Discovering Cell-targeting Ligands and the Cell Surface Receptors via Selection of DNA-encoded Chemical Libraries (DELs) against Cancer Cells without Predefined Targets

Yuhan Gui,[†] Rui Hou,[†‡] Yuchen Huang, [™] Yu Zhou, [†‡] Shihao Liu,[†] Ling Meng,[†‡] Ying Li,[†] Fong Sang Lam,[†] Clara Shania Wong,[†] Chan Tat Yin,[†] Gang Li,[™]* Xiaojie Lu,[§]* and Xiaoyu Li[†‡]*

[†] *Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China*
[‡] *Laboratory for Synthetic Chemistry and Chemical Biology Limited, Health@InnoHK, Innovation and Technology Commission, Units 1503-1511, 15/F., Building 17W, Hong Kong SAR, China*
[§] *State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 501 Haike Road, Zhang Jiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China*
[™] *Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China*

E-mail: xiaoyuli@hku.hk; xjlu@simm.ac.cn; ligang@szbl.ac.cn

Keywords: DNA-encoded chemical library; drug discovery; high-throughput screening; targeted drug delivery.

ABSTRACT:

Small molecule ligands that can specifically recognize the surface of cancer cells have wide utilities in cancer diagnosis and treatment. Screening large combinatorial libraries against live cells is an effective approach to discover cell-targeting ligands. In the past decade, DNA-encoded chemical library (DEL or DECL) has become a powerful technology in drug discovery and been successfully used in ligand discovery against numerous biological targets. However, nearly all DEL selections had predefined targets, whereas completely unbiased DEL selections interrogating the entire cell surface remain underexplored. In this report, we systematically optimized cell-based DEL selection method to perform unbiased selections against cancer cells without predefined targets. A 104.96-million-member DEL was selected against MDA-MB-231 and MCF7, a pair of breast cancer cell lines with high and low metastatic properties, respectively, and cell-specific small molecule ligands and ligand combinations ("clusters") have been identified. We further show that the ligand cluster could be optimized to improve the binding affinity and applied in cell-targeting applications including cancer photodynamic therapy and targeted drug delivery. Finally, we leveraged the DNA tag of the DEL compounds and identified the cell surface receptor of an individual ligand targeting MDA-MB-231 cells. Overall, this work provides an efficient method for discovering cell-targeting small molecules and demonstrated the potential of DELs as a tool for cancer biomarker discovery.

Introduction.

Cancer remains one of the leading causes of death worldwide, with the mortality rate expected to surpass cardiovascular diseases by 2030.[1] Besides surgery, chemotherapy, targeted therapy, and immunotherapy are the major treatment modalities. Regardless of the modality, the therapeutic agents need to reach the tumor site and bind to the cancer cells. The binding specificity is extremely important to avoid systemic toxicity and improve the efficacy of the anti-cancer drugs. Thus, the discovery of cell-specific ligands is an important task in cancer research.[2] Cell-targeting ligands are also useful tools for studying fundamental biology. The cell surface displays many biomolecules that are often present in higher order structural arrangements such as multimer assembly, complexes, and microdomains that are characterized by their cell types and cell properties. [3] Ligands that can recognize specific cell surface features can be potentially developed as molecular probes for biomarker discovery and exploring the underlying signaling pathways.[2] Antibodies have been widely used to target cell surface antigens;[4, 5] however, antibodies are large proteins that can suffer from stability issues, high manufacturing costs, incompatibility with oral administration, and organ toxicity and hypersensitivity reaction issues.[6, 7]

Many combinatorial libraries, including phage display, mRNA display, yeast display, bacterial display, and one-bead one compound (OBOC) libraries, have been used to discover cell-targeting ligands through unbiased selections against live cells, tissue samples, *ex vivo* and *in vivo* systems, and even in humans.[8-12] However, the identified ligands are mostly peptides or peptidomimetics with less favorable pharmacokinetic properties.[8] Cell-SELEX is a powerful technique to identify cell-binding aptamers,[13-15] but it is limited to nucleic acid structures. Small molecules have better drug-like properties and are more chemically tractable; however, efficient ways to identify cell-targeting small molecules are still lacking, especially in a high-throughput format. On another aspect, although numerous cell-targeting peptides and aptamers have been identified from unbiased selections against cancer cells and several them have already been approved for clinical uses, it is striking that only a very small fraction of the ligands have had their cellular receptors identified.[3, 8, 14, 16] Target identification falls far behind ligand discovery, representing a major issue in the biological and clinical development of these ligands.

Originally proposed by Brenner and Lerner in 1992,[17] DNA-encoded chemical library (DEL or DECL) has since become a powerful platform widely used in drug discovery.[18-37] In a DEL, each compound is encoded with a DNA tag, which allows the selection of all the library compounds against the target simultaneously. The selected binders are decoded by PCR amplification and next-generation sequencing (NGS) (Figure 1a). DELs can contain millions to billions of compounds. With the rapid development of DNA-compatible chemistry, the chemical space of DELs has been greatly expanded.[22, 32, 38] DELs have been used to interrogate numerous protein and nucleic acid targets. DEL selections on/inside live cells have also been realized.[39-47] The Bradley group pioneered PNA-encoded library screening on live cells;[48, 49] GSK reported the first cell-based DEL selection with cells overexpressing tachykinin receptor neurokinin-3;[39] the Krusemark group applied a crosslinking method[40] to both cell-surface and intracellular targets;[41-43] the Neri group extensively optimized the experimental conditions for cell-surface selections;[44] Vipergen microinjected both the library and the target's mRNA into frog oocytes to enable intracellular selection;[45] our group developed a method that can selectively label the target with a DNA tag for selections against endogenous membrane proteins;[46] and Kodadek, Paegel, and co-workers selected an OBOC-DEL library against antibodies in human sera from patients with active or latent tuberculosis (TB) to identify ligands specific for active TB IgG antibodies.[50] However, nearly all DEL selections have predefined targets, whereas completely unbiased selections with live cells remain largely unexplored.

We reasoned DEL selection may be an efficient approach to identify cell-targeting ligands. First, DELs contain high chemical diversity that can be sampled by the complex landscape of the cell surface.[51] Besides individual binders, we hypothesize that the selection may also identify ligand combinations that bind to multiple cell surface receptors. Second, the DNA conjugation site provides an exit vector that can be conveniently used for payload and label attachment for various biological applications. Third, DEL avoids the avidity effect from multivalent binding, which is often seen with phage display and OBOC libraries.[52, 53] Herein, we report a method to perform unbiased DEL selections without predefined targets on live cells. This method utilized photo-crosslinking to stabilize the ligand-receptor interactions on the cell (Figure 1b).[40-43, 46, 54] After systematically optimizing the selection conditions, we selected a 104.96-million-member DEL against MDA-MB-231 and MCF7, a pair of breast cancer cells with different metastatic properties.[3, 8, 16, 55-59] Small molecule ligands and ligand combinations ("clusters") specifically targeting the aggressive MDA-MB-231 cells have been identified. Optimization of the cluster composition led to the second-generation ligand cluster with improved binding affinity. Furthermore, the ligand cluster was used to guide the selective killing of MDA-MB-231 cells via

photodynamic therapy and targeted delivery of cytotoxic drug. Finally, leveraging the DNA tag of DEL compounds, we have identified the receptor of an individual ligand specific for MDA-MB-231 cells. Collectively, this study provides an efficient method for discovering cell-targeting small molecules by harnessing the power of DELs and also demonstrated that DELs can potentially be used for biomarker discovery.

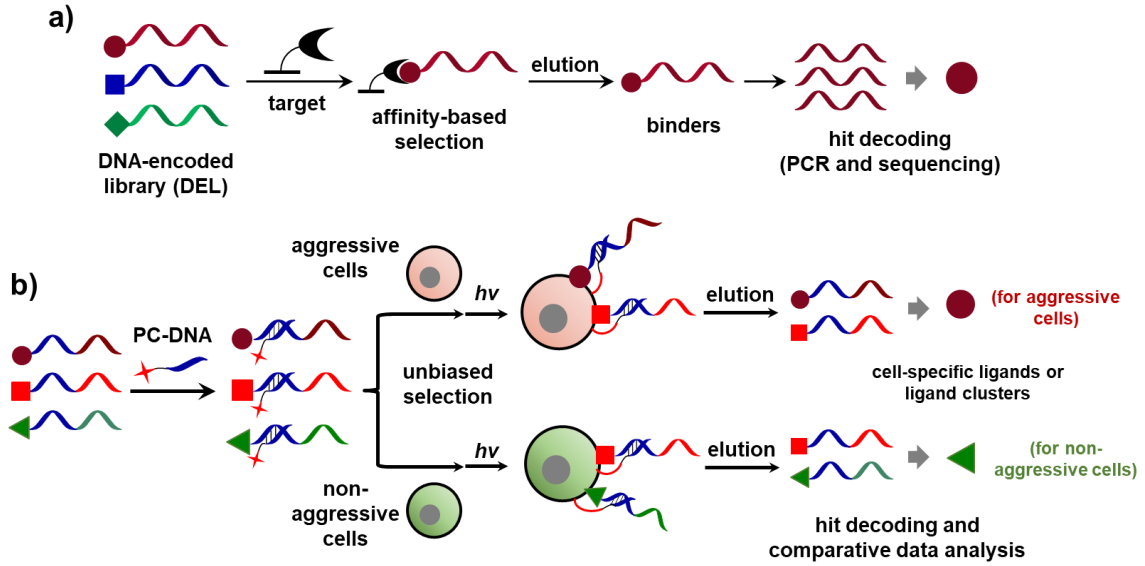

Figure 1: DEL selection against the immobilized targets and the proposed unbiased cell-based selection. a) Scheme of DEL selection with immobilized protein. b) (This work) Unbiased DEL selection against live cells without predefined targets. Photo-crosslinking is used to stabilize the ligand-receptor binding. Selection results from cells of different properties can be compared to identify cell-specific ligands or ligand clusters. PC-DNA: photo-crosslinking DNA; red star: photo-crosslinker.

Results.

Design and optimization of the unbiased DEL selection method on live cells.

Because of the low compound concentration in DELs, the binding equilibrium in the selection is driven by the relatively high target concentration. However, for the selection against the entire cell surface without a predefined target, the techniques to increase the effective target concentration, such as protein overexpression and tagging, can no longer be used. If incubating the cells with the library and then washing away the non-binders, only the strong binders to the targets of high abundance will be identified, whereas moderate binders and possibly even high-affinity ligands binding to low-abundance targets would be lost. To address this issue, we hybridized a photo-crosslinking DNA (PC-DNA) at the primer-binding site of the library (Figure 1b). Upon UV irradiation, the PC-DNA covalently captures the target, thereby stabilizing the ligand-receptor binding. After washing, the binders can be eluted from the cell under denaturing condition for hit decoding. Such a "crosslinking DEL" technique has been proven to be able to improve the library recovery and facilitate the identification of moderate/low-affinity binders both in buffer and on live cells. [40-43, 46, 54, 60-62] The library will be selected in parallel against cancer cells of different cell properties (e.g., aggressive *vs.* non-aggressive), and the enrichment profiles will be compared to identify the cell-specific ligands or ligand clusters (Figure 1b).

Although this study aims to achieve "targetless" selection on live cells, we reasoned that using a model system at first with a known target would facilitate methodology development. As shown in Figure 2a, CA-12 is a membrane carbonic anhydrase implicated in malignant cancers,[63] and **CBS** is a known ligand of CA-12 (K_d = 0.97 μM in DNA-conjugated form).[46, 47, 64] A549 cells expressing high level of CA-12 were used as the target cells (effective concentration: ~1.04 μM; Figure S1). First, **CBS** was covalently tethered to a 16-nt DNA as the binding probe (BP) (**BP-1**), and a complementary 16-nt DNA carrying the photo-reactive phenylazide group and a biotin tag was prepared as the capture probe (CP) (**bio-CP-1**; Figure 2a). Following our previous reported procedure,[47] the **BP-1**/**bio-CP-1** duplex was incubated with the cells (4 ºC, 90 min.) before a brief UV irradiation (365 nm, 1 min.) to trigger the crosslinking. After washing, the binders were eluted by heating,[46] isolated with avidin beads, and analyzed by using Western blot. The results showed the specific labeling of CA-12 with photo-crosslinking (Figure 2b and S2). By replacing the biotin with a fluorescein (FAM) (**fam-CP-1**),

flow cytometry analysis corroborated the specific labeling (Figure 2c). Next, we optimized the selection conditions, aiming to maximize the enrichment fold (EF) of the binder. **CBS** was conjugated to a 35-nt DNA (**BP-2**), and an amine-modified DNA with an orthogonal primer binding site was used as the background (**NP-2**). The DNAs were mixed (ratio 1:10) and hybridized with a 16-nt photo-reactive DNA (**CP-2**). The model library was selected against A549 cells (Figure 2d), and quantitative PCR (qPCR; Figure S3-S4) was used to quantify the enrichment of **CBS**. First, an incubation time of at least 1.5 h and a 2-step washing procedure appeared to be beneficial in improving library recovery and the EF values (Figure 2e-2f). The number of cells had little effect, which is consistent with the previous report (Figure 2g).[44] Interestingly, blocking agents[65] decreased the enrichment (Figure 2h). Running multiple rounds of selections, i.e., the eluted binders are directly used as the input in the next round, is often used in DELs to improve enrichment. We also observed that a two-round selection gave significantly better results (Figure 2i). Next, we increased the **BP-2**/**NP-2** ratio to 1:7,000 and performed the selection under the optimized conditions, and a ~200 EF was obtained after 2 rounds of selections (Figure 2j). Unfortunately, additional round of selection resulted in insufficient amount of the recovered library with qPCR C_T values beyond the reliable concentration range (Figure S4). Furthermore, a model library with chemical diversity was prepared by mixing a **CBS**-DNA conjugate with a 67,600-member dipeptide DEL at equal ratio (Figure 2k).[66] The 67,601-member library was selected against A549 cells. The selected library samples were PCR-amplified with NGS-compatible primers and submitted for sequencing (Figure S5). The raw sequencing data were processed by using a Python script to quantitatively tally the codons of each compound and then calculate the enrichment fold (post-selection% /pre-selection%), which generated a 2D scatter plot for each selection (enrichment fold *vs.* post-selection sequencing count).[67] The data points with both high enrichment folds and sequencing counts are considered to be potential hits. **CBS** gave low EF after the 1st round of selection but was significantly enriched after the 2nd round (Figure 2l). Low enrichment was observed without photo-crosslinking. Next, we performed a counter-selection against HEK293 cells, a cell line with low CA-12 expression. The *z*-scores of the enriched compounds[68, 69] with the two types of cells were plotted, and it showed the preferential enrichment of **CBS** on A549 cells (Figure 2m). In addition, we also tested the selection method with another model system. Folate receptor (FR) is a cell-surface glycoproteins implicated in many tumors, and folate acid (**FA**) is a small molecule binding to FR with nanomolar affinity.[46] A model library was prepared by mixing a **FA**-DNA conjugate with excess background DNA and selected against HeLa cells overexpressing FR,[46] and qPCR analysis showed significant enrichment of the **FA**-DNA conjugate (Figure S6).

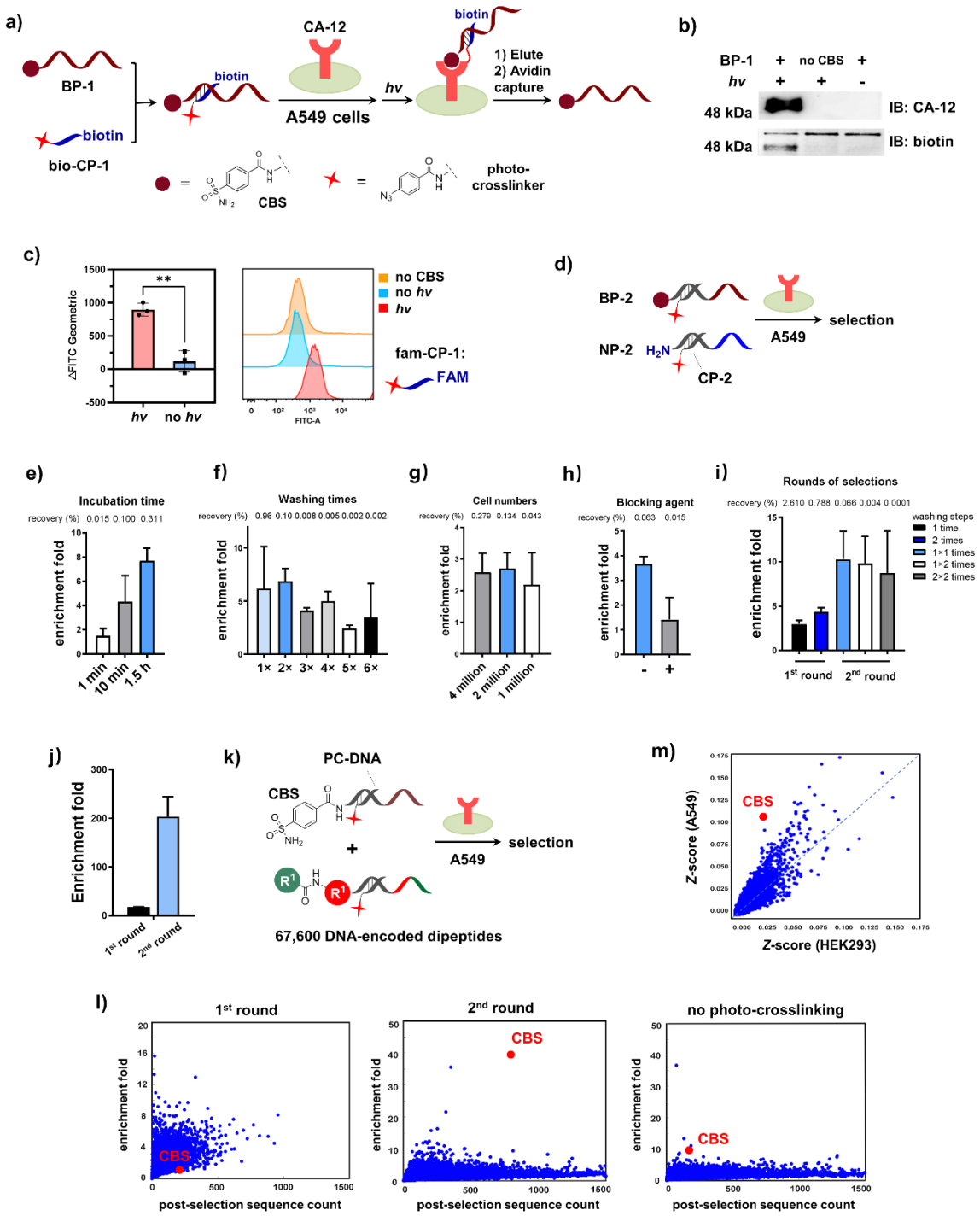

Figure 2: Optimization of the selection conditions. a)-c) BP-1 was hybridized with bio-CP-1/fam-CP-1. The duplexes were used to label CA-12 on A549 cells. After photo-crosslinking, washing, and elution, the labeled proteins or cells were analyzed by using b) Western blot and c) flow cytometry, respectively. In b), no CBS: no CBS in BP-1. Conditions: DNA, 5 μM; cell number, 200 million; buffer, 1x phosphate-buffer saline (PBS); UV, 365 nm, 60 sec. at 0 °C; IB, immunoblotting. d) A model library was selected against A549 cells under different conditions: e) incubation time; f) washing times; g) cell number; h) use of blocking agents (0.1% NaN_3 (g/100 mL) and 1% sheared salmon sperm DNA); and i) round of selections. The selections were analyzed with qPCR to calculate the library recovery and the enrichment fold (EF) of BP-2. j) Selection results of a 1:7,000 model library. k) A model library containing 67,000 dipeptides and CBS-DNA was selected against A549 cells. l) NGS (Illumina) sequencing results shown in scatter plots.[67] *y*-axis: enrichment fold; *x*-axis: post-selection sequencing counts; enrichment fold: post-selection% /post-selection% (naïve library). The data points of CBS-DNA are highlighted in red. m) Comparison of the selections with A549 and HEK293 cells; *x*-axis: normalized *z*-score of the enriched compounds in A549 selection; *y*-axis: normalized *z*-score of the enriched compounds in HEK293 selection.[68] $n = 2$ independent experiments; data are presented as mean values ± s.d.

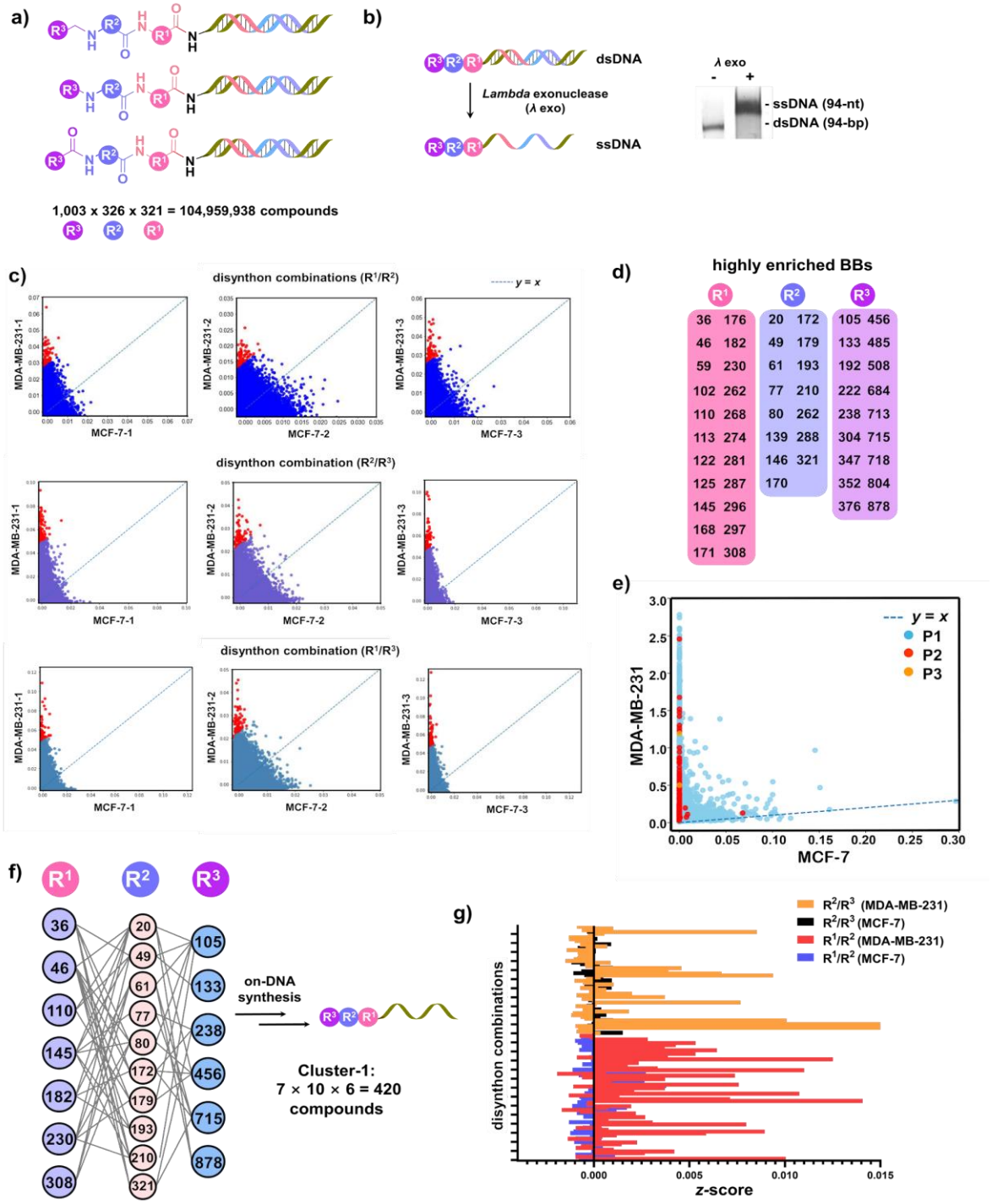

Figure 3: Selections against MDA-MB-231 and MCF7 cells. a) Structure and BB composition of the 104.96-million-compound DEL. b) Converting the dsDNA tag to ssDNA by using *Lambda* exonuclease. [47] **Gel image shows the band shift after the conversion. c) Disynthon (two BBs) *z*-score analysis of the MDA-MB-231 (*y*-axis) and MCF7 (*x*-axis) selection results. Top: R^1/R^2 disynthon; middle: R^2/R^3 disynthon; bottom: R^3/R^1 disynthon; triplicate selections are shown in the same row; the top 100 enriched disynthons are highlighted in red. d) Highly enriched BBs from the top 100 enriched disynthons in c). e) Tri-synthon (all three BBs) *z*-score analysis of the selections. Data points are grouped and highlighted in different colors based on the number of BBs in d). P1 (no or only 1 BB), P2 (2 BBs), and P3 (3 BBs). f) BBs of Cluster-1; the dashed lines indicate disynthon correlations. g) Disynthon *z*-score analysis and comparison of the BBs in Cluster-1 in the selection. See the Supporting Information for details on library synthesis, DNA sequences, BB structures, and data analysis methods. Selected BB structures in c)-g) are also provided in Figure S9-S11.**

Unbiased selections against MDA-MB-231 and MCF7 cells.

Two human breast cancer cell lines were chosen for the selection: MDA-MB-231 and MCF7. MDA-MB-231 is a highly aggressive and poorly differentiated cell line established from metastatic pleural effusion of mammary adenocarcinoma.[70] MCF7 cells were also derived from pleural effusion but are less aggressive.[71] The pair of cell lines are often used as the "positive" and "negative" cells in profiling and library screening studies.[3, 8, 16, 55-59] First, we validated the different invasiveness of the two cell lines (Figure S7). Next, a 104.96-million-compound DEL consisting of three sets of building blocks (BBs) was prepared.[65] The library was synthesized by first coupling 321 amino acids (R^1) to the amine-modified "headpiece" DNA,[65] followed by 326 amino acids (R^2) to form the dipeptides. For the R^3's, 640 carboxylic acids were coupled via amidation, 330 aldehydes were coupled via reductive amination, and 33 sulfonyl chlorides and heterocyclic chlorines were coupled via nucleophilic substitution reactions, respectively (Figure 3a and S8). The library was originally encoded with double-stranded DNA (dsDNA), but it was converted to single-stranded DNA (ssDNA) by *Lambda* exonuclease digestion to be compatible with the selection method (Figure 3b).[47] After hybridizing a 15-nt PC-DNA at the primer-binding site, the ssDNA library was selected against MDA-MB-231 and MCF7 cells, respectively. The eluted binders were PCR-amplified and decoded with NGS as described in Figure 2, and all selections were performed in triplicates.

We expect the selection will enrich combinations of multiple ligands binding to multiple targets on the cell. Therefore, we first used a disynthon-based approach for data analysis. In brief, the sequencing data were processed with a custom script (Supplement File 1)[68, 69, 72, 73] to calculate the *z*-scores of the three disynthon combinations (R^1/R^2, R^2/R^3, and R^1/R^3), respectively, and generated three sets of scatter plots (Figure 3c). The top 100 disynthons (red dots) enriched by MDA-MB-231 cells were selected for each combination (full lists in Figure S9); then, the BBs of these top disynthon combinations were tallied based on frequency, which gave a table of "highly enriched BBs" (cut-off: $R^1 \geq 4$, $R^2 \geq 4$, $R^3 \geq 3$; Figure 3d and full lists in Figure S10). Moreover, the selection data were also processed based on trisynthon combinations. We focused on the data points with post-sequencing counts (C_i^{Post}) ≥ 200 in the MDA-MB-231 selections to minimize statistical under-sampling (Figure 3e). By using the custom script (Supplement File 1), the enriched compounds in the trisynthon analysis were categorized into three groups based on the number of BBs that are present in the "highly enriched BBs" table: P1 (no or only 1 BB; blue dots), P2 (2 BBs in the table; red dots), and P3 (3 BBs in the table; orange dots). Finally, all the BBs in the P2 and P3 groups were tallied based on frequency (cut-off: $R^1 \geq 20$, $R^2 \geq 15$, $R^3 \geq 7$), which eventually generated a cluster containing 420 compounds (**Cluster-1**; Figure 3f and full BB structures in Figure S11). The dashed lines connecting the BBs indicate disynthon correlations. Finally, the *z*-score comparison of the original disynthon combinations also clearly showed that the BBs in **Cluster-1** were indeed more enriched by MDA-MB-231 cells (Figure 3g).

Cluster-1 was resynthesized on a 16-nt ssDNA following the same procedure as in the library synthesis but without the DNA tagging step (Figure 3f). A control DNA without the small molecules was also prepared (**DNA-OH**). The DNAs were hybridized with a FAM-labeled CP (**fam-CP-2**), and the duplexes were used to label MDA-MB-231 and MCF7 cells. Flow cytometry showed **Cluster-1** increased the fluorescence for MDA-MB-231 cells but not for MCF7 cells and **DNA-OH** did not show noticeable labeling of both cell lines (Figure 4a-4b). The specificity of **Cluster-1** was also observed with confocal microscopy (Figure 4c). The binding affinity of **Cluster-1** to MDA-MB-231 cells was estimated to be ~16.5 μM (K_d) by titration (Figure 4d). Next, a deletion strategy was implemented to improve the cluster's binding affinity.[74, 75] The heatmap in Figure 5a shows the *z*-scores of the **Cluster-1** BBs. First, we prepared five "truncated" clusters by deleting 1 to 5 BBs starting from the ones with the lowest *z*-score at the R^3 position (**Cluster-t3-1** to **Cluster-t3-5**; Figure 5b). In addition, three control clusters, in which the R^3 is replaced with two unrelated carboxylic acid or completely omitted, were also prepared (**Cluster-N1**, **N2**, and **N3**; Figure 5b). The clusters were hybridized with **fam-CP-2** and used to label MDA-MB-231 cells. The results showed that deleting the two BBs with the lowest *z*-scores improved the labeling (**Cluster-t3-1** and **Cluster-t3-2;** Figure 5c-5d), whereas deleting more BBs decreased the labeling (**Cluster-t3-3** to **Cluster-t3-5**). The three negative control clusters gave no or little labeling, supporting the importance of the remaining R^3 BBs (Figure 5c-5d). Next, the truncated **Cluster-t3-2** was subjected to similar deletions at the R^1 position. The results showed that the BB with the lowest *z*-score could be removed but not the next one (**Cluster-t1-1**; Figure S12). Finally, **Cluster-t1-1** was subjected to deletions at R^2. Interestingly, removing the three R^2 BBs with the lowest *z*-scores improved the labeling by ~50% (**Cluster-t2-3**; Figure S13a-c), but the next BB appeared to be highly important since its deletion significantly decreased the labeling (**Cluster-t2-4**). Finally, **Cluster-t2-3**, the cluster showing the highest labeling efficiency with 168 compounds, was renamed **Cluster-2** as the 2nd generation cluster (Figure 5e and S13d-e). **Cluster-2**

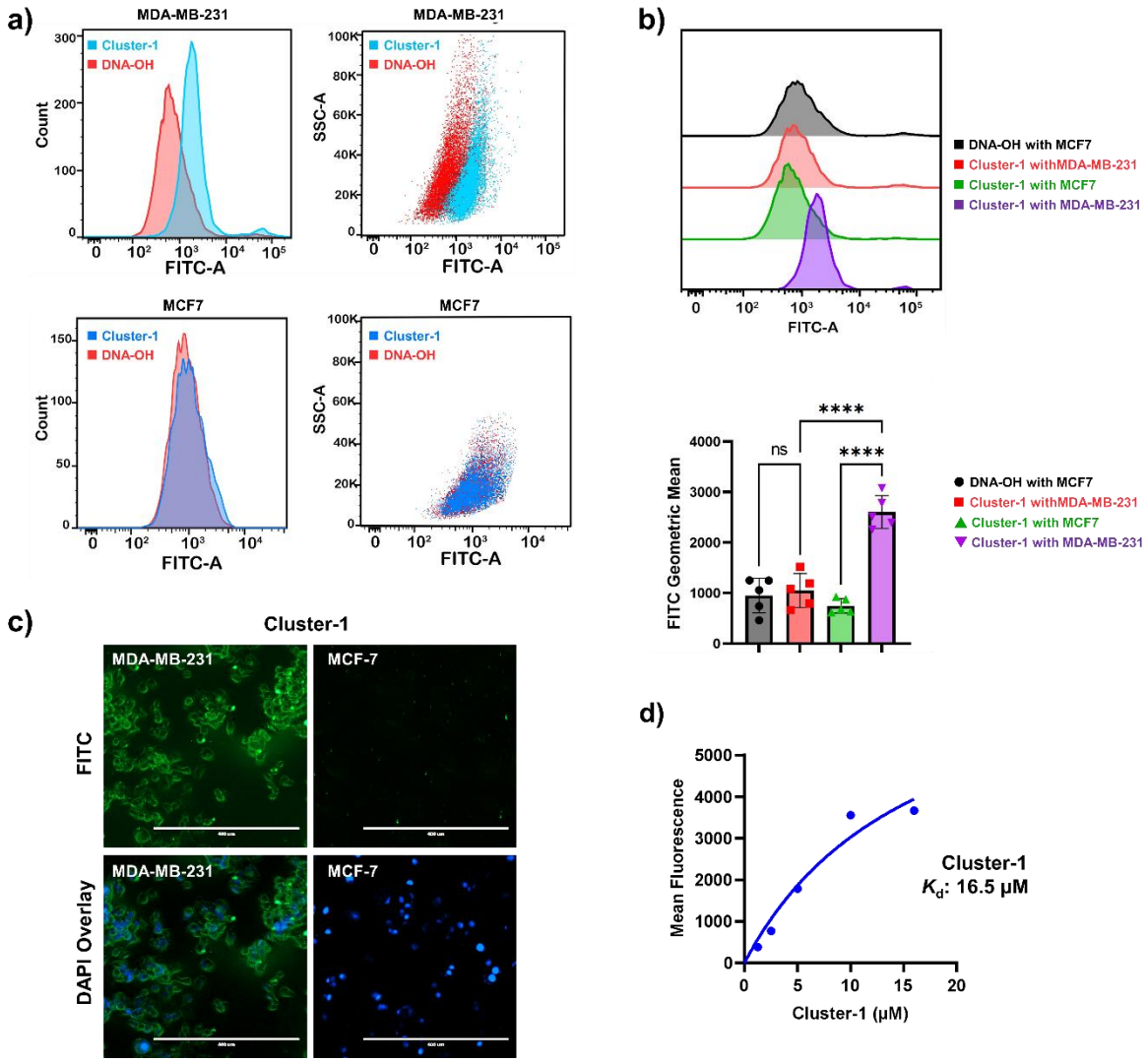

Figure 4: Cluster-1 specifically binds to MDA-MB-231 cells. a)-b) Flow cytometry analysis of Cluster-1 and the control DNA-OH's binding to MDA-MB-231 and MCF7 cells, respectively, with fam-CP-2 and photo-crosslinking. c) Confocal imaging analysis; DAPI: 4′,6-diamidino-2-phenylindole. d) Cluster-1's binding affinity to MDA-MB-231 cells was determined with flow cytometry. The labeling conditions are the same as in Figure 2. *n* = 5 independent experiments; data are presented as mean values ± s.d.

exhibited selective binding for MDA-MB-231 cells (Figure 5f) with a higher binding affinity (K_d: 4.1 μM; Figure 5g).

Collectively, these results demonstrated that the selected "ligand cluster" could be optimized by narrowing down to a smaller number of compounds. However, here, the deletion strategy is simply based on the *z*-scores and did not exhaustively examine all BB combinations. For instance, the deletion stopped when an important BB was identified without further examining its cooperativity with other BBs. Using the deconvolution strategies in traditional combinational chemical libraries, such as positional scanning and deletion synthesis deconvolution,[74, 75] may be able to further optimize the cluster to fewer compounds of higher binding affinity.

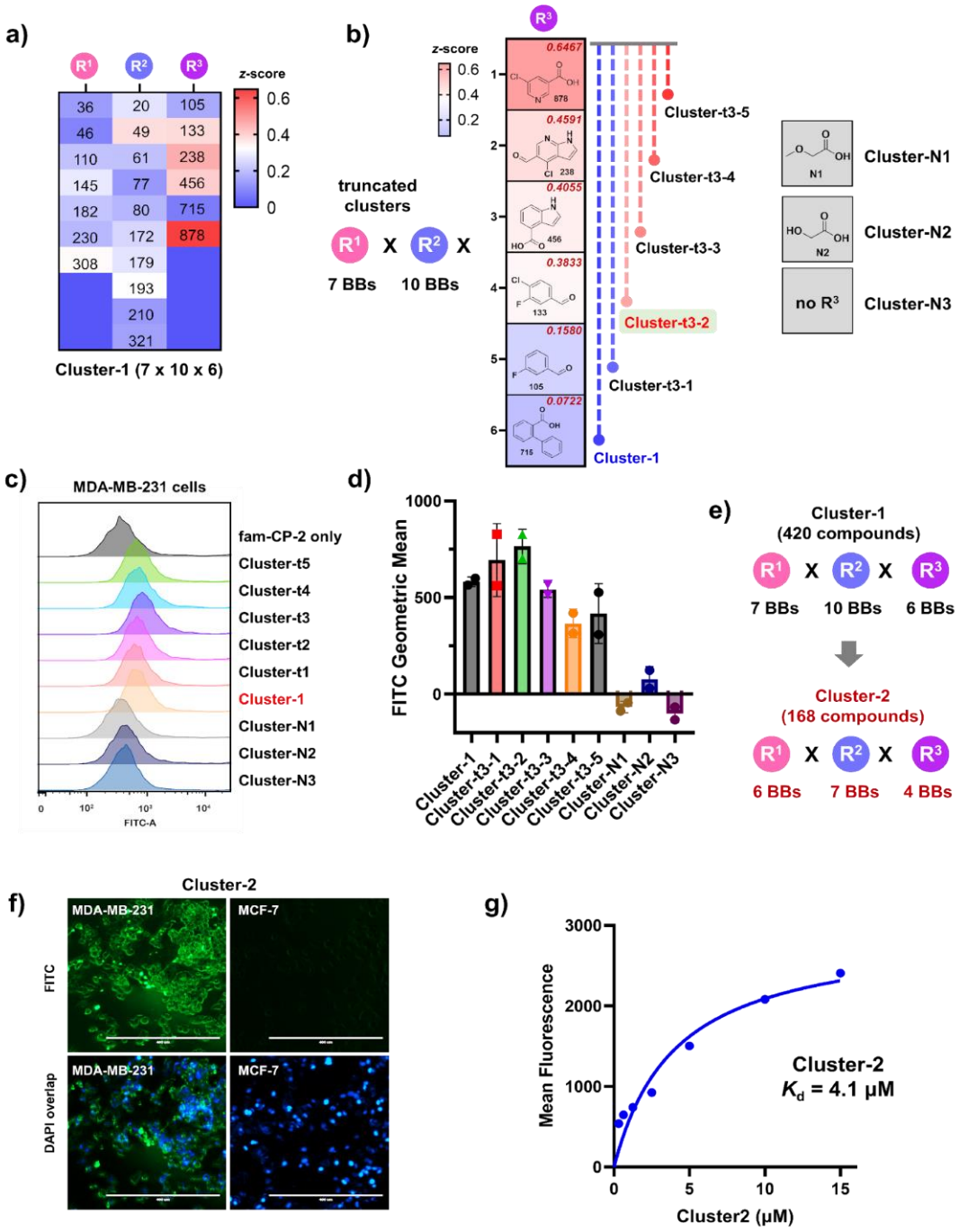

Figure 5: Optimization of the ligand cluster. a) Heatmap showing the *z*-scores of the Cluster-1 BBs; the BB numbers are shown. b) A series of truncated clusters were prepared by deleting the R^3 BBs progressively starting from ones with the lowest *z*-score, while keeping the R^1 and R^2 BBs unchanged; three negative controls (Cluster-N1, N2, and N3) were prepared and tested. c)-d) Flow cytometry analysis of the clusters' binding to MDA-MB-231 cells with fam-CP-2. *n* = 2 independent experiments; data are presented as mean values ± s.d. e) Similar deletion experiments were also performed with the R^1 and R^2 BBs to obtain the optimized cluster (Cluster-2). See Figure S12-S13 for details of the R^1 and R^2 deletion experiments and flow cytometry assays. f) Confocal imaging of Cluster-2's binding to MDA-MB-231 and MCF7 cells. g) Cluster-2's binding affinity to MDA-MB-231 cells was determined by using flow cytometry. The labeling conditions are the same as in Figure 2. Full BB structures of Cluster-1 and Cluster-2 are shown in Figure S11 and S13e.

Applying the ligand clusters for targeted delivery to cancer cells.

Cell-specific ligands can be used in various applications targeting cancer cells, such as drug delivery, gene therapy, detection and isolation of circulating tumor cells, imaging, cancer immunotherapy, targeted protein degradation.[2] Here, we demonstrate that the selected ligand clusters can be used for targeted drug delivery. Doxorubicin (Dox) is a widely used chemotherapy drug that intercalates the DNA double helix to inhibit

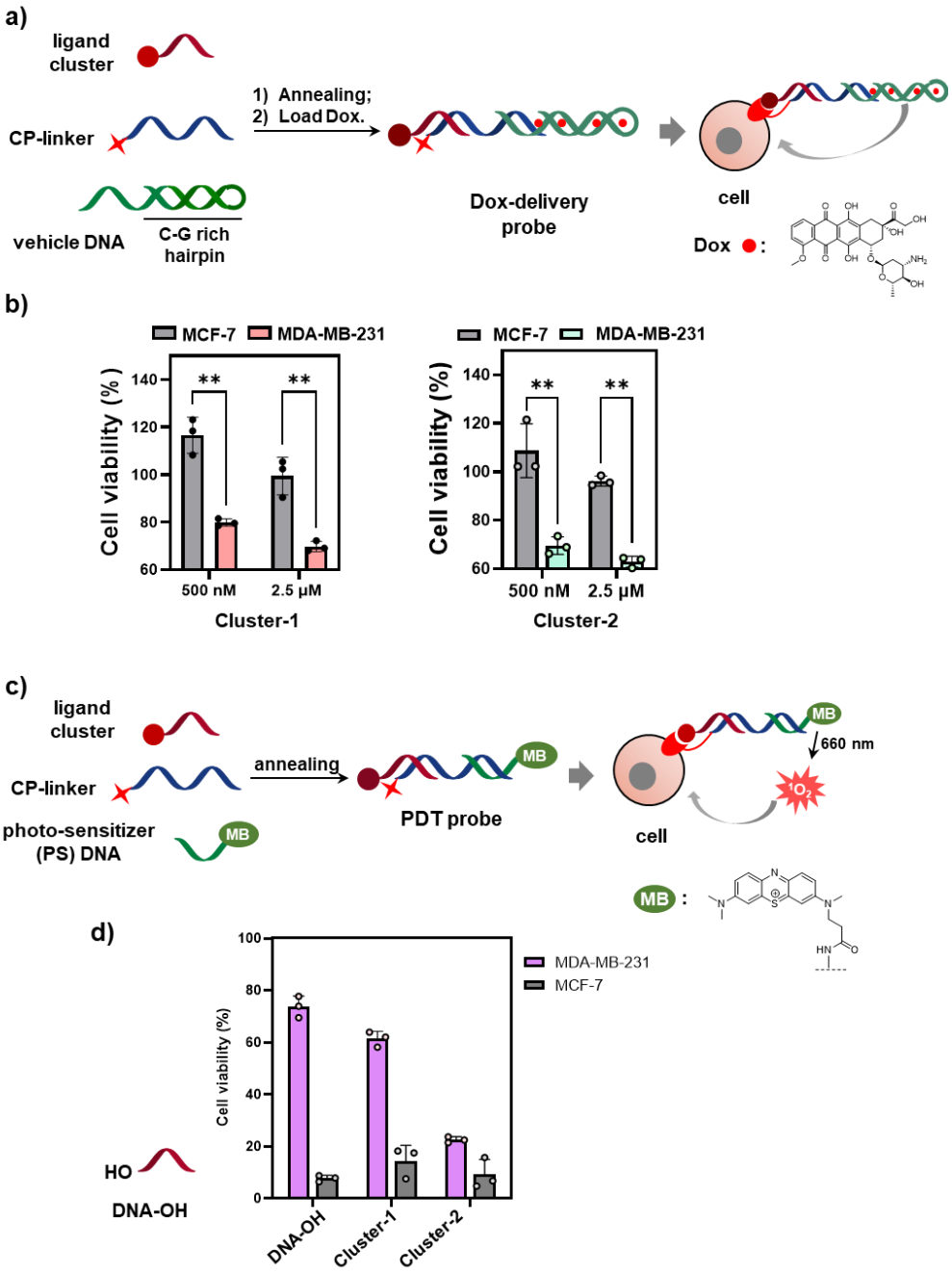

Figure 6: Applying the ligand clusters in targeting cancer cells. a) Delivery of Dox using the probe assembled from DNA-conjugated ligand cluster, a photo-reactive CP-linker DNA, and a C-G rich vehicle DNA. b) Cell viability after Dox delivery using the Cluster-1- (left panel) and Cluster-2- (right panel) based probes. c) PDT treatment of cells using the probes assembled from DNA-conjugated ligand cluster, a photo-reactive CP-linker DNA, and a PS-DNA. d) Cell viability after PDT treatment with a control probe without the small molecules (DNA-OH), Cluster-1 probe, and Cluster-2 probe. $n = 3$ independent experiments; data are presented as mean values ± s.d.

topoisomerase II progression, thereby inhibiting DNA replication; Dox has been used in numerous drug delivery applications.[76] First, we confirmed that MDA-MB-231 and MCF7 cells are both sensitive to Dox treatment with similar IC_{50} values and that the "unloaded" ligand clusters are not toxic to the cells (Figure S14). Next, a 16-nt DNA carrying the ligand cluster, a photo-reactive CP-linker DNA (31-nt), and a C/G-rich vehicle DNA hairpin (21-bp at the duplex region) were hybridized to form a probe (Figure 6a). The probe was incubated with Dox and a 1:10 (probe : Dox) loading ratio appeared to be optimal based on fluorescence quenching (Figure S15). The Dox-loaded probe was incubated with the cells (4 ºC, 1.5 hrs.), briefly UV-irradiated, washed with cold PBS buffer, and then cultured in drug free fresh medium at 37 ºC for up to 24 h before the viability of the treated cells was determined with the CCK-8 assay. As shown in Figure 6b, both **Cluster-1** and **Cluster-2**

showed selectively killing of MDA-MB-231 cells. Photodynamic therapy (PDT) has been used to treat cancer for many years.[77] PDT agents typically comprise a photosensitizer (PS) and a targeting moiety connected through a linker. The targeting moiety binds to the cell, and then light irradiation (600-800 nm) generates cytotoxic reactive oxygen species (ROS) to trigger cell death.[78] The generation of ROS can either be inside the cell after PS internalization or in proximity to the cell surface.[79-81] As shown in Figure 6c, a 16-nt DNA carrying the ligand cluster, a photo-reactive CP-linker DNA (31-nt), and a DNA with a photo-sensitizer methylene blue (MB) hybridized to form the probe. MB is a widely used PS that can generate cytotoxic single oxygen species ($^1O_2$; quantum yield: ~0.52).[78] We first tested the PDT sensitivity of MDA-MB-231 and MCF7 cells with either MB only (Figure S16) or in the presence of a control probe without the small molecules (DNA-OH; Figure 6d); however, MDA-MB-231 cells exhibited strong resistance towards MB-mediated PDT treatment, whereas MCF7 cells were sensitive. Nevertheless, under PDT condition, the probe with the **Cluster-2** ligands was able to sensitize MDA-MB-231 cells and induced significant cell death. Collectively, these experiments have demonstrated the potential of the selected ligand clusters in targeted drug delivery applications against cancer cells.

Identification of the cell surface receptor of a ligand targeting MDA-MB-231 cells.

Target identification of the ligand clusters may be challenging since the compounds may bind to multiple receptors on the cell surface. Therefore, we pursued the individual ligands identified in the selection. As shown in Figure 7a, the normalized *z*-scores of the MDA-MB-231 and MCF7 selections were plotted and seven individual hit compounds were chosen following these criteria: 1) with post-sequencing count (C_i^{Post}) ≥ 200 in the MDA-MB-231 selection; and 2) were reproducible in at least two out of three replicates. A total of 109 compounds fit the criteria, and the top 7 compounds with the highest average *z*-scores were selected for further validation (Figure 7b and full list in Figure S17). These compounds were resynthesized on-DNA, hybridized with **fam-CP-2**, and then used to label the cells. As shown in Figure 7c and S18-S19, they all exhibited selective labeling of MDA-MB-231 over MCF7 cells with micromolar binding affinities. Hit compounds **5** and **6** gave relatively low K_d values (Figure 7c-7d), but the iodobenzene motif in **6** might have stability issue; therefore, we chose compound **5** for target identification. First, we verified that the DNA conjugate of **5** selectively binds to MDA-MB-231 but not MCF7 cells and the binding depends on the small molecule motif, rather than the DNA tag (Figure 7e). Next, the DNA-**5** conjugate was incubated with MDA-MB-231 cells along with three complementary CPs having a biotin group, in which the photo-crosslinker in the CP is either protruding by three nucleobases (*n* = +3), side-by-side (*n* = 0), or recessive (*n* = -3) relatively to the small molecule (Figure 7f and S20). We have previously shown such a multi-probe approach is beneficial in identifying the specific targets of DNA-conjugated small molecules.[62, 82, 83] These affinity probes were incubated with the cells and briefly UV-irradiated, and the cells were lysed by heating. The captured proteins were isolated by using streptavidin beads and analyzed by Western blot. As shown in Figure 7g, the probes captured a few protein bands at ~50-65 kDa, which disappeared in the presence of free small molecule competitor. Furthermore, we performed quantitative proteomic analysis of the capture proteins by using Stable Isotope Labeling with Amino acids in Cell culture (SILAC). In brief, the cells were cultured with either "heavy" or "light" isotopes, respectively; then, the heavy/light cells were incubated with the DNA-based affinity probes either with or without free small molecule competitor for protein capture. The experiments were performed reciprocally in duplicates, and the samples were submitted for proteomic analysis. As shown Figure 7h, the results showed that a protein, α-enolase (ENO1), has been specifically captured in both experiments. α-enolase is a key glycolytic enzyme, but it is also expressed on the tumor cell surface and acts as a plasminogen receptor facilitating extracellular matrix degradation and cancer invasion.[84-86] In addition, its molecular weight (~48 kDa; with DNA: ~54 kDa) matches the band shift in Figure 7g. Indeed, using an anti-ENO1 antibody, it clearly showed that the isolated band at ~55 kD may α-enolase (Figure i). Currently, we are characterizing the binding interactions between compound **5** and α-enolase either *in vitro* or on the cell surface and identifying other possible cell surface receptors. Also, we will perform target identification studies of other individual cell-targeting ligands shown in Figure 7b.

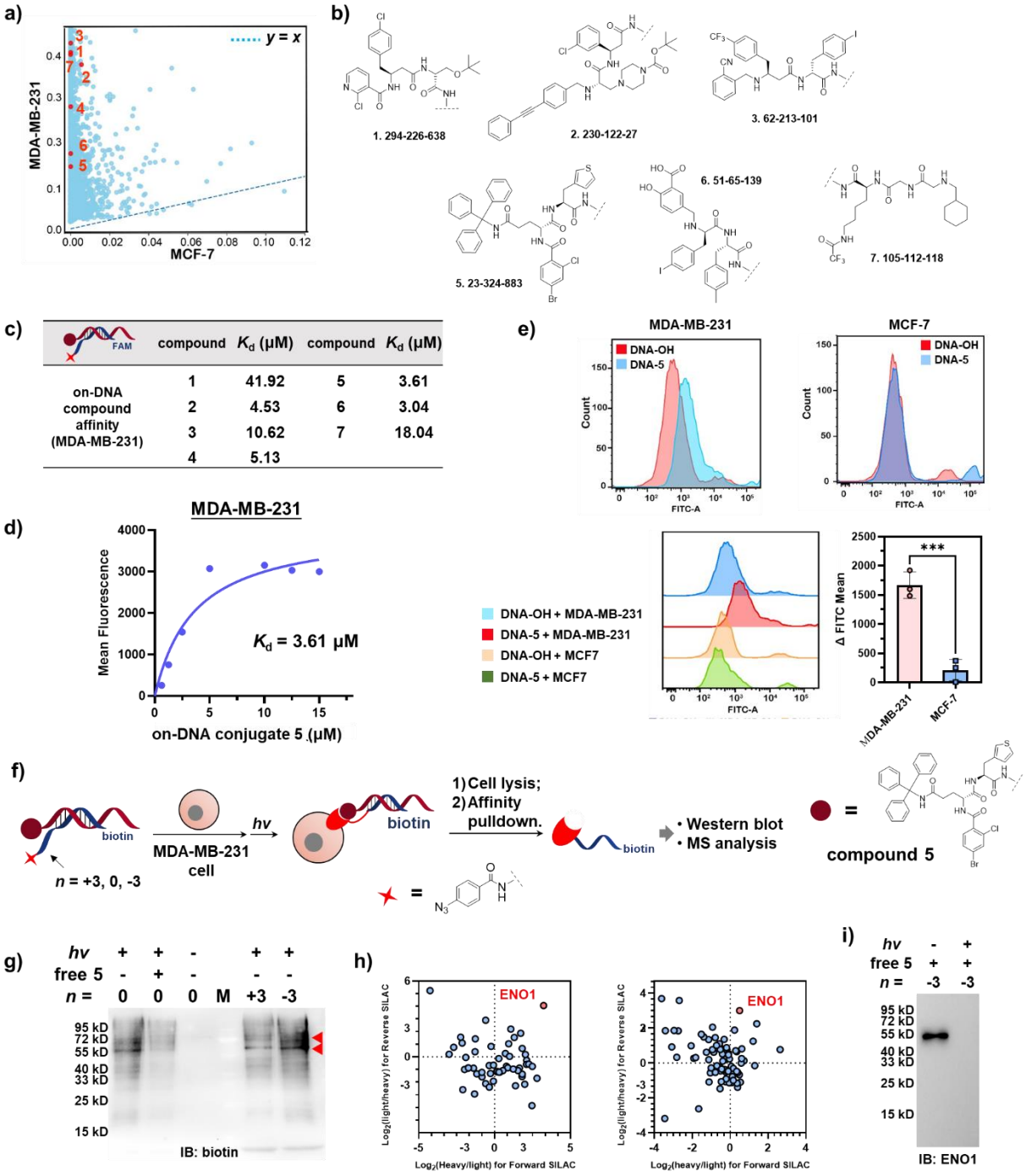

Figure 7: Identification and characterization of individual cell-targeting ligands. a) *z*-score analysis of the selections with MDA-MB-231 and MCF7 cells; the selected hit compounds targeting MDA-MB-231 cells are highlighted. b) Structures of the selected hit compounds. c) Flow cytometry analysis of the "on-DNA" compounds' binding affinities to MDA-MB-231 cells; full data are provided in Figure S19. d) Determining the binding affinity of on-DNA conjugate of 5 to MDA-MB-231 cells. e) Flow cytometry analysis of the DNA-5 conjugate and a control DNA without the small molecule (DNA-OH)'s labeling of MDA-MB-231 and MCF7 cells. $n = 3$ independent experiments; data are presented as mean values ± s.d. f) Compound 5 was conjugated to DNA, hybridized with three different bio-CP DNAs, and used to capture its cell-surface receptor using DNA-templated photo-crosslinking. f) Western blot analysis of the capture protein; M: marker lane; red triangle indicates the possible target band. g) SILAC-based proteomic MS analysis of the captured proteins (anti-biotin). i) Western blot analysis of the capture protein using the anti-ENO1 antibody. See the Supporting Information for experimental details.

Conclusion and discussion

In conclusion, we have shown that DELs can be used to interrogate the cell surface without a predefined target. The selections gave different enrichment profiles on cancer cells of different invasive properties, and the comparative analysis can identify either the ligand clusters or individual compounds specific for the target cells.

We also demonstrated that both the clusters and the individual ligands can be used for targeted cell delivery without the knowledge of their cellular receptors. Furthermore, we performed target identification of one of the identified ligands and identified α-enolase as one possible receptor of the ligand. Typically, after a DEL selection, the identified compounds will be resynthesized off-DNA for hit validation and biological testing. Here, we show that the DNA tag can be leveraged for not only assembling the targeted delivery vehicle but also target identification, which is expected to facilitate the biological and clinical developments of the cell-targeting ligands.

Several points need to be noted. First, unlike typical DEL selections, the abundance of the cell-surface proteins and other biomolecules may be a confounding factor in that the ligands with moderate binding affinities to the targets of high abundance may be more enriched than ligands with higher affinities binding to the targets with low abundance. Thus, the level of compound enrichment is a combination of both binding affinity and the effective target concentration on the cell. However, this may be less of a concern if the objective is to identify the ligands or ligand clusters that are enriched by the entire cell surface, rather than individual proteins. Second, when optimizing the cluster, we did not exhaustively test all BB combinations. **Cluster-2** contains 168 compounds, which is a reasonable number that all of them could be synthesized individually to identify the ones of high binding affinity and/or specificity to MDA-MB-231 cells. It would also be interesting to examine the cooperativities of the different BB combinations. Third, the ligands may bind to non-protein biomolecules on the cell-surface.[87-90] In principle, proteinase K or trypsin could be used to elute protein binders, which can be compared with the general elution conditions (heating or sonication).[91] However, identifying non-protein target is still challenging because it often requires prior knowledge of the putative targets.[88-90, 92] Fourth, the cell-surface selection data pattern is expected to be uniquely different from regular selections because of the target heterogeneity. Here, we used a "BB-centric" approach and relied on the normalized *z*-score as the metric for data analysis. In future, more sophisticated data analysis modalities, such as machine learning,[93-99] is expected to facilitate hit identification or even generate enrichment "fingerprints" for specific cell types or properties.

Currently, we are applying the selection method to a large panel of cancer cells of different types, properties, and cell states, and we expect broad utilities of this method in targeting cancer cells. We will also explore whether the selections would identify previously unknown cancer biomarkers.

Author Information

Corresponding Author

* xiaoyuli@hku.hk; xjlu@simm.ac.cn; ligang@szbl.ac.cn

Author Contributions

The manuscript was written through the contributions of all authors.

Notes

The authors declare no competing financial interests

Acknowledgment

This work was supported by grants from the Shenzhen Bay Laboratory, Shenzhen, China (SZBL2020090501008), NSFC of China (91953119, 91953203, 22377139), Research Grants Council of Hong Kong SAR, China (AoE/P-705/16, 17301118, 17111319, 17303220, 17300321, 17300423, C7005-20G, and C7016-22G), the GuangDong Basic and Applied Basic Research Foundation General Program (2023A1515010711), and Shanghai Municipal Science and Technology Major Project. We acknowledge the support from "Laboratory for Synthetic Chemistry and Chemical Biology" under the Health@InnoHK Program and State Key Laboratory of Synthetic Chemistry by Innovation and Technology Commission, Hong Kong SAR, China.

Statistics and Reproducibility

All experiments were performed at three or more times with independent samples unless otherwise noted. Quantitative values are expressed as the mean values ± SD (standard deviation). The number of replicates and details of statistics are provided in figure legends. All experiments could be reproduced with similar results.

Code availability

The custom Python scripts for sequencing data analysis have been made freely available for downloading at GiHub (https://github.com/uohiuR/OH-FFNNForDEL).

References

1. ReFaey, K. et al. Cancer Mortality Rates Increasing vs Cardiovascular Disease Mortality Decreasing in the World: Future Implications. *Mayo Clin. Proc. Innov. Qual. Outcomes* **5**, 645-653 (2021).
2. Manzari, M.T. et al. Targeted drug delivery strategies for precision medicines. *Nat. Rev. Materi.* **6**, 351-370 (2021).
3. Gray, B.P. & Brown, K.C. Combinatorial peptide libraries: mining for cell-binding peptides. *Chem. Rev.* **114**, 1020-1081 (2014).
4. Fu, Z., Li, S., Han, S., Shi, C. & Zhang, Y. Antibody drug conjugate: the "biological missile" for targeted cancer therapy. *Signal Transduct. Tar. Therapy* **7**, 93 (2022).
5. Lu, R.-M. et al. Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* **27**, 1 (2020).
6. Tabrizi, M.A., Tseng, C.M. & Roskos, L.K. Elimination mechanisms of therapeutic monoclonal antibodies. *Drug Discov. Today* **11**, 81-88 (2006).
7. Jefferis, R. Posttranslational Modifications and the Immunogenicity of Biotherapeutics. *J. Immunol. Res.* **2016**, 5358272 (2016).
8. Liu, R., Li, X., Xiao, W. & Lam, K.S. Tumor-targeting peptides from combinatorial libraries. *Adv. Drug Deliv. Rev.* **110-111**, 13-37 (2017).
9. Pung, H.S., Tye, G.J., Leow, C.H., Ng, W.K. & Lai, N.S. Generation of peptides using phage display technology for cancer diagnosis and molecular imaging. *Mol. Biol. Rep.* **50**, 4653-4664 (2023).
10. Komnatnyy, V.V., Nielsen, T.E. & Qvortrup, K. Bead-based screening in chemical biology and drug discovery. *Chem. Commun.* **54**, 6759-6771 (2018).
11. Arap, W. et al. Steps toward mapping the human vasculature by phage display. *Nat. Med.* **8**, 121-127 (2002).
12. Krag, D.N. et al. Selection of tumor-binding ligands in cancer patients with phage display libraries. *Cancer Res.* **66**, 7724-7733 (2006).
13. Tan, W.H., Donovan, M.J. & Jiang, J.H. Aptamers from Cell-Based Selection for Bioanalytical Applications. *Chem. Rev.* **113**, 2842-2862 (2013).
14. Bing, T., Zhang, N. & Shangguan, D. Cell-SELEX, an Effective Way to the Discovery of Biomarkers and Unexpected Molecular Events. *Adv. Biosyst.* **3**, e1900193 (2019).
15. Hu, Q. et al. DNAzyme-based faithful probing and pulldown to identify candidate biomarkers of low abundance. *Nat. Chem.* **16**, 122–131 (2023).
16. Wu, L. et al. Aptamer-Based Detection of Circulating Targets for Precision Medicine. *Chem. Rev.* **121**, 12035–12105 (2021).
17. Brenner, S. & Lerner, R.A. Encoded combinatorial chemistry. *Proc. Nat. Acad. Sci. USA* **89**, 5381-5383 (1992).
18. Conole, D., J, H.H. & M, J.W. The maturation of DNA encoded libraries: opportunities for new users. *Future Med. Chem.* **13**, 173-191 (2021).
19. Satz, A.L., Kuai, L. & Peng, X. Selections and screenings of DNA-encoded chemical libraries against enzyme and cellular targets. *Bioorg. Med. Chem. Lett.* **39**, 127851 (2021).
20. Song, M. & Hwang, G.T. DNA-Encoded Library Screening as a Core Platform Technology in Drug Discovery. Its Synthetic Method Development and Applications in DEL Synthesis. *J. Med. Chem.* **63**, 6578–6599 (2020).
21. Kunig, V.B.K., Potowski, M., Klika Skopic, M. & Brunschweiger, A. Scanning Protein Surfaces with DNA-Encoded Libraries. *ChemMedChem* **16**, 1048-1062 (2021).
22. Fitzgerald, P.R. & Paegel, B.M. DNA-Encoded Chemistry: Drug Discovery from a Few Good Reactions. *Chem. Rev.* **121**, 7155-7177 (2021).
23. Kodadek, T., Paciaroni, N.G., Balzarini, M. & Dickson, P. Beyond protein binding: recent advances in screening DNA-encoded libraries. *Chem. Commun.* **55**, 13330-13341 (2019).
24. Satz, A.L. et al. DNA-encoded chemical libraries. *Nat. Rev. Methods Primers* **2**, 3 (2022).

25. Plais, L. & Scheuermann, J. Macrocyclic DNA-encoded chemical libraries: a historical perspective. *RSC Chem. Biol.* **3**, 7-17 (2022).
26. Neri, D. & Lerner, R.A. DNA-Encoded Chemical Libraries: A Selection System Based On Endowing Organic Compounds with Amplifiable Information. *Annu. Rev. Biochem.* **87**, 479-502 (2018).
27. Reddavide, F.V., Thompson, M., Mannocci, L. & Zhang, Y.X. DNA-Encoded Fragment Libraries: Dynamic Assembly, Single-Molecule Detection, and High-Throughput Hit Validation. *Aldrichim Acta* **52**, 63-74 (2019).
28. Huang, Y.R., Li, Y.Z. & Li, X.Y. Strategies for developing DNA-encoded libraries beyond binding assays. *Nat. Chem.* **14**, 129-140 (2022).
29. Dockerill, M. & Winssinger, N. DNA-Encoded Libraries: Towards Harnessing their Full Power with Darwinian Evolution. *Angew. Chem. Int. Ed.* **62**, e202215542.
30. Yuen, L.H. & Franzini, R.M. Achievements, Challenges, and Opportunities in DNA-Encoded Library Research: An Academic Point of View. *ChemBioChem* **18**, 829-836 (2017).
31. Dixit, A., Barhoosh, H. & Paegel, B.M. Translating the Genome into Drugs. *Acc. Chem. Res.* **56**, 489-499 (2023).
32. Matsuo, B., Granados, A., Levitre, G. & Molander, G.A. Photochemical Methods Applied to DNA Encoded Library (DEL) Synthesis. *Acc. Chem. Res.* **56**, 385-401 (2023).
33. Sunkari, Y.K., Siripuram, V.K., Nguyen, T.L. & Flajolet, M. High-power screening (HPS) empowered by DNA-encoded libraries. *Trends Pharmacol. Sci.* **43**, 4-15 (2022).
34. Peterson, A.A. & Liu, D.R. Small-molecule discovery through DNA-encoded libraries. *Nat. Rev. Drug Discov.* **22**, 699–722 (2023).
35. Flood, D.T., Kingston, C., Vantourout, J.C., Dawson, P.E. & Baran, P.S. DNA Encoded Libraries: A Visitor's Guide. *Isr. J. Chem.* **60**, 268-280 (2020).
36. Ottl, J., Leder, L., Schaefer, J.V. & Dumelin, C.E. Encoded Library Technologies as Integrated Lead Finding Platforms for Drug Discovery. *Molecules* **24**, 1629 (2019).
37. Goodnow, R.A., Jr., Dumelin, C.E. & Keefe, A.D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **16**, 131-147 (2017).
38. Götte, K., Chines, S. & Brunschweiger, A. Reaction development for DNA-encoded library technology: From evolution to revolution? *Tetrahedron Lett.* **61**, 151889 (2020).
39. Wu, Z. et al. Cell-Based Selection Expands the Utility of DNA-Encoded Small-Molecule Library Technology to Cell Surface Drug Targets: Identification of Novel Antagonists of the NK3 Tachykinin Receptor. *ACS Comb. Sci.* **17**, 722-731 (2015).
40. Denton, K.E. & Krusemark, C.J. Crosslinking of DNA-linked ligands to target proteins for enrichment from DNA-encoded libraries. *Medchemcomm* **7**, 2020-2027 (2016).
41. Cai, B. et al. Selection of DNA-Encoded Libraries to Protein Targets within and on Living Cells. *J. Am. Chem. Soc.* **141**, 17057-17061 (2019).
42. Cai, B., Mhetre, A.B. & Krusemark, C.J. Selection methods for proximity-dependent enrichment of ligands from DNA-encoded libraries using enzymatic fusion proteins. *Chem. Sci.* **14**, 245-250 (2023).
43. Cai, B., El Daibani, A., Bai, Y., Che, T. & Krusemark, C.J. Direct Selection of DNA-Encoded Libraries for Biased Agonists of GPCRs on Live Cells. *JACS Au* **3**, 1076-1088 (2023).
44. Oehler, S. et al. Affinity Selections of DNA-Encoded Chemical Libraries on Carbonic Anhydrase IX-Expressing Tumor Cells Reveal a Dependence on Ligand Valence. *Chem. Eur. J.* **27**, 8985-8993 (2021).
45. Petersen, L.K. et al. Screening of DNA-Encoded Small Molecule Libraries inside a Living Cell. *J. Am. Chem. Soc.* **143**, 2751-2756 (2021).
46. Huang, Y. et al. Selection of DNA-encoded chemical libraries against endogenous membrane proteins on live cells. *Nat. Chem.* **13**, 77-88 (2021).
47. Gui, Y. et al. Converting Double-Stranded DNA-Encoded Libraries (DELs) to Single-Stranded Libraries for More Versatile Selections. *ACS Omega* **7**, 11491-11500 (2022).
48. Svensen, N., Diaz-Mochon, J.J. & Bradley, M. Decoding a PNA encoded peptide library by PCR: the discovery of new cell surface receptor ligands. *Chem. Biol.* **18**, 1284-1289 (2011).
49. Svensen, N., Diaz-Mochon, J.J. & Bradley, M. Encoded peptide libraries and the discovery of new cell binding ligands. *Chem. Commun.* **47**, 7638-7640 (2011).
50. Mendes, K.R. et al. High-throughput Identification of DNA-Encoded IgG Ligands that Distinguish Active and Latent Mycobacterium tuberculosis Infections. *ACS Chem. Biol.* **12**, 234-243 (2017).
51. Barry, M.A., Dower, W.J. & Johnston, S.A. Toward cell-targeting gene therapy vectors: selection of cell-binding peptides from random peptide-presenting phage libraries. *Nat. Med.* **2**, 299-305 (1996).

52. Lowman, H.B. Bacteriophage display and discovery of peptide leads for drug development. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 401-424 (1997).
53. Kodadek, T., Paciaroni, N.G., Balzarini, M. & Dickson, P. Beyond protein binding: recent advances in screening DNA-encoded libraries. *Chem. Commun.* **55**, 13330-13341 (2019).
54. Zhao, P. et al. Selection of DNA-encoded small molecule libraries against unmodified and non-immobilized protein targets. *Angew. Chem. Int. Ed.* **53**, 10056-10059 (2014).
55. Nomura, D.K. et al. Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell* **140**, 49-61 (2010).
56. Jessani, N., Liu, Y., Humphrey, M. & Cravatt, B.F. Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *Proc. Nat. Acad. Sci. USA* **99**, 10335-10340 (2002).
57. Aina, O.H. et al. From combinatorial chemistry to cancer-targeting peptides. *Mol. Pharm.* **4**, 631-651 (2007).
58. DeRosa, M.C. et al. In vitro selection of aptamers and their applications. *Nat. Rev. Methods Primers* **3** (2023).
59. Xie, S. et al. Aptamer-Based Targeted Delivery of Functional Nucleic Acids. *J. Am. Chem. Soc.* **145**, 7677-7691 (2023).
60. Ma, H. et al. PAC-FragmentDEL – photoactivated covalent capture of DNA-encoded fragments for hit discovery. *RSC Med. Chem.* **13**, 1341-1349 (2022).
61. Shi, B., Deng, Y., Zhao, P. & Li, X. Selecting a DNA-Encoded Chemical Library against Non-immobilized Proteins Using a "Ligate-Cross-Link-Purify" Strategy. *Bioconjug. Chem.* **28**, 2293-2301 (2017).
62. Zhang, J. et al. Identification of Histone deacetylase (HDAC)-Associated Proteins with DNA-Programmed Affinity Labeling. *Angew. Chem. Int. Ed.* **59**, 17525 (2020).
63. Vullo, D. et al. Carbonic anhydrase inhibitors. Inhibition of the transmembrane isozyme XII with sulfonamides-a new target for the design of antitumor and antiglaucoma drugs? *Bioorg. Med. Chem. Lett.* **15**, 963-969 (2005).
64. Huang, Y., Deng, Y., Zhang, J., Meng, L. & Li, X. Direct ligand screening against membrane proteins on live cells enabled by DNA-programmed affinity labelling. *Chem. Commun.* (2021).
65. Clark, M.A. et al. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **5**, 647-654 (2009).
66. Deng, Y. et al. Selection of DNA-encoded Dynamic Chemical Libraries for Direct Inhibitor Discovery. *Angew. Chem. Int. Ed.* **59**, 14965-14972 (2020).
67. Zhou, Y. et al. DNA-Encoded Dynamic Chemical Library and Its Applications in Ligand Discovery. *J. Am. Chem. Soc.* **140**, 15859-15867 (2018).
68. Faver, J.C. et al. Quantitative Comparison of Enrichment from DNA-Encoded Chemical Library Selections. *ACS Comb. Sci.* **21**, 75-82 (2019).
69. Dawadi, S. et al. Discovery of potent thrombin inhibitors from a protease-focused DNA-encoded chemical library. *Proc. Nat. Acad. Sci. USA* **117**, 16782-16789 (2020).
70. Chavez, K.J., Garimella, S.V. & Lipkowitz, S. Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer. *Breast Dis.* **32**, 35-48 (2010).
71. Lee, A.V., Oesterreich, S. & Davidson, N.E. MCF-7 cells--changing the course of breast cancer research and care for 45 years. *J. Natl. Cancer Inst.* **107** (2015).
72. Taylor, D.M. et al. Identifying Oxacillinase-48 Carbapenemase Inhibitors Using DNA-Encoded Chemical Libraries. *ACS Infectious Diseases* **6**, 1214-1227 (2020).
73. Chamakuri, S. et al. DNA-encoded chemistry technology yields expedient access to SARS-CoV-2 M(pro) inhibitors. *Proc. Nat. Acad. Sci. USA* **118** (2021).
74. Boger, D.L., Lee, J.K., Goldberg, J. & Jin, Q. Two Comparisons of the Performance of Positional Scanning and Deletion Synthesis for the Identification of Active Constituents in Mixture Combinatorial Libraries. *J. Org. Chem.* **65**, 1467-1474 (2000).
75. Boger, D.L., Chai, W. & Jin, Q. Multistep Convergent Solution-Phase Combinatorial Synthesis and Deletion Synthesis Deconvolution. *J. Am. Chem. Soc.* **120**, 7220-7225 (1998).
76. Zhao, N., Woodle, M.C. & Mixson, A.J. Advances in delivery systems for doxorubicin. *J. Nanomed. Nanotechnol.* **9** (2018).
77. Dolmans, D.E.J.G.J., Fukumura, D. & Jain, R.K. Photodynamic therapy for cancer. *Nat. Rev. Cancer* **3**, 380-387 (2003).

78. Pham, T.C., Nguyen, V.-N., Choi, Y., Lee, S. & Yoon, J. Recent Strategies to Develop Innovative Photosensitizers for Enhanced Photodynamic Therapy. *Chem. Rev.* **121**, 13454-13619 (2021).
79. Ma, W. et al. A Cell Membrane-Targeting Self-Delivery Chimeric Peptide for Enhanced Photodynamic Therapy and In Situ Therapeutic Feedback. *Adv. Healthcare Mater.* **9**, 1901100 (2020).
80. Zhang, W. et al. Amphiphilic Tetraphenylethene-Based Pyridinium Salt for Selective Cell-Membrane Imaging and Room-Light-Induced Special Reactive Oxygen Species Generation. *ACS App. Mater. Interf.* **11**, 10567-10577 (2019).
81. Bu, Y. et al. A NIR-I light-responsive superoxide radical generator with cancer cell membrane targeting ability for enhanced imaging-guided photodynamic therapy. *Chem. Sci.* **11**, 10279-10286 (2020).
82. Li, G., Liu, Y., Chen, L., Wu, S. & Li, X. Photoaffinity Labeling of Small-Molecule-Binding Proteins by DNA-Templated Chemistry. *Angew. Chem. Int. Ed.* **52**, 9544-9549 (2013).
83. Wang, D.Y. et al. Target Identification of Kinase Inhibitor Alisertib (MLN8237) by Using DNA-Programmed Affinity Labeling. *Chem. Eur. J.* **23**, 10906-10914 (2017).
84. Díaz-Ramos, A., Roig-Borrellas, A., García-Melero, A. & López-Alemany, R. α-Enolase, a multifunctional protein: its role on pathophysiological situations. *J. Biomed. Biotechnol.* **2012**, 156795 (2012).
85. Almaguel, F.A., Sanchez, T.W., Ortiz-Hernandez, G.L. & Casiano, C.A. Alpha-Enolase: Emerging Tumor-Associated Antigen, Cancer Biomarker, and Oncotherapeutic Target. *Front. Genetics* **11** (2021).
86. Hsiao, K.-C. et al. Surface α-Enolase Promotes Extracellular Matrix Degradation and Tumor Metastasis and Represents a New Therapeutic Target. *PLOS ONE* **8**, e69354 (2013).
87. Thapa, N. et al. Discovery of a phosphatidylserine-recognizing peptide and its utility in molecular imaging of tumour apoptosis. *J. Cell Mol. Med.* **12**, 1649-1660 (2008).
88. Peletskaya, E.N., Glinsky, V.V., Glinsky, G.V., Deutscher, S.L. & Quinn, T.P. Characterization of peptides that bind the tumor-associated Thomsen-Friedenreich antigen selected from bacteriophage display libraries. *J. Mol. Biol.* **270**, 374-384 (1997).
89. Laumonier, C. et al. A new peptidic vector for molecular imaging of apoptosis, identified by phage display technology. *J. Biomol. Screen.* **11**, 537-545 (2006).
90. Shao, R., Xiong, C., Wen, X., Gelovani, J.G. & Li, C. Targeting phosphatidylserine on apoptotic cells with phages and peptides selected from a bacteriophage display library. *Mol. Imaging* **6**, 417-426 (2007).
91. Shangguan, D. et al. Aptamers evolved from live cells as effective molecular probes for cancer study. *Proc. Nat. Acad. Sci. USA* **103**, 11838-11843 (2006).
92. Thapa, N. et al. Discovery of a phosphatidylserine-recognizing peptide and its utility in molecular imaging of tumour apoptosis. *J. Cell. Mol. Med.* **12**, 1649-1660 (2008).
93. Komar, P. & Kalinic, M. Denoising DNA Encoded Library Screens with Sparse Learning. *ACS Comb. Sci.* **22**, 410-421 (2020).
94. Lim, K.S. et al. Machine Learning on DNA-Encoded Library Count Data Using an Uncertainty-Aware Probabilistic Loss Function. *J. Chem. Inf. Model.* **62**, 2316-2331 (2022).
95. Hou, R., Xie, C., Gui, Y., Li, G. & Li, X. Machine-Learning-Based Data Analysis Method for Cell-Based Selection of DNA-Encoded Libraries. *ACS Omega* **8**, 19057-19071 (2023).
96. McCloskey, K. et al. Machine Learning on DNA-Encoded Libraries: A New Paradigm for Hit Finding. *J. Med. Chem.* **63**, 8857-8866 (2020).
97. Ahmad, S. et al. Discovery of a First-in-Class Small-Molecule Ligand for WDR91 Using DNA-Encoded Chemical Library Selection Followed by Machine Learning. *J. Med. Chem.* **66**, 16051-16061 (2023).
98. Li, A.S.M. et al. Discovery of Nanomolar DCAF1 Small Molecule Ligands. *J. Med. Chem.* **66**, 5041-5060 (2023).
99. Montoya, A.L. et al. Combining pharmacophore models derived from DNA-encoded chemical libraries with structure-based exploration to predict Tankyrase 1 inhibitors. *Eur. J. Med. Chem.* **246**, 114980 (2022).