nature chemistry

ARTICLES

https://doi.org/10.1038/s41557-020-00605-x

Selection of DNA-encoded chemical libraries against endogenous membrane proteins on live cells

Yiran Huang[1], Ling Meng[1], Qigui Nie[2], Yu Zhou[1], Langdong Chen[3], Shilian Yang[2], Yi Man Eva Fung[1], Xiaomeng Li[1], Cen Huang[2], Yan Cao[3]✉, Yizhou Li[2,4]✉ and Xiaoyu Li[1,5]✉

Membrane proteins on the cell surface perform a myriad of biological functions; however, ligand discovery for membrane proteins is highly challenging, because a natural cellular environment is often necessary to maintain protein structure and function. DNA-encoded chemical libraries (DELs) have emerged as a powerful technology for ligand discovery, but they are mainly limited to purified proteins. Here we report a method that can specifically label membrane proteins with a DNA tag, and thereby enable target-specific DEL selections against endogenous membrane proteins on live cells without overexpression or any other genetic manipulation. We demonstrate the generality and performance of this method by screening a 30.42-million-compound DEL against the folate receptor, carbonic anhydrase 12 and the epidermal growth factor receptor on live cells, and identify and validate a series of novel ligands for these targets. Given the high therapeutic significance of membrane proteins and their intractability to traditional high-throughput screening approaches, this method has the potential to facilitate membrane-protein-based drug discovery by harnessing the power of DEL.

Membrane proteins perform a myriad of biological functions in cells[1]. Not surprisingly, numerous human diseases are associated with aberrant membrane protein functions; indeed, membrane proteins account for over 60% of the targets of all approved small-molecule drugs[2]. Probably the most notable example is the G-protein coupled receptor protein superfamily; as the largest class of cell-surface receptors, they are the targets of ~34% of clinical drugs[3]. Recently, the immuno-checkpoint membrane proteins have been intensively pursued in cancer immunotherapy[4]. However, ligand discovery against membrane proteins is highly challenging. Unlike cytosolic proteins, membrane proteins are situated in the hydrophobic lipid bilayer of the cell membrane, which makes in vitro biochemical techniques, such as protein expression and purification, extremely difficult[2,5]. Outside their native contexts, membrane proteins may lose important biological features, such as post-translational modifications, cofactor binding and complex formation. Therefore, the development of ligand discovery methods compatible with natural cellular conditions is highly desirable for membrane proteins.

Originally proposed by Brenner and Lerner in 1992[6,7], DNA-encoded chemical library (DEL) has become a powerful ligand discovery technology in biomedical research[8–14]. In a DEL, each compound is conjugated with a unique DNA tag that encodes its chemical structure. All the library compounds are mixed together and selected against the target simultaneously. The selected binders are decoded with a polymerase chain reaction (PCR) amplification and DNA sequencing to read the barcodes (Fig. 1a). DELs can contain many billions or even trillions of compounds[15], and the library selection can be conducted in just a few hours. Today, DELs are widely adopted by the pharmaceutical industry[16,17].

DELs have been selected against the soluble domains of membrane proteins[18–22] and membrane proteins stabilized with detergent[23] or nanodiscs[24]; however, the target scope of DELs is mostly limited to purified proteins[9,12]. Previously, we and others developed DEL methods compatible with soluble proteins[25–32]: GlaxoSmithKline performed DEL selection against the NK3 receptor on live cells[33], the Bradley group selected peptide nucleic acid-encoded libraries against chemokine receptors[34,35], the Song group selected a glycan library against whole bacteria[36] and the Krusemark group reported DEL selection against the opioid receptor on live cells[37]. However, these methods rely on recombinant protein expression to obtain a high target abundance and/or the target needs to be fused to large tags. Target-specific DEL selection against endogenous membrane proteins on live cells has yet to be achieved. Here we report a method that enables the cell-based DEL selection against membrane proteins; this method is target specific, does not require protein overexpression or any other genetic manipulation and is compatible with endogenous proteins. Using a ligand-directed affinity labelling method, we show that membrane proteins can be labelled with a DNA tag, which serves as a homing beacon to guide DEL selection. To demonstrate the performance of this method, a 30.42-million-compound DEL was selected against three membrane proteins (folate receptor (FR), carbonic anhydrase 12 (CA-12) and epidermal growth factor receptor (EGFR)), and a series of novel ligands were identified for these targets.

To realize DEL selection on live cells, two important issues need to be addressed: target specificity and target concentration. The first issue is obvious, as the cell surface contains many other proteins and biomolecules. The second issue is also important: typically, DEL selection is performed with the library present at a minute scale

[1]Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Hong Kong SAR, China. [2]Chongqing Key Laboratory of Natural Product Synthesis and Drug Research, School of Pharmaceutical Sciences, Chongqing University, Chongqing, China. [3]School of Pharmacy, Second Military Medical University, Shanghai, China. [4]Key Laboratory of Biorheological Science and Technology, Ministry of Education, College of Bioengineering, Chongqing University, Chongqing, China. [5]Laboratory for Synthetic Chemistry and Chemical Biology of Health@InnoHK, Hong Kong SAR, China. ✉e-mail: caoyan@smmu.edu.cn; yizhouli@cqu.edu.cn; xiaoyuli@hku.hk

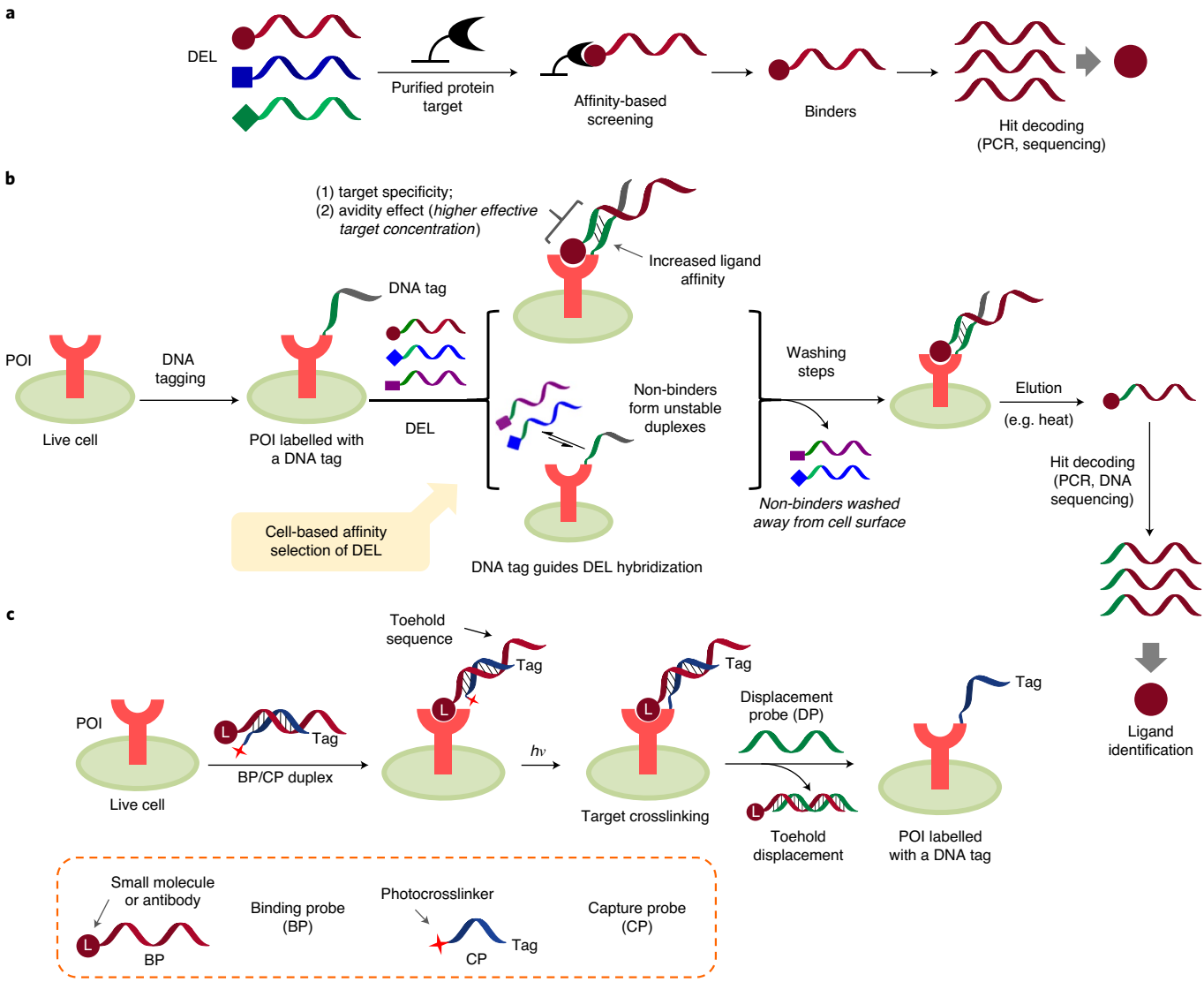


Fig. 1 | The proposed cell-based DEL selection strategy. a, DELs are typically selected against purified and immobilized proteins. **b**, The proposed cell-based selection method. The POI is labelled with a DNA tag, which guides DEL hybridization to achieve target specificity and increases the effective target concentration. After washing away the non-binders from the cell surface, the hit compounds were eluted (for example, by heating to denature the cells) and decoded with PCR amplification and DNA sequencing. **c**, Method to label membrane proteins with a DNA tag. L, a known ligand; tag, fluorophore, biotin and so on. The BP/CP duplex is preformed and then added to the cell. After the BP/CP-mediated affinity labelling, the BP is removed via toehold displacement.

(femtomoles–picomoles) and a relatively high target concentration (low-to-mid micromolar) is necessary to drive the binding equilibrium. However, the abundance of most membrane proteins is very low; without overexpression[33], it would be difficult to identify the ligands with a dissociation constant (K_d) higher than or close to the target concentration. We reasoned that, if the protein of interest (POI) could be labelled with a DNA tag, it might serve as a homing beacon to guide the library selection. On the one hand, the DNA tag provides target specificity in the selection by guiding library hybridization; on the other hand, it boosts the ligand affinity by avidity, and thereby drives the binding equilibrium and retains the binders on the cell in the affinity-based selection (Fig. 1b). The non-binders form relatively unstable DNA duplexes and could be washed away. After the elution step (heat denaturation, selective elution with free ligand and so on), the binders could be identified through the typical DEL decoding procedure using PCR amplification and DNA sequencing[38].

Many methods have been developed to covalently graft DNA tags to cells, such as the seminal work by Bertozzi, Francis and co-workers using the Staudinger reaction and strain-promoted alkyne–azide cycloaddition[39,40], the amidation reaction[41,42], oxidative cleavage[43,44] and so on, which are employed in many biological applications[45,46]. Recently, elegant methods for target-specific protein labelling with DNA aptamers were reported by Famulok[47], Tan[48–50], Bertozzi[51], Gothelf[52] and their respective co-workers. In addition, the Hamachi group developed non-DNA-based ligand-directed labelling methods for membrane proteins[53]. Previously, we reported a protein labelling method named DPAL (DNA-programmed affinity labelling)[54], which has been used to identify the targets for small molecules, nucleic acids and aptamers by us[54–56] and other groups[57,58]. With a similar concept, the Gothelf group developed an elegant DNA-templated protein conjugation method for site-specific antibody–DNA conjugation[52,59,60] and the Tan group used aptamers as templates to direct protein conjugation[48–50]. Here we propose to

use DPAL to deliver DNA tags to membrane proteins. As shown in Fig. 1c, a known ligand of the POI is conjugated with a DNA strand as the binding probe (BP). The BP forms a duplex with a capture probe (CP) that bears a photocrosslinker. After the BP/CP duplex engages the POI, ultraviolet light irradiation triggers the target crosslinking. Next, a displacement probe (DP) removes the BP through toehold displacement. The DNA tag originally from the CP could be used to guide the subsequent DEL selection.

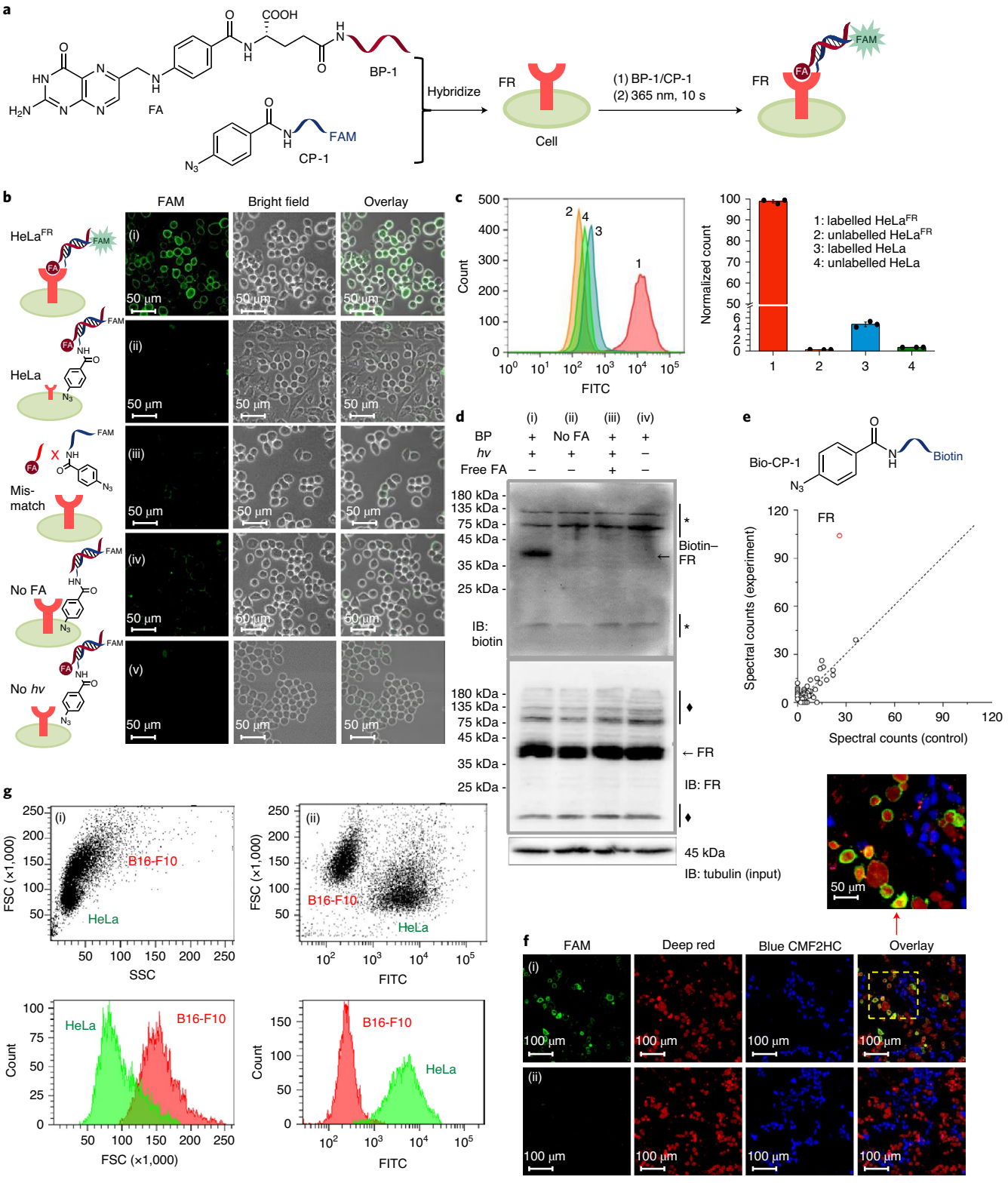
Results

Membrane proteins on live cells can be specifically labelled with DNA tags using DPAL. As an initial validation, we used the FR/folic acid (FA) system (Fig. 2a). FR is a membrane protein implicated in many tumours, and FA binds to FR at nanomolar affinity. HeLa cells cultured in FA-deficient medium with upregulated FR expression were used for labelling (HeLa^{FR}; Supplementary Fig. 1)[61], and regular HeLa cells were used as a control. First, FA was coupled with a 16-nucleotide DNA as the BP, and a complementary DNA with a phenylazide photocrosslinker and a fluorescein (FAM) group was prepared as the CP (BP-1/CP-1; Fig. 2a). The BP-1/CP-1 duplex was incubated with the cells before brief ultraviolet irradiation (365 nm, 10 s) to trigger crosslinking while minimizing cell damage. The cells were washed and visualized under a microscope. As shown in Fig. 2b, much stronger fluorescence was observed with the HeLa^{FR} cells (Fig. 2b(i)) than the control cells (Fig. 2b(ii)). As expected, the fluorescence mostly surrounded the cell membrane. The negative controls (mismatched CP (Fig. 2b(iii)), no FA on BP (Fig. 2b(iv)) and no ultraviolet irradiation (Fig. 2b(v))) showed low fluorescence, which proves that the labelling was specific and dependent on DNA-mediated photocrosslinking. Flow cytometry also showed that the HeLa^{FR} cells were more efficiently labelled (Fig. 2c). The number of DNA tags on a cell was quantified using fluorescent beads as a calibration standard, as in previous reports (Supplementary Fig. 2)[40,44]. On average, each cell was tagged with ~1.55×10^6 DNA tags, comparable to the amount used in other reported methods[44]. Longer irradiation did not result in substantial non-specific labelling (Supplementary Fig. 3), but a higher probe concentration led to a slightly higher background (Supplementary Fig. 4). We also verified that the labelled cells maintained similar viability to that of the untreated cells and did not undergo morphological changes (Supplementary Fig. 5)[62]. To validate the labelling specificity, the cells were labelled with BP-1 and a biotinylated CP (bio-CP-1), lysed and the biotinylated proteins were analysed with western blotting. The results showed that FR was specifically labelled (Fig. 2d(i)), whereas no labelling was observed without FA on BP ((Fig. 2d(ii)), with free FA competition ((Fig. 2d(iii)) or without ultraviolet irradiation ((Fig. 2d(iv)) (also see Supplementary Fig. 6). The labelled proteins were affinity purified with streptavidin beads, and mass spectrometry (MS) analysis confirmed the labelling specificity (Fig. 2e). Next, we tested whether the labelling was responsive to target abundance. HeLa cells were harvested after different passages to obtain cell batches with different FR expression levels. Using western blotting, the FR protein concentration in the labelling reaction was estimated to be 0.024–0.64 μM (Supplementary Fig. 7). These cell batches were labelled with BP-1/CP-1 and flow cytometry showed that the labelling efficiency correlated with the FR-expression level. Similar correlation was also observed with different cell types (Supplementary Figs. 8–10). As a test in a mixed cell population, HeLa^{FR} and B16-F10 cells, which have a low FR expression, were stained with red and blue dyes, respectively, co-cultured for 18 hours and labelled with BP-1/CP-1. As shown in Fig. 2f, nearly all the labelling occurred on HeLa^{FR} cells (red); in contrast, B16-F10 cells (blue) were barely labelled. Flow cytometry also corroborated the labelling selectivity (Fig. 2g). We further tested the labelling with two more proteins. First, integrin is a class of membrane proteins that regulates cell adhesion properties[63]. CycloRGDfK, an integrin $\alpha_v\beta_3$ ligand, was coupled to DNA as the BP (RGD–BP; Supplementary Fig. 11). A375 cells, a cell line with high $\alpha_v\beta_3$ expression, were labelled with RGD–BP/CP-1. The results showed a higher fluorescence with the labelled A375 cells than with the negative controls. Next, another membrane protein, CA-12, was tested. Carboxybenzene sulfonamide (CBS) was coupled to DNA as CBS–BP, which showed an affinity of 0.97 μM for CA-12 (Supplementary Fig. 12). The CA-12 concentration was measured to be ~0.20 μM in the labelling reaction (Supplementary Fig. 7) and, as shown in Extended Data Fig. 1, the efficiency and specificity of CA-12 labelling were also validated.

After tagging, we used toehold displacement to remove BP to make the tag available for library hybridization (Fig. 3a): a FAM–BP (BP-2) and a nine-nucleotide-shorter CP-2 were first used to label the cells, and a DP was added to displace BP-2 from the cell surface. The nine-nucleotide toehold length was chosen to promote a fast displacement[64]. Strong fluorescence was observed after labelling (Fig. 3b(i)); DP addition greatly reduced the signal (Fig. 3b(ii)), but not with a mismatched DP (Extended Data Fig. 2), which indicates that the signal loss was due to a DP-mediated displacement. To confirm that the tag could hybridize with DNA-conjugated small molecules, a six-nucleotide FAM-labelled BP-3 was added. Our previous studies showed that short DNA strands (5–7 bases) form unstable duplexes under physiological conditions, but can be stabilized by binding to the target protein[32,65]. Here, BP-3 formed an unstable duplex and could be washed off the cell surface unless it had a ligand bound to the target, which resulted in enhanced affinity due to the avidity effect. As shown in Fig. 3b(iii), BP-3 restored the cell fluorescence to a similar level as that in Fig. 3b(i), whereas using a BP without FA, a mismatched BP-3 or free FA competition, little fluorescence gain was observed (Extended Data Fig. 2). DP-mediated displacement was also observed with DNA-tagged CA-12 on A549 cells (Extended Data Fig. 2). These results showed that BP-3 could be washed off the cell when it does not hybridize with a DNA tag, despite the high affinity of FA ($K_d \approx 1$ nM) (ref. [66]). To investigate this, we determined the mean fluorescence intensity of the flow cytometry data, which is more quantitative than gating the histograms. The results showed that the mismatched BP-3 yielded only a slightly higher labelling than that with FA competition (1.2-fold); we also measured the binding affinity of BP-3 to FR and a K_d of

Fig. 2 | Labelling of FR on HeLa cells. a, Cells were labelled with BP-1/CP-1 (1.0 μM). **b**, Confocal images of the labelled cells: HeLa^{FR} (i); HeLa (ii); HeLa^{FR} with a mismatched CP (iii); no FA on the BP (iv) and no ultraviolet irradiation (v). **c**, Flow cytometry analysis of the labelled or unlabelled cells. Left, flow cytometry histogram; right, column graph of the analysis data; experiments 1–4 in both graphs are specified in the right panel. $n = 3$ biologically independent HeLa^{FR}/HeLa cell batches were examined. Data are presented as mean ± s.d. **d**, HeLa^{FR} cells were labelled with BP-1/bio-CP-1 and analysed with western blotting: affinity labelling experiment (i), no FA on BP (ii), free FA competition (iii), no irradiation (iv). Free FA, 50-fold FA competitor; no FA, no FA on the BP; diamonds, non-specific proteins; loading control, endogenous biotinylated proteins, marked with asterisks; a portion of each sample was separately blotted for tubulin as an additional sample processing control. **e**, MS analysis of the affinity-purified proteins. *y* axis from **d**(i) (experiment) and *x* axis from **d**(ii) (control). **f**, HeLa^{FR} and B16-F10 cells were stained with CellTracker Deep Red and Blue CMF2HC dyes, respectively, co-cultured and labelled with BP-1/CP-1 (i) and the negative control (no FA) (ii). Green, FAM; red, Cy5; blue, 4,6-diamidino-2-phenylindole. Inset (top right): enlargement of the image in the dashed square. **g**, Co-cultured HeLa^{FR} and B16-F10 cells were labelled with BP-1/CP-1 and analysed with flow cytometry by cell size and granularity (i) and by fluorescence (ii).

710 nM was obtained (Supplementary Fig. 13), much lower than that of free FA. We reasoned that the reduced affinity might be due to DNA conjugation, and may explain that BP-3 alone could be washed off the cell surface. Furthermore, using the CompareTm DNA calculator[67], the free-energy gain from BP-3/CP-2 hybridization was calculated to be 4.12 kcal mol^{-1}, which corresponds to a K_d of ~1 mM and an about three orders of magnitude affinity increase; thus, the enhanced affinity of BP-3 with the DNA tag was in the subnanomolar range, which may lead to the retention of BP-3 on the cell. Previously, we showed that a BP with a low micromolar affinity could efficiently label the target protein[54]. Here, both FA–DNA and CBS–DNA have a K_d of ~1 μM, and they, indeed, yielded an efficient labelling. For CBS–DNA, the relatively high CA-12 abundance may also contribute to the labelling efficiency. Several previous reports also showed that FA- and CBS-based probes were able to label FR and CA-12 on live cells, respectively, and the probes could be



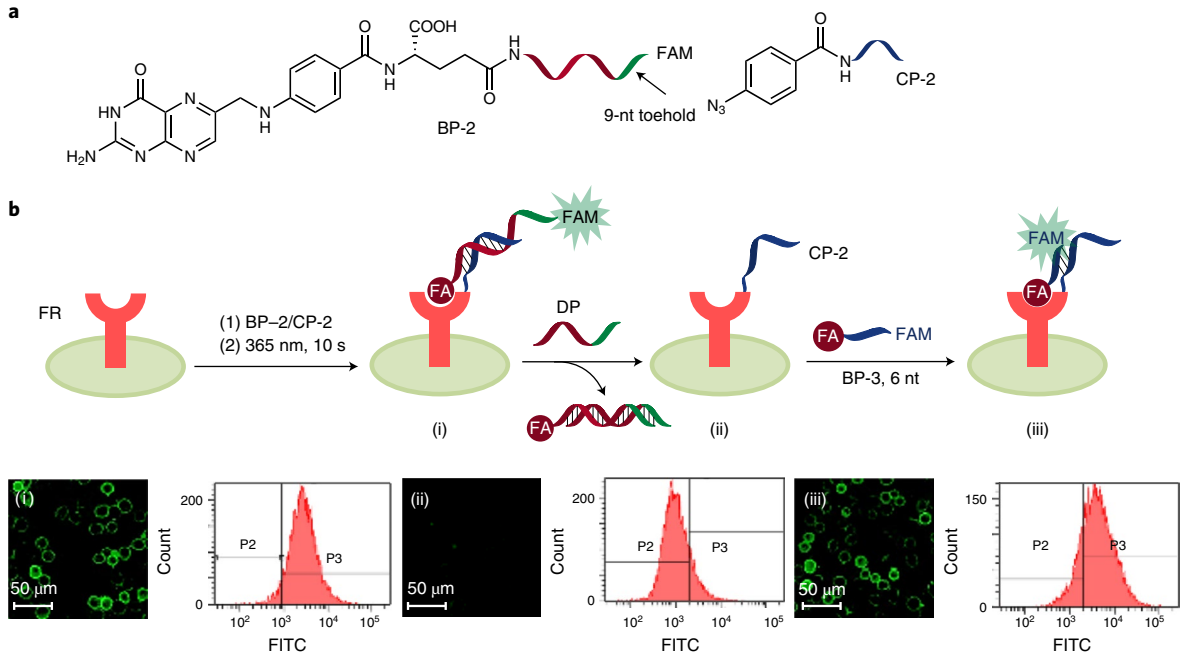


Fig. 3 | BP removal using toehold displacement after the labelling of FR. a, The structures of BP-2 and CP-2. **b**, HeLa^{FR} cells were subjected to BP-2/CP-2-mediated labelling (i), toehold displacement (ii) and BP-3 addition (iii). Each step was monitored with fluorescence imaging and flow cytometry: after BP-2/CP-2 labelling (i), after DP addition (ii) and after BP-3 addition (iii). Micrograph images show the cell fluorescence. The vertical lines in the flow cytometry histograms indicate gating thresholds and the horizontal lines are floating bars provided by the software for manually setting the gating threshold. See Extended Data Fig. 2 for the negative controls. FITC, fluorescein isothiocyanate.

feasibly washed off the cell after the labelling[66,68,69]. Collectively, these results show that small-molecule ligands can direct the labelling of membrane proteins with a DNA tag on live cells.

Antibody could also guide the labelling of membrane proteins with a DNA tag. Many membrane proteins do not have a known small-molecule ligand; thus, we investigated whether antibodies could be used, as they are more broadly available for membrane proteins. First, we chose EGFR, a transmembrane receptor tyrosine kinase, to test this approach. We coupled an antibody that recognized the extracellular domain of EGFR (Met1–Ser645) to a 5′-thiol-modified DNA through the heterobifunctional linker *N*-(β-maleimidopropyloxy)succinimide (BMPS) (Extended Data Fig. 3). To obtain a low DNA-to-antibody ratio, we modified a reported protocol[70] and found that a 1:5:2 antibody/BMPS/DNA ratio gave the optimal balance between conjugation efficiency and the DNA-to-antibody ratio. Gel analysis showed three major products, which were purified and characterized with MS along with the unlabelled antibody. The results confirmed that 1–2 DNAs were conjugated to the antibody (Extended Data Fig. 3 and Supplementary Fig. 14). The mono-DNA–antibody conjugate **b** (Supplementary Fig. 14) showed similar antigen-binding properties to those of the unmodified antibody (Supplementary Fig. 15)[59] and was used as the antibody–BP duplex (antibody–BP). Here, the DNA–antibody conjugate was prepared without site specificity, and different lysine residues were modified. We hypothesized that this might be advantageous for the selection, as a mixture of conjugates may deliver the DNA tag to different sites and the library selection may cover more regions on the target protein.

Owing to the antibody's large size, CP may crosslink to the antibody instead of the target protein, which may reduce the labelling efficiency. We proposed to address this issue by extending the crosslinker away from the antibody with a spacer of '*n*' bases (Fig. 4a); however, long CPs may also lower the labelling yield. Thus, we sought to find a spacer length that balanced these two factors. First, several biotinylated CPs with a wide range of lengths ($n = 3$, 9, 18, 25, or 30) were paired with an antibody–BP to label EGFR on A431 cells. As shown in Supplementary Fig. 16, all the probes could label EGFR, although short CPs yielded poor labelling, and $n = 18$ showed a relatively high labelling efficiency. As expected, considerable antibody self-labelling was observed, especially with the short CPs. Next, more detailed studies were conducted by centring the *n* value at 18. A series of CPs ($n = 9$, 12, 15, 18, 21, 25 and 30) were tested with three repeats for each probe (Fig. 4b,c and Supplementary Fig. 17). CPs with $n = 18$, 21 and 25 showed similar labelling efficiencies with relatively low levels of self-labelling, but self-labelling could not be completely avoided. Although the longest CP ($n = 30$) yielded the least self-labelling, it exhibited very large error bars. Finally, we chose $n = 18$ as the optimal spacer length in this study. The CPs with $n = 21$ and 25 performed similarly, but they were longer. To validate the labelling specificity, we performed several negative control experiments, and in all cases no or little labelling was observed (Fig. 4d). The captured proteins were affinity purified, and MS characterization also confirmed the labelling specificity (Fig. 4e). Finally, we confirmed the feasibility of removing the antibody–BP from the cell surface by using DP-mediated toehold displacement (Supplementary Fig. 18).

We further tested antibody-guided labelling with three more membrane proteins: FR, transferrin receptor 1 and CD44 (cluster of differentiation 44). For each protein, we used an antibody that recognized the extracellular domain of the target. DNA conjugation of the 3 antibodies exhibited a slightly higher DNA-to-antibody ratio than that of EGFR with 1–3 DNA strands installed on the antibody, presumably due to variations in the experimental conditions (Supplementary Fig. 19). Mono-DNA conjugates were purified, characterized, and used as the antibody–BPs, which were paired with a FAM–CP ($n = 18$) to label the respective targets[40]. As shown in Supplementary Fig. 20, these antibody–BPs efficiently installed DNA tags on the targets. Compared with FA-guided labelling, the FR antibody exhibited ~56% efficiency, possibly because of the small

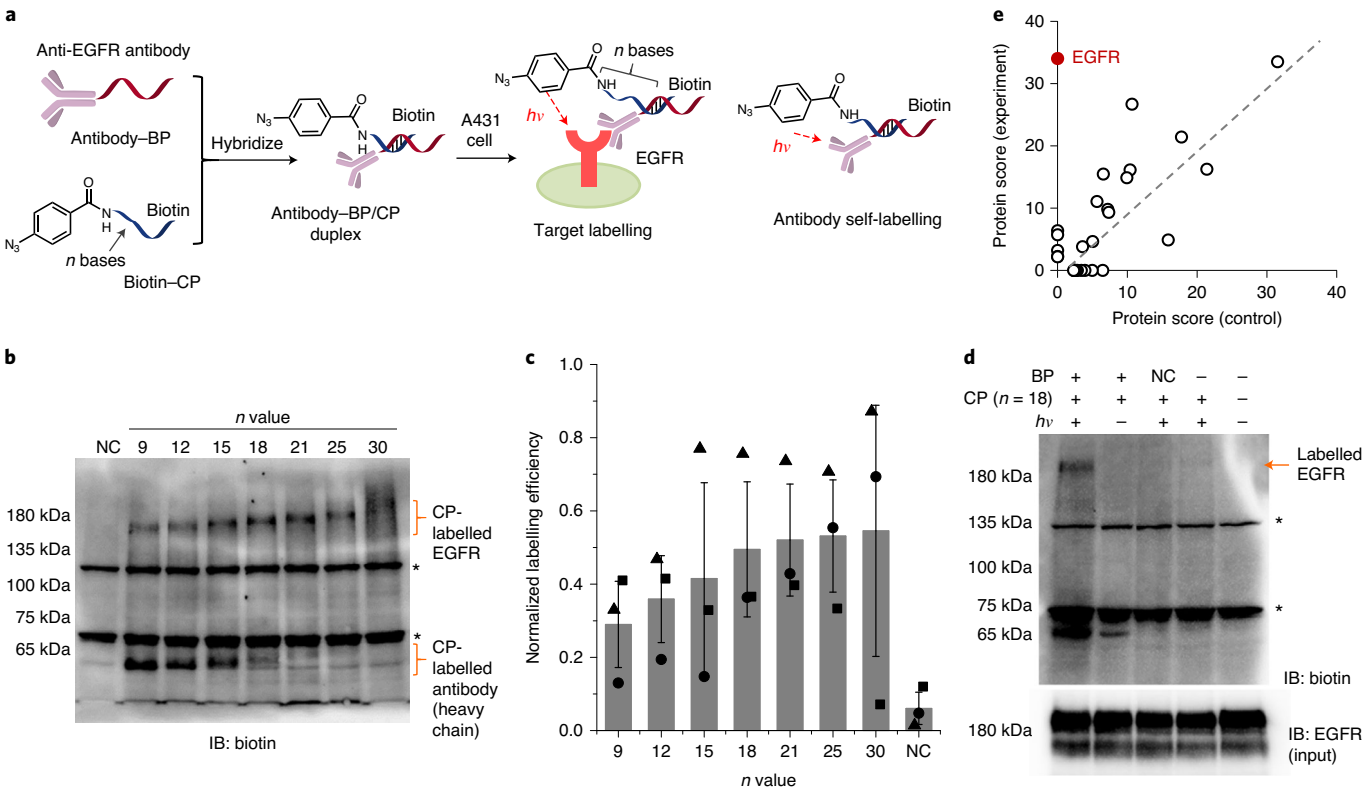


Fig. 4 | Cell-based labelling of EGFR with antibody-guided probes. a, The self-labelling issue could be addressed by inserting a spacer. **b**, The anti-EGFR-antibody–DNA conjugate was paired with biotin CPs with different space lengths and used to label EGFR on A431 cells. The biotinylated proteins were analysed with western blotting; the *n* values are as shown. NC, no antibody on BP; *, endogenously biotinylated protein; IB, immunoblotting. **c**, Column graph summarizing the labelling results. $n = 3$ biologically independent sample; data are presented as mean values ± s.d. **d**, The labelling experiment with antibody–BP/CP ($n = 18$) was performed with negative controls. The loading controls were internal endogenous biotinylated proteins, marked with an asterisk; a portion of each sample was separately blotted for EGFR as an additional sample processing control. **e**, MS analysis of the affinity-purified proteins from **b**. The *y* axis shows the experiments with the antibody–BP and the *x* axis (control) shows those with no antibody on BP. The few proteins with high scores in both experiments may be endogenously biotinylated proteins.

size of FR (~35 kDa). Using a short CP ($n = 3$), the labelling efficiency decreased to ~24%, which further confirms the importance of spacer length. For CD44 and transferrin receptor 1, higher labelling efficiencies were observed. Next, to validate the labelling specificity, a biotin–CP ($n = 18$) was paired with the antibody–BPs to label the cells; the biotinylated proteins were analysed with western blotting and characterized with MS after affinity purification, and specific labelling was observed for all the targets (Supplementary Fig. 20). Collectively, these results demonstrated that antibodies may be used for the membrane proteins without known small-molecule ligands; however, if available, small molecules are still more desirable, especially for proteins with small extracellular domains.

Selection of DELs against membrane proteins on live cells. The encoding DNA of DELs contains a pair of primer-binding sites with constant sequences. If the DNA tag is complementary to the site next to the library compound, it may direct the library hybridization and enable the selection. Again, the FA/FR system was used for validation. We prepared a 4,800-member tripeptide DEL and spiked in an FA–DNA as a positive control (Fig. 5a). HeLa cells were labelled with an FA-conjugated BP/CP probe pair (BP-4/CP-3; Fig. 5b) and incubated with the 4,801-member DEL. CP-3 had a seven-nucleotide sequence complementary to the primer-binding site and the hybridization increased the ligand affinity and stabilized the duplex; thus, the binders were retained on the cell, whereas the non-binders were washed away. Although other cell-surface proteins also interacted with the library compounds, the interactions lacked support from the DNA tag, which can be controlled by using unlabelled cells. After library incubation (30 °C, one hour), the cells were subjected to ten cycles of washing and resuspension to wash away the non-binders. Next, the cells were heated at 95 °C for ten minutes and centrifuged to elute the binders, and the supernatant was PCR-amplified and sequenced with next-generation sequencing (NGS)[38]. We used strong heating for the elution, similar to that in previous reports[33,37]; other approaches, such as selective elution by free ligand competition[31,71,72] or direct on-cell PCR[38], may also be used to obtain a higher decoding specificity. The sequencing data were processed with a script to quantitatively tally the codons for each compound and calculate the enrichment fold, and the results are shown in scatter plots[25,65]. In each plot, the compounds with a high enrichment fold and a high post-selection sequencing count were considered potential hits. As shown in Fig. 5c, with the labelled cells, FA was strongly enriched, whereas FA was not enriched with the unlabelled cells (Fig. 5d).

In a selection, both target concentration and ligand affinity affect the binding equilibrium, which dictates the affinity range of the hit compounds. To better understand this in the context of cell-based selection, we devised a method to estimate the target concentration—the average number of DNA tags on a cell could be quantified with flow cytometry (Supplementary Fig. 2)[40,44]. Assuming it represented the number of DNA-tagged target molecules, the target concentration could be calculated based on the number of cells, cell volume and the volume of the selection suspension. In the selection of Fig. 5, the number of DNA-tagged FRs on each cell was ~52,000,

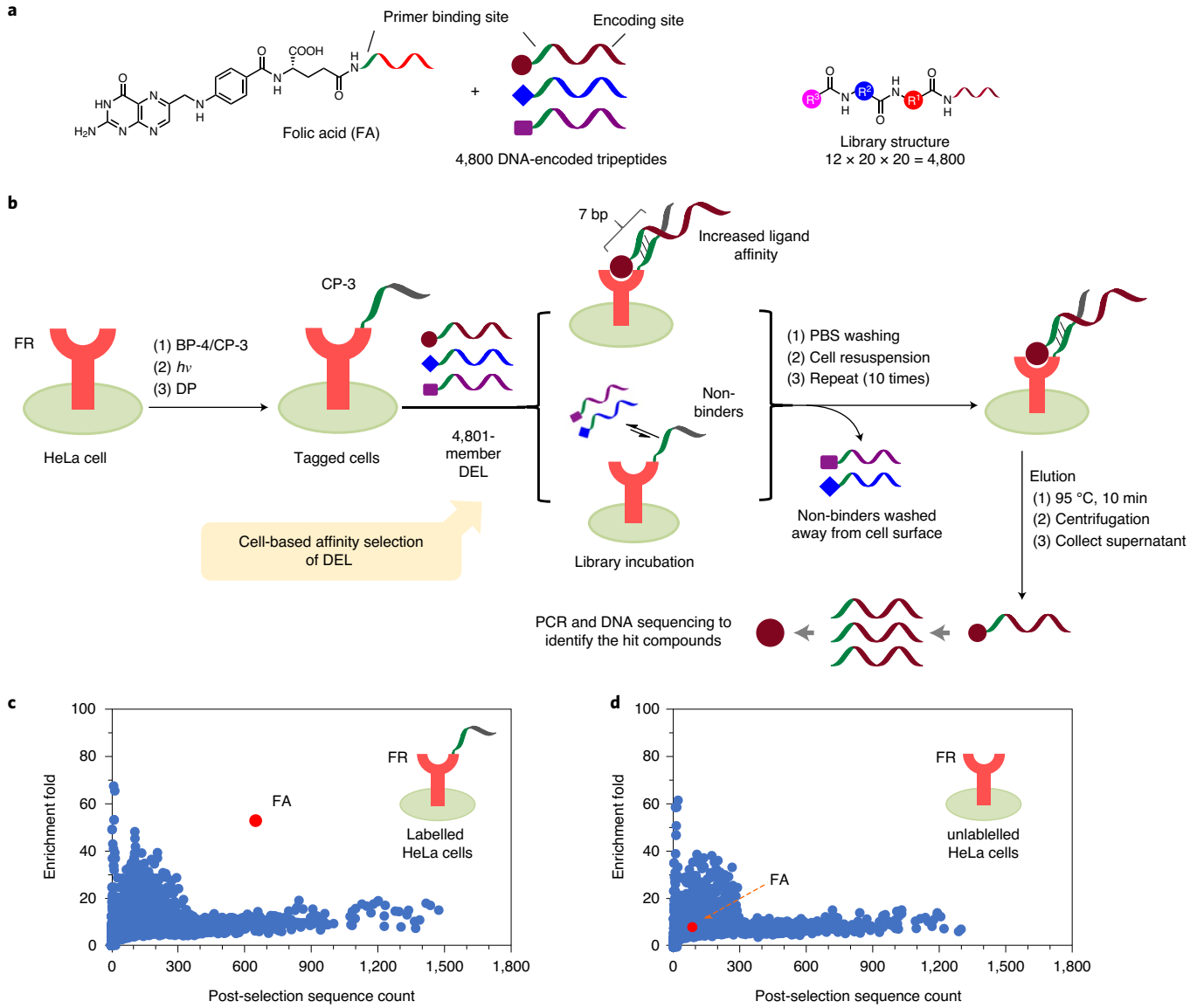


Fig. 5 | DEL selection against DNA-tagged FR on HeLa cells. a, The composition of the 4,801-member DEL. **b**, Regular HeLa cells were labelled with BP-4/CP-3 and incubated with the library. After repeated washing with PBS (ten times) to remove the non-binders from the cell surface, the binders retained on the cell were eluted by heating (95 °C, 10 min) and centrifugation. The selected DEL compounds in the supernatant were collected for hit decoding. **c**,**d**, Scatter plots of the selection results for the labelled (**c**) and unlabelled (**d**) HeLa cells. Enrichment fold = % post-selection/% preselection. The positive control FA is highlighted. bp, base pairs.

and HeLa cell has a size of ~3,000 μm³ (ref. [73]); thus, the target concentration was calculated to be ~5.1 nM in the selection suspension (Supplementary Section 9). Using the CompareTm DNA calculator, the CP-3 tag with a seven-nucleotide complementarity provided a free energy gain of 5.93 kcal mol^{-1}, which corresponded to an about 18,838-fold affinity increase[67]; thus, the K_d of FA–DNA would be ~0.037 nM, much lower than the target concentration and lead to the enrichment of FA–DNA. The non-binders hybridized with the DNA tag with a weak affinity of ~53.1 μM and therefore were not enriched. With the unlabelled cells, the FR concentration was ~24 nM (Supplementary Fig. 7) and there was no 'affinity boost' from the DNA tag, which makes the enrichment of FA–DNA very difficult, especially given the repetitive washing in the selection.

Furthermore, we performed two series of selections with varied target concentration and length of tag complementarity (Extended Data Fig. 4). Besides FA, methotrexate (MTX) was also added to the library; MTX binds to FR with a K_d of ~100 nM as a free ligand[74], but its conjugates have a much lower affinity (~5.2–40 μM) (refs [68,75]). Here, the affinity of MTX–DNA was measured to be ~26.6 μM. First, HeLa cells were harvested after different passages, and six batches with increasing FR expression (P1–P6) were fluorescently labelled. The target concentration of the cell batches was measured to be from ~5.1 to 110.9 nM. Next, the original cells were labelled with a tag with seven complementary bases and selected with the 4,802-member library. As shown in Extended Data Fig. 4, FA was highly enriched in all cases, whereas MTX was enriched at higher target concentrations (after P4), and the enrichment fold increase roughly followed the target concentration. Next, we calculated the ligand affinity increase with tags from six to ten complementary bases (Extended Data Fig. 5); the results showed that these tags provided a free-energy gain that ranged from 5.65 to 9.70 kcal mol^{-1}. Hybridization with the six- and seven-nucleotide tags gave a K_d of 84.5 and 53.1 μM, respectively, consistent with our observation that they formed dynamic DNA duplexes[32,65]; the eight-nucleotide tag

gave a K_d of 14.1 μM; and, as expected, the nine- and ten-nucleotide tags formed stable duplexes (K_d = 285 and 103 nM, respectively). Thus, we conducted the selections with the tag lengths of six, seven, eight and ten complementary bases. The ten-nucleotide tag represented a stable hybridization, and tags shorter than six nucleotides were not tested because they would have hybridization-specificity issues. Two cell batches, P1 and P4, were used in the selection. As shown in Extended Data Fig. 5, the enrichment of FA and MTX increased with more complementary bases. Interestingly, with the ten-nucleotide tag, the enrichment of FA and MTX decreased, but many other library compounds were enriched and the scatter plots exhibited a more 'spread-out' pattern (Supplementary Fig. 22). We reasoned this might be because the ten-nucleotide tag formed a stable duplex with all the library compounds and led to the enrichment of many low-affinity binders. We also performed selection against the membrane protein CA-12. GLCBS and CBS, two known CA-12 ligands (K_d = 0.3 and 1 μM, respectively) were added to the 4,800-member library and selected against DNA-tagged CA-12 on A549 cells. As shown in Extended Data Fig. 6, GLCBS and CBS were highly enriched with the tagged cells, but not with the unlabelled cells. The concentration of tagged CA-12 in the selection was estimated to be ~29.8 nM, and the seven-nucleotide tag would increase the affinity of GLCBS/CBS to the subnanomolar range, well below the target concentration.

Collectively, these experiments provided an approach to estimate the target concentration and the ligand affinity in the selection: fluorescent tagging could calculate the target concentration, whereas the tag length determined the affinity increase. As changing the membrane protein expression could be challenging in practice, varying the tag length is a more feasible way to tune the binding equilibrium. In principle, 6–8-nucleotide tags could increase the affinity of a weak binder with a high micromolar K_d to a low nanomolar one, which should be suitable for most applications. A tag longer than nine-nucleotide would lead to stable hybridization of all the library compounds and may mask the true binders. However, note that the calculation was still approximate, as ligand binding and DNA hybridization may not be simply additive. In fact, either cooperativity or a penalty factor would affect the avidity-mediated gain and/or loss of the apparent binding affinity[76], and DNA tagging may contribute additional non-specificity. Nevertheless, this method provides a useful and relatively quantitative guideline for conducting the DEL selection on live cells. In addition, it may also be applicable to non-cell-based selections, in which the target could be tagged to identify weak binders without high protein concentrations[29,30].

Next, we prepared a large-scale DEL with 30.42 million tripeptides (Fig. 6a). The library was selected against FR on either tagged (guided with FA–BP) or untagged HeLa cells. The selection with the tagged cells identified several distinctly enriched compounds (**H1**–**H5**; Fig. 6b), which were not enriched in control selection (Fig. 6c). **H1**–**H5** were resynthesized 'off-DNA' and assayed for FR-binding affinities with surface plasmon resonance (SPR; Fig. 6d). **H2**–**H5** showed low micromolar affinities, whereas **H1** had a relatively high binding affinity of 58 nM. This selection used a seven-nucleotide tag; thus, the binding affinity of **H5** would be 'boosted' to ~1.4 nM, well below the target concentration. In contrast, for the unlabelled cells, the FR concentration was 24.7 nM (Supplementary Fig. 7), which was not sufficient to enrich these hit compounds. **H1** and **H2** were also coupled with FAM (**H1**–FAM and **H2**–FAM), and their FR-binding affinities were determined with fluorescence polarization (FP). The results corroborated the SPR data (Fig. 6f). We also synthesized three 'hits' from the selection with the untagged cells (**C1**–**C3**), and SPR analysis showed that they were not FR binders (Fig. 6d), which suggests that they might bind other molecules on the cell surface. In addition, to control for DNA–cell interactions, a 'blank library' with the same DNA composition, but without the small molecules, was also selected, and none of the hit compounds became enriched (Supplementary Fig. 24). Furthermore, the strongest binder **H1** was coupled with a FAM-labelled DNA (**H1**–BP) and incubated with HeLa^{FR} cells. After very gentle washing due to the non-covalent binding, the cells were analysed with flow cytometry and fluorescence imaging. As shown in Fig. 6g,h, HeLa^{FR} cells could be stained with **H1**–BP (i), but not with FAM–DNA without **H1** (iii), and HeLa cells with a low FR expression also yielded a low fluorescence (ii).

Finally, we investigated whether a membrane protein labelled with an antibody-based probe could be subjected to DEL selection. Although many EGFR inhibitors are available as anticancer drugs, most of them target the intracellular kinase domain and it would be important to identify extracellular EGFR binders. First, A431 cells were labelled with the antibody–BP/CP (n = 18, seven-nucleotide tag complementarity). After toehold displacement, the tagged cells were subjected to selection with the 30.42-million-member library, and unlabelled A431 cells were also used as the control (Fig. 7a). The selected compounds were identified with PCR amplification and NGS. To better identify the compounds specifically enriched with the tagged cells, the post-selection sequence count and the enrichment fold were individually normalized, and then combined to generate a 'selection score' (SC) for each compound (Supplementary Information). The SC values of the compounds selected with the tagged cells were plotted against those from the selection with the untagged cells. As shown in Fig. 7b, the compounds specifically enriched with the tagged cells could be easily identified in the top left area of the plot. We selected five hits (**H6** to **H10**) for resynthesis; their binding affinities were determined using SPR, and the K_d values ranged from low- to mid-micromolars (Fig. 7c). The EGFR concentration in the suspension of the unlabelled cells was ~254 nM (Supplementary Fig. 7), which was too low to enrich micromolar binders. For the tagged cells, the seven-nucleotide DNA tag increased the ligand affinity substantially; for example, the weak binder **H10** would have an affinity of ~3.6 nM, well below the tagged EGFR concentration in the selection (~49.8 nM; Supplementary Fig. 18). We also synthesized a compound enriched with the unlabelled cells (**C4**) and a compound enriched with both the labelled and unlabelled cells (**C5**). SPR analysis showed that these compounds were not EGFR binders, which suggests that they might bind other molecules. In addition, a fluorescence polarization analysis also corroborated the SPR results (Supplementary Fig. 26). Collectively, these results demonstrated that membrane proteins could be tagged with either small-molecule or antibody-based BP/CP probes and subjected to DEL selection to identify novel ligands and that DNA tagging is essential to achieve target specificity and an increased ligand affinity for hit enrichment.

Discussion

In summary, we developed a method to select DELs against membrane proteins on live cells. Installing a DNA tag on the target provides a homing beacon to guide library hybridization and to promote ligand binding, despite the low target concentration. We showed that both small molecules and antibodies can be used to guide the labelling. Recently, DEL emerged as an important ligand discovery technology now widely adopted by the pharmaceutical industry in drug discovery[14]. However, the target scope of DEL was mostly limited to purified proteins, whereas cell-based DEL selection remained underexplored. This approach is expected to expedite ligand discovery for many membrane proteins. For example, classic drug targets, such as G protein-coupled receptors and ion channels, may be revisited in a cellular context by harnessing the vast diversity of DELs[23,24,77], and this method may also be used to interrogate membrane proteins intractable to traditional high-throughput screening approaches.

This method has several aspects to be further developed. First, it requires a known ligand to guide the labelling. The results showed

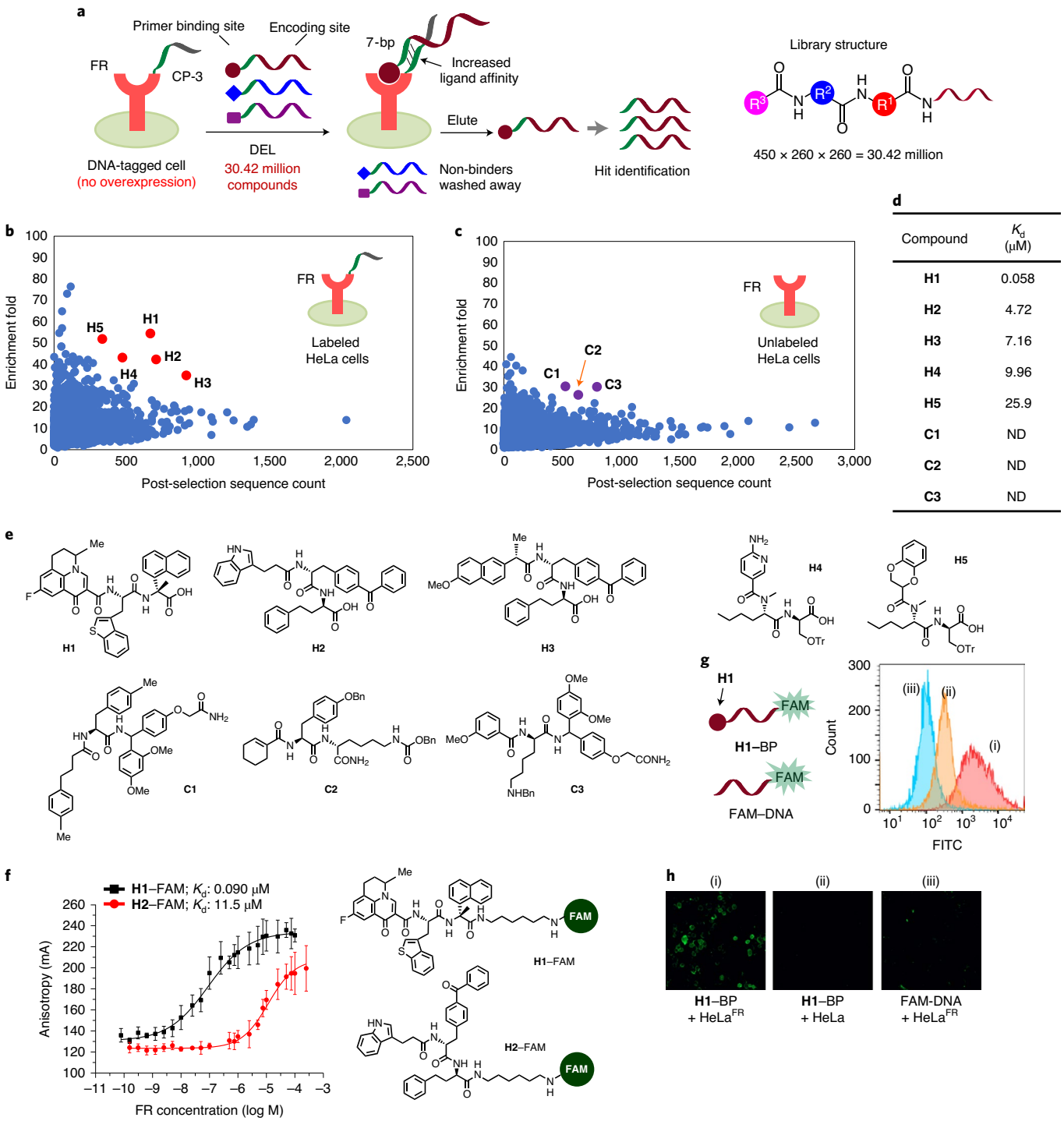
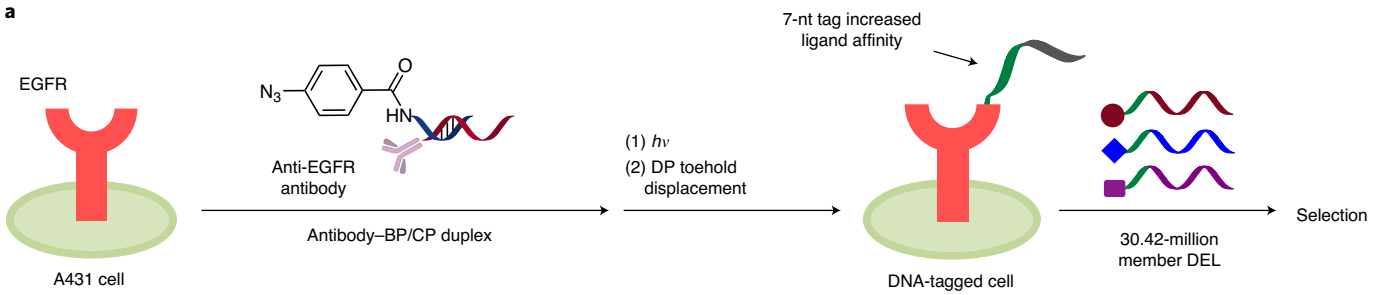
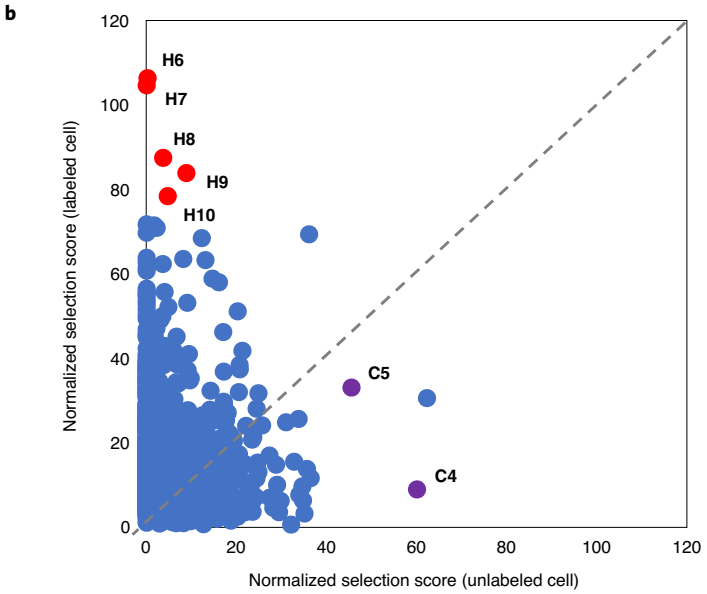


Fig. 6 | Selection of a 30.42-million DEL against FR on live cells. a, The selection scheme and library structure. The selection procedure was the same as that shown in Fig. 5. **b**,**c**, The results are shown in scatter plots of the labelled (**b**) and unlabelled (**c**) cells. The selected hits are highlighted. **d**, SPR analysis of the 'off-DNA' hit compounds from **b** and **c**. ND, affinity not detectable. Sensorgrams are shown in Supplementary Fig. 23. **e**, Structures of the hit compounds. **f**, Fluorescence polarization analysis of **H1**-FAM and **H2**-FAM. $n = 3$ biologically independent samples; data points are mean values ± s.d. **g**,**h**, Flow cytometry (**g**) and fluorescent imaging (**h**) of cells stained with **H1**-BP/HeLa^{FR} (i), **H1**-BP/HeLa (ii) and FAM–DNA/HeLa^{FR} (iii). Histograms are shown in Supplementary Fig. 25.

that small molecules are more straightforward to use, especially for targets with a small extracellular domain (for example, FR), but small molecule ligands may not be available for many membrane proteins. The ability to use antibodies provides a broader target coverage, and other types of ligands, such as peptides, small engineered proteins and aptamers, which had been used to graft DNAs onto cells[47–52], may also be employed to further expand the target scope. In principle, for any DEL selection against endogenous protein in a complex biological milieu, a guiding ligand will be required to achieve target specificity; otherwise, a completely different strategy would be necessary. Second, the selection is expected to be limited to the region that surrounds the DNA attachment point on the





c

Compound	K_d (μM)
H6	12.0
H7	4.2
H8	35.8
H9	18.0
H10	67.5
C4	>200
C5	>200

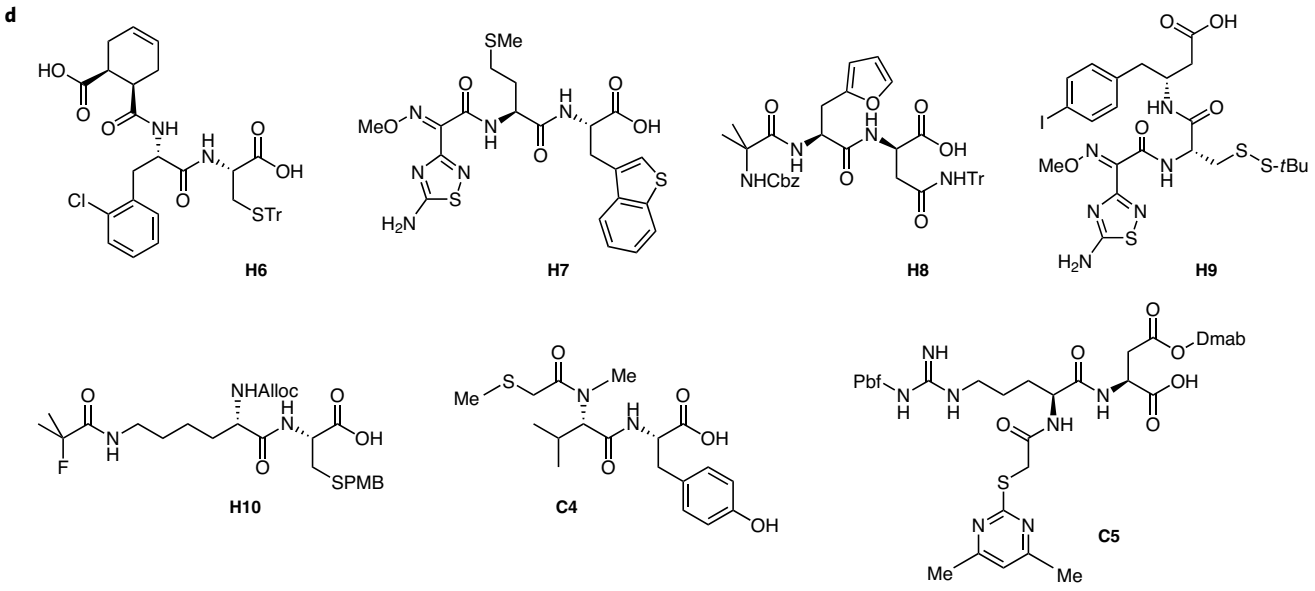


Fig. 7 | Selection of a 30.42-million DEL against DNA-tagged EGFR on live cells. a, EGFR on A431 cells was labelled with an anti-EGFR-antibody–BP/CP ($n = 18$, seven-nucleotide complementarity) and selected with the 30.42-million DEL. The selection procedure was the same as that used above (Figs. 5 and 6). **b**, The selection results are shown in a scatter plot: the *x* axis shows SC with unlabelled cells (control) and the *y* axis shows SC with labelled cells (experiment). The resynthesized hit compounds are highlighted. **c**, SPR analysis of the hit compounds. Sensorgrams are shown in Supplementary Fig. 26. **d**, Structures of the hit compounds.

target, but it could also be tuned by using CPs with extended cross-linkers. Alternatively, shifting the DNA tag/library hybridization position may be another way to explore a larger area of the targe, and it could also be exploited for site-specific ligand discovery. For example, orthosteric or allosteric ligands could be used to guide the labelling so that the selection may identify site-specific or even

functional-state-specific binders. Here the antibody–DNA conjugates were prepared without site specificity, which might be beneficial because the selection could cover more regions of the target protein. In addition, using antibody–BP with multiple DNA strands (for example, band c in Extended Data Fig. 3) may guide the library to explore multiple sites on the target simultaneously. Conversely, using site-specific antibody–DNA conjugates[59] should the control library selection in the area close to the antibody-binding site. Third, the strong heat denaturing condition in the elution step could be changed to alternative approaches to achieve a higher decoding specificity, such as selective elution by ligand competition[31,71,72], or direct on-cell PCR[38] after trypsinization to suspend the cells. Finally, this method is suitable for the extracellular domain of membrane proteins, and recent work to deliver DELs into cells using cell-penetrating peptide has demonstrated the potential for intracellular DEL selection[37]. We will perform more in-depth studies on these aspects and exploit their utilities in ligand discovery.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-020-00605-x.

Received: 19 January 2019; Accepted: 10 November 2020;
Published online: 21 December 2020

References

1. Cournia, Z. et al. Membrane protein structure, function, and dynamics: a perspective from experiments and theory. *J. Membr. Biol.* **248**, 611–640 (2015).
2. Yin, H. & Flynn, A. D. Drugging membrane protein interactions. *Annu. Rev. Biomed. Eng.* **18**, 51–76 (2016).
3. Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schioth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug. Discov.* **16**, 829–842 (2017).
4. Li, X. L., Shao, C. S., Shi, Y. F. & Han, W. D. Lessons learned from the blockade of immune checkpoints in cancer immunotherapy. *J. Hematol. Oncol.* **11**, 31 (2018).
5. Rawlings, A. E. Membrane proteins: always an insoluble problem? *Biochem. Soc. Trans.* **44**, 790–795 (2016).
6. Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *Proc. Natl Acad. Sci. USA* **89**, 5381–5383 (1992).
7. Nielsen, J., Brenner, S. & Janda, K. D. Synthetic methods for the implementation of encoded combinatorial chemistry. *J. Am. Chem. Soc.* **115**, 9812–9813 (1993).
8. 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).
9. Zhao, G., Huang, Y., Zhou, Y., Li, Y. & Li, X. Future challenges with DNA-encoded chemical libraries in the drug discovery domain. *Expert Opin. Drug Discov.* **14**, 735–753 (2019).
10. 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).
11. Dickson, P. & Kodadek, T. Chemical composition of DNA-encoded libraries, past present and future. *Org. Biomol. Chem.* **17**, 4676–4688 (2019).
12. 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).
13. Huang, Y., Savych, O., Moroz, Y., Chen, Y. Y. & Goodnow, R. A. DNA-encoded library chemistry: amplification of chemical reaction diversity for the exploration of chemical space. *Aldrichim. Acta* **52**, 75–87 (2019).
14. 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).
15. Halford, B. Breakthroughs with bar codes—DNA-encoded libraries help pharma find drug leads. *Chem. Eng. News* **95**, 28–33 (2017).
16. Arico-Muendel, C. C. From haystack to needle: finding value with DNA encoded library technology at GSK. *MedChemComm* **7**, 1898–1909 (2016).
17. Goodnow, R. A. *A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery* (John Wiley & Sons, 2014).
18. Buller, F. et al. Selection of carbonic anhydrase IX Inhibitors from one million DNA-encoded compounds. *ACS Chem. Biol.* **6**, 336–344 (2011).
19. Kollmann, C. S. et al. Application of encoded library technology (ELT) to a protein–protein interaction target: discovery of a potent class of integrin lymphocyte function-associated antigen 1 (LFA-1) antagonists. *Bioorg. Med. Chem.* **22**, 2353–2365 (2014).
20. Wichert, M. et al. Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. *Nat. Chem.* **7**, 241–249 (2015).
21. Leimbacher, M. et al. Discovery of small-molecule interleukin-2 inhibitors from a DNA-encoded chemical library. *Chem. Eur. J.* **18**, 7729–7737 (2012).
22. Richter, H. et al. DNA-encoded library-derived DDR1 inhibitor prevents fibrosis and renal function loss in a genetic mouse model of Alport syndrome. *ACS Chem. Biol.* **14**, 37–49 (2019).
23. Ahn, S. et al. Allosteric 'beta-blocker' isolated from a DNA-encoded small molecule library. *Proc. Natl Acad. Sci. USA* **114**, 1708–1713 (2017).
24. Ahn, S. et al. Small-molecule positive allosteric modulators of the β_2-adrenoceptor isolated from DNA-encoded libraries. *Mol. Pharmacol.* **94**, 850–861 (2018).
25. 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).
26. 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. *Bioconjugate Chem.* **28**, 2293–2301 (2017).
27. Blakskjaer P. et al. A method for making an enriched library patent. WO patent 2012041633 A1, 2012.4.5. (2012).
28. Bao, J. et al. Predicting electrophoretic mobility of protein–ligand complexes for ligands from DNA-encoded libraries of small molecules. *Anal. Chem.* **88**, 5498–5506 (2016).
29. McGregor, L. M., Jain, T. & Liu, D. R. Identification of ligand–target pairs from combined libraries of small molecules and unpurified protein targets in cell lysates. *J. Am. Chem. Soc.* **136**, 3264–3270 (2014).
30. 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).
31. Kim, D. et al. Application of a substrate-mediated selection with c-Src tyrosine kinase to a DNA-encoded chemical library. *Molecules* **24**, 2764 (2019).
32. Zhou, Y. et al. DNA-encoded dynamic chemical library and its applications in ligand discovery. *J. Am. Chem. Soc.* **140**, 15859–15867 (2018).
33. 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).
34. 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).
35. 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).
36. Yan, M. et al. Next-generation glycan microarray enabled by DNA-coded glycan library and next-generation sequencing technology. *Anal. Chem.* **91**, 9221–9228 (2019).
37. 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).
38. Decurtins, W. et al. Automated screening for small organic ligands using DNA-encoded chemical libraries. *Nat. Protoc.* **11**, 764–780 (2016).
39. Chandra, R. A., Douglas, E. S., Mathies, R. A., Bertozzi, C. R. & Francis, M. B. Programmable cell adhesion encoded by DNA hybridization. *Angew. Chem. Int. Ed.* **45**, 896–901 (2006).
40. Gartner, Z. J. & Bertozzi, C. R. Programmed assembly of 3-dimensional microtissues with defined cellular connectivity. *Proc. Natl Acad. Sci. USA* **106**, 4606–4610 (2009).
41. Furst, A. L., Smith, M. J. & Francis, M. B. Direct electrochemical bioconjugation on metal surfaces. *J. Am. Chem. Soc.* **139**, 12610–12616 (2017).
42. Zhao, W. et al. Mimicking the inflammatory cell adhesion cascade by nucleic acid aptamer programmed cell–cell interactions. *FASEB J.* **25**, 3045–3056 (2011).
43. El Muslemany, K. M. et al. Photoactivated bioconjugation between *ortho*-azidophenols and anilines: a facile approach to biomolecular photopatterning. *J. Am. Chem. Soc.* **136**, 12600–12606 (2014).
44. Vogel, K., Glettenberg, M., Schroeder, H. & Niemeyer, C. M. DNA-modification of eukaryotic cells. *Small* **9**, 255–262 (2013).
45. Meyer, R., Giselbrecht, S., Rapp, B. E., Hirtz, M. & Niemeyer, C. M. Advances in DNA-directed immobilization. *Curr. Opin. Chem. Biol.* **18**, 8–15 (2014).
46. Furst, A. L., Smith, M. J. & Francis, M. B. New techniques for the generation and analysis of tailored microbial systems on surfaces. *Biochemistry* **57**, 3017–3026 (2018).

47. Vinkenborg, J. L., Mayer, G. & Famulok, M. Aptamer-based affinity labeling of proteins. *Angew. Chem. Int. Ed.* **51**, 9176–9180 (2012).
48. Cui, C. et al. Recognition-then-reaction enables site-selective bioconjugation to proteins on live-cell surfaces. *Angew. Chem. Int. Ed.* **56**, 11954–11957 (2017).
49. Li, L. et al. Aptamer displacement reaction from live-cell surfaces and its applications. *J. Am. Chem. Soc.* **141**, 17174–17179 (2019).
50. Wang, R. et al. Using modified aptamers for site specific protein–aptamer conjugations. *Chem. Sci.* **7**, 2157–2161 (2016).
51. Robinson, P. V., de Almeida-Escobedo, G., de Groot, A. E., McKechnie, J. L. & Bertozzi, C. R. Live-cell labeling of specific protein glycoforms by proximity-enhanced bioorthogonal ligation. *J. Am. Chem. Soc.* **137**, 10452–10455 (2015).
52. Skovsgaard, M. B., Mortensen, M. R., Palmfeldt, J. & Gothelf, K. V. Aptamer-directed conjugation of DNA to therapeutic antibodies. *Bioconjugate Chem.* **30**, 2127–2135 (2019).
53. Tamura, T. & Hamachi, I. Chemistry for covalent modification of endogenous/native proteins: from test tubes to complex biological systems. *J. Am. Chem. Soc.* **141**, 2782–2799 (2019).
54. 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).
55. 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).
56. Liu, Y. et al. Photoaffinity labeling of transcription factors by DNA-templated crosslinking. *Chem. Sci.* **6**, 745–751 (2015).
57. Bai, X. et al. Development of a DNA-templated peptide probe for photoaffinity labeling and enrichment of the histone modification reader proteins. *Angew. Chem. Int. Ed.* **55**, 7993–7997 (2016).
58. Bai, X. et al. An integrated approach based on a DNA self-assembly technique for characterization of crosstalk among combinatorial histone modifications. *Anal. Chem.* **90**, 3692–3696 (2018).
59. Rosen, C. B. et al. Template-directed covalent conjugation of DNA to native antibodies, transferrin and other metal-binding proteins. *Nat. Chem.* **6**, 804–809 (2014).
60. Kodal, A. L., Rosen, C. B., Mortensen, M. R., Torring, T. & Gothelf, K. V. DNA-templated introduction of an aldehyde handle in proteins. *ChemBioChem* **17**, 1338–1342 (2016).
61. Kane, M. A. et al. Influence on immunoreactive folate-binding proteins of extracellular folate concentration in cultured human cells. *J. Clin. Invest.* **81**, 1398–1406 (1988).
62. Furst, A. L., Klass, S. H. & Francis, M. B. DNA hybridization to control cellular interactions. *Trends Biochem. Sci* **44**, 342–350 (2019).
63. Kim, C., Ye, F. & Ginsberg, M. H. Regulation of integrin activation. *Annu. Rev. Cell Dev. Biol.* **27**, 321–345 (2011).
64. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131**, 17303–17314 (2009).
65. Li, G. et al. Design, preparation, and selection of DNA-encoded dynamic libraries. *Chem. Sci.* **6**, 7097–7104 (2015).
66. Miki, T. et al. LDAI-based chemical labeling of intact membrane proteins and its pulse-chase analysis under live cell conditions. *Chem. Biol.* **21**, 1013–1022 (2014).
67. Weber, G. Optimization method for obtaining nearest-neighbour DNA entropies and enthalpies directly from melting temperatures. *Bioinformatics* **31**, 871–877 (2015).
68. Fujishima, S. H., Yasui, R., Miki, T., Ojida, A. & Hamachi, I. Ligand-directed acyl imidazole chemistry for labeling of membrane-bound proteins on live cells. *J. Am. Chem. Soc.* **134**, 3961–3964 (2012).
69. Mizusawa, K., Takaoka, Y. & Hamachi, I. Specific cell surface protein imaging by extended self-assembling fluorescent turn-on nanoprobes. *J. Am. Chem. Soc.* **134**, 13386–13395 (2012).
70. Coyle, M. P., Xu, Q., Chiang, S., Francis, M. B. & Groves, J. T. DNA-mediated assembly of protein heterodimers on membrane surfaces. *J. Am. Chem. Soc.* **135**, 5012–5016 (2013).
71. Wrenn, S. J., Weisinger, R. M., Halpin, D. R. & Harbury, P. B. Synthetic ligands discovered by in vitro selection. *J. Am. Chem. Soc.* **129**, 13137–13143 (2007).
72. Hansen, M. H. et al. A yoctoliter-scale DNA reactor for small-molecule evolution. *J. Am. Chem. Soc.* **131**, 1322–1327 (2009).
73. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**, D750–D753 (2010).
74. Rijnboutt, S. et al. Endocytosis of GPI-linked membrane folate receptor-α. *J. Cell Biol.* **132**, 35–47 (1996).
75. Wong, P. T. & Choi, S. K. Mechanisms and implications of dual-acting methotrexate in folate-targeted nanotherapeutic delivery. *Int. J. Mol. Sci.* **16**, 1772–1790 (2015).
76. Vauquelin, G. & Charlton, S. J. Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands. *Br. J. Pharmacol.* **168**, 1771–1785 (2013).
77. Brown, D. G. et al. Agonists and antagonists of protease-activated receptor 2 discovered within a DNA-encoded chemical library using mutational stabilization of the target. *SLAS Discov.* **23**, 429–436 (2018).

Methods

DNA sequence design. DNA sequences were designed by using several online DNA property calculation tools, which included OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html)[78] and OligoAnalyzer (https://www.idtdna.com). These tools were mostly used to calculate the thermal stability (T_m) values, the sequence complementarity and the potential mismatch issues of DNA sequences. For more detailed thermodynamic parameter calculations (entropy, enthalpy and free energy), as shown in Extended Data Fig. 5, the CompareTm calculator developed by Weber[67] was used; specific experimental conditions, such as the ionic strength (salt concentration, bivalent ion concentration and so on) and the probe concentration, were included in the calculation to obtain more accurate results.

For stable DNA duplexes (for example, BP/CP and the ten-nucleotide DNA tag), the sequences were designed to have a T_m substantially higher than the experimental temperature. In general, DNA duplexes with 12–15 base pairs and a 40–50% or higher CG content were used; such DNA duplexes usually have a $T_m > 50$ °C under physiological conditions. For unstable, dynamic DNA duplexes (for example, the 6- to 8-nt DNA tag), the sequences were designed to have a T_m lower than the experimental temperature. The sequence design was mostly based on our previous studies on DNA-encoded dynamic libraries[32,65]. Previously, we showed that DNA sequences with 6–7 bases were suitable for dynamic DNA hybridization under physiological conditions, and such a duplex length corresponded to a 16–27 °C T_m, depending on the specific buffer condition and the specific DNA sequences. Changing the C/G contents in these sequences could fine tune the thermal stability of the corresponding DNA duplexes. After the initial design, the sequences were evaluated for potential mismatch issues using OligoAnalyzer. If mismatches were identified, the problematic bases were replaced and the new sequences were checked for stability and specificity through another round of iteration. The BP/CP hybridization sequences were designed to have a T_m of 55–60 °C. The CP DNA was fully complementary to BP and formed a very stable DNA duplex under the experimental conditions, whereas CP was partially complementary to the primer-binding site of the DNA-encoded library with 6–8 bases to create a dynamic hybridization[32,65]. More details on sequence design are provided in the Supplementary Information.

DNA modification and purification. Small molecules were activated in situ with 100 mM *N*,*N*′-dicyclohexylcarbodiimide and *N*-hydroxysuccinimide and then coupled to amine-modified oligonucleotides. The conjugation products were purified by a size-exclusion column (NAP-5, Cytiva, no. 17-0853-01) and reverse-phase HPLC using a gradient of acetonitrile (5–80%) in 100 mM triethylammonium acetate (pH 7.0), followed by lyophilization. Oligonucleotides were quantitated based on the calculated extinction coefficient at 260 nm and characterized by MS.

Antibody–DNA conjugate preparation. The preparation of antibody–DNA conjugates followed a previous reported protocol[70] with modifications. In a typical reaction, 300 µg of antibody were buffer-exchanged into 1× PBS buffer (pH 7.2) to a concentration of 1 mg ml^{-1} and then treated with 5 equiv. heterofunctional BMPS linker (5 µl in dimethylsulfoxide) for 3 h at room temperature. The BMPS-modified antibody was desalted with NAP-5 and concentrated in Amicon Ultra 30000-NMWL concentrators. The 5′-thiol-modified DNA (2 equiv. in 5 µl of 0.5 M TAPS buffer, pH = 8.0) was treated with TCEP (tris(2-carboxyethyl)phosphine) (50 µl, 0.2 mM in 0.5 M TAPS buffer, pH = 8.0) for 90 min at room temperature, followed by ethanol precipitation to remove the excess TCEP. The DNA pellet was dissolved with the solution that contained the concentrated antibody directly. The mixture was incubated for 4 h at room temperature and buffer exchanged to 1× PBS buffer to remove the excess DNA. The antibody-modified DNA mixtures were analysed with non-denaturing polyacrylamide gel electrophoresis, purified with fast protein liquid chromatography and characterized with MS. The antibody-modified DNA can be stored at 4 °C.

Labelling of membrane proteins on live cells. In small-molecule-guided DNA labelling, the cell medium was removed and the cells were washed three times with 1× PBS buffer (pH = 7.4) before labelling. To prepare the labelling probes, small-molecule-conjugated BP and CP (each probe typically 1 µM unless specified otherwise) were mixed in 1× PBS buffer, heated to 90 °C for 3 min and then cooled down to room temperature slowly. To the cells were added the prepared DNA probes, and then incubated at 4 °C for 1.5 h (total volume of 200 µl), followed by ultraviolet irradiation over ice at 365 nm for 10 s by a Uvata UV LED point light source. After labelling, the cells were washed three times with 1× PBS buffer to remove the free probes. For the suspension cells, cells were harvested and washed three times with 1× PBS buffer by gentle resuspension and centrifugation at 500*g* for 5 min before labelling. The cells were resuspended in the prepared probes with the same concentration and volume, incubated at 4 °C for 1.5 h and subjected to ultraviolet irradiation at 365 nm for 10 s. The cells were then washed three times by gentle resuspension and centrifugation at 500*g* for 5 min. In antibody-guided labelling, the cells were harvested and washed three times. The cells were resuspended with a solution that contained 1 µM antibody–BP and 2 µM CP in 1× PBS buffer (prehybridized, no heating and a total volume of 200 µl) and incubated at room temperature for 2 h before ultraviolet irradiation on ice at 365 nm for 10 s by a Uvata UV LED point light source. After labelling, the cells were washed three times with 1× PBS buffer for at least 20 min each time to remove free probes. Mixed cells were stained with CellTracker Deep Red Dye and Blue CMF2HC Dye, respectively, cocultured on a coverslip for 18 h in a 24-well plate at 37 °C before labelling and analysis. The cells shown in Supplementary Figs. 3 and 4 were labelled following the same procedure as described above, except that varied probe concentrations and irradiation times were used.

Quantification of DNA molecules on cells. Cells were treated in suspension conditions and analysed with flow cytometry (Supplementary Fig. 2). To determine the average number of DNA molecules on a cell, the mean fluorescence intensity for each labelling condition was calculated and compared with a linear fitting curve of the mean fluorescence intensity of standard beads (Quantum Alexa Fluor 488 MESF; Bangs Laboratories Inc.) following the protocol provided by the manufacturer and previous reports[40,44].

Toehold displacement. For adherent cells, the labelled cells were treated with 1 µM DP in 1× PBS buffer for toehold displacement at room temperature for 20 min. The cells were then washed three times with 1× PBS buffer. For the suspension cells, after displacement, the cells were washed three times with 1× PBS buffer by gentle resuspension and centrifuged at 500*g* for 5 min. For antibody-based labelling, the cells were washed three times with 1× PBS buffer for at least 20 min each time to remove the free probes.

Selection of DNA-encoded libraries on live cells. After the cells were labelled and subjected to toehold displacement, the cells (~10^7 cells per selection) were resuspended in a 1× PBS buffer (200 µl) that contained ~500 pmol DNA-encoded library. The cell suspension was incubated at 30 °C for 1 h, followed by washing with 1× PBS buffer ten times to remove the unbound ligands by gentle resuspension and centrifugation at 500*g* for 5 min each time. Bound ligands were finally eluted by heating the cells in 1× PBS to 95 °C for 10 min, followed by centrifugation at 13,300 r.p.m.[33,37]. The supernatant that contained the bound library members was collected and ethanol precipitated (detailed protocol is provided in the Supplementary Information) before being subjected to PCR amplification using the long primers compatible with high-throughput sequencing. The PCR products were quantified and then submitted for high-throughput sequencing.

NGS. Sequencing experiments were performed on Illumina sequencers using standard 2 × 75 pair-end sequencing reagent kits and hybridization primers. After the Illumina sequencing, raw data were exported for processing with a custom Python script (https://github.com/cenhuang0916/sequencing-data-processing-script.git). We first applied a regular expression for data cleansing and extracted codon region sequences from the full DNA sequences, and then the sequence counts for each library member before and after the selection were tallied to calculate the enrichment fold for each compound, following the data-processing methods previously reported[25,79–81]. The enrichment folds were plotted against the post-selection sequence counts in the form of scatter plots. The compounds with very low post-selection sequence counts were not considered due to statistical under sampling[80,81]. To better identify the binders that were specific for the DNA-tagged cells, the post-selection sequence count and the enrichment fold for each compound were normalized to the same scale, and then summed to generate the SC for each compound. The SC values of the compounds identified with DNA-tagged cells are plotted against the values from the selection with untagged cells (Fig. 7). The binders specifically enriched with the DNA-tagged cells are located in the upper left corner of the plots. The compounds that were identified from the selection with the tagged cells but not with the untagged cells are included in the plots, but the compounds that were only identified from the untagged cells were not plotted as they are unlikely to be specific binders.

Calculation of the target protein concentration in the selection. The cells were first labelled with a CP with a fluorescein label. After labelling, the cells were analysed with flow cytometry along with the fluorescent beads as the calibration standard, following the manufacturer's instructions. The average number of DNA molecules on each cell was calculated by following a previously reported method[40,44]. Then, the target protein concentration could be calculated using the equation target = [(number of DNA molecules on each cell)/6.022×10^{23}] × (number of cells)/(volume of the cell suspension − total volume of the cells). A 200 µl suspension with 10^7 cells was used in the selections. The average size of the cell was based on literature reports: Milo et al. for HeLa cells[73], Jiang et al. for A549 cells[82] and Zhang et al. for A431 cells[83]. Detailed calculation results are provided in the Supplementary Information.

Preparation of the 4,800- and 30.42-million-member tripeptide libraries. The libraries were prepared following the previously reports[25,84], which were based on the DNA-templated synthesis method[85,86], combined with a typical split-and-mix library synthesis protocol[10]. Both libraries have a three-building-block peptide structure, as shown in Figs. 5 and 6. In both libraries, the first two sets of building blocks (R^1 and R^2) are amino acids, whereas the third set of building blocks

comprises simple monofunctional acids (R^3). The 4,800-member library was prepared in a 20×20×12 format and the 30.42-million library was prepared in a 260×260×450 format. The scheme for library preparation is shown in Supplementary Fig. 31. After library synthesis, the library was analysed with Sanger sequencing as a quality control process to ensure correct encoding (Supplementary Fig. 31).

Library DNA sequence. The sequence is 5′-CCTGAATTCC**NNNNNNNN**AATGCTCACATCTG**N NNNNNNN**CTCACTCTCGAAAT**NNNNNNNN**CCAAACTGCC-3′, where **N** denotes the mixed base sequences in the three encoding regions. The complete list of building block structures for both libraries is provided in separate files in the Supplementary Information.

Hit compound synthesis and characterization. The assembly of all the peptides was carried out using the Fmoc-based solid-phase peptide synthesis method manually in a glass reaction vessel fitted with a sintered glass frit. The first amino acid was assembled on 2-chlorotrityl resin by adding 2.0 equiv. *N*-Fmoc-protected amino acid (relative to the resin loading) and 4.0 equiv. DIPEA (*N*,*N*′-diisopropylethylamine) in DCM (*N*,*N*′-dimethylformamide) for 4 h. Coupling reactions were performed manually by using 2.0 equiv. *N*-Fmoc-protected amino acid (relative to the resin loading) activated in situ with 2.0 equiv. HTAU (*N*-[(dimethylamino)-1*H*-1,2,3-triazolo-[4,5-b]pyridin-1-ylmethylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide), 2.0 equiv. HOBt (*N*-hydroxylbenzotriazole) and 4.0 equiv. DIPEA in DMF (*N*,*N*′-dimethylformamide) for 4 h. The coupling efficiency in manual synthesis was assessed by TNBS tests. Fmoc-protecting groups were removed by treatment with a piperidine/DMF solution (1:4) for 10 min. The process was repeated three times and the completeness of deprotection verified by ultraviolet absorption of the piperidine washings at 299 nm. Synthetic linear peptides were recovered directly on acid cleavage. Before cleavage, the resin was washed thoroughly with dichloromethane. The peptide was released from the resin using a cleavage solution of 2,2,2-trifluoroethanol/acetic acid/dichloromethane (2:1:7, 2×30 min).

Statistics and reproducibility. All the experiments were performed three or more times with independent samples. Quantitative values are expressed as the mean values ± s.d. The number of replicates and details of statistics are provided in figure legends for all the column and scattered graphs. All the experiments were reproduced with similar results.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data supporting the findings of this study are available within the Article, the associated Source Data files, Supplementary Information and Extended Data files. All the published tools and packages used for data analysis are provided with the paper. The Human UniProt database (release-2016_05) used can be accessed at https://www.uniprot.org/proteomes/UP000005640, and the BioNumbers database can be accessed at https://bionumbers.hms.harvard.edu/. Source data are provided with this paper.

Code availability

The custom Python script for sequencing data analysis is freely available for downloading both as part of Supplementary Information and also at GitHub (https://github.com/cenhuang0916/sequencing-data-processing-script.git)

References

78. Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, W43–W46 (2007).
79. Clark, M. A. et al. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **5**, 647–654 (2009).
80. Kleiner, R. E., Dumelin, C. E., Tiu, G. C., Sakurai, K. & Liu, D. R. In vitro selection of a DNA-templated small-molecule library reveals a class of macrocyclic kinase inhibitors. *J. Am. Chem. Soc.* **132**, 11779–11791 (2010).
81. Mannocci, L. et al. High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc. Natl Acad. Sci. USA* **105**, 17670–17675 (2008).
82. Jiang, R. D., Shen, H. & Piao, Y. J. The morphometrical analysis on the ultrastructure of A549 cells. *Rom. J. Morphol. Embryol.* **51**, 663–667 (2010).
83. Zhang, F. et al. Quantification of epidermal growth factor receptor expression level and binding kinetics on cell surfaces by surface plasmon resonance imaging. *Anal. Chem.* **87**, 9960–9965 (2015).
84. Li, Y., Zhao, P., Zhang, M., Zhao, X. & Li, X. Multistep DNA-templated synthesis using a universal template. *J. Am. Chem. Soc.* **135**, 17727–17730 (2013).
85. Gartner, Z. J. et al. DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**, 1601–1605 (2004).
86. Usanov, D. L., Chan, A. I., Maianti, J. P. & Liu, D. R. Second-generation DNA-templated macrocycle libraries for the discovery of bioactive small molecules. *Nat. Chem.* **10**, 704–714 (2018).

Acknowledgements

This work was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (AoE/P-705/16, 17321916, 17302817, 17301118, 17111319 and 17303220), Laboratory for Synthetic Chemistry and Chemical Biology of Health@InnoHK of ITC, HKSAR, National Natural Science Foundation of China (21572014, 21877093, 81603067, 21907011 and 91953119), the Fundamental Research Funds for the Central Universities (project numbers 2019CDQYYX018 and 2020CQJQY-Z002), Chongqing Research and Frontier Technology (cstc2020jcyj-jqX0009) and Venture & Innovation Support Program for Chongqing Overseas Returnees (cx2019084) for Y.L. We thank the Centre for PanorOmic Sciences (CPOS) Genomics Core at HKU for NGS support and the CQU-Agilent Joint Lab on DNA-encoded Library for MS support.

Author contributions

Y.H., Y.L. and Xiaoyu Li conceived and designed the experiments. Y.H., L.M., Q.N., Y.Z., L.C., S.Y., Xiaomeng Li and C.H. carried out the experiments and analysed the data. Y.M.E.F. carried out the MS experiments and analysis. Y.C. designed and carried out the SPR experiments and analysis. Y.H., L.M., Q.N., Y.Z., L.C., S.Y., Y.M.E.F., Xiaomeng Li, C.H., Y.C., Y.L. and Xiaoyu Li co-wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

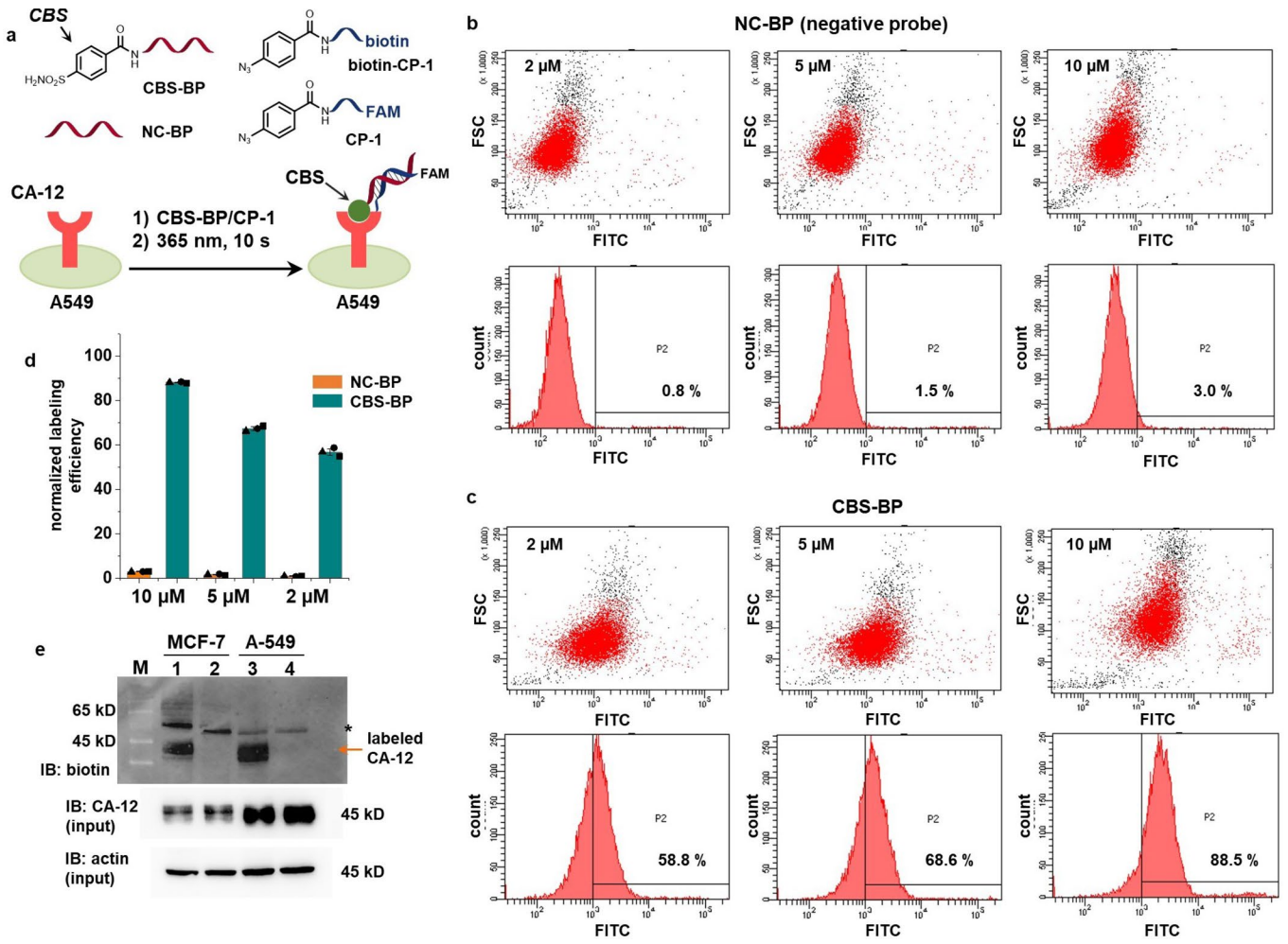
Extended data is available for this paper at https://doi.org/10.1038/s41557-020-00605-x.

Supplementary information is available for this paper at https://doi.org/10.1038/s41557-020-00605-x.

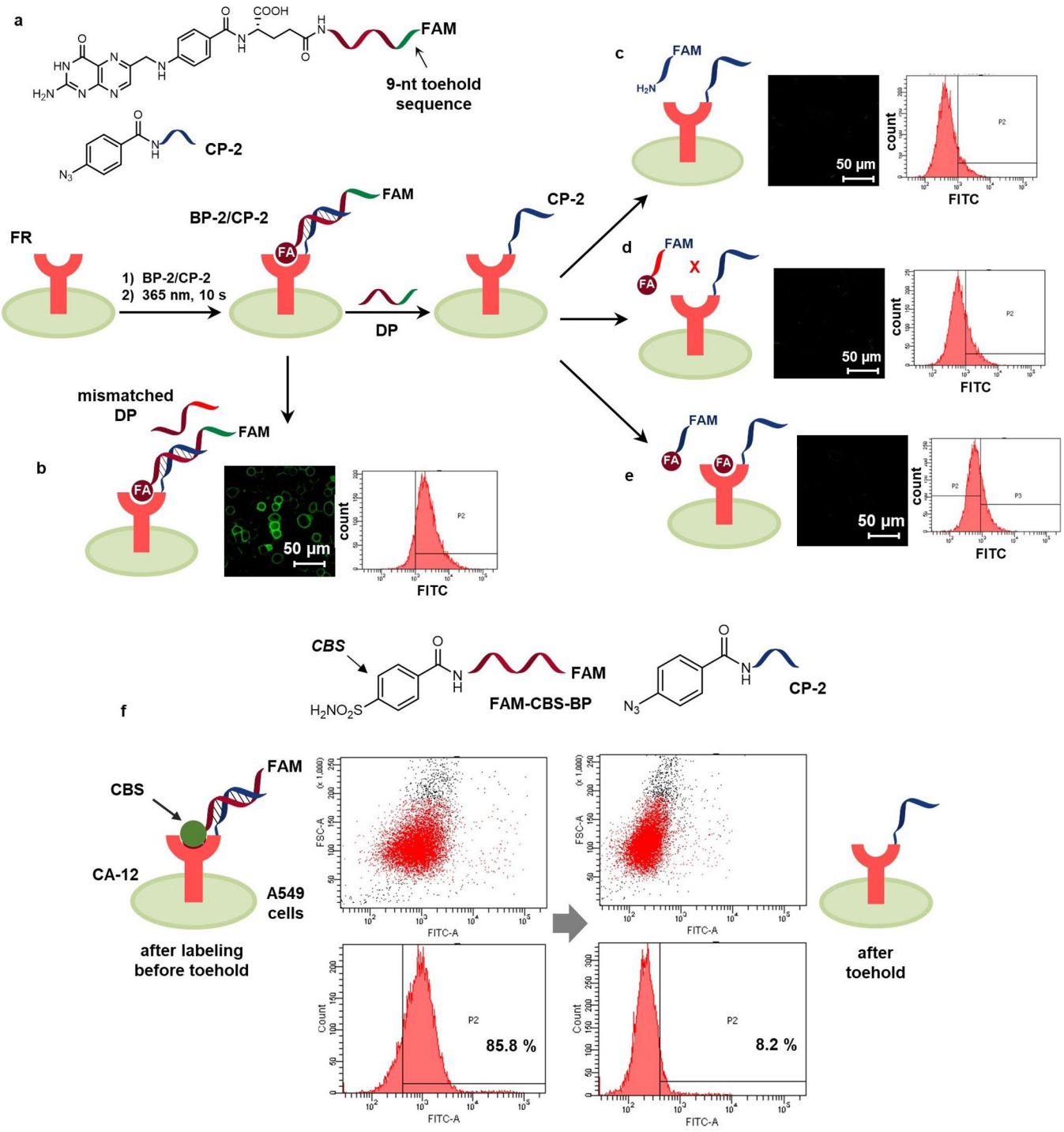
Correspondence and requests for materials should be addressed to Y.C., Y.L. or X.L.

Peer review information *Nature Chemistry* thanks the anonymous reviewers for their contribution to the peer review of this work.

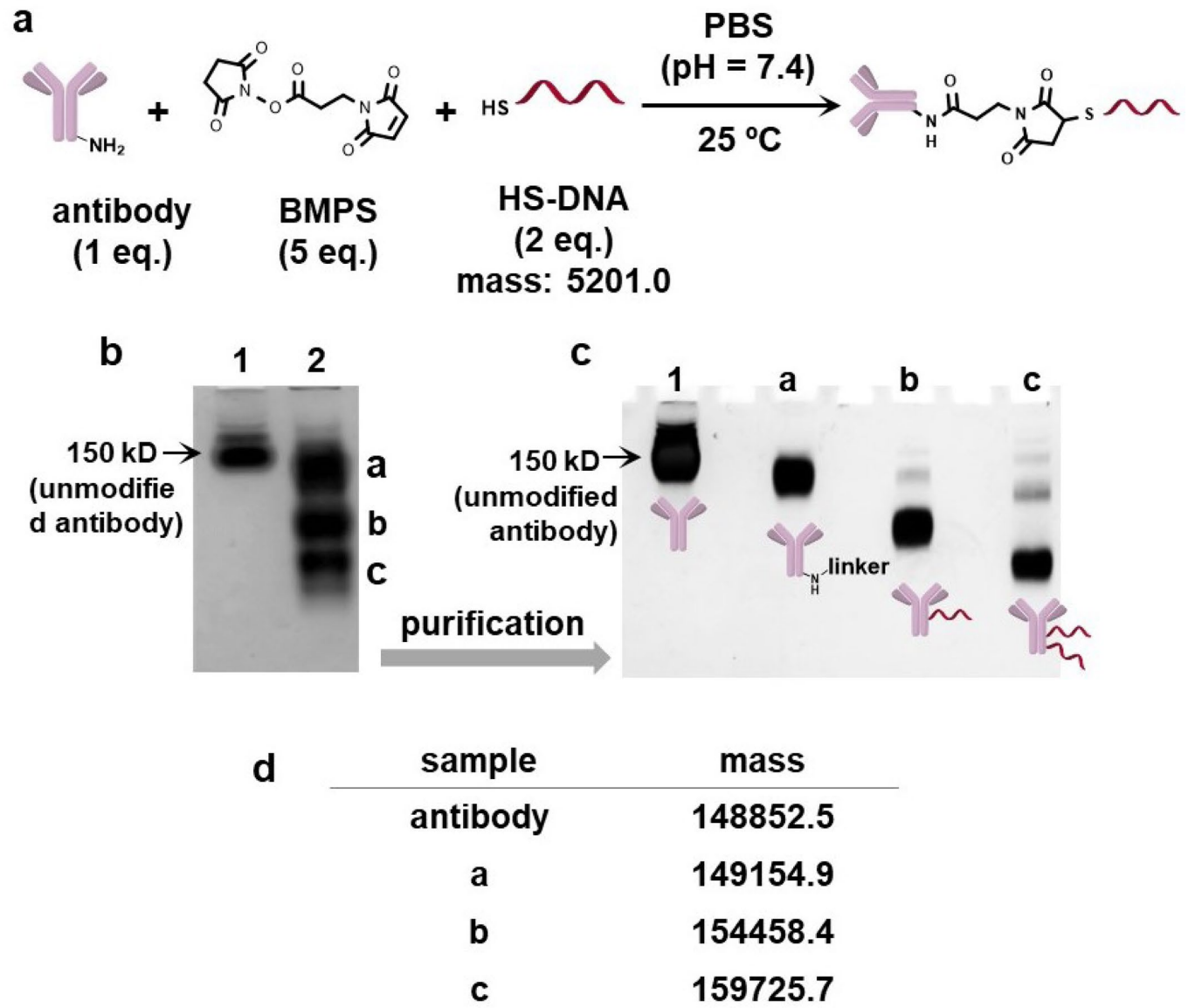
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Labelling of CA-12 on live cells using CBS-guided BP/CP probes. a) Structures of CBS-BP, NC-BP (a negative binding probe), and CP-1. A549 cells were labelled with CBS-BP/CP-1 and NC-BP (no CBS)/CP-1, respectively, targeting CA-12 on the cell surface. Experimental conditions are the same as in Fig. 2 of the main text. **b**)-**c**) Flow cytometry analysis of the labelled A549 cells. **d**) Column graph summarizing the flow cytometry results. **e**) CBS-BP and NC-BP were paired with biotin-CP-1, respectively, and used to label CA-12 on A549 and MCF-7 cells. After labelling, the cells were lysed and the biotinylated proteins were analysed with western blotting. M: marker; lanes 1 and 3: with CBS-BP; lanes 2 and 4: with NC-BP. *: endogenous biotinylated proteins. Loading control: internal endogenous biotinylated proteins, marked with *; a portion of each sample was separately blotted for CA-12 and actin as additional input/sample processing controls. In **d**), $n = 3$ biologically independent samples were measured; data are presented as mean values ± SD (standard deviation).



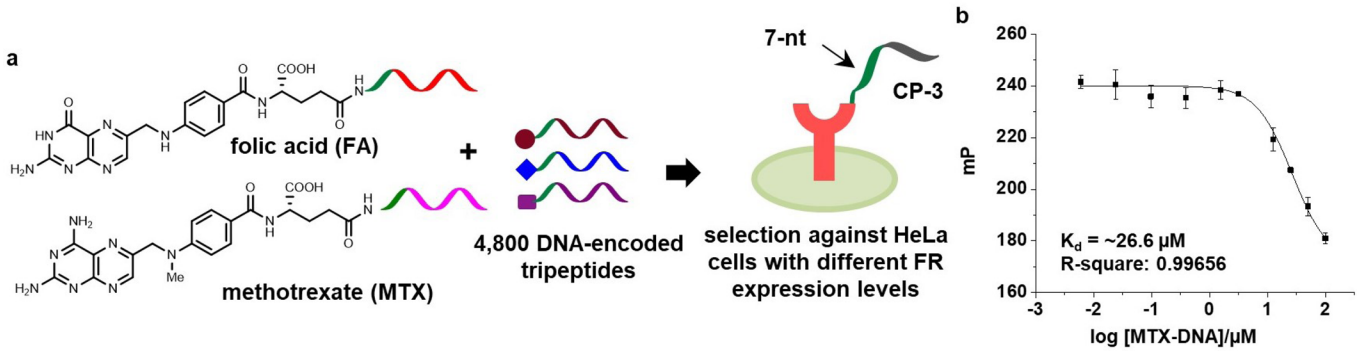
Extended Data Fig. 2 | Additional data on toehold displacement of BP after the labelling of FR on HeLa cells. a) Structures of BP-2 and CP-2 and the labelling scheme. Fluorescent imaging and flow cytometry were used to monitor a series of control experiments. **b**) Toehold displacement with a mismatched DP after labelling. Unlike the complementary DP shown in Fig. 3, the mismatched DP did not reduce cell fluorescence. **c**)-**e**) After toehold displacement with a complementary DP to remove the original BP, a series of control experiments were performed. **c**): with no FA on BP-3; **d**): with a mismatched BP-3; **e**): with free FA competitor (50-fold). Experimental conditions were the same as in Fig. 2 of the main text. f) The same labelling and toehold displacement experiments were performed with the CBS/CA-12 system. Flow cytometry histograms after labelling and before/after DP displacement are shown. Strong fluorescence reduction was observed.



d

sample	mass
antibody	148852.5
a	149154.9
b	154458.4
c	159725.7

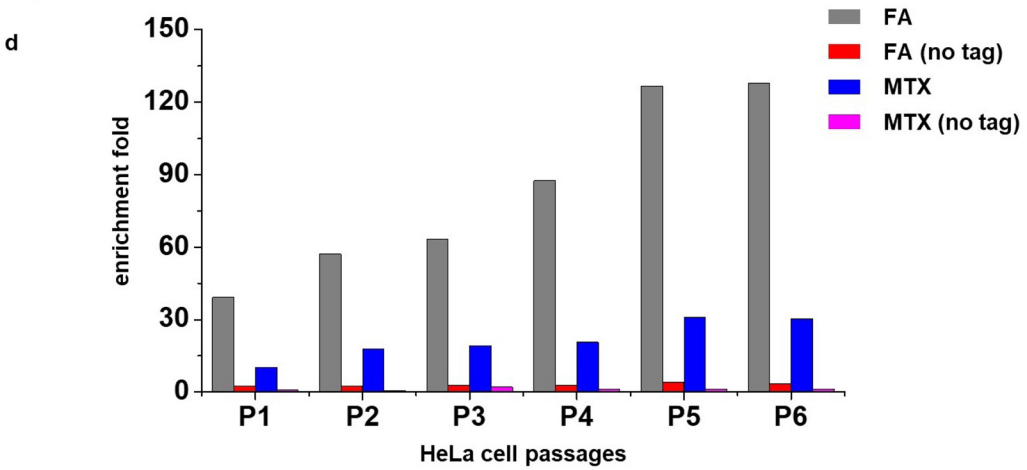
Extended Data Fig. 3 | Preparation of antibody–DNA conjugates. a) The conjugation reaction scheme. **b**) Native PAGE analysis of the reaction. Lane 1, an anti-EGFR antibody standard; lane 2: the reaction mixture. Marker is based on unmodified antibody. **c**) The bands a–c in **b**) were purified and analysed with native PAGE (polyacrylamide gel electrophoresis). **d**) The unlabelled antibody and purified bands a–c were characterized with ESI-MS; the results confirmed that conjugate b was the mono-DNA–antibody conjugate.



c

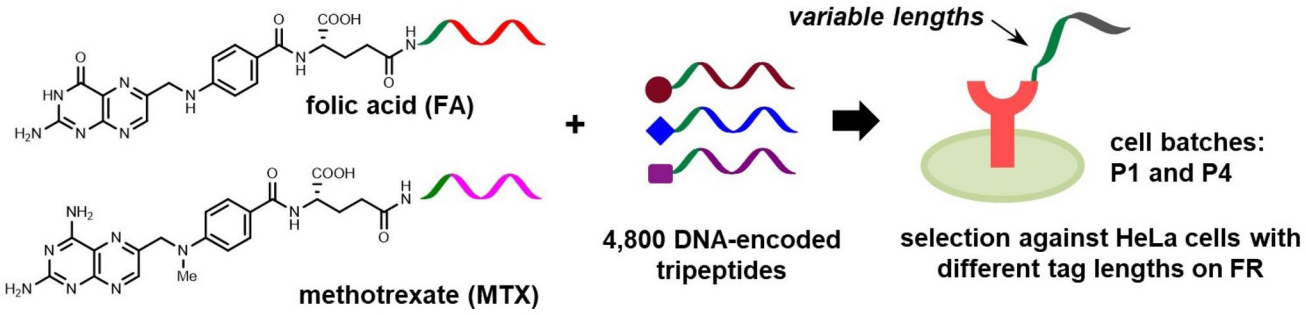
batch	# of DNA	cell size (L)	[FR] (nM)	EF (FA)	EF (MTX)	EF (FA) no tag	EF (MTX) no tag
P1	61,000	3.0e-12	33.8	39.2	10.3	2.5	1.1
P2	115,000	3.0e-12	63.7	57.1	17.9	2.6	0.7
P3	192,000	3.0e-12	106.3	63.2	19.1	3.0	2.1
P4	602,000	3.0e-12	333.2	87.5	20.6	2.7	1.4
P5	1,086,000	3.0e-12	601.1	126.7	31.0	4.1	1.2
P6	1,336,000	3.0e-12	739.5	127.8	30.5	3.6	1.3

* tag length: 7-nt



Extended Data Fig. 4 | DEL selection against HeLa cells with different FR expression levels. a) Besides FA, another FR ligand, methotrexate (MTX) was conjugated with a DNA strand and added to the 4,800-member library as shown in Fig. 5. This 4,802-member DEL was selected against the DNA-tagged FR on HeLa cells with different FR levels; a tag with 7-nt complementarity was used in the selections. **b**) FP analysis to measure the binding affinity of the MTX–DNA conjugate to the target protein FR, and a K_d of ~26.6 μM was obtained. n = 3 biologically independent FP samples were measured. Data are presented as mean values ± SD (standard deviation) based on biologically independent replicates. **c**) HeLa cells were cultured in FA-deficient medium and harvested after different passages. Six batches of the cells with different FR expression levels were fluorescently labelled with BP-1/CP-1 and analysed with flow cytometry for each batch. The average number of DNA molecules on each cell was measured; based on a size of 3,000 cubic μm for HeLa cells, the FR concentration for each cell batch was calculated (200 μL selection volume; see Section 9 for calculation method). The 4,802-member DEL was subjected to the same selection procedure as described in Fig. 5,6 against these cell batches, respectively; the selection results were processed also in the same way and summarized in the table. EF: enrichment factor. For each selection, a control without FR tagging was also conducted. **d**) Column graph summarizing the selection results shown in **c**).

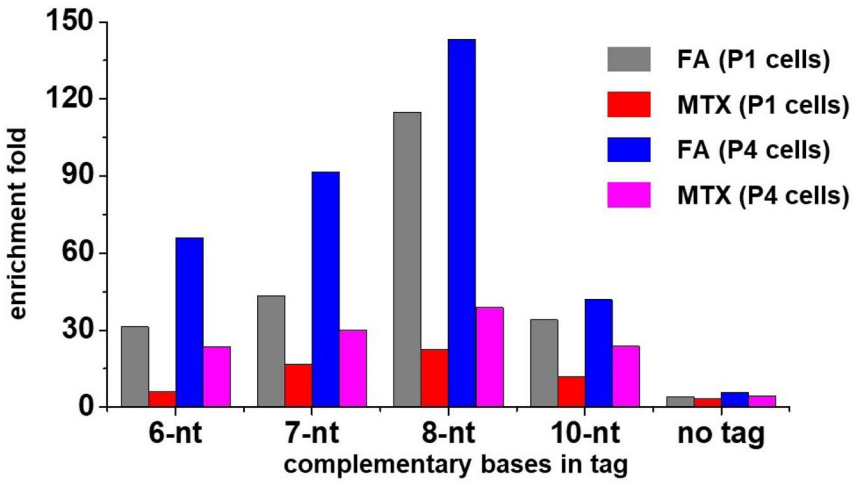
a



b

complementary bases in tag	sequence	ΔH (kcal/mol)	ΔS (cal/K.mol)	ΔG (kcal/mol)	fold of affinity increase	K_d/μM (tag/library DNA)
6-nt	CCTGAA	-42.46	-121.42	-5.65	11,835	84.5
7-nt	CCTGAAT	-42.93	-122.04	-5.93	18,838	53.1
8-nt	CCTGAATT	-50.46	-144.25	-6.73	71,081	14.07
9-nt	CCTGAATTC	-57.13	-158.51	-9.08	3,514,830	0.285
10-nt	CCTGAATTCC	-70.99	-202.19	-9.70	9,675,250	0.103

c

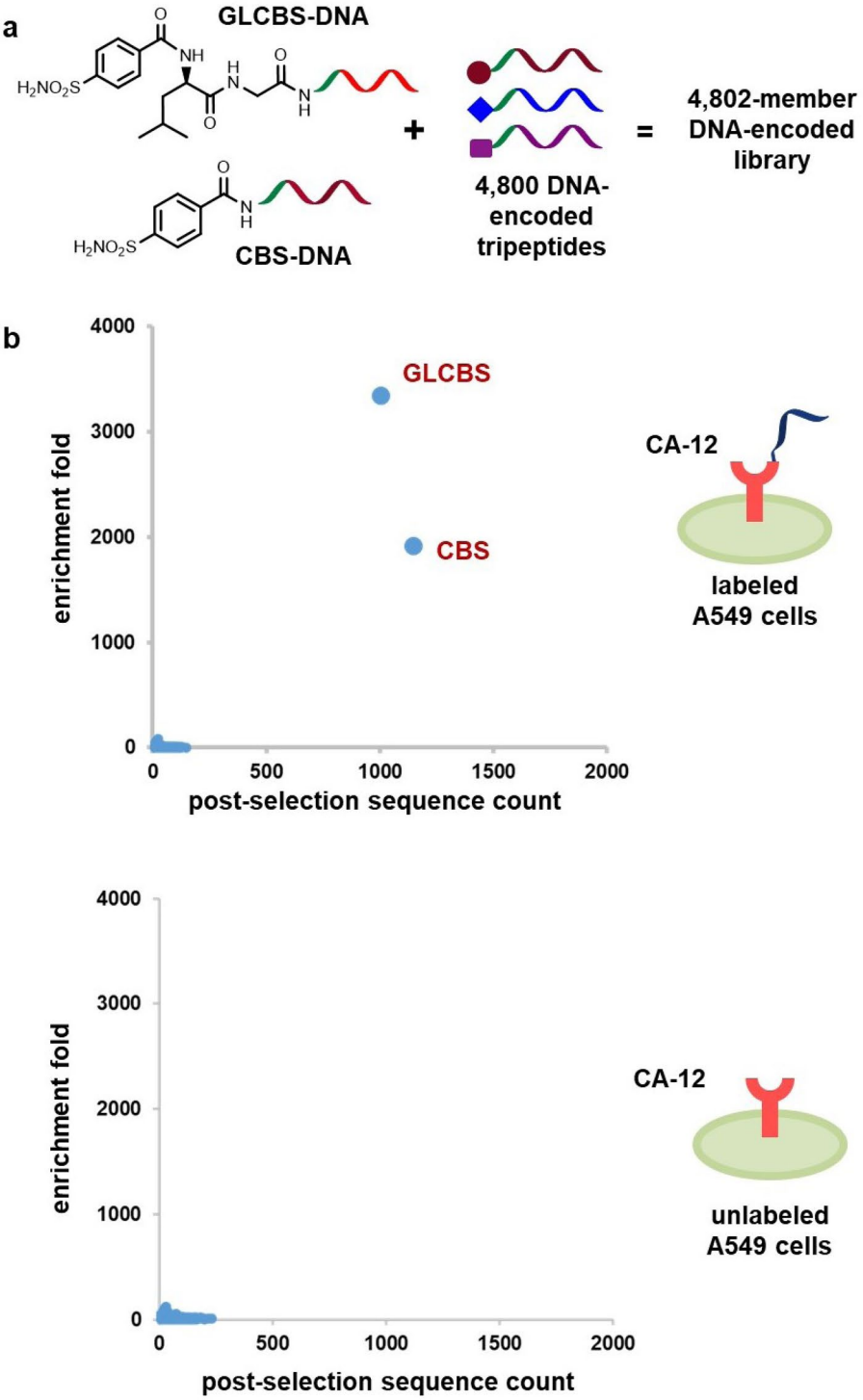


d

complementary bases in tag	sequence	P1 cells		P4 cells	
		EF (FA)	EF (MTX)	EF (FA)	EF (MTX)
6-nt	CCTGAA	31.3	5.92	66.0	23.5
7-nt	CCTGAAT	43.4	16.7	91.5	30.1
8-nt	CCTGAATT	114.9	22.6	143.2	38.9
10-nt	CCTGAATTCC	34.0	11.8	41.9	23.7
no tag	--	2.2	1.1	2.4	1.0

Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | DEL selection against HeLa cells labelled with the tags with different lengths of complementary bases. a) The 4,802-member DEL was selected against the DNA-tagged FR on HeLa cells with different lengths of complementary bases in the tag. Two cell batches (P1 and P4) were used in the selections. The selection procedure and data processing method were the same as in Figs. 5–6. **b**) The effects of different DNA tag lengths were calculated and summarized in the table; key parameters include: ΔH, ΔS, ΔG, fold of affinity increase, and K_d of the DNA tag/library DNA duplex. **c**)-**d**) Column graph and the table summarizing the selection results; EF: enrichment factor. The tag lengths from 6 to 10 bases corresponded to a free energy gain from 5.65 to 9.70 kcal/mol and an affinity increase of ~11,000 to 9-million folds. At 6- and 7-nt, the tag hybridized with library DNA at μM affinity; at 9- and 10-nt, the tag and the library DNA formed stable duplexes, which would increase the affinity of all library compounds to nM binders (K_d: 285 nM and 103 nM, respectively). The results also showed that the enrichment fold of FA and MTX dropped with the 10-nt tag, but many other library compounds were enriched. We reasoned this might be because the 10-nt tag formed stable DNA duplex with all library compounds and resulted in the enrichment of many low-affinity binders. DNA tag shorter than 6-base was not tested because it would have hybridization specificity issue at such a short length; the tag may hybridize with the other regions of the library DNA, instead of the primer-binding site.



c

	# of DNA/cell	cell size/L	[CA-12] /cell (nM)	suspension volume (μL)	[CA-12] in suspension (nM)
CA-12 on A549 cells	3.28×10^5	1.75×10^{-12}	312.3	200	29.8

Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | DEL selection against the DNA-tagged CA-12 on live cells. a) Structures of GLCBS–DNA and CBS–DNA, which were two positive controls added to the 4,800-member DEL for selection against CA-12 on A549 cells. **b**) Scatter plots of the selection results of the tagged A549 cells (top) and the untagged cells (bottom). The selection experiment condition and data processing protocol are the same as in Fig. 7. *x*-axis: post-sequencing counts; *y*-axis: enrichment fold = (post-selection %)/(pre-selection %) of each compound. The positive controls (GLCBS and CBS) are highlighted. **c**) Calculation of the DNA-tagged CA-12 concentration on A549 cells. The average number of DNA on each cell was determined with flow cytometry.

nature research

Corresponding author(s): Xiaoyu Li

Last updated by author(s): Sep 19, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Image Lab Software (version 2.2); PerkinElmer multi-label plate reader software (VICTOR X5); Flow Cytometer and Cell Sorter Workstation (BD FACS AriaIII) software; Confocal laser microscope (Carl Zeiss LSM 710 NLO) software; EVOS FL Cell Imaging System; NGS sequencer (Illumina® NextSeq 500 sequencer); Mass spectrometer (LCQ Orbitrap and LCQ Orbitrap Fusion)
Data analysis	Excel 2016 and 2019; Python (version 2.7.13); Origin Pro (version 8.6); ImageJ (version 1.51k); Biacore T200 evaluation software (version 2.0); FlowJo X 10.0.7 software; Zen 2.3 (blue edition); GraphPad Prism (version 8); DTASelect (version 2.0.47); ProLuCID algorithm.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that all data supporting the findings of this study are available within the article, the associated source data files, and the Supplementary Information figures and files. All data are also available from the corresponding author upon reasonable request. The author declare that all the published tools and packages used for data analysis have been provided with the paper. The Human UniProt database (release-2016_05) used can be accessed at: https://www.uniprot.org/proteomes/UP000005640, and the BioNumbers database can be accessed at: https://bionumbers.hms.harvard.edu/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Experiments were performed in biological triplicate n=3. No statistical tests were used to determine sample size. Sample sizes for experiments were determined based on previous studies with related experiments and we have determined that this sample size to be sufficient to ensure reproducibility and significance.
Data exclusions	No data were excluded from analysis.
Replication	Experiments were repeated in triplicates and we confirm that all attempts at replication were successful with similar results.
Randomization	Randomization was not relevant for this study. This manuscript did not involve any population study, and the techniques used to in the experiments are inherently unbiased and no further randomization was needed. Different cell batches were used for biological replicate in cell-based experiments.
Blinding	Not applicable for this study. All samples were processed identically through standard protocols and procedures, and automated procedures was used for DNA synthesis and DNA sequencing, which should not bias any specific outcomes by knowledge about sample identity.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*

Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	A full list of the antibodies with vendor, catalog number, dilution, and antigen information is provided in the Supplementary Information.
Validation	For the anti-FR antibody (Sino Biologics; 81073-T40, immunogen: recombinant rat folate receptor, Met1-Met231), the supplier validated the antibody with the extracts from K562 cells and rat kidney; validated applications: WB and IHC-P. Relevant citation: Senol S, Ceyran AB, Aydin A, et al. Int. J. Clin. Exp. Pathol. 2015; 8(5):5633-5641. For the anti-α tubulin antibody (Abcam; ab7291, immunogen: full length native protein corresponding to chicken alpha tubulin; aa 426-450), the supplier validated the antibody with the extracts from A549 cells, whole HeLa cells and human heart muscle tissue; validated applications: Flow Cyt, ICC/IF, IHC-P, WB. Relevant citations: Dirks ML et al. J. Physiol. 598:123-137 (2020); Y an J et al. J. Cell

Mol. Med. 24:814-829 (2020); Miao L et al. Theranostics 10:498-515 (2020).
For the anti-gamma actin antibody (Abcam; ab194952, immunogen: synthetic peptide within gamma actin aa 188-215 conjugated to Keyhole Limpet Haemocyanin), the supplier validated the antibody with the extracts from HeLa, C32, Sol8, and NIH/3T3 cells, and human upper stomach tissue; validated applications: WB, ICC/IF, Flow Cyt, IHC-P. Relevant citations: Williams RC Jr. et al., 1999. Anal Biochem. 275(2): 265-7; Nogales E. et al., 1998. Nat Struct Biol. 5(6): 451-8; Dutcher SK. 2001. Curr. Opin. Cell Biol. 13(1): 49-54.
For the anti-transferrin receptor antibody (Sino Biologics; 11020-MM04; immunogen: recombinant human transferrin receptor (TFRC), extracellular domain, Cys89-Phe760), the supplier validated the antibody with whole U937 and HeLa cells; validated applications: FCM, ICC/IF. Relevant citations: Douabin-Gicquel V., et al., 2001, Hum. Genet. 109:393-401; Ryschich, E. et al., 2004, Eur. J. Cancer. 40 (9):1418-22; Tosoni D., et al., 2005, Cell 123:875-888; Wollscheid B., et al., 2009, Nat. Biotechnol. 27:378-386.
For the anti-CD44 antibody (Sinobiologics, 12211-MM10, immunogen: recombinant human CD44, extracellular domain, Met1-Pro220), the supplier validated the antibody with the extracts from HeLa cells, whole HeLa cells, human skin and esophagus; validated applications: WB, ELISA, IP. Relevant citations: Bajorath J. et al. Proteins. 39(2): 103-11 (2000); Johnson P, et al. Biochem. Pharmacol. 59(5): 455-65 (2000); Martin TA, et al. The role of the CD44/ezrin complex in cancer metastasis. Crit. Rev. Oncol. Hematol. 46(2): 165-86 (2003); Johnson P, et al. CD44 and its role in inflammation and inflammatory diseases. Inflamm. Allergy Drug Targets. 8(3): 208-20 (2009).
For the anti-CA-12 antibody (Sinobiologics, 10617-RP02, immunogen: recombinant human carbonic anhydrase XII Protein; Met 1-Gln 291), the supplier validated the antibody with the extracts from human prostate carcinoma, colon, and kidney cells; validated applications: WB, IHC-P. Relevant citations: Sahin, U. et al., Proc. Natl. Acad. Sci. U.S.A. 92 (25): 11810–11813 (1996); Ivanov, S.V. et al., Proc. Natl. Acad. Sci. USA 95:12596 – 12601 (1998); Strausberg, R.L. et al., Proc. Natl. Acad. Sci. USA 99:16899 – 16903 (2002); Liao, S.Y. et al., J. Med. Genet. 40:257 – 262 (2003); Supuran, C. T. et al., Curr. Pharm Des. 14 (7): 601-602 (2008); Elleuche, S. et al., Curr. Genet. 55 (2): 211-222 (2009).
For the anti-EGFR antibody (Sinobiologics, 10001-MM08T, immunogen: recombinant human EGFR, extracellular domain, Met1-Ser645), the supplier validated the antibody with purified recombinant EGFR protein, extracellular domain, live A431 cells and cell extracts; validated applications: FCM, ICC/IF. Relevant citations: Schlessinger, J. Cell 103(2): 211-25 (2000); Giaccone, G. Ann. Oncol. 16(4): 538-48 (2005); Yarden, Y., et al. Nat. Rev. Mol. Cell. Biol. 2(2): 127-37 (2001).
We validated the anti-FOLR antibody, anti-tubulin antibody with the extracts from HeLa cells. We validated the anti-transferrin receptor antibody with MCF-7 cells. We validated the anti-CD44 antibody with PC-3 and U-87 MG cells. We validated the anti-EGFR unconjugated affibody with A-431 cells. We validated the anti-CA-12 antibody with A-549 cells.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HeLa (CCL-2), A549 (CCL-185), MCF-7 (HTB-22), A-431 (CRL-1555), B16-F10 (CRL-6475), A-375 (CRL-1619), HUVEC (CRL-1730), PC-3 (CRL-1435), and U-87 MG (HTB-14). All cell lines were purchased from ATCC.
Authentication	None of the cell lines used in this study were authenticated by the authors.
Mycoplasma contamination	The cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature,*

Field-collected samples	*photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	The cells were harvested and washed three times with 1x PBS buffer by gentle resuspension and centrifugation at 500 × g for 5 min. Then the cells were treated with corresponding probes and underwent UV irradiation on ice at 365 nm for 10 s. After further wash and fixation steps, the cells were resuspended for flow cytometry analysis.
Instrument	Flow Cytometer and Cell Sorter Workstation (BD FACS AriaIII).
Software	BD FACS AriaIII software and FlowJo X 10.0.7 software.
Cell population abundance	No sorting was involved in the study, so pre-/post-sorting cell population abundance is not relevant. The BD FACS AriaIII instrument sets a cell counting limit at 10,000 cells per experiment, and a typical 85%-95% of cells was obtained above the gating threshold.
Gating strategy	The FSC/SSC population was determined by the densest clustering part of the FSC-SSC scattered diagram, abandoning the scatters deviating from the part. The gate between “positive” and “negative” was usually set at 1000 (a.u.) on the FITC axis, but would also be set on the overlapping point of the two peaks to show the shift of the peaks between “positive” and “negative” in certain cases. A figure graphically showing the gating strategy for all FACS data is provided in the Supplementary Information.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis | *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*