Supporting Information for:

Synthetic Lectin Arrays for the Detection and Discrimination of Cancer Associated Glycans and Cell Lines

Kevin L. Bicker,[b] Jing Sun,[a] Morgan Harrell,[a] Yu Zhang,[a] Maria M. Pena,[a] Paul R. Thompson*[b] and John J. Lavigne*[a]

[a] *Department of Chemistry & Biochemistry, University of South Carolina, 631 Sumter Street, Columbia, SC, USA 29208. Fax: (803)-777-9521; Tel: (803)-777-5264; E-mail: JLavigne@chem.sc.edu*

[b] *Department of Chemistry, The Scripps Research Institute, Scripps Florida, 120 Scripps Way, Jupiter, Florida, USA 33458. Fax: (561)-228-3050; Tel: (561)-228-2860; E-mail: PThompso@scripps.edu*

Methods

SL Library and Hit Validation. SL library synthesis, screening, sequencing, and SL hit validation have already been described.[1]

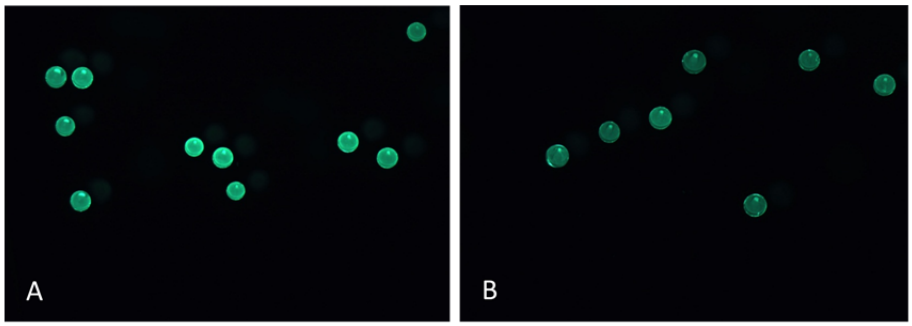

Figure S1. Representative images of an SL screening. SL5 (**A**) and the library (**B**) after binding with FITC-PSM.

Chemicals. FITC and FITC-streptavidin were purchased from Sigma-Aldrich (Milwaukee, WI). Fmoc-protected amino acids, O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate (HBTU), and rink amide resin were purchased from Novabiochem Corp (Gibbstown, NJ). Biotin labeled glycans were purchased from GlycoTech Corp. (Gaithersburg, MD). Human serum was purchased from PAA Laboratories, Inc. (Dartmouth, MA) and cell growth media and supplies were purchased from VWR International, LLC (West Chester, PA). All other chemicals were purchased from Acros Organics (Morris Plains, NJ) and used without further purification.

Microscopy. All fluorescent images were taken using a Leica MZ 16F microscope equipped with a GFP filter set (excitation 450-490 nm; emission filter 500-550 nm), and a QImaging

MicroPublisher 5.0 RTV digital camera. The images were analyzed with Adobe Photoshop, using the elliptical marquee tool to select the entire bead, obtaining luminosity values that represent the fluorescent intensity of individual beads. High average luminosity values correspond to brighter beads that presumably bind strongly to the analyte under study.

FITC-Labeled Cell Membrane Extracts. Cells were grown to ~80% confluence, at which point they were scraped and collected from four T75 flasks and washed with PBS. The Qiagen Plasma Membrane Protein Kit (Qiagen, Inc., Valencia, CA) was used to isolate the cell plasma membrane proteins and glycoproteins. After adjusting the pH of the resulting solution to 8.8, 28 µg of FITC (71.9 nmol) was added in DMF and the solution tumbled gently at 37 °C for 1 h. This solution was transferred to a 3.5 kDa Slide-A-Lyzer cassette (Thermo Fisher Scientific, Inc., Rockford, IL) and dialyzed against Screening Buffer (500 mL) at 4 °C for 6 h, changing the buffer twice more.

SL Screening in Human Serum. SL2 and SL5 were screened using the methods described previously[1] in the presence of human serum. Briefly, 0.1 mg/mL FITC-tagged analytes (i.e., OVA, BSM, PSM, and BSA) in varying concentrations of Screening Buffer containing diluted human serum (i.e., 0, 25, 50, and 95%), uniformly containing 1% BSA were added to individual tubes of SL resin (2 mg) and incubated at 23 ºC for 16 h. Library resin was screened in parallel to act as a control. Resin was subsequently washed, imaged, and analyzed as previously described (**Figure S2A and S2B**).[1]

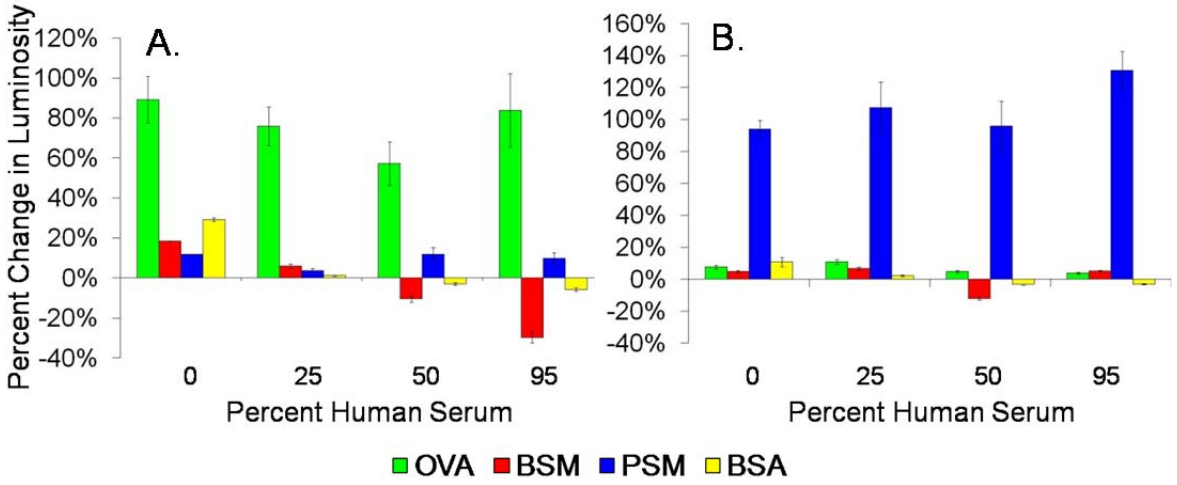

Figure S2. Plots of the relative percent change in luminosity for SL2 (**A.**) and SL5 (**B.**) in varying amounts of human serum, indicating that these SLs maintain their selectivity in complex media.

Human Serum Screening Controls. Parallel screenings to those done to test the selectivity of SL2 in human serum were done without any fluorescent glycoprotein to assure no component in the serum produced any appreciable background fluorescence (**Figure S3**).

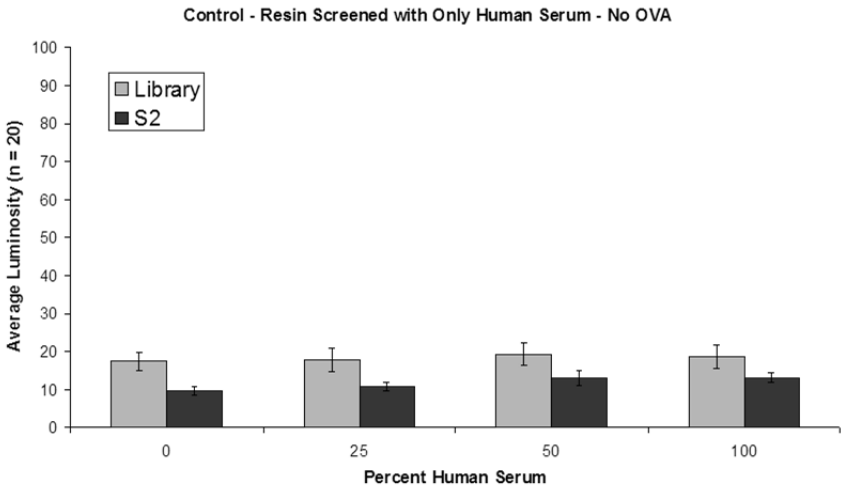

Figure S3. Human serum screening controls showing that the serum contributes no appreciable fluorescence.

Determination of Bead-Based SL5 K_d. Portions of SL5 (2 mg) were prepared and screened as described previously with varying concentrations of FITC-PSM (i.e., 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.25, 0.5, 1.0, 2.0, and 4.0 μM) for 6 h in Screening Buffer. Resin was subsequently washed, imaged, and analyzed to acquire average luminosities. This data was plotted as fraction bound versus FITC-PSM concentration and fit to a single site ligand binding model using GraFit 5.0.11 (**Figure S4**).[2]

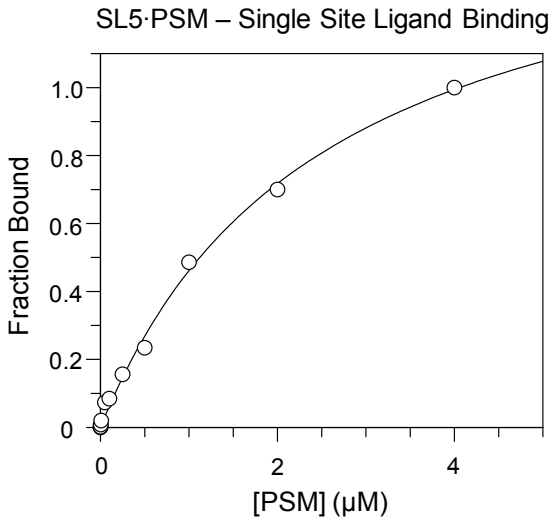

Figure S4. Data for treatment of bead based SL5 with varying concentrations of FITC-PSM is fit to a single site ligand binding model using GraFit to obtain a K_d of 2.5 ± 0.29 μM as an upper limit.

Synthesis of FITC-SL5. A peptide of the sequence Ac-RA**D***TRV**D***VBBRM**K***-resin, where B is beta-alanine, **D*** denotes a boronic acid functionalized Dab, and **K*** denotes a FITC functionalized Lys, was synthesized on rink amide resin (200-400 mesh; 0.71 mmol/g substitution) using Fmoc strategies.[1] After peptide synthesis, the N-terminus was acetylated in 10 mL of CH_2Cl_2 containing 5% acetic anhydride and 5% pyridine. The attachment of the phenylboronic acid moieties then proceeded as previously described.[1] Fmoc-Lys(Mtt)-OH was used on the C-terminus to provide an orthogonal handle to attach the fluorophore. The Mtt protecting group was removed by treating the resin 5 times with 15 mL of 1% TFA in CH_2Cl_2, 5 min each, washing with DCM between treatments. Resin was then treated with FITC (1.5 eq, 67.6 mg, 0.1737 mmol) in 10 mL of 5% NMM in DMF 2 h with tumbling. Resin was subsequently washed with DMF, MeOH, and DMF and cleaved from the resin with 95% TFA, 2.5% water, 2.5% TIS for 6 h. This solution was collected, the TFA bubbled off with nitrogen gas, the peptide precipitated by diethyl ether, and purified by preparative HPLC on a C18 column.

Fluorescence Polarization.[3] All fluorescence polarization (FP) work was performed on a Molecular Devices, SpectraMax M5 plate reader in 384 well microtiter plates (MTPs), exciting at 480 nm, reading the emission at 525 nm, and using a dichromatic cut-off of 495 nm. Note that all assays were performed in triplicate, in Screening Buffer containing 1% Pluronic F127, with a

sample volume of 30 µL. This assay was first set up at four different concentrations of FITC-SL5 (0.05, 0.5, 5.0, and 50.0 µM), each treated with 10 different concentrations of PSM (0, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0, and 10 µM) for 6 h, taking readings every 30 min. This work identified the optimal FITC-SL5 concentration (0.5 µM) and incubation time (1.5 h), which were used for all subsequent FP assays. This assay was then performed with varying concentrations of OVA (0, 0.1, 1.0, 3.0, 10, 30, 100, 300, 900, 1500, and 2250 µM), BSM (0, 0.1, 1.0, 3.0, 5.0, 10, 30, 57.2 µM), and BSA (0, 0.1, 1.0, 5.0, 10, 30, 100, 500, 800, 1046 µM) (**Figure S5**).

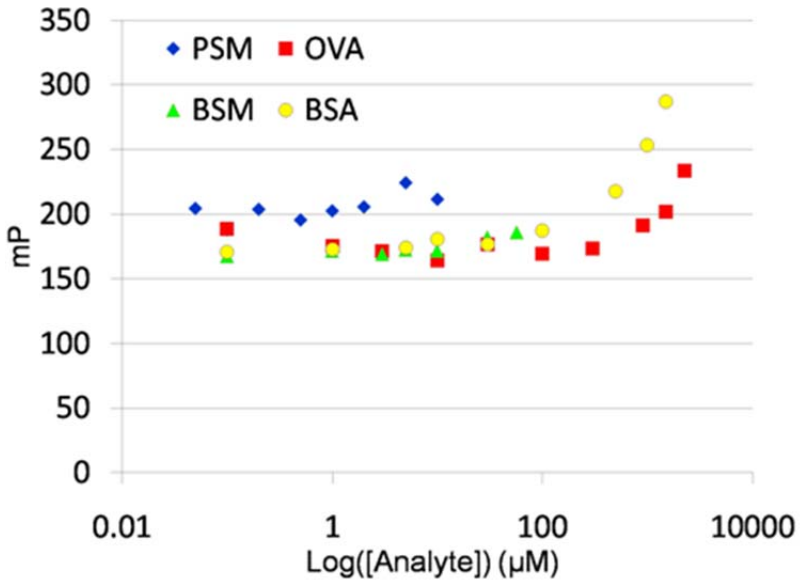

Figure S5. Plot of the fluorescence polarization assay used to monitor binding between monovalent FITC-SL5 and various analytes. Note that a response is only observed at higher concentrations of analyte, indicating that multivalency is an important component for strong binding with bead based SLs.

Glycan Competition Studies. Portions of Dab Fixed Library and SL resin (2 mg) were prepared and screened as previously described with their respective FITC-glycoprotein (0.1 mg/mL; 1 mL) in PBSG doped with varying concentrations (i.e. 0, 0.001, 0.01, 0.10, 1.0, 10, 100, and 1000 mM) of glycan (fructose, mannose, galactose, and N-acetylglucosamine for SL2; fructose, galactose, N-acetylglucosamine, N-acetylgalactosamine, fucose, and sialic acid for SL5). Resin was subsequently washed, imaged, and analyzed as described above.

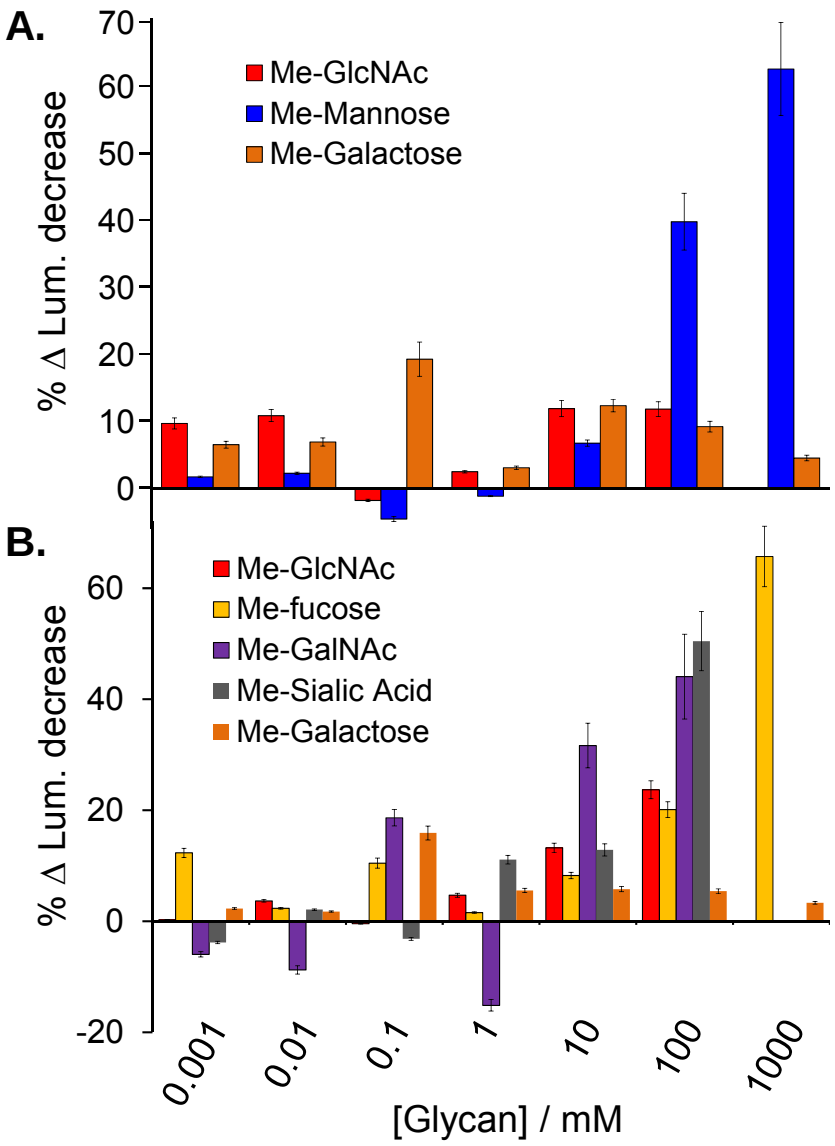

Figure S6. Percent change in luminosity for the non-reducing glycan competition studies used to explore the SL2-OVA (**A**) and SL5-PSM (**B**) binding interactions (analyte identification indicated in the legends above). Error bars represent the standard error of the percent change relative to the control as this propagated uncertainty is based on the variance between replicate measurements for the sample and control reference.

SL Based Array for Glycan Differentiation. Five separate portions of each SL used in the array (2 mg; SLs 1, 3, 4, and 5) were prepared as described previously[1] and each portion subsequently treated with 0.5 mL of a premixed solution of 1 μM FITC-streptavidin and 4 μM glycan (i.e., biotin tagged TF antigen, Le^a, Le^x, sLe^a, or sLe^x) in Screening Buffer (100 mM NaH_2PO4, 150 mM NaCl, 10% glycerol, pH 7.2) for 6 h at 23 ℃ with gentle tumbling. Resin was washed,

imaged, and analyzed as described previously.[1] Luminosity data was analyzed using commercially available feature selection algorithms. Systat 11.00.01was used to carry out all linear discriminant analysis (LDA) determinations and to obtain the general and leave-one-out cross validation classification accuracies.[4] Statistica was used as a graphical program and for training/test set analysis with data sets randomly assigned from the Normal Distribution using Excel.[5]

Group Frequencies

Lea	Lex	TF	sLea	sLex
15	15	15	15	15

Group means

	Lea	Lex	TF	sLea	sLex
VAR00002	0.84	0.71	0.694	0.874	0.638
VAR00003	0.822	0.837	0.766	0.93	0.604
VAR00004	0.723	0.915	0.782	0.787	0.763
VAR00005	0.768	0.909	0.643	0.892	0.57

Between groups F-matrix -- df = 4 67

	Lea	Lex	TF	sLea	sLex
Lea	0				
Lex	47.736	0			
TF	49.662	68.299	0		
sLea	22.861	47.964	120.103	0	
sLex	127.918	144.103	27.151	242.098	0

Wilks' lambda

Lambda =	0.009	df =	4 4 70	
Approx. F=	47.0237	df =	16 205	prob = 0

Classification functions

	Lea	Lex	TF	sLea	sLex
CONSTANT	-525.027	-554.428	-408.568	-635.707	-323.33
VAR00002	501.254	435.076	408.271	529.412	369.486
VAR00003	298.742	289.374	275.866	336.666	211.082
VAR00004	121.903	183.022	163.575	127.759	171.917
VAR00005	380.067	425.913	297.402	438.963	261.963

Included				Excluded		
Variable	F-to-remove	Tolerance		Variable	F-to-enter	Tolerance
VAR00002	21.87	0.9684				
VAR00003	12.97	0.975306				
VAR00004	8.16	0.937653				
VAR00005	35.1	0.921545				

Classification matrix (cases in row categories classified into columns)

	Lea	Lex	TF	sLea	sLex	%correct
Lea	14	0	0	1	0	**93**
Lex	0	15	0	0	0	**100**
TF	0	0	15	0	0	**100**
sLea	0	0	0	15	0	**100**
sLex	0	0	0	0	15	**100**
Total	**14**	**15**	**15**	**16**	**15**	**99**

Jackknifed classification matrix

	Lea	Lex	TF	sLea	sLex	%correct
Lea	13	0	0	2	0	**87**
Lex	0	14	0	1	0	**93**
TF	0	0	15	0	0	**100**
sLea	1	0	0	14	0	**93**
sLex	0	0	0	0	15	**100**
Total	**14**	**14**	**15**	**17**	**15**	**95**

Eigenvalues

17.855 3.182 0.403 0.002

Canonical correlations

0.973 0.872 0.536 0.043

Cumulative proportion of total dispersion

0.833 0.981 1 1

Wilks' lambda

Lambda = 0.009
Approx.F= 47.098 df = 16, 205 p-tail = 0

Pillai's trace

trace = 1.997
Approx.F= 17.451 df = 16, 280 p-tail = 0

Lawley-Hotelling trace

trace =	21.442		
Approx.F=	87.779	df = 16, 262	p-tail = 0

Canonical discriminant functions

	1	2	3	4
Constant	-27.507	4.208	4.121	-12.817
VAR00002	13.356	12.801	6.235	12.552
VAR00003	9.595	3.079	-17.515	-3.183
VAR00004	-3.378	-11.562	-4.343	14.044
VAR00005	16.602	-9.367	11.256	-6.933

Canonical discriminant functions -- standardized by within variances

	1	2	3	4
VAR00002	0.585	0.56	0.273	0.55
VAR00003	0.475	0.152	-0.867	-0.158
VAR00004	-0.184	-0.628	-0.236	0.763
VAR00005	0.746	-0.421	0.505	-0.311

Canonical scores of group means

	1	2	3	4
Lea	1.913	1.941	0.464	-0.058
Lex	2.007	-3.205	0.138	-0.021
TF	-2.853	0.379	-1.124	-0.014
sLea	5.251	0.812	-0.093	0.061
sLex	-6.319	0.073	0.614	0.032

Cell Culture. All cells were grown in monolayers in T75 flasks at 37℃ in 5% CO2 and 95% air. HCT116 and LoVo human colorectal carcinoma cells were grown in Roswell Park Memorial Institute (RPMI) media supplemented with 10% fetal bovine serum (FBS) and 1% antibiotic/antimycotic. HT29 human colorectal carcinoma cells and MCF7 human breast cancer cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and 1% antibiotic/antimycotic. NIH/3T3 murine fibroblast cells were grown in DMEM supplemented with 10% neonatal calf serum and 1% antibiotic/antimycotic.

SL Based Array for Cell Differentiation. SL resin was prepared as described for the glycan array and each portion subsequently treated with a 50-fold dilution of each FITC-cell membrane extract in Screening Buffer for 6 h at 23 °C with gentle tumbling. Resin was washed, imaged,

and analyzed as described above. To account for differences in fluorescent labeling and protein concentration, each individual SL luminosity for one cell type, was divided by the highest SL luminosity for that cell type. Luminosity data was analyzed using commercially available feature selection algorithms. Systat 11.00.01was used to carry out all linear discriminant analysis (LDA) determinations and to obtain the general and leave-one-out cross validation classification accuracies.[4] Statistica was used as a graphical program and for training/test set analysis with data sets randomly assigned from the Normal Distribution using Excel..[5]

Single layered analysis: SL-Array responding to all 7 cell lines

In this initial analysis, the data for each normalized cell line was considered individually to afford 7 classes of analytes.

Group Frequencies

3T3/NIH	CT-26	CT-26-F1	CT-26-FL3	HCT116	HT-29	LoVo
40	40	40	40	80	40	60

Group means

	3T3/NIH	CT-26	CT-26-F1	CT-26-FL3	HCT116	HT-29	LoVo
VAR00002	0.947	0.56	0.693	0.773	0.554	0.789	0.446
VAR00003	0.819	0.795	0.419	0.489	0.601	0.791	0.408
VAR00004	0.614	0.915	0.911	0.911	0.797	0.801	0.822
VAR00005	0.839	0.901	0.849	0.739	0.863	0.592	0.451

Between groups F-matrix -- df = 4 330

	3T3/NIH	CT-26	CT-26-F1	CT-26-FL3	HCT116	HT-29	LoVo
3T3/NIH	0						
CT-26	220.442	0					
CT-26-F1	285.772	239.399	0				
CT-26-FL3	226.529	282.742	27.097	0			
HCT116	250.841	73.187	120.634	185.571	0		
HT-29	84.871	230.024	220.7	125.836	254.145	0	
LoVo	507.747	500.279	148.108	110.028	388.655	256.22	0

Wilks' lambda

Lambda =	0.0034	df =	4 6 333	
Approx. F=	196.8729	df =	24 1152	prob = 0

Classification functions

	3T3/NIH	CT-26	CT-26-F1	CT-26-FL3	HCT116	HT-29	LoVo
CONSTANT	-156.695	-183.161	-120.806	-117.974	-134.333	-136.044	-73.862
VAR00002	36.702	-39.1	12.244	34.36	-24.991	35.854	10.066
VAR00003	179.783	192.498	88.468	96.021	145.102	165.169	81.688
VAR00004	52.271	94	102.48	103.078	82.224	84.144	98.057
VAR00005	113.638	161.204	116.463	87.462	145.873	70.721	56.335

Included			**Excluded**		
Variable	F-to-remove	Tolerance	Variable	F-to-enter	Tolerance
VAR00002	191.56	0.689043			
VAR00003	255.33	0.910461			
VAR00004	50.4	0.970885			
VAR00005	278.88	0.702439			

Classification matrix (cases in row categories classified into columns)

	3T3/NIH	CT-26	CT-26-F1	CT-26-FL3	HCT116	HT-29	LoVo	**%correct**
3T3/NIH	40	0	0	0	0	0	0	**100**
CT-26	0	39	0	0	1	0	0	**98**
CT-26-F1	0	0	33	7	0	0	0	**83**
CT-26-FL3	0	0	2	38	0	0	0	**95**
HCT116	0	9	3	0	67	1	0	**84**
HT-29	0	0	0	1	0	39	0	**98**
LoVo	0	0	0	0	0	0	60	**100**
Total	**40**	**48**	**38**	**46**	**68**	**40**	**60**	**93**

Jackknifed classification matrix

	3T3/NIH	CT-26	CT-26-F1	CT-26-FL3	HCT116	HT-29	LoVo	**%correct**
3T3/NIH	40	0	0	0	0	0	0	**100**
CT-26	0	39	0	0	1	0	0	**98**
CT-26-F1	0	0	33	7	0	0	0	**83**
CT-26-FL3	0	0	2	37	0	0	1	**93**
HCT116	0	9	5	0	65	1	0	**81**
HT-29	0	0	0	1	0	39	0	**98**
LoVo	0	0	0	0	0	0	60	**100**
Total	**40**	**48**	**40**	**45**	**66**	**40**	**61**	**92**

Eigenvalues

8.996	5.183	1.833	0.681

Canonical correlations

0.949	0.916	0.804	0.636

Cumulative proportion of total dispersion

0.539	0.849	0.959	1

Wilks' lambda

lambda =	0.003		
Approx.F=	196.949	df= 24, 1152	p-tail = 0

Pillai's trace

trace =	2.79		
Approx.F=	128.027	df= 24, 1332	p-tail = 0

Lawley-Hotelling trace

trace =	16.693		
Approx.F=	228.49	df= 24, 1314	p-tail = 0

Canonical discriminant functions

	1	2	3	4
Constant	10.124	1.053	-3.704	-10.949
VAR00002	3.478	10.714	7.071	2.339
VAR00003	-13.157	3.309	-7.551	2.9
VAR00004	3.341	-3.312	-0.268	10.893
VAR00005	-9.675	-9.832	5.203	-1.704

Canonical discriminant functions -- standardized by within variances

	1	2	3	4
VAR00002	0.31	0.956	0.631	0.209
VAR00003	-0.873	0.22	-0.501	0.192
VAR00004	0.286	-0.283	-0.023	0.931
VAR00005	-0.778	-0.79	0.418	-0.137

Canonical scores of group means

	1	2	3	4
3T3/NIH	-3.429	3.617	1.008	-1.098
CT-26	-4.052	-2.21	-1.302	1.095
CT-26-F1	1.86	-1.493	2.207	0.365
CT-26-FL3	2.267	0.67	1.667	0.939
HCT116	-1.548	-2.141	-0.049	-0.702
HT-29	-0.584	3.651	-1.228	0.911
LoVo	4.689	0.031	-1.503	-0.537

Analysis layer 1: Discriminate cell lines grouped by class (cancerous, metastatic and healthy)

Group Frequencies

cancerous	metastatic	healthy
160	140	40

Group means

	cancerous	metastatic	healthy
VAR00002	0.615	0.61	0.947
VAR00003	0.697	0.434	0.819
VAR00004	0.827	0.873	0.614
VAR00005	0.805	0.647	0.839

Between groups F-matrix -- df = 4 334

	cancerous	metastatic	healthy
cancerous	0		
metastatic	299.854	0	
healthy	110.096	329.356	0

Wilks' lambda

Lambda =	0.0883	df = 4 2 337	
Approx. F=	197.5482	df = 8 668	prob = 0

Classification functions

	cancerous	metastatic	healthy
CONSTANT	-57.659	-46.876	-59.767
VAR00002	-4.024	11.007	10.079
VAR00003	67.294	20.143	82.162
VAR00004	58.979	82.153	25.13
VAR00005	24.711	6.78	29.857

Included				**Excluded**		
Variable	F-to-remove	Tolerance	|	Variable	F-to-enter	Tolerance
VAR00002	94.91	0.72245	|			
VAR00003	404.85	0.708488	|			
VAR00004	139.1	0.793274	|			
VAR00005	71.01	0.725508	|			

Classification matrix (cases in row categories classified into columns)

	cancerous	meta	healthy	**%correct**
cancerous	152	7	1	**95**
metastatic	0	140	0	**100**
healthy	0	0	40	**100**
Total	**152**	**147**	**41**	**98**

Jackknifed classification matrix

	cancerous	meta	healthy	**%correct**
cancerous	150	8	2	**94**
metastatic	0	140	0	**100**
healthy	0	0	40	**100**
Total	**150**	**148**	**42**	**97**

Eigenvalues

5.207 0.825

Canonical correlations

0.916 0.672

Cumulative proportion of total dispersion

0.863 1

Wilks' lambda

Lambda =	0.088		
Approx.F=	197.548	df = 8, 668	p-tail = 0

Pillai's trace

trace =	1.291		
Approx.F=	152.493	df = 8, 670	p-tail = 0

Lawley-Hotelling trace

trace =	6.032		
Approx.F=	251.098	df = 8, 666	p-tail = 0

Canonical discriminant functions

	1	2
Constant	2.144	4.008
VAR00002	1.853	6.737
VAR00003	-10.903	-4.402
VAR00004	7.645	-5.349
VAR00005	-4.102	-1.818

Canonical discriminant functions -- standardized by within variances

	1	2
VAR00002	0.271	0.986
VAR00003	-1.046	-0.422
VAR00004	0.735	-0.514
VAR00005	-0.666	-0.295

Canonical scores of group means

	1	2
cancerous	-1.294	-0.809
metastatic	2.558	0.363
healthy	-3.778	1.968

Analysis layer 2: Discrimination within grouped cancerous/non-metastatic cell lines (CT-26, HCT116 and HT-29)

Group Frequencies

CT-26	HCT116	HT-29
40	80	40

Group means

	CT-26	HCT116	HT-29
VAR00002	0.56	0.554	0.789
VAR00003	0.795	0.601	0.791
VAR00004	0.915	0.797	0.801
VAR00005	0.901	0.863	0.592

Between groups F-matrix -- df = 4 154

	CT-26	HCT116	HT-29
CT-26	0		
HCT116	60.673	0	
HT-29	258.48	245.509	0

Wilks' lambda

Lambda =	0.0468	df = 4 2 157	
Approx. F=	139.3791	df = 8 308	prob = 0

Classification functions

	CT-26	HCT116	HT-29
CONSTANT	-157.365	-117.263	-122.438
VAR00002	-31	-7.95	79.89
VAR00003	140.045	105.125	109.769
VAR00004	106.948	89.619	72.31
VAR00005	133.929	118.357	58.914

Included				**Excluded**		
Variable	F-to-remove	Tolerance	|	Variable	F-to-enter	Tolerance
VAR00002	150.93	0.684638	|			
VAR00003	64.26	0.870438	|			
VAR00004	14.02	0.87492	|			
VAR00005	101.13	0.76337	|			

Classification matrix (cases in row categories classified into columns)

	CT-26	HCT116	HT-29	**%correct**
CT-26	39	1	0	**98**
HCT116	8	72	0	**90**
HT-29	0	0	40	**100**
Total	**47**	**73**	**40**	**94**

Jackknifed classification matrix

	CT-26	HCT116	HT-29	**%correct**
CT-26	39	1	0	**98**
HCT116	9	71	0	**89**
HT-29	0	0	40	**100**
Total	**48**	**72**	**40**	**94**

Eigenvalues

7.931 1.39

Canonical correlations

0.942 0.763

Cumulative proportion of total dispersion

0.851 1

Wilks' lambda

Lambda = 0.047

Approx.F= 139.379 df = 8, 308 p-tail = 0

Pillai's trace

trace =	1.47		
Approx.F=	107.376	df = 8, 310	p-tail = 0

Lawley-Hotelling trace

trace =	9.321		
Approx.F=	178.273	df = 8, 306	p-tail = 0

Canonical discriminant functions

	1	2
Constant	3.56	11.684
VAR00002	15.183	1.972
VAR00003	-1.932	-11.595
VAR00004	-3.959	-4.528
VAR00005	-10.273	-1.326

Canonical discriminant functions -- standardized by within variances

	1	2
VAR00002	1.038	0.135
VAR00003	-0.155	-0.928
VAR00004	-0.327	-0.374
VAR00005	-0.91	-0.117

Canonical scores of group means

	1	2
CT-26	-2.356	-1.766
HCT116	-1.204	1.054
HT-29	4.763	-0.341

Analysis layer 2: Discrimination within grouped cancerous/metastatic cell lines (CT-26-F1m CT-26-FL3 and LoVo)

Group Frequencies

CT-26-F1	CT-26-FL3	LoVo
40	40	60

Group means

	CT-26-F1	CT-26-FL3	LoVo
VAR00002	0.693	0.773	0.446
VAR00003	0.419	0.489	0.408
VAR00004	0.911	0.911	0.822

VAR00005	0.849	0.739	0.451

Between groups F-matrix -- df = 4 134

	CT-26-F1	CT-26-FL3	LoVo
CT-26-F1	0		
CT-26-FL3	37.349	0	
LoVo	189.756	106.889	0

Wilks' lambda

Lambda =	0.079	df = 4 2 137	
Approx. F=	85.6963	df = 8 268	prob = 0

Classification functions

	CT-26-F1	CT-26-FL3	LoVo
CONSTANT	-211.591	-202.242	-139.21
VAR00002	-59.623	-36.404	-35.736
VAR00003	197.378	214.085	184.7
VAR00004	227.164	226.566	201.511
VAR00005	203.546	161.453	113.474

Included				**Excluded**		
Variable	F-to-remove	Tolerance	|	Variable	F-to-enter	Tolerance
VAR00002	22.4	0.494287	|			
VAR00003	10.52	0.909455	|			
VAR00004	8.88	0.986348	|			
VAR00005	159.72	0.522001	|			

Classification matrix (cases in row categories classified into columns)

	CT-26-F1	CT-26-FL3	LoVo	**%correct**
CT-26-F1	35	5	0	**88**
CT-26-FL3	2	38	0	**95**
LoVo	0	0	60	**100**
Total	**37**	**43**	**60**	**95**

Jackknifed classification matrix

	CT-26-F1	CT-26-FL3	LoVo	**%correct**
CT-26-F1	33	7	0	**83**
CT-26-FL3	3	37	0	**93**
LoVo	0	1	59	**98**
Total	**36**	**45**	**59**	**92**

Eigenvalues

6.205 0.757

Canonical correlations

0.928 0.656

Cumulative proportion of total dispersion

0.891 1

Wilks' lambda

Lambda =	0.079		
Approx.F=	85.696	df= 8, 268	p-tail = 0

Pillai's trace

trace =	1.292		
Approx.F=	61.602	df= 8, 270	p-tail = 0

Lawley-Hotelling trace

trace =	6.962		
Approx.F=	115.746	df= 8, 266	p-tail = 0

Canonical discriminant functions

	1	2
Constant	13.53	-7.397
VAR00002	3.296	8.046
VAR00003	-3.426	10.19
VAR00004	-4.945	3.482
VAR00005	-15.054	-7.696

Canonical discriminant functions -- standardized by within variances

	1	2
VAR00002	0.384	0.939
VAR00003	-0.179	0.531
VAR00004	-0.332	0.234
VAR00005	-1.177	-0.602

Canonical scores of group means

	1	2
CT-26-F1	-2.898	-0.91
CT-26-FL3	-1.227	1.292
LoVo	2.75	-0.255

References

1. K. L. Bicker, J. Sun, J. J. Lavigne and P. R. Thompson, *ACS Comi. Sci.*, 2011, **13**, 232-243.
2. *GraFit*, Erithacus Software Limited, Version 5.0.11 edn., 2004, pp. Note that when the data were fit to a two-site binding model no significant differences in the calculated Kd values were apparent, Kd1 = 0.47 ± 40.51 μM and Kd42 =43.47 ± 41.40 μM. However, the errors are quite large for this later analysis while the single site model afforded a significantly better fit to the data.
3. N. J. Moerke, *Curr. Protoc. Chem. Biol.*, 2009, **1**, 1-15.
4. , Systat Software, Inc, Version 11.00.01 edn., 2004.
5. , StatSoft, Inc, 6.1 edn., 2003.