EXPERIMENTAL SECTION

Generation of a Stable GPC3-expressing 1MEA Cell Line. Full-length human GPC3 (hGPC3) was digested from cDNA clone (Open Biosystems, AL) using EcoRI and NotI REs and inserted into pcDNA3.1/His C vector (Thermo Fisher Scientific, CA). Recombinant vector was amplified in E.coli DH5α and transfected into 1MEA cells by Lipofectamine 2000 (ThermoFisher, CA) following the manual instructions. Briefly, recombinant vector (1 μg) was transfected into each of a 6-well plate containing 100,000 1MEA cells (80% confluent) with 5 μL of lipofectamine 2000 as the delivery reagent. Transfected 1MEA expressing hGPC3 ($1MEA^{hGPC3}$) were selected using specific selection marker (500 μg/mL G418, Geneticin®, Thermo Fisher Scientific, CA) for 3 weeks to obtain survival clones. The stable transfected cell line was further analyzed for expression level of hGPC3 by RT-PCR, WB or Immunocytochemistry (ICC) assays using the specific antibody for GPC3 (1G12, Santa Cruz Biotechnology, TX).

Chemical synthesis and purification of GACTZP libraries to support AEGIS-LIVE. All d**Z** and d**P** containing oligonucleotides (Table S1) were synthesized with a DNA synthesizer (ABI 394) using standard phosphoramidite chemistry on glass support (CPG). Protected d**Z** and d**P** phosphoramidites were from Firebird Biomolecular Sciences LLC (Alachua FL, www.firebirdbio.com). Standard phosphoramidites (Bz-dA, Ac-dC, dmf-dG, and dT) were purchased from Glen Research (Sterling VA). The DNA library oligonucleotides were designed to have forward and reverse primer binding segments (each 18 nucleotides in length) with a random region (35 nts) containing GACT**ZP** (six nucleotides) in a ratio of 1:1:1:1:2:2. Coupling times were 120 seconds.

To remove protecting groups, CPG-bound DMT-off DNA molecules were incubated with acetonitrile-triethylamine (1:1 v/v, 1.5 mL) for 1 hour at room temperature. Following removal of supernatant, the CPG-bound oligonucleotides were treated with another 1.5 mL of triethylamine-acetonitrile (1:1 v/v) for overnight at room temperature. After removal of supernatant, the CPG-bound oligonucleotides were incubated with 1.0 mL of 1M DBU in anhydrous acetonitrile solution at room temperature for 18 hours to remove the protecting groups on d**Z**. After removal of acetonitrile, d**Z** and d**P** containing oligonucleotides were retreated with ammonium hydroxide (55 °C, overnight). The product mixture was resolved by

denaturing PAGE (7 M urea), and extracted with TEAA buffer (0.2 M, pH=7.0). The product was then desalted by Sep-Pac® Plus C18 cartridges (Waters). All 5'-biotinylated d**Z** and d**P** containing potential aptamers were synthesized, deprotected, and purified in house based on the above methods. All standard 5'-biotinylated oligonucleotides were purchased from IDT and purified by HPLC.

Cell lines. Liver cancer cell lines except for Huh7, LH86 and Hu1545 were all purchased from the American Tissue Culture collection (ATCC). Huh7 was purchased from JCRB Cell Bank; LH86 was developed by our group from a resected, well-differentiated, hepatocellular carcinoma. Hu1545 was established by immortalizing primary hepatocyte (normal liver cell) with lentivirus carrying hTERT (human telomerase reverse transcriptase; the enzyme maintains telomere length at the end of chromosomes, thus enabling cells to grow and proliferate). Though equipped with extended lifespan, this cell still remains the characteristic, most importantly the protein profile, of normal liver cells compared to liver cancer cells. All cell lines were maintained in high glucose DMEM culture medium (Sigma-Aldrich, St. Louis, MO) supplemented with 10% Fetal Bovine Serum (Gibco®, Life Technologies, Carlsbad, CA) and 1% penicillin-streptomycin (Life Technologies, Carlsbad, CA). Cultures were incubated at 37°C with 5% CO_2. Since this cell line is adherent, 95-100% confluent cell culture dishes were used throughout the entire selection process.

Experimental procedure for AEGIS Cell-LIVE. To begin the AEGIS-LIVE experiment, $1MEA^{hGPC3}$ cells were seeded in 10 cm culture dishes (~10^6 cells). These cells adhered to the bottom of dishes and grew to about 97% coverage. Cells were washed with washing buffer (4.5 g/liter glucose, 5 mM $MgCl_2$ in Dulbecco's PBS). A sample of the DNA library (20 nmol) was dissolved in binding buffer (700 µL, 4.5 g/liter glucose, 5 mM $MgCl_2$, 0.1 mg/mL tRNA and 1 mg/mL BSA, all in Dulbecco's PBS). The GACT**ZP** DNA library was denatured by heating (85 °C, 5 min), and then "snap cooled" on ice for 10 min. The library was then incubated with the cells still adhering to the bottom of the culture dish (4 °C with rocking for 30-60 min). Cells were thrice gently washed with washing buffer to remove unbound sequences. Binding buffer (0.5 mL) was then added and the cells scraped off the plate using cell scraper to recover cell/DNA complexes.

Once the cells were scraped from the bottom of the dish into a suspension in PBS buffer, they were heated (85 °C for 10 min). The resulting mixture

was centrifuged at 14000 rpm to pellet the cell debris. The supernatant containing the ssDNA survivors was further incubated with counter cells (1MEA) attached to the dish bottom at 4 °C (with rocking, 1 h). The survivors were collected in the supernatant, and then amplified by six-nucleotide PCR using fluorescein- and biotin-labeled primers with all six nucleoside triphosphates (d**Z**TP, d**P**TP, dGTP, dATP, dCTP, and dTTP).

Different PCR cycles (from 8 cycles to 25 cycles) were tested to determine the optimum number of cycles for preparative PCR to produce maximal amount of amplicon with the least **Z**/P loss. Typical six-nucleotide PCR reagents and conditions are listed in Table S2. Upon the completion of six-nucleotide PCR, the FITC-labeled DNA strands were separated from the biotinylated strands by affinity purification with streptavidin-coated Sepharose beads (GE Healthcare Bio-Sciences Corp., Piscataway), followed by alkaline denaturation (with NaOH, 100 mM). The surviving ssDNA was desalted and resuspended in binding buffer to a final concentration of 0.5 μM. The survivors were denatured at 85 °C, snap cooled and used to perform the second round of selection using the same procedure as described for the first round of selection. The negative selections were added from Round 5. The entire selection process was repeated until a sustained significant enrichment was obtained at the 11th round. During the selection, the stringency of the selection was increased by decreasing the number cells and the incubation times (Table S3).

Deep sequencing of GACTZP DNA survivors using Next Generation sequencing technology. Solutions containing enriched GACT**ZP** DNA survivors after the 13th round of AEGIS Cell-LIVE were divided into two equal parts. These were separately converted into standard DNA under two conversion conditions using primers that carried barcodes for the Ion Torrent deep sequencing (Table S4). Following conversion, the samples were combined, purified by native agarose gel, and submitted for Ion Torrent "next generation" sequencing at the University of Florida, ICBR sequencing core facility. The products were aligned to identify sequences derived from a single common aptamer "ancestor", and the ancestral sequence was inferred (see below).

Inference of GACTZP aptamer sequences. Ion Torrent sequencing reads that did not contain exact matches to the barcode, forward and reverse priming sequences were discarded. To minimize miscalling, any read present in fewer than 20 copies was removed from the analysis. The

remaining reads (126,441 out of 887,092) were then clustered using software custom designed at the FfAME, which ignored differing barcodes during the clustering and accepted single-step changes within sequence reads. Clustered sequences were then separated by barcode, with variable sites being compared between each barcode (differentiating the two conversion conditions). The clustered sequences obtained under the first conversion conditions (Barcode 2, **Z** to C and P to G conversion) serve as reference for the clustered sequences obtained under the second conversion conditions (Barcode 4, **Z** to T/A and **P** to C/G conversion). Sites where C and T were found in approximately equal amounts after conversion under the second conditions were assigned as **Z** in their "parent". Sites where G and A were found in approximately equal amounts after conversion under the second conditions were assigned as P in their "parent".

Screening of potential aptamer candidates. Analysis of the Ion Torrent sequencing output identified with decreasing abundance several different aptamer species holding **Z** and/or P, as well as species containing only standard bases. Each sequence was chemically synthesized, labeled with biotin at the 5' end, and then purified by PAGE. These were quantified (UV 260/280) and diluted to standard concentrations. Flow cytometry binding assays were then done using the target $1MEA^{hGPC3}$ cells or other types of cells. To obtain suspended cells for flow cytometry, culture medium was removed from the cells and non-enzymatic dissociation buffer was added to cover the surface of the entire flask. This was placed in an incubator at 37 °C. After incubation (5 min), the cells were aspirated using a transfer pipette to remove them from the flask. These were washed twice by centrifugation and approximately 5.0 $x10^5$ cells were incubated separately with the aptamer candidates at a final concentration of 250 nM. After incubation, cells were washed. Streptavidin-PE conjugate (100 µL of 1:400 dilution, optimized) was then added, and the mixture was incubated at 4 °C for 10 min. Excess dye conjugates were removed by washing twice and the cell-DNA complexes were resuspended in 150 µL binding buffer. The aptamer binding signal was detected using flow cytometry (BD Accuri™ C6). Unselected library was used as a control to set the fluorescence background.

Determination of binding affinity. The binding affinity of the most abundant aptamers was measured by flow cytometry using biotin-labeled

aptamer, and the signal was detected with streptavidin-PE conjugate. HepG2 cells were dissociated using non-enzymatic dissociation buffer. Cells were washed and incubated with varying concentrations (0.1 nM - 1000 nM final concentration) of biotin- labeled aptamer in a 200 mL volume of binding buffer. After 20 min of incubation, cells were washed twice with washing buffer and then incubated with conjugate dye (100 mL, 1:400 dilution). This was incubated (10 min) and then washed twice (1300 µL each) with washing buffer. The cell pellets were resuspended in washing buffer (200 µL) and analyzed by flow cytometry.

The biotin-labeled unselected library was used as a negative control to determine the background binding. All binding assays were done in triplet. The mean fluorescence intensity of the unselected library was subtracted from that of the corresponding aptamer with the target cells to determine the specific binding of the labeled aptamer.

Competition assay. Ten times of concentration (2.5 uM) of biotin labeled DNA aptamers were first incubated with HepG2 cells at 4 ˚C for 30 min, followed by incubation with 250 nM of anti-hGPC3 antibody (R&D systems, MN). Anti-hGPC3 antibodies were stained by Alexa 488-labelled secondary antibody. Aptamers were stained by streptavidin-Alexa 633. Binding was measured using flow cytometry.

Confirmation of membrane protein target of selected aptamers. HepG2 cells (4×10^8 cells) were disassociated using non-enzymatic dissociation buffer, followed by washing three times at 4 °C with a PBS buffer. Collected cells were lysed in 5 mL of hypotonic buffer [50 mM Tris-HCl (pH 7.5) containing protease inhibitors (0.1 mM PMSF and 2 µg/mL pepstatin, leupeptin, and aprotinin)] at 4 °C for 30 min. After centrifugation, the debris was washed three times with 5 mL of hypotonic buffer and dissolved in 1.5 mL of lysis buffer (PBS containing 5 mM $MgCl_2$ and 1% Triton X-100) at 4 °C for 30 min. After centrifugation, the supernatant was incubated with 0.1 mg/mL salmon sperm DNA as nonspecific competitor. Biotin labeled aptamers (150 pmol) were added and incubated at 4 °C for 30 min. The protein-DNA complex was captured by incubating it with 2 mg (200 µL) of magnetic streptavidin beads (Thermo Fisher Scientific, CA) at 4 °C for 15 min and collected on a magnet stand. Collected magnetic beads were washed four times with 1 mL of PBS containing 5 mM $MgCl_2$. The proteins were eluted by heating in 30 µL of loading buffer and analyzed by

polyacrylamide gel electrophoresis (12%) (SDS-PAGE). Western blot was then performed using anti-hGPC3 antibody (R&D systems, MN).

Supplementary Figures

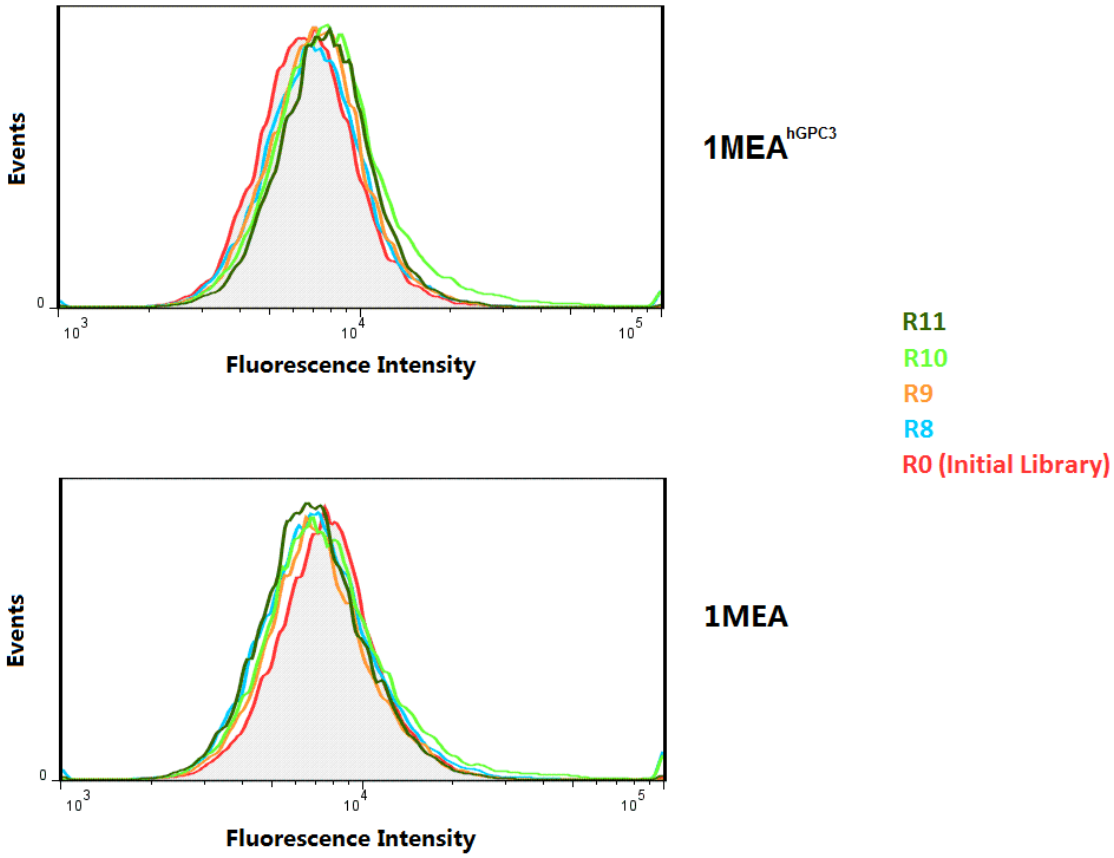


Figure S1. Monitoring the progress of Cell-SELEX using flow cytometry.
The fluorescence were from FITC labeled on each DNA sequences

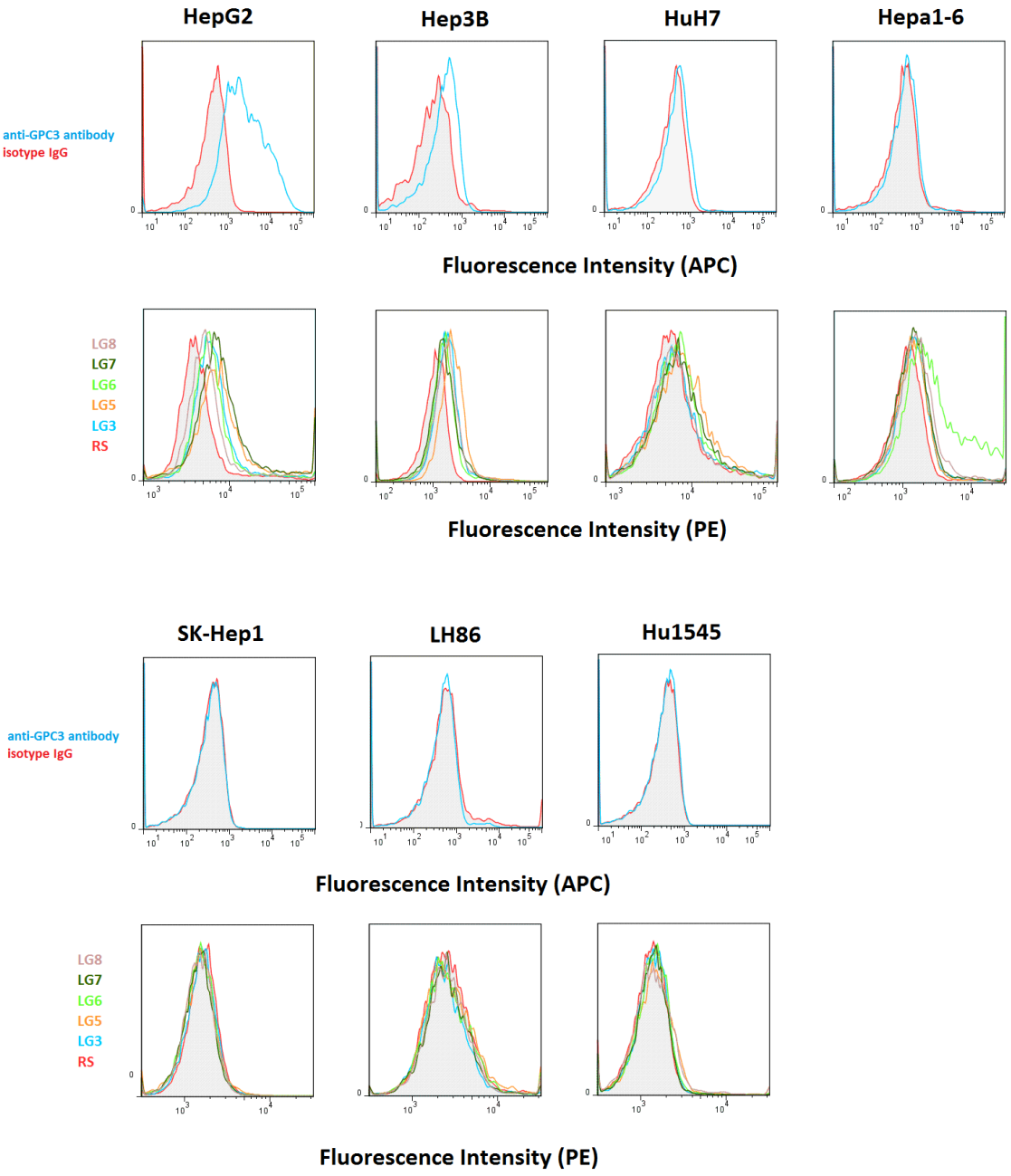


Figure S2. Flow cytometry test to confirm the expression level on seven cell lines. Antibody labeled with Allophycocyanin (APC) fluorophore. Aptamers labeled with phycoerythrin (PE).

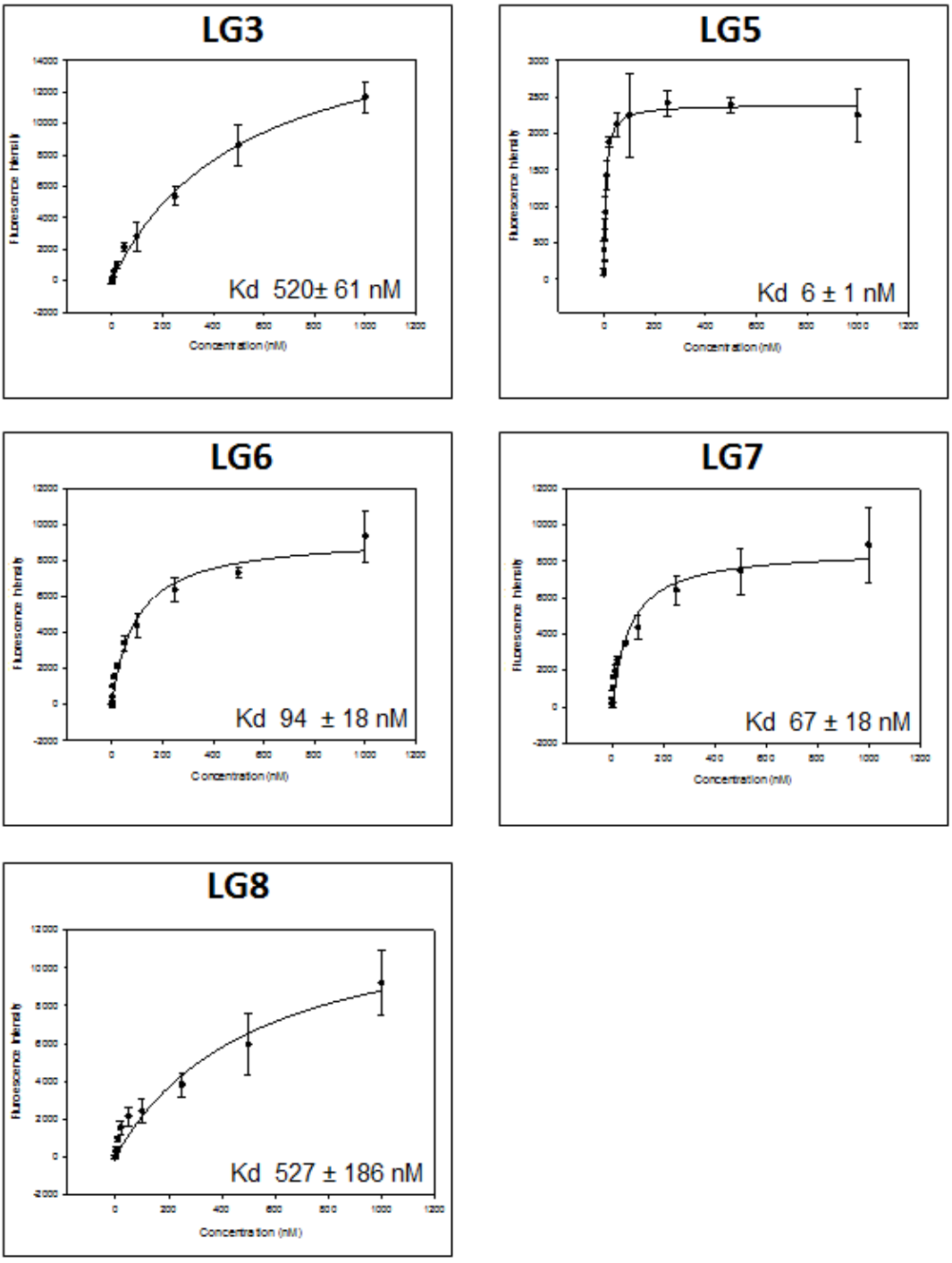


Figure S3. Binding curves to estimate the disassociate constants of aptamer candidates.

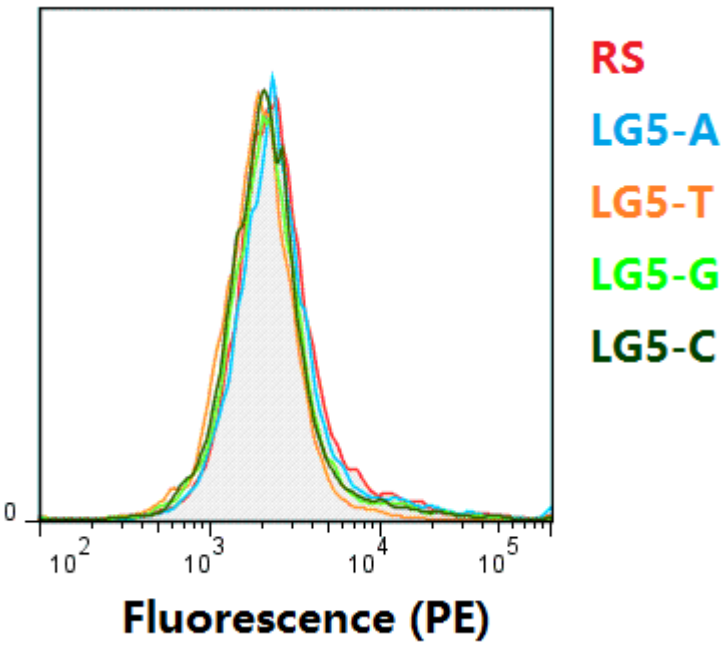


Figure S4. Binding of analogues of aptamer LG5 with Z replaced by standard nucleotides (A, T, G, C). The indicated aptamer analogues (250 nM) were incubated with target cell (4 °C for 30 min) and then analyzed using flow cytometry as described for Figure 3.

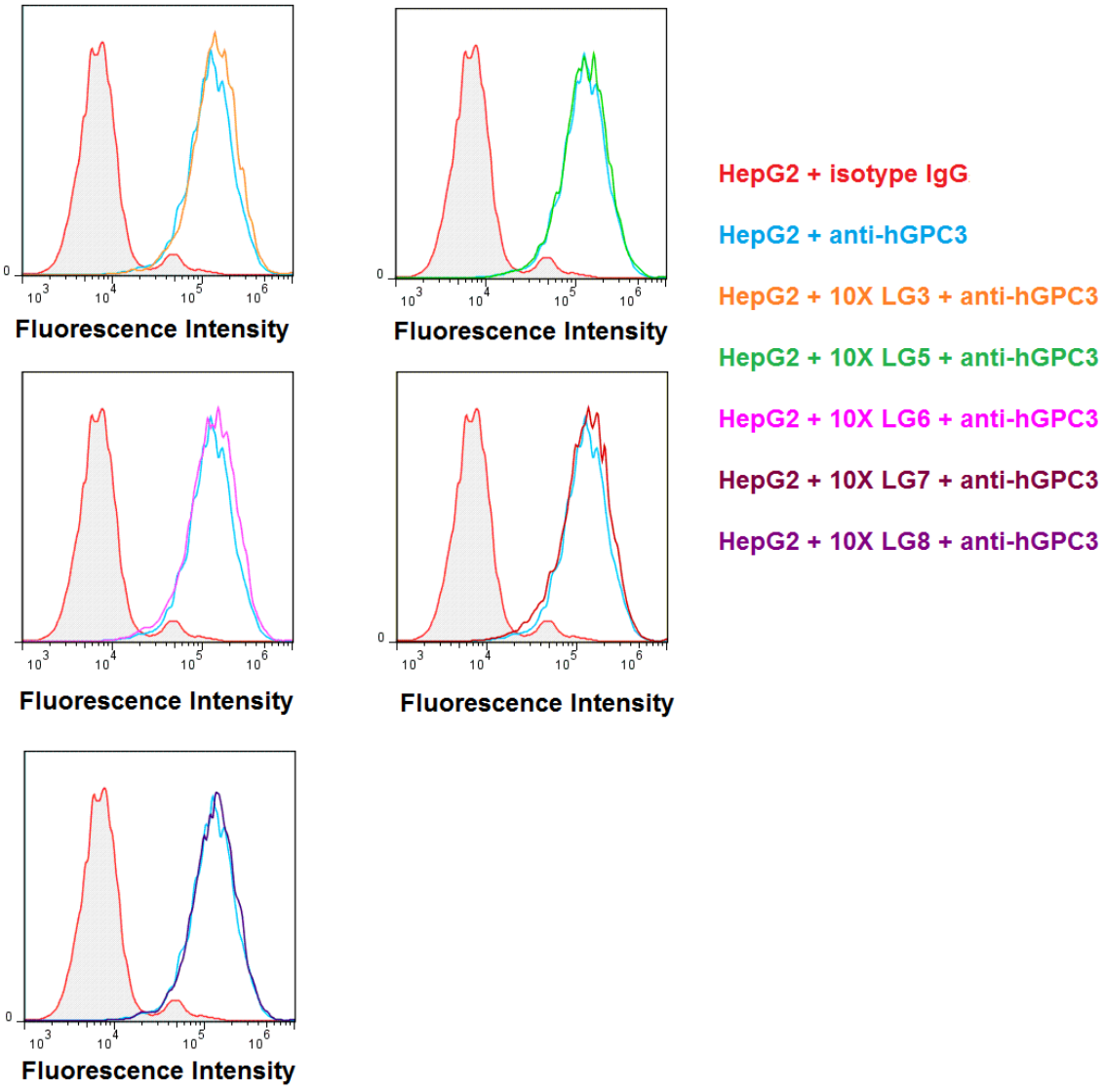


Figure S5. Competition test between anti-hGPC3 antibodies and aptamer candidates. Anti-hGPC3 antibodies were stained by Alexa 488 labelled secondary antibody. Aptamers were stained by streptavidin-Alexa 633. Isotype is the control antibody that does not bind to GPC3. Upper left hand panel is the antibody telling us that the HepG2 GPC3 positive cells are expressing GPC3.

Colorized Versions of Figures in Main Text

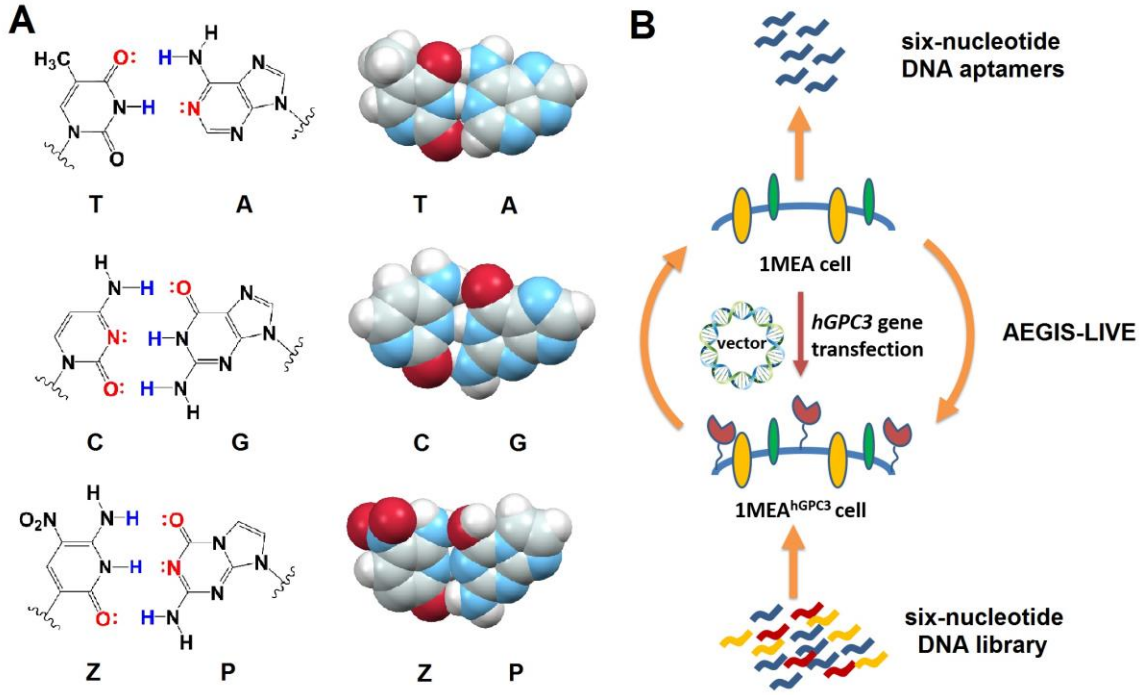


Figure S6. **(Figure 1 in main text) Chemical structures of six nucleotides and AEGIS-LIVE.** (A) Molecular structures (left) and space-filling models (right) of C:G, T:A and Z:P pairs showing their similarity (PDB ID: 4RHD). (B) Schematic of engineering hPGC3-overexpressing cells and AEGIS-LIVE procedure.

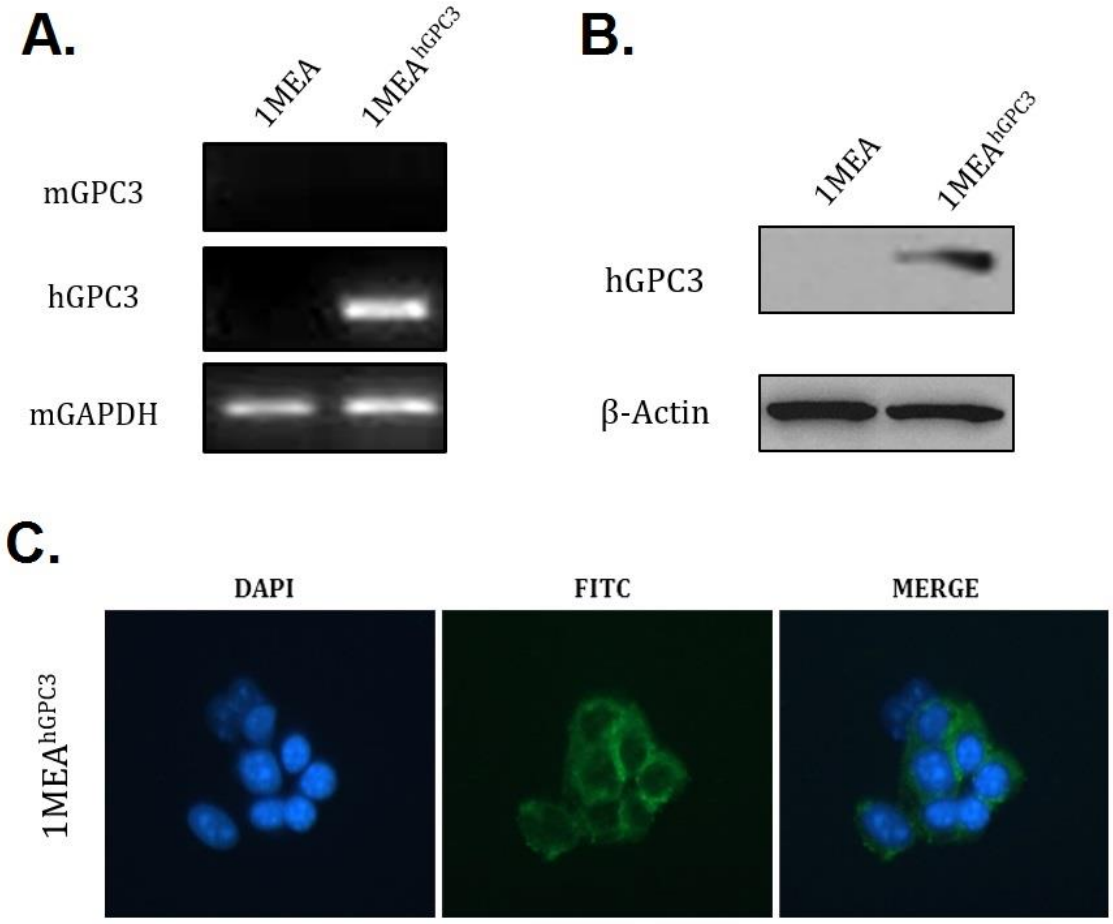
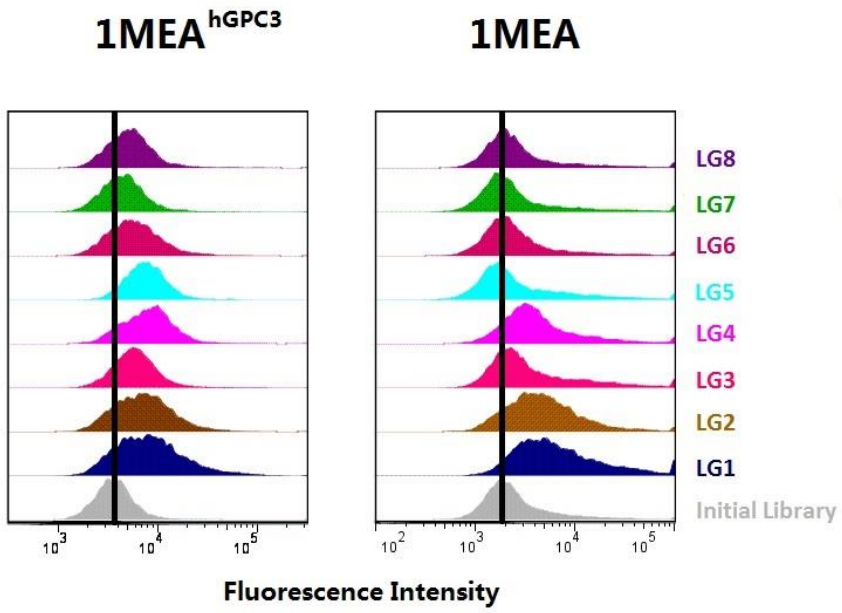


Figure S7. (Figure 2 in main text) Confirmation of human GPC3 expression on 1MEA cell surface in the engineered cells. (A) RT-PCR assessing hGPC3 gene transcription level after transfection. Mouse mGAPDH gene transcription was used as a positive control; mGPC3 was used as negative control. (B) Western blot showing expression of hGPC3 after transfection. Beta-actin was used as a loading control. (C) ICC assays using the 1G12 antibody specific for GPC3, stained by fluorescein-labeled secondary antibody. DAPI (blue) locates the nucleus; FITC (green) identifies antibody binding.

A.



B.

Name	Sequence	Percentage
LG1	~PGGTGGGCGGAGGTCTZGCTACAPGPTTTGGPGGC~	11.37%
LG2	~PGCCCGGGPTAPPGTGPTGGGTGTTCGCTATCCAG~	7.98%
LG3	~GGTAACTAGTAGTTGACCCTGPAGTGZTGTPTCTG~	6.01%
LG4	~GGCGGGGTZGPGTAAGGGGTCTAAGGCATTGGGTC~	4.48%
LG5	~GGAGGAAGTGGTCCTTGCTTTGCZTCGTATCTGGG~	2.57%
LG6	~GGTZGATTATTPGGTTCAATAACACPTCCTGGTGG~	1.96%
LG7	~PGCACAGTGTGZZCCATAGGTTGTAATGACPTZTG~	1.04%
LG8	~GGCAGCZCCTGPAGTPGAGTGTPATGGCTTATTCG~	0.91%

Figure S8. (Figure 3 in main text) Names, Sequences, percentage in pool, and specificity of selected aptamers. (A) Binding of DNA aptamer candidates. (Left) Binding to positive 1MEAhGPC3 cells in the selection. (Right) Binding to negative (original) 1MEA cells. Gray distributions were generated with the initial DNA library. (B) Sequences (randomized region) and their percentage contribution to the pool of surviving binders (AEGIS Z and P are shown in red and blue, respectively).

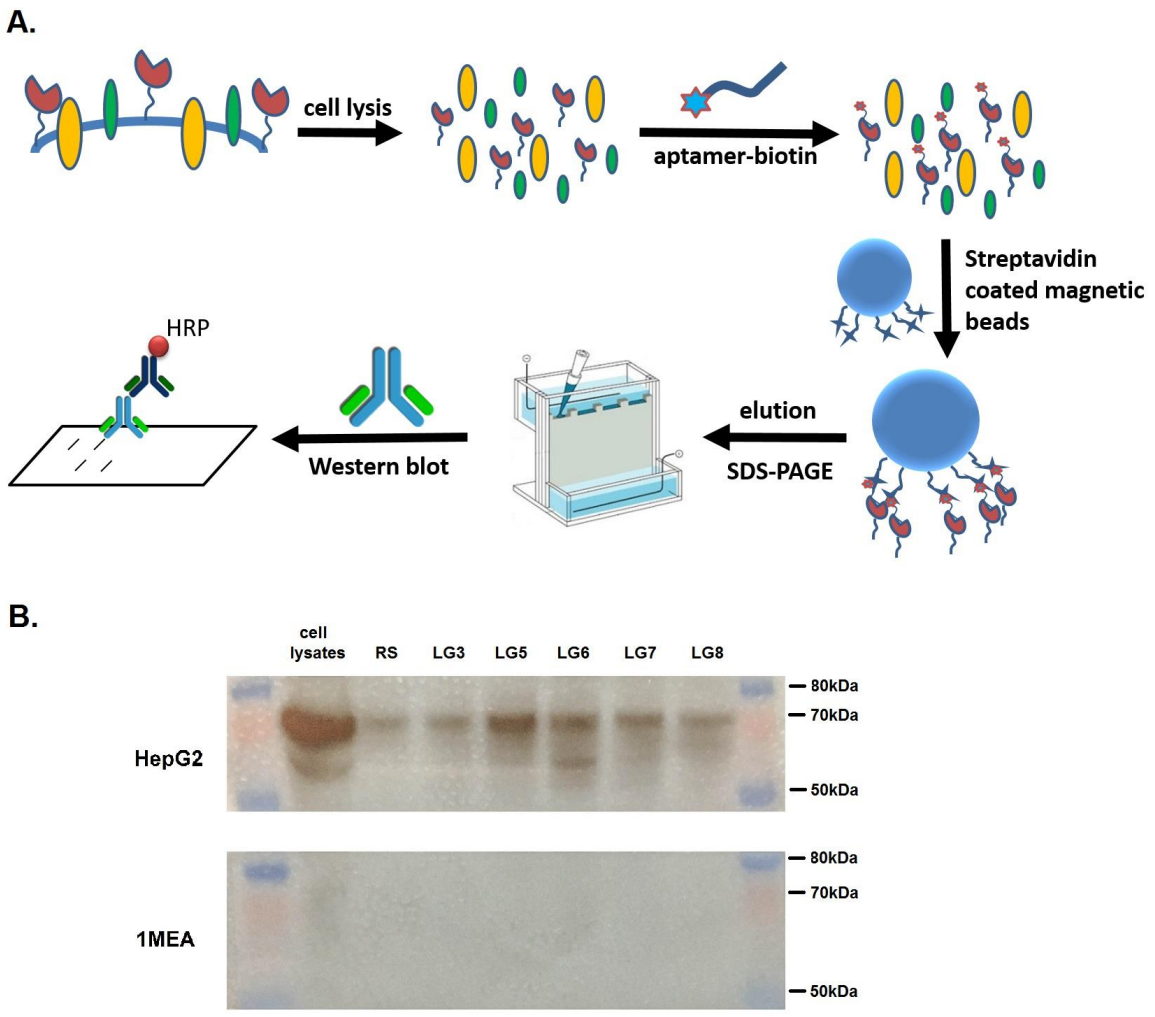


Figure S9. (Figure 4 in main text) Target confirmation of selected aptamers. (A) Schematic of procedure to identify the target of aptamer candidates. (B) Western blot assay of biotin-aptamer-target complex captured by streptavidin modified magnetic beads using anti-hGPC3 primary antibody.

Supplementary Tables

Table S1. Sequences synthesized in SELEX experiment.

Name	Sequence
Initial GACTZP DNA Library	5'-ATA GAC TGG ACT GTC GTC **$(N)_{35}$** TAG CAT CGG ATA CAG GTC-3` ***N*** = A, G, C, T, Z, and P nucleotides mixture with ratio of T:G:A:C:Z:P = 1:1:1:1:2:2
Forward Primer	5'- FITC-ATA GAC TGG ACT GTC GTC-3`
Reverse Primer	5'- Biotin-GAC CTG TAT CCG ATG CTA-3`
GPC3_Code2_F58_2	Adaptor A **Key** **Barcode 2** Forward Primer 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-**GATGATTGCC**- ATAGACTGGACTGTCGTC-3`
GPC3_Code4_F58_4	Adaptor A **Key** **Barcode 4** Forward Primer 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG- **CTTACACCAC** - ATAGACTGGACTGTCGTC-3'
GPC3_trP_R41	Adaptor trP1 Reverse Primer 5'-CCTCTCTATGGGCAGTCGGTGAT-GACCTGTATCCGATGCTA-3`

Table S2. Typical six-nucleotide PCR amplification of GACTZP DNA library

Reagents	**Volume (µL)**	**Final Concentration**
ddH_2O	30.5	
Forward and reverse primers mixture (each 10 µM)	2.5	0.5 µM
Six-Nucleotide Mix of 10x		
dA,T,G/TPs (1 mM of each)		0.1 mM of each
dCTP (2 mM)		0.2 mM
	5.0	
dZTP (1 mM)		0.1 mM
dPTP (6 mM)		0.6 mM
10x ThermoPol Reaction Buffer (pH = 8.0)	5.0	1x
GACTZP DNA library (survivors)	5.0	(10% of reaction volume)
AmpliTaq Gold DNA polymerase (5 uints/µL)	2	0.10 (U/µL)
Total volume (uL)	50.0	

Note: 1 x ThermoPol Reaction Buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.0 at 25 °C); PCR cycling conditions: one cycle of 90 °C for 1.5 min; 8 cycles ~ 25 cycles of (90 °C for 30 s, 55.5 °C for 30 s, 72 °C for 5 min); 72 °C for 10 min; 4 °C for extended times.

Table S3. Selection procedure of GPC3-overexpressed 1MEA cells SELEX.

Round	Cell involved	Culture dish size (diameter)	Incubation time
1	Positive selection only	100 mm	1 h
2	Positive selection only	100 mm	1 h
3	Positive selection only	100 mm	45 min
4	Positive selection only	100 mm	45 min
5	Positive selection and negative selection	100 mm for both	45 min for positive and 1h for negative
6	Positive selection and negative selection	100 mm for both	45 min for positive and 1h for negative
7	Positive selection and negative selection	60 mm for positive and 100 mm for negative	30 min for positive and 1h for negative
8	Positive selection and negative selection	60 mm for positive and 100 mm for negative	30 min for positive and 1h for negative
9	Positive selection and negative selection twice	60 mm for positive and 100 mm for negative	30 min for positive and 1h for negative
10	Positive selection and negative selection twice	60 mm for positive and 100 mm for negative	30 min for positive and 1h for negative
11	Positive selection and negative selection twice	60 mm for positive and 100 mm for negative	30 min for positive and 1h for negative

Table S4. Converting Z:P to C:G (barcode 2) or converting Z:P to T:A and C:G (barcode 4)

Components	Z:P to C:G conversion	Z:P to T:A and C:G conversion	Final Concentration
ddH2O	29 µl	29 µl	Total volume: 50 µl
GPC3_Code2_F58_2 (10 µM)	2 µl		0.4 µM
GPC3_trP_R41 (10 µM)	2 µl		0.4 µM
GPC3_Code4_F58_4 (10 µM)		2 µl	0.4 µM
GPC3_trP_R41 (10 µM)		2 µl	0.4 µM
10th-Round Survivors	5 µl	5 µl	
10x Five-Nucleotide Mix			
dZTP (0.1 mM)	5 µl		0.01 mM
dC,G/TPs (2 mM of each)			0.2 mM of each
dT,A/TPs (0.4 mM)			0.04 mM of each
10x Five-Nucleotide Mix			
dPTP (2 mM)		5 µl	0.2 mM
dC,G/TPs (0.4 mM of each)			0.04 mM of each
dT,A/TPs (2 mM)			0.2 mM of each
10x ThermoPol Buffer (pH 8.8)	5 µl		1x
10x JumpStart Buffer (pH 8.3)		5 µl	1x
JumpStart Taq (2.5 uints/µl, Sigma)	2 µl	2 µl	0.1 (U/µl)

Note: 1) 1 x ThermoPol Reaction Buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.8 at 25 °C). 2) 1 x JumpStart *Taq* Reaction Buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin, pH 8.3 at 25 °C). 3) PCR conditions: one cycle of 94 °C for 1 min; 12 cycles of (94 °C for 20 s, 57 °C for 30 s, 72 °C for 90 s); 72 °C for 10 min; 4 °C for extended times. 4) The desired PCR product is 134bp.

Table S5. 11 most abundant sequences in the library from 10th round of selection

Name	Sequence	Percentage
Oligo1	**~PGGGGACCTATGGGGCTGPAPCGAGTPGGTTCCTG~**	11.6%
Oligo2 (LG1)	**~PGGTGGGCGGAGGTCTZGCTACAPGPTTTGGPGGC~**	11.37%
Oligo3 (LG2)	**~PGCCCGGGPTAPPGTGPTGGGTGTTCGCTATCCAG~**	7.98%
Oligo4 (LG3)	**~GGTAACTAGTAGTTGACCCTGPAGTGZTGTPTCTG~**	6.01%
Oligo5 (LG4)	**~GGCGGGGTZGPGTAAGGGGTCTAAGGCATTGGGTC~**	4.48%
Oligo6	**~PGACGGGAGTTPTGTGAGZTGAPCACCTGGTGATG~**	3.62%
Oligo7 (LG5)	**~GGAGGAAGTGGTCCTTGCTTTGCZTCGTATCTGGG~**	2.57%
Oligo8 (LG6)	**~GGTZGATTATTPGGTTCAATAACACPTCCTGGTGG~**	1.96%
Oligo9 (LG7)	**~PGCACAGTGTGZZCCATAGGTTGTAATGACPTZTG~**	1.04%
Oligo10	**~PGAGAACGPGGTPZGGGGTGTGGGAAAGGTCCTAG~**	0.98%
Oligo11 (LG8)	**~GGCAGCZCCTGPAGTPGAGTGTPATGGCTTATTCG~**	0.91%