

1 **An Array of 60,000 Antibodies for Proteome-Scale Antibody Generation and**  
2 **Target Discovery**

3

4 **Abstract**

5 Antibodies are essential for elucidating the roles of genes decoded by genome  
6 sequencing. However, affordable technology for proteome-scale antibody  
7 generation does not exist. To address this, we developed the Proteome Epitope  
8 Tag Antibody Library (PETAL) and its array. PETAL consists of 62,208 mAbs  
9 against 15,199 peptides from diverse proteomes. PETAL harbors binders for a  
10 great multitude of proteins in nature due to antibody multispecificity, an intrinsic  
11 feature of an antibody. Distinctive combinations of 10,000-20,000 mAbs were  
12 found to target specific proteomes by array screening. Phenotype-specific mAb-  
13 target pairs were discovered for maize and zebrafish samples.  
14 Immunofluorescence and flow cytometry mAbs for human membrane proteins  
15 and ChIP-seq mAbs for transcription factors were identified from respective  
16 proteome-binding PETAL mAbs. Differential screening of cell surface proteomes  
17 of tumor and normal tissues discovered internalizing tumor antigens for antibody-  
18 drug conjugates. By discovering high affinity mAbs at a fraction of current time  
19 and cost, PETAL enables proteome-scale antibody generation and target  
20 discovery.

21

22

23

1 **Keywords:**

2 antibody; proteome-scale; multispecificity; antibody microarray; target discovery;  
3 antibody-drug conjugate (ADC)

## 1 INTRODUCTION

2 Facilitated by the ever-growing capability of current DNA sequencing  
3 technologies, more than 1,300 genomes of animals and 496 genomes of plants  
4 and many others have already been sequenced, representing millions of genes,  
5 and the number will increase faster from projects such as G10K, i5k and so on<sup>1</sup>.  
6 To understand the roles of these genes, the functions of the gene-coded proteins  
7 need to be explored, and antibodies, especially renewable monoclonal antibodies  
8 (mAbs) generated at a proteome-scale, are urgently needed. mAbs produced by  
9 hybridoma technologies for human proteins have long been recognized as the  
10 most direct tools for diagnostic and therapeutic target discovery<sup>2</sup>. Classic  
11 therapeutic targets Sialyl Lewis Y, prostate-specific membrane antigen (PSMA)  
12 and, more recently, a novel target for multiple myeloma were discovered by  
13 mAbs for cell surface proteins<sup>3-6</sup>.

14 Despite the power of mAbs and mAb-based discovery, large-scale  
15 generation of mAbs remains difficult since traditional hybridoma development is  
16 time-consuming (4-6 months starting from antigens), expensive (\$3,000-8,000  
17 per antigen), and challenging to scale. Further, mAb generation by immunization  
18 typically requires a milligram of purified antigens, a significant challenge for many  
19 protein targets, especially for membrane proteins of primary research interest.  
20 Even for human proteins, a majority of the 6,000 membrane proteins have not  
21 been directly explored as diagnostic or therapeutic targets due to a lack of high-  
22 quality antibodies for applications such as flow cytometry (FACS) and  
23 immunofluorescence (IF)<sup>7, 8</sup>.

1 The Human Protein Atlas (HPA) offers an alternative approach for proteome-  
2 scale antibody development. HPA has generated over 25,000 affinity-purified  
3 polyclonal antibodies against >17,000 human proteins covering more than 80%  
4 of the human proteome<sup>9, 10</sup>. However, it is impractical to replicate the success of  
5 HPA on a significant number of other species with a need for proteome-scale  
6 antibodies due to the great human and capital resources required for such a  
7 project. Furthermore, polyclonal HPA antibodies are not renewable, making  
8 reproduction of these antibodies with consistent quality difficult. Thus, proteome-  
9 scale antibody generation has remained elusive for most sequenced genomes.

10 Over the past few decades, numerous attempts have been made to address  
11 high cost and poor scalability of large-scale antibody generation by improving  
12 traditional hybridoma methods and developing better *in vitro* recombinant  
13 antibody libraries and more efficient screening technologies<sup>11-14</sup>. For *in vitro*  
14 methods, continuing development of novel display technologies<sup>15-17</sup> and  
15 improvements in library design and screening methods<sup>18</sup> have been attempted  
16 dating back twenty years. However, established synthetic antibody libraries for  
17 therapeutic antibody discovery aren't yet used for large-scale reagent generation  
18 due to the concern of high cost. Despite an attempt to generate research  
19 antibodies using phage display libraries<sup>19</sup>, it is not economical to use these  
20 resources for generating antibodies for nonclinical uses or for nonhuman  
21 proteomes.

22 Antibody microarray is a powerful platform for high-throughput, multiplexed  
23 protein profiling using a collection of immobilized antibodies<sup>20, 21</sup>. By using

1 antibody array, one can achieve low cost and fast antibody discovery by direct  
2 array screening. In one approach, a library of ~10,000 in silico-designed and -  
3 synthesized antibody fragments was used to build an antibody array for *de novo*  
4 antibody discovery<sup>22</sup>. The arrayed library was able to generate antibody leads  
5 with  $\mu$ M binding affinity for therapeutic protein targets, suggesting a spatially  
6 addressed library comprising tens of thousands of individual antibodies should be  
7 sufficient for antibody discovery. However, this synthetic antibody library  
8 screening approach did not achieve broader impact since required additional  
9 antibody affinity maturation and engineering limited its usefulness for routine  
10 research affinity reagent development and target profiling.

11 Here, we present a system integrating “industrial-scale” hybridoma  
12 development, antibody microarray, and affinity proteomics to overcome previous  
13 challenges for proteome-scale antibody development and target discovery. Our  
14 technology, called the Proteome Epitope Tag Antibody Library (PETAL), takes  
15 advantage of antibody multispecificity<sup>23-25</sup>, an intrinsic property of antibody  
16 molecules to bind to a large number of proteins unrelated to the original antigen  
17 that the antibody was raised against with high affinity and specificity. Antibody  
18 multispecificity is exemplified by an anti-p24 (HIV-1) peptide monoclonal antibody.  
19 Its epitope sequences, consisting of key interacting residues deduced from five  
20 unrelated peptides binding to CB4-1, were identified in hundreds of heterologous  
21 proteins, and those proteins that could be obtained were shown to bind CB4-1<sup>26</sup>.  
22 PETAL is a mouse monoclonal antibody library consisting of 62,208 mAbs made  
23 against 15,199 peptide antigens representative of 3,694 proteins from 418

1 proteomes. PETAL has the potential to bind to a large number of proteins in  
2 nature due to antibody multispecificity. An antibody microarray was fabricated  
3 using PETAL mAbs. Using cell lysates of diverse proteomes to screen the  
4 PETAL array, we demonstrated, for the first time, the feasibility for proteome-  
5 scale protein targeting by mAbs. Identified antibodies are capable of broad  
6 applications, as shown for human membrane and nuclear protein targets.  
7 Phenotype-specific mAb-target pairs were identified for maize and zebrafish from  
8 respective proteome-specific mAbs. Therapeutic target candidate discovery was  
9 demonstrated by differential screening of normal versus tumor membrane  
10 proteomes. An antibody targeting ProteinX was identified as a candidate for  
11 building an antibody-drug conjugate (ADC) for Lung Squamous Cell Carcinoma  
12 (LUSCC) *in vitro* and *in vivo*. By generating high affinity mAbs at a small fraction  
13 of current time and cost, PETAL enables affinity reagent generation and target  
14 discovery for proteomes with available genomic sequence information.

15

1 **RESULTS**

2

3 The immunological basis of PETAL is antibody multispecificity (**Fig. 1a**). PETAL  
4 is an antipeptide antibody library of 62,208 monoclonal antibodies designed to  
5 have the potential for harboring binders for a great number of proteins from  
6 diverse proteomes (**Fig. 1b**). When PETAL is immobilized in an array format (**Fig.**  
7 **1c**), it enables proteome-scale antibody generation and differential target  
8 discovery (**Fig. 1d**).

9

10 **Design and construction of PETAL**

11 A total of 15,199 peptide antigens, called Proteome Epitope Tags or PETs, were  
12 designed from 3,694 proteins representing 418 proteomes (**Fig. 1b**;  
13 **Supplementary Fig. 1a**). Within each proteome, PETs were selected from  
14 unique regions of protein sequence using the heuristic blastp algorithms optimal  
15 for short peptide sequence comparison<sup>27, 28</sup>. Sequence analysis showed PET  
16 sequences were random and diverse (**Supplementary Fig. 1b**).

17 To construct PETAL, PET antigens were synthesized and used to generate  
18 mouse mAbs by a large-scale monoclonal antibody development operation  
19 modeled after an assembly line (**Fig. 1b**). More than 30,000 mice were  
20 immunized at an average of 2 mice per PET. A total of 62,208 mouse monoclonal  
21 antibodies were generated. Each hybridoma cell line was used to prepare ascites

1 containing 1-20 mg of mouse IgGs with varying concentrations from 0.1 to 10  
2 mg/mL; most were in the 1-3 mg/mL range.

3

4 **PETAL diversity**

5 To evaluate the diversity of PETAL, particularly multiple mAbs generated using  
6 the same PET, antibody V-region was sequenced for 91 randomly selected  
7 hybridomas, including 68 hybridoma clones against 24 PETs with 2-5 mAbs/PET  
8 and 23 hybridomas against 23 unique PETs (**Supplementary Fig. 1c-e**).  
9 Sequences with  $\geq 2$  amino acid differences in the CDR regions were considered  
10 “unique”, although two antibodies with a single CDR amino acid difference were  
11 found to bind to different epitopes<sup>29</sup>. Close to 90% of sequences were unique  
12 (**Supplementary Fig. 1d**). Multiple mAbs generated against the same PET  
13 peptide antigen were mostly (80-100%) unique (**Supplementary Fig. 1e**),  
14 indicating the effective library size was close the total number of PETAL mAbs  
15 since framework differences could also contribute to different binding affinity and  
16 therefore specificity.

17

18 **Construction of PETAL microarray with 62,208 mAbs**

19 To use PETAL for multiplexed screening, an antibody microarray was  
20 constructed with all 62,208 antibodies (**Fig. 1c**). Array printing quality was  
21 assessed using an anti-mouse secondary antibody conjugated with Cy5 (**Fig. 1c**,

1 left two panels). Antibodies of 10-1000 pg per spot (mostly 100-300 pg) were  
2 detected with a fluorescent intensity ranging from 500 to >60,000  
3 (**Supplementary Fig. 1f**).

4 Typically, biotinylated antigens such as 1-10 mg of proteome samples (**Fig.**  
5 **1c**, right two panels) were used for array screening. Array screening consistency  
6 was investigated using a human plasma sample in three independent replicates.  
7 An average of 20,000 antibodies showed positive binding ((signal intensity-  
8 background)/background > 3) with a coefficient of variation (CV) of 7%. More  
9 than 90% of array-positive antibodies had a fluorescent intensity of 500-10,000.  
10 The Pearson's correlation coefficient (r) among triplicate experiments was 0.98  
11 (**Supplementary Fig. 1g**). These data establish the PETAL array as a  
12 reproducible platform for screening.

13 To test the library/array for generating mAbs, a total of 81 recombinant  
14 proteins were used to screen the PETAL array (**Supplementary Fig. 2, Table**  
15 **S3**). Approximately half (47%, 38/81) of the proteins produced an average of 3.7  
16 (141/38) mAbs/antigen after ELISA screening of array binding mAbs using a  
17 detection limit of  $\leq 1\mu\text{g/mL}$  of antigen. When tested in immunoblotting assays, 31%  
18 (25/81) and 26% (21/81) of targets were successful on recombinant proteins and  
19 endogenous samples, respectively (**Supplementary Fig. 2, Table S3**).

20

21 **PETAL targets diverse proteomes for antibody and target discovery**

1 To apply PETAL for targeting broad and diverse proteomes, 11 proteome  
2 samples of plants, animals and bacteria were used to probe PETAL arrays (**Fig.**  
3 **2**). The number of array-positive antibodies for each proteome was 10,000-  
4 20,000. A selection of ~1,000/proteome of array-positive antibodies across the  
5 fluorescent intensity scale was used to probe endogenous samples to estimate  
6 the number of antibodies suitable for immunoblotting (**Fig. 2a**). Typically, 20-30%  
7 of antibodies produced specific (single or predominant single band)  
8 immunoblotting results, giving a proteome of 2,000-6,000 potential  
9 immunoblotting mAbs. Although proteome-binding mAbs overlapped significantly  
10 among proteomes (30-60%), the same antibodies often specifically recognized  
11 targets of different size in different proteomes (**Fig. 2b**), likely to be unrelated  
12 proteins.

13 Proteome-specific mAbs were used to identify phenotype-specific mAbs and  
14 corresponding protein targets for maize seed development and zebrafish  
15 regeneration. A panel of 335 immunoblotting-positive mAbs were produced from  
16 1000 array-positive binding mAbs using a protein sample consisting of maize  
17 seeds at two developmental stages, day after pollination 3 (DAP3) and DAP17  
18 (**Fig. 2c**)<sup>30</sup>. These included 206 mAbs with single or predominant single bands  
19 (Class I, 21%, 206/1000) and an additional 129 mAbs with more than one  
20 (typically 2-3) dominant bands (Class II) on immunoblotting (**Fig. 2c, Supporting**  
21 **data for Fig. 2**). After probing DAP3 and DAP17 protein samples separately  
22 using Class I mAbs, a panel of 70 differentially expressed mAbs were selected  
23 for target identification by immunoprecipitation (IP) and liquid chromatography-

1 tandem mass spectrometry (LC-MS/MS). Nineteen protein targets paired with 23  
2 mAbs were identified, and six examples are shown in **Fig. 2c**, including Sdh, Bt2  
3 and Sbe1, which have previously been shown to be involved in maize seed  
4 development<sup>30, 31</sup>.

5 In another example, a panel of 45 binding mAbs identified from the screening  
6 of protein lysates of zebrafish heart was further characterized by IF assays using  
7 samples of heart before injury and 7 days post injury (**Fig. 2d**). Six of them  
8 showed up- or downregulation in injured heart. Protein targets of these six  
9 antibodies were identified by IP and LC-MS/MS (**Fig. 2d**, right panel), including  
10 proteins with known heart injury protection function (for example, aldh2.2 in **Fig.**  
11 **2d**, left and middle panels)<sup>32</sup>.

12

13 **Proteome-scale antibody generation for human membrane and nuclear**  
14 **proteins**

15 To apply PETAL for antibody generation for organelle proteomes, PETAL arrays  
16 were screened with total protein extracts of membrane and nucleus from human  
17 PC9, HepG2, THP-1 and Jurkat cell lines (**Fig. 3a**). The total number of positive  
18 mAbs for each sample was in the range of 12,000-18,000.

19 To further screen for application-specific antibodies and to identify their  
20 cellular targets, array positive antibodies with high (>10,000), medium (2,000-  
21 10,000) and low (500-2,000) fluorescent intensity (**Supplementary Fig. 3a**) were  
22 selected. A total of 1,439 positive antibodies for membrane proteomes and 379

1 for nuclear proteomes were subjected to immunoblotting, IF/FACS and IP assays  
2 (**Fig. 3b and c, Supplementary Fig. 3b, Movies S1-S6, Supporting data for**  
3 **Fig. 3**) and their cellular targets were identified by LC-MS/MS (**Fig. 3b**). A total of  
4 149 antibodies representing 107 proteins were identified from membrane  
5 proteome screening (**Tables S4 and S5**), including known CD and RAB (small  
6 GTPase) molecules CD3e, CD49d, CD71 (TFRC), CD222, CD5, CD2,  
7 PROTEINX, RAB1B and RAB14. For nuclear proteomes, a total of 55 antibodies  
8 representing 42 proteins were identified (**Tables S4 and S5**), including  
9 transcriptional regulators NONO, NFIC, TRIM28, CSNK2A1, MTA2, SATB1,  
10 SFPQ, and SMARCC1. Approximately 20% of the targets had at least two  
11 independent antibodies that yielded similar IP and immunoblotting results,  
12 strengthening the antibody validation quality<sup>33, 34</sup> (**Fig. 3d**). The success rate of  
13 target identification was consistent over a wide fluorescent intensity range  
14 (**Supplementary Fig. 3a**), suggesting that more than 1,000 proteins could be  
15 covered with 12,000-18,000 proteome-binding antibodies.

16 Only 20-30% of the array-positive antibodies were successful in  
17 immunoblotting assays likely due to the native protein states in the screening  
18 samples. For antibodies without successful immunoblotting data (size unknown),  
19 overexpression or knock-down experiments were necessary to determine their  
20 cellular targets after immunoprecipitation (IP) and mass-spectrum (MS). For  
21 example, recognition of an antibody (Pb2795) to a multi-pass ion channel,  
22 PIEZO1<sup>35</sup>, was confirmed by the colocalization of IF staining signals with  
23 overexpressed PIEZO1-GFP (**Supplementary Fig. 3c**).

1 To investigate if proteins targeted by antibodies were biased toward specific  
2 types, protein abundance and function class (GO annotations) were examined for  
3 87 Jurkat proteins (67 membrane and 20 nuclear) (**Fig. 3e, Supplementary Fig.**  
4 **3d**). Abundance distribution of these proteins was similar to that of total  
5 membrane or nuclear proteins in Jurkat cells according to the PaxDb database<sup>36</sup>.  
6 Proteins identified by antibodies were from diverse protein families and functional  
7 groups similar to total human membrane and nuclear proteins (**Fig. 3e,**  
8 **Supplementary Fig. 3d**) according to the HPA database, suggesting that  
9 antibodies targeted broad classes of proteins without obvious bias.

10 To test antibodies for additional applications, we performed protein complex  
11 enrichment and chromatin immunoprecipitation-sequencing (ChIP-seq) with  
12 selected antibodies. Protein complex enrichment results are shown in **Fig. 3f**. For  
13 example, Pb51585 against PCCA could pull down PCCA-interacting proteins  
14 similar to its previously mapped interactome according to STRING<sup>37</sup> analysis.

15 ChIP-seq assays in HepG2 for antibodies against SMRC1, SATB1 and NFIC  
16 were carried out following previous studies<sup>38</sup>. A total of 93.9 million sequencing  
17 reads were generated, and 53.7% were uniquely mapped to the human reference  
18 genome. These reads were further processed and yielded 46,380 peaks,  
19 representing 29,441, 3,296, and 13,643 binding sites for SMRC1, SATB1 and  
20 NFIC, respectively (**Fig. 3g**). Further validation by comparing to commercial ChIP  
21 antibody, or with ChIP-qPCR assays, analyses of binding sites and enriched  
22 motifs all confirmed that antibodies against those transcription factors could be  
23 applied for ChIP-seq (**Supplementary Fig. 3e-h**).

1

2 **ADC Antibody/Target Discovery by Differential Array Screening**

3 ADC selectively eliminates cancer cells by linking a toxin, for example  
4 Monomethyl auristatin E (MMAE), with an antibody targeting an internalizing  
5 tumor-associated antigen characterized by higher expression in tumors than in  
6 normal tissues<sup>39</sup>. To identify candidate targets suitable for ADC, we screened  
7 PETAL arrays with normal and tumor cell proteomes (**Fig. 4**). More than 3,000  
8 antibodies out of 15,000 lung membrane proteome-positive antibodies showed a  
9 fluorescent intensity fold change of >1.5 between tumor and normal samples (**Fig.**  
10 **4a**). Four antibodies were screened to show both internalization and indirect  
11 cytotoxicity from a selection of 500 antibodies with a fold-change ranging from  
12 1.5 to 5 (**Fig. 4b; Supplementary Fig. 4a**). One antibody, Pb44707, with a signal  
13 2.4-fold higher in tumor tissue, was internalized with a half internalization time of  
14 2.5 hours and IC<sub>50</sub> of ≤100 pM for cell cytotoxicity in PC9 cells. IP and LC-MS/MS  
15 identified ProteinX, a putative cancer stem cell marker, as the most likely target  
16 protein (data not shown). Target protein was further verified by siRNA blocking  
17 experiments in which ProteinX-targeting siRNA caused the loss of surface  
18 fluorescent signal in FACS (**Fig. 4d and e**). The cellular binding affinity (EC<sub>50</sub>) of  
19 Pb44707 with PC9 was 832 pM (**Supplementary Fig. 4b**) as determined by  
20 antibody titration using FACS<sup>40</sup>.

21 To evaluate the expression of ProteinX in tumor and normal tissues,  
22 Pb44707 was used in immunohistochemistry (IHC) to probe tissue arrays.

1 ProteinX was markedly overexpressed in 60% of NSCLC patients, and more  
2 strikingly, in close to 90% of Squamous Cell Carcinoma (LUSCC) (**Fig. 4d**,  
3 **Supplementary Fig. 4c-d**). Furthermore, mRNA of ProteinX showed higher  
4 expression in all stages of LUSCC (**Supplementary Fig. 4e**). Most normal  
5 tissues were negatively stained with Pb44707 except in skin (**Fig. 4e, and f**),  
6 which showed both high and low expression in the population (**Supplementary**  
7 **Fig. 4f**). To build Pb44707 into an ADC molecule, Pb44707 was conjugated to  
8 vc-MMAE with an average drug-antibody-ratio (DAR) of 4.23. Pb44707-MMAE  
9 (AMT707) demonstrated *in vitro* cytotoxicity in lung cancer cell lines, and its  
10 potency was correlated with the ProteinX expression level in these cells (**Fig. 4g**  
11 **and h**). The *in vivo* efficacy of AMT707 was tested in both the cell line-derived  
12 xenograft (CDX) model and patient-derived xenograft (PDX) model. For the  
13 xenograft model with the gefitinib-resistant NSCLC cell line NCI-H1975 (**Figure**  
14 **4i**), AMT707 administered at 3 and 10 mg/kg suppressed tumor growth  
15 completely until days 47 and 57, respectively, whereas for treatment controls, i.e.,  
16 PBS, gefitinib and control ADC, there was no or a much lower effect. In a  
17 ProteinX-positive LUSCC PDX model (**Fig. 4j and k**), AMT707 halted tumor  
18 growth in a dose-dependent manner, and 10 mg/kg AMT707 suppressed tumor  
19 growth completely. Thus, by array screening, we identified multiple endocytic cell  
20 surface targets and at least one antibody and its target with a potential to build a  
21 lead ADC molecule.

22

## 1 DISCUSSION

2 To enable proteome-scale monoclonal antibody development, we constructed  
3 PETAL, a hybridoma library consisting of 62,208 mAbs and its corresponding  
4 antibody array, the largest antibody microarray reported so far. Taking advantage  
5 of antibody multispecificity, PETAL may harbor binders for a large number of  
6 proteins in nature<sup>23-26</sup>. Combining PETAL with the global screening capability of  
7 antibody microarrays, this platform enables mAb generation at a fraction of the  
8 current time and cost. High affinity and desirable specificity of selected mAbs was  
9 ensured by a fit-for-purpose workflow. We have demonstrated initial application  
10 of PETAL in large-scale antibody generation, affinity proteomics and therapeutic  
11 target discovery.

12 With its capacity to target a large number of proteins in a proteome (**Fig. 2**),  
13 our technology provides a solution for proteome/sub-proteome-scale generation  
14 of antibodies<sup>41</sup>. We demonstrated that antibodies targeting cell surface and  
15 nuclear proteins of cancer and immune cells were efficiently identified. Many of  
16 these antibodies were well-suited for IF and IP/ChIP. From an input of 1,818  
17 antibodies, ~200 antibodies capable of immunoblotting/IF/FACS/IP targeting 149  
18 independent membrane and nuclear proteins were identified. Given the  
19 input/target ratio of 12:1 (1,818/149), 10,000-20,000 array-positive antibodies  
20 would yield more than 1,000 targets.

21 Antibody-based functional proteomics has not been available for nonhuman  
22 species. Here, we demonstrated that broad proteome samples including plants,

1 animals and insects tested so far have all identified binding antibodies  
2 corresponding to a large number of protein targets for each proteome. PETAL  
3 immediately provided immunoblotting mAbs for initial characterization of proteins  
4 in organisms with genomic sequencing information by proteome sample  
5 screening and target identification described here. Thus, identified proteome-  
6 specific mAb-protein pairs were further used to probe samples from distinctive  
7 phenotypes to discover phenotype-specific mAb-target pairs. As demonstrated  
8 here, both new and previously reported phenotype-specific targets were  
9 discovered for zebrafish heart regeneration and maize seed development. For  
10 each protein target, mAbs capable of immunoblotting and IP were obtained. The  
11 overall success rate of antibody-target deconvolution was similar to that of  
12 human membrane/nuclear screening. Thus, our technology provides a  
13 straightforward and productive path to carry out affinity proteomics for numerous  
14 proteomes currently without available affinity reagents.

15 When used for differential profiling of cell membrane proteomes, PETAL  
16 provides a “fit-for-purpose” approach for therapeutic target discovery. PETAL  
17 relishes the full potential of antibody-based target discovery compared to that of  
18 other functional genomics (RNAi or CRIPSR) or MS-based approaches because  
19 it delivered the target and lead antibodies at the same time. As demonstrated in  
20 this study, differential array screening comparing tumor (NSCLC) with normal  
21 lung tissue membrane proteomes identified ProteinX-targeting antibody for  
22 building ADC molecule AMT707. More ADC candidate antibodies/targets are

1 expected to be discovered when all 3,000 “tumor high” antibodies are subjected  
2 to the screening process.

3 There are some limitations of the PETAL strategy. The construction of  
4 PETAL was a significant investment of time and effort that is difficult to repeat.  
5 However, PETAL is now a premade resource that can be readily accessed by  
6 researchers to generate mAbs for specific antigens or proteomes or to  
7 differentially screen protein samples to identify phenotype-specific protein targets  
8 (**Supplementary Fig. 5**). Even though PETAL was made as an antipeptide  
9 library, it is likely that other types of antigens could produce equally useful  
10 antibody libraries<sup>6</sup>. It is our hope that this work will not only serve as an  
11 immediate resource but also stimulate new hybridoma libraries for even more  
12 applications. For this purpose, our established industrial-scale monoclonal  
13 antibody development capability can be taken advantage of to efficiently build  
14 hybridoma libraries/arrays on the scale of tens of thousands of mAbs. Current  
15 PETAL only yielded high affinity mAbs for 20-30% of antigens, and this could be  
16 improved by increasing the size of the library, which in turn may also increase the  
17 success rate of application screening and target identification of PETAL mAbs.

18 Taken together, PETAL represents a significant improvement over previous  
19 antibody array and library approaches. A workflow for array screening and  
20 antibody-target deconvolution has been established to eliminate high cost and  
21 the long development time of previous methods. We expect PETAL to accelerate  
22 functional proteomics by enabling proteome-scale antibody generation and target  
23 profiling. We believe PETAL will stimulate the preparation of other antibody

1    libraries; thus, the strategy could be adopted and explored by many other  
2    researchers.

3

4

1 **Figure legends**

2 **Figure 1. Construction and application of PETAL and its array for**  
3 **antibody/target discovery.**

4 (a) Antibody multispecificity. An antibody binds to an epitope/mimotope found  
5 within a variety of proteins from different species, leading to high affinity,  
6 specific binding of this antibody to a large number of proteins in nature.

7 (b) PETAL construction. PETAL is a library of 62,208 mouse monoclonal  
8 antibodies (mAbs) derived from immunization of more than 30,000 mice  
9 against 15,199 diverse peptide antigens. PETAL has the potential to  
10 recognize a great number of proteins in nature.

11 (c) PETAL microarray construction. PETAL is printed into an antibody microarray  
12 as a high-throughput platform for antibody/target discovery (left panel). Right  
13 panel shows the design/layout of the array (red, visualized by a Cy5-  
14 conjugated anti-mouse antibody) and an array hybridization result using a  
15 protein sample (positive-binding mAb spots are shown as green).

16 (d) Workflow for proteome-scale antibody generation and target discovery. Two  
17 array screening applications are shown: direct screening to identify proteome-  
18 specific mAbs and subsequent antibody application screening and target  
19 identification or differential screening to discover mAbs and their cellular  
20 targets associated with a specific phenotype. Differential screening can also  
21 be performed with proteomic-specific mAbs (light green arrow) as in the  
22 examples of maize and zebrafish data.

1

2 **Figure 2. PETAL targets diverse proteomes for antibody and target**  
3 **discovery.**

4 (a) Proteome-targeting PETAL mAbs for immunoblotting. Successful rates  
5 (labeled as %) of immunoblotting (producing single/predominant single bands)  
6 were shown for 11 organisms by using a panel of ~1,000 proteome-binding  
7 mAbs for each organism to probe proteome samples. The total number of  
8 binding mAbs for a proteome was in the range of 10,000-20,000. The specific  
9 tissues for immunoblotting were cow (breast, ovary, and liver), cotton (ovule),  
10 peach (leaf or fruit), grape (nuclear fraction of seed), sugarcane (stalk), maize  
11 (seed), *Pseudomonas aeruginosa* (whole cell), silkworm (larva), zebrafish  
12 (embryo, heart or other tissues), axolotl (regenerating limb), chick (cell lines  
13 and tissues).

14 (b) Different targets detected by the same mAb in two different proteomes. Four  
15 example antibodies each recognized a specific band with different size in  
16 different proteomes upon immunoblotting.

17 (c) Identification of maize seed development stage-specific mAb-target pairs.  
18 PETAL array screening using a proteome sample consisting of total protein  
19 extracts from maize seeds DAP3 and DAP17 produced a total of 12,427  
20 binding mAbs. A selection of 1,000 mAbs to probe DAP3 and DAP17 samples  
21 yielded 206 Class I mAbs (single or predominant single band) and additional  
22 129 Class II mAbs (multiple bands) upon immunoblotting (**Supporting data**  
23 **for Fig. 2**). Seventy differentially expressed mAbs between DAP3 and DAP17

1 were used to IP their cellular protein targets for MS analysis, resulting in the  
2 identification of 19 proteins paired with 23 mAbs. Six proteins are shown in  
3 the right panel. Gel pictures from left to right: Class I mAb immunoblotting  
4 examples: silver staining (SS) of IP products and immunoblotting (IB) of input  
5 and IP products for Pb21831.

6 (d) Identification of heart injury-related proteins from zebrafish. From left to right:  
7 IF staining, silver staining of IP products and immunoblotting of input and IP  
8 products for Pb28030, and summary of the identified protein targets.

9

10 **Figure 3. Proteome-scale antibody generation for human membrane and**  
11 **nuclear proteomes.**

12 (a) Protein target identification for human membrane and nuclear proteome-  
13 specific PETAL mAbs.

14 (b) An example (TFRC/CD71) for antibody target identification. From left to right:  
15 silver staining (SS) for IP product, immunoblotting blot (IB) of input and after  
16 IP samples, coverage map of MS-identified peptides, IF and FACS data. The  
17 cell line for IF was selected according to HPA data. Membrane or nuclear  
18 targets were labeled MEM or NUC. Negative controls for FACS included  
19 staining with blank and irrelevant IgG.

20 (c) Examples of IF data for endogenously expressed membrane and nuclear  
21 proteins. ACTN4 and ACTP5B were stained under nonpermeable conditions.

22 Other targets were stained under permeable staining conditions. (Also see  
23 **Movies S1-S6.**)

1 (d) Targets with two independent IP and immunoblotting mAbs. Panel label (SS  
2 and IB) was the same as in (b). Nuclear targets were labeled in blue.  
3 (e) Abundance and function distribution of proteins identified from the Jurkat cell  
4 membrane proteome.  
5 (f) Protein interactome example using Pb51585 against PCCA. Protein-protein  
6 interacting map (right) analyzed by STRING with the mass-identified proteins.  
7 (g) Snapshot of the Integrative Genomics Viewer (IGV) showing sequencing  
8 read density of ChIP-seq data generated with antibodies against SMRC1,  
9 SATB1 and NFIC in HepG2 cells.

10

11 **Figure 4. Differential array screening for ADC therapeutic target/antibody**  
12 **discovery.**

13 (a) Differential antibody screening for ADC targets. Inset MA scatter plot shows  
14 differential distribution of antibody signal intensity of NSCLC and normal lung.  
15 N = 1. More than 3,000 tumor-high antibodies were identified.  
16 (b) An antibody candidate, Pb44707, for ADC. Antibody ID was labeled on the left  
17 of the IF image. IF (0 and 4 hours) image (antibody labeled green and DAPI  
18 labeled blue) time course of normalized surface fluorescence in FACS and  
19 cell cytotoxicity data is shown. Internalization half time ( $t_{1/2}$ ) and mean percent  
20 growth inhibition  $\pm$  SEM (n = 3) of the antibody is labeled. IF scale bar = 50  
21  $\mu$ m.

22 (c) siRNA targeting ProteinX specifically decreases the FACS signal of Pb44707

1 compared to control siRNA. Normalized mean fluorescent intensity (MFI)  $\pm$   
2 SEM of PC9 cells as detected by Pb44707.

3 (d) Representative images of the immunohistochemistry staining of Pb44707 in  
4 lung squamous cell carcinoma (LUSCC) tumor tissue and para-cancerous  
5 tissue from 1 patient. Scale bar = 300  $\mu\text{m}$ .

6 (e) Images of the immunohistochemistry staining of Pb44707 in representative  
7 vital normal human tissues. No expression of ProteinX is detected in these  
8 tissues. Scale bar = 400  $\mu\text{m}$ .

9 (f) Images of the immunohistochemistry staining of Pb44707 in normal human  
10 skin tissues from two healthy persons. Scale bar = 400  $\mu\text{m}$ .

11 (g) Quantification of total binding sites of Pb44707 on the plasma membrane of a  
12 variety of tumor cell lines.

13 (h)  $\text{IC}_{50}$  of Pb44707-ADC on tumor cell lines.  $N = 1-3$  for different cell lines,  
14 respectively.

15 (i) Growth curves of the NCI-H1975 CDX tumors of different treatment groups ( $n$   
16 = 7/group). Treatment with AMT707, control ADC, or gefitinib (intraperitoneal  
17 injection, dosing once a day) was initiated 7 days after tumor inoculation and  
18 administered as indicated by arrows.

19 (j) Representative images of the immunohistochemistry staining of Pb44707 in a  
20 LUSCC patient tumor tissue (top) and passage 1 (P1) PDX tumor tissue  
21 derived from the same patient. Scale bar = 50  $\mu\text{m}$

22 (k) Growth curves of the PDX tumors of different treatment groups ( $n = 6$ /group).  
23 Treatment with AMT707 or control ADC was initiated 35 days after tumor

1 inoculation and administered as indicated by arrows.

2

1    **Supplementary Figure legends**

2

3    **Supplementary Figure 1. PETAL antigens, antibody diversity and array**  
4    **performance.**

5    (a) Proteome, protein and peptide for PETAL construction. Peptide length is 10-  
6    12 amino acids. Artificially designed peptides, such as tag peptides and  
7    random-generated peptides, are labeled Synthetic. The table on the right  
8    showed the number of proteomes, proteins and peptides.

9    (b) Diversity of the PET library. PET sequence identity comparison with two  
10   randomly created peptide libraries. One from human protein sequences was  
11   labeled Random Human Peptides, whereas another from all NCBI sequences  
12   of the same length was labeled Random Peptides. All three peptide sets  
13   showed a similar sequence identity profile by pairwise alignment, suggesting  
14   that the PET library is random and diverse.

15   (c) Homology matrix of the CDR sequences of 91 PETAL mAbs shown as heat  
16   map. Homology scale is shown on the right side of the panel, in which red is  
17   100% homology and green is 0% homology. Overall, more than 90% of the  
18   sequenced mAb sequences were unique.

19   (d) Diversity of PETAL mAbs. The relationship between the number/percentage  
20   of unique PETAL mAb sequences and the number of CDR amino acid  
21   differences is shown.

22   (e) Unique mAbs against the same PET immunogens. Overall, more than 90% of  
23   mAb sequences against the same peptide antigen were unique, implying that

1 these antibodies likely either recognize different amino acid epitopes or bind  
2 to the same epitope with different affinities.

3 (f) Fluorescence profile of printed antibodies with the distribution of foreground  
4 and background fluorescent intensities of printed antibodies.

5 (g) Three independent array-hybridization experiments (marked as #1, #2 and #3)  
6 were performed using a proteome sample (total protein extract from a human  
7 cancer cell line). The signal intensities were then plotted, and the  
8 reproducibility, i.e., the coefficient values (R) were calculated among these 3  
9 experiments (labeled as #1 vs. #2 and #2 vs. #3).

10

11 **Supplementary Figure 2. PETAL screen by protein antigens.**

12 (a) mAb discovery for recombinant proteins by PETAL array. Approximately 30%  
13 of recombinant proteins yielded successful immunoblotting mAbs using  
14 recombinant (31%) or endogenous protein samples (26%).

15 (b) (top panel) Immunoblotting (IB) validation of Pb19964 against *E. coli* protein  
16 lpdA. Purified recombinant proteins of serial dilutions and total lysate of *E. coli*  
17 with lpdA overexpression were immunoblotted with the corresponding  
18 antibodies. The total lysate of an *E. coli* strain with the expression vector only  
19 was included as control. The arrowheads on the right marked the His-tagged  
20 protein and endogenous protein. (middle and bottom panels) Immunoblotting  
21 validation of Pb14410 and Pb3579 against *E. coli* proteins ldcC and fixB,  
22 respectively. Purified recombinant proteins of serial dilutions and total lysate  
23 of *E. coli* from ldcC ( $\Delta$ ldcC), fixB ( $\Delta$ fixB) knockout strains were immunoblotting

1 tested with the corresponding antibodies. The total lysate of WT *E. coli* strain  
2 was included as control. The arrowheads on the right mark the His-tagged  
3 protein and endogenous protein.

4 (c) Immunoblotting of antiprotein PETAL mAbs on additional *E. coli* proteins.  
5 Protein name is labeled on the left side of each panel. The corresponding  
6 PETAL mAbs IDs are labeled on the right.

7 (d) Immunoblotting of antiprotein PETAL mAbs on endogenous cell lysates from  
8 *E. coli*; 20 µg of total lysate was loaded on each lane.

9 (e) Immunoblotting of antiprotein PETAL mAbs on additional human recombinant  
10 proteins

11 (f) Immunoblotting of antiprotein PETAL mAbs on endogenous cell lysates from  
12 human cell lines; 20 µg of total lysate was loaded on each lane. #: NEDD8 is  
13 a ubiquitin-like protein; therefore, the ladder-like pattern and aggregation in  
14 large molecular weight were judged to be a positive result. Red asterisks  
15 marked the predicted position of NEDD8 monomer, dimer, trimer, and  
16 tetramer.

17

18 **Supplementary Figure 3. Proteome-scale antibody generation for human**  
19 **membrane and nuclear proteins.**

20 (a) Distribution of fluorescent intensity of array-positive antibodies from Jurkat cell  
21 membrane proteome screening (blue). Approximately 15,000 antibodies were  
22 positive for binding with a fluorescent signal intensity in the range of 500-  
23 15,000. Red dots represent antibodies used in the IP assay, and green dots

1 represent those successful for application screening  
2 (Immunoblotting/IF/FACS/IP) and target identification, showing similar  
3 success rates across the antibody fluorescence intensity scale.

4 (b) FACS analyses for endogenously expressed membrane and nuclear proteins.

5 Figure labels are the same as in Fig. 3b for TFRC.

6 (c) Overexpression validation of anti-PIEZO1 antibody Pb2795. Silver staining of  
7 Pb2795 IP product (lane IP) is shown on the left side. Lane ctr was loaded  
8 with IP product from an irrelevant antibody. Arrowhead marks the dominant IP  
9 band of PIEZO1. On the right side, HUVECs were transfected with PIEZO1-  
10 GFP (green) expression vector and stained with Pb2795 (red). The result  
11 shows fluorescent overlap of GFP and antibody (yellow in the right panel).  
12 Cells were stained with DAPI (blue).

13 (d) Similar to **Fig. 4e**, data for nuclear protein targets. The abundance and  
14 function distribution of the proteins identified from Jurkat cell nuclear  
15 proteome.

16 (e) Snapshot of IGV browser showing ChIP-seq read density at the REEP3 locus.  
17 The purple peaks were derived from PETAL NFIC antibody, and the orange  
18 ones were generated with a commercial NFIC antibody<sup>42</sup>. Antibody for NFIC  
19 identified known palindromic sequence TTGGCNNNNNGCCAA, whereas the  
20 commercial antibody did not enrich this sequence.

21 (f) ChIP-qPCR validation performed on four randomly chosen peaks from each  
22 ChIP-seq assay of SMRC1, SATB1 or NFIC antibody. The results were  
23 normalized to input chromatin, showing all chosen regions were fairly

1 enriched.

2 (g) Genomic location distribution of binding sites of SATB1, SMARC1 and NFIC  
3 relative to the nearest transcription units. The percentages of binding sites at  
4 the respective locations were shown.

5 (h) Predominant binding motifs of SATB1, SMRC1 and NFIC. These motifs were  
6 identified from the ChIP-seq datasets of SATB1, SMRC1 and NFIC,  
7 respectively. Known NFIC motif (TTGGCNNNNNGCCAA) was recovered  
8 using an antibody for NFIC.

9

10 **Supplementary Figure 4. Differential array screening for ADC therapeutic**  
11 **target/antibody.**

12 (a) Antibody candidates for ADC. Antibody ID was labeled on the left of the IF  
13 image. For each antibody, IF (0 and 4 hours) image (antibodies labeled  
14 green and DAPI labeled blue) time course of normalized surface  
15 fluorescence in FACS and cell cytotoxicity data is shown. Internalization half  
16 time ( $t_{1/2}$ ) and mean percent growth inhibition  $\pm$  SEM ( $n = 3$ ) for each  
17 antibody is labeled. IF scale bar = 50  $\mu$ m.

18 (b) Pb44707 cellular binding affinity. Representative image showing the binding  
19 affinity of Pb44707 with PC9 cells. The EC<sub>50</sub> is 832 pM. N = 3.

20 (c) Examples of IHC of Pb44707 on lung cancer tissue array. Subtypes of  
21 NSCLC were labeled squamous cell carcinoma (SCC), adenocarcinoma  
22 (AC), and large cell carcinoma (LCC). Small cell lung carcinoma was labeled  
23 SCLC. For each tumor type, IHC data from two patients and their papa

1 tumor tissues are shown. ProteinX was found overexpressed in 40% of AC  
2 and LCC, in 90% of LUSCC and not expressed in SCLC.

3 (d) ProteinX expression in LUSCC. Pb44707 was used to probe a tissue array  
4 of LUSCC patients with tumor (labeled as T) and para-cancerous (labeled as  
5 P) samples; 85% of LUSCC patients had overexpression of ProteinX.

6 Intensity of ProteinX overexpression on IHC was graded as high (+++, 36%),  
7 medium (++, 37%) and low (+, 12%). See (g).

8 (e) ProteinX mRNA expression analyzed at different stages of LUSCC using the  
9 TCGA database.

10 (f) Summary of ProteinX expression in NSCLC and skin.

11

12 **Supplementary Figure 5. Two approaches to access PETAL library/array.**

13 (a) PETAL arrays are provided to labs to run their own screenings. Any lab with  
14 access to a microarray facility, especially if equipped with a microarray  
15 scanner, is able to perform the microarray experiment by themselves using  
16 the PETAL microarrays following the procedure that we described. Once the  
17 array screening is performed with antigens of interest, a list of the identified  
18 antibodies, or simply the array hybridization result, could be sent back.  
19 Antibodies can then be provided at any point dependent upon the  
20 researcher's antibody need.

21 (b) Standard screening service is provided. Tissue samples/total protein lysates  
22 or other types of proteome/subproteome samples are provided for  
23 screenings following the process described in this report. As the final

1 outcome, scientists will then be provided with a set of validated monoclonal  
2 antibodies. In both approaches, arrays/services will be provided by Abmart  
3 and third party suppliers through two options: Option 1, scientific  
4 collaboration; Option 2, commercial service.

5

6

## 1    References

2

- 3    1. Koepfli, K.P., Paten, B. & O'Brien, S.J. The Genome 10K Project: a way forward. *Annual*  
4    *review of animal biosciences* **3**, 57-111 (2015).
- 5    2. Carter, P., Smith, L. & Ryan, M. Identification and validation of cell surface antigens for  
6    antibody targeting in oncology. *Endocrine-related cancer* **11**, 659-687 (2004).
- 7    3. Williams, A.F., Galfre, G. & Milstein, C. Analysis of cell surfaces by xenogeneic myeloma-  
8    hybrid antibodies: differentiation antigens of rat lymphocytes. *Cell* **12**, 663-673 (1977).
- 9    4. Hellstrom, I., Garrigues, H.J., Garrigues, U. & Hellstrom, K.E. Highly tumor-reactive,  
10   internalizing, mouse monoclonal antibodies to Le(y)-related cell surface antigens. *Cancer*  
11   *Res* **50**, 2183-2190 (1990).
- 12   5. Israeli, R.S., Powell, C.T., Fair, W.R. & Heston, W.D. Molecular cloning of a complementary  
13   DNA encoding a prostate-specific membrane antigen. *Cancer Res* **53**, 227-230 (1993).
- 14   6. Hosen, N. et al. The activated conformation of integrin beta7 is a novel multiple  
15   myeloma-specific target for CAR T cell therapy. *Nature medicine* **23**, 1436-1443 (2017).
- 16   7. Benicky, J., Hafko, R., Sanchez-Lemus, E., Aguilera, G. & Saavedra, J. Six Commercially  
17   Available Angiotensin II AT1 Receptor Antibodies are Non-specific. *Cell Mol Neurobiol* **32**,  
18   1353-1365 (2012).
- 19   8. Michel, M.C., Wieland, T. & Tsujimoto, G. How reliable are G-protein-coupled receptor  
20   antibodies? *Naunyn-Schmiedeberg's archives of pharmacology* **379**, 385-388 (2009).
- 21   9. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347** (2015).
- 22   10. Thul, P.J. et al. A subcellular map of the human proteome. *Science* **356** (2017).
- 23   11. Jin, A. et al. A rapid and efficient single-cell manipulation method for screening antigen-  
24   specific antibody-secreting cells from human peripheral blood. *Nature medicine* **15**,  
25   1088-1092 (2009).
- 26   12. Love, J.C., Ronan, J.L., Grotengbreg, G.M., van der Veen, A.G. & Ploegh, H.L. A  
27   microengraving method for rapid selection of single cells producing antigen-specific  
28   antibodies. *Nat Biotech* **24**, 703-707 (2006).
- 29   13. Di Cristina, M. et al. An antigen microarray immunoassay for multiplex screening of  
30   mouse monoclonal antibodies. *Nature protocols* **5**, 1932-1944 (2010).
- 31   14. Reddy, S.T. et al. Monoclonal antibodies isolated without screening by analyzing the  
32   variable-gene repertoire of plasma cells. *Nature biotechnology* **28**, 965-969 (2010).
- 33   15. Clackson, T., Hoogenboom, H.R., Griffiths, A.D. & Winter, G. Making antibody fragments  
34   using phage display libraries. *Nature* **352**, 624-628 (1991).
- 35   16. Feldhaus, M.J. et al. Flow-cytometric isolation of human antibodies from a nonimmune  
36   *Saccharomyces cerevisiae* surface display library. *Nature biotechnology* **21**, 163-170  
37   (2003).
- 38   17. Harvey, B.R. et al. Anchored periplasmic expression, a versatile technology for the  
39   isolation of high-affinity antibodies from *Escherichia coli*-expressed libraries. *Proceedings*  
40   *of the National Academy of Sciences of the United States of America* **101**, 9193-9198  
41   (2004).
- 42   18. Schaffitzel, C., Hanes, J., Jermytus, L. & Pluckthun, A. Ribosome display: an in vitro  
43   method for selection and evolution of antibodies from libraries. *Journal of*  
44   *immunological methods* **231**, 119-135 (1999).
- 45   19. Mersmann, M. et al. Towards proteome scale antibody selections using phage display.  
46   *New biotechnology* **27**, 118-128 (2010).

1 20. de Wildt, R.M., Mundy, C.R., Gorick, B.D. & Tomlinson, I.M. Antibody arrays for high-  
2 throughput screening of antibody-antigen interactions. *Nature biotechnology* **18**, 989-  
3 994 (2000).

4 21. Wingren, C. Antibody-Based Proteomics. *Advances in experimental medicine and biology*  
5 **926**, 163-179 (2016).

6 22. Mao, H. et al. Spatially addressed combinatorial protein libraries for recombinant  
7 antibody discovery and optimization. *Nature biotechnology* **28**, 1195-1202 (2010).

8 23. Notkins, A.L. Polyreactivity of antibody molecules. *Trends in immunology* **25**, 174-179  
9 (2004).

10 24. James, L.C., Roversi, P. & Tawfik, D.S. Antibody multispecificity mediated by  
11 conformational diversity. *Science* **299**, 1362-1367 (2003).

12 25. Mariuzza, R.A. Multiple paths to multispecificity. *Immunity* **24**, 359-361 (2006).

13 26. Kramer, A. et al. Molecular basis for the binding promiscuity of an anti-p24 (HIV-1)  
14 monoclonal antibody. *Cell* **91**, 799-809 (1997).

15 27. Altschul, S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database  
16 search programs. *Nucleic acids research* **25**, 3389-3402 (1997).

17 28. Berglund, L., Andrade, J., Odeberg, J. & Uhlén, M. The epitope space of the human  
18 proteome. *Protein Science* **17**, 606-613 (2008).

19 29. Kabat, E.A. & Wu, T.T. Identical V region amino acid sequences and segments of  
20 sequences in antibodies of different specificities. Relative contributions of VH and VL  
21 genes, minigenes, and complementarity-determining regions to binding of antibody-  
22 combining sites. *The Journal of Immunology* **147**, 1709-1719 (1991).

23 30. Jiang, L. et al. Multigene engineering of starch biosynthesis in maize endosperm  
24 increases the total starch content and the proportion of amylose. *Transgenic research*  
25 **22**, 1133-1142 (2013).

26 31. de Sousa, S.M., Paniago Mdel, G., Arruda, P. & Yunes, J.A. Sugar levels modulate sorbitol  
27 dehydrogenase expression in maize. *Plant molecular biology* **68**, 203-213 (2008).

28 32. Budas, G.R., Disatnik, M.H. & Mochly-Rosen, D. Aldehyde dehydrogenase 2 in cardiac  
29 protection: a new therapeutic target? *Trends in cardiovascular medicine* **19**, 158-164  
30 (2009).

31 33. Berglund, L. et al. A gene-centric Human Protein Atlas for expression profiles based on  
32 antibodies. *Molecular & cellular proteomics : MCP* **7**, 2019-2027 (2008).

33 34. Uhlen, M. et al. A proposal for validation of antibodies. *Nature methods* **13**, 823-827  
34 (2016).

35 35. Romac, J.M., Shahid, R.A., Swain, S.M., Vigna, S.R. & Liddle, R.A. Piezo1 is a mechanically  
36 activated ion channel and mediates pressure induced pancreatitis. *Nature  
37 communications* **9**, 1715 (2018).

38 36. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of  
39 eleven common cell lines reveals ubiquitous but varying expression of most proteins.  
40 *Molecular & cellular proteomics : MCP* **11**, M111.014050 (2012).

41 37. Szklarczyk, D. et al. The STRING database in 2017: quality-controlled protein-protein  
42 association networks, made broadly accessible. *Nucleic acids research* **45**, D362-d368  
43 (2017).

44 38. Zhao, X.D. et al. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations  
45 reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**,  
46 286-298 (2007).

47 39. Beck, A., Goetsch, L., Dumontet, C. & Corvaia, N. Strategies and challenges for the next  
48 generation of antibody-drug conjugates. *Nat Rev Drug Discov* **16**, 315-337 (2017).

1 40. Willuda, J. et al. Preclinical Antitumor Efficacy of BAY 1129980-a Novel Auristatin-Based  
2 Anti-C4.4A (LYPD3) Antibody-Drug Conjugate for the Treatment of Non-Small Cell Lung  
3 Cancer. *Molecular cancer therapeutics* **16**, 893-904 (2017).

4 41. Stoevesandt, O. & Taussig, M.J. Affinity proteomics: the role of specific binding reagents  
5 in human proteome analysis. *Expert Rev Proteomics* **9**, 401-414 (2012).

6 42. Gertz, J. et al. Distinct properties of cell-type-specific and shared transcription factor  
7 binding sites. *Molecular cell* **52**, 25-36 (2013).

8

9

1    **References for Supplementary Figure**

2    1.    Gertz, J. et al. Distinct properties of cell-type-specific and shared transcription factor  
3        binding sites. *Molecular cell* **52**, 25-36 (2013).  
4

1 **Methods**

2 **Peptide antigen selection**

3 A total of 15,199 peptide antigens, called Proteome Epitope Tags or PETs, were  
4 designed from 3,694 proteins representing 418 proteomes. Within each proteome,  
5 PETs were selected from unique regions of protein sequence using the heuristic  
6 blastp algorithms optimal for short peptide sequence comparison<sup>1, 2</sup>. Peptide  
7 antigens representative of predicted surface epitopes from a protein sequence  
8 were selected<sup>3, 4</sup>. Peptides were mostly 10-12 amino acids in length to contain 2-  
9 3 potential antibody epitopes<sup>5</sup>. Predicted peptide sequences with secondary  
10 structures including alpha-helix and beta-sheet were omitted<sup>6</sup>. Special sequences  
11 including transmembrane motif, signal peptide, and post-translational  
12 modification motif were also not selected. Only disordered or surface-looped  
13 regions were selected. Hydrophobic peptides were not selected, and peptide  
14 hydrophilicity was calculated by Hopp and Woods method<sup>7</sup>. Finally, peptides with  
15 more than one cysteine in the sequence were omitted to avoid synthesis  
16 difficulties.

17 All the peptide antigens were chemically synthesized by GL Biochem  
18 (Shanghai) Ltd. The purity and molecular weight of each peptide was evaluated  
19 with HPLC and mass spectrum (MS).

20

21 **Diversity analysis of PET library**

1 The diversity of the PET peptide library (15,199 peptides, see **Table S1** for detail)  
2 is evaluated by comparing the sequence similarity of all peptides against each  
3 other. The sequence identity (%) between a peptide and its closest homologs  
4 within the library was recorded. PET sequence similarity to two random peptide  
5 libraries generated computationally was also compared. The first library was a  
6 collection of 15,199 peptide sequences randomly sampled from all species  
7 without considering amino acid preference in different species<sup>8</sup>. The second  
8 library was constructed by random sampling the entire human proteome (all  
9 consensus coding sequences).

10

11 **Construction of PETAL monoclonal antibody library**

12 Monoclonal antibodies were developed using a large-scale monoclonal antibody  
13 development operation modeled after an assembly line. In the antibody assembly  
14 line, each of the close to 100 highly trained technicians performs 1-3 discrete  
15 steps (for example, plating fusion cells onto 96-well plates or cell transfer from  
16 96- to 384-well plates) for making hybridomas. An internally built informatics and  
17 data system (Antibody Assembler) is used for tracking materials and project  
18 status. More than 90% of all materials used are bar-coded to minimize hand  
19 labeling. Many steps have automatic data analysis and decision making (for  
20 example, clone picking). Together, Antibody Assembly Line is scalable and cost-  
21 efficient. PETAL is a premade library built by this highly efficient process.  
22 Following traditional hybridoma protocol<sup>9</sup>, after the immunization and fusion, a

1 series of ELISA screens were performed to ensure only the hybridoma clones  
2 with the highest affinity to peptide antigens were selected. IgG mAbs were  
3 selected using a Sigma antibody isotyping kit (#11493027001). Four to six IgG  
4 hybridomas per peptide antigen were selected for multiple rounds of limited  
5 dilution subcloning to ensure stability and monoclonality. Each hybridoma cell line  
6 was used to prepare milliliters of ascites containing 1-10 mg of mouse IgGs.  
7 Mouse strains used for immunization and ascite production were BALB/c and F1  
8 from Shanghai Super-B&K laboratory animal Corp. Ltd. The procedures for care  
9 and use of animals were approved by the Abmart Institutional Animal Care and  
10 Use Committee.

11

12 **Hybridoma V-region sequencing**

13 A mouse IgG primer set from Novagen (#69831-3) was used to amplify the IgG  
14  $V_H$  and  $V_L$  regions from selected hybridoma clones. In brief,  $1 \times 10^6$  cells were  
15 collected for each cell line. Total RNA was extracted using TRIzol reagent  
16 (Thermo Fisher, #15596026). The first-strand cDNA was amplified using  
17 PrimeScript RT-PCR kit from Takara (#RR104A). PCR products with the  
18 expected size (~400 bp for  $V_H$ , and 360 bp for  $V_L$ ) were sequenced. The  
19 sequences of the PCR products were analyzed by IMGT/V-Quest  
20 (<http://www.imgt.org>)<sup>10, 11</sup> to define the  $V_H$  or  $V_L$  regions and the corresponding  
21 subelements. The uniqueness of antibody sequences was evaluated by  
22 comparing full-length V ( $V_H$  and  $V_L$ ), frame or CDR sequences using clustal

1 algorithm. The homology matrixes were shown in the heat-map format, and that  
2 of the combined CDR sequences was shown in **Supplementary Fig. 1c** as an  
3 example.

4

5 **PETAL array construction and quality evaluation**

6 Ascites of 62,208 PETAL mAbs were prepared in 162 384-well plates and printed  
7 onto nitrocellulose-coated slides (FAST<sup>TM</sup>, Maine Manufacturing, #10486111) in  
8 a high-density microarray format (named as PETAL array) using the Marathon  
9 system (Arrayjet Ltd, UK). Approximately 100 pL of ascites was printed for each  
10 antibody per spot. The array and block layout are shown in **Figure 1c**. A total of  
11 110 blocks were aligned into 10 rows and 11 columns. Each block contains a  
12 subarray of  $48 \times 12 = 576$  individual antibody spots, except the subarrays in the  
13 last row were printed with  $40 \times 12 = 480$  (3 blocks) or  $39 \times 12 = 368$  (8 blocks)  
14 spots. Additional control rows, including a positioning fluorescent spot (Cy3) and  
15 a biotin-BSA gradient (0.4-50 pg) of 8 spots, were also printed for each block  
16 similar to previous antibody arrays<sup>12, 13</sup>. Biotin-BSA was prepared by saturated  
17 labeling of BSA with Thermo Fisher Sulfo-NHS-LC-Biotin (#21336) labeling  
18 reagent. PETAL arrays were stored at -80 °C.

19 To evaluate PETAL array quality, the slides were blotted directly with a  
20 mixture of a Streptavidin-Cy3 (Sigma, #S6402) and a Cy5-labeled goat anti-  
21 mouse IgG (Jackson ImmunoResearch, #115-175-146), both at 1:3000 dilution in  
22 1x PBS. Fluorescence of Cy3 and Cy5 were recorded using 532 nm and 635 nm

1 channels by a GenePix 4200A Microarray Scanner (Molecular Devices, LLC.).  
2 Images were analyzed by GenePix Pro 6.0 software to give fluorescent  
3 intensities of each spot and its corresponding background. Missing or distorted  
4 spots, typically controlled under 5% of the total spots, were automatically marked  
5 by the software.

6 Reproducibility of array experiments was evaluated by incubating a triplicate  
7 of the same sample with three PETAL arrays. The fluorescent intensity of each  
8 spot was normalized using biotin-labeled BSA signal in each block and array.  
9 The normalized fluorescent intensities were plotted between every experimental  
10 pair. Pearson product-moment correlation coefficient value (R value) was  
11 calculated for each data pair in which the r value of +1 means total positive  
12 correlation and 0 is no correlation<sup>14</sup>.

13

#### 14 **Protein sample preparation**

15 Cell lines, i.e., A431, A549, HEK293T (293T), H1975, H226, Hela, HepG2, HL60,  
16 HUVEC, Jurkat, K562, MCF7, PC3, PC9, THP1 and U937, were purchased from  
17 ATCC or Stem Cell Bank, Chinese Academy of Sciences (Shanghai, China).  
18 Cells were grown or maintained in DMEM or RPMI1640 media following the  
19 ATCC cell culture guide. When cells were grown to ~80% confluence, they were  
20 dissociated from culture plates by treatment with 1x PBS and 1 mM EDTA for 10-  
21 20 minutes. Trypsin was not used to avoid damages on cell surface proteins.  
22 Membrane or nuclear fractions of cell lysates were then prepared as described

1 previously<sup>15</sup>. For whole cell lysis, 1x PBS containing 1% NP40, 5 mM EDTA, and  
2 protease inhibitor cocktail (Calbiochem, #539134) was added to cells directly and  
3 incubated on ice for 30 min. An ultrasonication step was performed before  
4 collecting the supernatant. The prepared membrane fraction, nuclear fraction and  
5 whole cell lysate were labeled following the cell lines named as MEM, NUC and  
6 WCL, respectively. The enrichments of marker proteins in MEM, NUC, and WCL  
7 fractions were evaluated with anti-ATP5B (Abmart, #M40013), Histone3.1  
8 (Abmart, #P30266), and beta-Tubulin (Abmart, #M30109) mAbs, respectively.

9 For tissue samples from plants or animals, whole cell lysates were prepared  
10 following a protocol described previously<sup>16</sup>. In brief, frozen tissues were  
11 powdered in liquid nitrogen with a pestle, suspended in 10 mL/3 g of tissue  
12 extract protein extraction buffer (10 mM Tris, pH 8.0, 100 mM EDTA, 50 mM  
13 Borax, 50 mM Vitamin C, 1% Triton X-100, 2% -Mercaptoethanol, 30% sucrose)  
14 and incubated for 10 min. An equal volume of Tris-HCl pH 7.5-saturated phenol  
15 was then added and vortex mixed for 10 min at room temperature. The phenolic  
16 phase separated by centrifugation was recovered and re-extracted twice with 10  
17 mL of extraction buffer. Proteins in the final phenolic phase were precipitated  
18 overnight at -20°C with 5x volumes of saturated ammonium acetate in methanol.  
19 Protein pellets collected by centrifugation were washed twice with ice-cold  
20 methanol and once with ice-cold acetone. Pellets were then dried and dissolved  
21 with 500 mM triethylammonium bicarbonate containing 0.5% SDS, pH 8.5.  
22 Bacteria lysates were prepared using an ultrasonic apparatus.

23

1    **Patient samples**

2    All tissue microarray chips were purchased from Shanghai Outdo Biotech Co.,  
3    Ltd. Tumor and para-cancerous tissues (normal) were freshly excised from a  
4    patient with Non-Small-Cell Lung Carcinoma (**NSCLC**) undergoing surgery.  
5    Tumor tissue and matched para-cancerous tissue were homogenized<sup>17</sup>. Briefly,  
6    the specimens were cut into 0.5 mm sections before digestion with 0.1%  
7    collagenase IV (Gibco, #17104019) for 1 hour at 37°C. The cells were then  
8    passed through a 70 µm cell strainer (BD, #352350) and collected by  
9    centrifugation for 15 min at 400 g. Plasma membrane proteome extracts were  
10   prepared from single cell suspensions of tissues.

11

12   **Screening PETAL array with recombinant protein antigens**

13   Recombinant protein antigens were first labeled with biotin using EZ-Link™ NHS-  
14   LC-Biotin reagent (Thermo Fisher, #21366) and then hybridized with the PETAL  
15   array. Array-bound proteins were incubated with Streptavidin-Cy3. The  
16   fluorescent intensity of mAb spots was then recorded by GenePix 4200A  
17   Microarray Scanner. Array-positive spots were defined as (signal-  
18   background)/background > 3. Protein-binding PETAL mAbs selected from array  
19   experiments were further screened through protein-mAb ELISA using a detection  
20   limit of 1 µg/mL of protein antigen.

21

1 **Screening PETAL array with proteomic antigens**

2 Proteomic antigens including membrane, nuclear, or whole cell lysates were  
3 labeled with biotin using EZ-Link™ NHS-LC-Biotin reagent and incubated with  
4 the PETAL arrays. Following a similar procedure as described above, antibody  
5 spots positive in three independent experiments were then ranked by the  
6 averaged fluorescence intensities. A limited number of array positive antibodies  
7 (1,000-2,000 according to the expected output of each screening) with high  
8 (>10,000), medium (2,000-10,000) and low (500-2,000) fluorescent intensity were  
9 selected as candidate antibodies for further validation assays.

10

11 **Immunoblotting assays of PETAL mAbs**

12 For recombinant protein immunoblotting, selected mAbs were used to probe 50,  
13 10, 2, and 0.4 ng of recombinant protein antigens.

14 For immunoblotting of endogenous human protein samples by PETAL mAbs,  
15 cell lines were selected according to the protein expression profile from HPA and  
16 Uniprot databases. For membrane or nuclear protein targets, corresponding  
17 cellular fractions were prepared for immunoblotting. Typically, 20 µg of protein  
18 was loaded onto each lane. Support-positive immunoblotting results were  
19 evaluated following the criteria described by Antibodypedia  
20 ([http://www.antibodypedia.com/text/validation\\_criteria#western blot](http://www.antibodypedia.com/text/validation_criteria#western blot)) and HPA<sup>2</sup>.  
21 Basically, an antibody was qualified as immunoblotting positive when the size of  
22 a single or predominant single band on immunoblotting matched or was within 10%

1 of the predicted antigen molecular weight. In some cases, an immunoblotting-  
2 positive conclusion was enhanced when the same predicted protein band was  
3 detected in two or more different cell lysates. Some antibodies detected multiple  
4 bands with different sizes, but the predicted size protein band was also detected.

5

## 6 **IF, FACS validation**

7 A cell line with target protein expression (HPA data) was selected for IF and  
8 FACS assays. Known/predicted subcellular localization of the target protein was  
9 also obtained from HPA or Uniprot (**Table S4**). For cell surface protein targets, IF  
10 and FACS assays were performed under nonpermeable conditions without  
11 detergent in the buffers. For intracellular targets, the permeable condition with 0.1%  
12 Triton added to the buffers was used throughout. Antibody binding signal was  
13 detected using Alexa Fluor 488 and 594 goat anti-mouse IgG secondary  
14 antibodies (Jackson ImmunoResearch, #115-545-003 and #115-585-003). In  
15 brief, cells attached on cover slips (IF assays) or suspended in 1x PBS (FACS  
16 assays) were first fixed in 4% PFA for 10 min. PFA was then removed, and cells  
17 were rinsed three times with 1x PBS. Cells were blocked overnight at 4°C in the  
18 blocking buffer (1x PBS containing 10% normal goat serum, 0.1% Triton was  
19 added for intracellular targets). After removing the blocking buffer, cells were  
20 incubated with primary antibody (dilution in the blocking buffer at 1:100-1000) for  
21 3 hours at room temperature. Cells were rinsed 6 times in 1x PBS before being  
22 incubated with fluorescence-labeled secondary antibody (diluted in blocking

1 buffer at 1:500 dilution ratio with 1/10000 Hoechst 33258, Sigma #94403) for 1  
2 hour. Finally, cells were rinsed 3 times with 1x PBS. IF images were recorded  
3 with the Nikon confocal system A1Si. The 3D reconstruction of the IF results  
4 were performed in ImageJ (NCBI free software). IF staining patterns were  
5 compared with HPA data to confirm the subcellular localization of the target  
6 proteins. The FACS data were collected using the BD Accuri C6 Plus system. A  
7 control sample without primary antibody and another sample with isotype control  
8 antibody were used.

9

10 **IP and Mass-spectrum assays to identify antibody binding target**

11 Immunoprecipitation (IP) assays were performed using CNBr-activated  
12 Sepharose 4B (GE, #17-0430-02) by following the user's manual. In brief, 200 µg  
13 of purified PETAL mAbs was cross-linked to 20 µL of hydrolyzed CNBr beads  
14 and used to pull-down target protein from 1 mg of cell membrane or nuclear  
15 protein samples. Typically, an excessive amount of antibodies was used. A  
16 similar procedure was developed following the instructions described previously<sup>18</sup>  
17 to identify the binding protein targets of the tested antibodies.

18 Essentially, target identification for immunoblotting-successful (yielding single  
19 or predominant single band) mAbs used for IP was done by comparing the silver  
20 staining result of the IP product with immunoblotting data on samples before and  
21 after IP. Expected size band (matched on silver staining and immunoblotting)  
22 was selected for MS analysis. For some mAbs, more than one band on SDS-

1 PAGE was selected for MS identification; several identified protein targets could  
2 be the binding target of an antibody. For antibodies that failed in the  
3 immunoblotting assay, their IP products separated on silver-stained SDS-PAGE  
4 were compared to IP products from other mAbs. One predominant specific band  
5 or several stoichiometric specific bands were selected for MS analysis

6 Once one or more bands were selected for MS, 20  $\mu$ L of IP product was  
7 separated in an SDS-PAGE gel and stained with Coomassie blue R-250. The  
8 selected target bands were excised and sent to MS facilities (Instrumental  
9 Analysis Center of SJTU, or Biological Mass Spectrometry Facility at Robert  
10 Wood Johnson medical school and Rutgers, the state university of New Jersey)  
11 for target identification using LC-MS/MS on Thermo Q Exactive HF or Thermo  
12 Orbitrap-Velos Pro.

13 Mascot distiller (version 2.6, Matrix Science) or Protein Discovery software  
14 (version 2.2) was used to convert raw to mgf or mzML format for downstream  
15 analysis. The LC-MS/MS data were searched against Uniprot human (557,992  
16 proteins) for the human cell culture sample, Uniprot zebrafish (61,675) for  
17 zebrafish tissue sample or Uniprot maize (137,106) for corn tissue sample.  
18 Enzyme specificity was set as C-terminal to Arg and Lys and allowed for two  
19 missed cleavages. Furthermore,  $\pm$ 10 ppm and 0.02 Da (Thermo Q Exactive HF)  
20 or 1 Da (Thermo Orbitrap-Velos Pro) were used as tolerance for precursor (MS)  
21 and product ions (MS/MS), respectively. Carbamidomethylated cysteine was set  
22 as complete modifications. N-Terminal protein acetylation and oxidation of  
23 methionine were set as potential modifications. Deamidation at asparagine and

1 glutamine and oxidation at methionine and tryptophan were specified as variable  
2 modifications.

3 To ensure MS data quality, we used a threshold of 20 total identified peptide  
4 number or 5 nonredundant peptide number to achieve high confidence of the  
5 identified protein target. In analyzing the MS result for the antibody, the identified  
6 protein list was first prioritized using the total identified peptide number. Proteins  
7 that were identified in multiple different antibodies were excluded. For most  
8 antibodies in this study, a unique protein with the highest total identified peptide  
9 number and matched protein size detected on the silver staining and  
10 immunoblotting was selected as the target protein.

11 The mass spectrometry proteomics data have been deposited to the  
12 ProteomeXchange Consortium via the PRIDE<sup>19</sup> partner repository with the  
13 dataset identifier PXD011629 (Reviewer account details: Username:  
14 reviewer41517@ebi.ac.uk, Password: 7ZqfVOM8).

15

## 16 **Abundance distribution and molecular function analysis**

17 The identified membrane, nuclear and other proteins were from the reference  
18 database (Nucleoplasm protein database and Nuclear membrane plus Plasma  
19 membrane protein database from Human Proteome Atlas). Expression  
20 abundance information of human proteins was obtained from the PaxDb  
21 database. Function distributions were clustered using the PANTHER  
22 classification system<sup>20</sup> depending on the molecular function.

1

2 **ChIP-Seq assay**

3 The ChIP and input DNA libraries were prepared as previously described<sup>21, 22</sup>.  
4 Briefly, 10 million HepG2 cells were cross-linked with 1% formaldehyde for 10  
5 min at room temperature and then quenched with 125 mM glycine. The  
6 chromatin was fragmented and then immunoprecipitated with Protein A + G  
7 magnetic beads coupled with antibodies against SMRC1, SATB1 and NFIC. After  
8 reverse crosslinking, ChIP and input DNA fragments were used for library  
9 construction with NEBNext Ultra Ligation Module (E7445, NEB). The DNA  
10 libraries were amplified and subjected to deep sequencing with an Illumina  
11 sequencer. The ChIP-seq data processing was performed as we reported  
12 recently<sup>22</sup>. *Cis*-regulatory sequence elements that mediate the binding of SMRC1,  
13 SATB1 or NFIC were predicted with MEME-ChIP<sup>23</sup>.

14

15 **Internalization Assay**

16 For the IF assay, live PC9 cells were cultured on coverslips and incubated with  
17 10 µg/mL of mAbs for 1 hour on ice before being washed 3 times with PBS. Cells  
18 were then cultured at 37°C for 0, 2, or 4 hours before fixation with 4% PFA.  
19 Antibodies were then labeled with Alexa Fluor 488 conjugated anti-Mouse  
20 antibody. Images were acquired by Nikon confocal system A1Si.

1 For the FACS assay, live PC9 cells were incubated with 10 µg/mL of mAbs  
2 for 0.5 hour on ice before being washed 3 times with PBS. Cells were then  
3 cultured at 37°C for up to 4 hour before fixation with 4% PFA. Cells were then  
4 stained with Alexa Fluor 488-conjugated anti-Mouse antibody and analyzed with  
5 FACS. Surface MFI was calculated. Surface MFI, which represented surface  
6 localization of mAbs, was measured by FACS.

7

8 **Indirect cytotoxicity assay**

9 PC9 cells were cultured in 96-well plates at 2000/well confluence overnight. Cells  
10 were treated with serial dilution of mAbs together with 2 µg/mL of Monomethyl  
11 auristatin E (MMAE)-conjugated secondary goat anti-mouse IgG antibody for 72  
12 hours. Cell number was then calculated by Cell Counting Kit-8 (CCK8, Dojindo,  
13 #CK04-20). Antibody-drug conjugation services were provided by Levena  
14 Biopharma, Nanjing.

15

16 ***In vivo* tumor models**

17 For the cell line-derived xenograft (CDX) model,  $5 \times 10^6$  NCI-H1975 cells were  
18 suspended in Matrigel (BD Biosciences, #354234) and injected subcutaneously  
19 (s.c.) to the right flank of female BALB/c nude mice (jsj-lab). For studies with the  
20 patient-derived xenograft (PDX) model, the tumor fragments from patients with  
21 lung squamous cell carcinoma were passaged twice in NOD-SCID mice (Beijing

1      Vital River Laboratory Animal Technology Co., Ltd.). Tumor fragments obtained  
2      from *in vivo* passage were then implanted s.c. in the right flank of female BALB/c  
3      nude mice (jsj-lab). Body weight and tumor volume (0.5 x length x width<sup>2</sup>) were  
4      measured every 3 days. Mice were randomized into control and treatment groups  
5      based on the primary tumor sizes (median tumor volume of approximately 100  
6      mm<sup>3</sup>). Pb44707-ADCs and control ADCs were administered intravenously (i.v.)  
7      every third day and repeated for a total of three times (Q3Dx3). Gefitinib (Selleck,  
8      ZD1839) was administered intraperitoneally (i.p.) every day.

9

10     **siRNA knockdown and overexpression**

11     PC9 was transfected with siRNA targeting human PROTEINX (sense, 5'-CUA  
12     CUU UAC UGG AAG GUU Att-3'; and antisense, 5'-UAA CCU UCC AGU AAA  
13     GUA Gtt-3'), which has been reported previously<sup>24</sup> or control siRNA (sense,  
14     5'-UUC UCC GAA CGU GUC ACG Utt-3'; and antisense, 5'-ACG UGA CAC  
15     GUU CGG AGA Att-3') by Lipofectamine 2000 (Thermo Fisher, #11668019) 48  
16     hr before performing experiment.

17     PIEZO1-GFP plasmid used for overexpression validation was a gift from Prof.  
18     David Beech, which was described previously<sup>25</sup>. PIEZO1-GFP was transfected  
19     into HUVEC cells by lippofectamine200. Cells were fixed and stained with anti-  
20     PIEZO1 antibody following IF procedure described above.

21

1 **Peptide blocking assay**

2 1 mg/ml Pb44707 was pre-incubated with 1 mg/ml PROTEINXs recombinant  
3 protein (Abcam) or PROTEINX peptide in 1:1 ratio at 4°C overnight before used  
4 in FACS analysis.

5

6 **Antibody cellular binding site quantification**

7 The antibody binding sites on cell lines were determined with the QIFIKIT (Dako,  
8 # K0078) according to the manufacturer's instructions.

9

## 1 References for Methods

- 2 1. Altschul, S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database  
3 search programs. *Nucleic acids research* **25**, 3389-3402 (1997).
- 4 2. Berglund, L. et al. A gene-centric Human Protein Atlas for expression profiles based on  
5 antibodies. *Molecular & cellular proteomics : MCP* **7**, 2019-2027 (2008).
- 6 3. Hopp, T.P. & Woods, K.R. Prediction of protein antigenic determinants from amino acid  
7 sequences. *Proceedings of the National Academy of Sciences of the United States of  
8 America* **78**, 3824-3828 (1981).
- 9 4. Welling, G.W., Weijer, W.J., van der Zee, R. & Welling-Wester, S. Prediction of sequential  
10 antigenic regions in proteins. *FEBS letters* **188**, 215-218 (1985).
- 11 5. Getzoff, E.D., Tainer, J.A., Lerner, R.A. & Geysen, H.M. The chemistry and mechanism of  
12 antibody binding to protein antigens. *Advances in immunology* **43**, 1-98 (1988).
- 13 6. Chou, P.Y. & Fasman, G.D. Prediction of protein conformation. *Biochemistry* **13**, 222-245  
14 (1974).
- 15 7. Hopp, T.P. Use of hydrophilicity plotting procedures to identify protein antigenic  
16 segments and other interaction sites. *Methods in enzymology* **178**, 571-585 (1989).
- 17 8. Moura, A., Savageau, M.A. & Alves, R. Relative amino acid composition signatures of  
18 organisms and environments. *PLoS one* **8**, e77319 (2013).
- 19 9. de StGroth, S.F. & Scheidegger, D. Production of monoclonal antibodies: strategy and  
20 tactics. *Journal of immunological methods* **35**, 1-21 (1980).
- 21 10. Brochet, X., Lefranc, M.P. & Giudicelli, V. IMGT/V-QUEST: the highly customized and  
22 integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic  
23 acids research* **36**, W503-508 (2008).
- 24 11. Giudicelli, V., Brochet, X. & Lefranc, M.P. IMGT/V-QUEST: IMGT standardized analysis of  
25 the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. *Cold Spring  
26 Harbor protocols* **2011**, 695-715 (2011).
- 27 12. Delfani, P. et al. Technical Advances of the Recombinant Antibody Microarray Technology  
28 Platform for Clinical Immunoproteomics. *PLoS one* **11**, e0159138 (2016).
- 29 13. Borrebaeck, C.A. & Wingren, C. Design of high-density antibody microarrays for disease  
30 proteomics: key technological issues. *J Proteomics* **72**, 928-935 (2009).
- 31 14. Cohen, J. Statistical Power Analysis. *Current Directions in Psychological Science* **1**, 98-101  
32 (1992).
- 33 15. Hagen, J.v. & Michelsen, U. in Methods in Enzymology, Vol. Volume 533. (ed. L. Jon) 25-  
34 30 (Academic Press, 2013).
- 35 16. Wang, Y. et al. Tomato nuclear proteome reveals the involvement of specific E2  
36 ubiquitin-conjugating enzymes in fruit ripening. *Genome biology* **15**, 548 (2014).
- 37 17. Zhang, D.G., Jiang, A.G., Lu, H.Y., Zhang, L.X. & Gao, X.Y. Isolation, cultivation and  
38 identification of human lung adenocarcinoma stem cells. *Oncology letters* **9**, 47-54  
39 (2015).
- 40 18. Marcon, E. et al. Assessment of a method to characterize antibody selectivity and  
41 specificity for use in immunoprecipitation. *Nature methods* **12**, 725-731 (2015).
- 42 19. Vizcaino, J.A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic acids  
43 research* **44**, 11033 (2016).
- 44 20. Mi, H., Muruganujan, A., Casagrande, J.T. & Thomas, P.D. Large-scale gene function  
45 analysis with the PANTHER classification system. *Nature protocols* **8**, 1551 (2013).
- 46 21. Zhao, X.D. et al. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations

1 reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**,  
2 286-298 (2007).

3 22. Zhang, X.L. et al. Integrative epigenomic analysis reveals unique epigenetic signatures  
4 involved in unipotency of mouse female germline stem cells. *Genome biology* **17**, 162  
5 (2016).

6 23. Machanick, P. & Bailey, T.L. MEME-ChIP: motif analysis of large DNA datasets.  
7 *Bioinformatics* **27**, 1696-1697 (2011).

8 24. Kobayashi, K. et al. Clinical significance of PROTEINX variant 9 expression as a prognostic  
9 indicator in bladder cancer. *Oncology reports* **36**, 2852-2860 (2016).

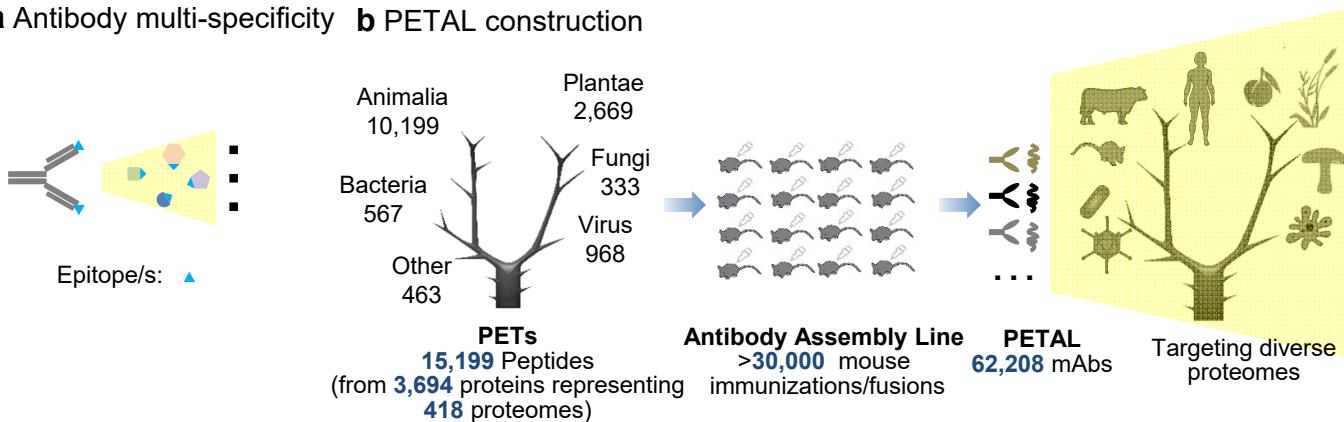
10 25. Li, J. et al. Piezo1 integration of vascular architecture with physiological force. *Nature*  
11 **515**, 279-282 (2014).

12

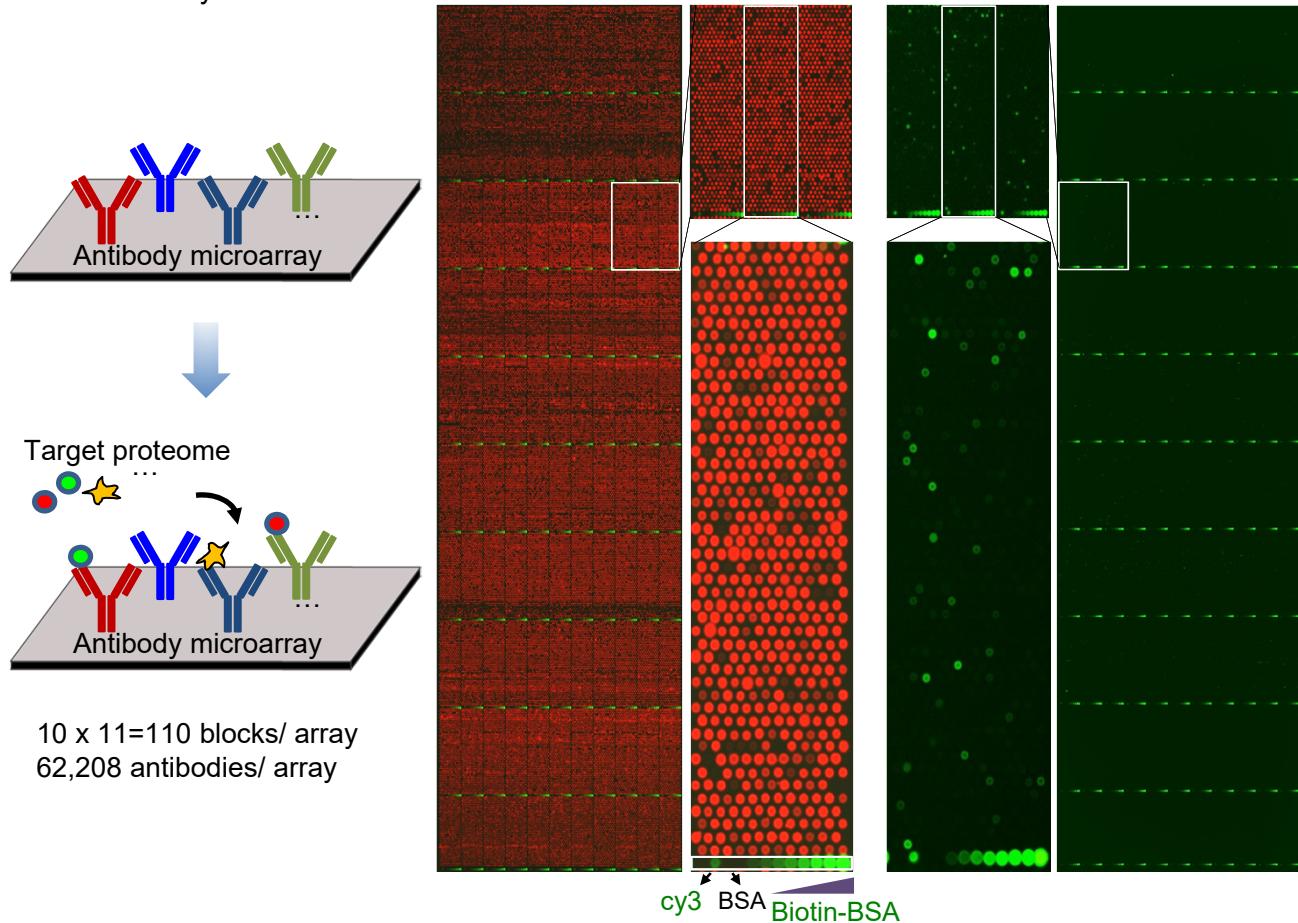
13

14

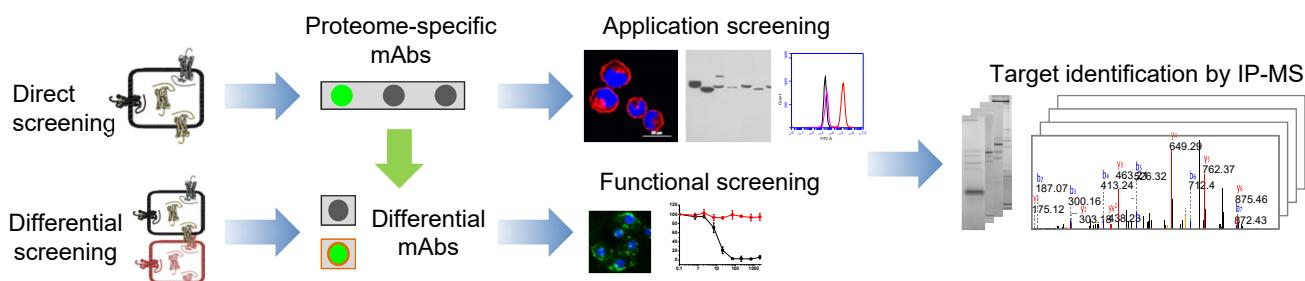
**a Antibody multi-specificity** **b PETAL construction**



**c PETAL microarray construction**

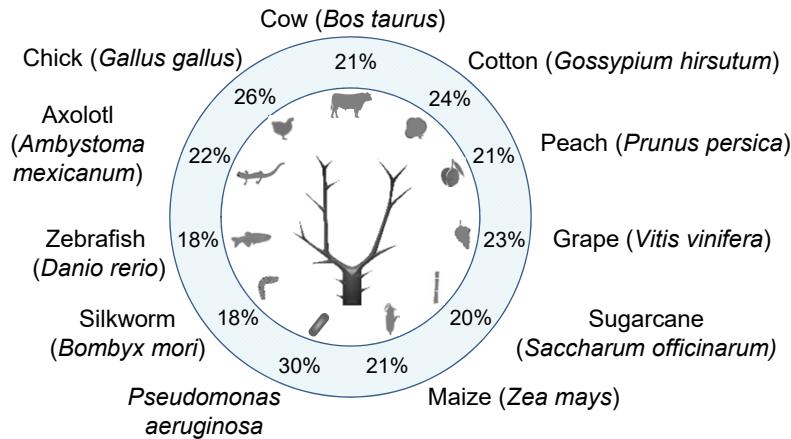


**d Workflow for proteome-scale antibody generation and target discovery**

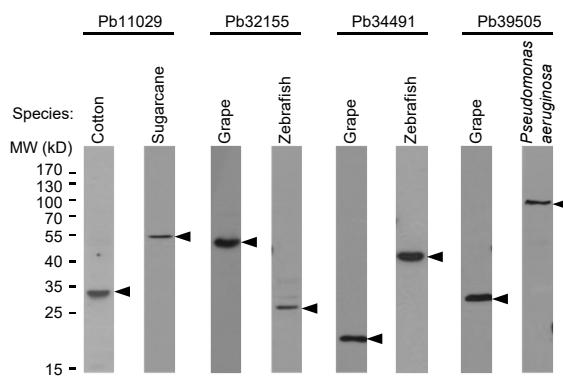


**Figure 1. Construction and application of PETAL and its array for antibody/target discovery**

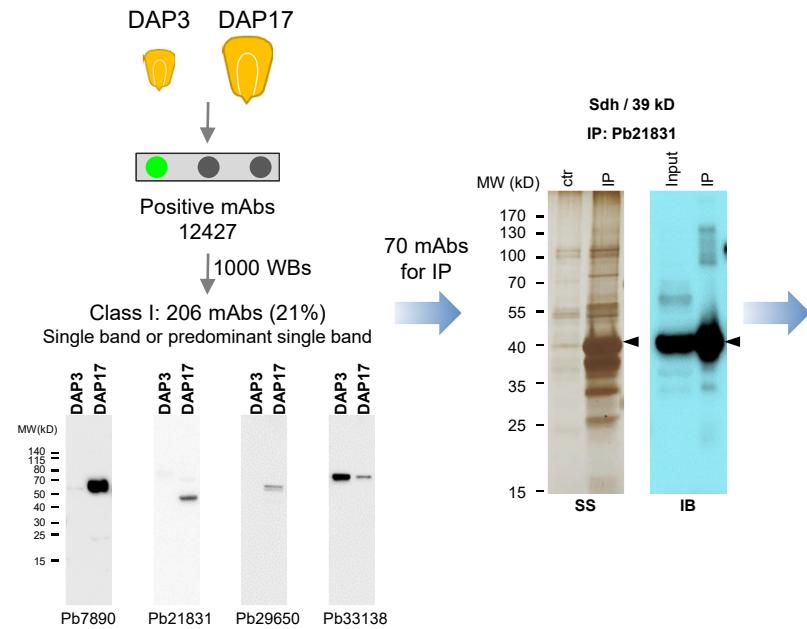
**a** Immunoblotting antibodies for proteomes



**b** Same mAb detected bands with different MW sizes in two different proteomes



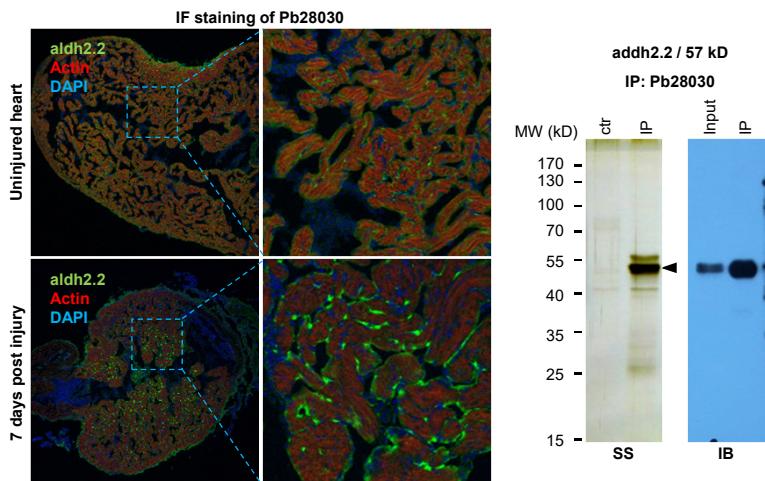
**c** Identification of maize seed development stage-specific mAb-target pairs



Total: 19 proteins paired with 23 mAbs, 6 are shown

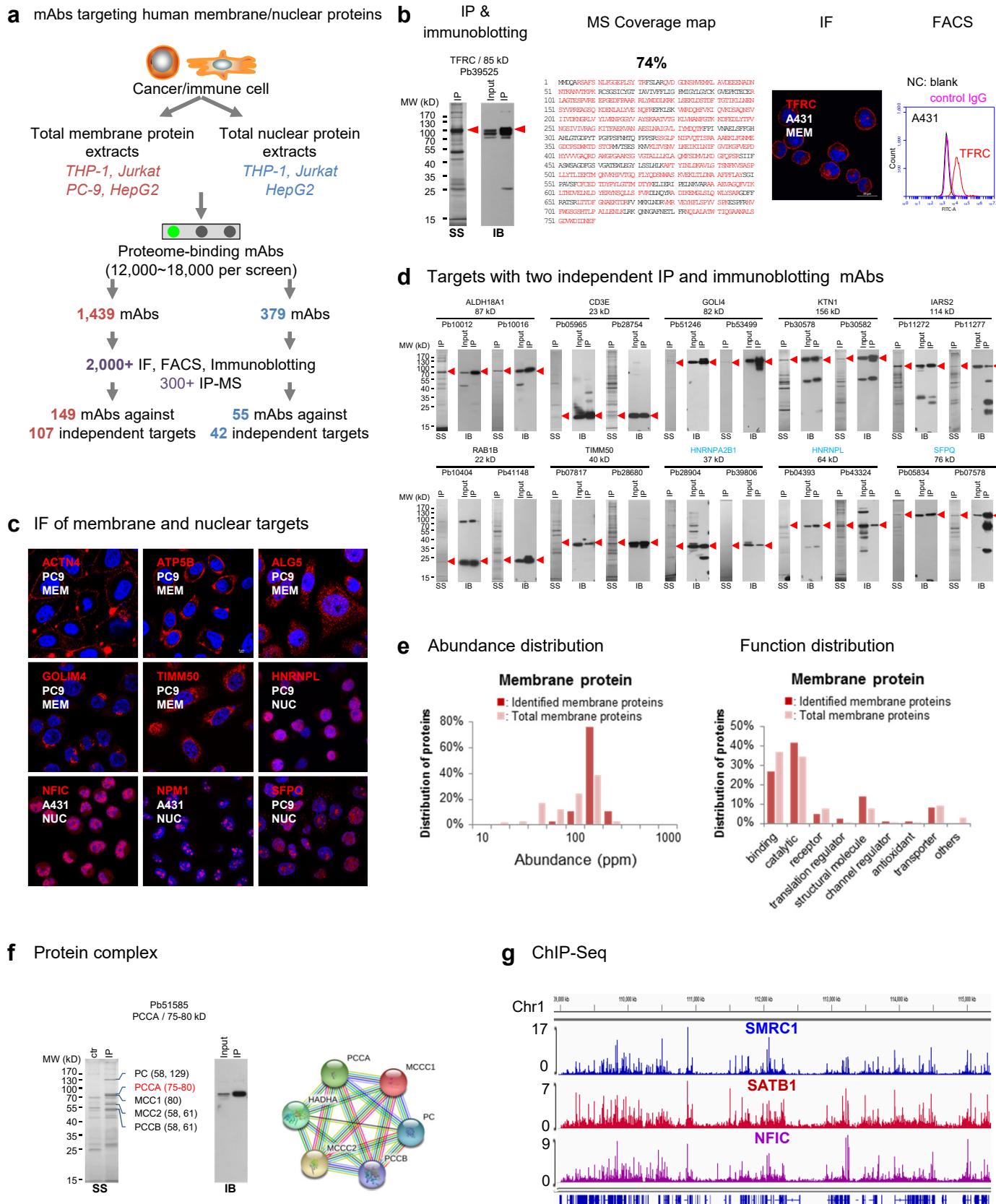
mAb ID	IP band size (kD)	IP-MS identified Target	Uniprot ID	Regulation in DAP17
Pb7890	53	Leg1; Legumin 1	Q946V2	Up
Pb29650	52	Bt2; Glucose-1-phosphate adenyllyltransferase	D3YKV1	Up
Pb31074	91	Sbe1; 1 4-alpha-glucan-branched enzyme 2 chloroplastic/amloplastic	Q08047	Up
Pb21831	39	Sdh; Sorbitol dehydrogenase	A0A1D6K L30	Up
Pb33138	65	Phgdh; D-3-phosphoglycerate dehydrogenase	B4FWI4	Down
Pb30103	35	Gid12; Gibberellin receptor GID1L2	A0A1D6H B17	Up

**d** Identification of proteins related to heart injury in zebrafish

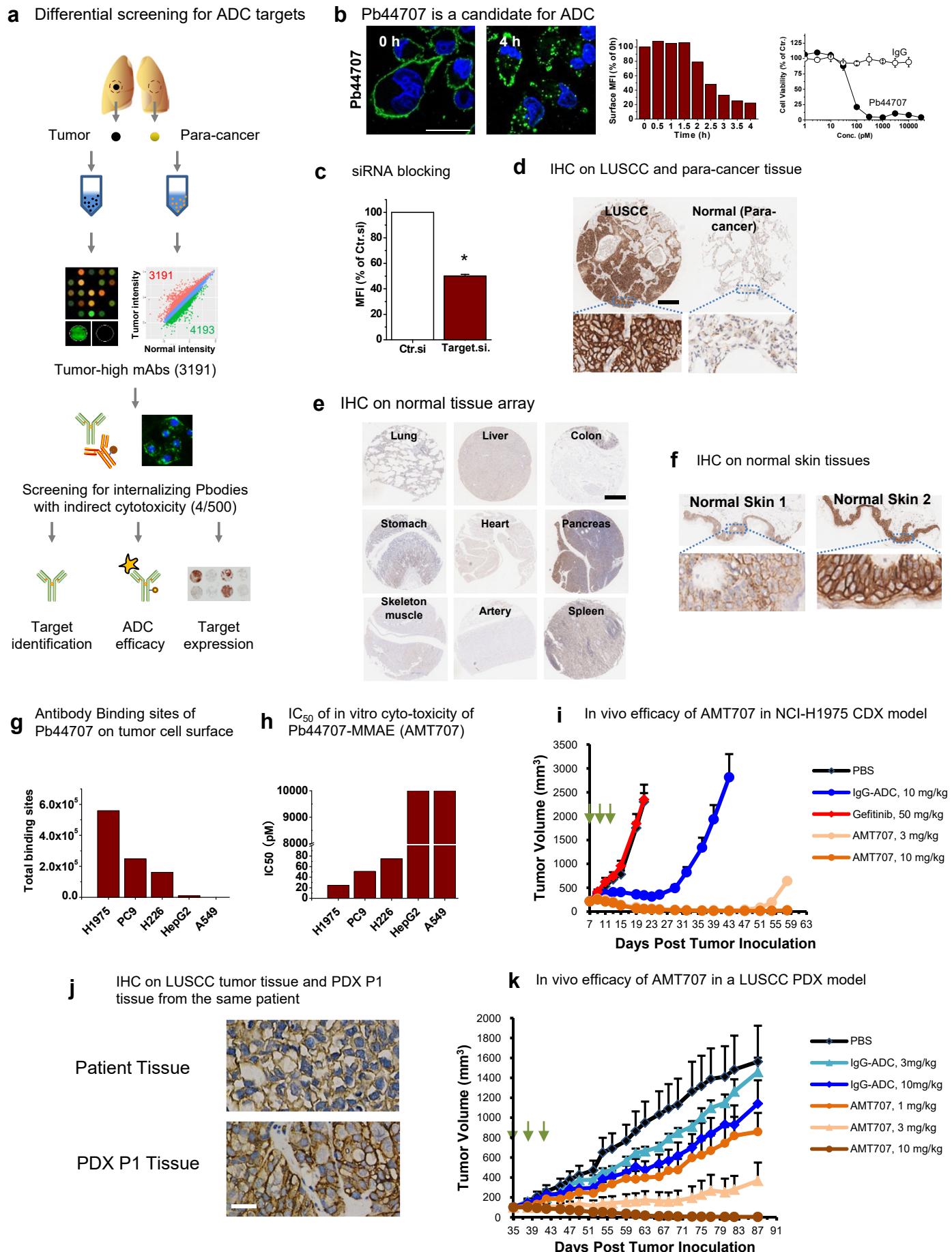


mAb ID	IP band size (kD)	IP-MS identified Target	Uniprot ID	Regulation in injured heart
Pb28030	55	aldh2.2; Aldehyde dehydrogenase 2 family tandem duplicate 2	F1QZU7	UP
Pb27552	45, 55	krt4/krt8; Keratin 4 and Keratin 8	F1QK60 and Q6NWF6	UP
Pb2775	40	acadsb; Acyl-CoA dehydrogenase short-branched chain	F1QFQ7	DOWN
Pb7308	80	postnb; Periostin, osteoblast specific factor	Q75U66	UP
Pb4124	90	cast; Calpastatin	A0A0R4I LE0	UP
Pb7542	70	aldh16a1; Aldehyde dehydrogenase 16 family, member A1	F1QGP1	UP

**Figure 2. PETAL targets diverse proteomes for antibody and target discovery**

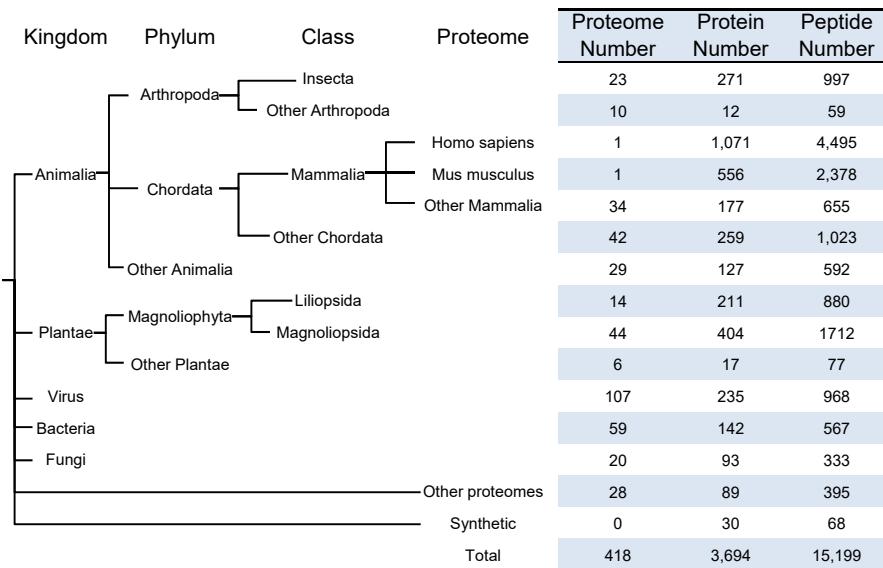


**Figure 3. Proteome scale antibody generation for human membrane and nuclear proteins**

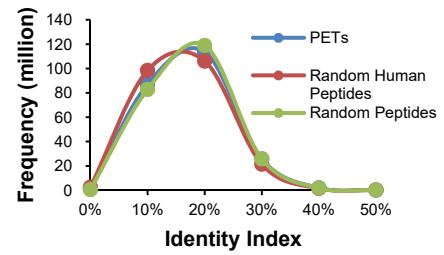


**Figure 4. Differential array screening for ADC therapeutic target/antibody**

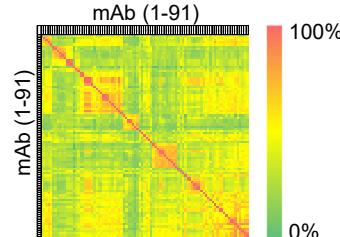
**a** Proteome, protein and peptide for PETAL construction



**b** Diversity of the PETs



**c** Homology matrix of 91 PETAL mAbs (CDR)



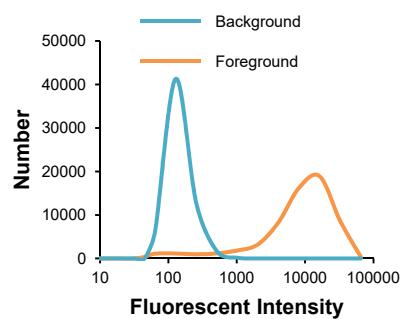
**d** Diversity of PETAL mAbs (the 91 sequenced mAbs)

Amino acid difference in CDR	Unique mAbs Number	Ratio
0	86	93.4%
1	84	92.3%
2	82	90.1%
3	81	89.0%
4	79	86.8%

**e** Unique mAb against the same PET immunogen (the 91 sequenced mAbs)

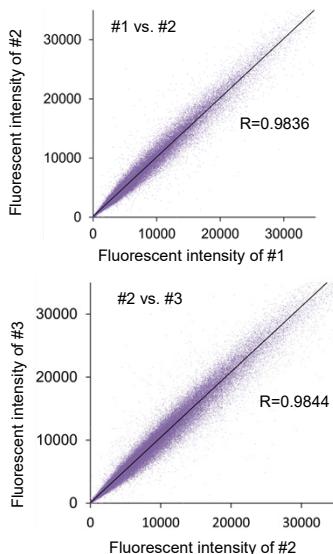
mAbs/PET	PETs	mAbs	Unique mAbs	
			Number	Ratio
5	1	5	5	100.0%
4	4	16	13	81.3%
3	9	27	26	96.3%
2	10	20	18	90.0%
1	23	23	23	100.0%
Sum	47	91	85	93.4%

**f** Antibody fluorescent profile



	Foreground	Background
Minimum	40	41
Maximum	51,486	1,572
Median	7,239	98
Mean	8,785.4	112.7
S.D.	7,004.5	61.3

**g** Array screening reproducibility

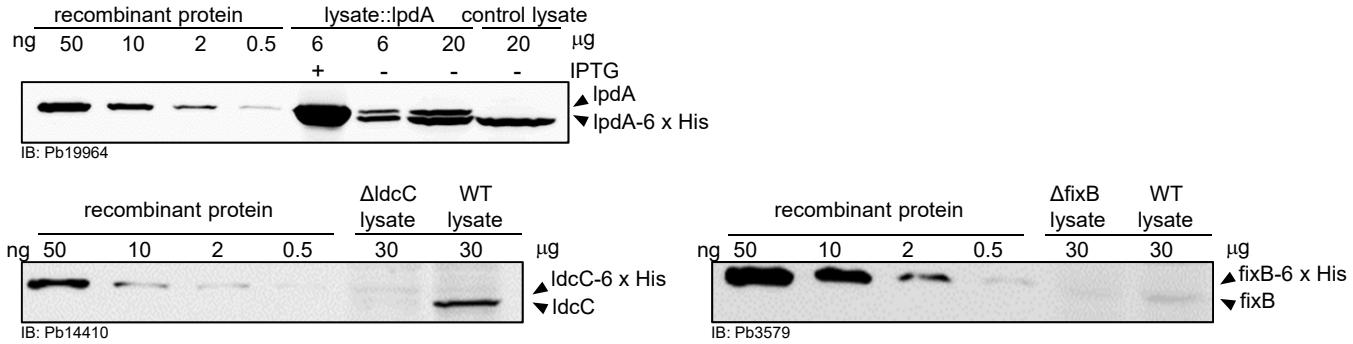


**Supplementary Figure 1. PETAL antigens, antibody diversity and array performance**

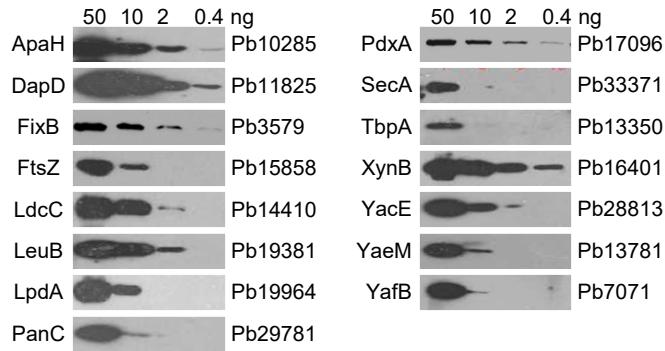
**a PETAL array screening of recombinant proteins**

	Input	ELISA positive (<=1 µg/ mL)	IB succeed on recombinant protein (<50 ng)	IB succeed on endogenous protein
Proteins	81	38	25 (31%)	21 (26%)
Positive mAbs	-	141	72	34

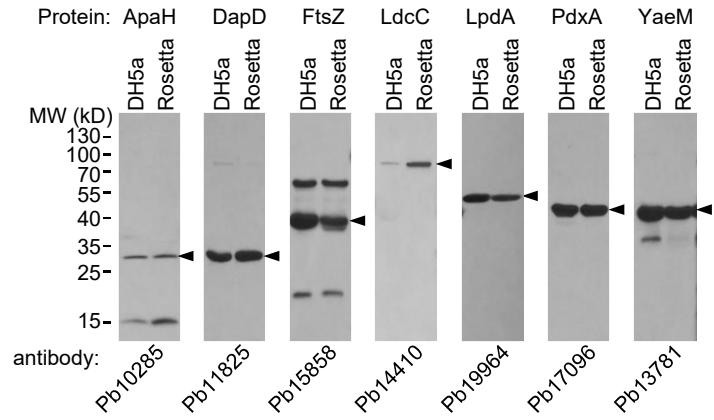
**b Immunoblotting (*E. coli*)**



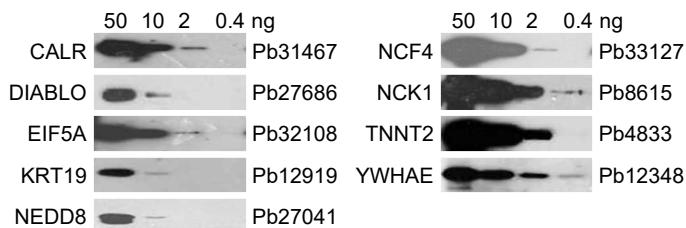
**c Immunoblotting on recombinant protein (*E. coli*)**



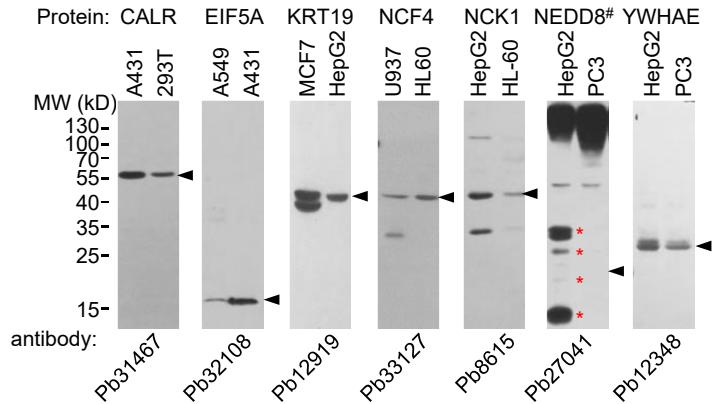
**d Endogenous immunoblotting (*E. coli*)**



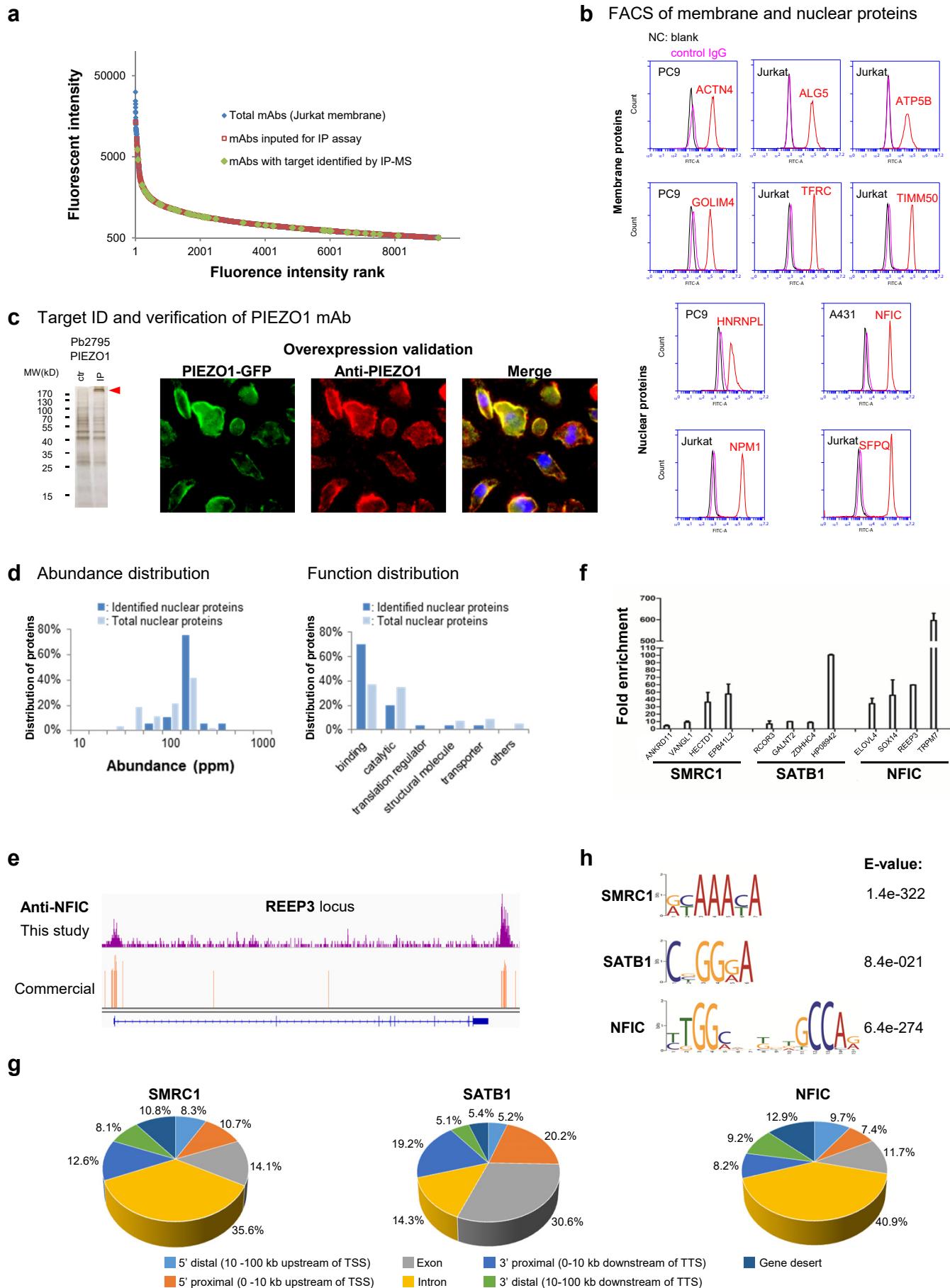
**e Immunoblotting on recombinant protein (human)**



**f Endogenous immunoblotting (human)**

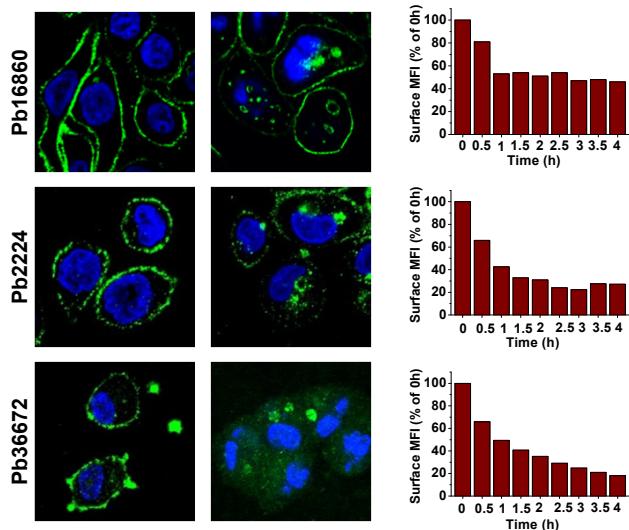


**Supplementary Figure 2. PETAL screen by protein antigens**

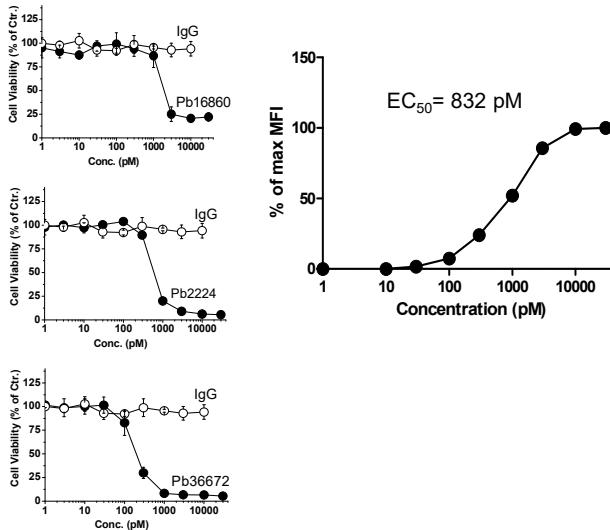


**Supplementary Figure 3. Proteome scale antibody generation for membrane and nuclear proteins**

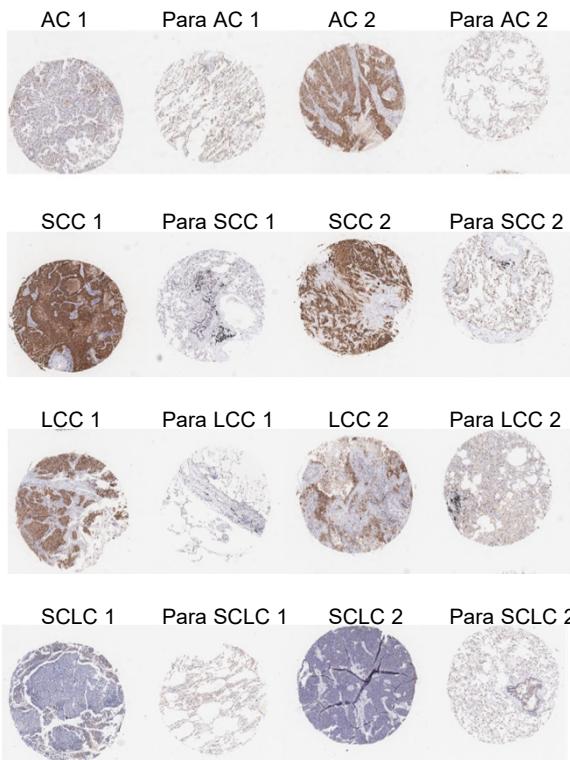
**a Internalizing antibodies**



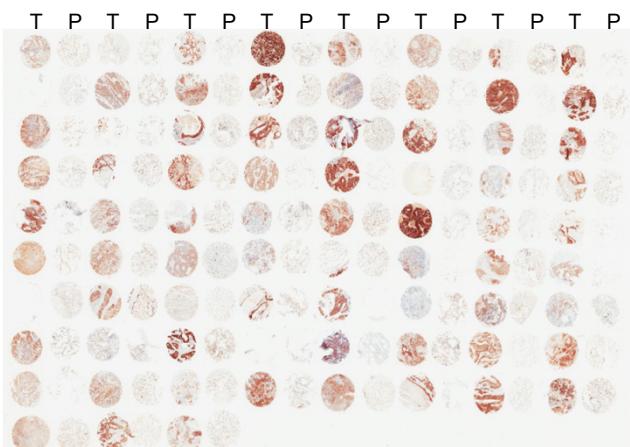
**b Pb44707 affinity**



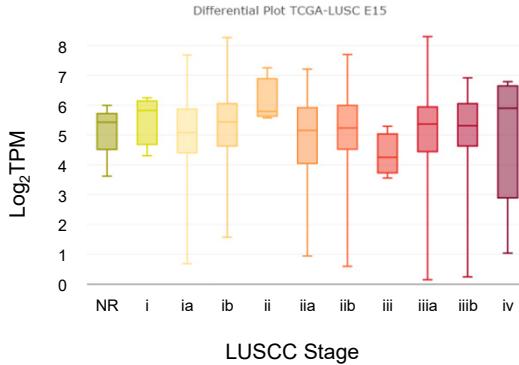
**c IHC on lung cancer tissue array**



**d IHC on LUSCC tissue array**



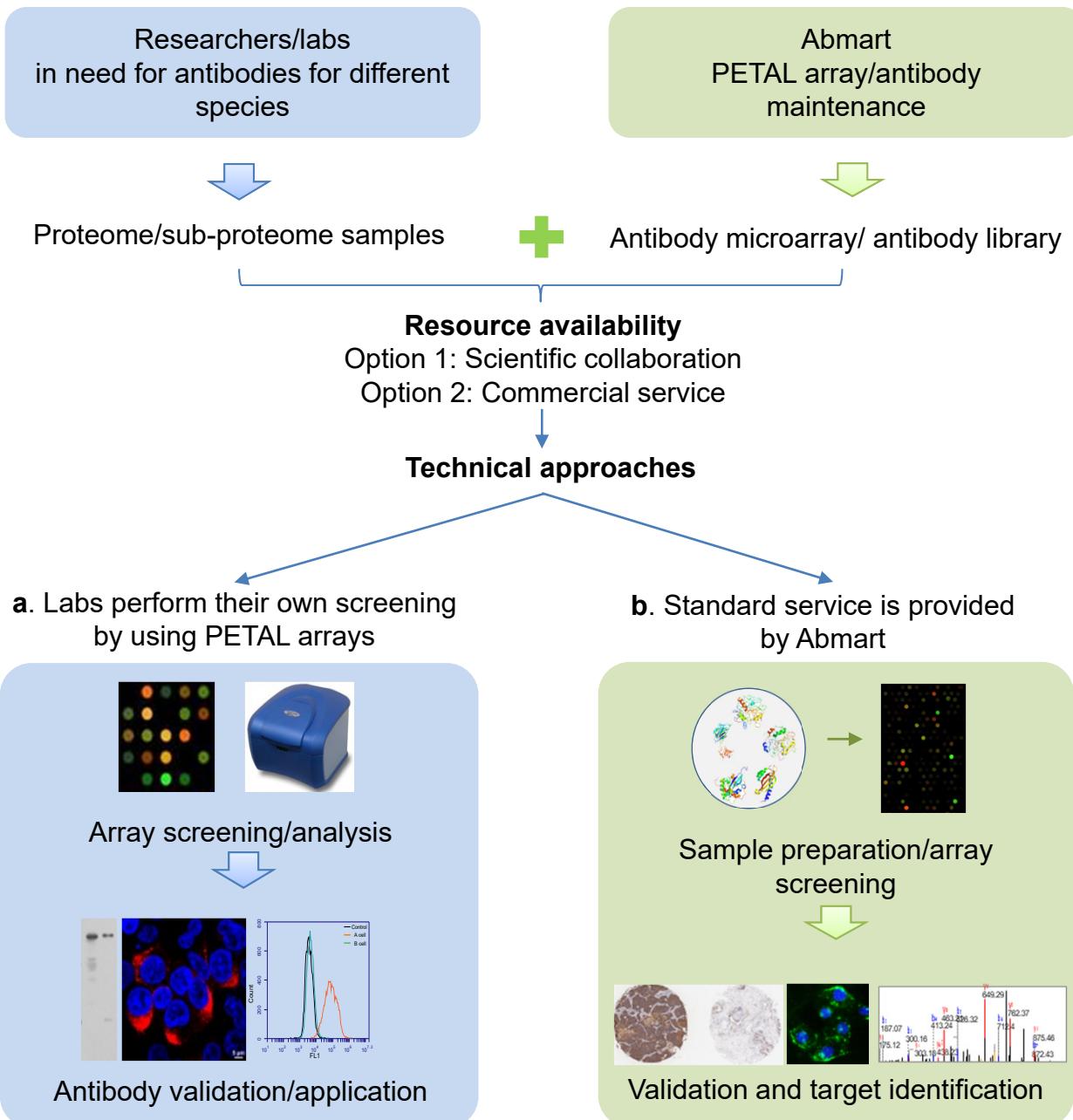
**e Analysis of CD44v9 mRNA expression in different stages of LUSCC (TCGA data)**



**f Summary of the IHC result of Pb44707**

	0	1+	2+	3+
<b>AC</b>	10/26	4/26	10/26	2/26
<b>SCC</b>	11/89	10/89	31/89	37/89
<b>LCC</b>	2/5	2/5	1/5	0/5
<b>SCLC</b>	8/8	0/8	0/8	0/8
<b>Normal Skin</b>	0/20	6/20	8/20	6/20

**Supplementary Figure 4. Differential array screening for ADC therapeutic target/antibody**



**Supplementary Figure 5. Two approaches to access PETAL library/array**