$ CD38$^-$ CD90$^-$ 45RA$^+$ MPPs. The next quality control step further leaved us with 90 potentially therapeutic surface membrane proteins and their 243 target-indication combinations (Supplementary Table 3).

Clinical responses are more frequently observed in patients with tumours displaying high levels of target antigens. To refine our selection, we analyzed target antigen expression patterns in both tumour and normal tissues using publicly available gene expression data. This step resulted in the identification of 87 candidates and their corresponding 225 target-indication

combinations. To address heterogeneity in gene expression within the patient population, we leveraged comprehensive oncology repositories with paired genomic and transcriptomic data. This enabled the exclusion of target antigens exhibiting significant decreases in gene expression among patients bearing specific gene mutations, resulting in 84 candidates and 213 target-indication combinations. Additionally, we assessed tumour heterogeneity by predicting protein overexpression rates across 22 different tumour subtypes, guiding our selection for subsequent immunohistochemistry validation. Utilizing this algorithm, we ultimately pinpointed 75 molecules with characteristics conducive to ADC targeting, encompassing 165 target-indication combinations (Supplementary Table 4). In the final stage, we conducted a thorough evaluation based on twelve criteria across four dimensions: expression profiling (including differential expression, absolute levels in tumour, expression homogeneity, normal tissue expression, and expression in hematopoietic stem cells and multipotential progenitors), internalization (including endocytosis and trafficking route), complexity of target antibody discovery (including single/multiple transmembrane, extracellular domain size, and extracellular domain homology), and biological functions (encompassing genetic basis and cancer stem cell features).

Identified ADCs and their target antigens in clinical trials

A total of 159 ADCs were identified from 760 clinical trials in ClinicalTrials.gov, 406 of which are actively undergoing evaluation in various stages across hematologic malignancies and solid cancers (Supplementary Table 2). Specifically focusing on solid cancer types, 72 ADCs targeting 36 antigens identified from 227 interventional studies are currently under evaluation (Fig. 1a). Meanwhile, 118 trials for 51 ADCs are completed, 46 trials regarding 31 ADCs are terminated, and statuses of 6 trials on 5 ADCs are unknown (Supplementary Table 2). The overwhelming majority of ADC trials for solid cancers was initiated in 2008 or later. Notably, BRCA stands out as the only indication with new clinical trials initiated every year since 2006 (Fig. 1b). Among the various solid cancer types, BRCA had the highest number of evaluated ADCs and trials ($n_{ADC} = 28$, $n_{study} = 139$), followed by pan-cancer ($n_{ADC} = 73$, $n_{study} = 113$), NSCLC and SCLC ($n_{ADC} = 21$, $n_{study} = 43$), BLCA ($n_{ADC} = 7$, $n_{study} = 18$), OV ($n_{ADC} = 12$, $n_{study} = 18$), STAD ($n_{ADC} = 11$, $n_{study} = 18$), and PRAD ($n_{ADC} = 9$, $n_{study} = 18$) (Fig. 1b and Supplementary Table 2). However, in the case of LIHC, the lone trial assessing the safety and tolerability of the GPC3-directed ADC in patients was initiated in 2016 and subsequently terminated without publicly disclosed results. Similarly, the investigation of ADC for CESC exclusively revolves around the TF target (also known as F3). Additionally, there are no clinical trials focused on ESCA and THCA, despite them being common cancer types accounting for 3.1% and 3.0% of total cancer cases, respectively (https://gco.iarc.fr/today/home).

Iver the past two decades, a total of 397 clinical trials have been conducted for solid cancers across various stages, with 163 in Phase I, followed by 127 in Phase II, 65 in Phase I-II, 37 in Phase III, 4 in Phase II-III, and 1 in Phase IV. These trials have involved 67 unique antigens, 32 of which have been studied in two or more trials (Fig. 1c). Among these, twenty-one ADCs are directed against HER2, followed by TROP2 ($n = 5$), MSLN ($n = 4$), and FOLR1 ($n = 3$), each of which has attracted at least three different ADCs (Fig. 1c). HER2 is the only target that has been the focus of at least 20 clinical studies in each of Phase I, II, and III. Moreover, HER2, TROP2, MSLN, FOLR1, and DLL3 are the five antigens targeted by at least one ADC in more than ten trials (Fig. 1c). Currently, ten ADCs are undergoing Phase III trials, including the registered trastuzumab emtansine (T-DM1, HER2-directed), enfortumab vedotin-ejfv (ASG-22CE, NECTIN4-directed), trastuzumab deruxtecan (DS-8201a, HER2-directed), sacituzumab govitecan (IMMU-132, TROP2-directed), and six other ADCs. Among these, two target

CEACAM5 and FOLR1 for NSCLC and OV, respectively. Furthermore, twenty ADCs are being evaluated in Phase II trials, while two ADCs are in Phase II/III trials. Nine ADCs that reached Phase II did not progress to Phase III, and 27 ADCs that reached Phase I did not advance to Phase II. As for the ADCs currently in Phase II or Phase II/III trials, their target antigens, such as F3 (TF), ROR1, FOLR1, CD56 (NCAM1), ENPP3, MET, CD166 (ALCAM), HER3 (ERBB3), MSLN, and SLC39A6 (LIV1), have not yet been introduced to the ADC market (Fig. 1a). According to the expression dataset, the target antigens of drugs on the market (i.e., HER2, EGFR, NECTIN4, TROP2) used to treat solid cancers show limited expression in normal tissues. However, target antigens currently under investigation in clinical trials exhibit relatively high expression in normal tissues. For instance, CD166 (ALCAM) and CD56 (NCAM1) show high expression in the kidney, liver, lung, and heart (Fig. 1a). Overall, the target antigens in drugs that have been approved or are under active trials demonstrate high expression in various solid cancers, particularly in LUAD. In contrast, their expression is significantly lower in LIHC (Fig. 1d). This suggests that most ADCs are developed to target LUAD, while LIHC lacks suitable target antigens. This distinction may be attributed to the fact that LIHC exhibits relatively unique genome-wide global expression profiles with almost no overlap with other cancer types [29].

Differential gene expression patterns between tumour and normal tissues

Out of the 90 target antigen candidates identified through our algorithm, it was observed that CELSR1, GPR87, OR51E1, SLC2A12, and VANGL1 do not exhibit medium to high expression in the 32 specified tissues of interest. Additionally, there are 26 targets with medium to high expression in one tissue, 18 in two tissues, and 14 in three tissues (Fig. 2). CD276, ERBB2, ERBB3, CMTM4, and GPNMB exhibit medium or medium-high expression in more than nine normal tissues, but CD276, ERBB2, and ERBB3 are not highly expressed in any tissues except for the brain. Among the 243 target-indication combinations, MUC17-STAD is the combination that has the highest differential ratio ($\log_2$ transformed) of MUC17 expression level in gastric tumour tissue to that of normal gastric tissue, reaching 9.8, followed by VTCN1-UCEC, DSC3-LUSC, GPR87-LUSC, GJB6-LUSC, and GPC3-LIHC, with ratios that reach 9.1, 8.1, 8.1, 8.1, and 7.8, respectively. Similarly, the differential expression ratio of MUC17-ESCA is also as high as 12.3, but the absolute gene expression level of MUC17 in ESCA is too low for this pair to be a target-indication combination. It's important to note that while the differences in ERBB2 expression levels between 19 solid cancer tissues and their adjacent normal tissues may not be highly pronounced, it is still the target with the highest accumulation of ADCs. We will delve deeper into this issue in the subsequent discussion. TACSTD2 (TROP2) accounts for the second largest number of ADC, while the first approved indication for TROP2-directed sacituzumab govitecan is BRCA, the pair with the lowest differential expression ratio. For several other indications such as BLCA, NSCLC, PRAD, and UCEC with relatively high differential expression ratios, five Phase II trials and two Phase III trials were launched in the past three years 2018–20 (Supplementary Table 2).

In addition to the analysis of differential expression relative to adjacent tissue, we also calculated the differential expression ratios of target gene expression level in the paired indication to that of other normal tissues (Fig. 3). Our findings reveal that the esophagus is the least conducive environment, as in 14.5% of the target-indication combinations, the absolute gene expression levels of targets in their paired indications are lower than their expression levels in the normal esophagus. On the other hand, the liver, gallbladder, and pancreas are the three most favorable tissues, as in none of the target-indication combinations is the target gene expression level lower than that in these three normal tissues. While the differential expression ratios for ERBB2 may not be particularly striking,
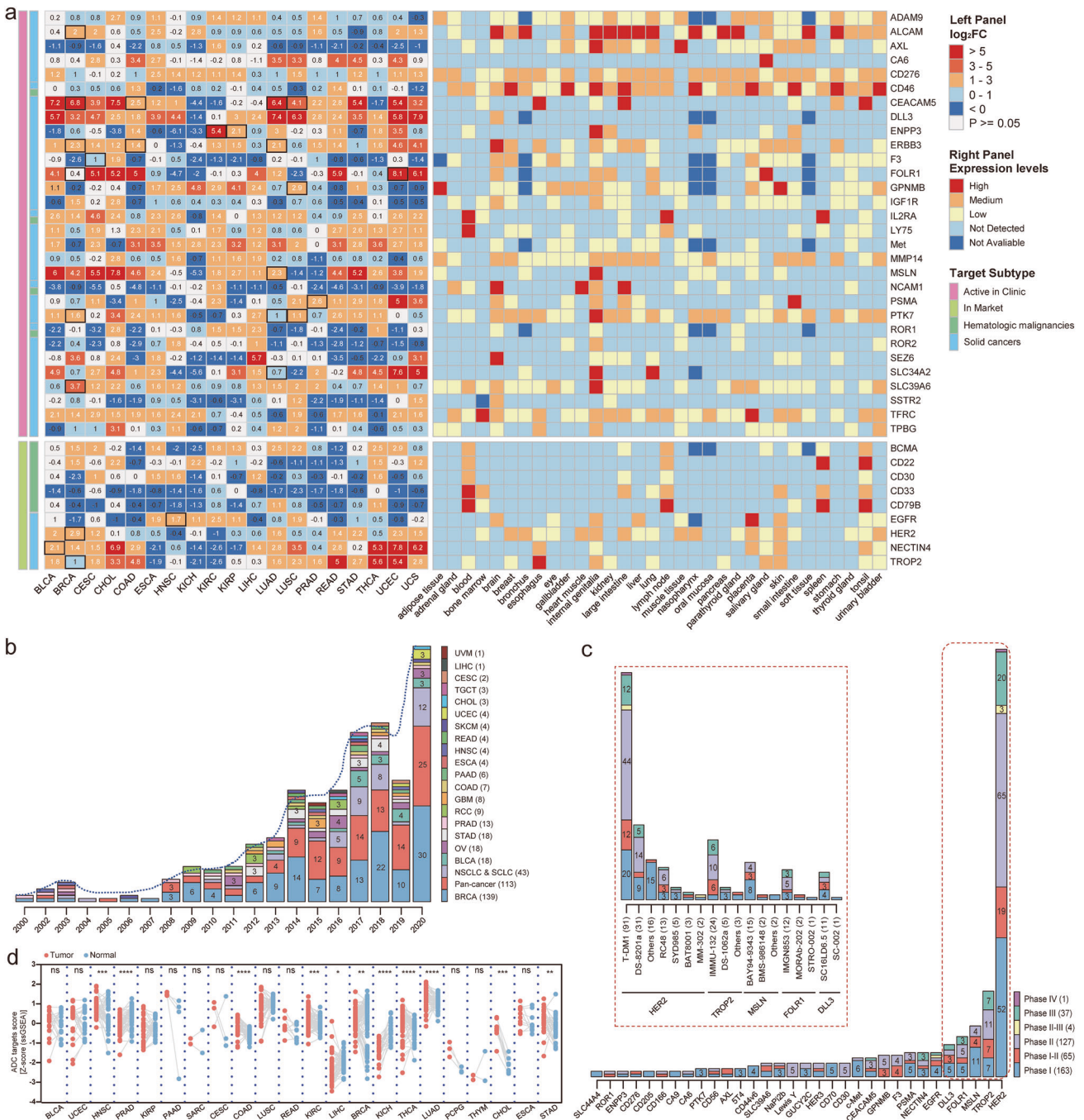
Fig. 1 Expression profile and differential expression profile of ADC targets and overview of the clinical development pipelines and current ADC targets. a Differential gene expression profile of target antigens in active clinical trials and approved ADCs among 19 types of solid cancer (left panel). log$_2$FC is the effect size estimate between tumour and adjacent normal tissue produced from DESeq2. The black box in the left panel encircles the paired indications of targets. Protein expression of a given target set by normal tissue groups (right panel). **b** Year of clinical study initiation for ADCs among various solid cancers. The dotted trend line fitting the total number of studies per year. **c** Targets of clinical-stage ADCs ranked by the total number of ADCs directed to. The histogram in the box represents the clinical stage distribution of ADCs directed to the top five targets. **d** Gene expression enrichment scores of targets in active clinical trials and approved ADCs in pan-cancer.

its expression level in each paired indication is at least double that of each normal tissue. For TACSTD2, it exhibits significant differential expression between numerous tumor tissues and their paired normal tissues, often more than fivefold. However, its performance is relatively subpar in the esophagus and salivary gland. CD276 is associated with many target-indication combinations, yet its expression level in each paired indication

is less than double and sometimes even lower than its expression level in the adrenal gland, breast, cervix, prostate, and uterus. Furthermore, TNFRSF21 and TPBG show higher expression in the urinary bladder than in their paired indications, and DLK1 expression level in LIHC is lower than its expression level in the normal adrenal gland, bone marrow, ovary, and testis tissues.
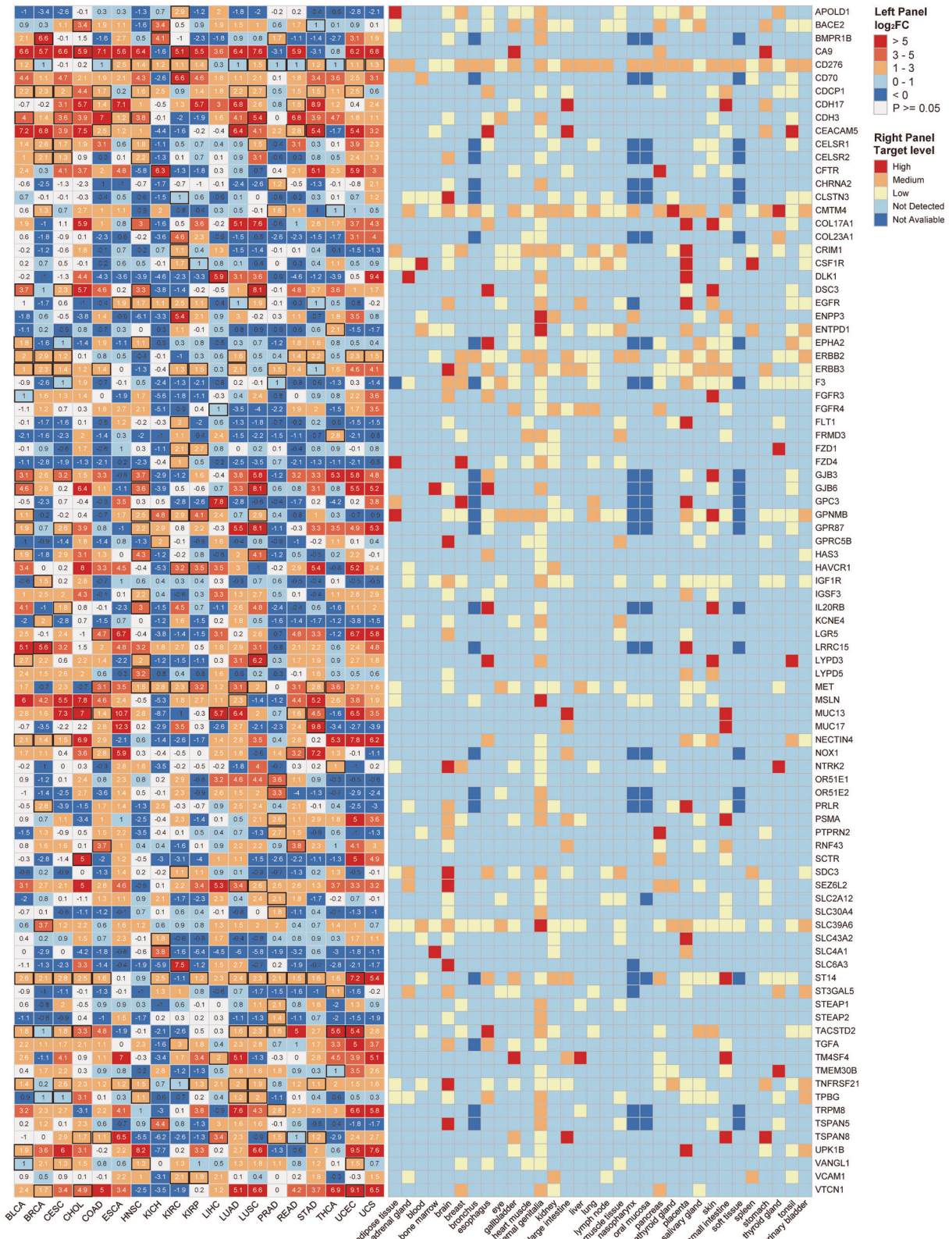
J. Fang et al.

Fig. 2 Expression profile and differential expression profile of selected target antigens screened from the algorithm. Differential gene expression profile of ADC target candidates in 19 types of solid cancer (left panel), and log₂FC is the effect size estimate between tumour and adjacent normal tissue produced from DESeq2. The black box in the left panel encircles the paired indications of targets. Protein expression profile of ADC target candidates by the rearranged 32 normal tissue groups is shown in the right panel.
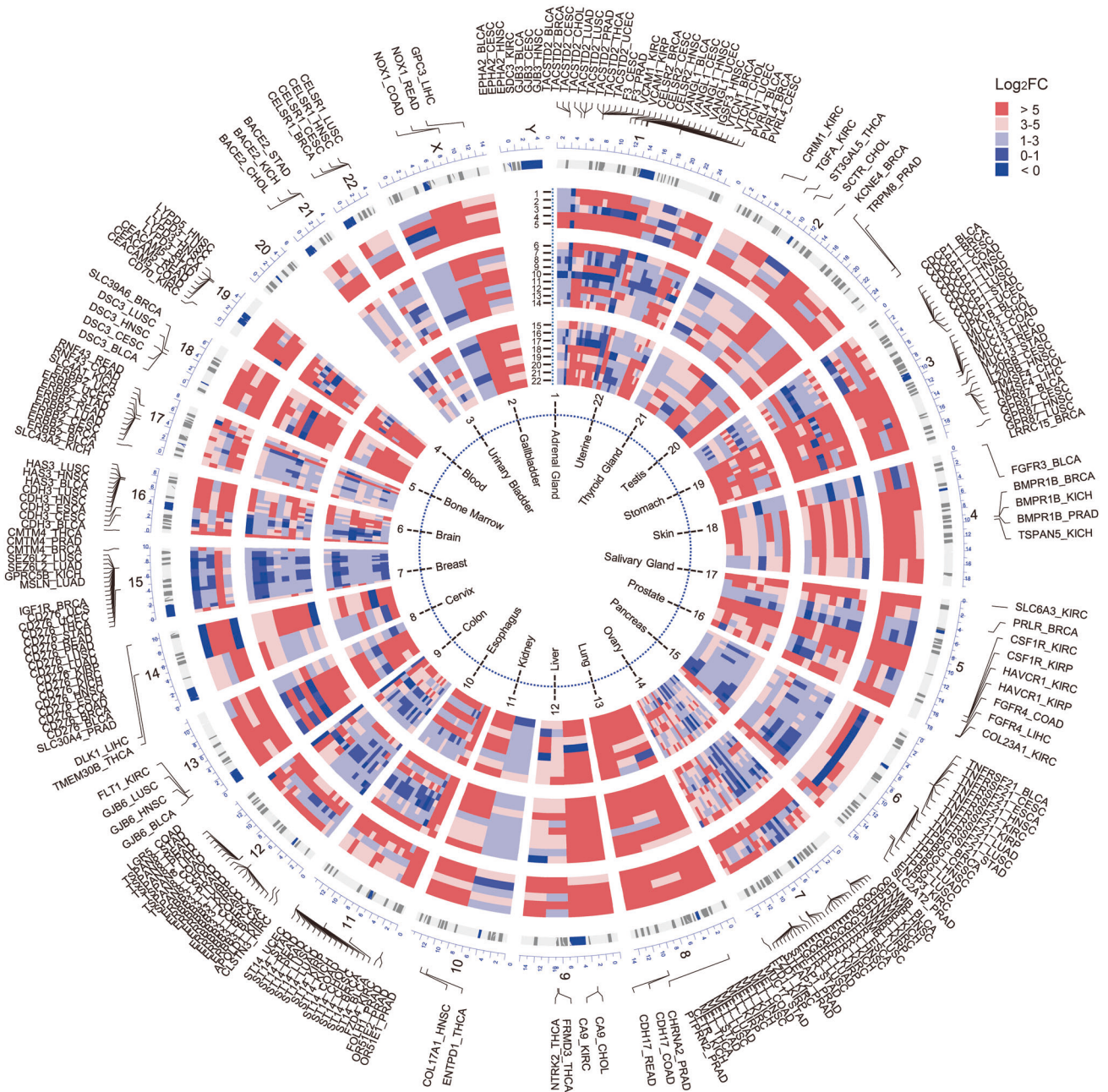
Fig. 3 Circular visualization of the differential gene expression profile of target antigens in normal tissues. Specific chromosomal location of each gene is shown by lines coming from each target-indication combination pointing to a specific position on each chromosome, with cytobands also included. Heatmap weights are based on the differential value of each target gene expression level in its paired indication to that in each normal tissue. log₂FC is calculated by the non-parametric Mann–Whitney U test analysis. FC refers to the ratio of gene expression value in the paired indication to that in normal tissues. The numbers 1-22 arranged vertically represent normal tissues.

Heterogeneous gene expression patterns in solid cancers

Homogeneous gene expression in clinical indications is a crucial characteristic for antigens to be considered as potential therapeutic targets. Through a comprehensive analysis of paired genome and transcriptome data, we identified that among 13 types of solid cancer, there are 146 target-indication combinations where 65 mutated genes exhibit a significant relationship with the gene expression level of 51 targets (Fig. 4a and Supplementary Table 5). Among them, *TP53* mutations exhibit the strongest correlation with target gene expression. Specifically, in 21 out of the 146 combinations, *TP53* mutations, both in coding and disruptive forms, have a significant impact on the expression of target genes (Fig. 4b). For instance, in

34.3% of BRCA samples with mutated *TP53*, the expression level of *BMPR1B* is 4.9 times lower compared to samples without *TP53* mutation. Similarly, *KCNE4* and *SLC39A6* exhibit a reduction in expression levels by 76.1% and 72.5%, respectively. Another example is the elevated expression of *TACSTD2* in 58.8% of THCA samples bearing a *BRAF* mutation, showing a 3.9-fold increase. Additionally, *FGFR3* expression is upregulated by 3.4 times in 13.4% of BLCA patients with a self-gene mutation. Several other notable correlations include *ERBB2_PTEN* (57.4% of BLCA samples, FC = 3.2, direction = down), *MSLN_TP53* (47.4% of LUAD samples, FC = 2.6, direction = down), *MET_-BRAF* (58.8% of THCA samples, FC = 2.5, direction = up), and *GPNMB_VHL* (49.1% of KIRC samples, FC = 2.3, direction =
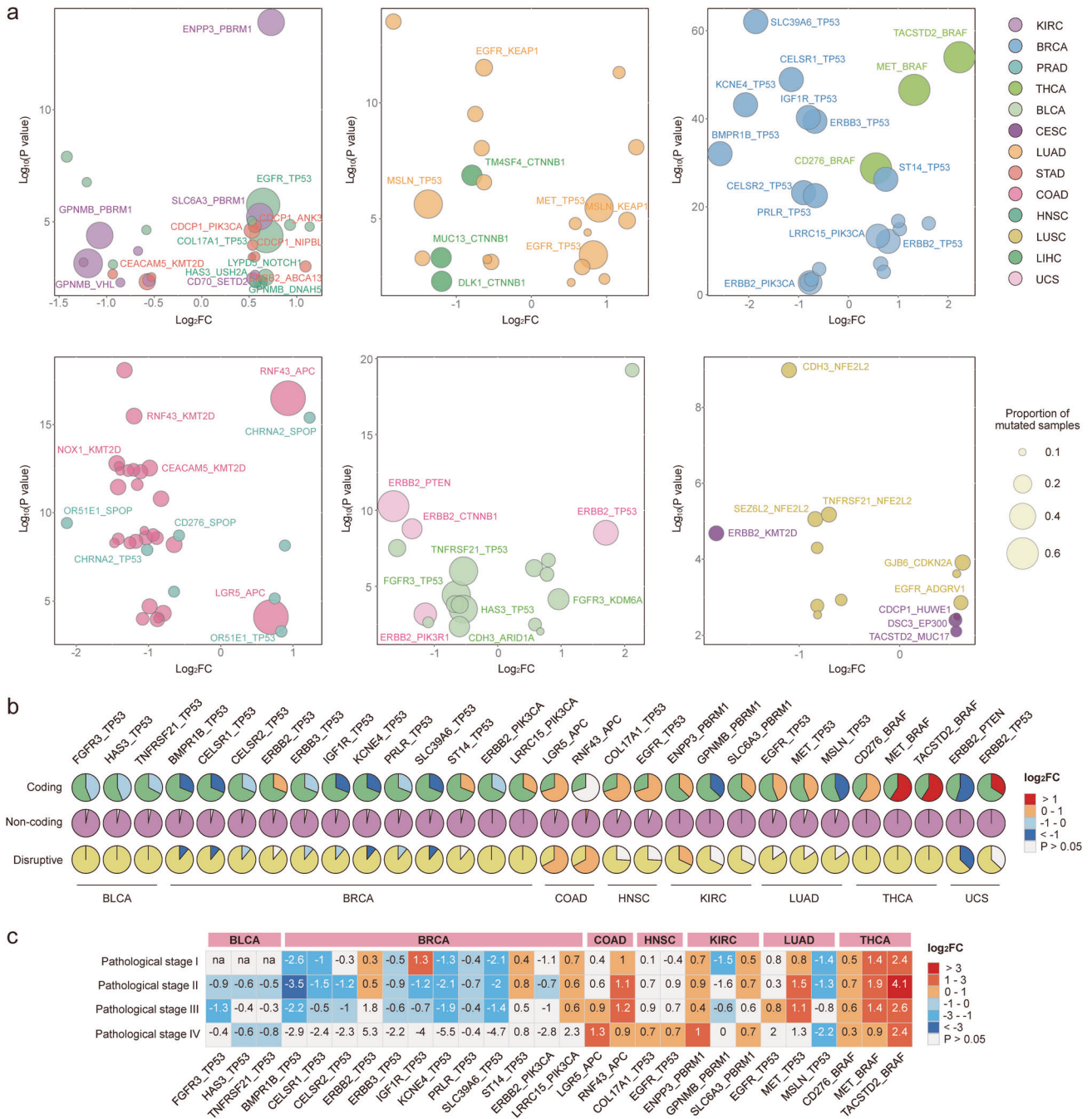
Fig. 4 Heterogeneous gene expression pattern of several target candidates. a Bubble-volcano plot represents the fold change of target gene expression value in the mutant samples to that in the wild samples among different solid cancer types, with the proportion of patients harbouring a gene mutation also included. The color represents different cancer types, and the size of the bubbles represents the proportion of patients bear a gene mutation. Six subplots are used to ensure clarity. **b** Pie chart represents the fold change of target gene expression level in the mutant samples to that in the wild samples by gene mutation types. The size of the fan-shaped patches painted with the color indicating fold change represents the proportion of patients harbouring each mutation type of a given gene. **c** Rank-heatmap represents the fold change of target gene expression value in the mutant samples to that in the wild samples by pathological stages. log₂FC and *P* values are calculated by the non-parametric Mann–Whitney *U* test analysis. The numerical values on the rank-heatmap represent log₂FC.

down) (the former is the target gene, and the latter is the mutated gene). Moreover, the correlation between gene mutation and the expression levels of certain target genes varies across different pathological stages (Fig. 4c). For example, *TP53* mutation has the greatest correlation with the downregulated *BMPR1B* expression in BRCA pathological stage II, and *BRAF* mutation is most related to the upregulation of *TACSTD2* expression in the pathological stage II of THCA.

Notably, *IGF1R* expression is upregulated in the pathological stage I of BRCA patients harbouring mutated *TP53*, while it is downregulated in stage II and stage III.

Irrespective of gene mutations, the expression levels of genes also exhibit variations across different pathological stages and tumor sizes, and further differ in cases of tumor invasion, peripheral lymph node metastasis, and distant metastasis (Supplementary Figs. 1–5). For instance, as cancer progresses from pathological stage I to stage
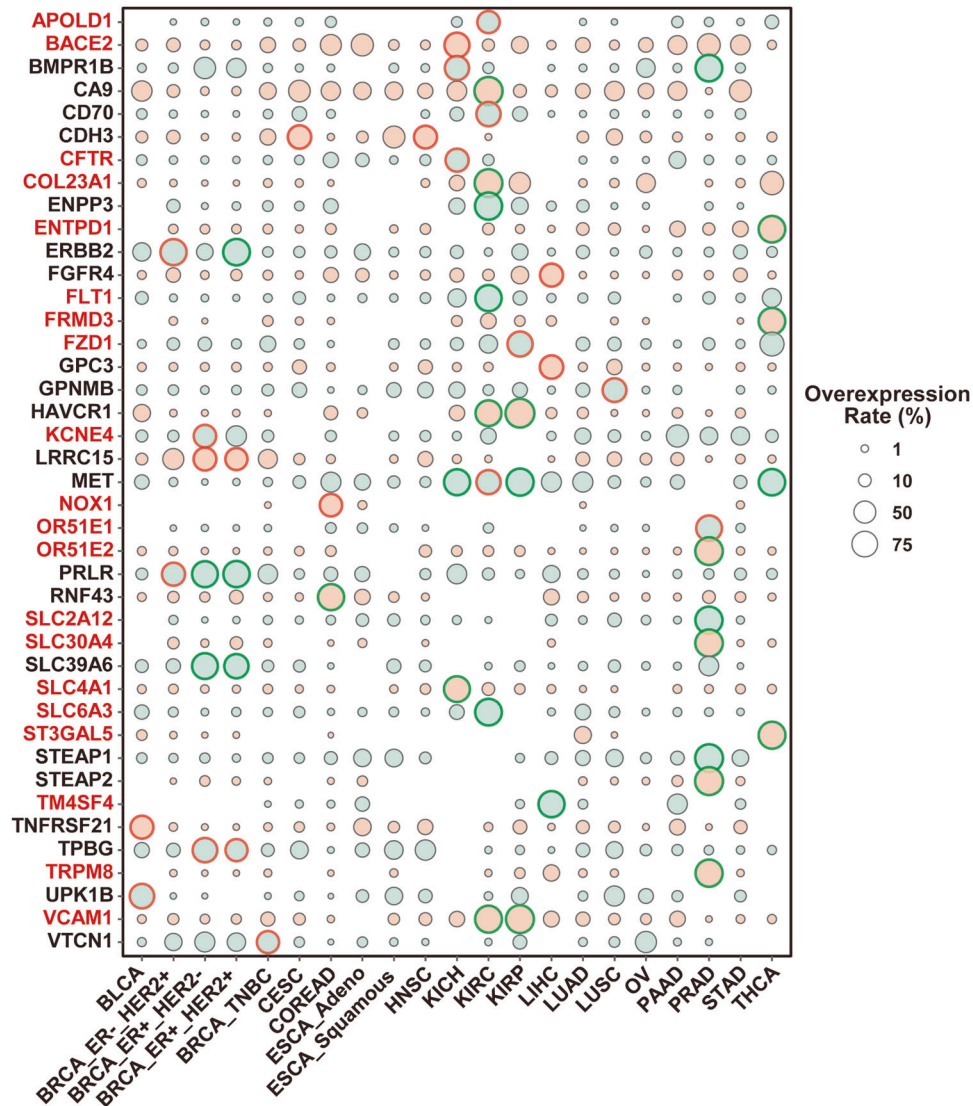
Fig. 5 Predicted target overexpression analysed by functional genomic mRNA profiling. Predicted protein overexpression rates per tumour (sub)types are represent as dots. The size of dots indicates the percentage of patient-derived tumour samples. The green circle around dots represents a percentage >75% and the red circle represents >50% and <75%. The y-axis represents these unreported target antigens (brown) and current ADC targets (black) identified from our algorithm and discovery platform, which possess a predicted overexpression rate greater than 50% in at least one tumour (sub)types.

IV, the expressions of *APOLD1* and *ST14* undergo a gradual down-regulation in KIRC and KIRP, respectively (Supplementary Figs. 2, 3). Conversely, in THCA, the expressions of *MET* and *TACSTD2* experience an initial down-regulation followed by an up-regulation (Supplementary Figs. 4, 5). Notably, the expression level of *CDH3* in ESCA at pathological stage II is approximately double that in the other stages (Supplementary Fig. 1). As tumors grow and spread into nearby tissue, the expressions of *APOLD1* and *ST14* in KIRC and KIRP tend to decrease. In contrast, the expressions of *MET* and *TACSTD2* in THCA show gradually upregulation. As tumors spread to nearby lymph nodes and distant metastatic sites, the expressions of *APOLD1* and *ST14* show a downward trend in KIRC and KIRP, respectively. Conversely, in THCA, the expressions of *MET* and *TACSTD2* are notably upregulated as the tumor spreads to nearby lymph nodes and distant sites.

Predicted protein overexpression rates of target antigens
To further characterize the tumour heterogeneity of the target antigen, we utilized expression profiles to predict protein overexpression rates for the 75 unique target antigens across 22

different tumor (sub)types (Supplementary Table 6). Notably, over 75% of ADC targets in their respective 243 target-indication combinations exhibited a predicted overexpression rate of >10% of samples. Additionally, a predicted protein overexpression rate >75% of samples was observed for 24 ADC targets in 30 target-indication combinations, and >50% for 41 in 54 combinations (Fig. 5).

In BRCA (sub)types, a predicted overexpression rate of >77% of samples for HER2 was observed in ER-negative/HER2-positive or ER-positive/HER2-positive breast cancer. In ER-positive/HER2-negative and triple-negative breast cancer (TNBC), the over-expression rate dropped to 23% and 5%, respectively. For BMPR1B, CELSR1, ERBB3, IGF1R, KCNE4, SLC39A6, and TPBG, their overexpression rates in ER-positive breast cancer samples are significantly higher than that in ER-negative samples, while for CD276, LRRC15, PRLR, and VTCN1, each protein was observed to show overexpression independent of breast cancer intrinsic subtypes. In TNBC, the highest predicted overexpression rate with 52% was observed for VTCN1 (B7-H4), followed by 43% for CD276 (B7-H3). VTCN1 is a target antigen that has not been clinically

explored in breast cancer. In kidney cancer (sub)types, it was observed that VCAM1 and HAVCR1 are two targets highly overexpressed in both KIRC and KIRP, while MET was observed to show overexpression independent of kidney cancer intrinsic subtypes. In KIRC, the highest predicted overexpression rate with 93% was observed for CA9, while in KIRP and KICH, 100% and 78% were observed for HAVCR1 and MET, respectively. In addition, several other unreported targets screened from our algorithm, including APOLD1, BACE2, CFTR, COL23A1, FLT1, FZD1, SLC4A1, SLC6A3, were observed to be overexpressed in > 50% kidney cancer (sub)type samples. In BLCA we observed the highest predicted overexpression rate of 67% for UPK1B, followed by 56% for TNFRSF21. In COREAD, RNF43 ranked the highest with a predicted overexpression rate of 78%, followed by NOX1 (53%). In LIHC, it was observed that, three target antigens -TM4SF4, GPC3, FGFR4- show the top three highest overexpression rates. PRAD was observed to be the cancer type with the largest number of target antigens overexpressed in more than 75% samples, TRPM8, SLC2A12, SLC30A4, and OR51E2 among them have not been reported for ADC application.

Genetic basis of target antigen candidates

We systematically analyzed the genetic basis of the 90 targets, and found that *ERBB2*, *EGFR*, *MET*, and *FGFR3* as oncogenic driver are frequently altered in a wide type of solid cancers. The most common alterations in oncogene *EGFR* are mutation (5.5%), amplification (2.6%), exon 19 mutation (1.8%), exon 19 deletion (1.6%), and exon 21 mutation (1.3%). When it comes to LUAD, these proportions rise to 26.1%, 5.3%, 11.5%, 10.7%, and 8.6%, respectively [30] (Fig. 6a). The annotation about *EGFR* alterations curated by OncoKB shows that exon 19 deletion, L858R, and T790M, as well as amplification are all to be oncogenic and have gain-of-function effects. *ERBB2* as another oncogene alters in 13.9% of BRCA patients with amplification and mutation present in 12.6% and 3.4% of all patients, respectively (Fig. 6a). *FGFR3* mainly alters in BLCA patients, and mutations account for the overwhelming proportion of its alteration. S249C is the main source the mutations, followed by Y373C, G370C, and G380R (Fig. 6a). *MET* amplification and mutation occur in many solid tumours, but the alteration frequency in each cancer type does not exceed 5% (Fig. 6a). Copy number variations (CNVs) are usually propagated to the proteome level, while post-transcriptional mechanisms attenuate this impact. For *ERBB2*, the greatest agreements between its transcript and CNVs and transcript and protein, are observed in BRCA (Fig. 6b). Renal cell carcinoma, however, are with poor correlations. Similarly, *EGFR* shows good agreement in transcript/ CNV and transcript/protein in multiple cancer types (Supplementary Fig. 6). We applied radar charts to exhibit various indicators under different measurement systems. The performance of ERBB2 on BRCA and BLCA is the most prominent among all target-indication combinations, although it is moderately expressed in many normal tissues (Fig. 6c). Compared with the ERBB2-BRCA or ERBB2-BLCA combination, performance of EGFR on LUAD falls behind in terms of the expression profiling. But its performance is superior to FGFR3-BLCA with regard to cancer stem cell features. Regardless of the two factors of genetic basis and cancer stem cell features, the performance of NECTIN4 on BLCA is impeccable; so is the performance of MET in LUAD, except for its expression heterogeneity (Fig. 6c).

DISCUSSION

The conventional approach to researching ADC target antigens typically centers on comparing their expression in tumor tissues with their corresponding normal counterparts, often neglecting the comprehensive assessment of systemic expression in potential target candidates. This oversight can lead to an underestimation of the potential risks of toxicity across normal tissues. To address this, we compiled an extensive human proteome map by integrating various proteomics databases. This integration, which combines data from immunohistochemical analysis and mass spectrometry, enhances the confidence in the reliability of identifying low-level expression. Drawing on lessons learned from clinical experiences, we recognize that achieving a balanced level of on-target toxicity between normal and tumor cells is crucial for the success of clinical trials. Therefore, we systematically analyzed the differential expression profiles of candidate target antigens not only in relation to their paired indications and normal counterparts but also in comparison to other normal tissues. By emphasizing antigens with both high expression in tumor tissues and low expression in normal tissues, we identified targets with significant differential expression profiles. Additionally, targets with relatively uniform expression patterns among patients with a given indication are more likely to benefit a larger population. Through the utilization of extensive oncology repositories, we excluded antigens whose gene expression levels experienced significant decreases in a large proportion of patients bearing a gene mutation. Furthermore, we conducted a thorough annotation of the tumor heterogeneity of target antigens based on functional genomic mRNA profiling. It's worth noting that target selection should also take into account the expression in hematopoietic stem cells (HSCs) and multipotential progenitors (MPPs), as restricted expression in these cell populations could enable them to consistently replenish, even after depletion by ADCs. Overall, starting from 20,242 HUGO gene symbols across 19 solid cancer types, our algorithm identified 75 candidate antigens with features potentially suitable for ADC targeting and their 165 target-indication combinations.

In the RTK/RAS/MAP-Kinase signaling pathway, several particularly interesting receptor tyrosine kinases in our selection list (Supplementary Fig. 7), including ERBB2, ERBB3, EGFR, FGFR3, and MET, are recurrently altered by mutation, amplification and/or overexpression in a large proportion of patients with various types of cancer. Cancer type-specific alterations of these oncogenic genes have different levels of clinical actionability. As in the biomarker-drug association case for *ERBB2* amplification in BRCA, the first-line treatment is 'dual HER2 blockade' with trastuzumab and pertuzumab plus a taxane. The mechanism of action of T-DM1 is relevant to not only the DM1 payload, but also the trastuzumab moiety since it retains all the mechanism of action of the naked antibody [31]. The results of DESTINY-Lung01 showed that the overall confirmed ORR of DS-8201 in the HER2 overexpression cohort was 24.5%, while in the *HER2* mutation cohort, the confirmed ORR was as high as 61.9% [32]. Furthermore, cancer stem cells that are tumourigenic may be resistant to conventional cancer therapies, and efforts are now being directed towards exploring precision medicine to target these cell types. Several target antigens in our target atlas, including LGR5, EGFR, ERBB2, CSF1R, MET, IGF1R, and TPBG (5T4), are cell surface antigens with characteristics of cancer stem cells, and thus might enable the ADCs directed to these targets to suppress cancer relapse and metastasis.

It is commonly agreed that faster ADC endocytosis and deeper penetration into late lysosomes are favorable, some targets exhibit remarkably slow endocytosis without precluding them from being considered as ADC targets. Conversely, the HER2 double epitope ADC (MEDI4276) demonstrates exceptionally rapid endocytosis, but its toxicity levels are too high to be well-tolerated. The strict internalization requirement in ADC development has been challenged in the past decade [33]. Non-internalizing ADC were shown to induce a potent lineage ablation for hematologic malignancies and solid tumours [34–36]. Target antigen candidates that lack high degree of credibility regarding endocytosis and endocytic trafficking routes were therefore retained in our selection list. Regarding the debate between fixed-point and random coupling, the consensus leans towards fixed-point
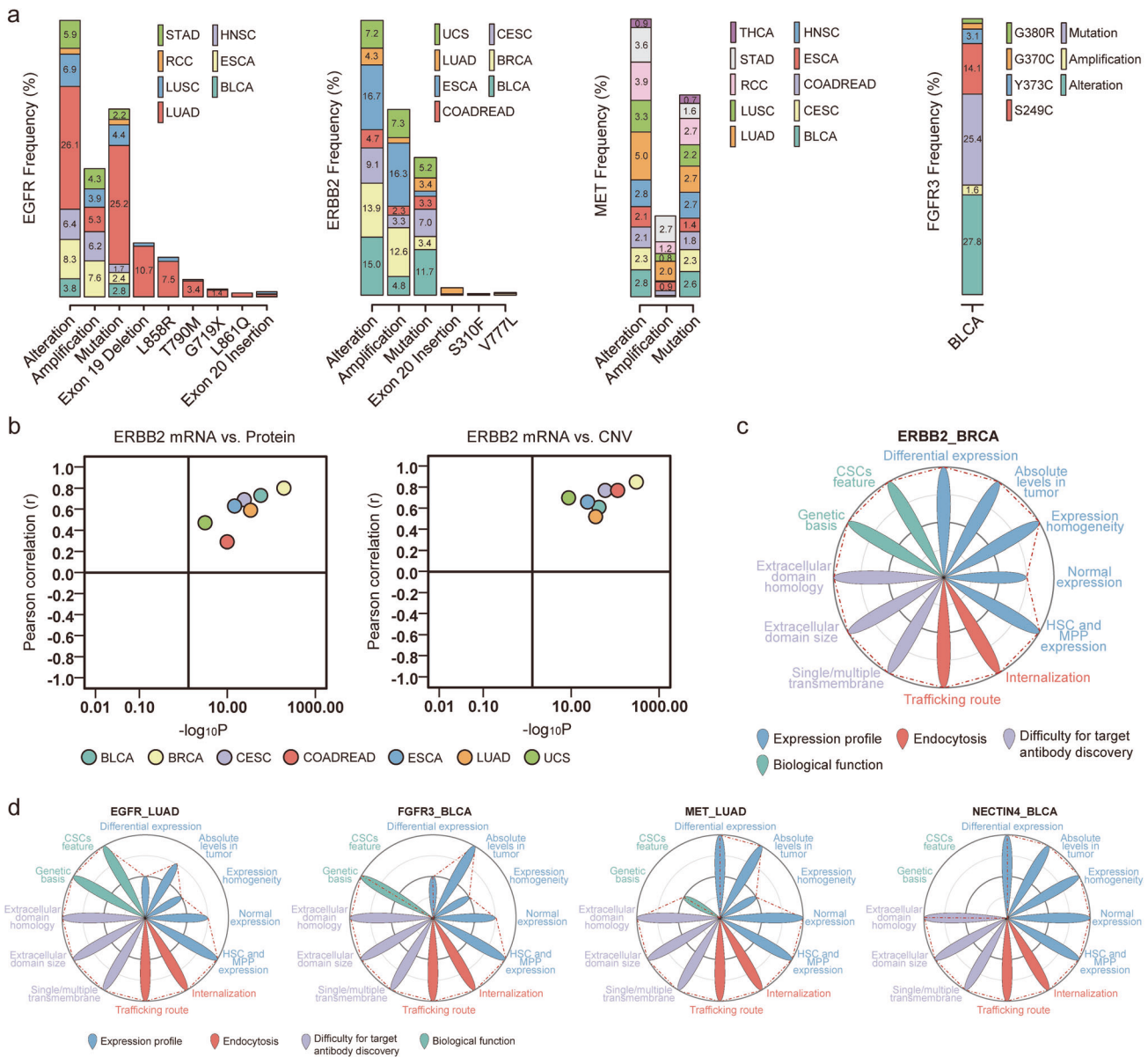
Fig. 6 **The common genetic alterations of oncogenes and a combinatorial analysis of their coding protein as ADC targets. a** Genetic alteration frequencies of EGFR, ERBB2, MET, and FGFR3 in different types of solid cancer. The numerical values on the bar chart represent genetic alteration frequencies. **b** Correlations of ERBB2 mRNA expression with its protein level or copy number variation events calculated via Pearson's correlation analysis. **c, d** Radar charts show the combinatorial scoring of five representative target-indication combinations focusing on four dimensions and twelve aspects.

coupling as superior. However, when the fixed-point coupling HER2 ADC climbs to 1.5 mg/kg, its on-target cardiotoxicity becomes strikingly apparent, necessitating dosage limitations. Meanwhile, the random coupling RC48, also targeting HER2, can be administered at 2.5 mg/kg with considerably lower cardiotoxicity. Furthermore, ADC dosages are influenced by factors like payload potency, DAR, effective internalization, linker cleavage, payload release, and target cell responsiveness. Adjusting a single parameter, like the effective internalization rate, would lead to an increase in local payload concentration within the tumor and healthy tissue. Consequently, the starting dosages along with the maximum tolerated dosages are reduced. Therefore, ADC development poses a complex challenge, wherein modifying a parameter will result in corresponding adjustments in the dosages.

This study, leveraging a comprehensive annotation database built on transcriptomics, proteomics, and genomics, represents the inaugural effort in uncovering pan-cancer ADC targets. The extensive landscape of ADC targets compiled herein serves as a valuable resource for cancer researchers, clinical oncologists, and the wider scientific and pharmaceutical community invested in the rapidly evolving field of ADC oncology therapeutics for solid cancers.

DATA AVAILABILITY
The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

CODE AVAILABILITY
The custom computer codes utilized during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. WHO. World cancer report. France: Imprimerie Faurite; 2020.
2. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: Retrospective cohort study of drug approvals 2009-13. BMJ. 2017;359:j4530.
3. Beck A, Goetsch L, Dumontet C, Corvaïa N. Strategies and challenges for the next generation of antibody-drug conjugates. Nat Rev Drug Discov. 2017;16:315–37.
4. Chau CH, Steeg PS, Figg WD. Antibody–drug conjugates for cancer. Lancet. 2019;394:793–804.
5. Modi S, Saura C, Yamashita T, Park YH, Kim SB, Tamura K, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Positive Breast Cancer. N. Engl J Med. 2020;382:610–21.
6. Lyon R. Drawing lessons from the clinical development of antibody-drug conjugates. Drug Discov Today Technol 2018;30:105–9.
7. Van Geel R, Wijdeven MA, Heesbeen R, Verkade JMM, Wasiel AA, Van Berkel SS, et al. Chemoenzymatic conjugation of toxic payloads to the globally conserved N-glycan of native mabs provides homogeneous and highly efficacious antibody-drug conjugates. Bioconjug Chem. 2015;26:2233–42.
8. Junutula JR, Raab H, Clark S, Bhakta S, Leipold DD, Weir S, et al. Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. Nat Biotechnol. 2008;26:925–32.
9. Panowski S, Bhakta S, Raab H, Polakis P, Junutula JR. Site-specific antibody drug conjugates for cancer therapy. MAbs. 2014;6:34–45.
10. Strop P, Delaria K, Foletti D, Witt JM, Hasa-Moreno A, Poulsen K, et al. Site-specific conjugation improves therapeutic index of antibody drug conjugates with high drug loading. Nat Biotechnol. 2015;33:694–6.
11. Axup JY, Bajjuri KM, Ritland M, Hutchins BM, Kim CH, Kazane SA, et al. Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. Proc Natl Acad Sci USA. 2012;109:16101–6.
12. Tsuchikama K, An Z. Antibody-drug conjugates: recent advances in conjugation and linker chemistries. Protein Cell. 2018;9:33–46.
13. Donaghy H. Effects of antibody, drug and linker on the preclinical and clinical toxicities of antibody-drug conjugates. MAbs. 2016;8:659–71.
14. Anami Y, Yamazaki CM, Xiong W, Gui X, Zhang N, An Z, et al. Glutamic acid-valine-citrulline linkers ensure stability and efficacy of antibody-drug conjugates in mice. Nat Commun. 2018;9:2512.
15. Devay RM, Delaria K, Zhu G, Holz C, Foletti D, Sutton J, et al. Improved Lysosomal Trafficking Can Modulate the Potency of Antibody Drug Conjugates. Bioconjug Chem. 2017;28:1102–14.
16. De Goeij BECG, Vink T, Ten Napel H, Breij ECW, Satijn D, Wubbolts R, et al. Efficient payload delivery by a bispecific antibody-drug conjugate targeting HER2 and CD63. Mol Cancer Ther. 2016;15:2688–97.
17. Tolcher AW. Antibody drug conjugates: lessons from 20 years of clinical experience. Ann Oncol. 2016;27:2168–72.
18. Saleh MN, Sugarman S, Murray J, Ostroff JB, Healey D, Jones D, et al. Phase I trial of the anti-Lewis Y drug immunoconjugate BR96-doxorubicin in patients with Lewis Y-expressing epithelial tumors. J Clin Oncol. 2000;18:2282–92.
19. Tijink BM, Buter J, De Bree R, Giaccone G, Lang MS, Staab A, et al. A phase I dose escalation study with anti-CD44v6 bivatuzumab mertansine in patients with incurable squamous cell carcinoma of the head and neck or esophagus. Clin Cancer Res. 2006;12:6064–72.
20. Damelin M, Zhong W, Myers J, Sapra P. Evolving Strategies for Target Selection for Antibody-Drug Conjugates. Pharm Res. 2015;32:3494–507.
21. Perna F, Berman SH, Soni RK, Mansilla-Soto J, Eyquem J, Hamieh M, et al. Integrating Proteomics and Transcriptomics for Systematic Combinatorial Chimeric Antigen Receptor Therapy of AML. Cancer Cell. 2017;32:506–19.
22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
23. Hu Y, Yan C, Hsu CH, Chen QR, Niu K, Komatsoulis GA, et al. Omiccircos: A simple-to-use R package for the circular visualization of multidimensional Omics data. Cancer Inf. 2014;13:13.
24. Fehrmann RSN, Karjalainen JM, Krajewska M, Westra HJ, Maloney D, Simeonov A, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. Nat Genet. 2015;47:115–25.
25. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. BMC Bioinforma. 2013;14:7.
26. Arora S, Pattwell SS, Holland EC, Bolouri H. Variability in estimated gene expression among commonly used RNA-seq pipelines. Sci Rep. 2020;10:2734.
27. Wang Q, Armenia J, Zhang C, Penson AV, Reznik E, Zhang L, et al. Data Descriptor: Unifying cancer and normal RNA sequencing data from different sources. Sci Data. 2018;5:180061.
28. Bagger FO, Kinalis S, Rapin N. BloodSpot: A database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. Nucleic Acids Res. 2019;47:881–5.
29. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. Science. 2017;357:eaan2507.
30. Sweeney SM, Cerami E, Baras A, Pugh TJ, Schultz N, Stricker T, et al. AACR project genie: Powering precision medicine through an international consortium. Cancer Discov. 2017;7:818–31.
31. Junttila TT, Li G, Parsons K, Phillips GL, Sliwkowski MX. Trastuzumab-DM1 (T-DM1) retains all the mechanisms of action of trastuzumab and efficiently inhibits growth of lapatinib insensitive breast cancer. Breast Cancer Res Treat. 2011;128:347–56.
32. Smit EF, Nakagawa K, Nagasaka M, Felip E, Goto Y, Li BT, et al. Trastuzumab deruxtecan (T-DXd; DS-8201) in patients with HER2-mutated metastatic non-small cell lung cancer (NSCLC): Interim results of DESTINY-Lung01. J Clin Oncol. 2020;38:9504.
33. Staudacher AH, Brown MP. Antibody drug conjugates and bystander killing: is antigen-dependent internalisation required. Br J Cancer. 2017;117:1736–42.
34. Perrino E, Steiner M, Krall N, Bernardes GJL, Pretto F, Casi G, et al. Curative properties of noninternalizing antibody-drug conjugates based on maytansinoids. Cancer Res. 2014;74:2569–78.
35. Gébleux R, Stringhini M, Casanova R, Soltermann A, Neri D. Non-internalizing antibody–drug conjugates display potent anti-cancer activity upon proteolytic release of monomethyl auristatin E in the subendothelial extracellular matrix. Int J Cancer. 2017;140:1670–9.
36. Giansanti F, Capone E, Ponziani S, Piccolo E, Gentile R, Lamolinara A, et al. Secreted Gal-3BP is a novel promising target for non-internalizing Antibody–Drug Conjugates. J Control Release. 2019;294:176–84.

AUTHOR CONTRIBUTIONS
JF: conceptualization, investigation, methodology, data curation, data visualization, writing—original draft, writing—review and editing; LG: supervision, project administration, funding acquisition; YZ: investigation, methodology, software; QG: data curation, software; MW, XW: supervision, methodology, project administration.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41417-023-00701-3.

Correspondence and requests for materials should be addressed to Lei Guo, Ming Wang or Xiaoxiao Wang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.