Check for updates

RESEARCH ARTICLE SUMMARY

APTAMERS

SPARK-seq: A high-throughput platform for aptamer discovery and kinetic profiling

Guoyan Luo, Jia Song, Yongbo Fu, Yuanjie Jiang, Yuan Gao, Zhixing Zhong, Ling Li, Yong Wei, Hao-Ran Jia, Lu Guo, Ting Fu, Qin Wu*, Weihong Tan*

Full article and list of author affiliations: https://doi.org/10.1126/science.adv6127

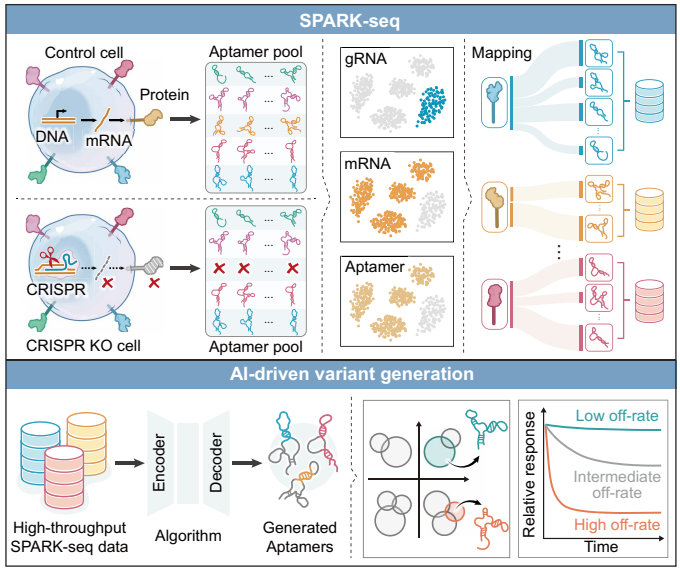
INTRODUCTION: Cell surface proteins represent the majority of clinically actionable targets and are vital to cellular communication, signaling, and homeostasis. However, methods for generating high-affinity molecular probes such as aptamers against these targets remain limited. Current approaches are low-throughput and often compromise native protein conformations, leaving most of these targets underexplored and hindering aptamer-based diagnostic and therapeutic development.

RATIONALE: We hypothesized that a multimodal approach combining CRISPR-based genetic perturbation with single-cell multiomics could overcome the limitations of conventional aptamer screening. By concurrently profiling genetic perturbations, gene expression, and protein binding within one single cell, we established single-cell perturbation-driven aptamer recognition and kinetics sequencing (SPARK-seq), a platform that enables high-throughput mapping of aptamer-target interactions in native cellular contexts. This integrated design not only facilitates the identification of binders to low-abundance targets but also directly couples discovery with kinetic profiling, accelerating the development of precision molecular tools.

RESULTS: We applied SPARK-seq to systematically interrogate aptamer-target interactions by combining multiplexed CRISPR knockout with single-cell mRNA and aptamer sequencing. Following four rounds of enrichment with Cell-SELEX (systematic evolution of ligands by exponential enrichment), the aptamer library was screened against a pooled population of cells subjected to CRISPR knockout of 13 surface proteins. Single-cell sequencing enabled simultaneous detection of gRNAs, transcriptomes, and aptamer binding events, linking genetic perturbations to aptamer-binding profiles.

Using our computational pipeline single-cell perturbation-aptamer recognition and targeted aptamer-generation algorithm (SPARTA), we analyzed 8466 high-quality single cells and clustered the top 10,000 unique aptamer sequences into 1906 families. Differential binding analysis identified 5535 aptamer sequences targeting eight distinct surface proteins. Orthogonal validation using flow cytometry, surface plasmon resonance, and microscale thermophoresis confirmed target specificity and nanomolar affinities. The approach also resolved complex binding patterns, including an integrin-targeting aptamer recognizing ITGA3/ITGB1 and a single aptamer binding multiple PTPR paralogs.

SPARK-seq preferentially enriched aptamers with slow dissociation rates (k_{off}). Across multiple targets, binding difference ($-\log_2$FC) correlated strongly with k_{off} ($R^2 = 0.8$ to 0.9) but not with equilibrium affinity (K_D), underscoring the platform's capacity to isolate aptamers with prolonged target engagement, a critical feature for therapeutic and diagnostic applications. Leveraging these data, SPARTA's convolutional neural network–based classifier achieved approximately ~97% accuracy in predicting PTK7-binding sequences, and its generative module produced functional variants with improved kinetic profiles.



Spark-seq for aptamer discovery and variant design. A CRISPR-based Cell-SELEX workflow identifies aptamers by differential binding between control and knockout cells, followed by large-scale profiling of aptamer target interactions through single-cell sequencing. A deep learning framework then predicts and designs aptamer variants with improved binding kinetics.

CONCLUSION: SPARK-seq is an integrated platform that unites CRISPR-mediated genetic perturbation, single-cell transcriptomics, and sequence-based aptamer profiling, augmented by the deep learning framework SPARTA. This work establishes a foundation for "Aptomics," a multi-omics enabled framework for the large-scale, sequencing-driven decoding of aptamer binding specificities and target properties. By converting millions of aptamer-binding events into high-dimensional sequencing data anchored to genetic and transcriptional states, SPARK-seq enables direct genotype-phenotype-ligand mapping across thousands of individual cells. This multimodal strategy expands the dynamic range for detecting low-abundance and conformation-sensitive targets, resolves kinetic and enrichment patterns at sequence resolution, and supports rapid, scalable discovery and rational optimization of high-specificity aptamers, paving the way for advanced diagnostic and therapeutic applications. □

*Corresponding author. Email: tan@him.cas.cn (W.T.); wuqin@him.cas.cn (Q.W.) Cite this article as G. Luo *et al.*, *Science* 391, eadv6127 (2026). DOI: 10.1126/science.adv6127

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

APTAMERS

SPARK-seq: A high-throughput platform for aptamer discovery and kinetic profiling

Guoyan Luo[1,2]†, Jia Song[3,4]†, Yongbo Fu[3,5], Yuanjie Jiang[1,6], Yuan Gao[4], Zhixing Zhong[3,4], Ling Li[1], Yong Wei[1], Hao-Ran Jia[1,4], Lu Guo[1], Ting Fu[1], Qin Wu[1,7]*, Weihong Tan[1,2,4]*

Cell surface proteins are key disease biomarkers and therapeutic targets, yet high-throughput methods for aptamer discovery targeting these proteins in situ remain limited. We introduce single-cell perturbation-driven aptamer recognition and kinetics sequencing (SPARK-seq), a high-throughput platform integrating single-cell messenger RNA and aptamer sequencing with CRISPR-based surface protein perturbation. In a single experiment, SPARK-seq simultaneously mapped 5535 distinct aptamers to eight surface proteins, capturing interactions across more than two orders of magnitude in protein abundance and spanning diverse biophysical classes. The method discriminated closely related paralogous proteins with no detectable cross-reactivity and provided kinetic information that enabled the prioritization of aptamers with slow dissociation rates. Leveraging this kinetic diversity, we engineered variants with improved off-rate properties. SPARK-seq establishes a platform for high-efficiency discovery and rational variant design of aptamers and functional nucleic acids, unlocking possibilities in diagnostics and therapeutics.

Identified through the systematic evolution of ligands by exponential enrichment (SELEX) technique, aptamers are short, single-stranded DNA or RNA molecules, typically 20 to 100 nucleotides (nt) in length, with notable target specificity (*1*, *2*). To date, SELEX has produced thousands of aptamers against diverse targets, including small molecules, proteins, bacteria, viruses, and live cells (*3*-*6*). Unlike antibodies, which require in vivo generation, aptamers are selected in vitro, allowing for highly customized design. Aptamers are nucleic acids and thus provide advantages such as facile synthesis, precise chemical modification, and high stability across experiments (*6*). These characteristics have expanded the applications of aptamers from basic research to clinical diagnostics (*7*), targeted drug delivery (*8*), and molecular imaging (*9*). Aptamers have thus become indispensable tools in advancing precision medicine and innovative therapeutic strategies, offering transformative potential across scientific and clinical applications.

Traditional SELEX, however, favors homogeneous proteins, limiting its use for cell surface proteins such as receptors, transporters, and receptor tyrosine kinase proteins, key disease biomarkers and targets of more than 60% of therapeutic drugs (*10*, *11*). This limitation arises from challenges in preparing cell surface proteins in a homogeneous form (*12*). To address this, cell-based selection methods have been developed using whole living cells as targets. This approach provides a key advantage by selecting aptamers within their natural cellular environment, enabling the identification of aptamers that bind to cell surface proteins in their native conformations (*13*). Furthermore, Cell-SELEX facilitates the discovery of new biomarkers without prior knowledge of target proteins, offering a distinct advantage over traditional protein-based screening methods.

Despite the widespread adoption of Cell-SELEX, efficiently identifying the molecular targets of cell-specific aptamers remains a major challenge (*14*). The current standard method involves aptamer immunoprecipitation followed by mass spectrometry (Apt IP-MS) whereby proteins in a sample are first captured by specific aptamer binding through aptamer immunoprecipitation. The captured proteins are then analyzed by MS to determine identity and quantity (*15*, *16*). Typically, aptamers for IP-MS validation are selected empirically based on their high copy numbers in the enriched library, a process that usually requires multiple rounds of selection (generally >10) to achieve sufficient enrichment, demanding substantial time and resources (*17*, *18*). Because this empirical approach favors highly abundant aptamer sequences, the identified target proteins are often nucleic acid–binding proteins or those with high expression levels, whereas low-abundance proteins are frequently overlooked (*15*, *19*–*22*). This limitation slows the discovery process and narrows the range of potential targets. Additionally, polymerase chain reaction (PCR) biases during selection rounds can result in the overrepresentation of sequences with suboptimal binding properties, potentially excluding aptamers with superior binding affinity and specificity (*17*, *20*, *22*). More importantly, the current low-throughput method identifies only one or two aptamers per experiment, leaving millions of aptamer sequences unexplored and severely limiting the efficiency of biomarker discovery (*14*, *19*).

To address these challenges, we leveraged the sequenceable nature of aptamers to develop single-cell perturbation-driven aptamer recognition and kinetics sequencing (SPARK-seq), a high-throughput method for aptamer target identification that integrates CRISPR-based gene perturbations with single-cell mRNA and aptamer sequencing (*23*–*25*). This approach enables concurrent perturbation of multiple genes, alongside in-depth profiling of gene expression changes and cell surface aptamer binding across thousands of individual cells. Then, leveraging the extensive aptamer-target interaction data generated by SPARK-seq, we developed a deep learning–based framework, termed single-cell perturbation-aptamer recognition and targeted aptamer-generation algorithm (SPARTA), for aptamer prediction and variant generation. SPARTA is a specialized algorithm that pairs aptamer-binding data with guide RNA (gRNA) perturbation and cellular transcriptomic profiles, allowing us to uncover specific relationships between distinct aptamer sequences and gene-knockout cell populations, pinpointing potential target antigens.

Using SPARK-seq, we accurately mapped the interactions between 5535 distinct aptamers and eight surface proteins in a single experiment. Beyond mapping aptamer-target pairs, this method enables precise tracking of copy number variations and the enrichment patterns of aptamer sequences influenced by individual target perturbations at single-cell resolution. This feature facilitates the rapid identification of aptamers with diverse binding affinities, dissociation rates, and specificity. Compared with traditional target identification methods, SPARK-seq allows for simultaneous identification of multiple target proteins and aptamer sequences, facilitating high-throughput screening of optimal aptamers and multiplexed biomarker discovery on the cell surface. Then, using SPARTA, we can accurately predict the binding potential of aptamers to specific proteins and generate aptamer variants tailored to individual targets. Together, SPARK-seq and SPARTA represent a powerful, integrated platform that accelerates

[1]Department of Pathology, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang, P.R. China. [2]Molecular Science and Biomedicine Laboratory (MBL), State Key Laboratory of Chemo and Biosensing, College of Chemistry and Chemical Engineering, College of Biology, Aptamer Engineering Center of Hunan Province, Hunan University, Changsha, Hunan, P.R. China. [3]Institute for Developmental and Regenerative Cardiovascular Medicine, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, P.R. China. [4]Institute of Molecular Medicine (IMM), Renji Hospital, Shanghai Jiao Tong University School of Medicine, and College of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai, P.R. China. [5]School of Pharmaceutical Sciences, Xiamen University, Xiamen, Fujian, P.R. China. [6]Department of Chemistry, University of Science and Technology of China, Hefei, Anhui, P.R. China. [7]School of Molecular Medicine, Hangzhou Institute for Advanced Study, UCAS, Hangzhou. Zhejiang. P.R. China. ***Corresponding author. Email: tan@him.cas.cn (W.T.); wuqin@him.cas.cn (Q.W.)** †These authors contributed equally to this work.

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

aptamer discovery, paving the way for more precise targeting strategies in therapeutic and diagnostic applications.

SPARK-seq workflow

SPARK-seq integrates multiplexed CRISPR-mediated gene inactivation with single-cell RNA sequencing, allowing for comprehensive profiling of gene expression phenotypes resulting from each specific perturbation (*23*, *26*). To adapt this platform for identifying aptamers that target cellular phenotypes affected by specific gene inactivation, we first designed an aptamer library compatible with single-cell sequencing. This library contains a central 46-nt randomized region flanked by a 32-nt 5′ PCR handle and a 22-nt 3′ capture sequence (Fig. 1A). Through four rounds of Cell-SELEX, we enriched the library for aptamers that bind specifically to targeted cells (Fig. 1B).

Next, the enriched aptamer library was incubated with a mixed-cell population, each cell having undergone CRISPR knockout targeting distinct surface proteins and then subjected to 10X Genomics 5′-end sequencing. In this assay, gRNAs, aptamers, and mRNA were captured through a CRISPR capture primer, template-switch oligo (TSO), and poly(T) sequences (T denotes thymidine), respectively, before processing for single-cell sequencing (Fig. 1C).

Following the sequencing step, cell barcodes from each omics layer were identified and corrected to map each cell and its corresponding perturbation to an aptamer-binding profile (Fig. 1D). Unique DNA barcodes assigned to each gRNA and aptamer allowed the deconvolution of pooled perturbations and aptamer libraries, thereby enabling precise linkage of cell phenotypic outcomes to both gRNA and aptamer sequences (Fig. 1E). Meanwhile, we validated the reliability and reproducibility of aptamer sequencing using three spike-in sequences (~6% error rate, fig. S1A) and triplicate runs (>99% reproducibility across the top 10,000 unique sequences; fig. S1, B to E), respectively.

To investigate the effects of each perturbation on the binding of specific aptamer families and uncover potential cell surface protein aptamer interactions, we developed the SPARTA algorithm. This tool integrates Markov clustering (MCL) and statistical methods to identify aptamer families and analyze their binding dynamics under perturbations. This workflow ultimately enabled high-throughput identification of aptamer interactions with cell surface antigens and simultaneous discovery of multiparameter cellular markers.

Proof-of-concept of SPARK-seq

To verify that SPARK-seq can effectively establish a direct association among gene knockout, cell phenotype, and aptamer binding, we conducted a proof-of-concept study using single-gene knockout cell lines incubated with a single specific aptamer sequence. We selected cell surface protein tyrosine kinase 7 (PTK7) for CRISPR-based knockout and used PTK7-specific aptamer sgc8c to demonstrate the capabilities of SPARK-seq (*13*, *27*).

To increase cellular diversity within a single SPARK-seq run, we employed distinct gRNAs (data S1) targeting *PTK7* to perform knockout in both human (SUM159PT) and mouse (EMT6) cell lines (*28*, *29*). Results confirmed a marked reduction of *PTK7* expression levels in both cell lines (Fig. 2A and fig. S2A), accompanied by a substantial decrease in sgc8c binding in PTK7-knockout cells (Fig. 2B and fig. S2B). Furthermore, flow cytometry confirmed that sgc8c reliably discriminated between PTK7-expressing and PTK7-knockout cells, demonstrating the aptamer's specificity for PTK7-positive cell populations (fig. S2C). To adapt sgc8c for compatibility with single-cell sequencing, we modified it by appending a 43-nt capture sequence (equal to that of the gRNA) at the 3′ end and 21-nt at the 5′ end, creating sgc8c-S (fig. S2D). Surface plasmon resonance (SPR) binding assay confirmed that modified sgc8c-S maintained binding affinity comparable to that of its native counterpart (fig. S2, D and E). Finally, human and murine cells with varying *PTK7* expression levels were mixed and incubated with sgc8c-S for SPARK-seq to profile *PTK7*-gRNA, mRNA, and sgc8c-S binding simultaneously (Fig. 2C).

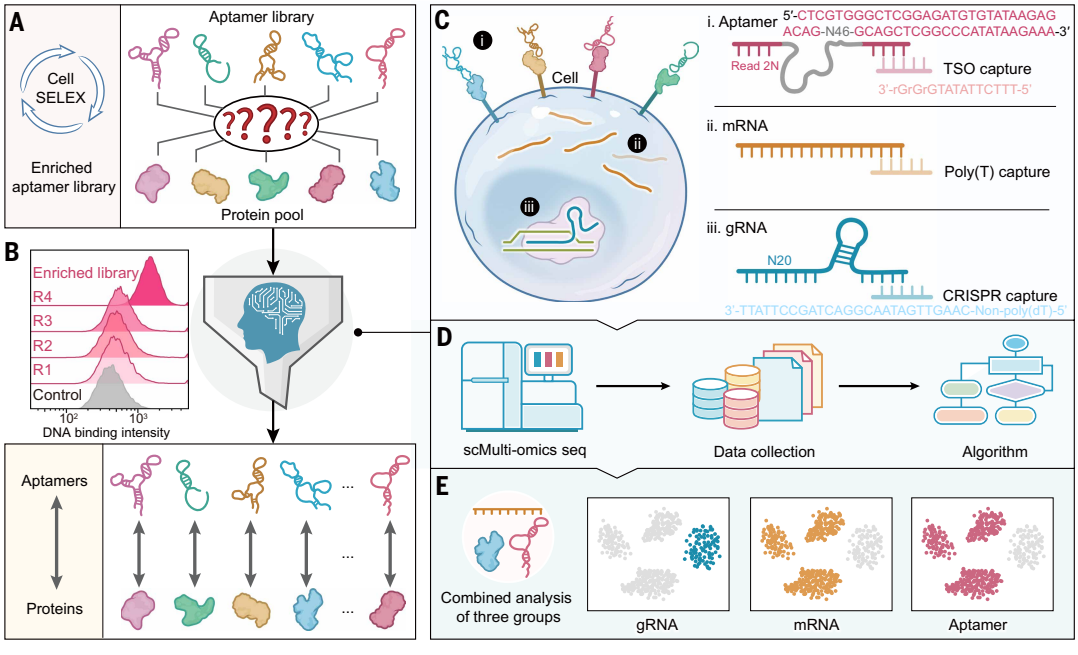


Fig. 1. SPARK-seq workflow overview. (**A**) Schematic of Cell-SELEX for enrichment of aptamers that bind native cell-surface proteins. (**B**) Flow cytometry validation of enriched library binding under physiological conditions. R1 to R4 denote libraries from selection rounds 1 to 4. (**C**) Architecture of the single-cell sequencing assay. The aptamer library features a 46-nt randomized core flanked by a 32-nt 5′ primer handle and a 22-nt 3′ primer handle, of which 13 nts constitute a capture sequence for single-cell sequencing. mRNA is captured through poly(A) tails. Each gRNA includes a 20-nt genomic targeting guide sequence (N20) followed by a scaffold containing a 26-nt capture motif. poly(dT) denotes poly(deoxythymidine). (**D**) Computational pipeline for differential analysis, integrating single-cell aptamer counts, gRNA identity, and transcriptome profiles. (**E**) Integrated single-cell readout reveals aptamer-protein interactions by correlating aptamer abundance with CRISPR perturbations and gene expression changes.

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

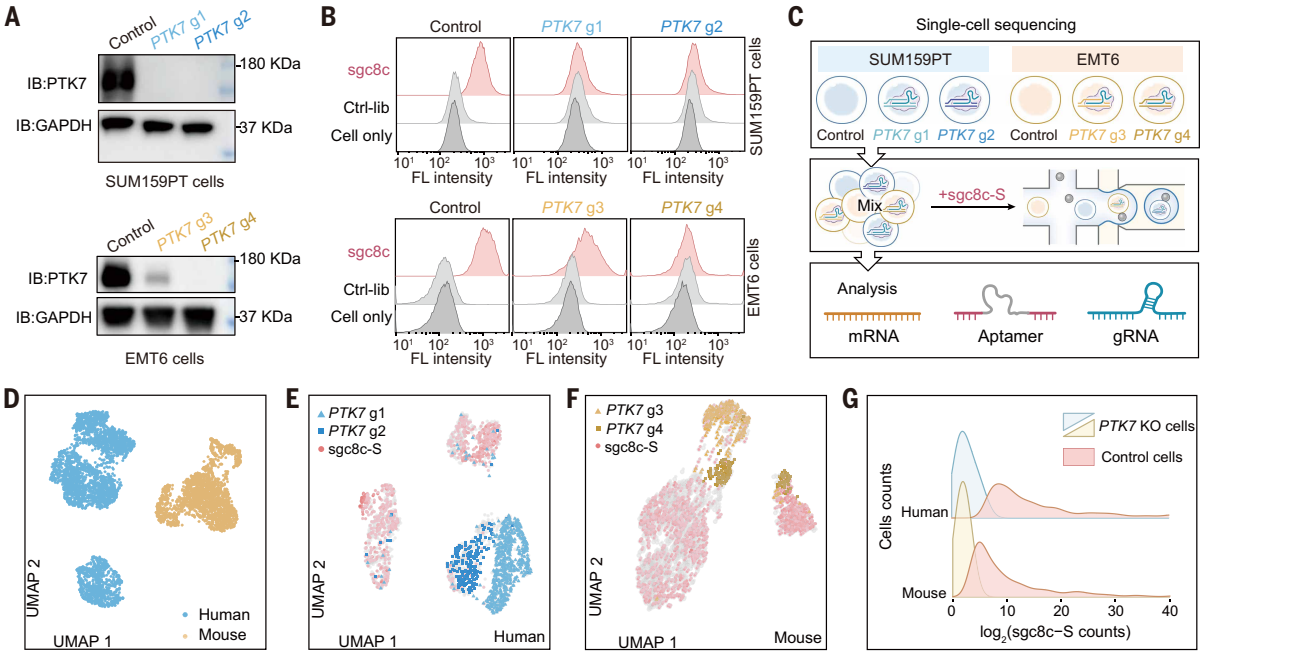


Fig. 2. Proof of concept for SPARK-seq with sgc8c aptamer. (**A**) Western blot analysis of PTK7 protein expression in control and CRISPR-Cas9 gene-edited cells. GAPDH serves as a loading control. Lanes represent cells transduced with distinct gRNA sequences targeting control or *PTK7* (g1 to g4 indicate distinct gRNA sequences). (Top) Human SUM159PT cells; (bottom) mouse EMT6 cells ($n = 3$ biological replicates). (**B**) Flow cytometry analysis of sgc8c aptamer binding to SUM159PT cells (transduced with Control gRNA, *PTK7* g1, *PTK7* g2) and EMT6 cells (transduced with Control gRNA, *PTK7* g3, *PTK7* g4). Fluorescence (FL) intensity reflects cell-surface aptamer binding. A nonbinding aptamer library (Ctrl-lib) was used as a negative control ($n = 3$ biological replicates). (**C**) Schematic illustrating the single-cell 5′-end sequencing workflow performed on six cell populations after aptamer incubation, enabling integrated mRNA, aptamer, and gRNA triomics analysis. (**D**) Uniform Manifold Approximation and Projection (UMAP) plot of mixed SUM159PT and EMT6 cells based on single-cell transcriptomes. Human (blue) and mouse (yellow) cells form distinct clusters, demonstrating effective species separation. (**E** and **F**) UMAP overlays of *PTK7* gRNA expression and sgc8c-S aptamer binding in SUM159PT (E) or EMT6 (F) cells. Each plot maps gRNA identity and corresponding sgc8c-S aptamer binding levels to their respective perturbed cell populations. (**G**) Quantification of sgc8c-S aptamer binding across PTK7 knockout and control cells by single-cell sequencing.

Species-specific mRNA analysis distinguished human from mouse cells, and intraspecies doublets were removed (Fig. 2D and fig. S2F). After removing low-quality cells and potential doublets expressing two or more gRNAs, a Gaussian mixture model (GMM) was applied to distinguish the gRNA-expressing and sgc8c-S binding groups from the control cells (fig. S2, G to N) (*23*, *24*). Uniform manifold approximation and projection (UMAP) mapping of the sgc8c-S binding profile (Fig. 2, E and F) revealed high binding in cells without *PTK7* gRNAs whereas cells with *PTK7* gRNA expression exhibited a significantly reduced sgc8c-S-binding level (Fig. 2G) across the cell lines of both species. These results confirm that SPARK-seq not only allows for the simultaneous detection of aptamer, gRNA, and mRNA levels in single cells but also facilitates pairing aptamers with specific knockout cell populations, linking the aptamer-binding profile to genetically target perturbed cell types.

SPARK-seq maps aptamer-target interactions

We then sought to explore whether SPARK-seq could achieve high-throughput target identification of an aptamer library using pooled genetically perturbed cells. To achieve this, we first constructed a single-stranded DNA library compatible with a single-cell sequencing platform, which we then used for screening against SUM159PT cells. After four rounds of selection, we obtained an enriched aptamer library, termed round 4 (R4), which exhibited strong binding to SUM159PT cells (Fig. 1B and fig. S3A).

Given the capacity of the 10X Genomics single-cell platform to profile approximately 10,000 cells per sample, we balanced cell number per gRNA-transduced cluster with the number of proteins targeted for perturbation. To refine our targeted protein panel, we performed a pull-down assay and MS analysis on the enriched libraries from each screening round (R0, R1, R2, R3, and R4) to identify potential aptamer targets (data S2). MS results showed a progressive enrichment of specific target proteins across selection rounds (fig. S3B), leading to selection of the top nine enriched membrane proteins (NRP1, PTK7, PTPRF, NRP2, PTPRS, CD151, ITGB1, PTPRD, and ITGA3), along with four proteins distributed across different enrichment intervals (SLC25A5, RAB5C, RAB23, and CDCP1) as perturbation targets (fig. S3C). We then transfected SUM159PT cells with a pooled library of 46 gRNAs, targeting the selected 13 proteins with three distinct gRNAs per protein, including seven negative control gRNAs (data S3) (*29*). These mixed-cell populations were subsequently incubated with the enriched R4 aptamer library, followed by SPARK-seq to assess target interactions at single-cell resolution (*24*).

To investigate the effects of each perturbation on the binding of specific aptamers and to uncover potential cell surface protein-aptamer interactions, we developed the SPARTA algorithm to analyze the multiomics data obtained from SPARK-seq. Based on the transcriptomic data, we removed empty cells and low-quality cells (fig. S4A), retaining 14,275 high-quality cells. We then denoised the gRNA data (fig. S4B) (*30*), classifying cells into distinct populations based on gRNA expression, and removed doublets with more than one gRNA (fig. S4C). A total of 8466 single cells with clear gRNA expression were identified, including 473 control cells (NC, negative control) (fig. S4D). Cell populations with fewer cells such as those with RAB23- and RAB5C-knockout were filtered out, leaving 11 protein-perturbed populations (population of 33 gRNAs) and 1 NC population (population of seven control gRNAs) for further analysis (fig. S4E).

To obtain aptamer-binding profiles for cells in these cell populations, we analyzed the aptamer sequences, identifying 4.027×10^7 distinct sequences and generating 1.72×10^8 reads. The top 10,000 unique sequences accounted for 68.84% of the total reads and expanding the analysis to the top 50,000 or 100,000 sequences only marginally improved coverage by <1% (fig. S4F). Accordingly, downstream analyses

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

focused on this high-coverage subset of 10,000 sequences. To strengthen the reliability of subsequent differential analysis, we grouped similar aptamer sequences, which were assumed to bind to the same target protein (*3*, *31*, *32*), into 1906 families using the BLAST (Basic Local Alignment Search Tool)-short-MCL algorithm (Fig. 3, A and B) (*33*). Briefly, we performed aptamer family classification using all-vs-all BLAST alignment followed by MCL, yielding objective, data-driven family boundaries that naturally group related sequence variants. Overall, we obtained 8466 cells across 34 cell populations with different gRNA expression, along with their aptamer-binding profiles for 1906 aptamer families (Fig. 3, C and D).

To identify aptamer-protein interactions, SPARTA integrates binding difference calculations, quality control, and a scoring system (Fig. 3E). First, the median abundance difference for each aptamer family was calculated within each gRNA-defined cell population relative to the NC population to quantify the relative potential binding changes resulting from target knockout (defined as binding differences, see methods). Aptamer families with nearly consistent binding differences across the gRNA-defined cell populations were excluded, as were gRNAs causing large binding differences across half of aptamer families, to minimize false-positive signals (see methods for additional quality controls). Next, Gaussian mixture modeling was applied to the $\log_2$-transformed binding differences to distinguish between specific and nonspecific binding, with a threshold defined to identify potential specific binding events (fig. S5). The variation in binding differences declined sharply beyond the top 20 families (fig. S6, A and B), which together accounted for 98% of the reads from the top 10,000 unique sequences (Fig. 3B). We therefore restricted downstream analyses to this high-coverage subset. Subsequently, proteins with <2 corresponding gRNA cell populations were excluded to minimize the impact of random variability. For the remaining proteins, a single candidate was assigned as the target, and for multiple candidates a scoring system was applied using score ratios and proportions to prioritize the most likely binding targets (Fig. 3E, as detailed in methods). This approach effectively filtered out nonspecific signals and strengthened the reliability of aptamer-target protein pairings.

Additionally, to rule out artifacts from incomplete gene disruption (fig. S6C), we implemented Mixscape to assess whether gene expression in individual cells was truly affected by the intended knockout (*24*) and filtered out gRNA-positive cells that did not exhibit transcriptional evidence of perturbation. This targeted filtering uncovered aptamer Family 4 against PTPRS (fig. S6D). Finally, we note that expanding the analysis window from top 10,000 to top 50,000 or 100,000 unique aptamer sequences did not uncover additional targets (fig. S6E) but did introduce substantial statistical noise, increased computation time by 9- to 34-fold, and raised memory usage by 4- to 6-fold (fig. S6F). These findings underscore that under our conditions, the dominant sequence fraction sufficiently represents the library's functional potential.

Consequently, SPARK-seq identified binding differences for 12 aptamer families across nine different perturbed cell populations (Fig. 3F). Specifically, aptamers from families 5, 13, and 17 bind to CDCP1; aptamers from family 3 bind to the ITGA3 and ITGB1; families 2, 6, and 7 bind to NRP1; families 11 and 14 bind to NRP2; families 1 and 15 bind to PTK7; and family 4 binds to PTPRD, PTPRF, and PTPRS proteins (Fig. 3G and fig. S6A).

Validation of aptamer-target interactions

To validate the specificity of aptamer-target interaction candidates identified from SPARK-seq and SPARTA, we evaluated copy number differences among individual aptamers within each aptamer family by comparing with both target-knockout and control cells. For each of the protein targets, we observed variable changes in aptamer-binding levels upon target knockout, suggesting that these aptamers exhibit differing sensitivities to target expression levels (Fig. 4, A, D, G, J, M, P, S, and V). To confirm that the aptamers with varying copy number changes specifically bind to their target, we selected aptamers with distinct binding capacities for validation. For instance, within the PTK7-binding aptamer family, we selected Apt-1-1 (Aptamer family number and aptamer number, respectively), showing larger fold change in copy number with $\log_2$ fold change ($\log_2$FC) = 2.3060, and Apt-15-27, showing smaller fold change with $-\log_2$FC = 0.1049, for evaluation. Flow cytometry results confirmed that both aptamers specifically bound to PTK7-expressing cells but failed to bind PTK7-knockout cells, demonstrating their specificity for PTK7 (Fig. 4B and fig. S7A). SPR and microscale thermophoresis (MST) assays further validated the binding affinity of aptamers to PTK7 protein in vitro in the nanomolar range, further supporting their high specificity and binding strength for PTK7 (Fig. 4C and fig. S7, B and C). Similarly, for targets CDCP1, NRP1, and NRP2, we selected aptamers with varying copy number changes and conducted flow cytometry, SPR, and MST assays (Fig. 4, D to L; fig. S7, D to G; fig. S8, A and B; and data S4) (*34*). These results consistently confirmed that aptamers identified through SPARTA specifically bound their targets, underscoring the accuracy of SPARK-seq in identifying precise aptamer-target interactions.

Beyond the above single target mappings, SPARK-seq also predicted multitarget binding for aptamer families 3 and 4 (ITGA3/ITGB1 and PTPRD/PTPRF/PTPRS, respectively) (Fig. 3F). For aptamer family 3, the representative aptamer Apt-3-3 failed to bind cells with either ITGB1 or ITGA3-knockout, reinforcing the specificity of Apt-3-3 for both targets (Fig. 4N and fig. S9A). These results were further validated by SPR, which demonstrated that Apt-3-3 bound specifically to the integrin complex (Fig. 4O). Further subunit analysis revealed that ITGA3-knockout did not affect ITGB1 protein levels, whereas ITGB1-knockout significantly reduced ITGA3 protein levels (fig. S9, B to D). These findings suggest that ITGB1 regulates the protein expression of ITGA3. Crucially, the inability of Apt-3-3 to bind ITGA3 knockout cells, despite intact ITGB1 expression, demonstrates that Apt-3-3 directly targets ITGA3 (Fig. 4N, fig. S9, A and D).

For aptamer family 4, flow cytometry showed that knockout of each paralog (PTPRD, PTPRF, PTPRS) impaired binding of the representative aptamer Apt-4-5 (Fig. 4, Q, T, and W), whereas antibody staining confirmed unaltered expression of the remaining two paralogs in each knockout background (fig. S9, E and F), supporting the conclusion that Apt-4-5 directly binds all three proteins. SPR assays further validated this, revealing nanomolar binding affinities to each purified target (Fig. 4, R, U, and X, and data S4). Although these paralogs share 49.92% sequence identity (fig. S9G) and AlphaFold predicted strong structural homology (fig. S9H), SPR-based kinetic analyses showed distinct equilibrium dissociation constants (K_D), association rate constant (k_{on}), and dissociation rate constant (k_{off}) values (fig. S9, I to L), indicating that Apt-4-5 recognizes a conserved epitope but discriminates subtle structural or contextual differences.

To further validate the binding of aptamer across all identified targets under physiological conditions, we performed flow cytometry binding assays at 37°C on a subset of representative aptamers and observed that these sequences retain robust binding at physiological temperature (fig. S10A). Moreover, kinetic analyses in live-cell contexts using flow cytometry and real-time interaction cytometry (RT-IC) experiments (*35*) demonstrate that although the K_D measured in the live-cell environment are not identical to those obtained from traditional SPR using purified proteins, they are broadly comparable (Fig. 4Y, fig. S10, B and C, and data S4). These results confirm that aptamer binding kinetics observed in vitro are generally reflective of their behavior in physiological cellular contexts.

In summary, high-throughput SPARK-seq enabled mapping of interactions between 5535 aptamer sequences and eight surface proteins (PTK7, ITGA3, CDCP1, NRP1, NRP2, PTPRD, PTPRF, PTPRS) (Fig. 4Z and data S5). Of the 13 proteins initially targeted, RAB5C and RAB23 populations were excluded as a result of insufficient cell numbers and ITGB1 was omitted because it functions as a co-receptor rather than a direct

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

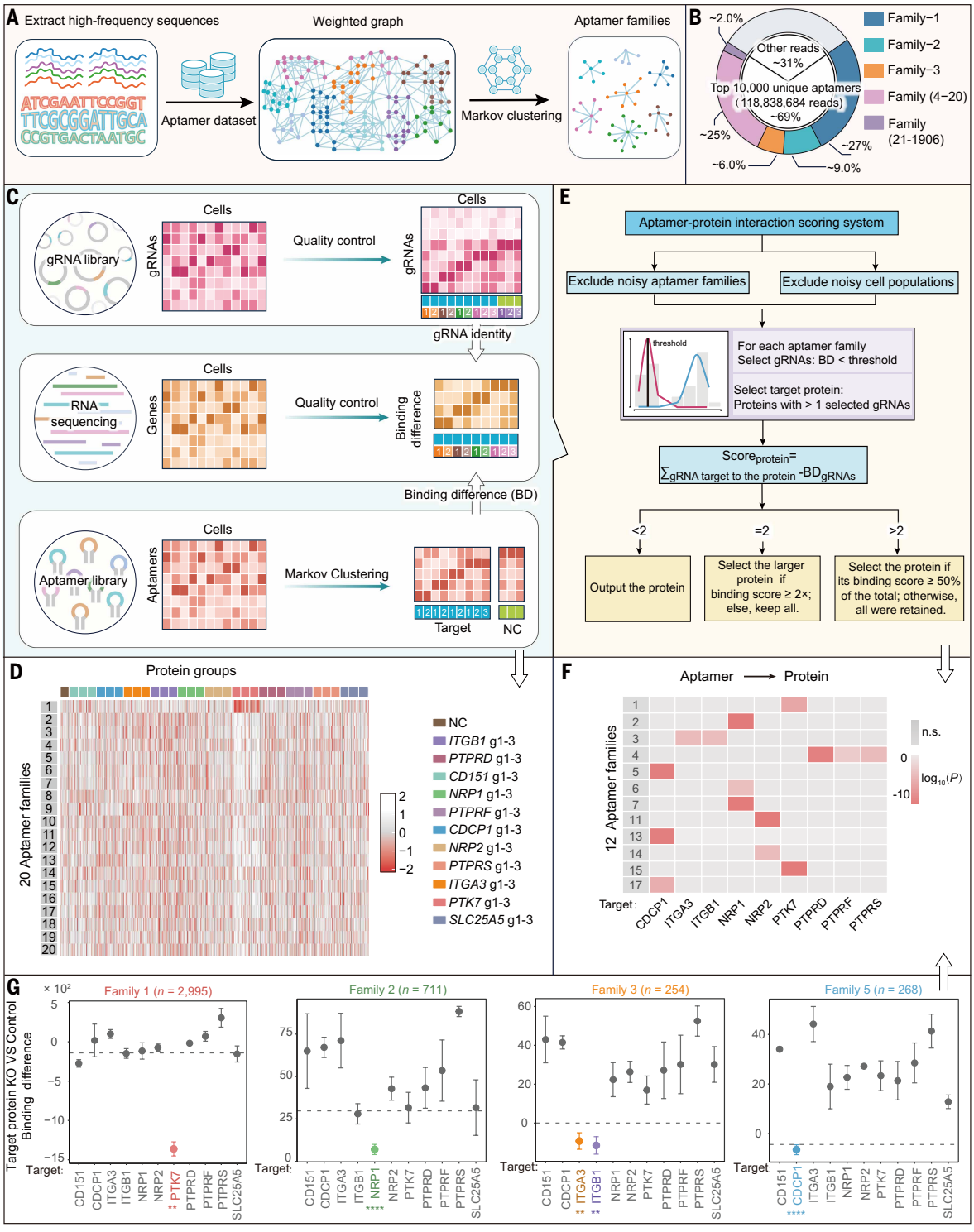


Fig. 3. High-throughput aptamer-target mapping pipeline. (**A**) Schematic of aptamer family definition. High-frequency sequences are extracted from the enriched library, represented as nodes in a weighted graph, and clustered into families by Markov clustering. (**B**) Read count distribution among aptamer families clustered from the top 10,000 unique sequences. (**C**) Single-cell multiomics quality control and data integration. From a pooled CRISPR perturbed cell population, gRNA identities (top) and transcriptomes (middle) undergo parallel QC to remove low-quality or ambiguous cells. Aptamer counts (bottom) are then assigned to families by sequence clustering. Each cell is represented by its gRNA, gene expression, and aptamer-binding difference (abundance difference) relative to nontargeting controls (NC). (**D**) Heatmap of median differential binding (log_2 foldchange relative to NC) for the 20 most abundant aptamer families across 12 protein-knockout populations (color bar at top). Values < 0 (red) indicate reduced binding in knockouts, consistent with target-specific aptamer loss. (**E**) SPARTA scoring algorithm for predicting aptamer–protein pairs. Candidate aptamer–protein pairs are scored using differential binding, Gaussian mixture modeling and cross-family filtering to yield high-confidence predictions. (**F**) Heatmap of predicted interactions between aptamer families (rows) and candidate membrane proteins (columns). Color intensity indicates $log_{10}(P)$ value of binding differences. n.s., not significant. (**G**) Dot plots of quantitative binding differences of aptamer family for indicated targets versus control; *n* indicates family size. Asterisks denote significance by two-tailed *t* test (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$).

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

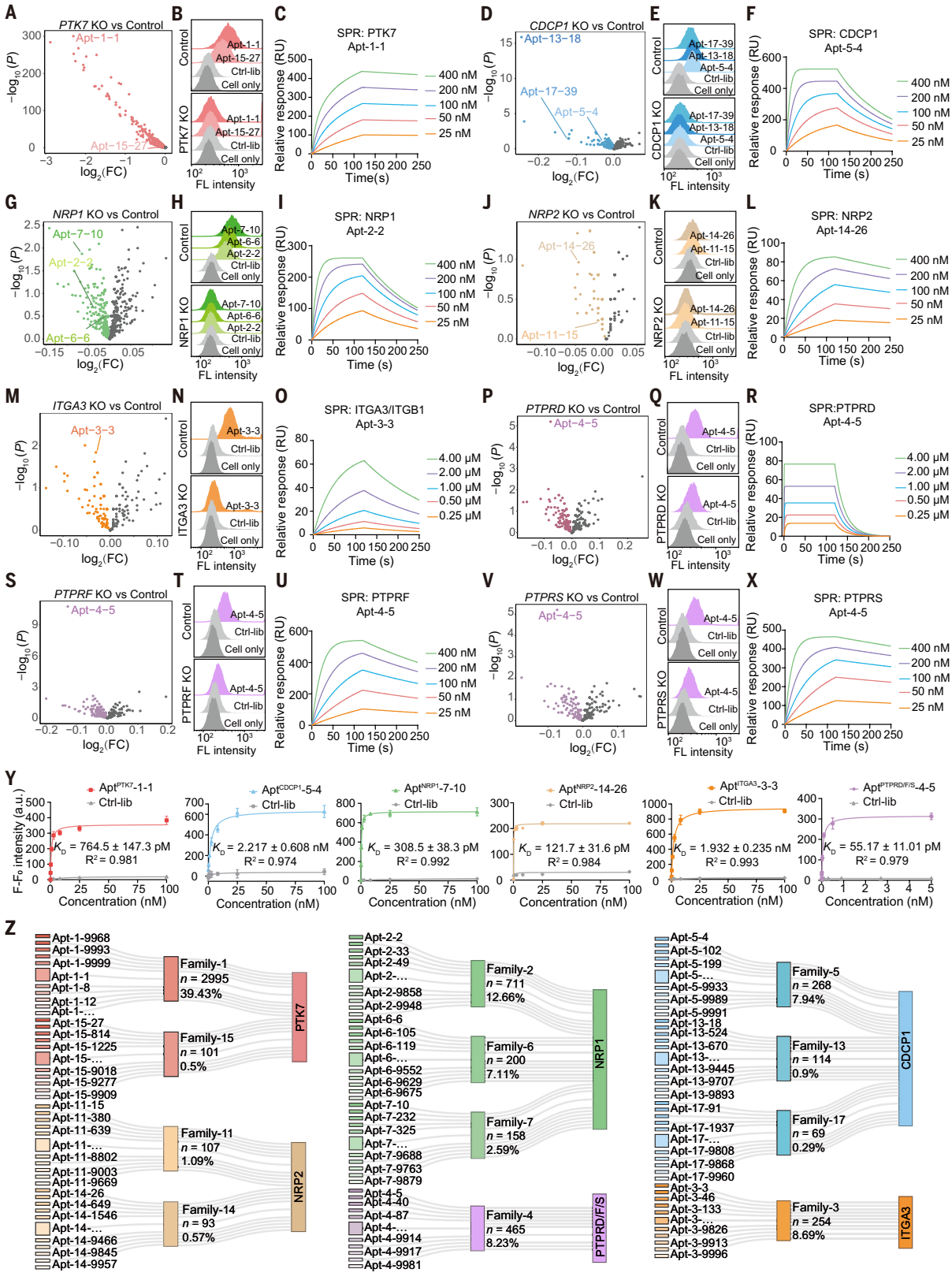


Fig. 4. Aptamer target validation. (**A** to **X**) For each of the eight surface proteins (PTK7, CDCP1, NRP1, NRP2, ITGA3/ITGB1, PTPRD, PTPRF, and PTPRS): Volcano plots (left) display $\log_2$fold change in aptamer abundance (knockout/control) versus $-\log_{10}(P)$ values (empirical Bayes moderated t-test) obtained from differential expression analysis using limma on single-cell SPARK-seq data; Flow cytometry histograms (middle) compare FAM-labeled aptamer binding to control and knockout cells, a nonbinding control aptamer (Ctrl-lib) was used as a negative control ($n = 3$ biological replicates); SPR sensorgrams (right) of concentration-dependent binding titrations against purified proteins. ($n = 3$ biological replicates). (**Y**) Determination of aptamer binding affinities for target proteins via flow cytometry ($n = 3$ biological replicates). (**Z**) Sankey diagram linking the 5535 unique sequences to eight validated protein targets. Family sizes (n) and percentage of total reads are indicated.

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

aptamer binder. SPARK-seq also precisely excluded CD151 and SLC25A5 as indirect interactors. Although CD151 co-immunoprecipitated (Co-IP) with ITGA3 (a validated Apt-3-3 target; fig. S10D and data S6), CD151 knockout did not alter Apt-3-3 binding and SPR confirmed no direct interaction (fig. S10, E to G). Likewise, SLC25A5 associated with NRP2 by Co-IP (fig. S10H and data S6) but its knockout had no effect on NRP2-binding aptamers (fig. S10, I and J). These results demonstrate that SPARK-seq reliably identifies aptamers with high specificity and affinity for direct targets while effectively filtering out indirect associations, enabling high-confidence, high-throughput discovery of aptamer target pairs.

Orthogonal validation of aptamer specificity

To confirm the molecular specificity of our SPARK-seq derived aptamers, we employed a three-tiered validation strategy. First, in flow cytometry-based competition assays, cells were incubated with a fluorescently labeled "hot" aptamer then challenged with increasing concentrations of an unlabeled competitor. Homologous competition yielded a clear dose-dependent reduction in fluorescence (fig. S11, A and B), demonstrating saturable, high affinity binding to the intended surface antigen. By contrast, a 10-fold excess of noncognate aptamer failed to attenuate the "hot" aptamer signal (Fig. 5A), confirming that off-target aptamer sequences do not disrupt binding. Second, biochemical pull-down from cell lysates with biotinylated aptamers followed by target-specific immunoblotting enriched the predicted proteins, confirming selective recognition within the native proteome (fig. S11C). Finally, we performed SPR cross binding assays on twelve representative aptamer families, collectively covering all eight identified SPARK-seq targets. Each aptamer was tested against its cognate protein and seven unrelated proteins. High-affinity interactions were observed exclusively with the intended target whereas signals for nontargets matched negative controls (Fig. 5B). Together, these orthogonal assays compellingly demonstrate the precise, molecular level specificity of our aptamer repertoire.

Building on this specificity, we assessed SPARK-seq's discrimination power within a highly conserved protein family. Despite high sequence and structural homology between NRP1 and NRP2 (Fig. 5C), SPARK-seq identified distinct aptamers that bind exclusively to either NRP1 or NRP2 with no cross-reactivity (Fig. 5D). The NRP1 aptamer competitively blocked VEGF165 binding (Fig. 5E), suggesting that it engages the canonical VEGF-binding epitope, whereas NRP2 aptamer bound to a noncanonical site that did not compete with VEGF165, VEGF-C, or TGF-β1 (Fig. 5F), offering a probe for NRP2-specific biology and therapeutic targeting.

SPARK-seq aptamers capture diverse protein targets

To investigate whether SPARK-seq captures proteins across a spectrum of physicochemical properties, we profiled eight validated targets. Predicted isoelectric points spanned 4.5 to 8.5 (fig. S12A), and hydrophobicity did not correlate with aptamer capture efficiency (fig. S12B), indicating that SPARK-seq performance is independent of target charge and overall primary sequence hydrophobicity. Structural models from AlphaFold revealed large, modular extracellular domains: immunoglobulin folds in PTK7 and PTPR family proteins and fibronectin repeats in integrins (fig. S12, C and D). These architectures likely present stable binding surfaces for nucleic acid ligands. Surface electrostatic potential maps further identified positively charged patches on these domains, providing a plausible basis for electrostatic complementarity with aptamers (fig. S12E).

Finally, we assessed target abundance at both the total whole-cell and surface levels. Mining published MS datasets (6447 quantified proteins) revealed five targets (e.g., PTK7) ranked in the top 30% of cellular protein abundance, NRP1 fell near the 30th percentile, whereas NRP2, PTPRD, and PTPRS were below detection (fig. S12F) (*36*). High-sensitivity MS platform (10,970 detected proteins) recovered NRP2 and PTPRS in the bottom third whereas PTPRD remained undetected, suggesting ultralow expression (fig. S12G and data S7). Calibrated flow cytometry confirmed these trends at the cell surface, quantifying over 10^5 copies per cell for ITGA3 versus <1000 copies per cell for PTPRD and PTPRF (Fig. 5G and fig. S12H). Together, these findings demonstrate that SPARK-seq can reliably discover aptamers against cell-surface proteins spanning more than two orders of magnitude in expression and diverse physicochemical features.

In comparison to traditional methods that focus on high-abundance sequences and rely on pull-down experiments, SPARK-seq's single-cell differential screening offers a more robust and sensitive approach. It effectively identifies aptamer-target interactions, such as CDCP1, that may be missed by conventional biochemical techniques as a result of issues such as protein denaturation or ionization inefficiencies (fig. S13 and data S6). This functional, cell-based strategy highlights SPARK-seq's ability to uncover targets that are sensitive to in vitro artifacts and conformational changes, providing a powerful complementary tool for aptamer discovery.

SPARK-seq prioritizes slow dissociation-rate aptamers

To determine why certain aptamers are preferentially enriched by SPARK-seq, we quantified binding differences as the $\log_2$FC in aptamer abundance between protein-knockout and nontargeting control populations. Focusing on 3096 aptamers targeting PTK7, we selected eight aptamers with varying $\log_2$FC values, ranging from −3 to 0, for further investigation. Flow cytometry confirmed the specificity of their binding and revealed a strong correlation between $-\log_2$FC values and fluorescence intensity ($R^2 = 0.95$, $P < 0.0001$) (Fig. 6, A and B), indicating that higher $-\log_2$FC values correspond to increased cell surface binding. Kinetic analysis further demonstrated that the k_{off} correlates strongly with $-\log_2$FC ($R^2 = 0.84$, $P < 0.0001$), whereas the K_D showed no significant association ($R^2 = 0.0001$, $P = 0.96$), highlighting k_{off} as the primary determinant of SPARK-seq enrichment (Fig. 6C and fig. S14, A and B).

To systematically dissect the kinetic basis of this enrichment, we measured K_D, k_{on}, and k_{off} by SPR. Notably, k_{off} showed a robust association with $-\log_2$FC ($R^2 = 0.88$, $P < 0.0001$) (Fig. 6, D and E), whereas K_D and k_{on} exhibited weaker or no correlations with $-\log_2$FC ($R^2 = 0.68$, $P < 0.0001$; $R^2 = 0.21$, $P = 0.0162$, respectively) (fig. S14, C and D), demonstrating that SPARK-seq selectively captures aptamers with slower off-rates. To validate these findings in a more physiological setting, we performed live-cell binding assays by RT-IC. Again, k_{off} dominated the strongest correlation with $-\log_2$FC values (Fig. 6F; $R^2 = 0.87$, $P = 0.0007$), and K_D and k_{on} remained poor predictors of enrichment ($R^2 = 0.12$, $P = 0.32$; $R^2 = 0.23$, $P = 0.12$, respectively) (fig. S14, E to H). Importantly, this strong correlation between $-\log_2$FC values and k_{off} held across additional targets, including CDCP1, NRP1, NRP2, and PTPRF, even when their $-\log_2$FC differences were modest (Fig. 6, G to J). By contrast, correlations between K_D or k_{on} and $-\log_2$FC varied inconsistently across targets, underscoring that SPARK-seq is especially sensitive to aptamers with slow off-rates (fig. S14, I to L, and data S4).

To further investigate whether SPARK-seq prioritizes the identification of slow off-rate aptamers, we compared this approach with the conventional selection method, which typically identifies aptamers based on their highest enrichment levels, i.e., the most abundant sequences in the library. Across all targets analyzed, we observed that aptamers with the highest $-\log_2$FC values did not overlap with those with the highest copy numbers (Fig. 6, K to M, fig. S15, A to C, and data S8). More importantly, aptamers with the highest $-\log_2$FC values consistently exhibited significantly slower dissociation rates compared with those identified by conventional methods (Fig. 6, N to P, and fig. S15, D to F). For targets, such as PTK7, NRP1, and NRP2, the dissociation rates of aptamers identified through SPARK-seq were ~10 times slower than those obtained through traditional selection approaches (Fig. 6, N to P). These findings underscore the advantage of SPARK-seq in identifying slow off-rate aptamers for diverse targets.

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

Capitalizing on the diversity of aptamer kinetics captured in the SPARK-seq dataset, we implemented, within the SPARTA framework, a two-module platform for aptamer prediction and variant generation (fig. S16A) (*37*, *38*). The prediction module employs a convolutional neural network (CNN)-based deep learning model trained on SPARK-seq–derived aptamer families to classify sequences by binding potential. In a PTK7 proof of concept, training on 2395 PTK7-binding versus 2395 nonbinding sequences yielded ~97% accuracy when tested on seven known PTK7 aptamers and 24 unrelated controls (fig. S16, B to D, and data S9). Moreover, by correlating SPARK-seq $-\log_2$FC values with measured dissociation rates, we stratified aptamers into nonbinders, strong binders (low k_{off}), and intermediates, and used these classifications to build a three-class model whose kinetic predictions were validated by SPR (fig. S16, E and F). The generation module employs a variational autoencoder, trained on the PTK7 aptamer family to create variant sequences. Of 15 generated variants, 12 bound PTK7 in vitro, with one variant (SPARK-Apt-18) exhibiting the slowest dissociation observed (fig. S16G, and fig. S17A). Thus, SPARTA combines single-cell resolved kinetic profiling with deep generative modeling to rapidly generate variants with distinct slow off-rates, a property that enhances target occupancy and signal persistence and offers clear advantages for in vivo imaging, targeted drug delivery, and therapeutic blockade.

Discussion

SPARK-seq integrates CRISPR-based gene perturbation with single-cell aptamer and transcriptome profiling to map interactions between 5535 aptamers and eight target proteins at single-cell resolution. By directly linking perturbation, gene expression and binding phenotype in intact cells, the approach enables unbiased discovery of physiologically relevant aptamer-target pairs.

However, the platform does generate high-dimensional, sparse data that pose analytical challenges. To address these challenges, we employed a "human-in-the-loop" framework. This involved retaining the top 10,000 most abundant unique sequences, clustering them into sequence-informed families, and retaining multiple statistically significant protein associations per family rather than enforcing a single winner. By comparing aptamer enrichment in knockout versus control cells, SPARK-seq directly measures target-dependent binding and overcomes the bias of traditional methods toward abundant proteins and minimizes PCR bias or pull-down artifacts, enriching for high-affinity, target-specific aptamers. Furthermore, SPARK-seq's cell-native readout and compatibility with the 10X platform allow seamless integration into Cell-SELEX workflows (typically 1 to 4 rounds), facilitating adoption across research laboratories.

These technical features translate into functional strengths. We validated aptamers across a >100-fold dynamic range of surface abundance and across diverse biophysical classes and resolved closely related paralogs (for example, NRP1 versus NRP2) with no detectable cross-reactivity. Moreover, SPARK-seq reflects kinetic parameters (e.g., k_{off}), enabling prioritization of slow off-rate binders, a property that underpins recent engineered slow-off aptamers (SOMAmers) and is highly desirable for diagnostic and therapeutic applications (*39*–*43*). Beyond immediate experimental utility, the large, labeled interaction atlas produced by SPARK-seq provides a rich substrate for predictive modeling. Whereas most machine-learning efforts to date have focused on transcription factors or nucleic-acid binders (*38*, *44*), SPARK-seq directly links aptamer sequences to native cell-surface targets and kinetics, opening the way for AI-driven aptamer design and rapid optimization against clinically relevant receptors.

We acknowledge certain limitations, wherein our approach is most effective for aptamers binding individual proteins and is less suited to epitopes requiring intact multiprotein complexes, which may be disrupted by single knockouts. Broader transcriptomic or proteomic integration following perturbation could help uncover such indirect effects. Additional constraints include the limited chemical diversity in unmodified DNA libraries and the throughput constraints of droplet-based platforms, which can capture only ~20,000 cells per run, limiting CRISPR screening to 10 to 15 targets (*45*, *46*). Selections were performed at 4°C, in line with standard SELEX practices, to minimize endocytosis and cell internalization (*17*, *47*), though optimization at 37°C will be important for in vivo applications. Future improvements will focus on modified libraries, higher-throughput platforms, optimized CRISPR designs, more effective knockdown strategies, and competitive selection to sharpen discovery and functional mapping.

In summary, SPARK technology establishes a platform for high-throughput, high-specificity identification of multiplexed aptamer-target interactions in native cellular contexts. SPARK-seq offers broad opportunities for expansion. Strong, unexplained binding phenotypes may highlight uncharacterized or mislocalized antigens, suggesting that genome-scale CRISPR screens could systematically map aptamers to both known and unannotated surface proteins. Future adaptations such as cell permeabilization and stress-induced mislocalization models will enable aptamer identification for intracellular and mislocalized proteins (*48*). Additionally, this methodology provides a versatile framework for the discovery and optimization of a broad spectrum of functional nucleic acids, further broadening SPARK-seq's utility for target discovery and therapeutic probe development.

Materials and methods

Plasmid construction and cell line generation

For proof-of-concept experiments, we utilized the pLentiCRISPRv2 plasmid (Addgene, 52961) as the vector. In the high-throughput identification phase, we employed pLentiCas9-BFP (Addgene, 78545) and lentiGuide-Puro (Addgene, 52963) to target specific genes. Guide RNAs (gRNAs) were designed using the CRISPOR tool, or sourced from the literature (*29*), ensuring minimal off-target effects. Each gRNA sequence (data S1) was synthesized and cloned into the BsmBI restriction site of the pLentiCRISPRv2 plasmid using the ZhangLab assembly method (*28*). Following ligation, constructs were transformed into Stbl3 competent cells via heat shock. Plasmids were purified using the EndoFree Mini Plasmid Kit (TIAN GEN, DP118-02), and successful insertion of the correct gRNA was confirmed by Sanger sequencing.

For lentiviral production, HEK293T cells (from ATCC) were co-transfected with lentiCas9-Blast, lentiGuide-Puro containing gRNAs, or pLentiCRISPRv2 plasmids, along with packaging plasmids pRSV-Rev (Addgene, 12259), pMDLg/pRRE (Addgene, 12251), and pMD2.G (Addgene, 12253), using Lipofectamine 3000 (Thermo Fisher Scientific). Forty-eight hours post-transfection, viral supernatants were collected, filtered, and used to transduce target cells in the presence of 8 μg/mL polybrene. Successfully transduced cells (pLentiCRISPRv2) were selected with puromycin (1 μg/mL) for 48 hours. For pLentiCas9-BFP, after transfection, cells were selected using a fluorescence sorter and then expanded. LentiGuide-Puro viruses containing different gRNAs were introduced into target cells in the presence of 8 μg/mL polybrene. Transduced cells were then selected using puromycin (1 μg/mL) for 48 hours. knockout (KO) efficiency was evaluated by Western blotting, flow cytometry or RT-qPCR.

Cell culture

All parental cell lines, including HEK293T, SUM159PT, A549, EMT6, and MCF-10A, were obtained from ATCC. Cells were cultured in RPMI-1640, F12K, DMEM medium (Gibco), or MCF-10A Cell Complete Medium (Procell, CM-0525), each selected according to the manufacturer's recommended formulation for the respective cell line, and supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin-streptomycin (100 U/mL, Gibco). Cultures were maintained at 37°C in a humidified atmosphere with 5% CO2. For passaging, cells were washed twice with Dulbecco's Phosphate-Buffered Saline (DPBS) and detached using 0.05% trypsin-EDTA (Gibco). Cells were passaged when 80 to 90% confluent

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

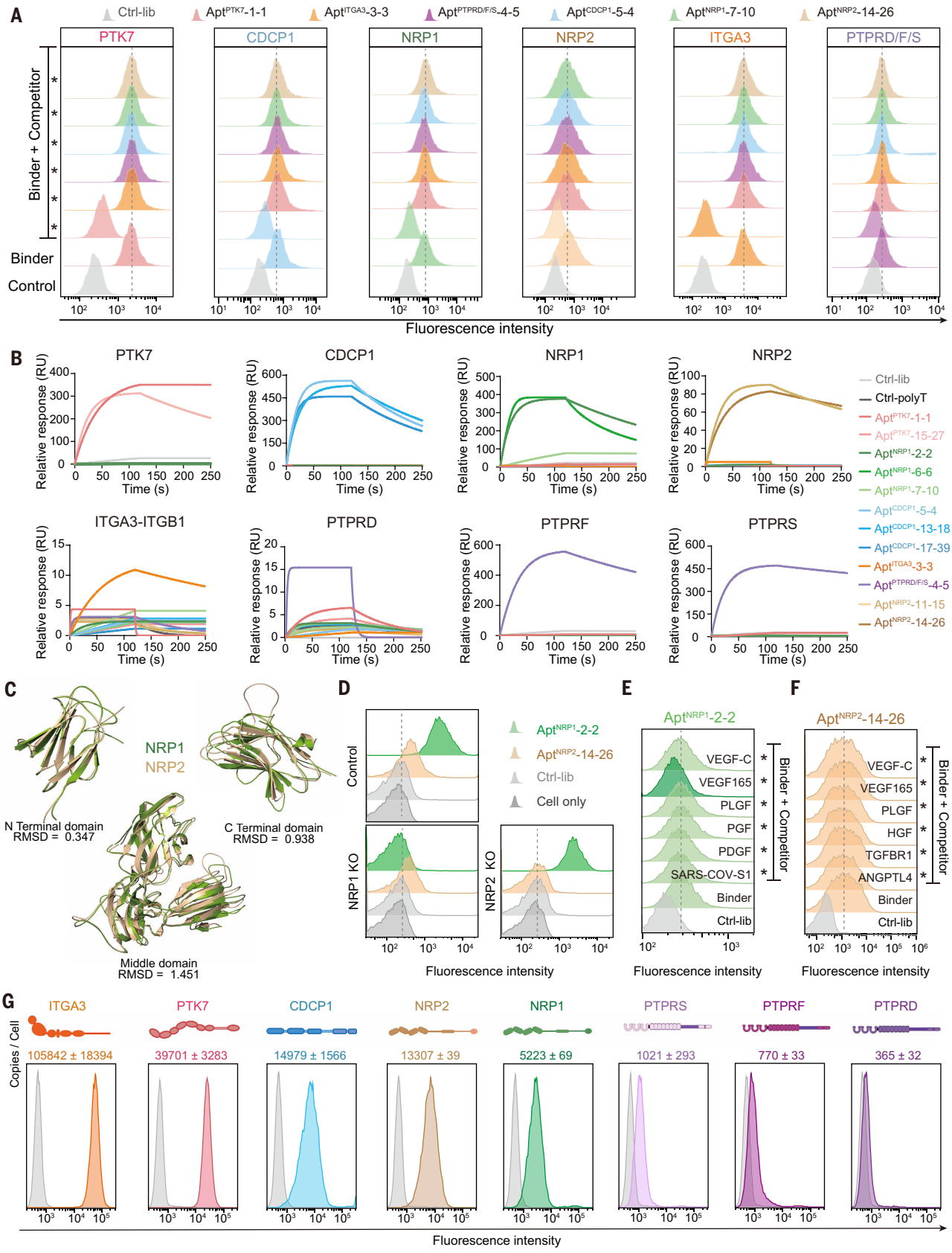


Fig. 5. Aptamer specificity validation. (**A**) Flow cytometry competition assays. From bottom to top: Ctrl-lib (gray, nonbinding aptamer control); FAM-labeled aptamer binding (Binder); and FAM-labeled aptamer binding in the presence of a 10-fold molar excess of unlabeled noncognate competitor aptamers (colored traces, asterisks mark competitor conditions). (**B**) SPR sensorgrams for distinct aptamers (200 nM) binding to indicated protein. Ctrl-lib and polyT serve as negative controls ($n = 3$ biological replicates). (**C**) Structural alignment of NRP1 (green) and NRP2 (tan) extracellular domains using Alphafold (https://alphafold.com/). RMSDs, root means square deviations. (**D**) Histograms of

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

Apt^{NRP1}-2-2 (green) and Apt^{NRP2}-14-26 (orange) binding to control versus NRP1 or NRP2 knockout cells, alongside Ctrl-lib (light gray) and cell-only (dark gray) controls ($n = 3$ biological replicates). (**E**) Flow-cytometric competition of Apt^{NRP1}-2-2 with indicated NRP1 ligands. From bottom to top: Ctrl-lib (gray, nonbinding aptamer control); FAM-labeled Apt^{NRP1}-2-2 (binder); and FAM-labeled Apt^{NRP1}-2-2 in the presence of a 10-fold molar excess of inidcated NRP1 ligands (asterisks). ($n = 3$ biological replicates). (**F**) Flow-cytometric competition of Apt^{NRP2}-14-26 with known NRP2 ligands (A549 cells). From bottom to top: Ctrl-lib (gray, nonbinding aptamer control); Alexa Fluor 647-labeled aptamer binding (binder); and Alexa Fluor 647-labeled aptamer binding in the presence of a 10-fold molar excess of known NRP1 ligands (indicated by asterisks). ($n = 3$ biological replicates). (**G**) Quantitative SUM159PT cell surface calibration. Upper: ribbon diagrams of the eight validated target proteins. Bottom: fluorescence-based calibration curves convert mean fluorescence intensity to quantify protein copy number per cell, revealing target abundance across the panel ($n = 3$ biological replicates; mean ± SD).

and seeded for subsequent experiments. Mycoplasma contamination was regularly checked using aptamer ZY3A (FAM-labeled) via flow cytometry (*49*); no contamination was detected at any stage.

Western blotting

Antibodies against PTK7 (CST, 25618), NRP1 (HUABIO, ET1609-69), ITGA3 (HUABIO, HA600100), ITGB1 (HUABIO, ET1601-17), and GAPDH (HUABIO, EM1101) were used, as well as HRP-conjugated secondary antibodies from the same supplier. Western blot experiments followed standard protocols: cells were lysed in 1 mL RIPA buffer (APPLYGEN, C1053) supplemented with protease inhibitor cocktail (Roche, 81410500), phosphatase inhibitor cocktail (Beyotime, P1082), and universal nuclease (Beyotime, DT121-250U) for 15 min at 4°C. Lysates were centrifuged at 13,000 × g for 10 min at 4°C to obtain clarified total protein extracts. Protein concentrations were determined using the BCA Protein Assay Kit (Thermo Fisher Scientific, 23225), following the manufacturer's instructions. Equal amounts of protein (normally 20 μg) were separated by 4–20% SDS-PAGE gradient gel electrophoresis (GenScript, M42012C). A pre-stained protein marker was used to indicate molecular weight (ABclonal, RM19001). Proteins were then transferred to a PVDF membrane (Millipore, 0000377714) at 80 V for 120 minutes in an ice bath using wet transfer. PVDF membranes were pre-activated in 100% methanol for 1 to 2 minutes before use. Blocking was performed with 5% (w/v) nonfat dry milk in TBST (10 mM Tris-HCl pH = 7.5, 150 mM NaCl, 0.05% Tween-20) for 1 hour at room temperature. Membranes were incubated overnight at 4°C with primary antibodies at a recommended dilution. The following day, membranes were washed three times with TBST and then incubated with HRP-conjugated secondary antibodies (CST, 1:5000 dilution) at room temperature for 1 hour. Signals were detected using an enhanced chemiluminescence (ECL) system, according to the manufacturer's instructions, and images were captured using a digital imaging system.

Flow cytometry (FCM) analysis

Cells were detached using enzyme-free dissociation solution (Shyuanye, R20103). After washing with washing buffer (DPBS supplemented with 1 mM $MgCl_2$, 4.5 g/mL D-Glucose), cells were resuspended in binding buffer (washing buffer supplemented with 0.1 mg/mL salmon sperm DNA (Thermo Fisher, 15634017), and 1 mg/mL BSA). For antibody staining, cells were incubated with primary antibodies, including Anti-Human ITGA3 (1:200; Thermo Fisher, MA5-28565), Anti-Human ITGB1 (1:200; Abcam, ab30394), PE Mouse Anti-Human CDCP1 (1:200, BD Biosciences, 566656), Anti-Human NRP2 (1:200; Abcam, ab273584), Anti-Human PTPRS (1:200; Proteintech, 13008-1-AP), Anti-Human PTPRD (1:200; Biograde tech, BGT-ANT-55876), and Anti-Human PTPRF (1:200; Thermo Fisher, MA5-38615), on ice for 30 minutes with gentle mixing. After washing with washing buffer, cells were incubated with Alexa Fluor 488-conjugated secondary antibodies (1:400, Abcam) for 20 minutes with gentle mixing. For aptamer staining, cells were incubated with FAM or Alexa Fluor 647-labeled aptamers at a final concentration of 100 nM (200 nM for enriched aptamer library) in Binding buffer at 4°C for 25 minutes, followed by washing with washing buffer. For physiological temperature assays, aptamer binding was assessed at 37°C under identical conditions (25 min incubation), unless otherwise specified. Samples were analyzed using a flow cytometer, and data were processed with FlowJo software. For affinity measurements, cells were incubated with increasing concentrations of fluorescently labeled aptamers (0.01 nM to 100 nM). The geometric mean fluorescence intensity was plotted against aptamer concentration, and dissociation constants (K_D) were calculated using GraphPad Prism through nonlinear regression fitting. Dissociation rate (k_{off}) = Fluorescence variation per unit time / Initial fluorescence value.

Cell-SELEX

All selection incubations were conducted at 4°C to minimize active internalization of surface proteins and reduce nonspecific uptake, in accordance with standard Cell-SELEX procedures.

Round 1 selection: 1. Library Preparation: Dissolve 1 nmol of TSO Single-Stranded DNA (ssDNA) library (synthesized by Sangon Biotech) (5′-CTCGTGGGCTCGGAGATGTGTATAAGAGACAG-N46-GCAGCTC-GGCCCATATAAGAAA-3′) in 100 μL DPBS buffer. Denatured ssDNA at 95°C for 10 minutes, and rapidly cooled on ice for 10 minutes to allow proper refolding.

2. Cell Preparation: Culture SUM159PT (target) and MCF-10A (counter-selection) cells to approximately 85% confluency. Harvest about 5×10^7 cells of each line for the first round selection.

3. Pre-washing: Rinse both cell populations twice with 1 mL washing buffer.

4. Negative selection: Combine the refolded ssDNA library with 900 μL of binding buffer. Incubate with MCF-10A cells at 4°C on a shaker for 30 minutes and collect the supernatant containing unbound sequences.

5. Positive selection: Transfer the supernatant to SUM159PT cells and incubate at 4°C for 25 minutes. Discard supernatant and wash SUM159PT cells three times (3 minutes each) with 1 mL washing buffer, discard each wash.

6. Elution of bound sequences. Add 300 μL of DNase-free ultrapure water to the adherent cells, scrape the cells from the dish, and transfer the cell suspension into a 1.5 mL microcentrifuge tube. Heat the mixture at 95°C for 10 minutes to release the bound DNA, then collect the supernatant containing the eluted sequences.

7. PCR Amplification:

(1) PCR1 Setup: The first-round PCR reaction consisted of 300 μL lysate supernatant, 100 μL of 5 × TSO PCR mix, and 100 μL ultrapure water. The 5 × TSO PCR mix was composed of 1 mL 10× PCR buffer, 600 μL of 2.5 mM dNTPs, 100 μL of 50 μM forward primer (TSO-For-FAM; FAM-CTCGTGGGCTCGGAGATGT-3′), 100 μL of 50 μM reverse primer (TSO-Rev-biotin; Biotin-TTTCTTATATGGGCCGAGCTGC-3′), 50 μL of Taq DNA polymerase (TaKaRa Ex Taq^{TM}, with Mg^{2+}, TKR-RR001A), and 150 μL of ddH_2O. All reagents were stored at −20°C. The annealing temperature was initially set at 60°C, but optimal conditions were determined empirically using gradient qPCR with nonfluorescent primers (TSO-For and TSO-Rev) across multiple temperatures (56, 58, 60, 62, and 64°C). The earliest amplification onset was used to select the optimal annealing temperature. A total of 11 cycles was used for PCR1.

(2) PCR2 Cycle Number Optimization: To optimize the number of amplification cycles, PCR2 was performed with 75 μL ddH_2O, 5 μL of PCR1 product, and 20 μL of 5 × TSO PCR mix. The mix was aliquoted into five reaction tubes. Amplifications were performed at the optimized annealing temperature (60°C), using 8, 10, 12, 14, or 16 cycles. The PCR

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

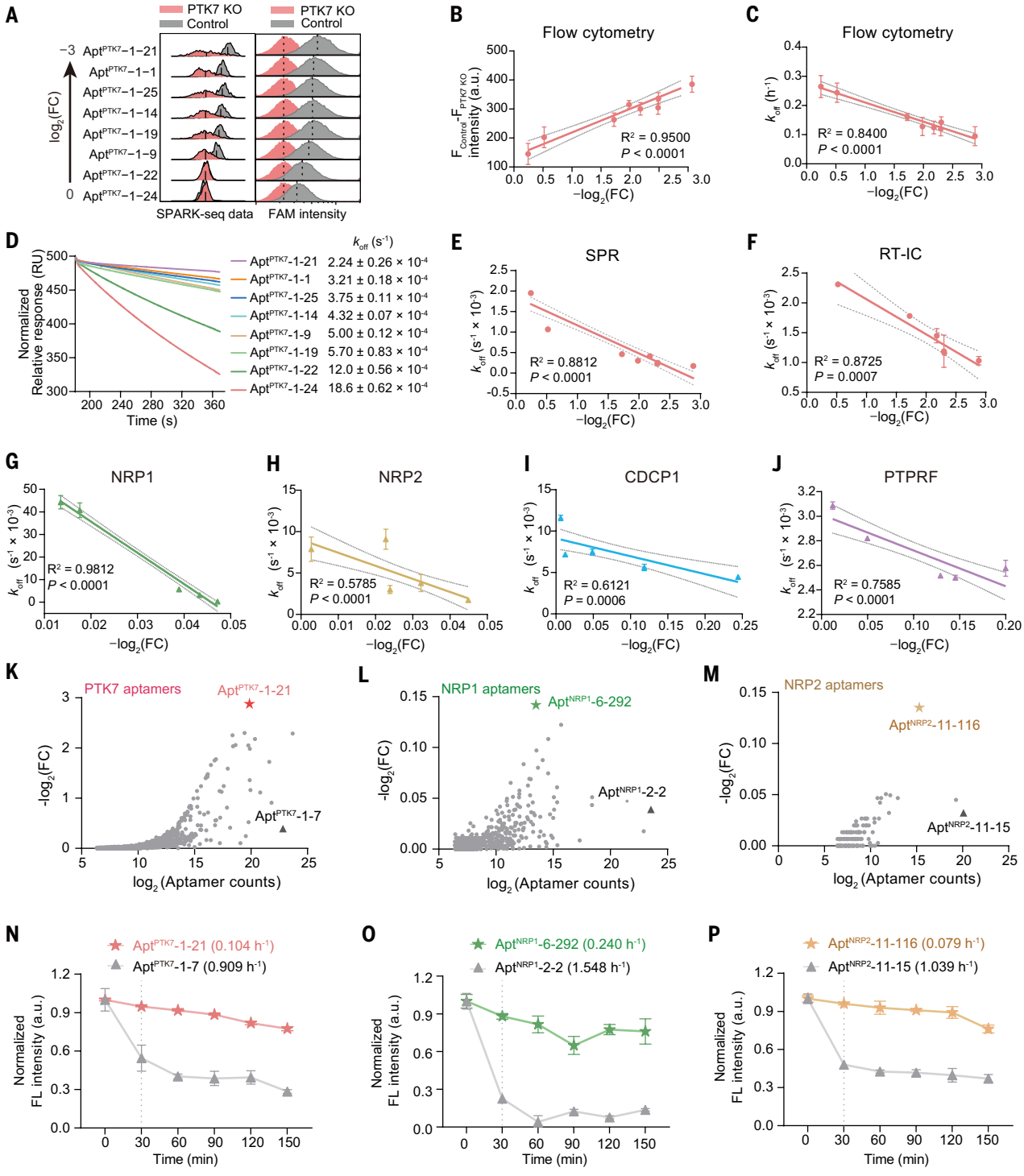


Fig. 6. Identification of aptamers with distinct dissociation kinetics. (**A**) Differential binding of PTK7 targeting aptamers. (Left) log_2FC from SPARK-seq data for indicated aptamers. (Right) Flow cytometry analysis of indicated aptamer binding in PTK7-knockout versus control cells. (**B** and **C**) Correlation between SPARK-seq-derived differential binding (log_2FC) and flow cytometry binding intensity (B) and Dissociation rate constants (k_{off}) (C). Each point represents an individual aptamer clone ($n = 3$ biological replicates, mean ± SD). Statistical significance was calculated using Spearman's rank correlation (two-tailed). (**D**) Representative dissociation curves of selected PTK7-binding aptamers and corresponding fitted k_{off} values measured using SPR ($n = 3$ biological replicates; mean ± SD). (**E** and **F**) Correlation between SPARK-seq-derived differential binding (Log_2FC) and k_{off} values from SPR (E) or RT-IC (F). Each point represents an individual aptamer clone ($n = 3$ biological replicates, mean ± SD). Statistical significance was calculated using Spearman's rank correlation (two-tailed). (**G** to **J**) Correlation between SPARK-seq-derived differential binding (Log_2FC) and SPR k_{off} for selected aptamers targeting NRP1 (G), NRP2 (H), CDCP1 (I), and PTPRF (J). Each point represents an individual aptamer clone ($n = 3$ biological replicates, mean ± SD). Statistical significance was calculated using Spearman's rank correlation (two-tailed). (**K** to **M**) Scatter plots of aptamer abundance (*x*-axis) versus $-log_2FC$ (*y*-axis) for aptamers targeting PTK7 (K), NRP1 (L) and NRP2 (M). Star indicates aptamer with highest $-log_2FC$ value, and triangle indicates aptamer with highest copy number based on aptamer bulk sequencing. (**N** to **P**) Time-resolved dissociation curves of indicated aptamer targeting PTK7 (N), NRP1 (O), and NRP2 (P), measured by flow cytometry. Data represent mean ± SD from three biological replicates.

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

products were analyzed on agarose gels to determine the optimal cycle number based on the presence of a single strong band without nonspecific amplification.

(3) PCR3 Amplification: PCR3 was conducted under the optimized cycling conditions (determined from PCR2). The reaction composition was the same as in PCR2. PCR3 products were used for subsequent ssDNA preparation.

8. ssDNA recovery:

(1) Add 70 μL of streptavidin agarose bead (cytiva, 17511301) slurry into a custom-packed 1 mL microcolumn. Wash the beads once with 1 mL of DPBS.

(2) Load the entire PCR3 product onto the column and wash with 1 mL of DPBS.

(3) Elute the bound biotinylated single-stranded DNA by flushing with 200 μL 0.2 M NaOH (supplemented with 0.2 M NaCl), collecting the flow-through.

(4) Desalt the eluate using a NAP-5 desalting column (cytiva, 17085302) pre-equilibrated with 15 mL of deionized water. Load with 200 μL ssDNA sample, rinse with 300 μL ultrapure water, then elute ssDNA with 700 μL ultrapure water.

(5) Quantify ssDNA by NanoDrop. Dry the ssDNA under vacuum and store at −20°C.

Rounds 2 to 4 of Cell-SELEX Selection

1. Library preparation:Heat the recovered ssDNA at 95°C for 10 minutes, then chill on ice before resuspending in binding buffer.

2. Increasing stringency: To enrich for high-affinity, highly specific aptamers, tighten selection criteria in each successive round by:

(1) Reducing target cell input. Rounds 2 and 3 use 2×10^7 SUM159PT cells; Round 3 uses 5×10^6; Round 4 uses 3×10^6.

(2) Modulating counter-selection: Include 2×10^7 MCF-10A cells in Round 1; omit in later rounds.

(3) Lowering ssDNA load. Reduce both the total amount and concentration of input ssDNA. Round 1: 3030 pmol; Round 2: 50 pmol; Round 3: 20 pmol; Round 4: 10 pmol ssDNA.

(4) Strengthening washes. Here, washing buffer is used to remove the unbound sequences. In Round 1, cells are washed twice. The number of washes increases by two each round. Additionally, blocker concentrations can be adjusted if needed.

3. Validation. After each round, analyze the ssDNA pool on an 8% denaturing polyacrylamide gel (PAGE). If off-target bands appear, excise the correct band, purify with a spin-column gel extraction kit (Sangon Biotech, B610357), and reamplify by PCR to regenerate the pool.

Aptamer library pulldown and LC-MS/MS-based target identification (Apt library IP-MS)

To enable aptamer library-based protein pulldown, the standard FAM-labeled ssDNA library was converted to biotin-labeled form. The workflow used in this study was as follows:

1. Biotinylated Aptamer Preparation

For each selection round, 20 μL of PCR product was mixed with 400 μL of 5× biotin-TSO PCR mix, which consisted of 1 mL 10× PCR buffer, 600 μL 2.5 mM dNTPs, 100 μL 50 μM forward primer TSO-For-biotin (biotin-CTCGTGGGCTCGGAGATGT-3′), 100 μL 50 μM reverse primer TSO-Rev-polyA20 (AAAAAAAAAAAAAAAAAAAA/iSp18/TTTCTTATATGGG-CCGAGCTGC-3′), 50 μL Taq DNA polymerase (TaKaRa, TKR-RR001A), and 150 μL ddH_2O (stored at −20°C). The final volume was adjusted with sterile water, and PCR was carried out for 20 cycles at 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s.

2. DNA Concentration and Purification

The PCR product was transferred to a 15 mL centrifuge tube, then added 4-5 volumes of butanol for DNA precipitation. The mixture was inverted to mix thoroughly. If no turbidity formed, 50–100 μL of ultrapure water was added. Samples were centrifuged at 7,500 × g for 10 minutes (tube strength pre-validated to prevent cracking). After centrifugation, the solution became clear and phase-separated. The lower PCR-enriched phase was transferred to 1.5 mL tubes and centrifuged at 13,000 × g for 2 min. The remaining butanol was aspirated, leaving ~100 μL concentrated product.

3. Denaturing PAGE and ssDNA Purification

2 × TBE-Urea Sample Buffer (Sangon Biotech, C506046) was added to the concentrated product. Samples were loaded onto an 8% denaturing PAGE and electrophoresed at 300 V for 15 minutes. Based on a DNA ladder, the 100 nt band was excised (a faint FAM signal at ~100 nt may be visible under UV due to residual FAM-labeled ssDNA; care was taken to avoid the 120 nt antisense strand band containing TSO-Rev-polyA20). The band was extracted using the EZ-10 Spin Column DNA PAGE Gel Extraction Kit (Sangon Biotech, B610357), and purified DNA was quantified using a NanoDrop spectrophotometer.

4. LC-MS/MS-Based Target Identification

Target proteins interacting with the enriched aptamer libraries were identified via liquid chromatography-tandem MS (LC-MS/MS).

a. Membrane Protein Extraction: SUM159PT ($> 5 \times 10^7$) cells were digested using enzyme-free cell dissociation buffer at 37°C for 6 minutes. After removing the supernatant, membrane proteins were extracted using the Membrane and Cytosol Protein Extraction Kit (Beyotime, P0033).

b. Aptamer Library-Based Pulldown: For each of rounds R1-R4, 100 nM biotinylated enriched aptamer (with the biotin-labeled initial library R0 as a control) was heat-denatured at 95°C and refolded at 4°C. Each aptamer library was then incubated with membrane proteins (pre-blocked with binding buffer containing 10% FBS for 1 h) at 4°C with gentle rotation for 1 h. Subsequently, 50 μL streptavidin (SA) agarose beads (cytiva, 17511301), pre-blocked in TBST with 5% BSA for 1 h, were added and incubated at 4°C for an additional hour. After centrifugation, beads were washed three times with washing buffer to remove unbound proteins.

c. Protein Elution and SDS-PAGE: Equal volumes of protein loading buffer (Sangon Biotech, C506025) were added to the bead pellets. Samples were denatured at 100°C for 10 minutes, cooled at 4°C for 10 minutes, and centrifuged to collect the supernatant. SDS-PAGE was performed using a 4–20% gradient gel (GenScript, M42012C) in GenScript electrophoresis buffer. Wells were cleaned using electrophoresis buffer prior to loading. Samples (20 μL) were loaded; if more sample was needed, an additional 20 μL was loaded after 10 minutes at 60 V. Gels were run at 60 V for 30 minutes, then at 120 V for 70 minutes until the bromophenol blue front reached the bottom. Gels were stained with Coomassie Blue Staining Solution (Sangon Biotech, E607056) for over 2 hours and destained using Coomassie Destaining Solution (Sangon Biotech, E607057) until bands were clearly visible. Gel bands were excised and sent to the facility core of Hangzhou Institute of Medical Sciences-CAS for proteomic analysis.

5. Data Analysis and Filtering

To avoid interference from false signals, nonmembrane proteins were filtered out from the protein list using the UniProt database of membrane proteins, leaving 239 membrane proteins (see data S2). In the negative control experiment, the natural library, also referred to as R0, biotin-labeled and attached to streptavidin-coated magnetic beads, was used to eliminate DNA-binding and bead-binding proteins. Proteins were ranked based on differences between each round and R0. We considered only proteins with high confidence in MS identification, i.e., proteins bound in R4 with counts greater than 8. After comprehensive evaluation, 13 target proteins were selected for subsequent identification: PTPRF, NRP1, PTK7, ITGA3, NRP2, ITGB1, PTPRS, PTPRD, CD151, SLC25A5, RAB23, RAB5C, and CDCP1.

Aptamer affinity purification and LC-MS/MS (Apt IP-MS)

To identify the target proteins of individual aptamer candidates, we employed immunoprecipitation coupled with MS (IP-MS) to analyze the proteins enriched by each aptamer. The detailed procedure is as follows:

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

a. Membrane Protein Extraction: SUM159PT cells ($> 5 \times 10^7$) were dissociated using enzyme-free dissociation buffer at 37°C for 6 minutes. After removing the supernatant, membrane proteins were extracted using the Membrane and Cytosol Protein Extraction Kit (Beyotime, P0033).

b. Aptamer–Protein Affinity Purification: Each biotin-labeled aptamer (100 nM) was heat-denatured at 95°C and refolded at 4°C. As a control, the initial biotin-labeled library (R0) was included. The refolded aptamers were incubated with the membrane protein extracts (pre-blocked with binding buffer containing 10% FBS for 1 h) at 4°C with gentle rotation for 1 h. Streptavidin MagPoly Beads (smart-Lifesciences, SM017010), pre-blocked with TBST buffer containing 5% BSA for 1 h, were added (50 μL per sample), and the mixture was further incubated at 4°C for an additional hour. After incubation, beads were collected by centrifugation and washed three times with washing buffer to remove unbound proteins. To reduce nonspecific interactions, the beads were optionally washed once with 0.01% dextran sulfate sodium (Beyotime, D5143; dissolved in washing buffer) for 2-5 minutes.

c. Protein Elution and MS: The precipitated protein–aptamer complexes were resuspended in 100 μL ddH_2O and denatured at 95°C for 10 minutes. Following magnetic separation, the supernatant was collected and submitted for LC-MS/MS analysis (Shanghai Bioprofile Technology Company Ltd; HFX, 60 minutes).

d. To identify high-confidence membrane-bound targets, nonmembrane proteins were filtered out using the UniProt database, retaining only plasma membrane proteins. Proteins were further selected based on the vendor-recommended Score threshold (Score ≥ 20), which reflects the match quality between the observed MS/MS spectra and theoretical peptide spectra—higher scores indicating greater confidence. For downstream analysis, we compared aptamer-enriched samples against Ctrl-lib IP-MS results using the following formula: $\log_2$[(Aptamer Intensity Average + 1) / (Ctrl-lib Intensity Average + 1)]. Proteins were then ranked by fold change, and the top-ranked protein was designated as the primary candidate target of the aptamer.

Affinity purification and LC-MS/MS of antibodies

To immunoprecipitate target proteins and their interacting partners from cell lysates, cells were washed with DPBS and lysed using a gentle lysis buffer containing protease and phosphatase inhibitors. Lysates were incubated with 3 μg of specific antibodies against the target proteins overnight at 4°C on a rotating shaker. The next day, Protein A/G magnetic beads (Thermo Scientific, 88803) were washed and added to the antibody-protein mixtures, followed by incubation at room temperature for 1 hour. Beads were washed three times with freshly prepared DPBST to remove nonspecific proteins. Bound proteins were eluted by adding 40 μL of 2× loading buffer and heating at 95°C for 10 minutes. Proteins were analyzed by Western blotting using specific antibodies to detect target proteins and their interactors.

For MS, proteins were digested directly on the beads with trypsin. Peptides were eluted, desalted using C18 columns, and analyzed by LC-MS/MS.

Proteomic analysis of SUM159PT cells

More than 2×10^7 SUM159PT cells were dissociated using enzyme-free cell dissociation buffer, washed once with DPBS, and the cell pellets were flash-frozen in liquid nitrogen. Samples were shipped on dry ice to a commercial service provider (Westlake Omics) for proteomic analysis using the Astral DIA 24-min workflow. Three independent biological replicates were analyzed (see data S7), and the averaged protein intensity values were used for downstream quantification and visualization.

Real-time interaction cytometry (RT-IC)

RT-IC was performed using the heliXcyto® system (RT-IC; Dynamic Biosensors), which integrates permeable microfluidic cell traps with time-resolved fluorescence detection, enabling kinetic analysis of aptamer-target binding on the surface of live cells. SUM159PT or A549 cells were seeded in culture dishes 48 h prior to the experiment. Upon reaching ~90% confluence, cells were detached using enzyme-free dissociation buffer and washed once with Binding buffer. Cells were then resuspended in buffer A (washing buffer supplemented with 2% FBS, 0.1 mg/mL BSA, and 0.01 mg/mL salmon sperm DNA), filtered through a 30 μm cell strainer, and gently introduced into the microfluidic traps (15-25 μm) of the heliXcyto® chip. Trapped cells were equilibrated at 25°C in DPBS containing 1 mM $MgCl_2$. FAM-labeled aptamers (10 nM) were injected under continuous flow. The association phase typically lasted 420 s, followed by a dissociation phase of 1200 s using buffer only. Fluorescence signals were recorded in real time for individual cells. The association flow rate was 25 μL/min, and the dissociation flow rate was 50 μL/min. Raw sensorgrams were double-referenced by subtracting signals from cell-free traps and baseline buffer signals. Data were normalized to account for autofluorescence and nonspecific binding. Curve fitting was performed using GraphPad Prism software, applying a global nonlinear least-squares fit to a 1:1 Langmuir binding model over the entire association and dissociation phases. Extracted kinetic parameters included the association rate constant (k_{on}), dissociation rate constant (k_{off}), and the K_D.

Surface plasmon resonance (SPR)

Instrument software settings were configured according to the manufacturer's instructions. Control molecules such as His-tag peptides or IgG were immobilized on Channel 1, while the protein of interest was immobilized on Channel 2. Protein immobilization was performed using default instrument parameters. Specifically, equal volumes of 0.4 M EDC and 0.1 M NHS were pre-mixed and added to a 96-well microplate to activate carboxyl groups on the CM5 sensor chip. Target proteins were diluted in 10 mM sodium acetate buffer and covalently immobilized for 900 s. The chip surface was then blocked with ethanolamine. All immobilization steps were conducted at a flow rate of 10 μL/min. Aptamers were first diluted to 20 μM in DPBS, heat-denatured at 95°C and refolded at 4°C. They were then further diluted to 200 nM or serially diluted into a concentration gradient using SPR running buffer (DPBS supplemented with 1 mM $MgCl_2$). Binding was assessed using a Biacore SPR instrument with an association time of 120 s and dissociation time of 180 s. Regeneration between cycles was performed with 1.5 M NaCl at a flow rate of 30 μL/min. Kinetic parameters including the K_D, k_{on}, and k_{off} were calculated using Biacore Insight Evaluation Software (v5.0.18.22102) with a 1:1 Langmuir binding model. Raw sensorgram data were exported as PowerPoint plots and further visualized using GraphPad Prism. Affinity and dissociation values were determined based on response units (RU) from three biologically independent replicates. All SPR related experiments, including immobilization and binding, were performed at 25°C unless otherwise specified.

Regarding SPR experiments, we now specify that recombinant human PTK7 protein (Met1–Thr704; Sino Biological, 19399-H08H) and other recombinant proteins, including CDCP1 (Phe30–Leu666; R&D Systems, 10402-CU); NRP1 (Phe22–Lys644; R&D Systems, 3870-N1), NRP2 (Gln23–Tyr855; R&D Systems, 2215-N2), ITGA3 (Phe33–Glu991)/ITGB1 (Glu21–Asp728) heterodimer (R&D Systems, 2840-A3), PTPRD (Glu21–Ser1174; R&D Systems, 9995-PR), PTPRF (Ala27–Glu1251; R&D Systems, 9377-PF), and PTPRS (Glu30–Gly1260; R&D Systems, 3430-PR) were purchased from the indicated suppliers.

Microscale thermophoresis (MST)

Binding affinities between aptamers and target proteins were measured using microscale thermophoresis (MST) on the Monolith NT.115 instrument (NanoTemper Technologies). Aptamers were 3′-labeled with Alexa Fluor 647 or Cy5 (hydrophilic) and resuspended in MST buffer (DPBS supplemented with 1 mM $MgCl_2$) to a final concentration

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

of 10 nM or 15 nM. Recombinant target proteins (Same as SPR) were serially diluted in MST buffer across 16 concentrations. Labeled aptamers were incubated with protein dilutions for 25 minutes at 4°C to ensure stable complex formation. After incubation, mixtures were loaded into standard treated capillaries (NanoTemper, MO-K022). Thermophoresis measurements were recorded at 22°C with LED power set to 40% (or 20%).

K_D values were determined using the built-in K_D model in MO.Affinity Analysis software (NanoTemper). Additionally, binding curves were replotted in GraphPad Prism using the "Dose-response - Stimulation" model with log(agonist) vs. response - Variable slope (four parameters). All measurements were performed in at least three independent biological replicates, and results are expressed as mean ± SD.

Aptamer immunoprecipitation assay (Apt IP) and western blot analysis

To investigate aptamer-target protein interactions, biotinylated aptamers were subjected to immunoprecipitation followed by Western blotting. Biotin-labeled aptamers were dissolved in DPBS, heat-denatured at 95°C and refolded at 4°C. A total of 300 pmol Biotin-Aptamer was mixed with an equal volume of binding buffer (containing 5 mM $MgCl_2$) and briefly incubated on ice for pre-blocking. Each aptamer was paired with one 15 cm dish of SUM159PT cells (at ~90% confluency > 10^8 cells). Cells were harvested by enzymatic-free dissociation and centrifuged at 300 × g for 5 min. The pellet was washed twice with washing buffer (containing 5 mM $MgCl_2$), followed by 300 × g centrifugation for 5 minutes after each wash. Cells were then lysed in 1 mL RIPA buffer (APPLYGEN, C1053) supplemented with protease inhibitors (Roche, 81410500) and phosphatase inhibitors (Beyotime, P1082) at 4°C for 15 min. Lysates were centrifuged at 13,000 × g for 10 min at 4°C, and the supernatant (total cell lysate) was collected; a portion was retained as input control for IP-WB. Lysates were blocked at a 2:1 volume ratio with binding buffer containing 20% FBS at 4°C for 1 hour with gentle inversion. The pre-blocked Biotin-Aptamer was incubated with the lysate at 4°C for 1 hour to allow formation of aptamer-target complexes.

In parallel, streptavidin magnetic beads (Smart-Lifesciences, SM017010) were prepared at 30 μL per aptamer sample. Beads were washed with 500 μL TBST. (For antibody-based control IP: To 500 μL of lysate, add 1:50 (v/v) of target-specific IP antibody or isotype IgG, incubate overnight at 4°C with inversion. Use 30 μL Protein A/G magnetic beads per sample). Beads were blocked in 0.5 mL TBST containing 5% BSA for 1 h at 4°C, washed with 1 mL TBST, and resuspended in 30 μL RIPA buffer. These were then incubated with the aptamer-protein complexes at 4°C for 1 hour. After incubation, beads were separated, and the supernatant was discarded. Beads were washed once with 1 mL TBST on ice and once with 1 mL washing buffer on ice, then resuspended in 30 μL RIPA buffer. Samples were denatured by adding SDS loading buffer (Sangon Biotech, C506032-0005) and heating at 95°C for 10 min. Proteins were resolved by 4–20% SDS-PAGE gradient gel electrophoresis. A pre-stained protein marker was used to indicate molecular weight. Proteins were then transferred to a PVDF membrane (Millipore, 0000377714) at 80 V for 120 minutes in an ice bath using wet transfer. PVDF membranes were pre-activated in 100% methanol for 1–2 minutes before use. Blocking was performed with 5% (w/v) nonfat dry milk (Beyotime, P0216-300G) in TBST for 1 hour at room temperature. Membranes were incubated overnight at 4°C with primary antibodies at a recommended dilution. The following day, membranes were washed three times with TBST and then incubated with HRP-conjugated secondary antibodies (1:5000 dilution) at room temperature for 1 hour. Following washes, membranes were incubated with ECL substrate (biosharp, BL520B), and chemiluminescent signals were detected using an Amersham ImageQuant 800 imaging system.

Immunoblotting (IB) was performed with the indicated antibodies: anti-PTK7 (CST, 25618, 1:2000), anti-CDCP1 (CST, 4115S, 1:2000), anti-NRP1 (HUABIO, ET1609-69, 1:2000), anti- NRP2 (Proteintech, 27193-1-AP, 1:2000), anti- ITGA3 (HUABIO, HA600100, 1:1000), anti-PTPRF (Proteintech, 14138-1-AP, 1:1000). Co-immunoprecipitation (Co-IP) analysis of lysates from SUM159PT cells. Lysates were immunoprecipitated (IP) with anti-ITGA3 antibody (Invitrogen, MA5-28565), anti-NRP2 antibody (Proteintech, 27193-1-AP), or control IgG (Wanleibio, WLA125). Immunoblotting was performed with the indicated antibodies: anti- ITGA3 (HUABIO, HA600100, 1:1000), anti-CD151 (Proteintech, 66567-1, 1:10000), anti- NRP2 (Proteintech, 27193-1, 1:2000), anti-SLC25A5 (Abclonal, A15639, 1:2000).

Flow cytometry: Cross-competition among targets

SUM159PT or A549 cells were seeded into 15 cm cell culture dishes 48 hours in advance. Once the cultures reached 90% confluency, cells were harvested by enzymatic-free dissociation and centrifuged at 300 × g for 5 minutes. Cell pellet was washed sequentially with DPBS and binding buffer. After another 5 min centrifugation at 300 × g, the pellet was resuspended in binding buffer based on a final suspension volume of 50 μL per sample.

Fluorescently labeled aptamers (Alexa Fluor 647 or FAM-labeled Binder-Apt) and unlabeled competitor aptamers (Competitor-Apt) were dissolved in DPBS, denatured at 95°C for 10 minutes, and rapidly cooled at 4°C for 30 minutes to allow proper refolding. Here, Competitor-Apt stands of aptamers targeting distinct proteins. To initiate the competition assay, 20 pmol Binder-Apt (or a saturating concentration of aptamer) was added to 50 μL of binding buffer. After a short incubation, the aptamer was mixed with 50 μL of cell suspension and incubated for at least 10 minutes. Subsequently, 200 pmol Competitor-Apt—either homologous aptamers at varying concentrations or excess (≥20-fold) competitor proteins—was added, gently mixed, and incubated on ice in the dark for 30 minutes with gentle mixing. Following incubation, cells were centrifuged at 300 × g for 5 minutes. The supernatant was removed, and cells were washed once with washing buffer. Cells were then resuspended in 100 μL washing buffer, filtered through a 40 μm cell strainer, and analyzed by flow cytometry to assess fluorescence intensity.

The protein competitors used in this assay included: VEGF165 (11066-HNAH), TGFBR1 (10459-H02H), ANGPTL4 (10563-H01H), VEGF-C (10542-H08H), SARS-CoV-2 (2019-nCoV) Spike S1 (40591-V08H), HGF (10463-HNAS), PDGFRA (10556-H02H), and PGF (10274-H05H), all obtained from SinoBiological.

Flow cytometric quantification of surface protein expression with fluorescence calibration

Primary antibodies used in these experiments were: Anti-Human ITGA3 (Thermo Fisher, MA5-28565), Anti-Human ITGB1 (Sangon Biotech, D194637), Anti-Human PTK7 (NOVUS, NBP2-01486), Anti-Human CDCP1 (Sangon Biotech, D199544), Anti-Human NRP1 (Proteintech, 60067-1-Ig), Anti-Human NRP2 (Abcam, ab273584), Anti-Human PTPRS (Proteintech, 13008-1-AP), Anti-Human PTPRD (Biograde Tech, BGT-ANT-55876) and Anti-Human PTPRF (Thermo Fisher, MA5-38615). Alexa Fluor 647–conjugated species-matched secondary antibodies were used for detection. The Quantum MESF Kit (Bangs Laboratories) was used for quantitative fluorescence calibration.

SUM159PT were seeded in 15 cm cell culture dishes 48 hours prior to the assay. Once cultures reached 90% confluency, cells were harvested by enzymatic-free dissociation and centrifuged at 300 × g for 5 min. The supernatant was discarded, and the cell pellet was washed once with DPBS, followed by a second centrifugation at 300 × g for 5 minutes. Cells were then resuspended in flow cytometry binding buffer (DPBS containing 2% FBS and 0.05% NaN_3), with the volume adjusted based on a final suspension of 50 μL per sample. Primary antibodies were diluted at 1:50 using flow cytometry binding buffer and added to the cell suspension at 50 μL per sample, resulting in a final antibody dilution of 1:100. Cells were incubated at 4°C for 30 minutes with gentle mixing. After incubation, cells were centrifuged at 300 × g for 5 minutes, and the supernatant was discarded. Alexa Fluor 647-conjugated

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

secondary antibodies (species-matched) were diluted at 1:500 in flow cytometry binding buffer and used to resuspend the cell pellet in 100 μL per sample. Cells were incubated in the dark at 4°C for 20 minutes with gentle mixing, followed by centrifugation at 300 × g for 5 minutes. The supernatant was discarded. Cells were then washed once with 200 μL of washing buffer and finally resuspended in 100 μL of washing buffer. Samples were filtered through a 40 μm cell strainer and analyzed by flow cytometry to measure fluorescence intensity.

For quantitative fluorescence calibration, the Quantum MESF Kit (Bangs Laboratories) was used. Calibration beads were diluted in washing buffer and analyzed on the same flow cytometer. Data analysis was performed using FlowJo software in combination with the QuickCal® v3.0 quantitative analysis template from Bangs Laboratories (www.bangslabs.com/quickcal).

5′ end single-cell perturbation sequencing

For concept validation in Fig. 2, aptamer sgc8c was denatured, renatured, and then dissolved in binding buffer. Cells (SUM159PT, SUM159PT PTK7 g1, SUM159PT PTK7 g2, EMT6, EMT6 PTK7 g3, and EMT6 PTK7 g4) were detached using enzyme-free dissociation solution and resuspended in binding buffer. Equal amounts of cells (1×10^6 each) were incubated with 200 nM sgc8c at 4°C for 30 minutes. Cells were centrifuged, the supernatant was discarded, and cells were washed three times. Finally, cells were resuspended in washing buffer containing 1 mM $MgCl_2$ and passed through a 40 μm filter to prepare single-cell suspensions. After 30 minutes at 4°C, single-cell RNA-seq libraries were generated using the 10X Genomics Next GEM Single Cell 5′ Reagent Kits v2 (Dual Index), following the manufacturer's protocol. CRISPR-perturbed sequences were integrated into the library by capturing the gRNA transcript barcode (5′-GGCTAGTCCGTTATCAACTTG-3′), and aptamer sequences were integrated into the library by capturing the sgc8c-S barcode (identical to the gRNA barcode). Libraries were pooled and sequenced on an Illumina NovaSeq 6000 platform. Data were processed using the Cell Ranger pipeline, and differential expression analysis was performed using RStudio. Cells were quality-controlled and separated by species, human or mouse, through mRNA analysis. Cell clustering was confirmed by analyzing guide RNA expression (*23*).

For target identification of the enriched aptamer library, the R4 library dissolved in DPBS was denatured at 95°C for 10 minutes, renatured at 4°C for 10 minutes, and then diluted in binding buffer to a final concentration of 8 μM. Pre-mixed cells perturbed with different genes (a total of 46 gRNAs; cells were mixed equally 3 days in advance) were detached using enzyme-free dissociation solution and washed once with washing buffer. Then, 1×10^6 cells were diluted in 50 μL binding buffer. After 10 minutes, 50 μL of the R4 library (8 μM) were added and incubated at 4°C for 30 minutes. Cells were centrifuged, washed three times, and resuspended in DPBS containing 1 mM $MgCl_2$. Single-cell suspensions were prepared by passing cells through a 40 μm filter. After 30 minutes at 4°C, single-cell RNA-seq libraries were generated by LC-Bio Technologies (Hangzhou, China) using the 10X Genomics Next GEM Single Cell 5′ Reagent Kits v2 (Dual Index), following the manufacturer's protocol. CRISPR-perturbed sequences were integrated into the library by capturing the gRNA transcript barcode (5′-GGCTAGTCCGTTATCAACTTG-3′), and aptamer sequences were integrated into the library by capturing the aptamer library barcode (5′-GCAGCTCGGCCCATATAAGAAA-3′). Libraries were pooled and sequenced on an Illumina NovaSeq 6000 platform. Data were processed using the Cell Ranger pipeline, and differential expression analysis was performed using RStudio.

Reverse transcription quantitative PCR (RT-qPCR)

Primers were designed to target regions downstream of gRNA target sites, with one primer spanning two exons to ensure specificity. Primer sequences (5′–3′) were as follows: CD151-For CGGAGCTCAAGGAGAACCTG, CD151-Rev CGGATCCACTCACTGTCTCG; SLC25A5-For GACACTGCAAAGGGAATGCTT, SLC25A5-Rev GCAGTGACAGTCTGTGCGAT; and GAPDH-For TTCCAGGAGCGAGATCCCT, GAPDH-Rev GGCTGTTGTCATACTTCTCATGG. Total RNA was extracted from cells using the FastPure Cell/Tissue Total RNA Isolation Kit V2 (Vazyme Biotech Co., Ltd., RC102). RNA (1 μg) was reverse-transcribed using the HiScript III All-in-one RT SuperMix Perfect for qPCR kit. Quantitative PCR (qPCR) was performed using Taq Pro Universal SYBR qPCR Master Mix (Vazyme, Q712-02) on a CFX96 real-time PCR detection system (Bio-Rad). Relative mRNA expression levels were normalized to GAPDH, and fold changes were calculated using the $2^{-\Delta\Delta Ct}$ method. Data were plotted using GraphPad Prism.

Statistical analysis

All statistical analyses were conducted using GraphPad Prism 9 (GraphPad Software). Data are expressed as mean ± standard error of the mean (SEM) based on at least three independent biological replicates unless otherwise stated. Comparisons between two groups were performed using an unpaired, two-tailed Student's *t* test. For comparisons involving more than two groups, a one-way or two-way analysis of variance (ANOVA) was employed, followed by Tukey's post hoc test to account for multiple comparisons. A significance level of $P < 0.05$ was considered statistically significant for all analyses.

Analysis of human-mouse hybrid data

To process 10X Genomics single-cell sequencing data containing both human and mouse cells, we used Cell Ranger software (version 3.0.2). We first integrated human (Homo_sapiens.GRCh38.dna_sm.toplevel.fa) and mouse (Mus_musculus.GRCm38.dna_sm.toplevel.fa) genome assemblies into a combined reference following the Cell Ranger guidelines. Sequencing reads were then aligned and quantified based on the integrated annotation file, which combined Homo_sapiens.GRCh38.94.gtf and Mus_musculus.GRCm38.94.gtf.

Downstream analysis was performed using the Seurat package (version 4.2.1). To distinguish human from mouse cells in the mixed dataset, we calculated the proportion of human and mouse gene expression for each cell. Cells expressing more than 90% of either human or mouse genes were classified as human or mouse. High-quality cells retained for further analysis had to meet the following criteria: (1) mitochondrial RNA content <25% and (2) detection of > 500 genes. RNA expression data were normalized using the NormalizeData function. After scaling and centering the data, we performed principal component analysis (PCA). UMAP dimensionality reduction was conducted using the RunUMAP function with the top 20 principal components and default parameters.

Owing to inherent random errors during the sequencing process and minor discrepancies in the cell barcode process, false positives can occur in the detection of gRNA and aptamer expression. To address this issue, we applied GMM to accurately identify cells expressing the corresponding gRNA and aptamer. Specifically, for each cell, we $\log_2$-transformed the expression levels of gRNAs and the sgc8c DNA aptamer targeting PTK7. A two-component GMM was fitted using the 'normalmixEM' function from the mixtools package (version 2.0.0) to estimate appropriate thresholds. The threshold was set to the mean (μ) of the second Gaussian distribution minus 1.96 times the standard deviation (σ). Cells with expression levels exceeding the threshold were considered positive for the corresponding gRNA and aptamer and were subsequently visualized on a UMAP plot.

Data preprocessing of SPARK-seq

The raw data were processed using Cell Ranger (v.6.0.0), including the extraction and analysis of cell barcodes and UMI barcodes, followed by alignment to the reference genome 'Homo_sapiens.GRCh38.dna_sm.toplevel.fa'. This process generated gene expression and gRNA UMI count matrices through the Gene Expression Output and CRISPR Guide Capture Output modules, respectively. Aptamer sequences

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

(45–47 bp), located upstream of the adapter primer (GCAGCTCGG-CCCATATAAGAAA), were extracted from Read2 of the aptamer omics data, while cell barcode and UMI information were obtained from the corresponding Read1 sequence. We subsequently performed UMI-based deduplication to remove PCR duplicates. This enabled generation of the UMI count matrix for the aptamer sequences.

Cell quality assessment and filtering

Low-quality cells were filtered based on the following criteria: high mitochondrial gene content (>10%), low gene detection (<200), and low aptamer detection (<100). Downstream analysis was performed using the Seurat package (version 4.2.1), enabling integrated analysis of multi-modal single-cell datasets (transcriptomics, aptamer, and gRNA). Gene expression counts were log-normalized according to the standard Seurat workflow.

gRNA identity assignment in single cells

Based on gRNA expression data, each cell was assigned a corresponding gRNA identity. Specifically, cells with fewer than 200 counts for all gRNAs were classified as "Negative." For the remaining cells, the gRNA enrichment ratio was calculated as the proportion of counts from the most abundant gRNA relative to the total gRNA counts in the cell. Due to technical or biological variability, individual cells may carry one, two, or even more gRNA species. As a result, the distribution of dominant gRNA fractions across cells exhibits a mixed pattern, reflecting underlying heterogeneity. To identify cells predominantly expressing a single gRNA species for downstream analysis, we fitted a three-component (k = 3) GMM to the expression distribution. This analysis identified three distinct components with means ± SD of 0.199 ± 0.088, 0.554 ± 0.088, and 0.908 ± 0.088, respectively. The first two components represent cells with ambiguous or mixed gRNA signals, whereas the third component, characterized by dominant gRNA fractions close to 1 (mean ~0.91), likely corresponds to cells harboring a single, clearly dominant gRNA species. To establish a conservative and reliable threshold for defining "singlet" cells, we set the cutoff at 0.70, which corresponds to the 1st percentile of the third (high-confidence) Gaussian component. This ensures that cells included above this threshold fall within the high-confidence, single-gRNA group. In this way, cells with a ratio exceeding 70% were assigned to the corresponding gRNA; those not meeting this threshold were classified as "multiple". "Negative" and "multiple" cells were excluded from further analysis. Additionally, gRNA categories with fewer than 50 detected cells, or those without other gRNAs targeting the same protein, were removed. Finally, cells were grouped into two categories: those with effective gRNA-mediated knockout of the target proteins and control cells with nontargeting gRNAs (7 NC gRNAs).

Assessment of aptamer sequencing accuracy and reproducibility

We further assessed aptamer sequencing accuracy using three spiked-in control sequences, treating all reads within an edit distance of ≤10 as potential errors. All three controls exhibited consistent error rates of approximately 6%, indicating high sequencing fidelity. Sequencing reproducibility was assessed using three independent sequencing replicates. The top 50 high-frequency sequences showed consistent abundance across replicates. Pairwise comparisons of the abundances of the top 10,000 sequences were carried out by Pearson correlation analysis and yielded correlation coefficients >99%, confirming high technical reproducibility.

Construction of the aptamer family binding profile matrix

We analyzed the aptamer sequences, identifying 4.027×10^7 distinct sequences and generating 1.72×10^8 reads for which the top 10,000 sequences represented 68.84% of total reads. These top 10,000 sequences were selected for downstream analysis. For a detailed explanation of the rationale behind this selection, please refer to the section "Selection of Aptamer Sequence Subsets for Downstream Analysis.". These sequences were classified into 1906 families using the "BLAST-short-MCL" strategy from SMART-Aptamer (*33*), a tool we previously developed. In detail, BLAST searches were performed with BLAST v2.7.1+ (https://blast.ncbi.nlm.nih.gov/), using an all-versus-all nucleotide-nucleotide search (blastn-short mode) with an e-value threshold of 0.05. The BLAST results were used to construct an orthogroup graph wherein cognate aptamer pairs were connected by edges weighted according to normalized BLAST bit scores. Finally, aptamers were clustered into discrete families using MCL, and the aptamer family abundance matrix was generated based on the UMI count matrix of the aptamer sequences.

Aptamer-target interaction prediction

The 20 aptamer families with the highest overall binding abundance were selected for downstream analysis. For each aptamer family, a differential binding value was computed for every gRNA-defined cell group by subtracting the median binding abundance of the corresponding control group from that of the gRNA-defined group. Aptamer families exhibiting minimal variation in differential binding—defined as having no more than three unique differential values across all gRNA-defined groups—were excluded from further analysis, such as Families 18, 19, and 20. For each remaining aptamer family, the top three gRNA-defined cell groups with the highest differential binding values were considered to exhibit strong binding differences. gRNA-defined cell groups that appeared among the top three in more than half of the aptamer families were excluded, as they were considered likely to represent nonspecific or promiscuous binding effects.

For each remaining aptamer family, the binding differential values were further analyzed using GMM with the 'normalmixEM' function from the mixtools package (version 2.0.0). Specifically, to quantify the target-specific binding of each aptamer family, we first calculated the binding difference for each target-knockout cell as its expression minus the median expression in gRNA-negative control (NC) cell groups, and then defined the binding difference (BD) as the median of these per-cell differences. This metric captures the extent to which an aptamer's signal is altered upon removal of its target protein. Across candidate targets, the distribution of binding differences exhibited a clear bimodal pattern, consistent with two underlying biological scenarios: (1) aptamer–target pairs showing strong, specific binding—evidenced by a substantial signal decrease upon target knockout (i.e., lower values), and (2) aptamer–target pairs lacking specific interaction, displaying minimal or no change. Due to the high dispersion of the original value distribution, a $\log_2$transformation was applied prior to GMM fitting to improve component separation. We modeled this distribution using a two-component GMM, which revealed a component with lower values corresponding to aptamer–target pairs that responded strongly to target knockout—indicative of specific and meaningful binding. Due to noise factors, such as nonspecific binding, the overlap between the two peaks was relatively large. To minimize false positives, we selected the mean of the first component as the threshold. By using this mean as the cutoff, we prioritize aptamer–target pairs with binding differences more likely driven by true biological specificity, while minimizing the inclusion of cases driven by background noise or off-target effects.

For the remaining gRNA-defined cell groups, a scoring system was applied, aggregating the differential values of the gRNA groups for each potential target protein into a composite score ($\text{Score}_{\text{protein}} = \sum_{gRNA\ target\ to\ the\ protein} - BD_{gRNAs}$.). This score represented the binding potential of the protein for the aptamer family such that a higher score indicated a greater likelihood of binding. Predictions were made as follows:

If only one protein remained, it was considered the target of the aptamer family.

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

If two proteins remained, the protein with the larger binding score (≥ 2×) was selected; otherwise, both proteins were retained.

multiple proteins remained, the protein was selected if its binding score was ≥50% of the total; otherwise, all proteins were retained.

Target knockout refinement

We first quantified mRNA expression levels of predicted target genes in both control and gRNA-positive cells. For each gene, the mean expression within the interquartile range (25th–75th percentile) was used to calculate fold changes between knockout and control groups, minimizing the impact of outliers. Statistical significance was assessed using a two-sided Wilcoxon rank-sum test. Because the gRNA targeting PTPRS exhibited suboptimal knockout efficiency, we incorporated an additional filtering step prior to the standard analysis pipeline. Specifically, Mixscape was applied (default parameters, except the number of differentially expressed genes was set to 2) to distinguish gRNA-positive cells with genuine transcriptional perturbation from those without. By removing cells lacking evidence of effective gene silencing, we improved the resolution of downstream analyses, which ultimately allowed the identification of aptamer Family 4 as specifically targeting PTPRS.

Selection of aptamer sequence subsets for downstream analysis

To evaluate how many of the most abundant aptamer sequences should be selected for downstream analysis, we systematically tested subsets of varying sizes from our SPARK-seq dataset. Specifically, we analyzed the top 2000, 5000, 8000, 10,000, 20,000, 50,000, and 100,000 distinct sequences, ranked by read abundance. The top 10,000 sequences alone captured 68.84% of total reads and yielded the maximal number of validated protein targets. Increasing the number of included sequences to 50,000 or 100,000 only slightly improved cumulative read coverage (to 69.5% and 69.8%, respectively), but resulted in disproportionately high computational costs—including 9- to 34-fold longer runtimes and 4- to 6-fold greater memory usage—without identifying any additional targets. Based on these results, we selected the top 10,000 most abundant sequences as the optimal balance between coverage, efficiency, and biological relevance for subsequent analyses.

Aptamer classification/prediction module

To identify binding aptamers for specific targets, we developed a deep neural network-based classification module capable of distinguishing aptamers that bind to the target from those that do not. The neural network architecture comprises two main submodules: a feature extraction submodule and a linear classification submodule. The feature extraction submodule utilizes a series of convolutional and pooling layers to capture hierarchical features from aptamer sequence data (*37*). Briefly, one-hot encoded aptamer sequences are processed through two sequential 1-dimensional convolutional layers, each followed by a ReLU activation layer and a pooling layer, to extract multilevel features effectively. These extracted features are then upsampled and integrated before being passed to the linear classification submodule. The classification submodule performs the final task, accurately identifying aptamers that bind to the target.

Aptamer variant generation module

To facilitate aptamer generation, SPARTA incorporates Raptgen, a framework designed to model sequence motifs. Raptgen achieves this using a decoder that combines Hidden Markov Models (HMM) and Variational Autoencoders (VAE), effectively capturing motif patterns through a profile HMM-based architecture (*38*). The generation process comprises two key stages: the training stage and the generation stage. During the training stage, aptamer sequences are input into the VAE, and the encoder embeds them into a latent space representation. The decoder then reconstructs the embeddings back into aptamer sequences, ensuring that the latent space captures the essential features of the input data. In the generation stage, the trained encoder is utilized to derive latent embeddings for specific target-binding aptamers. New samples are generated by performing sampling around these embeddings in the latent space. The sampled points are then passed through the trained decoder, producing new aptamer sequences with potential binding affinity to the target.

In this study, we embedded sequences of the PTK7 aptamer family, along with some nonbinding sequences, into a latent space. Clustering within this latent space yielded 28 clusters, 15 of which were primarily composed of sequences from the PTK7 aptamer family. We selected the centers of these 15 clusters to generate 15 nucleic acid sequences as candidate PTK7 aptamer sequences. Additionally, we randomly selected points outside these 15 clusters to generate noncandidate sequences for subsequent experimental studies.

Data visualization and statistical tests

To effectively select proteins for knockout, we calculated the growth ratios of rounds R1–R4 relative to R0 based on proteomics data and identified membrane proteins with increasing trends across these rounds. Data visualization was performed using ComplexHeatmap (v2.16.0). The differential aptamer abundance induced by each gRNA was visualized with ggplot2 (v3.5.1), and t tests were used to assess differences between gRNAs targeting proteins and others. Differential aptamer expression analysis was conducted using the limma package (version 3.54.2). For both the target knockout and NC groups, we selected 150 cells with binding signals around the median. These cells were subjected to differential expression analysis using limma, where the $\log_2$ fold change ($\log_2$FC) for each aptamer was calculated from the fitted linear model, and statistical significance was assessed via the empirical Bayes moderated t-test. The edit distance between sequences was calculated using the Python library editdistance (v0.8.1).

REFERENCES AND NOTES

1. D. L. Robertson, G. F. Joyce, Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467–468 (1990). doi: 10.1038/344467a0; pmid: 1690861
2. C. Tuerk, L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990). doi: 10.1126/science.2200121; pmid: 2200121
3. M. C. DeRosa *et al.*, In vitro selection of aptamers and their applications. *Nat. Rev. Methods Primers* **3**, 54 (2023). doi: 10.1038/s43586-023-00238-7; pmid: 37969927
4. D. Chang *et al.*, A high-dimensional microfluidic approach for selection of aptamers with programmable binding affinities. *Nat. Chem.* **15**, 773–780 (2023). doi: 10.1038/s41557-023-01207-z; pmid: 37277648
5. N. K. Singh *et al.*, High-affinity one-step aptamer selection using a nonfouling porous hydrogel. *Nat. Biotechnol.* **42**, 1224–1231 (2024). doi: 10.1038/s41587-023-01973-8; pmid: 37798416
6. K. Yang *et al.*, A functional group-guided approach to aptamers for small molecules. *Science* **380**, 942–948 (2023). doi: 10.1126/science.abn9859; pmid: 37262137
7. C. Ye *et al.*, A wearable aptamer nanobiosensor for noninvasive female hormone monitoring. *Nat. Nanotechnol.* **19**, 330–337 (2024). doi: 10.1038/s41565-023-01513-0; pmid: 37770648
8. A. Mullard, FDA approves second RNA aptamer. *Nat. Rev. Drug Discov.* **22**, 774 (2023). doi: 10.1038/d41573-023-00148-z; pmid: 37673979
9. M. Sunbul *et al.*, Super-resolution RNA imaging using a rhodamine-binding aptamer with fast exchange kinetics. *Nat. Biotechnol.* **39**, 686–690 (2021). doi: 10.1038/s41587-020-00794-3; pmid: 33574610
10. J. P. Overington, B. Al-Lazikani, A. L. Hopkins, How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006). doi: 10.1038/nrd2199; pmid: 17139284
11. C. Y. Lu, V. Terry, D. M. Thomas, Precision medicine: Affording the successes of science. *npj Precis. Onc.* **7**, 3 (2023). doi: 10.1038/s41698-022-00343-y
12. F. Katzen, T. C. Peterson, W. Kudlicki, Membrane protein expression: No cells required. *Trends Biotechnol.* **27**, 455–460 (2009). doi: 10.1016/j.tibtech.2009.05.005; pmid: 19616329
13. D. Shangguan *et al.*, Aptamers evolved from live cells as effective molecular probes for cancer study. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11838–11843 (2006). doi: 10.1073/pnas.0602615103; pmid: 16873550
14. L. Zhang *et al.*, Molecular Elucidation of Disease Biomarkers at the Interface of Chemistry and Biology. *J. Am. Chem. Soc.* **139**, 2532–2540 (2017). doi: 10.1021/jacs.6b10646; pmid: 28121431

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026

15. K. N. Morris, K. B. Jensen, C. M. Julin, M. Weil, L. Gold, High affinity ligands from in vitro selection: Complex targets. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2902–2907 (1998). doi: 10.1073/pnas.95.6.2902; pmid: 9501188
16. M. Blank, T. Weinschenk, M. Priemer, H. Schluesener, Systematic Evolution of a DNA Aptamer Binding to Rat Brain Tumor Microvessels: SELECTIVE TARGETING OF ENDOTHELIAL REGULATORY PROTEIN PIGPEN. *JBC*, **276**, 16464–16468 (2001). doi: 10.1038/nprot.2010.66; pmid: 20539292
17. K. Sefah, D. Shangguan, X. Xiong, M. B. O'Donoghue, W. Tan, Development of DNA aptamers using Cell-SELEX. *Nat. Protoc.* **5**, 1169–1185 (2010). doi: 10.1038/nprot.2010.66; pmid: 20539292
18. M. McKeague *et al.*, Analysis of In Vitro Aptamer Selection Parameters. *J. Mol. Evol.* **81**, 150–161 (2015). doi: 10.1007/s00239-015-9708-6; pmid: 26530075
19. N. Rifai, M. A. Gillette, S. A. Carr, Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nat. Biotechnol.* **24**, 971–983 (2006). doi: 10.1038/nbt1235; pmid: 16900146
20. X. Wu *et al.*, Elucidation and Structural Modeling of CD71 as a Molecular Target for Cell-Specific Aptamer Binding. *J. Am. Chem. Soc.* **141**, 10760–10769 (2019). doi: 10.1021/jacs.9b03720; pmid: 31185171
21. L. Kelly, K. E. Maier, A. Yan, M. Levy, A comparative analysis of cell surface targeting aptamers. *Nat. Commun.* **12**, 6275 (2021). doi: 10.1038/s41467-021-26463-w; pmid: 34725326
22. X. Wu *et al.*, Efficient Strategy to Discover DNA Aptamers Against Low Abundance Cell Surface Proteins in Scarce Samples. *JACS*, **146**, jacs.4c03129 (2024). doi: 10.1038/s41592-019-0392-0; pmid: 31011186
23. E. P. Mimitou *et al.*, Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019). doi: 10.1038/s41592-019-0392-0; pmid: 31011186
24. E. Papalexi *et al.*, Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat. Genet.* **53**, 322–331 (2021). doi: 10.1038/s41588-021-00778-2; pmid: 33649593
25. P. Zhou *et al.*, Single-cell CRISPR screens in vivo map T cell fate regulomes in cancer. *Nature* **624**, 154–163 (2023). doi: 10.1038/s41586-023-06733-x; pmid: 37968405
26. J. M. Replogle *et al.*, Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020). doi: 10.1038/s41587-020-0470-y; pmid: 32231336
27. D. Shangguan *et al.*, Cell-specific aptamer probes for membrane protein elucidation in cancer cells. *J. Proteome Res.* **7**, 2133–2139 (2008). doi: 10.1021/pr700894d; pmid: 18363322
28. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013). doi: 10.1038/nprot.2013.143; pmid: 24157548
29. J. G. Doench *et al.*, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016). doi: 10.1038/nbt.3437; pmid: 26780180
30. S. R. Srivatsan *et al.*, Massively multiplex chemical transcriptomics at single-cell resolution. *Science* **367**, 45–51 (2020). doi: 10.1126/science.aax6234; pmid: 31806696
31. H. Yu, O. Alkhamis, J. Canoura, Y. Liu, Y. Xiao, Advances and Challenges in Small Molecule DNA Aptamer Isolation, Characterization, and Sensor Development. *Angew. Chem. Int. Ed.* **2021**, 16800–16823 (2021). doi: 10.1002/anie.202008663; pmid: 33559947
32. J. Lv *et al.*, Generation of an Aptamer Targeting Receptor-Type Tyrosine-Protein Phosphatase F. *Anal. Chem.* **95**, 1228–1233 (2023). doi: 10.1021/acs.analchem.2c03988; pmid: 36594741
33. J. Song *et al.*, A Sequential Multidimensional Analysis Algorithm for Aptamer Identification based on Structure Analysis and Machine Learning. *Anal. Chem.* **92**, 3307–3314 (2020). doi: 10.1021/acs.analchem.9b05203; pmid: 31876151
34. T. Li *et al.*, Blocker-SELEX: A structure-guided strategy for developing inhibitory aptamers disrupting undruggable transcription factor interactions. *Nat. Commun.* **15**, 6751 (2024). doi: 10.1038/s41467-024-51197-w; pmid: 39117705
35. V. Morath *et al.*, PET-based tracking of CAR T cells and viral gene transfer using a cell surface reporter that binds to lanthanide complexes. *Nat. Biomed. Eng.* **9**, 1886–1906 (2025). doi: 10.1038/s41551-025-01415-7; pmid: 40514433
36. M. H. Uddin *et al.*, Proteomic Analysis Identifies p62/SQSTM1 as a Critical Player in PARP Inhibitor Resistance. *Front. Oncol.* **12**, 908603 (2022). doi: 10.3389/fonc.2022.908603; pmid: 35847859
37. Q. Zhang *et al.*, Locating transcription factor binding sites by fully convolutional neural network. *Brief. Bioinform.* **22**, bbaa435 (2021). doi: 10.1093/bib/bbaa435; pmid: 33498086
38. N. Iwano, T. Adachi, K. Aoki, Y. Nakamura, M. Hamada, Generative aptamer discovery using RaptGen. *Nat. Comput. Sci.* **2**, 378–386 (2022). doi: 10.1038/s43588-022-00249-6; pmid: 38177576
39. S. Kraemer *et al.*, From SOMAmer-Based Biomarker Discovery to Diagnostic and Clinical Applications: A SOMAmer-Based, Streamlined Multiplex Proteomic Assay. *PLOS ONE* **6**, e26332 (2011). doi: 10.1371/journal.pone.0026332; pmid: 22022604
40. A. C. Pan, D. W. Borhani, R. O. Dror, D. E. Shaw, Molecular determinants of drug-receptor binding kinetics. *Drug Discov. Today* **18**, 667–673 (2013). doi: 10.1016/j.drudis.2013.02.007; pmid: 23454741
41. R. A. Copeland, The drug–target residence time model: A 10-year retrospective. *ACS Energy Lett.* **15**, 87–95 (2016). doi: 10.1038/nrd.2015.18; pmid: 26678621
42. T. Ji *et al.*, Delivery of local anaesthetics by a self-assembled supramolecular system mimicking their interactions with a sodium channel. *Nat. Biomed. Eng.* **5**, 1099–1109 (2021). doi: 10.1038/s41551-021-00793-y
43. O. Alkhamis, J. Canoura, L. Wang, Y. Xiao, Nuclease-assisted selection of slow-off rate aptamers. *Sci. Adv.* **10**, eadl3426 (2024). doi: 10.1126/sciadv.adl3426; pmid: 38865469
44. F. Wong *et al.*, Deep generative design of RNA aptamers using structural predictions. *Nat. Comput. Sci.* **4**, 829–839 (2024). doi: 10.1038/s43588-024-00720-6; pmid: 39506080
45. B. Adamson *et al.*, A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867–1882.e21 (2016). doi: 10.1016/j.cell.2016.11.048; pmid: 27984733
46. A. Dixit *et al.*, Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866.e17 (2016). doi: 10.1016/j.cell.2016.11.038; pmid: 27984732
47. X. Wang *et al.*, Molecularly engineered aptamers targeting tumor tissue and cancer cells for efficient in vivo recognition and enrichment. *CCS Chem.* **4**, 2581–2587 (2022). doi: 10.31635/ccschem.021.202101337
48. Q. Su-Tobon *et al.*, CRISPR-Hybrid: A CRISPR-Mediated Intracellular Directed Evolution Platform for RNA Aptamers. *Nat. Commun.* **16**, 595 (2025). doi: 10.1038/s41467-025-55957-0; pmid: 39799111
49. Y. Zhang *et al.*, Nucleic acid aptamer controls mycoplasma infection for inhibiting the malignancy of esophageal squamous cell carcinoma. *Mol. Ther.* **30**, 2224–2241 (2022). doi: 10.1016/j.ymthe.2022.02.018; pmid: 35189346
50. G. Luo *et al.*, SPARK-seq: a high-throughput platform for aptamer discovery and kinetic profiling, v1.0.0, Zenodo (2025); doi: 10.5281/zenodo.17361123

ACKNOWLEDGMENTS

We also thank X. Li, Y. Du, J. He, and G. Hu from HIM for their technical guidance. We gratefully acknowledge the support provided by the Scientific Experiment Center, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences. **Funding:** This work was funded by the following: Strategic Priority Research Program of the Chinese Academy of Sciences (XDB1020000); National Key R&D Program of China (2023YFF1205905, 2019YFA0905800, 2024YFC3405600, 2022YFA1104200); National Natural Science Foundation of China (T2188102, YQ2022HW01, 22577133, 82372741 and 82103287, 82503388, 22304177, 22474075); Innovative research team of high-level local universities in Shanghai (SHSMU-ZLCX20212601); Zhejiang Provincial Natural Science Foundation of China (LR22B050001, Y21C050001, 2023SDYXS0003, YXD24B0701); Hangzhou Institute of Medicine Chinese Academy of Sciences (2024ZZBS01; 2024ZZBS06); Research Funds of Hangzhou Institute for Advanced Study, UCAS (2024HIAS-V008). **Author contributions:** Conceptualization: W.T., Q.W., and G.L. Methodology: G.L., J.S., Y.F., and Z.Z. Investigation: G.L., Y.J., and L.L. Data curation: G.L., Y.F., J.S., L.L., and Y.J. Bioinformatics analysis: J.S., Y.F., Z.Z., G.L., L.G., and Y.G. Writing – original draft: Q.W., G.L., H.J., and J.S. Writing – review & editing: W.T., Q.W., J.S., Y.W., and G.L. Visualization: G.L., Y.G., Y.F., Y.W., and Z.Z. Supervision: W.T., Q.W., and J.S. Project administration: W.T., Q.W., and T.F. Funding acquisition: W.T. and Q.W. All authors have read and approved the final manuscript. **Competing interests:** W.T., Q.W., G.L., J.S., Y.F., L.L., and Z.G. have (and expect) patents and patent applications regarding aptamers and the methods presented in this work and their uses (patent application CN2024119392732). **Data, code, and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. High-throughput sequencing data have been deposited in the NCBI Gene Expression Omnibus under accession number: PRJNA1202227. Source code generated and used in this manuscript is publicly available at GitHub (https://github.com/fuyboo/sparta) and has been archived at Zenodo (*50*). Plasmids generated in this study will be available on Addgene (ID: 249196-249245); all other materials are described in the methods and/or will be available from the authors upon reasonable request. **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/content/page/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adv6127
Figs. S1 to S18; Data S1 to S9; MDAR Reproducibility Checklist

Submitted 29 December 2024; resubmitted 24 August 2025; accepted 31 October 2025
10.1126/science.adv6127

Downloaded from https://www.science.org at Shanghai Normal University on January 02, 2026