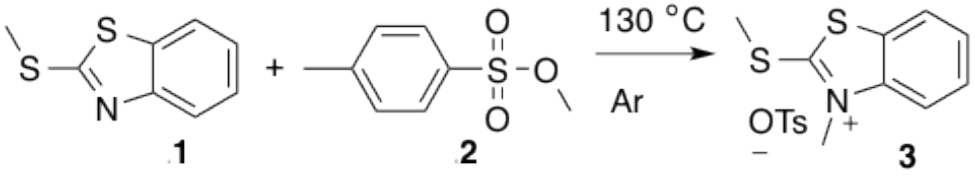
Supplementary Information

1. Synthesis

Chemicals were purchased from Sigma-Aldrich, Fisher Scientific, or Acros Organics and used without further purification. All solvents were of reagent grade or better and purchased from Fisher Scientific. NMR spectra were collected on a Varian 400MHz NMR using isotope-enriched solvents from Cambridge Isotopes. Mass spectrometry was performed in water/methanol or water/acetonitrile mixtures on an Agilent instrument.

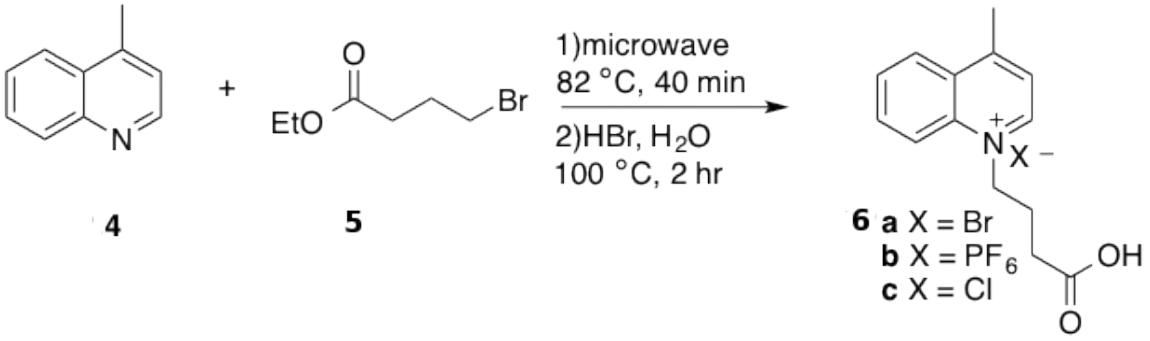
1. Synthesis of a Carboxylic Acid Thiazole Orange Derivative (TO1)

Benzothiazolium salt **3** was synthesized in good yield from commercially available starting materials according to the procedures of Bethge *et al*.[1] (Scheme S1).

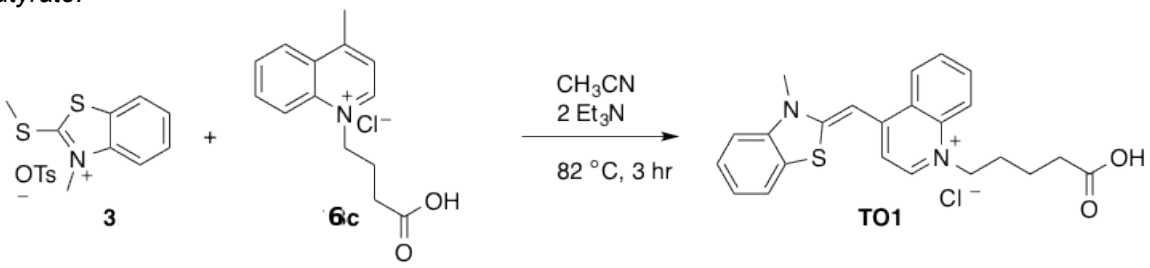


*Scheme S1. The reaction of **1** and **2** to form thiazolium salt **3**.*

Quinolinium salt **6** was synthesized using 4-methylquinoline (**4**) and ethyl 4-bromobutyrate (**5**) (Scheme S2). The ethyl ester was hydrolyzed to the acid, and water soluble salt **6a** was precipitated with KPF_6 to obtain **6b**. The counter ion was exchanged for chloride, and then **6c** was reacted with **3** in acetonitrile to obtain **TO1** (Scheme S3). **TO1** was purified by recrystallization from methanol/water.



*Scheme S2. The synthesis of quinolinium salts **6a-c** by reaction of 4-methylquinoline and ethyl 4-bromobutyrate.*



*Scheme S3. Reaction of thiazolium salt **3** with quinolinium salt **6c** to form a carboxylic acid thiazole orange derivative **TO1**.*

*1.1.1 Synthesis of 3-methyl-2-(methylthio)-benzothiazolium tosylate (**3**)*

3.7 g (20 mmol) of 2-methylthiobenzothiazole and 4.5 g (24 mmol) of methyl *p*-toluene sulfonate were added to a dry 100 mL round bottom flask and heated under inert atmosphere to 130 °C. After 1 hour, the solution turned yellow and the temperature was reduced to 70 °C. Acetone was added to the solution until a yellow solid appeared, and the solid was heated for an additional 30 min. After cooling to RT, white crystals were isolated by vacuum filtration and washed with acetone, then dried under high vacuum overnight, yielding 5.71 g of **3** (76% yield).

^{1}H-NMR (400 MHz, CD_3OD): δ 2.34 (s, 3H), 3.11 (s, 3H), 4.13 (s, 3H), 7.18 (d, *J*=7.9, 2H), 7.66 (d, *J*=8.2, 2H), 7.71 (td, *J*=1.0, 8.2, 1H), 7.83 (td, *J*=1.2, 8.5, 1H), 8.06 (d, *J*=8.5, 1H), 8.21 (d, *J*=8.2, 1H).

^{13}C-NMR (100 MHz, CD_3OD): δ 18.64, 21.44, 37.00, 116.60, 124.77, 127.03 (2C), 128.63, 129.89, 130.02 (2C), 130.85, 141.68, 143.86, 144.21, 183.35.

*1.1.2 Synthesis of 1-(3-carboxypropyl)-4-methylquinolin-1-ium chloride (**6c**)*

1.4 g of lepidine (9.8 mmol) and 2.2 g of 4-bromobutyrate (11.2 mmol) were added to a dry 5 mL microwave tube equipped with a stirbar and microwaved for 40 minutes at 82°C. The microwave tube was rinsed with MeOH, and the resulting purple liquid was poured into 100 mL of Et_2O and stirred until the ether became clear, then the ether was decanted. Upon combining the purple liquid with 50 mL of H_2O in a 100 mL rbf, the solution turned burnt orange. Then 2 mL of 48% HBr was added to the water and the solution was refluxed overnight. The solution volume was reduced by evaporation, and a solution of saturated KPF_6 was added dropwise until the remaining liquid became cloudy. After precipitation in the fridge overnight, an off-white solid, **6b**, was obtained from vacuum filtration. The product was dissolved in 3:1 MeCN:H_2O and passed through an Amberlite® 402 Cl ion exchange column and lyophilized. A white fluffy solid was isolated but turned into a hard taupe solid upon exposure to moisture. Yield of **6c** was 0.171 g (7%).

6b: ^{1}H-NMR (400MHz, CD_3CN): δ 2.27 (m, 2H), 2.53 (t, 2H), 3.00 (d, 3H), 4.92 (m, 2H), 7.85 (d, 1H), 8.02 (m, 1H), 8.48 (m, 2H), 8.90 (d, 2H).

^{19}F-NMR (400 MHz, CD_3CN): δ -74.58, -72.02.

^{13}C-NMR (100 MHz, CD_3CN): δ 20.85, 25.89, 30.84, 58.18, 120.28, 124.02, 128.50, 131.06, 131.12, 136.79, 138.78, 149.00, 161.16, 174.50.

MP: 100-105 °C

6c: 1H-NMR (400MHz, CD3CN): δ 2.57 (t, 3H), 2.99 (s, 3H), 4.95 (2H), 7.86 (d, 1H), 7.99 (t, 1H), 8.21 (t, 1H), 8.47 (d, 1H), 8.54 (d, 1H), 8.98 (d, 1H).

MP: 196-198 °C

*1.1.3 Synthesis of **TO1***

The thiazolium tosylate salt **3** (0.735 g, 2.0 mmol) and quinolinium chloride **6c** (0.534 g, 2.0 mmol) were combined into a dry 100 mL rbf. After adding dry CH_3CN (40 mL), the flask was equipped with a condenser, and the mixture heated to 82°C. 600 µL triethylamine (0.44 g, 4.3 mmol) was injected into the mixture, causing the orange solution to immediately change to a

dark red. The solution was heated under Ar for three hours. The flask was cooled and 100 mL Et_2O added into the suspension. The solid was vacuum filtered and rinsed with Et_2O and then recrystallized from MeOH/H_2O. The orange-red solid weighed 0.442 g (54% yield).

^{1}H-NMR (400 MHz, DMSO-d_6): δ 1.07 (t, *J*=7.0, 1H), 2.11-2.01 (m, 2H), 2.41 (t, *J*=7.1, 2H), 3.37 (dd, *J*=7.0, 14.0, 3H), 3.98 (s, 3H), 4.55 (t, *J*=7.6, 2H), 6.88 (s, 1H), 7.34 (d, *J*=7.2, 1H), 7.40 (t, *J*=7.6, 1H), 7.59 (t, *J*=7.8, 1H), 7.73 (t, *J*=7.1, 2H), 7.98 (dd, *J*=7.8, 17.2, 2H), 8.16 (d, *J*=8.6, 1H), 8.55 (d, *J*=7.2, 1H), 8.75 (d, *J*=8.2, 1H).

^{13}C-NMR (150 MHz, DMSO-d_6): δ 24.11, 30.30, 33.70, 53.41, 88.03, 107.87, 112.88, 117.83, 122.74, 123.77, 124.11, 124.45, 125.68, 126.74, 128.11, 133.20, 136.96, 140.31, 144.17, 148.50, 159.98, 173.69.

HRMS-ESI (*m/z*): $[M]^+$ calcd for $(C_{22}H_{21}N_2O_2S)^+$ 377.1324; found: 377.1324.

MP: 237-239 °C

1.2 Peptide Coupling

Synthesis of all peptides was accomplished using an automated peptide synthesizer (Protein Technologies) using Fmoc chemistries and Wang resins. All deprotection and coupling steps were repeated twice. Resins and protected amino acids were obtained from NovaBioChem. HPLC-grade solvents were purchased from Sigma-Aldrich and prepared by addition of 0.1% TFA (v/v), filtration through a 0.2 micron filter, and then degassed using sonication. HPLC purification was performed on a Shimadzu HPLC with a preparative C-18 column using water and acetonitrile as the mobile phase.

1.2.1 General Procedure

For a 200 μmol scale reaction, an Fmoc-protected Wang resin (100-200 mesh) coupled to the C-terminal residue was swelled in DMF. The resin was deprotected with a 20% piperidine solution then washed sequentially with DMF and DCM. A DMF solution of the next peptide in the sequence (100 mM), with appropriate protecting groups, was added to the reaction vessel for coupling between the amine and carboxylic acid. The coupling reagent used was PyBOP (300

mM) with DIPEA (1.2 M) in DMF. After coupling to the final residue in the peptide sequence, the resin was washed with DMF and DCM and removed from the synthesizer. The resin was manually deprotected using 20% piperidine until a positive Kaiser test resulted. At this point, a solution of **TO1** in HOBT/HBTU (300 mM each) with1.2 M DIPEA in DMF was added, and the resin shaken until a negative Kaiser test resulted. The resin turned blood red and retained a red stain after washing with 20 mL DMF, 20 mL DCM, and 20 mL MeOH. The resin was rinsed with 15 mL acetic acid and vacuum dried overnight. Peptides were cleaved from the resin using a TFA cleavage cocktail (95% TFA, 2.5% TIPS, 1.5% EDT, 1% H_2O) and shaking for 4 hours. The TFA solution was separated from the resin by filtration under aspirator pressure, the resin was washed with TFA, and the filtrate collected. TFA was removed from the filtrate using rotary evaporation. Precipitation of the peptides using cold ether resulted in an orange solid.

1.2.2 HPLC Purification and Isolation

Solid peptides were taken up in a mixture of 50/50 DMSO/H_2O and purified using preparative HPLC. The method was as follows: 0-35 min of 5-30% MeCN in H_2O, 35-45 min of 30% MeCN, 45-60 min of 90% MeCN. The absorbance at 500 nm was monitored and only fractions containing this absorbance were collected.

Fractions containing the desired peptide were verified using LC/MS then combined, concentrated under vacuum to remove acetonitrile, and lyophilized to remove water. Low resolution mass spectrometry was used for each peptide (Table S1).

TO1-peptide	Expected Mass (m/z)			Found (m/z)
	$(M+2H)^{3+}$	$(M+H)^{2+}$	(M^{+})	
TO1-1	315.5	472.7	944.5	315.6, 472.6, 944.1
TO1-2	282.8	423.7	846.4	282.9, 423.7, 846.3
TO1-3	292.8	438.7	876.4	292.9, 438.7, 876.3
TO1-4	310.8	465.7	930.4	310.9, 465.7, 930.3
TO1-5	292.8	438.7	876.4	292.9, 438.4, 876.3
TO1-6	305.4	457.7	914.3	457.6, 914.0
TO1-7	302.6	453.7	906.4	302.8, 453.6, 907.1
TO1-8	273.1	409.2	817.4	409.2
TO1-9	298.1	446.7	892.4	298.2, 446.6, 892.0

Table S1. FIA-ESI + MS (low-res) on HPLC purified peptides.

2. DNA Isolation and Amplification

2.1 General Procedures

All oligonucleotides were ordered from Integrated DNA Technologies (IDT), were dissolved in nuclease-free water to a concentration of 100 μM, and stored at -20C. Polymerase chain reaction (PCR) was performed under one of two following conditions. Using Taq Polymerase from NEB: 2.5U enzyme per 50ß μL reaction, 1X ThermoPol buffer, 0.2 mM dNTPs, 4 μM primers, 0.1 ng/μL template or 10 ng/μL template (genomic) were added. Thermal cycling steps were 1) 98 °C for 2 min, 2) 98 °C for 30 sec, 3) Annealing T for 30 sec, 4) 68 °C for 1 min/kb, 5) hold at 4 °C, and steps 2-4 were repeated for as many cycles as necessary to yield

optimal amplification. Using Q5 high fidelity polymerase from NEB 1X buffer, 0.2 mM dNTPs, 4 μM primers, 1.25 U/25 μL polymerase, and 0.1 ng/μL template were added to 50 or 100 μL reactions. Thermal cycling steps were 1) 98 °C for 30 sec, 2) 98 °C for 10 sec, 3) Annealing T for 30 sec, 4) 72 °C for 25 sec/kb, 5) 72 °C for 2 min, 6) hold at 4 °C, with steps 2-4 repeated as needed.

Concentrations of dsDNA were estimated from absorbance values measured at 260 nm on a nanodrop. Sequences of dsDNA were input into the spreadsheet provided from Tautarov *et al*. in order to estimate the sequence and length-dependent molar absorptivity of each strand at 260 nm. The concentrations of each dsDNA solution were calculated using the absorbance values at 260 nm and the estimated molar absorptivity.

Gel electrophoresis was performed using agarose gels of variable concentration (3-4% for less than 100 bp, 1-2% for 1000 bp or higher) in 1x TAE buffer. Samples were loaded with glycerol loading dyes and run against a Fermentas GeneRuler DNA Ladder Mix or O'RangeRuler 10 bp DNA ladder (<150 bps). Ethidium bromide was added to the gels upon preparation in order to stain the DNA. Gels were run at 15 V/cm until appropriate loading dyes ran at least half the length of the gel, and then gels were imaged with UV illumination.

2.2 1000-mer

Genomic DNA was isolated from buccal cells collected after rinsing the mouth with 0.1 M saline for 2 minutes. The buccal cells were lysed and extracted following the procedures of the Sigma-Aldrich GenElute™ miniprep kit. Isolated genomic DNA was stored at -20°C. Gel electrophoresis showed that the DNA had not been sheared to smaller than 15,000 bps (Figure S2). Concentration of the genomic DNA was measured as 58.4 ng/uL on a nanodrop.

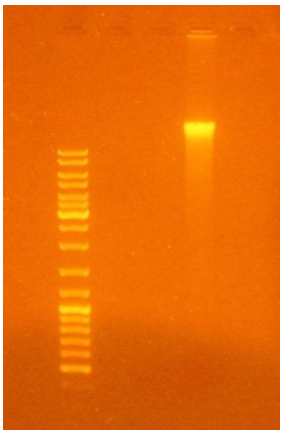


Figure S1. Genomic DNA isolated from buccal cells run against GeneRuler ladder on 1% agarose in 1X TAE.

A search through the NIH genome database Nucleotide identified a highly conserved genomic sequence of 2500 bps that codes for the β2 adrenergic receptor. Primers for a 1000 bp product from the gene were developed using Primer-3 and crosschecked in the Primer-BLAST program from NIH. Primer-BLAST showed that the only product that the primers would give from a genomic template was the intended 1000 bp sequence:

5'CACCAACTACTTCATCACTTCACTGGCCTGTGCTGATCTGGTCATGGGCCTGGCAGTGGTGCCCTTTGGGGCCGCCCATATTCTTA

TGAAAATGTGGACTTTTGGCAACTTCTGGTGCGAGTTTTGGACTTCCATTGATGTGCTGTGCGTCACGGCCAGCATTGAGACCCTGT

GCGTGATCGCAGTGGATCGCTACTTTGCCATTACTTCACCTTTCAAGTACCAGAGCCTGCTGACCAAGAATAAGGCCCGGGTGATC

ATTCTGATGGTGTGGATTGTGTCAGGCCTTACCTCCTTCTTGCCCATTCAGATGCACTGGTACCGGGCCACCCACCAGGAAGCCATC

AACTGCTATGCCAATGAGACCTGCTGTGACTTCTTCACGAACCAAGCCTATGCCATTGCCTCTTCCATCGTGTCCTTCTACGTTCCC

CTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCAGCTCCAGAAGATTGACAAATCTGAGGGCCGCTT

CCATGTCCAGAACCTTAGCCAGGTGGAGCAGGATGGGCGGACGGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAAGGAG

CACAAAGCCCTCAAGACGTTAGGCATCATCATGGGCACTTTCACCCTCTGCTGGCTGCCCTTCTTCATCGTTAACATTGTGCATGTG

ATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTCCTAAATTGGATAGGCTATGTCAATTCTGGTTTCAATCCCCTTATCTAC

TGCCGGAGCCCAGATTTCAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTATGGGAATGGCTACTC

CAGCAACGGCAACACAGGGGAGCAGAGTGGATATCACGTGGAACAGGAGAAAGAAAATAAACTGCTGTGTGAAGACCTCCCAGG

CACGGAAGACTTTGTGGGCCATCAAGGTACTGTGCCTAGCGATAACATTGAT-3'

Primers were ordered from IDT and dissolved in nuclease-free water to a concentration of 100 μM. The conditions for PCR were optimized by running a temperature gradient to find the ideal annealing temperature of the primers to the genomic sequence, 58.5 °C. The product band at 1000 bp was gel-extracted and purified, and the number of cycles for this template optimized for use in further PCR reactions (Figure S3). For amplification with NEB Taq, the annealing temperature was 58.5 °C; for NEB Q5, the annealing temperature was the recommended 3 °C above the lower primer T_m. 56 °C. PCR products were purified using a QIAquick PCR purification kit, and the A_{260} of the resulting DNA measured by nanodrop and its concentration calculated from ε=15791142 M^{-1} cm^{-1}.

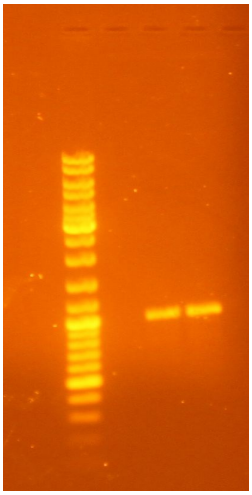


Figure S2. PCR amplified 1 kb dsDNA after optimization of annealing temperature and cycles. Samples were run on a 1.5% agarose gel in 1X TAE against GeneRuler ladder.

3. Peptide Conjugates of Thiazole Orange

Tripeptide sequences for **TO1-2**, **TO1-3**, **TO1-5**, **TO1-7**, and **TO1-8** were chosen using data from Kolonin *et al.* on association of phage-displayed peptides with cancer cells from the NCI-60 panel (Figure S3).[2] Chosen sequences showed high association for one or more cell lines as well as variable association responses for other cell lines in the panel. Sequences used in **TO1-1**, **TO1-4**, **TO1-6**, and **TO1-9** were identified from the literature and used based on their ability to bind to molecules overexpressed on cancer cell surfaces.

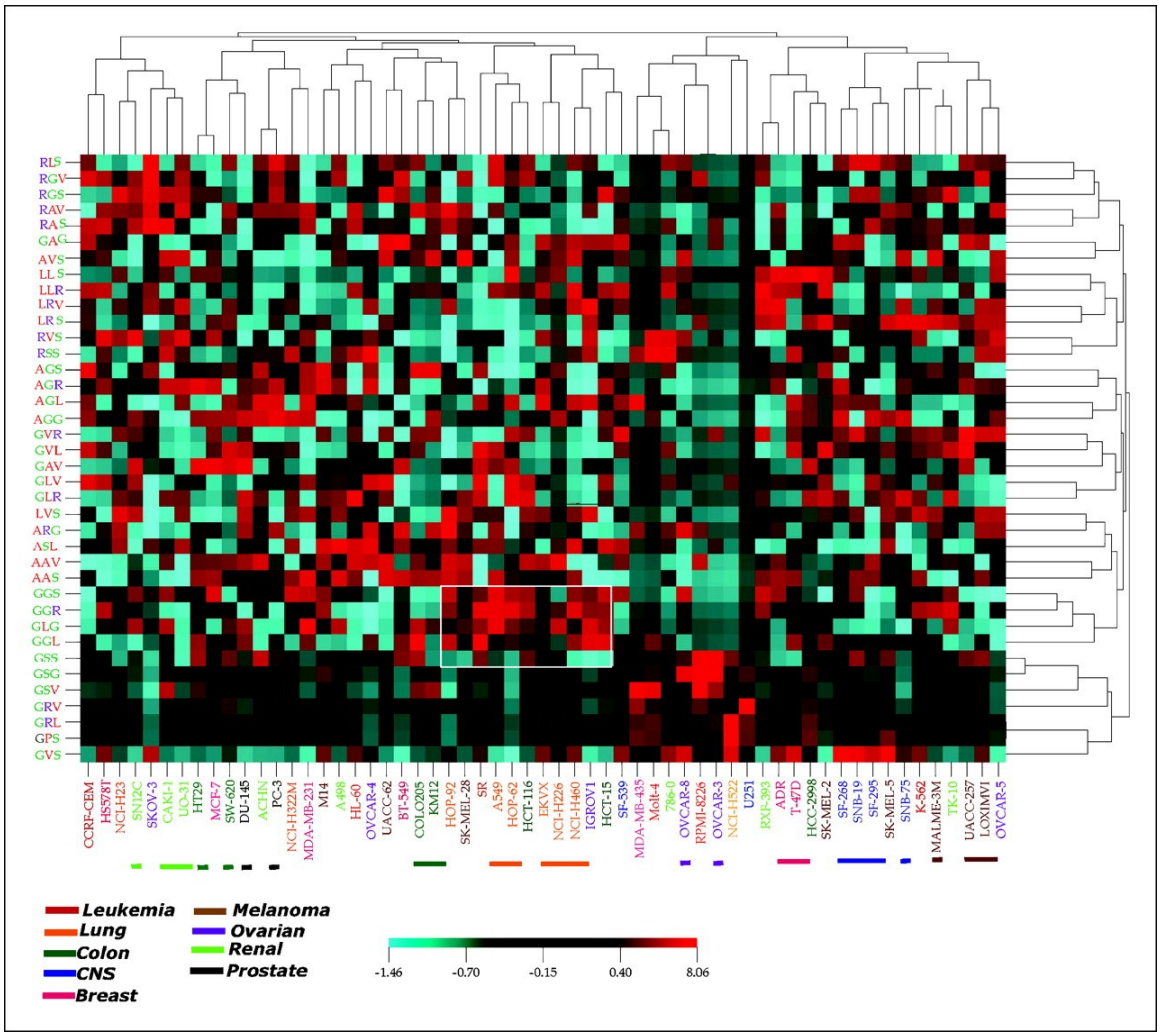


Figure S3. Tripeptide motifs identified as highly associated or not associated with particular cancer cell lines from the NCI-60 panel. Figure from Kolonin *et al.*[2]

To form a supramolecular assembly between **TO1-peptide** conjugates and the chosen dsDNA backbone, solutions of peptide conjugates of were titrated into DNA at varying equivalences in 96-well plates and fluorescence emission was monitored. Well plates were measured on a BioTek Synergy 2 Multimode Microwell-plate reader, with fluorescence filters as follows: excitation – 485/20, emission – 528/20. Well plates for the cell-patterning reproduction were measured on a BioTek Cytation3 Microwell-plate reader, using monochromators set for excitation at 485 nm and emission at 528 nm.

Titrations were undertaken in Dulbecco's phosphate buffered saline without calcium or magnesium (Gibco® Life Technologies). The formulation is 200 mg/L KCl, 200 mg/L KH_2PO_4, 8 g/L NaCl, and 2.16 g/L $Na_2HPO_4 \cdot 7H_2O$ which results in a 10 mM phosphate buffer with 150 mM NaCl at pH = 7.4.

3.1 Titrations of DNA and TO1-peptides

A stock solution of dsDNA was diluted to give a final concentration of 4 nM (1 kb strand) in appropriate wells of 96-well black, clear bottom plates (Costar® 3631) filled with DPBS. Stock solutions of peptides were added to each well at different volumes in order to provide different concentrations and thus equivalents. Fluorescence emission was measured and subtracted from DPBS blank with dsDNA in order to give a delta emission value for each well.

Titration of each peptide derivative of **TO1** was undertaken with 1 kb dsDNA in DPBS (Figures S4-S12). Differences in equivalences of peptide required to saturate the fluorescence signal, whether saturation was reached under the tested concentrations, as well as differences in the highest emission intensity reached, were observed for the variable peptide sequences. For example, titrations of **TO1-4** and **TO1-7** with 4 nM 1000mer dsDNA lead to different emission intensities and saturation behaviors. The emission of **TO1-4** reached saturation at 2000-2500 equivalents (2 – 2.5 bp equivs) with a maximum relative emission at 250,000 (Figure S8). In contrast, the titration with **TO1-7** reached saturation at approximately 1000 equivalents (1 bp equivalent) with a maximum relative emission of 350,000, which is 1.4x higher than **TO1-4** (Figure S11). Although both peptides contain a single lysine, which allows for greater association with the negatively charged DNA, the placement of lysine in **TO1-7** is closer to the site of intercalation than that of **TO1-4**. The additional positive charge closer to the intercalator may allow for greater intercalation of **TO1-7** compared to **TO1-4**, resulting in greater fluorescence emission and saturation at a lower concentration.

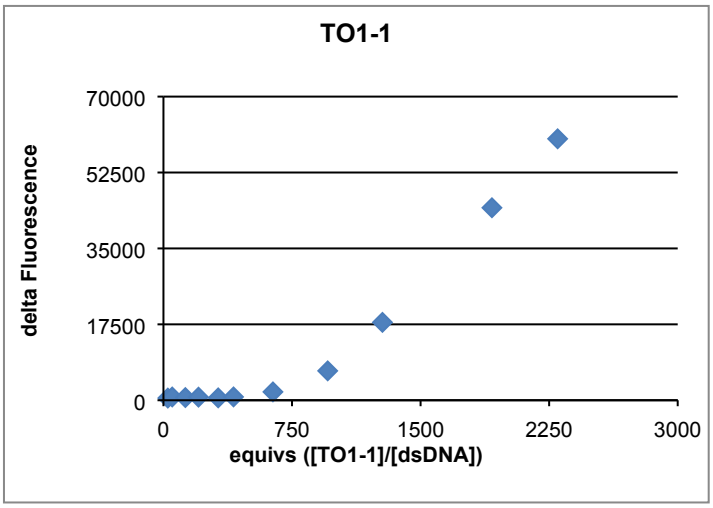


Figure S4. Fluorescence emission of titration of 1kb dsDNA (4 nM) with **TO1-1** in DPBS, pH = 7.4. Maximum fluorescence is seen beyond 2300 equivalents, or 9.2 μM, **TO1-1**.

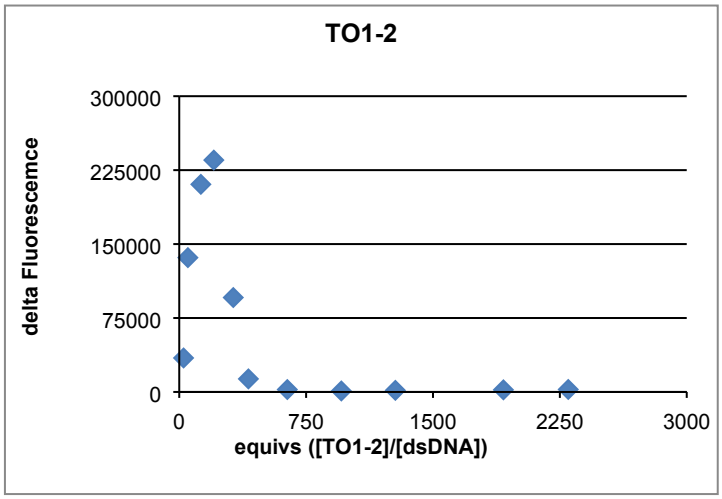


Figure S5. Fluorescence emission of titration of 1kb dsDNA (4 nM) with **TO1-2** in DPBS, pH = 7.4. Maximum fluorescence is seen at 200 equivalents, or 0.8 μM, **TO1-2**.

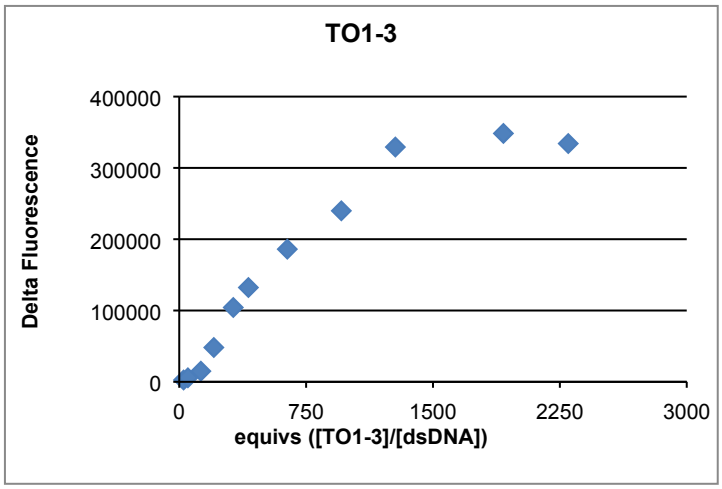


Figure S6. Fluorescence emission of titration of 1kb dsDNA (4 nM) with **TO1-3** in DPBS, pH = 7.4. Fluorescence saturation is seen at 1500 equivalents, or 6 μM, **TO1-3**.

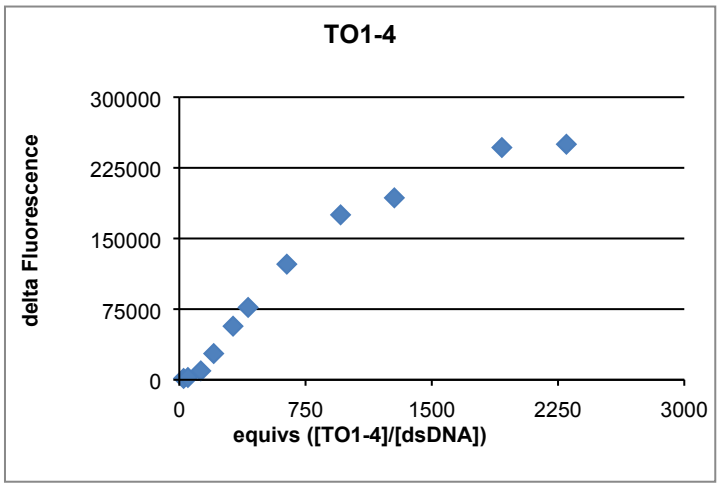


Figure S7. Fluorescence emission of titration of 1kb dsDNA (4 nM) with **TO1-4** in DPBS, pH = 7.4. Fluorescence saturation is seen at 2500 equivalents, or 7.8 μM, **TO1-4**

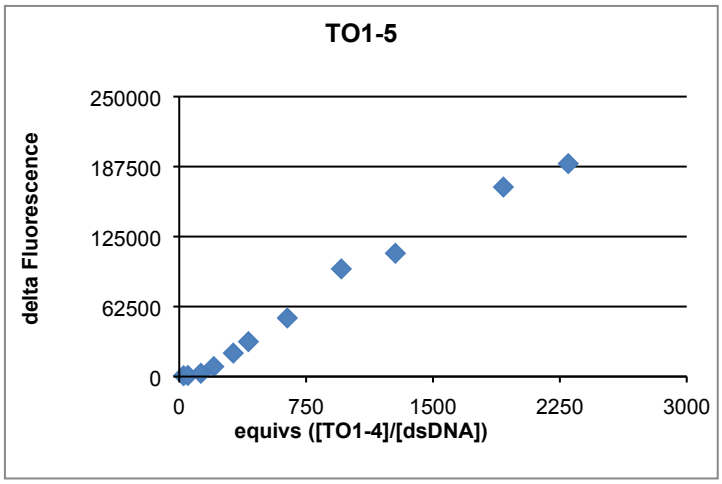


Figure S8. Fluorescence emission of titration of 1kb dsDNA (4 nM) with **TO1-5** in DPBS, pH = 7.4. Maximum fluorescence is seen beyond 2300 equivalents, or 9.2 µM, **TO1-5**.

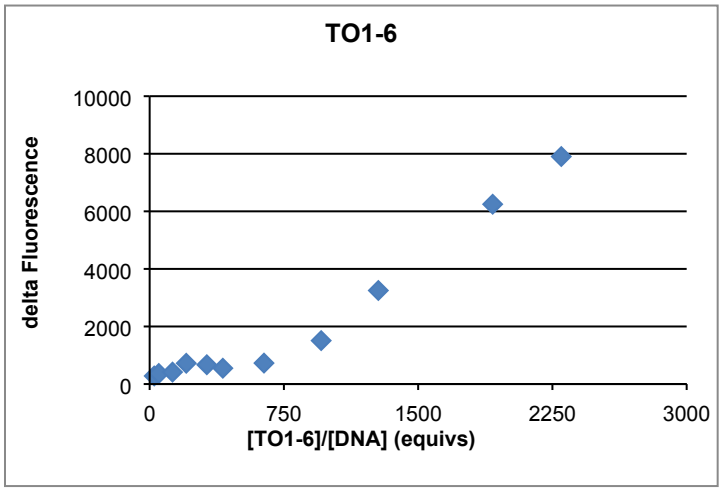


Figure S9. Fluorescence emission of titration of 1kb dsDNA (4 nM) with **TO1-6** in DPBS, pH = 7.4. Maximum fluorescence is seen beyond 2300 equivalents, or 9.2 µM, **TO1-6**.

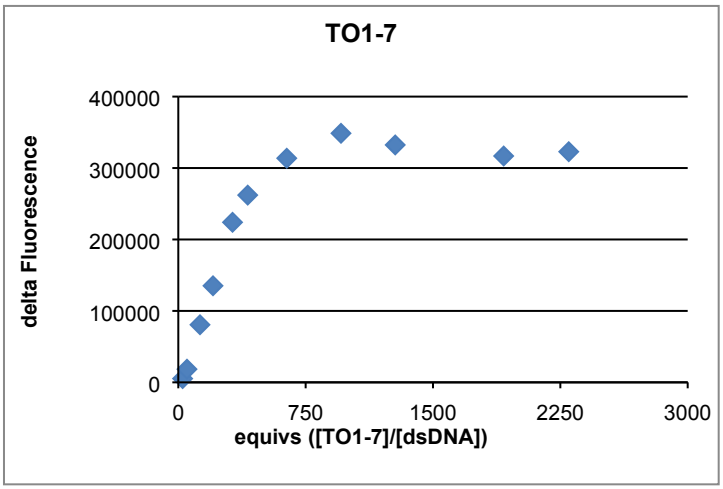


Figure S10. Fluorescence emission of titration of 1kb dsDNA (4 nM) with **TO1-7** in DPBS, pH = 7.4. Maximum fluorescence is seen at 1000 equivalents, or 3.7 μM, **TO1-7**.

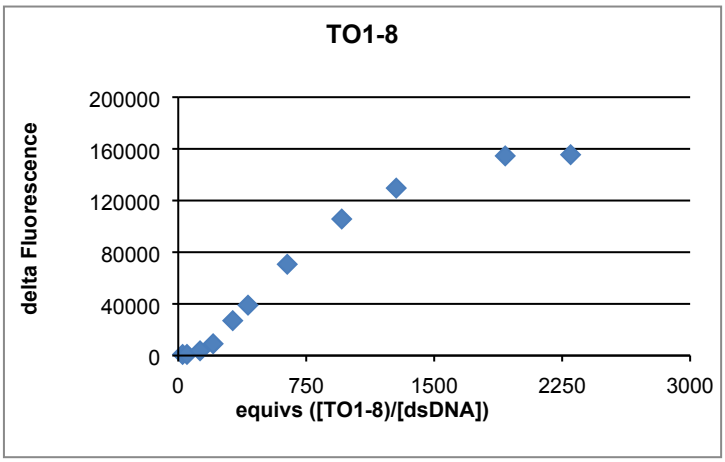


Figure S11. Fluorescence emission of titration of 1kb dsDNA (4 nM) with **TO1-8** in DPBS, pH = 7.4. Maximum fluorescence is seen at 1900 equivalents, or 7.6 μM, **TO1-8**.

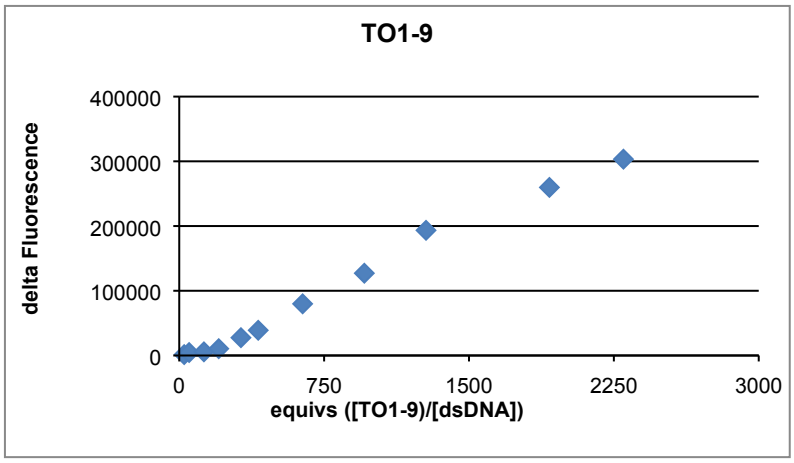


Figure S12. Fluorescence emission of titration of 1kb dsDNA (4 nM) with **TO1-9** in DPBS, pH = 7.4. Maximum fluorescence is seen beyond 2300 equivalents, or 9.2 µM, **TO1-9**.

3.2 Peptide concentration for cell patterning

Peptide concentrations used in cell patterning experiments were determined, using the titrations, at the concentration which resulted in maximum fluorescence emission (Table S2). Where saturation was not reached, a maximum concentration of 10 uM was used.

TO1-peptide	**Concentration (µM)**
TO1-1	10.0
TO1-2	0.8
TO1-3	5.8
TO1-4	7.8
TO1-5	10.0
TO1-6	10.0
TO1-7	3.7
TO1-8	7.8
TO1-9	10.0

Table S2. Concentrations of peptides used in cell patterning experiments, as determined by titration with 4 nM 1 kb dsDNA.

4. Cell Patterning

Cancer cells chosen from the NCI-60 panel exhibited different tissue origins as well as surface epitopes. MOLT-4 is a leukemia line isolated from peripheral blood of a 19-year-old male in relapse for acute lymphoblastic leukemia and is a suspension cell line of T lymphoblasts. SKMEL28 is a malignant melanoma isolated from a 51-year-old male. A549 is a non-small cell lung carcinoma consisting of alveolar basal epithelial cells isolated from a 58-year-old male. HCT15 is a colorectal adenocarcinoma isolated from colon tissue. SKOV3 is an adenocarcinoma isolated from ovarian ascites of a 64-year-old female.[3] The cells are tumorigenic and, although they express estrogen receptors, they are growth resistant to estrogen[4] and overexpress human epidermal growth factor receptor 2 (erbB-2).[5] MDAMB231 is an adenocarcinoma of mammary gland tissue derived from a metastatic site from a 51-year-old female that expresses EGFR and transforming growth factor alpha receptors (TGFαR). DU145 is a carcinoma of prostate tissue derived from a metastatic site from a 69-year-old male. 786O is a renal cell adenocarcinoma derived from the kidney of a 58-year-old male.

The cell line used for novelty detection, U87MGΔVIII, is a glioblastoma isolated from the brain of a 44-year-old male; it has been transfected with a deletion variant of epidermal growth factor receptor 1 (EGFR) to emulate *in vitro* the endogenous expression seen in tumors.[6, 7] The deletion variant EGFRvIII has a truncated extracellular domain, resulting in a constitutively active protein that does not contain binding sites for known ligands.

Solutions of **TO1-peptides** were added to cells grown to confluency as outlined in the manuscript experimental. Fluorescence values were measured for solutions both with and without DNA. Normalized fluorescence values for **TO1-peptides** without DNA are shown for three cell lines in Figures S13 and S14, which can be compared to the normalized fluorescence values for **TO1-peptides** with DNA (Figures 2 and 3). The peptide percentages shown in Figure S13 show

that emissions of peptides that have not been exposed to cells are more similar to each other than those of unexposed peptides intercalated into DNA (Figure 2). The emission intensities after exposure to cells are not dependent on the original emission intensities of the peptides, which are representative of the concentrations. Differences in fluorescence intensity of unexposed **TO1-peptide** conjugates were observed depending on peptide sequence and the presence or absence of DNA. Additionally, the change in emission intensity for peptides without DNA was extremely large after exposure to cells, as the non-radiative relaxation of the fluorophore was inhibited when bound to cells as compared to in free solution. Therefore, normalization to the total fluorescence of the nine sensors — as outlined in the manuscript — was used. As shown for three cell lines, the delta percent values show that there are slight differences in the response of each peptide without DNA to the different cells; however, the response across the peptide array is similar for each cell line.

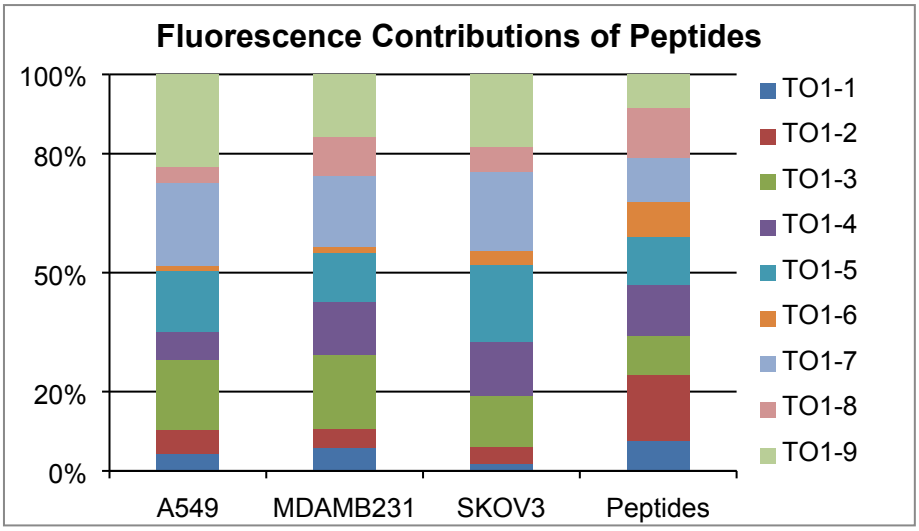


Figure S13. Average percent contribution of each peptide for each cell line, as compared to the contribution of the peptides exposed to cells.

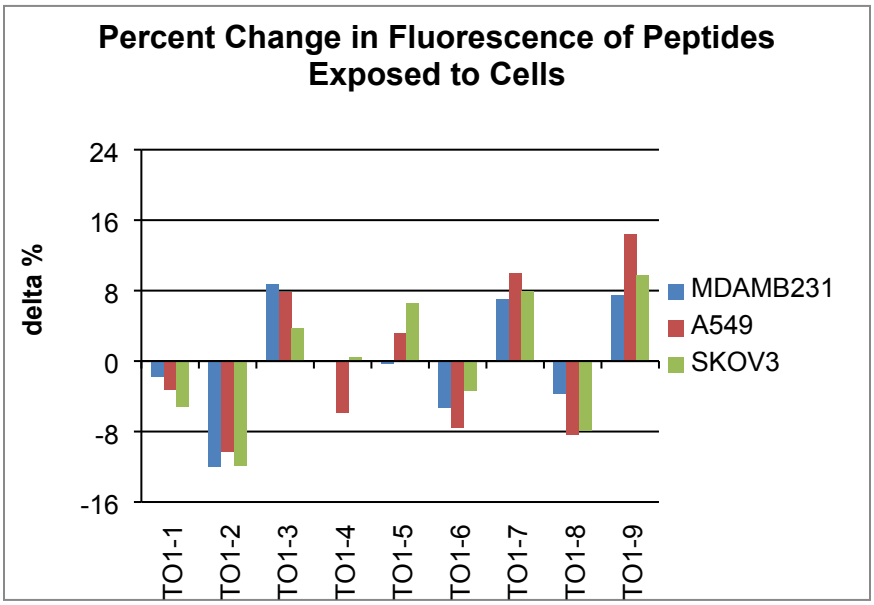


Figure S14. Average delta percent response of the three cell lines to the peptide array.

4.1 Linear Discriminant Analysis

The R package MASS was used for linear discriminant analysis with the lda function. Prior probabilities were calculated based on input class proportions. Scores for each cell line were calculated from the response of the sample to all of the variables that contributed to the classification axis. The placement of each cell line on the score plot is dependent on the loading of variables on the classification axis and the response of the cell sample to those variables, which is reflected in Figure 4. Due to the cross-reactive nature of the array, each classification axis is not determined by a single variable; instead, several variables contribute to the position of each cell line.

4.1.1. LDA classification of eight cell lines

The LDA model built from the data of the initial eight cell lines resulted in five significant linear discriminants. The plot of the first two most significant linear discriminants is seen in Figure 4 in the manuscript.

4.1.2 LDA classification of a novel cell line

Because LDA is inherently a biased classification method, each data point in Figure S15 was classified as belonging to one of the eight pre-defined classes based on its proximity. For both U87MGΔVIII(◆■) experiments, the data were most frequently classified as A549(△) (77.7%), followed by SKOV3(✱) (16.7%) and MDAMB231(◇) (5.6%). The two experiments exhibit similar scores, illustrating the reproducibility of the method. Although visual discrimination of U87MGΔVIII from A549 is not clear from Figure S16, it becomes apparent on higher dimensions, as shown in Figure S18. The fifth discriminant axis accounts for 6.8% of the classification in the original model, and it provides for better clustering of the two U87MGΔVIII data sets while showing improved visual discrimination from the other cell lines.

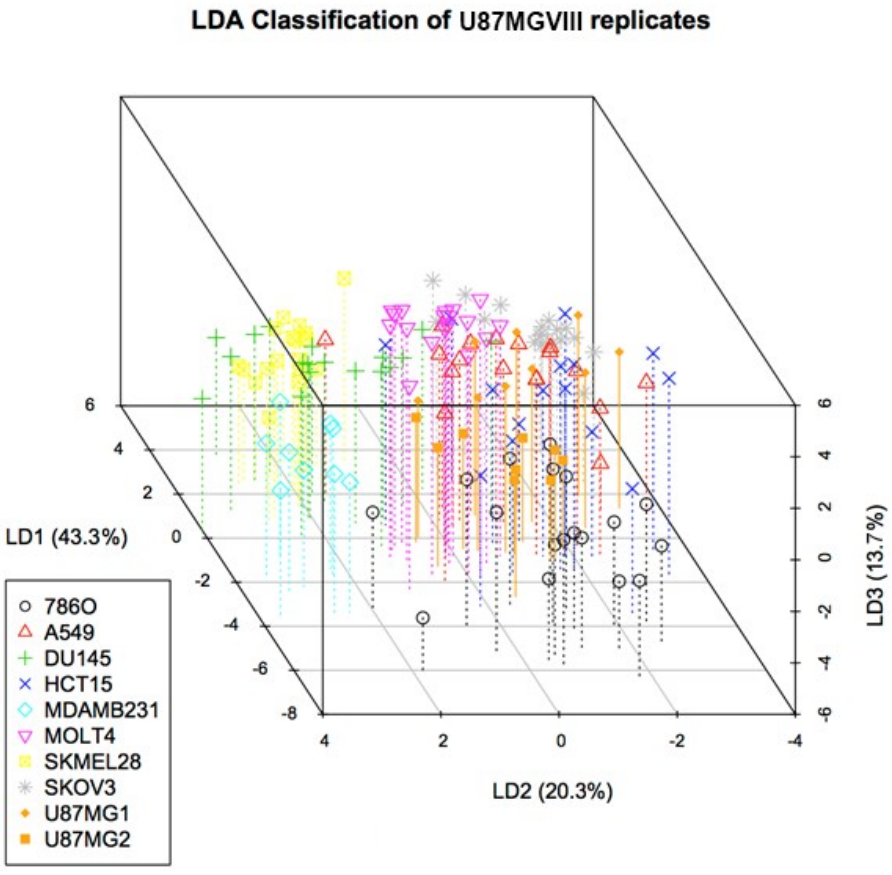


Figure S15. LDA score plot of two U87MG cell line experiments input into the eight cell line model.

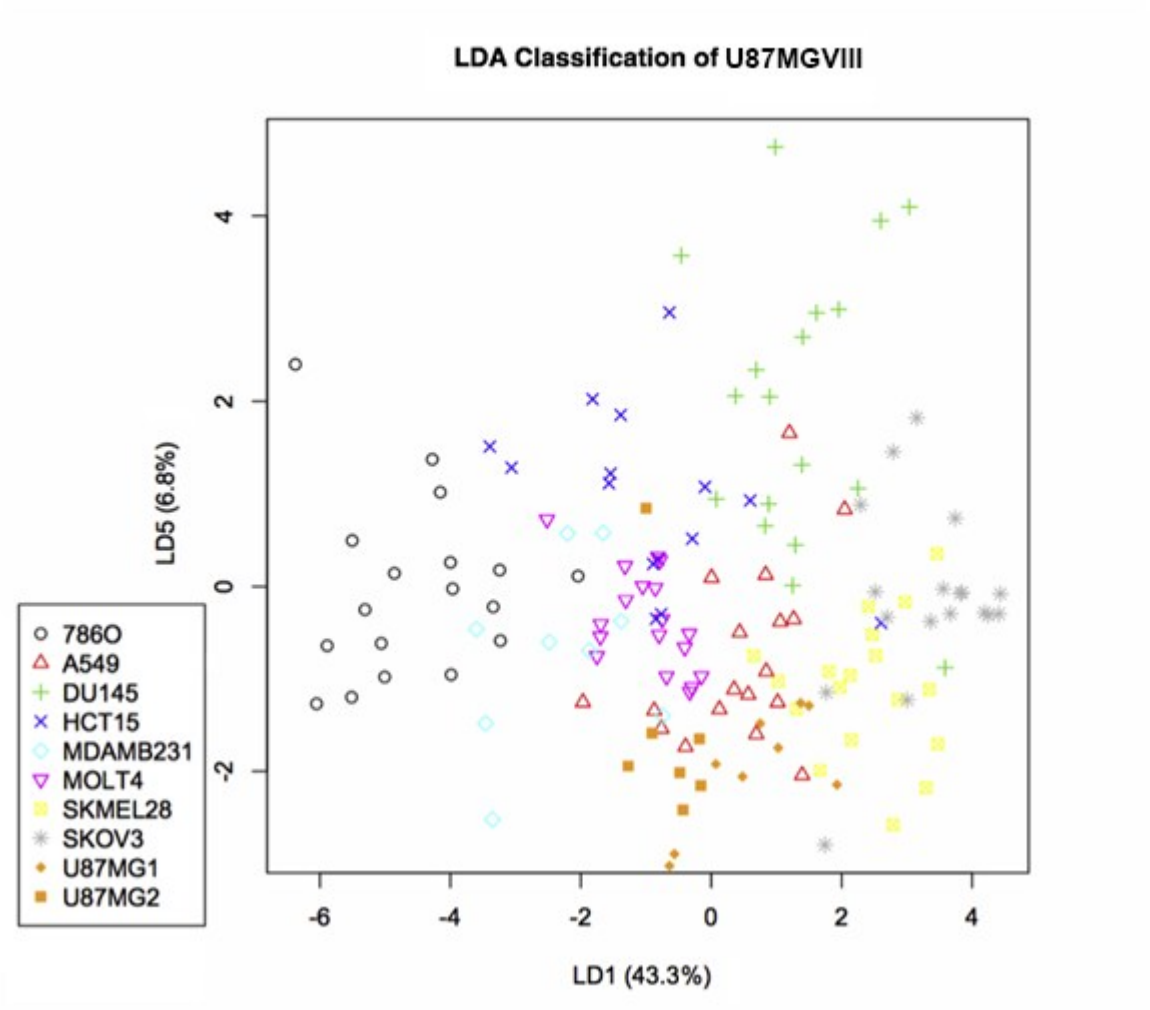


Figure S16. Clustering of U87MGVIII replicates and visual discrimination from the nearest neighbor, A549, is improved by plotting the scores on the fifth discriminant axis.

On LD5, both sets of U87MGΔVIII data score negatively. DNATO1-6, which contains a tripeptide mimic of mannose anticipated to bind to cell surface carbohydrate receptors, contributes negatively to this discriminant axis and to the score of the cell line. Sensors containing the sequence of TO1-9, which was included in the array to bind to erbB-family proteins, did not contribute significantly to the score of the deletion mutant EGFR expressing

glioblastoma, as may be expected. The TO1-3 sequence also contributes to the negative score of this cell line. This sequence had been reported to bind to SKOV3 and HCT15 cells, but in the context of this sensor, it behaves in a cross-reactive manner to contribute to classification of a cell line it was not previously shown to bind to.

4.2 Support Vector Machine Analysis

The R package e1071 was used for support vector machine analysis with the *svm* function. This package uses LIBSVM, a library for support vector machines. Classification was determined using one-against-one binary classifiers for pairs of cell lines with nu-classification and a second-degree polynomial kernel. Parameters were determined prior to training and prediction by parameter tuning on the full data set, resulting in nu = 0.1, gamma = 0.2, and coef0 = 0.3. Partitions were created using the *createDataPartition* function of the R package caret. Statistics from 100 random partitions of the data into training (60%) and test (40%) sets are shown in Table S3. The example training set resulted in 62 support vectors. Feature weights for the one-versus-one classification of this training set were calculated by multiplying the support vectors for a particular class by the coefficients of those support vectors, which resulted in weight vectors of the eighteen variables for separation of a cell type from each other cell type. The absolute values of these weights were summed for each cell type to give an indication of how important each variable was for separation of that cell type from other cell types. (Table S3).

test #	**LOOCV**	**test accuracy**	**train accuracy**	test #	**LOOCV**	**test accuracy**	**train accuracy**
1	0.875	0.980769231	1	51	0.9125	0.961538462	1
2	0.875	0.884615385	1	52	0.9	0.903846154	1
3	0.9	0.903846154	1	53	0.9625	0.769230769	1
4	0.9125	0.846153846	1	54	0.8875	0.942307692	1
5	0.9125	0.942307692	1	55	0.9	0.903846154	1
6	0.9125	0.923076923	1	56	0.95	0.846153846	1
7	0.8875	0.923076923	1	57	0.8875	0.923076923	1
8	0.9	0.942307692	1	58	0.925	0.865384615	1
9	0.8625	0.942307692	1	59	0.925	0.846153846	1
10	0.9	0.923076923	1	60	0.875	0.942307692	1
11	0.9375	0.903846154	1	61	0.9	0.961538462	1
12	0.9	0.846153846	1	62	0.8875	0.884615385	1
13	0.8875	0.942307692	1	63	0.925	0.884615385	1
14	0.9125	0.942307692	1	64	0.875	0.884615385	1
15	0.9125	0.923076923	1	65	0.9	0.903846154	1
16	0.975	0.846153846	1	66	0.9	1	1
17	0.9125	0.961538462	1	67	0.9	0.923076923	1
18	0.8625	0.846153846	1	68	0.925	0.884615385	1
19	0.925	0.903846154	1	69	0.9	0.903846154	1
20	0.8875	0.865384615	1	70	0.85	0.942307692	1
21	0.8875	0.884615385	1	71	0.8875	0.942307692	1
22	0.8625	0.942307692	1	72	0.9375	0.865384615	1
23	0.875	0.923076923	1	73	0.9375	0.846153846	1
24	0.9125	0.865384615	1	74	0.9	0.903846154	1
25	0.9375	0.826923077	1	75	0.9375	0.846153846	1
26	0.9375	0.923076923	1	76	0.925	0.903846154	1
27	0.9	0.884615385	1	77	0.9	0.865384615	1
28	0.9	0.903846154	1	78	0.85	0.980769231	1
29	0.95	0.846153846	1	79	0.9125	0.942307692	1
30	0.875	0.923076923	1	80	0.8875	0.942307692	1
31	0.9	0.884615385	1	81	0.9125	0.923076923	1
32	0.9125	0.884615385	1	82	0.9375	0.826923077	1
33	0.9125	0.807692308	1	83	0.8625	0.884615385	1
34	0.925	0.846153846	1	84	0.8875	0.961538462	1
35	0.9375	0.865384615	1	85	0.95	0.846153846	1
36	0.8875	0.942307692	1	86	0.925	0.846153846	1
37	0.925	0.903846154	1	87	0.8875	0.846153846	1
38	0.8875	0.980769231	1	88	0.875	0.903846154	1
39	0.8625	0.865384615	1	89	0.8875	0.961538462	1
40	0.8875	0.923076923	1	90	0.9125	0.903846154	1
41	0.875	0.865384615	1	91	0.875	0.903846154	1
42	0.8875	0.942307692	1	92	0.8625	0.961538462	1
43	0.9375	0.923076923	1	93	0.925	0.865384615	1
44	0.8875	0.903846154	1	94	0.8875	0.942307692	1
45	0.9125	0.884615385	1	95	0.8875	0.923076923	1
46	0.925	0.903846154	1	96	0.95	0.884615385	1
47	0.925	0.923076923	1	97	0.925	0.884615385	1
48	0.875	0.961538462	1	98	0.8625	0.961538462	1
49	0.95	0.846153846	1	99	0.9	0.942307692	1
50	0.925	0.865384615	1	100	0.9	0.903846154	1
				mean	0.9045	0.900769231	1
				median	0.9	0.903846154	1
				minimum	0.85	0.769230769	1

Table S3. Statistics for LOOCV, test set accuracy, and training set accuracy for 100 partitions of eight cell line data evaluated with SVM.

	SKMEL28	DU145	A549	MOLT4	786O	HCT15	SKOV3	MDAMB 231
DNATO1-1	0.329	**7.294**	2.752	2.817	2.383	2.064	3.538	0.237
DNATO1-2	0.642	0.718	1.419	3.019	0.921	1.558	0.474	**3.818**
DNATO1-3	0.682	0.875	3.288	1.985	1.397	**2.770**	0.808	2.303
DNATO1-4	1.495	1.064	4.491	1.227	0.544	2.130	0.592	2.200
DNATO1-5	1.901	0.947	2.122	1.255	0.922	1.899	0.652	**3.090**
DNATO1-6	**3.535**	2.119	2.418	2.093	3.559	0.374	**9.394**	1.816
DNATO1-7	0.746	**4.607**	**4.829**	0.930	3.828	0.551	2.649	1.401
DNATO1-8	1.153	2.120	4.299	1.056	0.314	0.743	4.144	0.381
DNATO1-9	2.000	**5.130**	4.284	0.621	**4.212**	1.677	**4.424**	1.543
TO1-1	1.271	3.991	3.319	2.108	4.199	2.629	0.579	1.321
TO1-2	**4.392**	0.534	0.879	2.484	**5.722**	0.843	3.217	2.003
TO1-3	0.065	1.929	2.045	2.715	1.596	0.547	1.417	2.287
TO1-4	0.551	2.868	**6.795**	0.782	0.695	0.729	1.481	0.871
TO1-5	1.520	2.302	2.558	0.902	3.514	**4.111**	0.702	2.380
TO1-6	**4.831**	0.978	**5.019**	**3.379**	1.868	**2.899**	1.352	1.837
TO1-7	0.565	2.074	2.074	0.955	2.214	1.119	0.842	2.586
TO1-8	1.882	1.633	3.654	1.965	0.480	0.606	0.765	0.126
TO1-9	0.488	3.559	4.023	1.357	**4.971**	1.170	**5.379**	**2.854**

Table S4. Summed feature weights of variables for eight-cell classification. The three largest feature weights for each cell line are indicated in bold.

Novelty detection was determined using one-classification with a second-degree polynomial; parameters were nu = 0.1, gamma = 0.07, and coef0 = 0.2. The partition fractions of eight cell line data into training and test sets were made such that the negative test set (20 cases) would be comparable in size to the positive test set (18 cases). Statistics for 100 partitions of the eight cell line data compared with two experimental sets of novel cell line data are presented in Table S5.

Partition #	train accuracy	LOOCV	positive accuracy	negative accuracy	Partition #	train accuracy	LOOCV	positive accuracy	negative accuracy
1	0.857	0.777	0.750	0.778	51	0.839	0.813	0.700	0.556
2	0.848	0.813	0.800	0.667	52	0.848	0.786	0.750	0.556
3	0.866	0.804	0.850	0.611	53	0.893	0.813	0.850	0.722
4	0.848	0.795	0.850	0.778	54	0.866	0.768	0.800	0.778
5	0.857	0.777	0.750	0.722	55	0.839	0.777	0.700	0.556
6	0.866	0.777	0.700	0.778	56	0.857	0.795	0.750	0.889
7	0.857	0.768	0.800	0.611	57	0.866	0.813	0.900	0.611
8	0.875	0.795	0.700	0.667	58	0.884	0.795	0.800	0.778
9	0.821	0.750	0.800	0.611	59	0.857	0.795	0.800	0.833
10	0.830	0.759	0.850	0.611	60	0.839	0.804	0.900	0.667
11	0.857	0.813	0.750	0.667	61	0.893	0.795	0.950	0.611
12	0.821	0.795	0.800	0.556	62	0.857	0.804	0.900	0.667
13	0.893	0.777	0.800	0.778	63	0.813	0.714	0.650	0.722
14	0.839	0.795	0.600	0.722	64	0.866	0.804	0.750	0.722
15	0.857	0.795	0.850	0.778	65	0.866	0.786	0.800	0.556
16	0.866	0.777	0.650	0.667	66	0.857	0.777	0.850	0.778
17	0.866	0.768	0.700	0.556	67	0.866	0.786	0.700	0.778
18	0.848	0.786	0.750	0.611	68	0.830	0.804	0.850	0.833
19	0.893	0.795	0.950	0.611	69	0.830	0.768	0.800	0.778
20	0.848	0.750	0.750	0.722	70	0.857	0.777	0.750	0.778
21	0.830	0.768	0.850	0.722	71	0.857	0.795	0.800	0.722
22	0.866	0.804	0.750	0.778	72	0.848	0.795	0.750	0.667
23	0.848	0.750	0.550	0.778	73	0.857	0.804	0.700	0.556
24	0.875	0.768	0.700	0.778	74	0.830	0.768	0.750	0.722
25	0.884	0.795	0.900	0.556	75	0.857	0.804	0.700	0.778
26	0.875	0.795	0.750	0.667	76	0.857	0.768	0.800	0.722
27	0.866	0.777	0.750	0.611	77	0.866	0.786	0.850	0.722
28	0.857	0.777	0.800	0.722	78	0.830	0.777	0.800	0.556
29	0.875	0.786	0.750	0.667	79	0.875	0.777	0.750	0.667
30	0.866	0.795	0.750	0.611	80	0.875	0.804	0.750	0.778
31	0.866	0.786	0.950	0.556	81	0.866	0.821	0.750	0.778
32	0.839	0.768	0.700	0.611	82	0.866	0.777	0.750	0.722
33	0.857	0.777	0.800	0.778	83	0.839	0.741	0.750	0.667
34	0.875	0.795	0.700	0.778	84	0.848	0.768	0.900	0.556
35	0.884	0.795	0.800	0.778	85	0.848	0.795	0.750	0.889
36	0.848	0.795	0.700	0.778	86	0.866	0.768	0.800	0.611
37	0.848	0.777	0.800	0.611	87	0.866	0.804	0.700	0.722
38	0.857	0.759	0.600	0.944	88	0.866	0.813	0.700	0.778
39	0.875	0.786	0.700	0.667	89	0.839	0.786	0.850	0.611
40	0.857	0.795	0.850	0.611	90	0.839	0.786	0.550	0.778
41	0.857	0.786	0.850	0.778	91	0.848	0.795	0.850	0.722
42	0.884	0.813	0.700	0.667	92	0.866	0.759	0.650	0.778
43	0.866	0.786	0.900	0.778	93	0.875	0.777	0.900	0.556
44	0.884	0.786	0.900	0.667	94	0.884	0.821	0.750	0.722
45	0.875	0.804	0.700	0.611	95	0.830	0.804	0.900	0.611
46	0.884	0.786	0.900	0.722	96	0.848	0.777	0.650	0.833
47	0.839	0.759	0.850	0.889	97	0.821	0.804	0.750	0.778
48	0.839	0.786	0.750	0.667	98	0.857	0.795	0.650	0.667
49	0.848	0.768	0.750	0.778	99	0.839	0.750	0.800	0.611
50	0.830	0.804	0.800	0.778	100	0.839	0.786	0.950	0.556
					average	0.857	0.785	0.777	0.698
					mean	0.857	0.786	0.750	0.722
					minimum	0.813	0.714	0.550	0.556

Table S5. Statistics for training set, LOOCV, positive test set, and negative test set accuracy for novelty detection of 100 partitions of eight cell line data with a novel cell line.

References

1. L. Bethge, D. V. Jarikote and O. Seitz, *Bioorgan Med Chem*, 2008, **16**, 114-125.
2. M. G. Kolonin, L. Bover, J. Sun, A. J. Zurita, K.-A. Do, J. Lahdenranta, M. Card√≥-Vila, R. J. Giordano, D. E. Jaalouk, M. G. Ozawa, C. A. Moya, G. R. Souza, F. I. Staquicini, A. Kunyiasu, D. A. Scudiero, S. L. Holbeck, E. A. Sausville, W. Arap and R. Pasqualini, *Cancer Research*, 2006, **66**, 34-40.
3. D. P. M. Hughes, D. G. Thomas, T. J. Giordano, L. H. Baker and K. T. McDonagh, *Cancer Research*, 2004, **64**, 2047-2053.
4. W. Hua, T. Christianson, C. Rougeot, H. Rochefort and G. M. Clinton, *J Steroid Biochem Mol Biol*, 1995, **55**, 279-289.
5. L. K. Shawver, E. Mann, S. S. Elliger, T. C. Dugger and C. L. Arteaga, *Cancer Research*, 1994, **54**, 1367-1373.
6. H. S. Friedman and D. D. Bigner, *New England Journal of Medicine*, 2005, **353**, 1997-1999.
7. H. K. Gan, A. N. Cvrljevic and T. G. Johns, *FEBS Journal*, 2013, **280**, 5350-5370.