

Supplementary Material for:

Characterization of Cell Lines Derived from Breast Cancers and Normal Mammary Tissues for the Study of the Intrinsic Molecular Subtypes

Alex Prat, Olga Karginova, Joel S. Parker, Lisa Bixby, Cheng Fan, Xiaping He, Chuck J. Harrell, Erick Roman, Barbara Adamo, Melissa Troester and Charles M. Perou*

*To whom correspondence should be addressed. E-mail: cperou@med.unc.edu

This PDF file includes:

- **Table S1.** Clinical-pathological characteristics of cell lines grown in 2D.
- **Table S2.** Percentages of different antigenic phenotypes using EpCAM and CD49f in a panel of breast cell lines, hMSCs and hESCs.
- **Figure Legends**
- **Fig. S1.** Summary of the tumor and cell line data sets analyzed in this study.
- **Fig. S2.** Summary of the molecular subtyping approach of cell lines.
- **Fig. S3.** Mean gene expression correlation between identical cell lines from different data sets.
- **Fig. S4.** Identification of breast cancer cell line-specific genes.
- **Fig. S5.** Differentiation scores of the UNC337 data set after removal of the potential tumor stromal genes.
- **Fig. S6.** Correlation between the protein and gene expression of EpCAM and CD49f in the normal breast sorted subpopulations.
- **Fig. S7.** EpCAM/CD49f and CD24/CD44 FACS analyses and Keratin5/Keratin8 and Vimentin IF imaging for each cell line.
- **Fig. S8.** Dual keratin 5/keratin 8 and vimentin IF imaging of HCC38 sorted fractions.

Figure Legends

Figure S1. Summary of the tumor and cell line data sets analyzed in this study. We used one breast sample data set (A) as published in Prat et al. which includes 320 tumor samples and 17 normal breast samples (UNC337, GSE18229). For cell lines, we used one UNC cell line data set (UNC31) that includes (B) UNC plated cell lines, (C) UNC sorted reduction mammaplasties and (D) UNC sorted SUM149PT, HCC1143 and HCC38 cell lines. As independent sets (INDCL) we used previously published cell line data sets of Neve et al., Hollestelle et al., and Kao et al. (E). Molecular subtyping was performed on the combined UNC337-UNC31-INDCL as described in materials and methods section and Figure S2. Throughout the manuscript, different sections of the combined set were used.

Figure S2. Summary of the molecular subtyping approach of cell lines.

Figure S3. Mean gene expression correlation between identical cell lines from different data sets. In red color, the 10 cell lines in common in all four data sets (BT474, BT549, HCC1937, MCF7, MDAMB231, MDAMB436, MDAMB468, SKBR3, SUM149PT and T47D).

Figure S4. Identification of breast cancer cell line-specific genes. (A) Statistically significant differentially expressed genes (up- and down-regulated) between BCCLs and their respective tumor subtype counterpart. On the right, overlap of the four down- and up-regulated cell line-specific gene lists. (B) Supervised hierarchical clustering of BCCLs and tumors using the overlap signatures identified in (A). The enrichment for gene ontology biological processes terms are shown below.

Figure S5. Differentiation scores of the UNC337 data set after removal of the potential tumor stromal genes identified in Figure S4 (downregulated cell line specific-genes).

Figure S6. Correlation between the protein and gene expression of EpCAM and CD49f in the normal breast sorted subpopulations. Cor=correlation.

Figure S7. EpCAM/CD49f and CD24/CD44 FACS analyses and Keratin5/Keratin8 and Vimentin IF imaging for each cell line.

Figure S8. Dual keratin 5/keratin 8 and vimentin IF imaging of HCC38 sorted fractions.

Figure S9. Dual keratin 5/keratin 8 and vimentin IF imaging of HCC1143 sorted fractions.

Figure S10. Dual keratin 5/keratin 8 and vimentin IF imaging of normal breast sorted fractions.

Figure S11. Dual keratin 5/keratin 8 and vimentin IF imaging of SUM149 sorted fractions.

Figure S12. Characterization of cell lines. Unsupervised hierarchical clustering of all cell lines and sorted fractions using the most variable genes (n=17,077). RMF, reduction mammaplasty fibroblasts; MSC, mesenchymal stem cells; hESC, human embryonic stem cells; HMEC, human mammary epithelial cells.

Table S1. Clinical-pathological characteristics of cell lines grown in 2D.

Cell Line	Molecular Subtype	ER status	HER2 status	Patient Age	Ethnicity (W=white; B=black; A=asian; WH=white hispanic)	Tumor type (IBC=inflammatory breast cancer; ANC=anaplastic carcinoma; DC=suctal carcinoma; IDC=invasive ductal carcinoma; AC=adenocarcinoma; pap=papillar; ILC=invasive lobular carcinoma; MC=metaplastic carcinoma)	Source (M=Metastasis; P=primary; N=normal breast; F=fibrocystic disease)	Culture Media
BT20	Basal-like	-	-	74	W	IDC	P	DMEM+10%FBS
Cal51	Basal-like	-	-	45	-	AC	M	RPMI + 10%FBS
DU4475	Basal-like	-	-	70	W	IDC	M	RPMI + 10%FBS
HCC1008	Basal-like	-	+	67	B	DC	M	RPMI + 10%FBS
HCC1143	Basal-like	-	-	52	W	DC	P	RPMI + 10%FBS
HCC1187	Basal-like	-	-	41	W	DC	P	RPMI + 10%FBS
HCC1569	Basal-like	-	+	70	B	MC	M	RPMI + 10%FBS
HCC1599	Basal-like	-	-	44	W	DC	P	RPMI + 10%FBS
HCC2157	Basal-like	-	-	48	B	DC	P	RPMI + 10%FBS
HCC3153	Basal-like	-	-	-	-	-	P	RPMI + 10%FBS
HCC70	Basal-like	-	-	49	B	DC	P	RPMI + 10%FBS
HMECB2	Basal-like	-	-	-	-	-	N	HUMEC
HMECBL2	Basal-like	-	-	-	-	-	N	HUMEC
HMECBW	Basal-like	-	-	-	-	-	N	HUMEC
HMECBX	Basal-like	-	-	-	-	-	N	HUMEC
HMLE	Basal-like	-	-	-	-	-	N	HUMEC
MCF12A	Basal-like	-	-	60	W	-	F	DMEM/F12 + Supl
SUM229PE	Basal-like	-	-	-	-	AC	M	HAM'S-F12 + 5%FBS+Supl
SUM102PT	Basal-like/Claudin-low ¹	-	-	-	-	IDC	M	HUMEC
SUM149PT	Basal-like/Claudin-low ¹	-	-	-	-	IBC ¹	P	HAM'S-F12 + 5%FBS+Supl ³
HCC1937	Basal-like/HER2-enriched ¹	-	-	23	W	IDC	P	RPMI + 10%FBS
MDAMB468	Basal-like/HER2-enriched ¹	-	-	51	B	AC	M	Leibovitz's L-15+10%FBS
184A1	Claudin-low	-	-	21	-	-	N	MEBM + Suppl
BABE	Claudin-low	-	-	-	-	-	N	HUMEC
BT549	Claudin-low	-	-	72	W	IDC, pap	P	RPMI + 10%FBS
HBL100	Claudin-low	-	-	27	-	-	N	DMEM+10%FBS
HCC1395	Claudin-low	-	-	43	W	IDC	P	RPMI + 10%FBS
HCC38	Claudin-low	-	-	50	W	DC	P	RPMI + 10%FBS
HMECA1	Claudin-low	-	-	-	-	-	N	HUMEC
HMLE_TWIST1	Claudin-low	-	-	-	-	-	N	HUMEC
HMLE_SNAIL1	Claudin-low	-	-	-	-	-	N	HUMEC
HS578T	Claudin-low	-	-	74	W	MC	P	DMEM+10%FBS
HTERT	Claudin-low	-	-	-	-	-	N	RPMI + 10%FBS
MCF10A	Claudin-low	-	-	36	W	-	F	DMEM/F12 + Supl
MDAMB157	Claudin-low	-	-	44	B	MDC	M	DMEM+10%FBS
MDAMB231	Claudin-low	-	-	51	W	AC	M	DMEM+10%FBS
MDAMB435	Claudin-low	-	-	31	W	IDC	M	DMEM+10%FBS
MDAMB436	Claudin-low	-	-	43	W	IDC	M	Leibovitz's L-15+10%FBS
ME16C	Claudin-low	-	-	-	-	-	N	HUMEC
HMF1	Claudin-low	-	-	-	-	-	N	DMEM/F12 + 10%FBS
HMF2	Claudin-low	-	-	-	-	-	N	DMEM/F12 + 10%FBS
HMF3	Claudin-low	-	-	-	-	-	N	DMEM/F12 + 10%FBS
HMF1714	Claudin-low	-	-	-	-	-	N	-
HMF1299	Claudin-low	-	-	-	-	-	N	-
HMF1632	Claudin-low	-	-	-	-	-	N	-
HMF1394	Claudin-low	-	-	-	-	-	N	-
SKBR7	Claudin-low	-	-	-	-	-	M	RPMI + 10%FBS
SUM1315	Claudin-low	-	-	-	-	IDC	M	HAM'S-F12 + 5%FBS+Supl ²
SUM159PT	Claudin-low	-	-	-	-	ANC	P	HAM'S-F12 + 5%FBS+Supl ³
AU565	HER2-enriched	-	+	43	W	AC	M	RPMI + 10%FBS
HCC1954	HER2-enriched	-	+	61	A	DC	P	RPMI + 10%FBS
HCC202	HER2-enriched	-	+	82	W	DC	P	RPMI + 10%FBS
HCC2185	HER2-enriched	-	-	49	WH	LC	M	RPMI + 10%FBS
HCC2218	HER2-enriched	-	+	38	W	IDC	P	RPMI + 10%FBS
HCC2688	HER2-enriched	-	-	-	-	-	P	RPMI + 10%FBS
SKBR3	HER2-enriched	-	+	43	W	AC	M	RPMI + 10%FBS
SKBR5	HER2-enriched	-	+	-	-	-	M	HAM'S-F12 + 10%FBS
SUM190PT	HER2-enriched	-	+	-	-	IBC	P	HAM'S-F12 + 5%FBS+Supl ³
SUM225CWN	HER2-enriched	-	+	-	-	IDC	M	HAM'S-F12 + 5%FBS+Supl ³
UACC893	HER2-enriched	-	+	57	W	IDC	P	RPMI + 10%FBS
BT474	HER2-enriched/Luminal B ¹	+	+	60	W	IDC	P	RPMI + 10%FBS
600MPE	Luminal B	+	-	-	-	IDC	M	DMEM+10%FBS
BT483	Luminal B	+	-	23	W	IDC, pap	P	RPMI + 10%FBS
CAMA1	Luminal B	+	-	51	W	AC	M	DMEM+10%FBS
EVSAT	Luminal B	-	-	-	-	-	M	RPMI + 10%FBS
HCC1007	Luminal B	+	+ ⁵	67	B	DC	P	RPMI + 10%FBS
HCC1419	Luminal B	-	+	42	W	IDC	P	RPMI + 10%FBS
HCC1428	Luminal B	+	-	49	W	AC	M	RPMI + 10%FBS
HCC712	Luminal B	+	-	-	-	-	P	RPMI + 10%FBS
LY2	Luminal B	+	-	69	W	IDC	M	DMEM+10%FBS
MCF7	Luminal B	+	-	69	W	IDC	M	RPMI + 10%FBS
MDAMB134	Luminal B	+	-	47	W	IDC	M	DMEM+10%FBS
MDAMB175	Luminal B	+	-	56	B	IDC	M	DMEM+10%FBS
MDAMB330	Luminal B	+	+	43	W	-	M	Leibovitz's L-15+20%FBS
MDAMB361	Luminal B	+	+	40	W	AC	p	Leibovitz's L-15+20%FBS
MDAMB415	Luminal B	+	-	38	W	AC	M	DMEM+10%FBS
MDAMB453	Luminal B	-	-	48	W	AC	M	DMEM+10%FBS
OCUBF	Luminal B	-	+	53	A	-	M	DMEM+10%FBS+0.5mM SP ⁴
SUM185PE	Luminal B	-	-	-	-	DC	M	HAM'S-F12 + 5%FBS+Supl ³
SUM44PE	Luminal B	+	-	-	-	AC	M	HAM'S-F12 + 5%FBS+Supl ³
SUM52PE	Luminal B	+	-	-	-	AC	M	HAM'S-F12 + 5%FBS+Supl ³
T47D	Luminal B	+	-	54	-	IDC	M	RPMI + 10%FBS
ZR751	Luminal B	+	-	63	W	IDC	M	RPMI + 10%FBS
ZR7530	Luminal B	+	+	47	B	IDC	M	RPMI + 10%FBS
ZR75B	Luminal B	+	-	-	-	-	M	RPMI + 10%FBS
HCC1500	Luminal B/Basal-like ¹	+	-	23	B	IDC	P	RPMI + 10%FBS
UACC812	Luminal B/HER2-enriched ¹	+	+	43	-	IDC	P	RPMI + 10%FBS

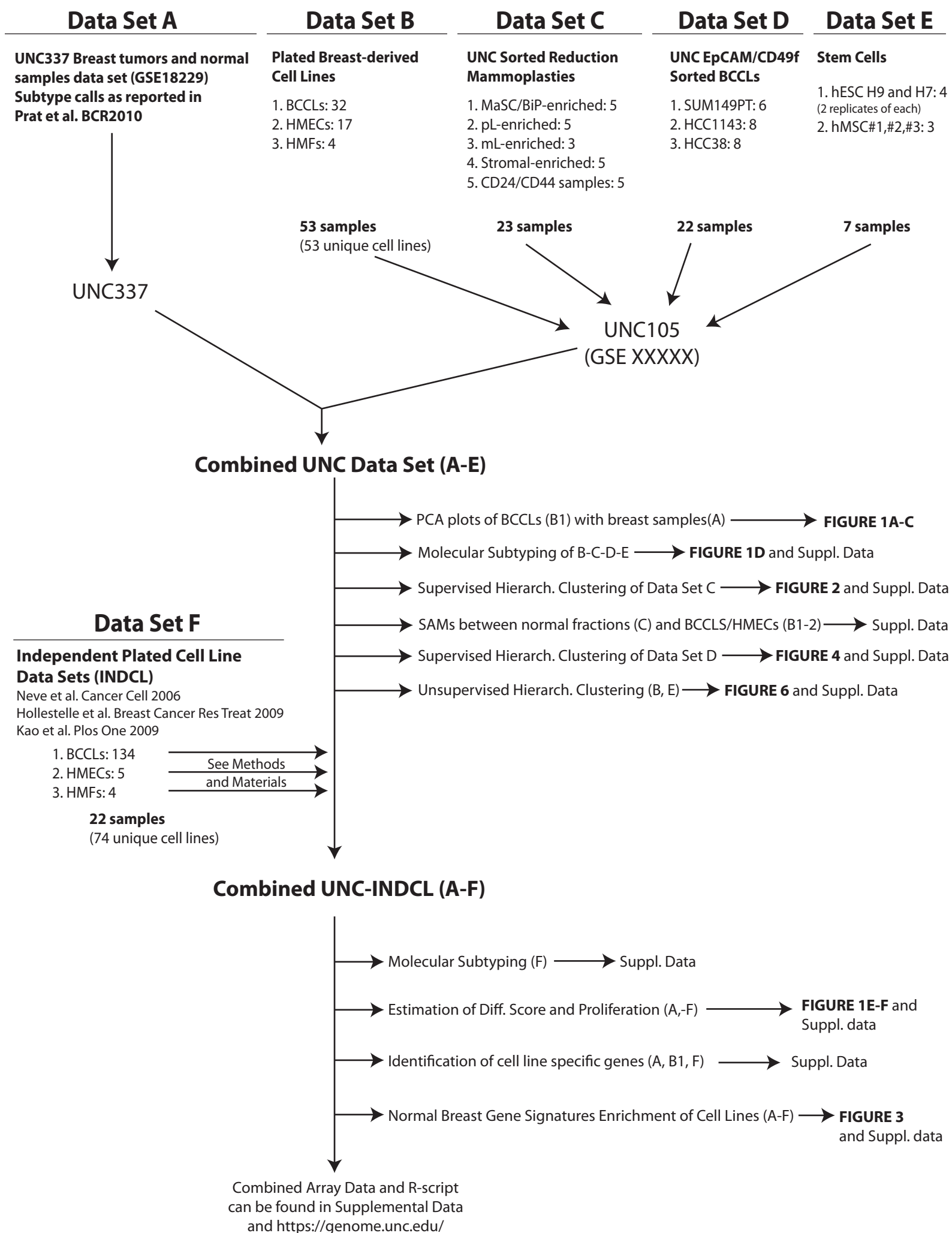
*1, Cell lines whose molecular subtype differed in at least one of the data sets evaluated (see Supplemental Material). *2, Insulin and epidermal growth factor; *3, Insulin and hydrocortisone; *4, sodium pyruvate; *5, This cell line has been identified as HER2+ in Kao et al., and HER2- in Neve et al.

Table S2. Percentages of different antigenic phenotypes using EpCAM and CD49f in a panel fo breast-derived cell lines, hMSC and hESC.

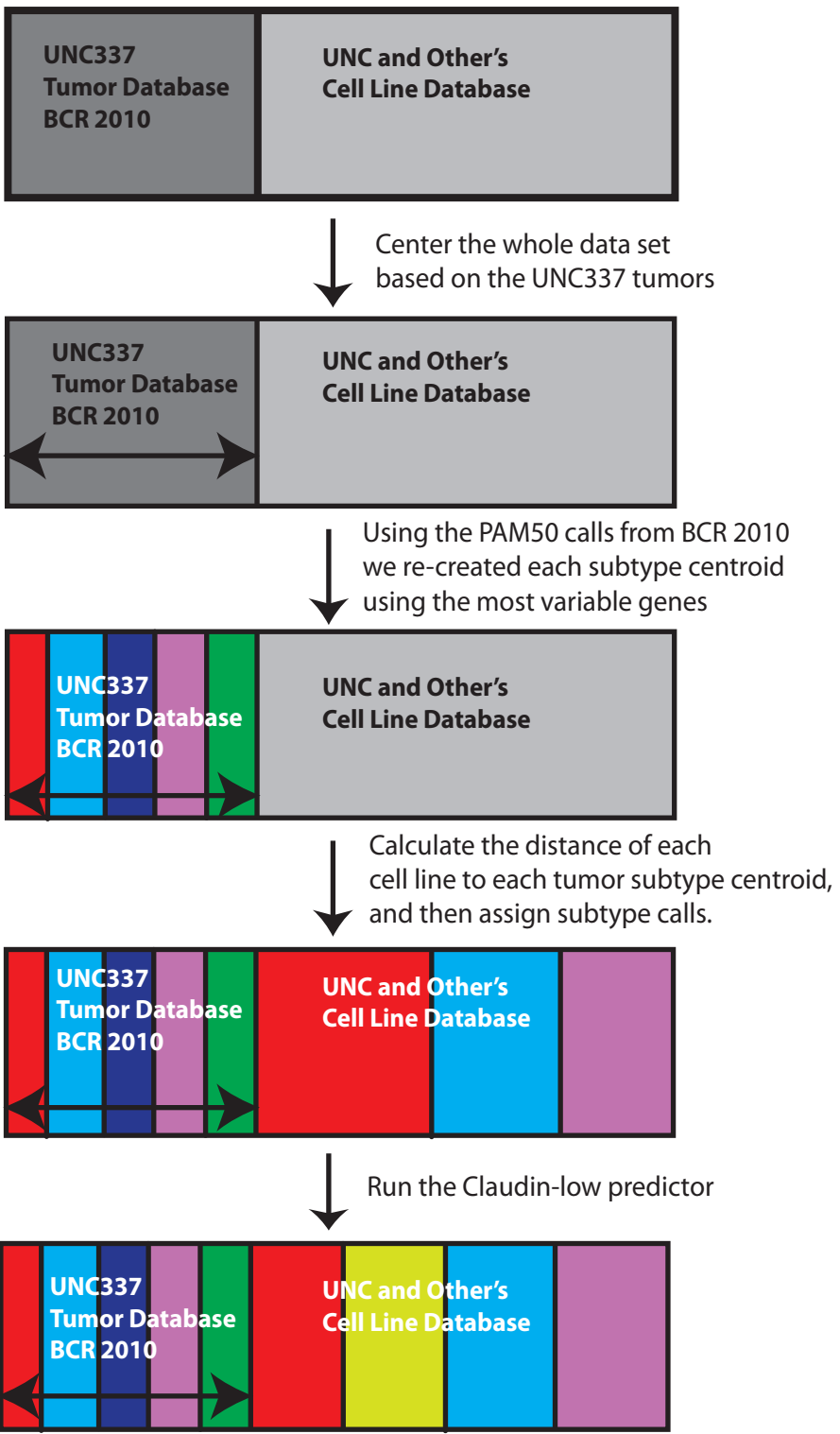
	EpCAM- CD49f- Stromal	EpCAM-/low CD49f+ Stromal/MaSC	EpCAM-/low CD49fhigh MaSC	EpCAMhigh CD49f+ pL	EpCAMhigh CD49f-/low mL
Basal-like					
HCC1143*			40%	60%	
HCC1187					96%
HCC1937				99%	
MDAMB468				100%	
SUM149PT*		10%			85%
HER2-enriched					
BT474	3%				91%
SKBR3	2%				97%
Luminal B					
HCC1428	3%				98%
MCF7	6%				94%
T47D				98%	
ZR751				98%	
HCC1500				99%	
Claudin-low					
BT549			92%		
SUM102PT			97%		
SUM1315			97%		
MDAMB231			97%		
MDA-MB436			97%		
HCC1395			98%		
SUM159PT			98%		
HS578T		97%			
HCC38*			79%	18%	
HMEC					
BABE			100%		
HMECBL			95%		
HMECBW			97%		
HMECBX			98%		
HMLE			98%		
ME16C			98%		
HMF					
HMF1		98%			
hMSC					
hMSC#1		98%			
hMSC#2		98%			
hMSC#3		98%			
hESC					
hESC H9				98%	
hESC H7				98%	

*, Cell lines with two clear distinct subpopulations.

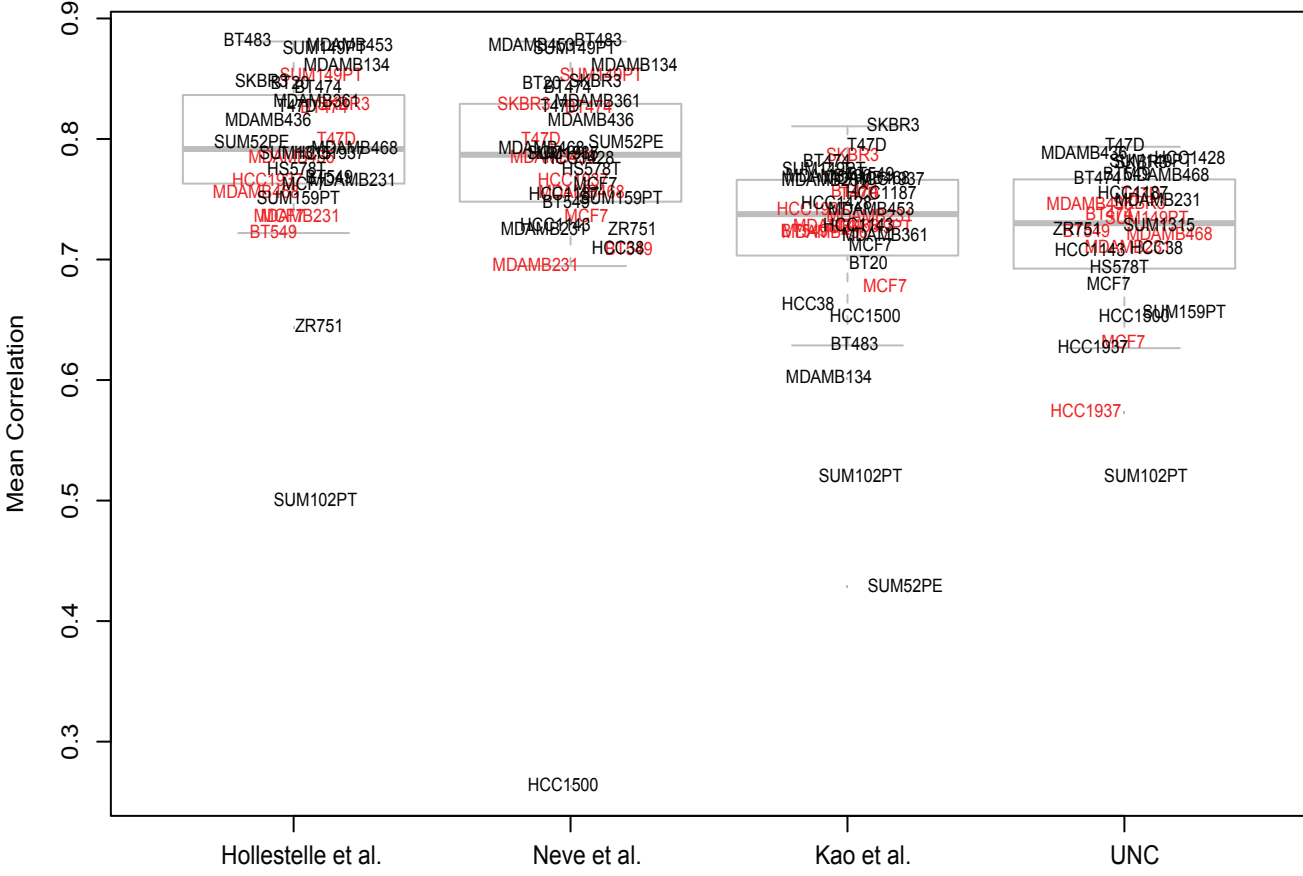
Supplemental Figure 1



Supplemental Figure 2

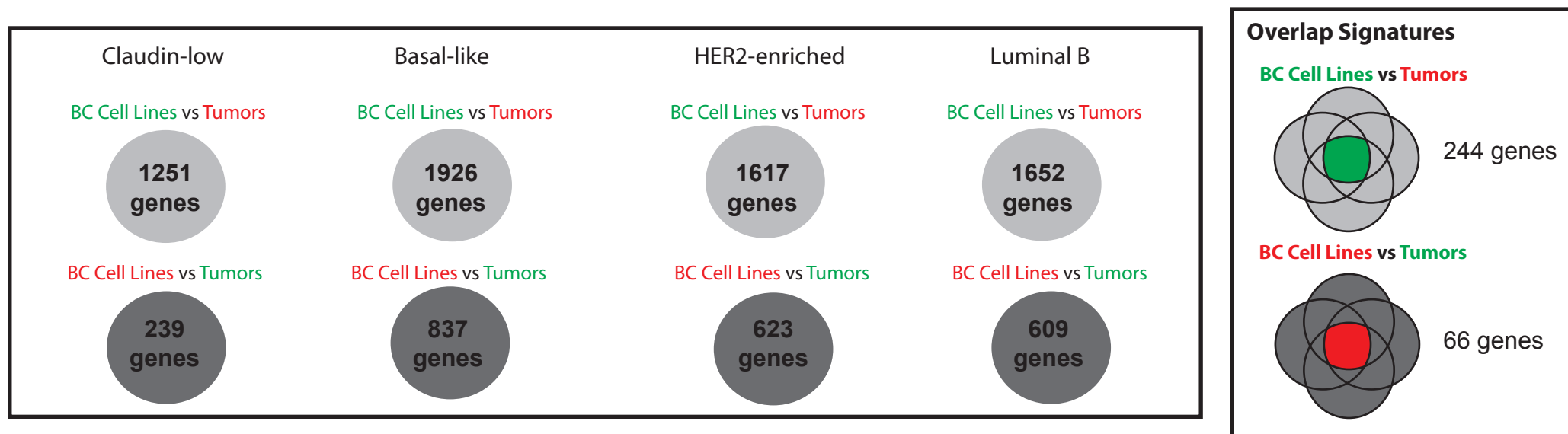


Supplemental Figure 3



Supplemental Figure 4

A



B

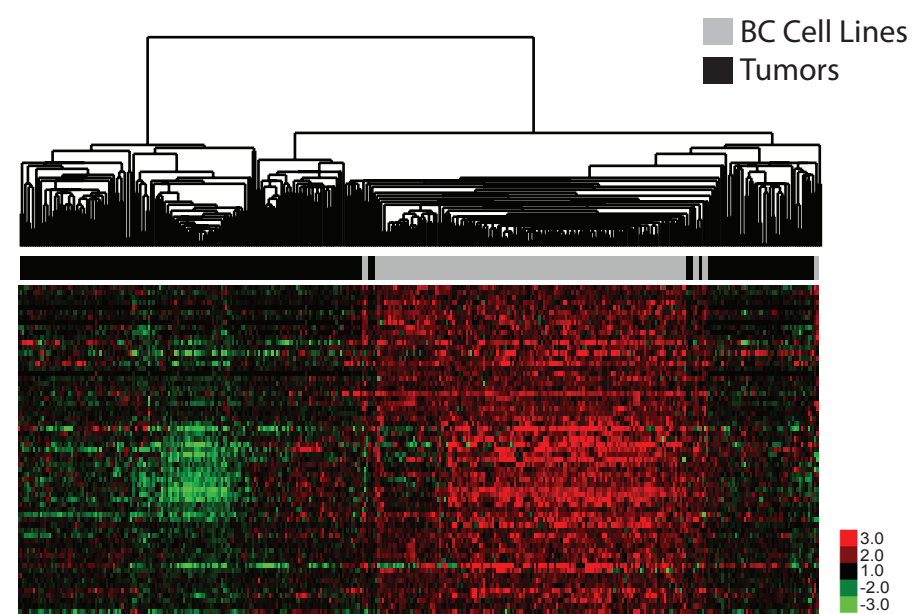
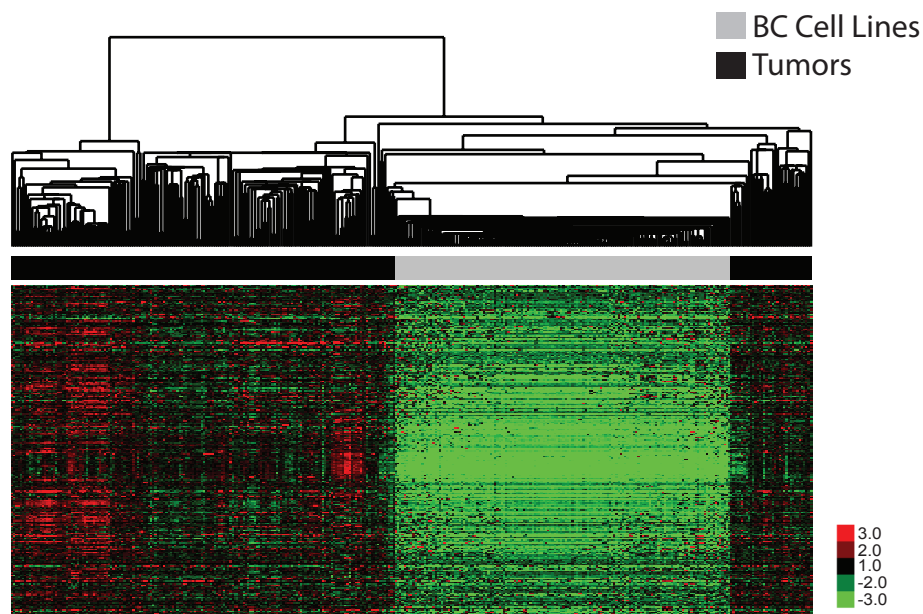


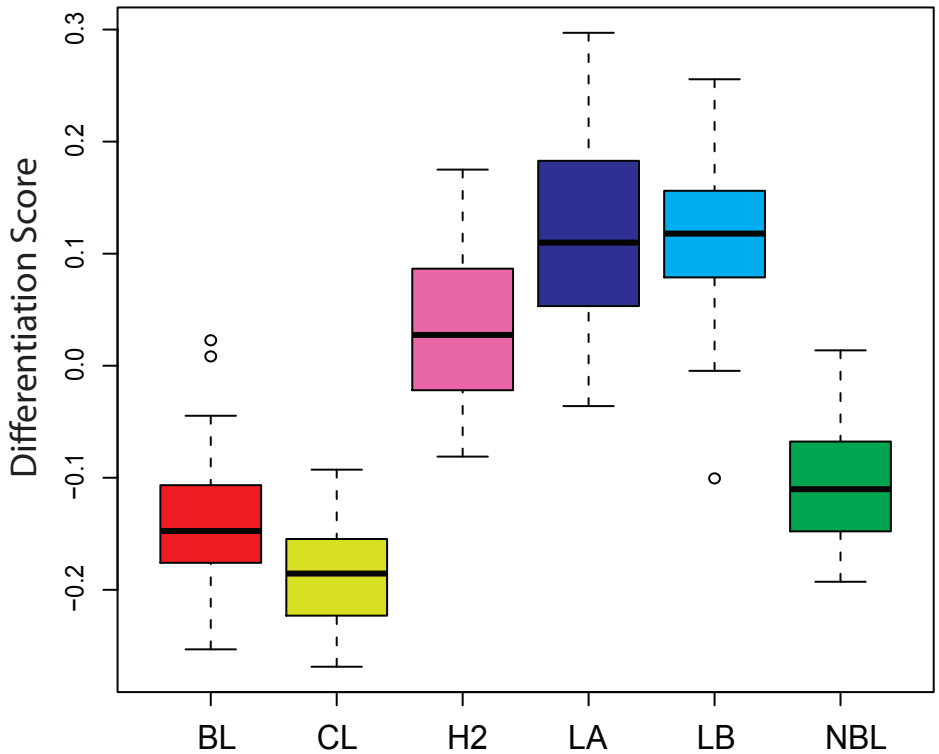
Table. GO Biological Processes Enriched in the Downregulated Cell Lines Gene List

Term	Count	%	PValue
GO:0009611~response to wounding	34	14	3.21E-12
GO:0009605~response to external stimulus	43	18	5.37E-11
GO:0006950~response to stress	54	22	9.46E-08
GO:0006954~inflammatory response	21	9	9.78E-08
GO:0048583~regulation of response to stimulus	25	10	1.35E-07
GO:0065008~regulation of biological quality	49	20	1.47E-07
GO:0048856~anatomical structure development	70	29	1.60E-07
GO:0032502~developmental process	81	33	2.21E-07
GO:0042221~response to chemical stimulus	44	18	3.70E-07
GO:0007275~multicellular organismal development	75	31	4.40E-07
GO:0048731~system development	65	27	4.54E-07
GO:0006952~defense response	28	11	5.48E-07
GO:0002682~regulation of immune system process	21	9	1.44E-06
GO:0048513~organ development	52	21	1.50E-06
GO:0050776~regulation of immune response	16	7	1.63E-06

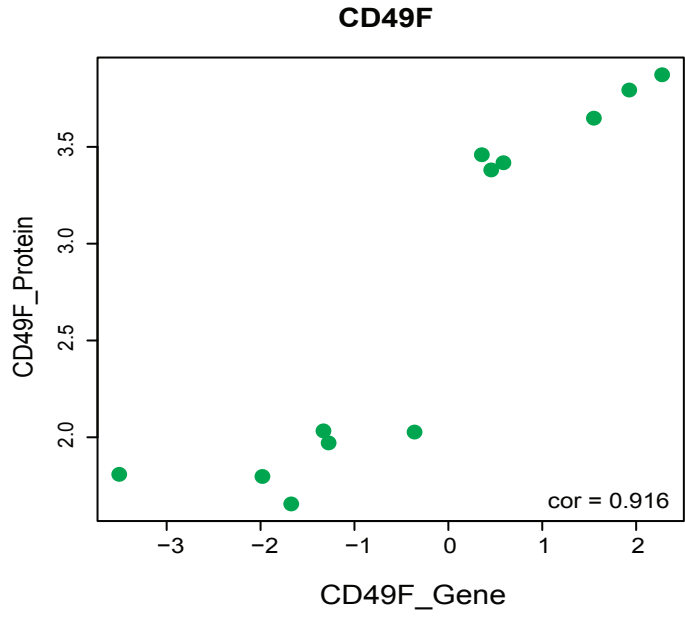
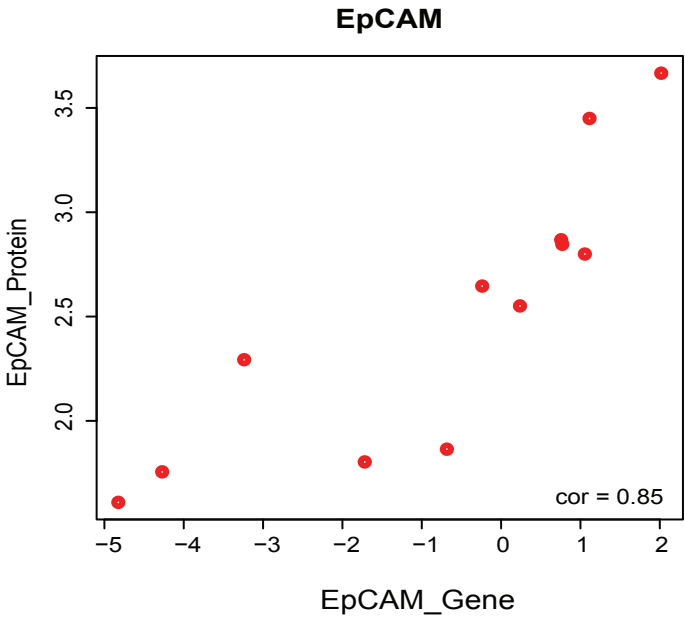
Table. GO Biological Processes Enriched in the Upregulated Cell Lines Gene List

Term	Count	%	PValue
GO:0019222~regulation of metabolic process	24	36	1.45E-03
GO:0031323~regulation of cellular metabolic process	23	35	2.00E-03
GO:0080090~regulation of primary metabolic process	22	33	2.56E-03
GO:0060255~regulation of macromolecule metabolic process	21	32	5.53E-03
GO:0045449~regulation of transcription	18	27	6.29E-03
GO:0010556~regulation of macromolecule biosynthetic process	19	29	6.34E-03
GO:0010468~regulation of gene expression	19	29	6.99E-03
GO:0031326~regulation of cellular biosynthetic process	19	29	9.75E-03
GO:0009889~regulation of biosynthetic process	19	29	1.04E-02
GO:0019219~regulation of nucleobase, nucleoside, nucleotide	18	27	1.38E-02
GO:0051171~regulation of nitrogen compound metabolic process	18	27	1.51E-02
GO:0009987~cellular process	45	68	1.70E-02
GO:0051049~regulation of transport	6	9	1.84E-02

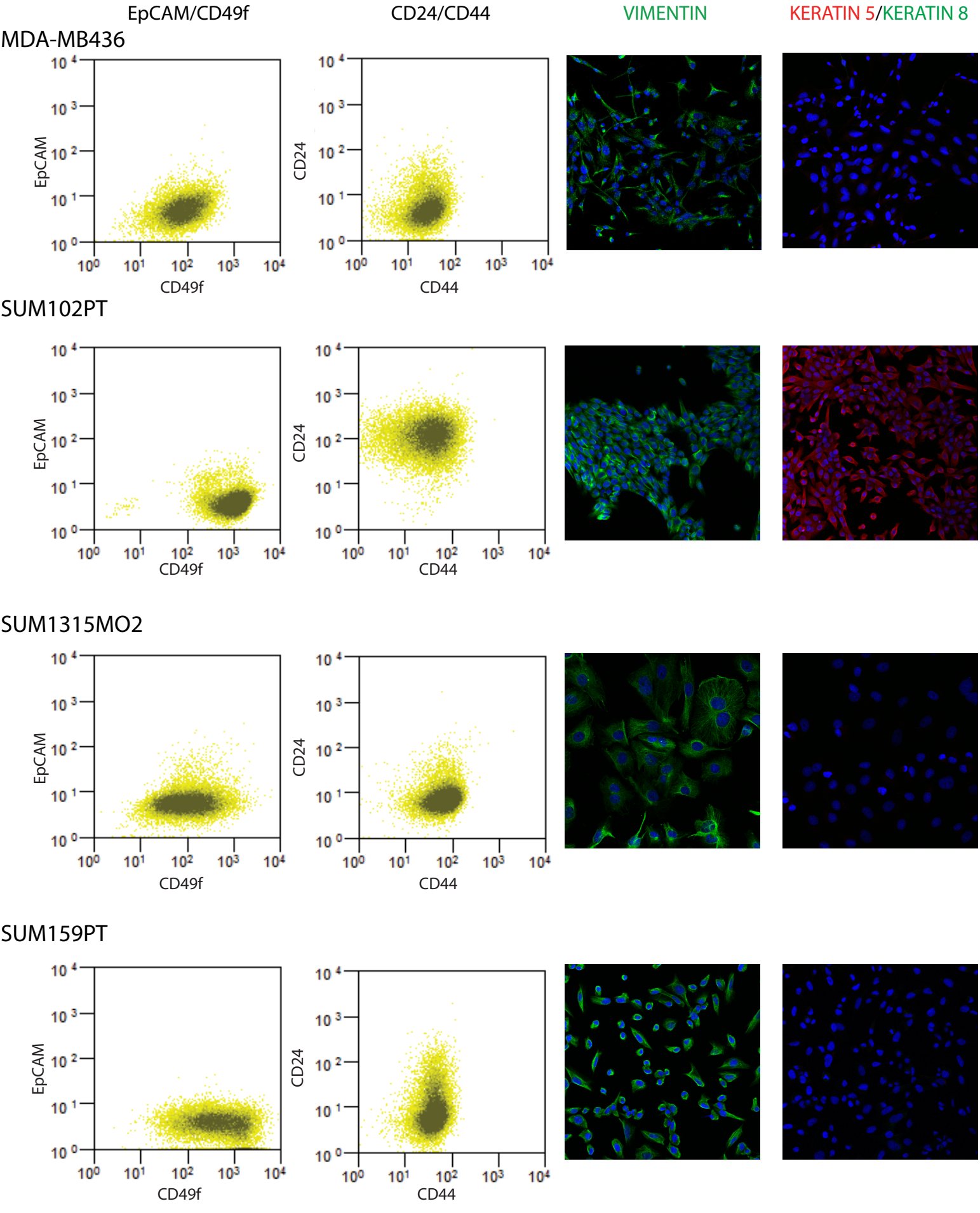
Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7



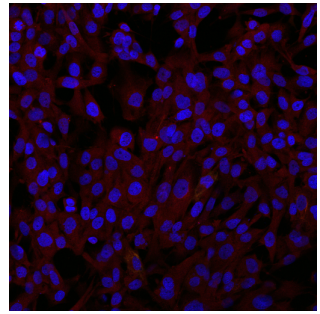
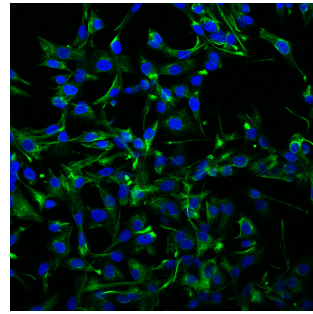
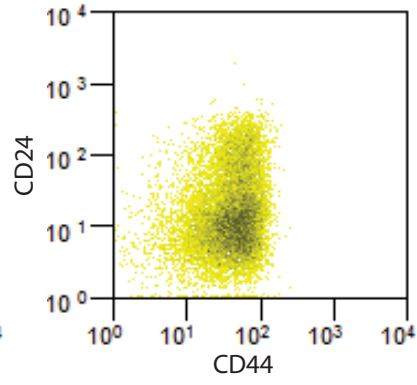
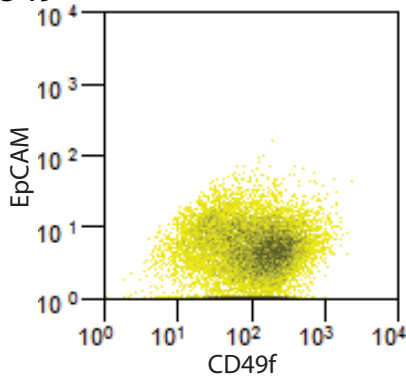
EpCAM/CD49f

CD24/CD44

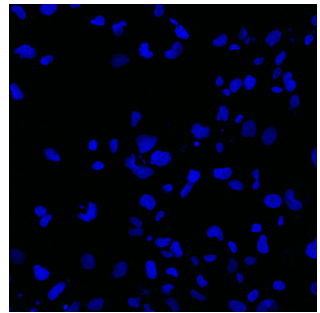
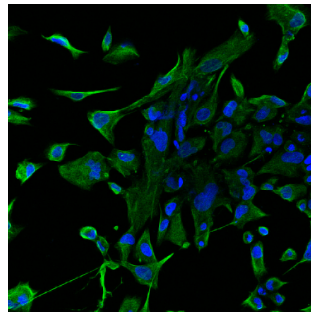
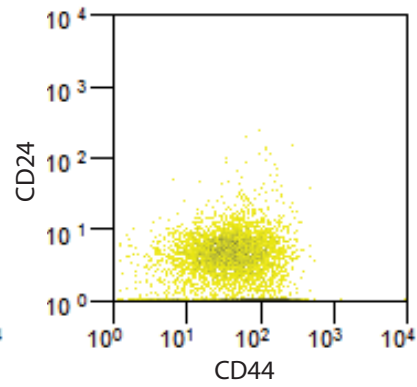
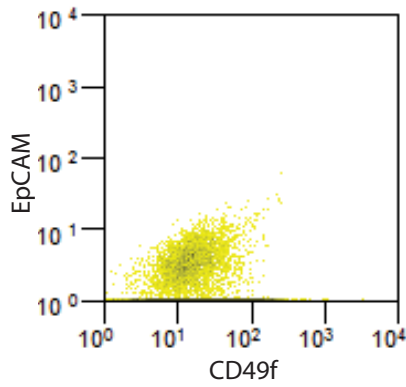
VIMENTIN

KERATIN 5/KERATIN 8

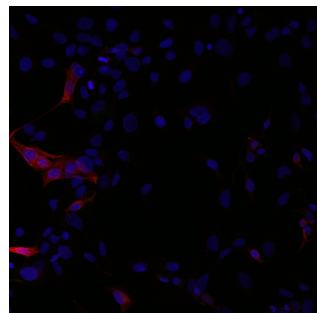
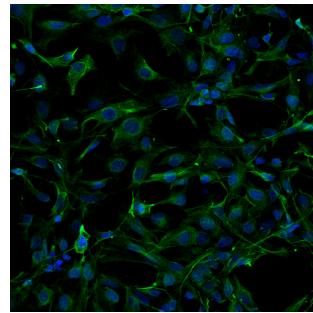
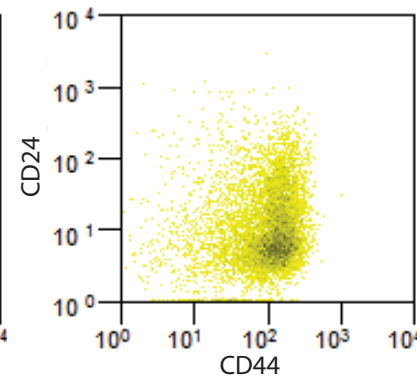
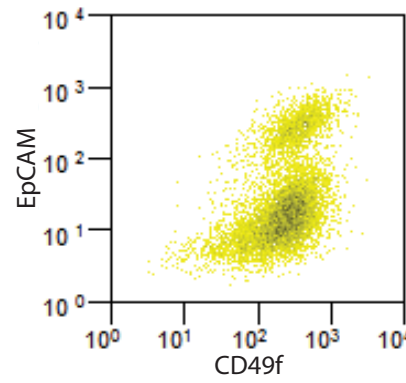
BT549



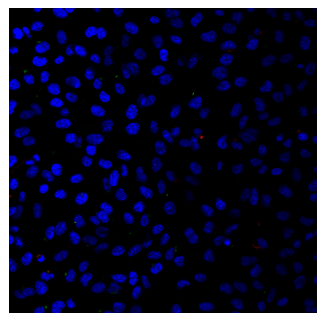
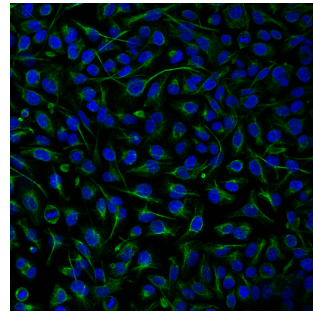
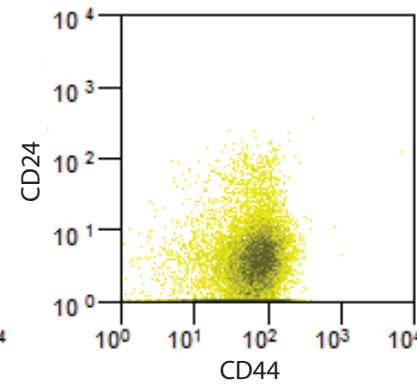
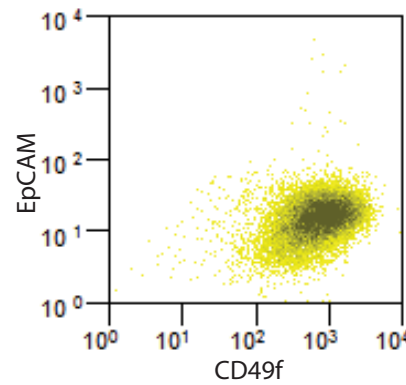
Hs578T



HCC38



MDA-MB231



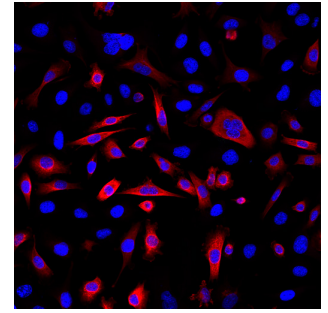
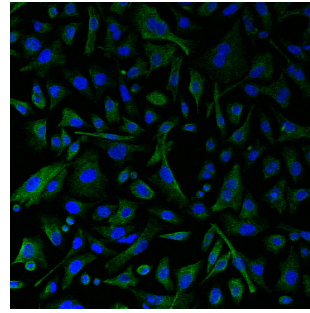
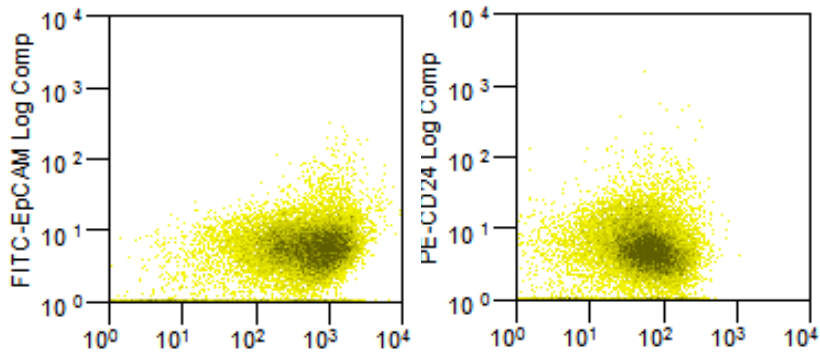
EpCAM/CD49f

CD24/CD44

VIMENTIN

KERATIN 5/KERATIN 8

HCC1395



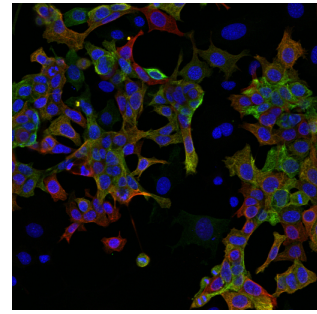
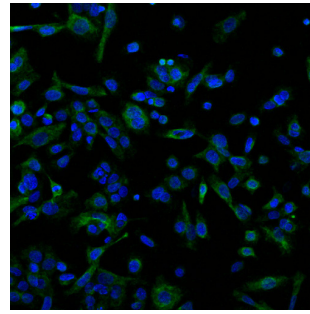
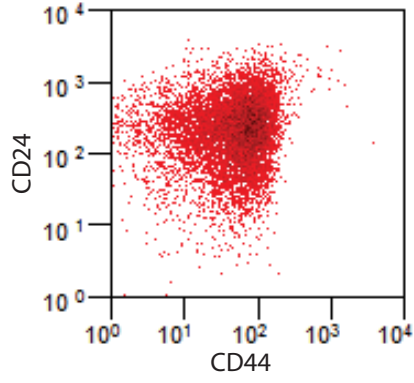
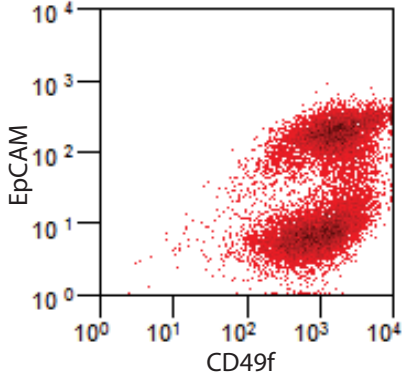
EpCAM/CD49f

CD24/CD44

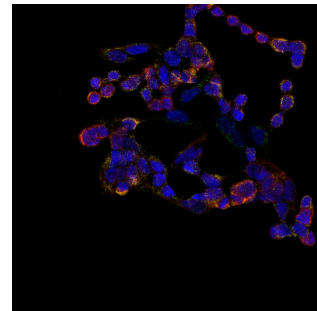
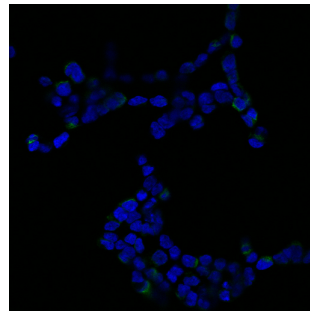
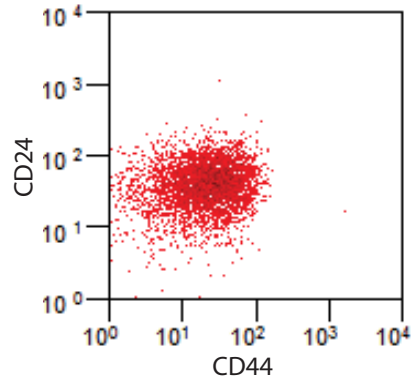
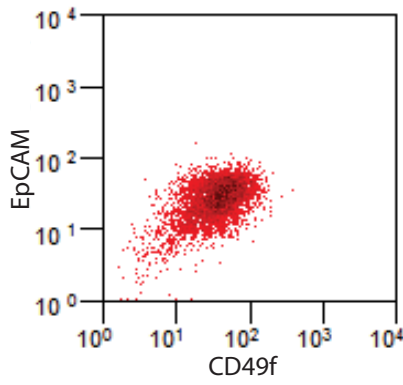
VIMENTIN

KERATIN 5/KERATIN 8

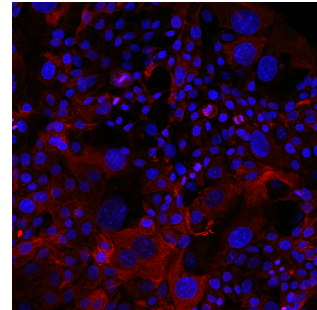
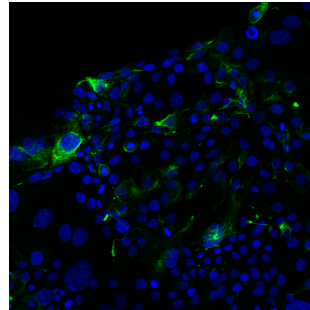
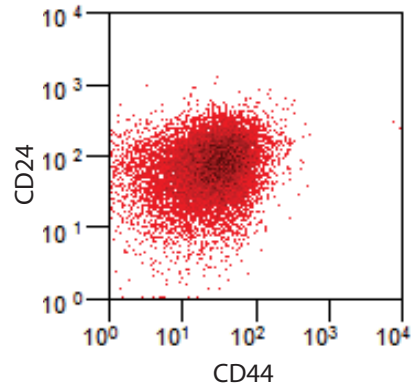
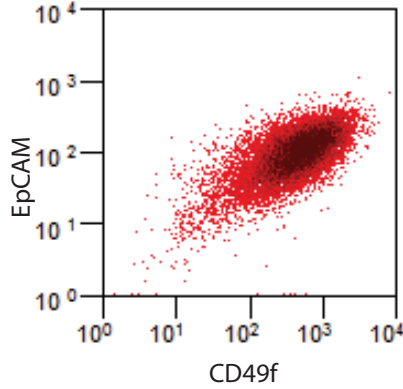
HCC1143



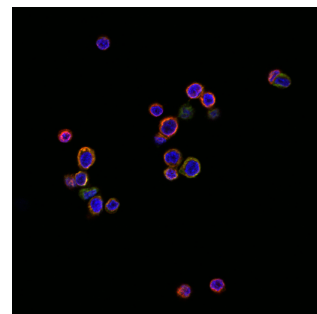
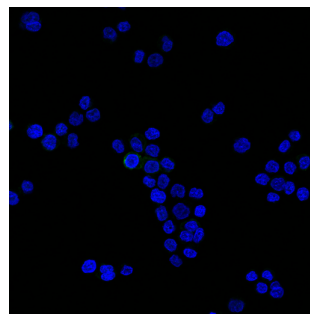
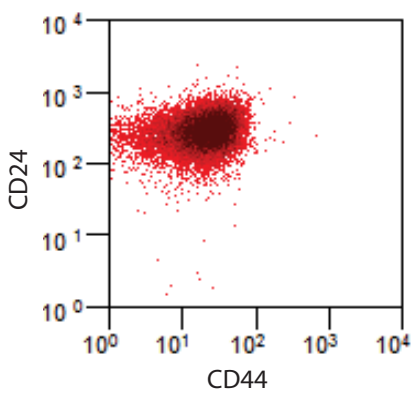
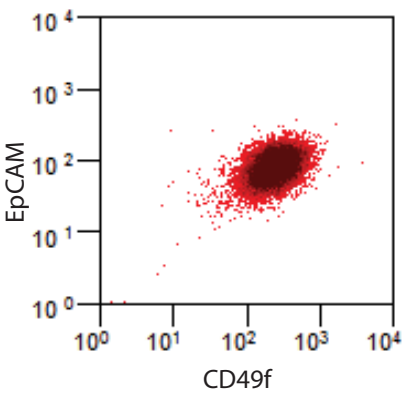
HCC1187



HCC1937



MDA-MB468



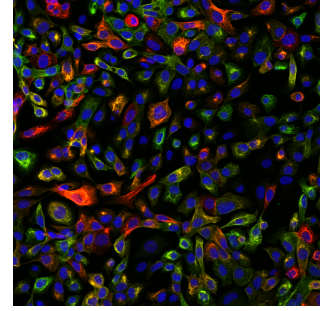
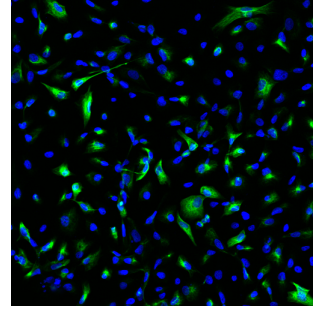
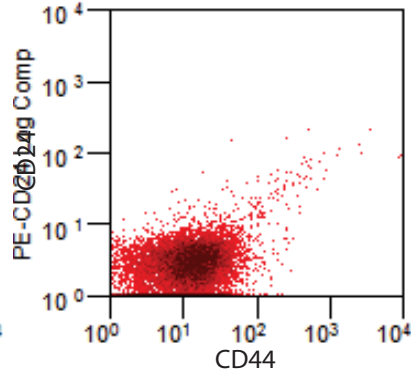
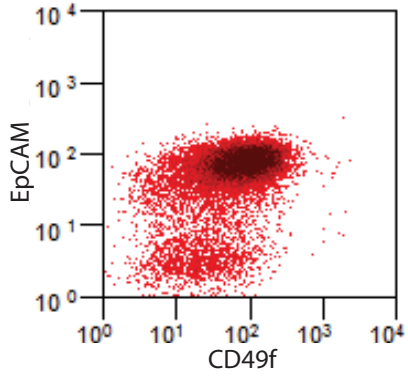
EpCAM/CD49f

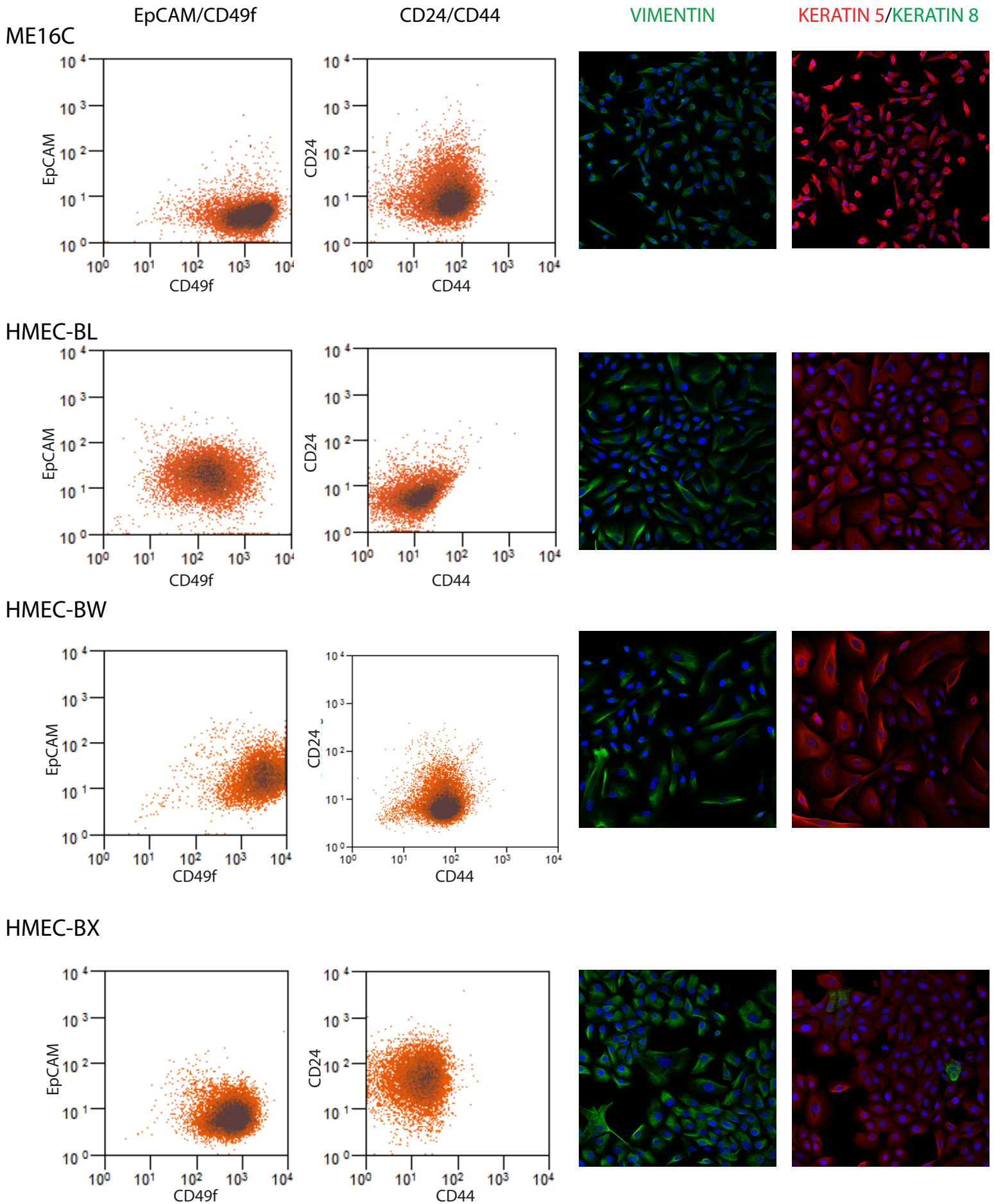
CD24/CD44

VIMENTIN

KERATIN 5/KERATIN 8

SUM149PT





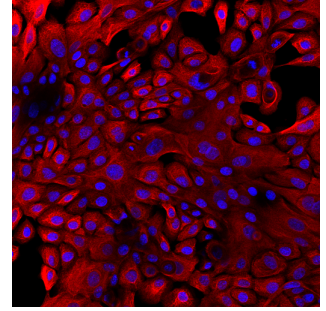
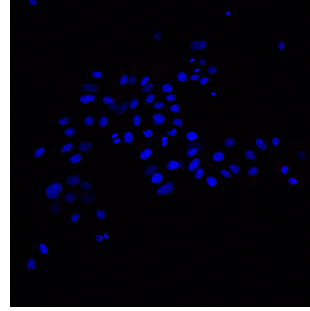
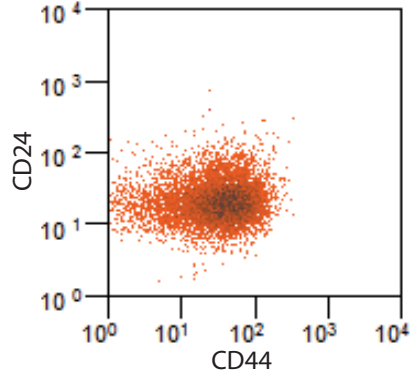
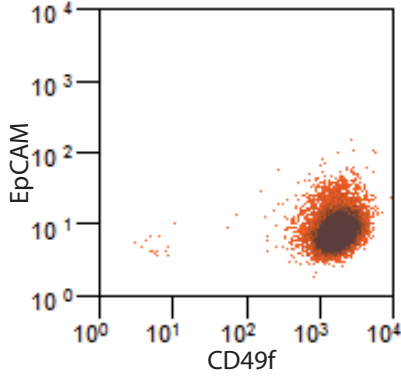
EpCAM/CD49f

CD24/CD44

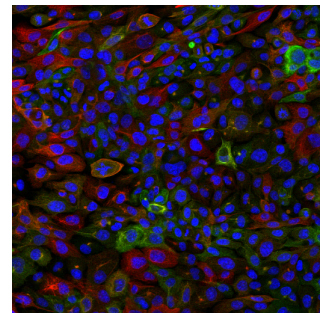
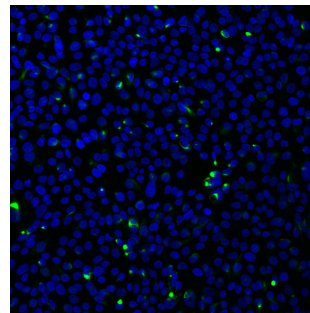
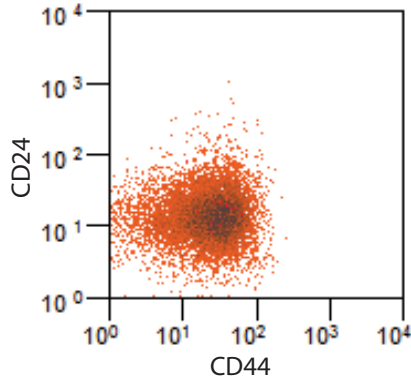
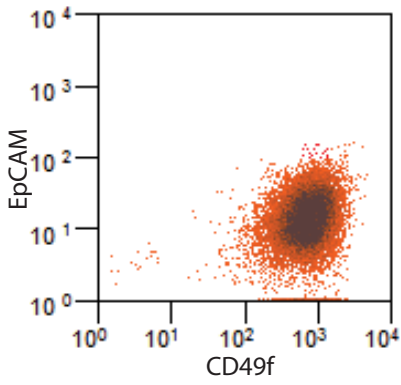
VIMENTIN

KERATIN 5/KERATIN 8

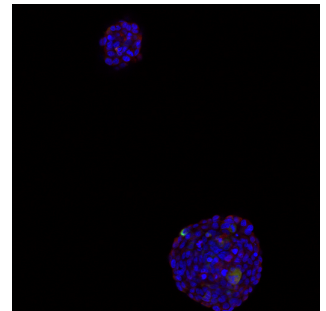
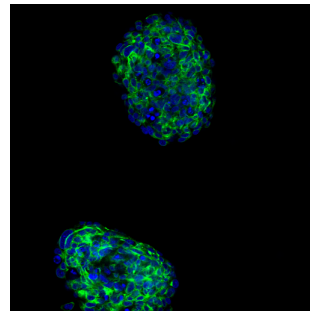
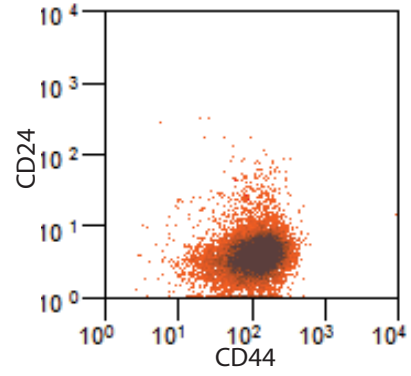
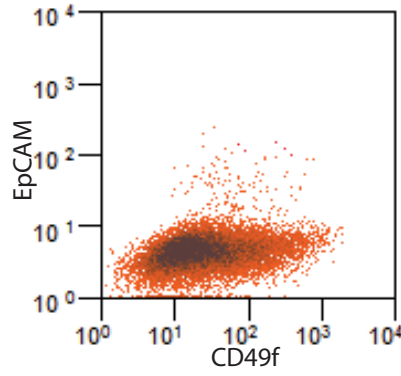
BABE



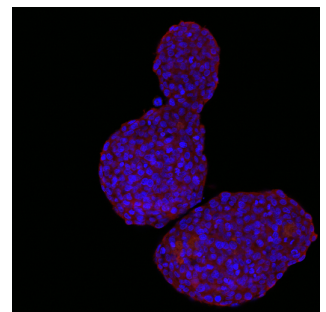
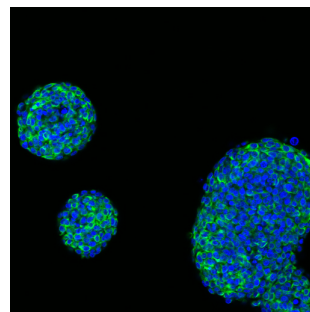
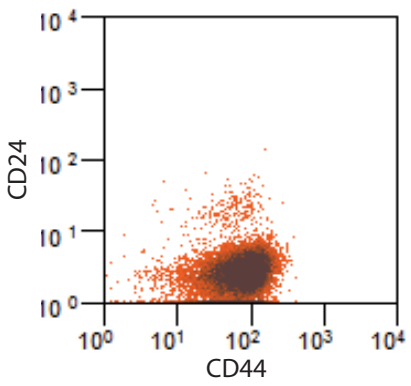
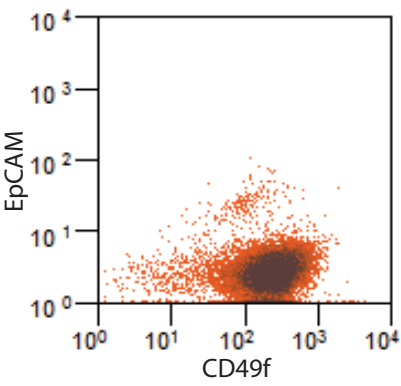
HMLE



HMLE-SNAIL



HMLE-TWIST1



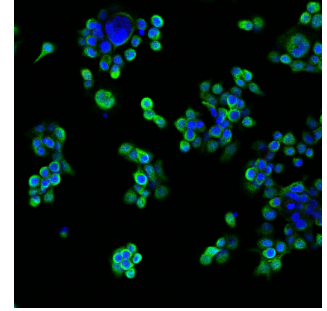
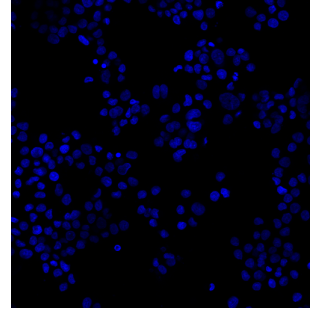
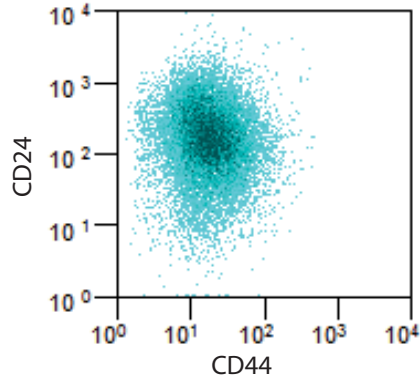
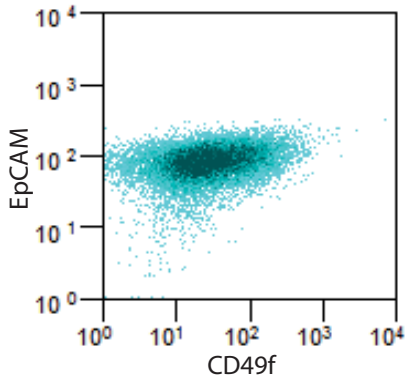
EpCAM/CD49f

CD24/CD44

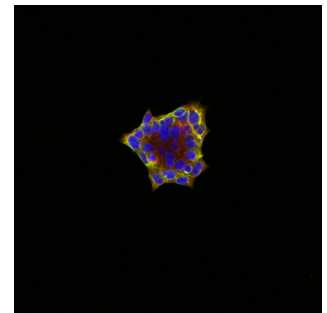
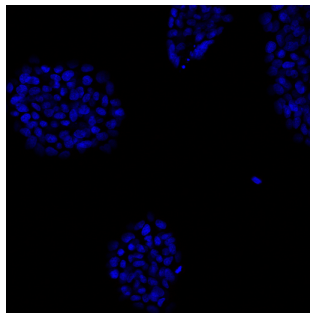
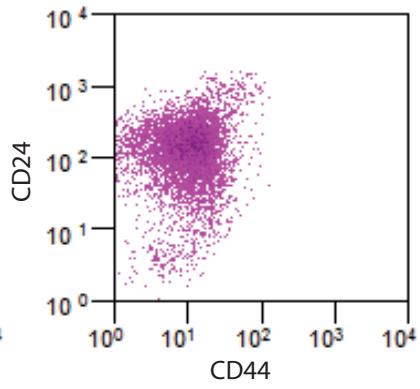
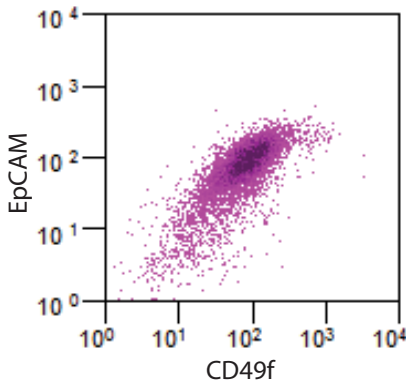
VIMENTIN

KERATIN 5/KERATIN 8

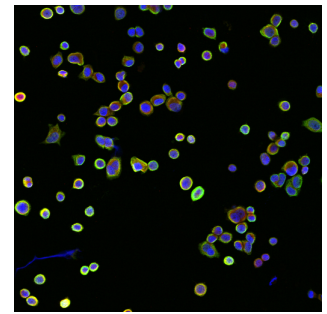
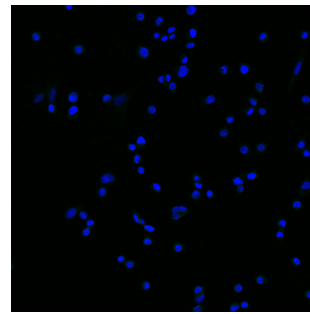
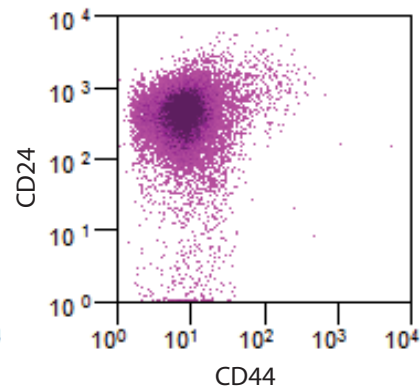
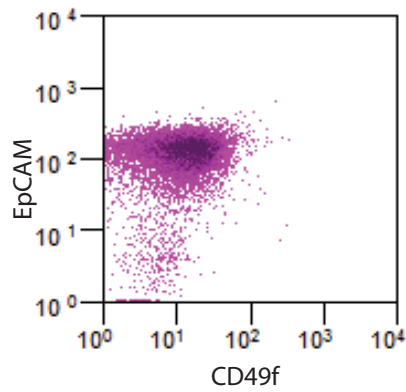
HCC1428



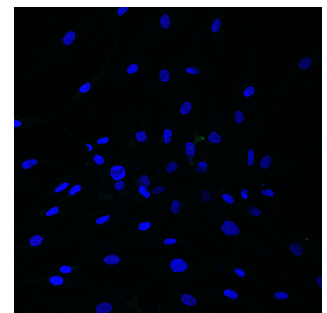
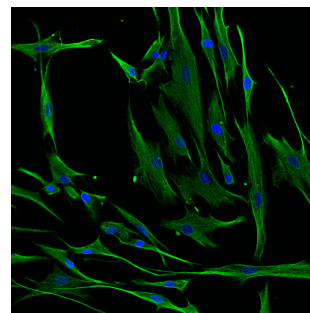
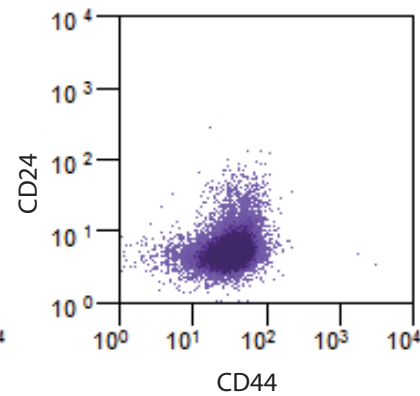
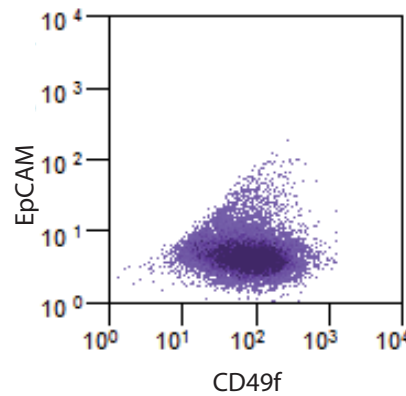
BT474



SKBR3



HMF1



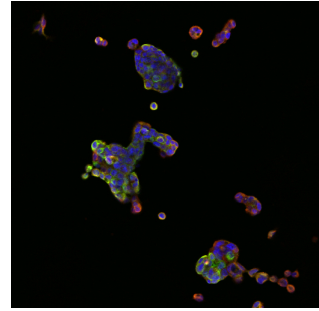
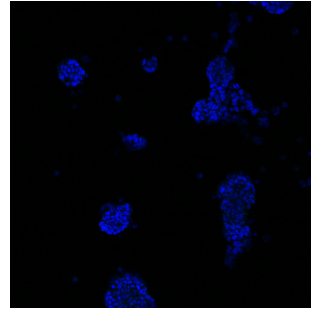
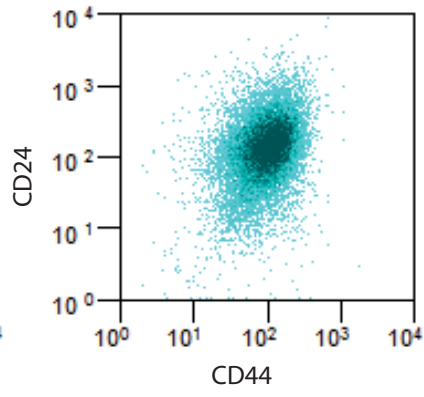
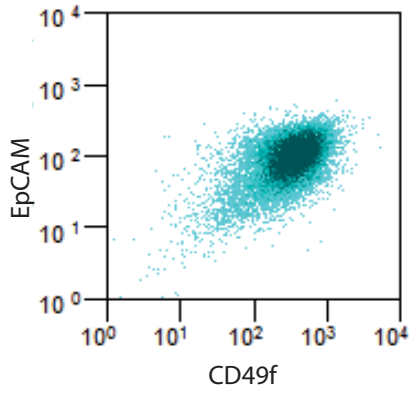
EpCAM/CD49f

CD24/CD44

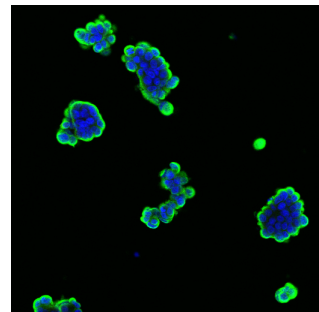
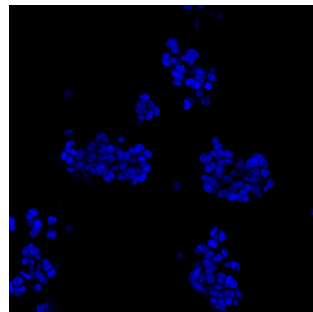
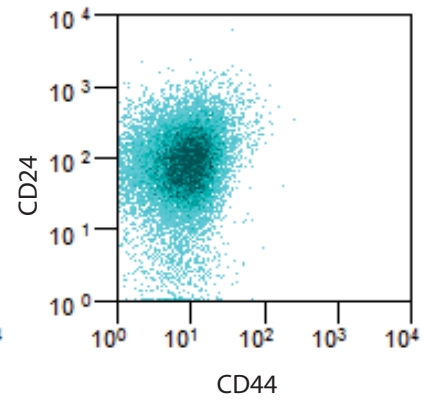
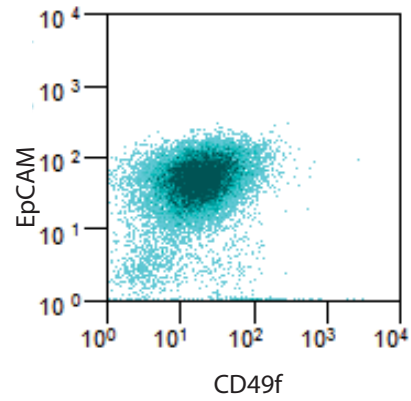
VIMENTIN

KERATIN 5/KERATIN 8

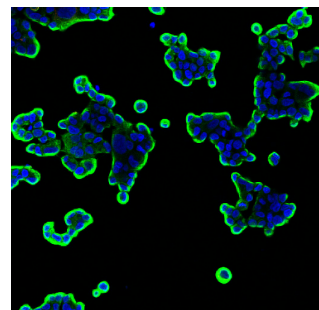
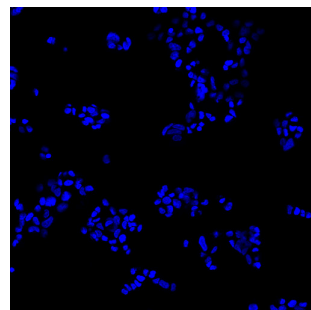
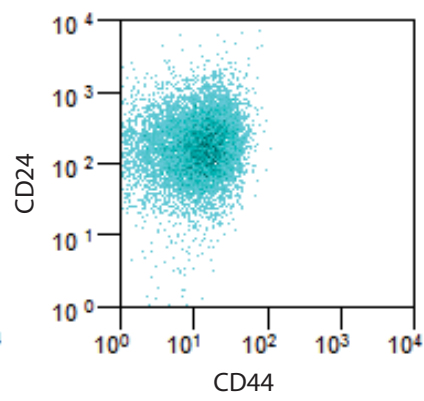
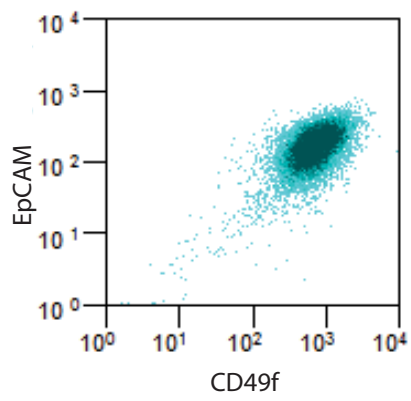
HCC1500



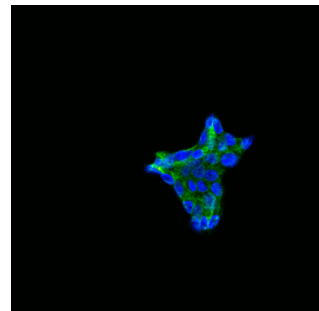
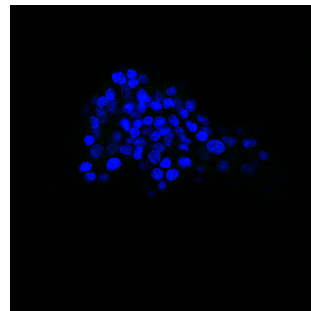
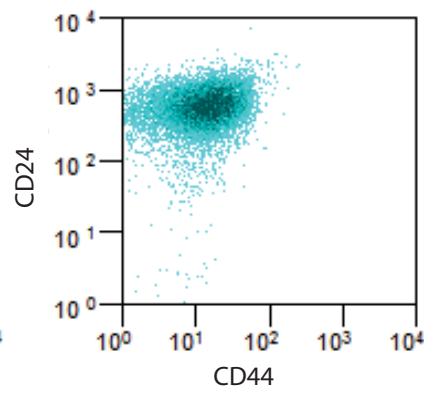
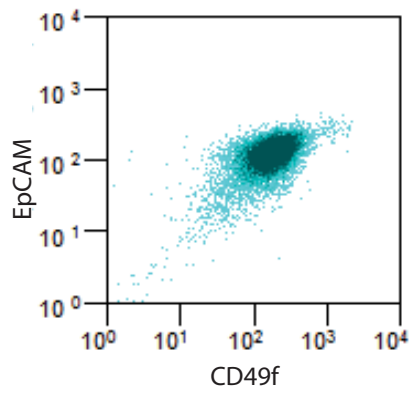
MCF7



T47D



ZR75-1



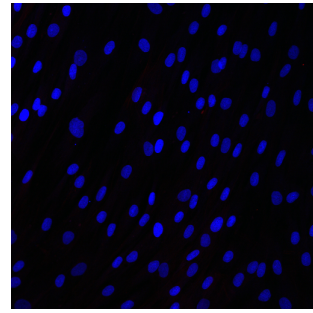
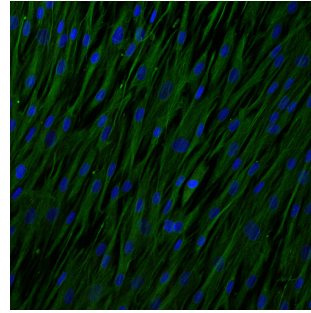
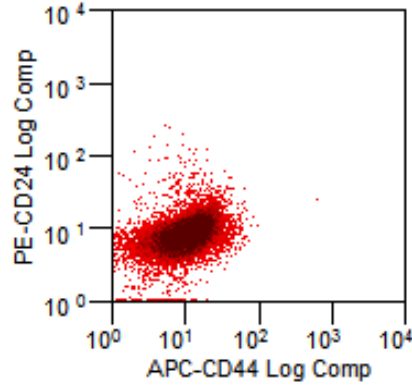
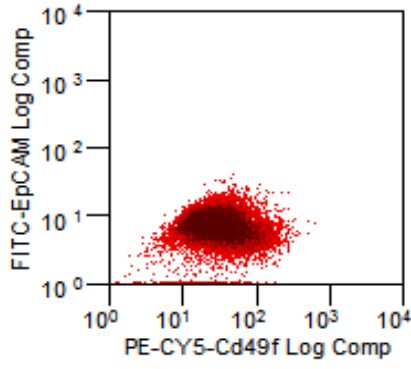
EpCAM/CD49f

CD24/CD44

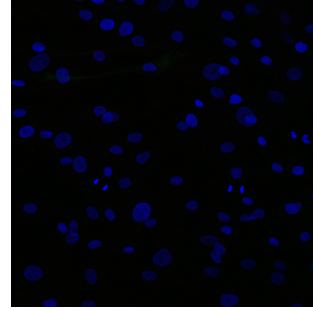
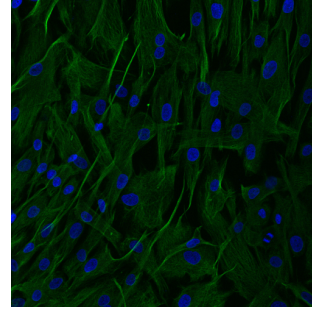
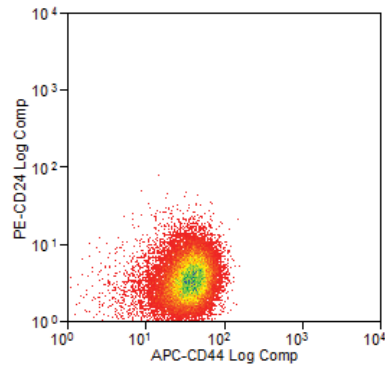
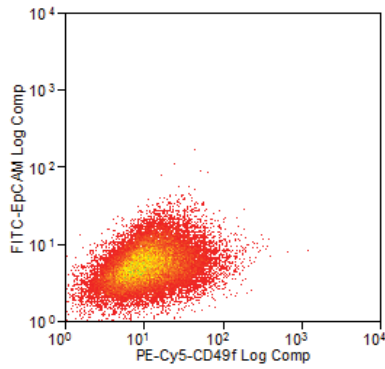
VIMENTIN

KERATIN 5/KERATIN 8

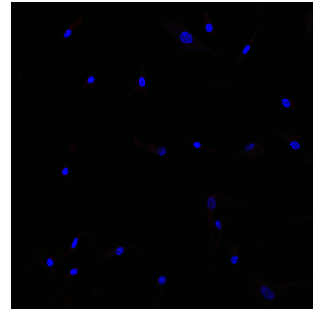
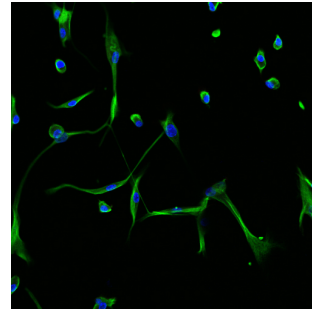
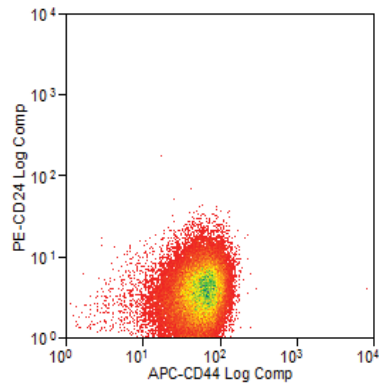
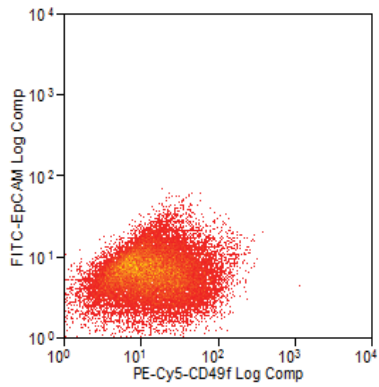
hMSC#1



hMSC#2



hMSC#3



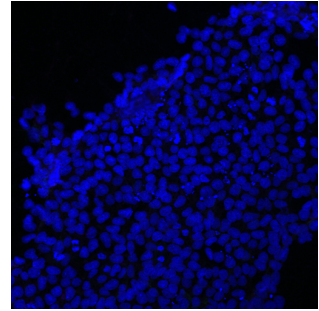
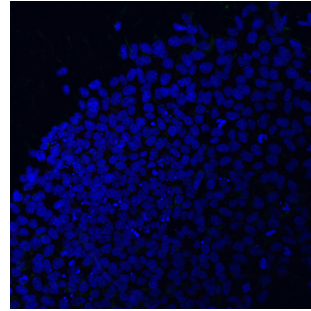
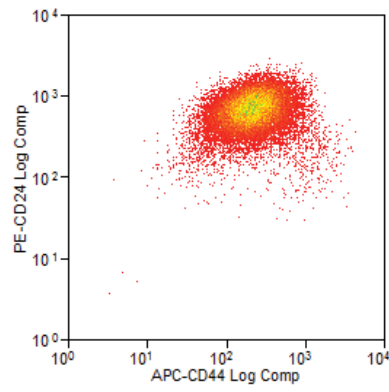
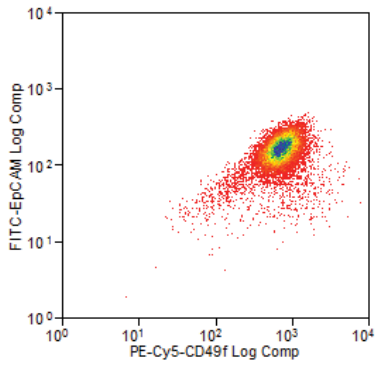
EpCAM/CD49f

CD24/CD44

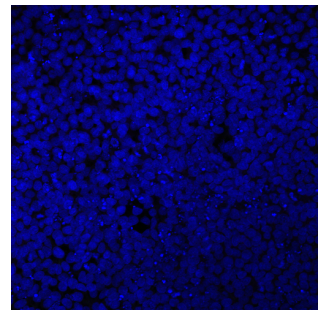
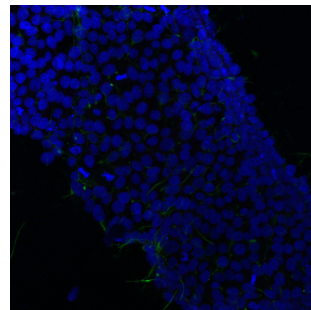
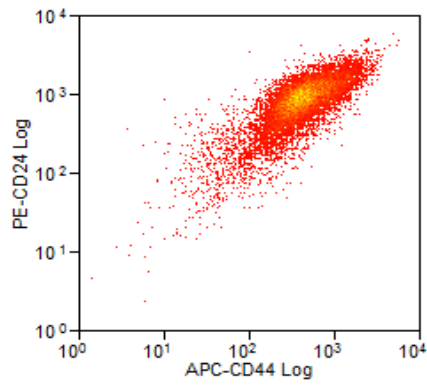
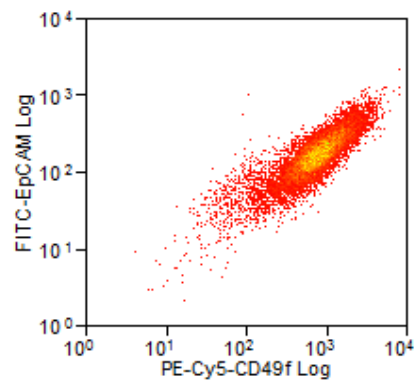
VIMENTIN

KERATIN 5/KERATIN 8

hESC H9

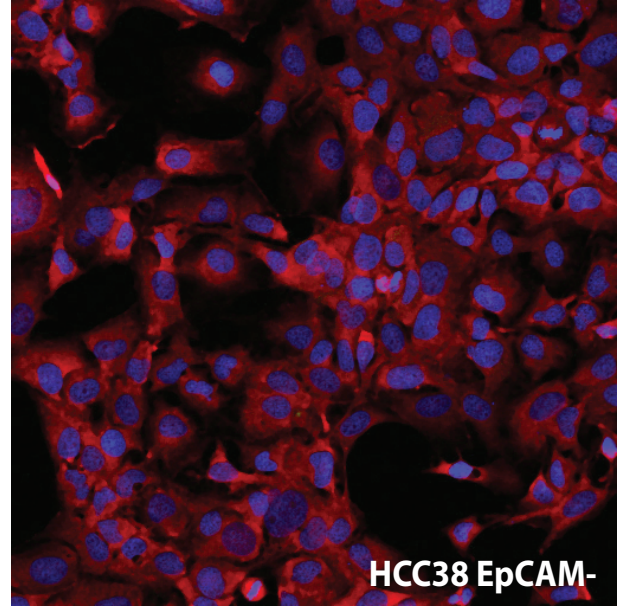
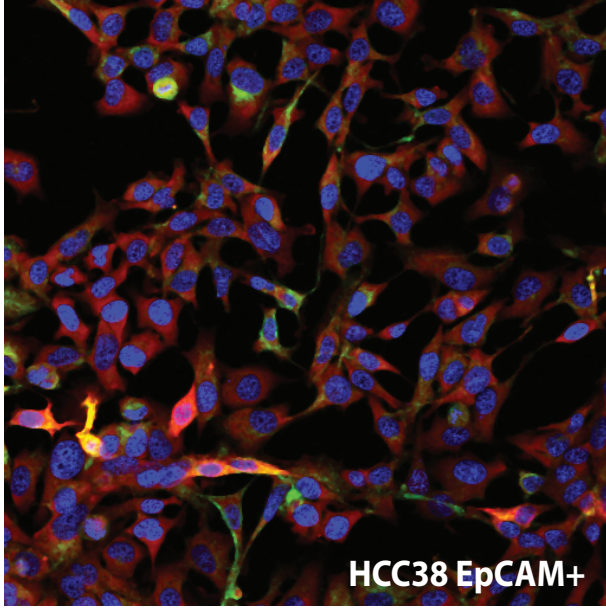


hESC H7

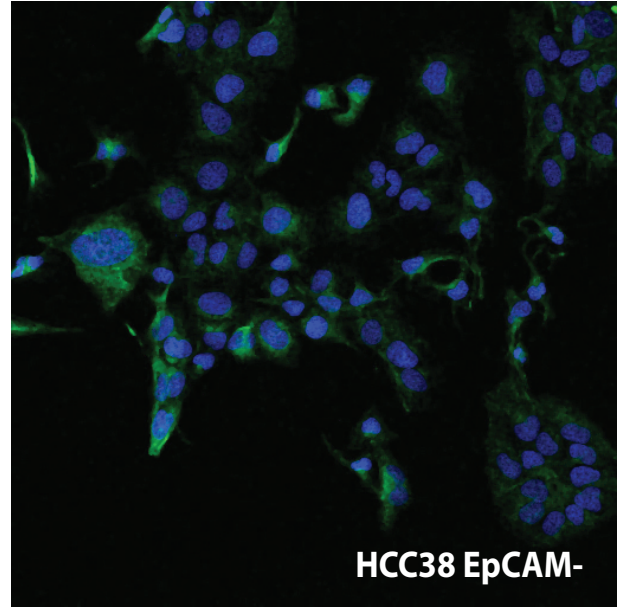
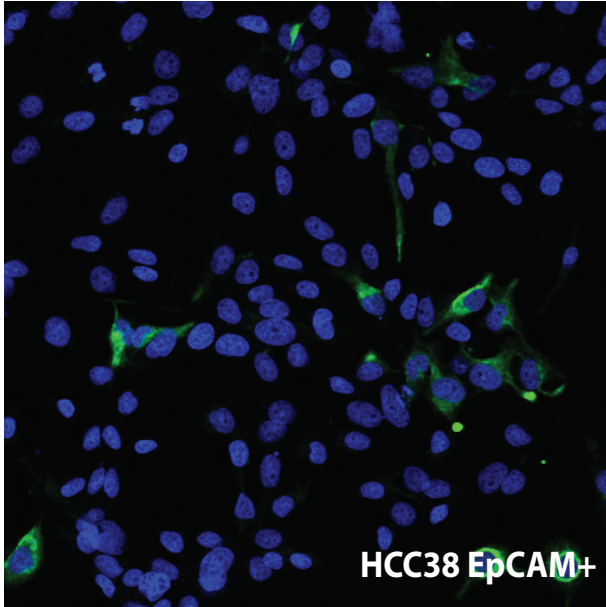


Supplemental Figure 8

Keratin5/Keratin8

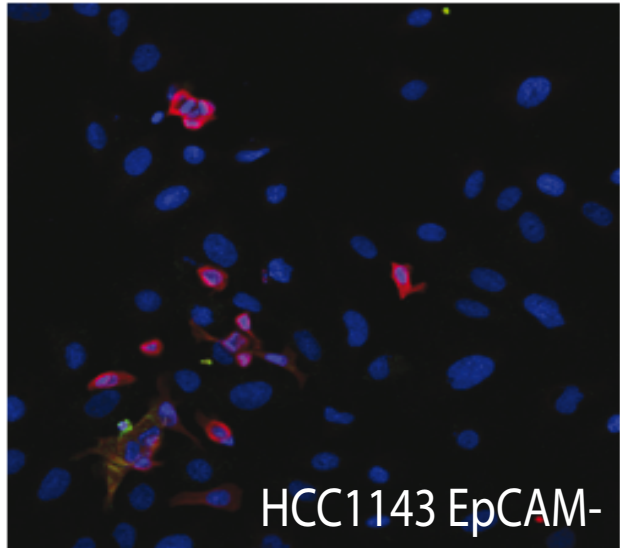
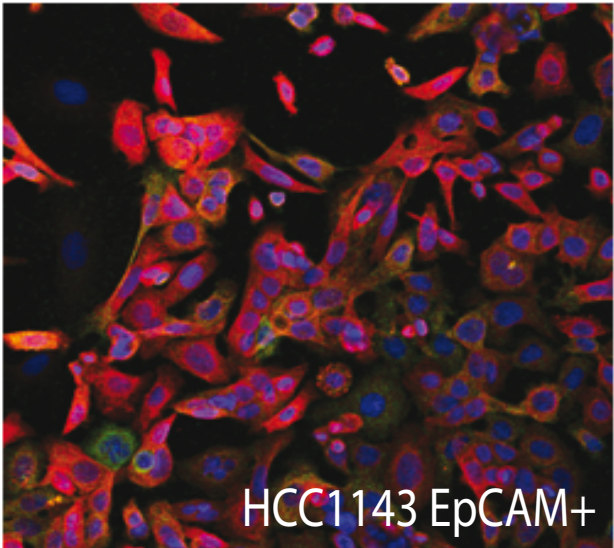


Vimentin

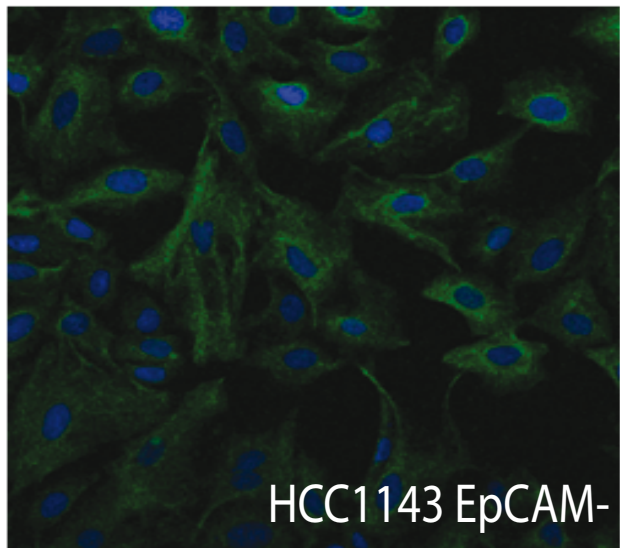
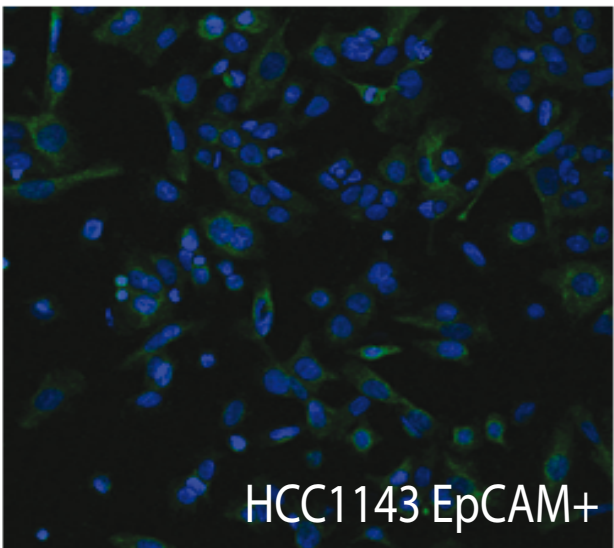


Supplemental Figure 9

Keratin5/Keratin8



Vimentin



Supplemental Figure 10

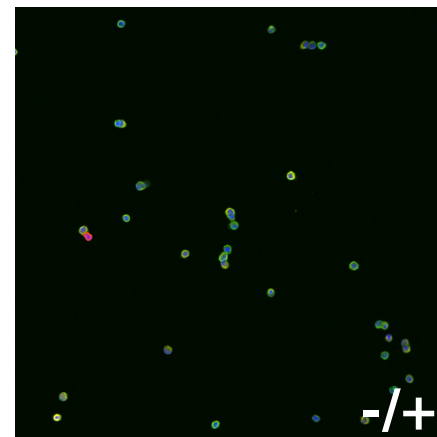
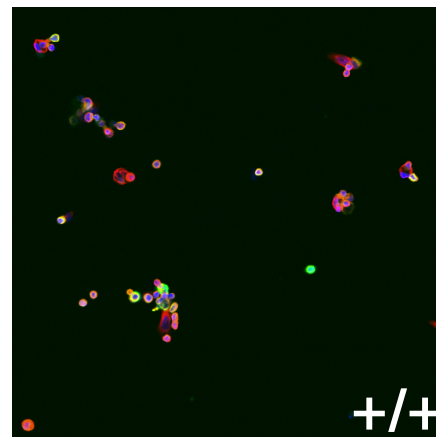
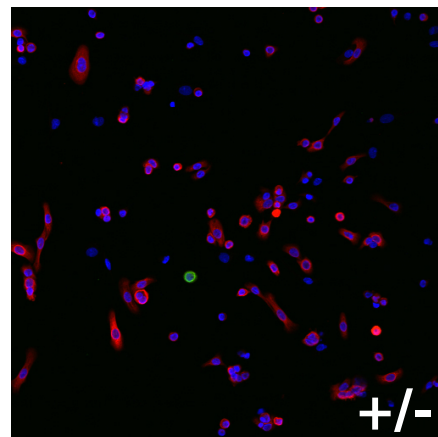
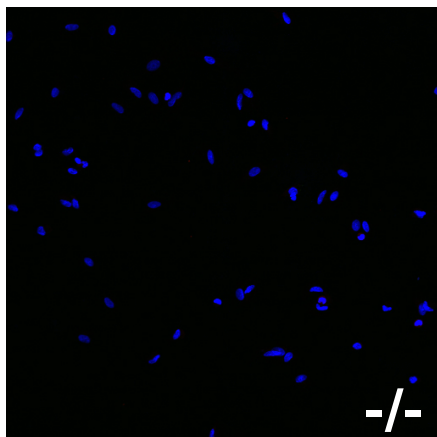
Stromal

MaSC/BiP

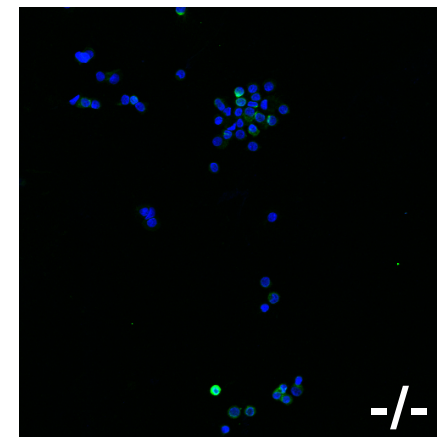
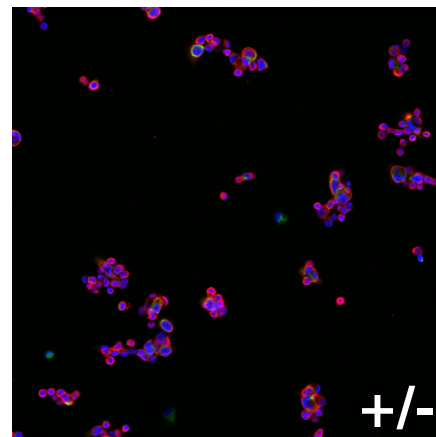
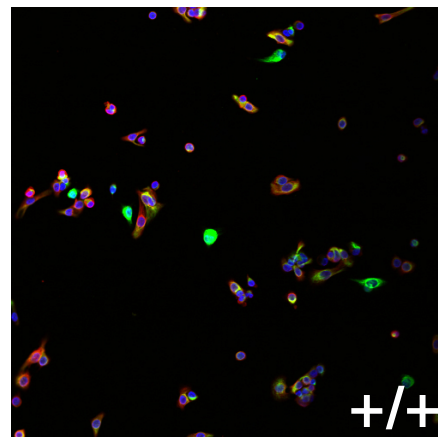
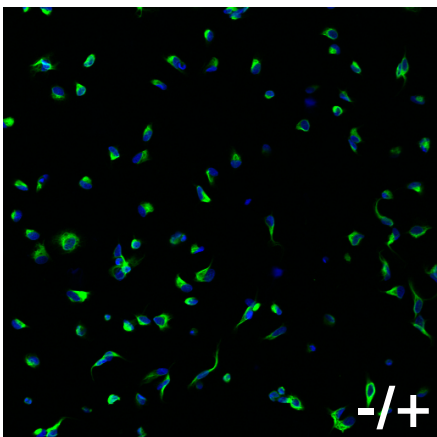
pL

mL

Keratin 5/Keratin 8



Keratin 5/Vimentin



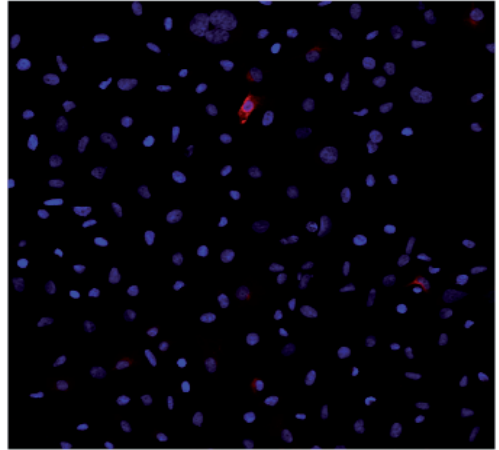
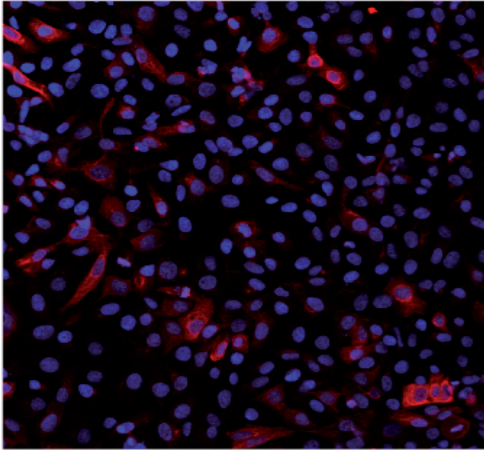
Supplemental Figure 11

SUM149
CD49F_HIGH

SUM149
EPCAM_LOW

KERATIN5

KERATIN8



KERATIN5

VIMENTIN

