Cell Reports, Volume *42*

Supplemental information

Resistance to mesenchymal reprogramming sustains clonal propagation in metastatic breast cancer

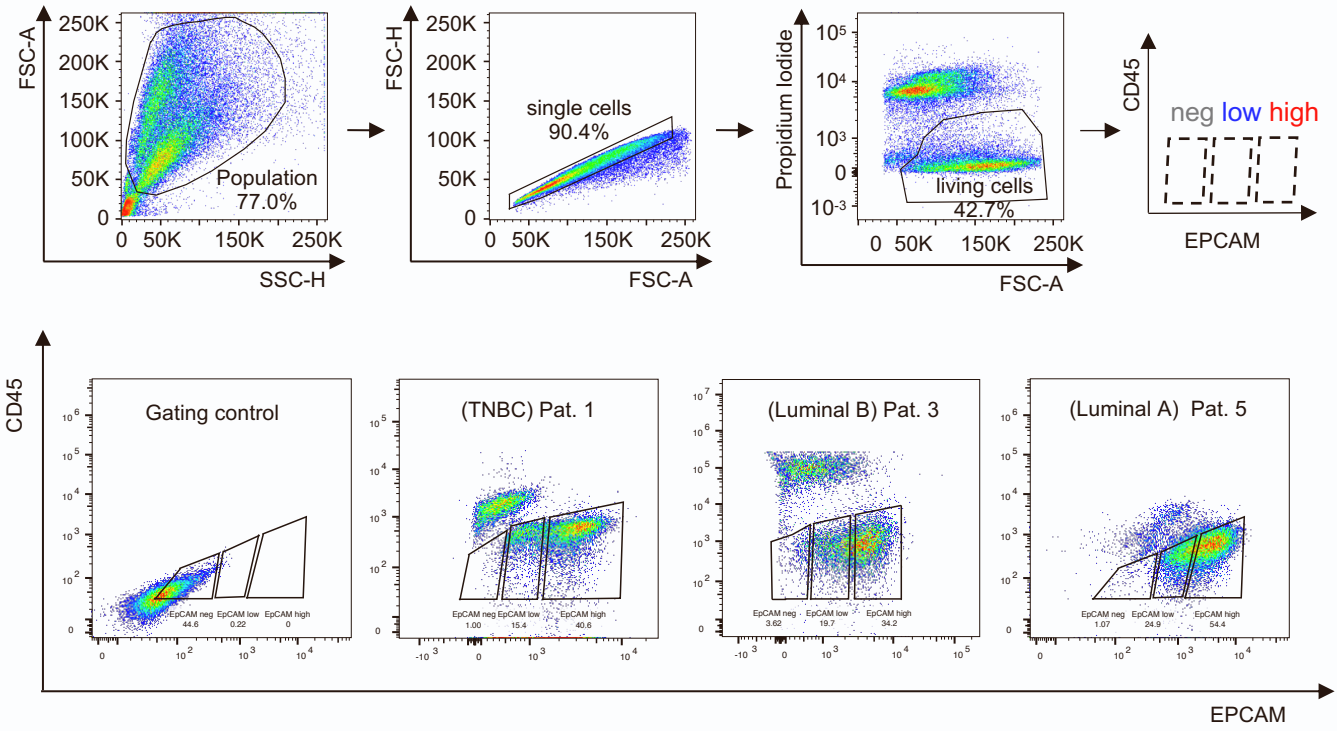
Massimo Saini, Laura Schmidleitner, Helena Domínguez Moreno, Elisa Donato, Mattia Falcone, Johanna M. Bartsch, Corinna Klein, Vanessa Vogel, Roberto Würth, Nicole Pfarr, Elisa Espinet, Mareike Lehmann, Melanie Königshoff, Manuel Reitberger, Simon Haas, Elisabeth Graf, Thomas Schwarzmayr, Tim-Matthias Strom, Saskia Spaich, Marc Sütterlin, Andreas Schneeweiss, Wilko Weichert, Gunnar Schotta, Maximilian Reichert, Nicola Aceto, Martin R. Sprick, Andreas Trumpp, and Christina H. Scheel

Supplementary Figure 1: High EPCAM levels mark disease-propagating cells in MBC biopsies

a

Patient ID	Age	ER status	Clinical subtype and characteristics	Histological type	TNM classification
1	78	neg (-)	Triple negative, inflammatory breast cancer	Invasive lobular carcinoma	cT4d cNx cM1
2	55	pos (+)	Luminal A, peritoneal metastatic	Invasive lobular carcinoma	pT3(m) pN3a M1
3	47	pos (+)	HER2-amplified Luminal B	Invasive ductal carcinoma	ypT4 pN1 M0
4	62	neg (-)	Triple negative	-	pT2 pN0 M0
5	59	pos (+)	Luminal A, liver metastatic	Invasive ductal carcinoma	pT3 pN1a(1/13) M0
6	48	pos (+)	Luminal A	Invasive lobular carcinoma	pT2 pN1a(1/8) cM0
7	66	pos (+)	Luminal B	Invasive ductal carcinoma	pT1 N0(0/13) cM0
8	73	pos (+)	Claudin-Low subtype	-	pT1b(m) pN1b(2/16) 0/5
9	58	pos (+)	Luminal A	Invasive ductal carcinoma	pT4b pN1 cM0
10	67	neg (-)	Triple negative	Invasive ductal carcinoma	pT1c pN3a(21/22) cM0

b



c

Injected cells	1-5 x10^3			1-5 x10^4			1-5 x10^5		
Patient	**EPCAM^{neg}**	**EPCAM^{low}**	**EPCAM^{high}**	**EPCAM^{neg}**	**EPCAM^{low}**	**EPCAM^{high}**	**EPCAM^{neg}**	**EPCAM^{low}**	**EPCAM^{high}**
#1	0/5	1/5	5/5	1/5	3/5	5/5	3/3	5/5	5/5
#2	0/2	0/5	2/5	0/2	1/3	2/3	0/2	2/2	2/2
#3	0/3	0/5	3/5	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
#4	0/5	n.a.	1/5	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
#5	0/2	0/4	0/4	0/2	0/3	1/3	0/2	1/4	3/4
#6	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
#7	0/3	0/3	0/3	0/3	0/5	1/5	0/3	n.a.	n.a.
#8	0/5	n.a.	2/5	0/3	n.a.	n.a.	0/3	n.a.	n.a.
#9	0/3	n.a.	0/3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
#10	0/2	0/2	0/2	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

d

Overall test for differences in TPC frequencies between any of the groups								
ER$^+$ cancers			ER$^-$ cancers			all cancers		
Chisq	DF	p-value	Chisq	DF	p-value	Chisq	DF	p-value
35.7	2	1.78E-08	45.9	2	1.10E-10	50	2	1.41E-11

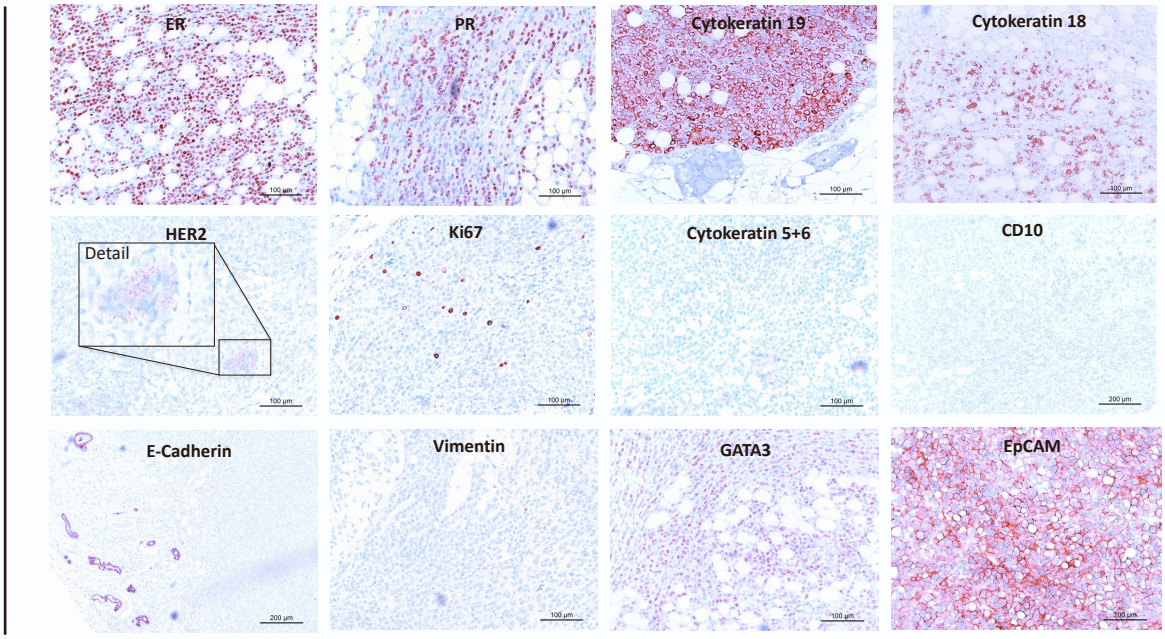
Pairwise tests for differences in TPC frequencies									
	ER$^+$ cancers			**ER$^-$ cancers**			**all cancers**		
	Chisq	DF	Pr (>Chisq)	Chisq	DF	Pr (>Chisq)	Chisq	DF	Pr (>Chisq)
neg vs. high	34.9	1	3.52E-09	43.9	1	3.45E-11	49.8	1	1.68E-12
neg vs. low	7.59	1	0.00588	9.56	1	0.00199	10.1	1	0.00147
low vs. high	10.6	1	0.00113	9.19	1	0.000243	13.9	1	0.000189

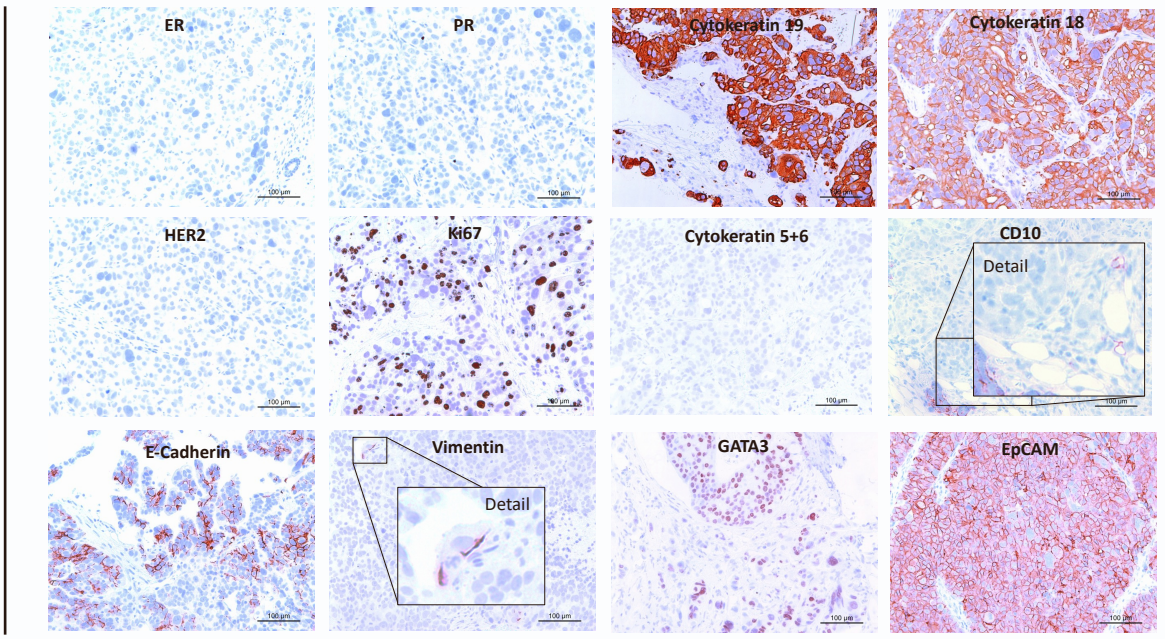
e

IHC stain	EPCAM (BerEP4)			Cytokeratin 19			GATA3			E-Cadherin		
Patient	EPCAM^{neg}	EPCAM^{low}	EPCAM^{high}	EPCAM^{neg}	EPCAM^{low}	EPCAM^{high}	EPCAM^{neg}	EPCAM^{low}	EPCAM^{high}	EPCAM^{neg}	EPCAM^{low}	EPCAM^{high}
#1	diffuse(++)	diffuse(++)	diffuse(++)	every(+++)	every(+++)	every(+++)	diffuse(++)	diffuse(++)	diffuse(++)	every(+++)	every(+++)	every(+++)
#2	n.a.	every(+++)	every(+++)	n.a.	every(+++)	every(+++)	n.a.	diffuse(+)	diffuse(+)	n.a.	negative	negative
#3	n.a.	n.a.	every(+++)	n.a.	n.a.	every(+++)	n.a.	n.a.	diffuse(++)	n.a.	n.a.	every(+++)
#4	n.a.	n.a.	every(+++)	n.a.	n.a.	every(+++)	n.a.	n.a.	diffuse(+)	n.a.	n.a.	diffuse(++)
#5	n.a.	diffuse(++)	diffuse(++)	n.a.	diffuse(++)	diffuse(++)	n.a.	diffuse(+)	diffuse(+)	n.a.	focal (+++)	focal (+++)
#6	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
#7	n.a.	n.a.	diffuse(++)	n.a.	n.a.	every(+++)	n.a.	n.a.	focal (++)	n.a.	n.a.	focal (+++)
#8	n.a.	n.a.	diffuse(+)	n.a.	n.a.	diffuse(++)	n.a.	n.a.	diffuse(++)	n.a.	n.a.	negative
#9	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
#10	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

IHC stain	Vimentin			ZEB1			CD10			Cytokeratin 5/6		
Patient	EPCAM^{neg}	EPCAM^{low}	EPCAM^{high}	EPCAM^{neg}	EPCAM^{low}	EPCAM^{high}	EPCAM^{neg}	EPCAM^{low}	EPCAM^{high}	EPCAM^{neg}	EPCAM^{low}	EPCAM^{high}
#1	disperse(+)	disperse(+)	disperse(+)	negative	negative	negative	negative	negative	negative	very rare	very rare	very rare
#2	n.a.	negative	negative	n.a.	negative	negative	n.a.	negative	negative	n.a.	negative	negative
#3	n.a.	n.a.	negative	n.a.	n.a.	negative	n.a.	n.a.	negative	n.a.	n.a.	negative
#4	n.a.	n.a.	negative	n.a.	n.a.	negative	n.a.	n.a.	negative	n.a.	n.a.	very rare
#5	n.a.	negative	negative	n.a.	negative	negative	n.a.	negative	negative	n.a.	negative	negative
#6	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
#7	n.a.	n.a.	negative	n.a.	n.a.	negative	n.a.	n.a.	negative	n.a.	n.a.	negative
#8	n.a.	n.a.	diffuse(++)	n.a.	n.a.	negative	n.a.	n.a.	very rare	n.a.	n.a.	negative
#9	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
#10	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

f





g

	Pat.1 (TNBC) EPCAM high	Pat.1 (TNBC) EPCAM low	Pat.1 (TNBC) EPCAM neg	Pat.6 (Luminal A) EPCAM high	Pat.6 (Luminal A) EPCAM low	Pat.6 (Luminal A) EPCAM neg
Epithelial cell lineage genes						
EPCAM	1.59	1.51	1.23	1.32	1.25	0.63
CDH1	1.78	1.61	1.37	0.49	0.46	0.40
ESRP1	1.66	1.69	1.29	1.50	1.43	0.70
ESRP2	0.99	0.93	0.71	0.72	0.42	-0.05
GRHL2	1.12	1.02	0.76	1.45	1.43	0.83
CD24	3.10	3.07	2.75	2.55	2.42	1.62
JUP	1.64	1.60	1.43	1.53	1.44	0.88
OVOL2	0.48	0.23	0.05	0.47	0.32	-0.18
GATA3	1.59	1.33	0.96	2.01	1.78	1.12
FOXA1	1.42	1.28	0.99	1.33	1.20	0.39
SPDEF	1.44	1.30	1.02	1.62	1.08	0.69
KRT19	2.55	2.31	2.33	2.39	2.33	2.24
KRT15	1.20	0.68	0.35	0.94	0.68	0.31
ALCAM	1.32	1.29	1.08	1.63	1.32	0.99
Mesenchymal and stromal genes						
CDH2	-1.11	-1.03	0.58	-0.36	-0.21	1.14
DCN	-1.24	-1.07	1.36	-1.03	-0.05	1.76
ITGA5	-0.71	-0.03	0.64	-0.78	-0.72	0.85
LAMA3	0.45	0.73	0.55	0.34	0.46	0.23
LAMA5	-0.10	0.26	0.09	0.36	0.47	0.09
LAMB1	0.62	0.33	1.09	-0.37	0.25	1.31
LAMB2	0.23	0.57	1.24	1.26	1.37	1.27
LAMB3	0.21	0.63	0.53	-0.36	0.03	0.13
LAMC1	-0.51	-0.16	0.91	0.19	0.69	1.08
LAMC2	1.05	1.34	0.96	-0.04	0.67	0.40
COL1A1	-0.14	0.08	1.82	-0.18	0.24	2.38
COL4A1	-1.85	-2.06	0.06	-1.92	-1.41	0.22
COL4A5	0.04	0.12	0.53	0.77	0.79	0.76
COL12A1	-1.25	-0.51	-0.11	-0.03	0.87	0.31
COL15A1	-1.12	-1.89	-0.38	0.34	0.47	0.80
COL27A1	-0.60	-0.70	0.01	-0.22	0.07	0.51
SERPINA1	1.19	1.57	1.38	1.25	1.29	0.36
SERPINB4	1.16	1.62	1.76	-0.66	-0.43	1.74
SERPINB8	0.51	0.75	1.00	0.24	0.64	0.94
SERPINE3	-0.09	0.00	-0.20	-0.61	-0.40	-0.41
Cell cycle progression and self-renewal promoting genes						
MKI67	0.95	-0.08	-0.19	0.42	-0.72	0.00
PLK1	0.81	0.09	-0.11	0.55	-0.39	0.17
CCNB1	1.94	1.17	0.91	1.35	0.64	1.17
CCNE1	0.45	-0.43	-0.52	0.35	-0.23	0.15
CCNE2	-0.11	-0.72	-0.63	0.29	-0.37	-0.54
AURKA	1.48	0.89	0.51	0.64	0.05	0.27
AURKB	1.07	0.02	-0.19	0.69	-0.36	0.20
BIRC5	1.37	0.28	0.17	0.93	-1.10	0.60
BMI1	0.88	0.75	0.74	1.11	0.89	0.91
MYB	0.67	0.45	0.06	0.75	0.18	-0.14
MYCN	0.44	-0.55	-1.21	-0.86	-1.27	-0.95
PCNA	1.89	1.55	1.50	1.52	1.32	1.58
Cell cycle Inhibition and growth arrest genes						
CDKN2B	-0.30	0.12	0.08	-0.89	0.02	0.16
CDKN1B	0.32	0.38	0.49	1.19	1.23	0.79
CDKN1C	-0.17	0.03	0.71	0.03	0.59	1.39
CDKN2A	-0.83	-0.80	-0.85	-1.12	-0.78	-0.58
GADD45A	0.31	0.69	0.98	0.40	0.84	1.47
GADD45B	1.00	1.30	1.31	1.06	1.39	1.22
GADD45G	-0.56	-0.37	-0.67	0.09	0.25	-0.54
TGF-β signaling genes						
SOX4	1.50	1.97	1.89	2.01	2.36	1.94
KLF10	0.83	1.87	1.83	1.41	1.76	1.81
TGFBI	0.32	1.14	1.34	-0.51	0.53	1.30
TGFBR2	0.57	0.83	1.19	-0.45	-0.15	1.37
EDN1	0.90	1.66	1.43	0.77	1.64	0.53
NEDD9	0.45	0.83	0.80	0.49	1.13	0.34
WNT5A	-0.35	0.28	0.42	0.33	0.54	0.69
FOXC1	-0.62	-0.17	0.24	-0.78	-0.20	0.69
ENG	-0.02	0.40	0.57	0.81	0.91	0.48
IL6R	0.25	0.86	0.78	-0.30	0.61	0.30
BCL6	0.58	1.25	1.09	0.57	0.98	0.85
TGM2	1.45	2.24	2.08	0.76	1.26	1.36
TFPI2	-0.21	0.37	0.92	-0.29	0.81	1.25

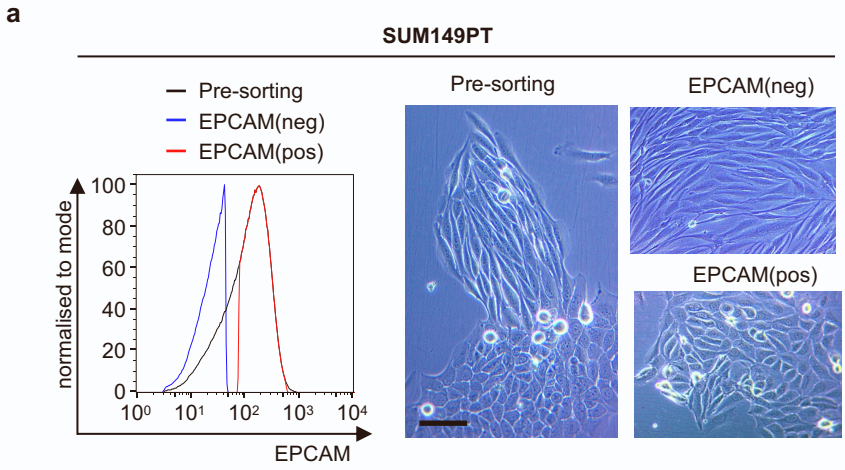
Expression as RPKM, Log(10)

row min — row max

Supplementary Figure 1: High EPCAM levels mark disease-propagating cells in MBC biopsies, related to Figure 1

(a) Clinical characteristics referring to patient samples presented in Figure 1. **(b)** Gating strategy for FACS sorting presented in Figure 1. **(c)** Summary table of limiting dilution analyses presented in Figure 1b. **(d)** Statistical tests of the limiting dilution MFP injection:(middle table) overall test for goodness of fit , (bottom table) individual tests for direct comparison between two conditions.**(e)** Summary of the results obtained from the immunohistochemical staining of mammary fat pad tumor xenografts presented in Figure 1b and 1c; “+++”, “++”,“+”= strong, moderate,faint expression, respectively; “Every” =positivity for all tumor cells; “Diffuse” = positivity on a majority of tumor cells; “Focal” = positivity of minority of tumor cells, localized in a focal area; “Disperse” = positivity for a minority of sparse, individually distributed tumor cells. “Very rare” = positivity for a very rare minority of sparse, individually distributed tumor cells. At least one complete tumor section scored per condition. **(f)** Representative IHC staining of markers in MFP tumor xenografts obtained from EPCAM(high) cells sorted from one ER^+ and one ER^- MBC case; **(g)** Heatmaps highlighting row differences in mRNA expression of relevant genes from each indicated category; Log-transformed read counts (RPKM) of EPCAM-sorted fractions from one ER^- and one ER^+ MBC case.

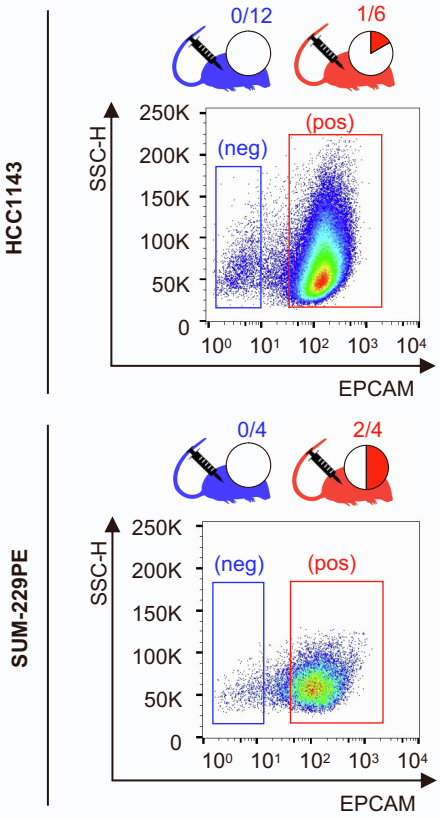
Supplementary Figure 2: Divergent response to EMT stimulation generates heterogeneity fo EPCAM

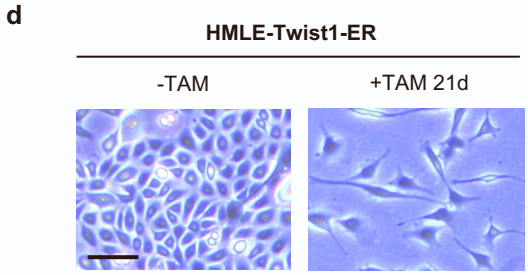


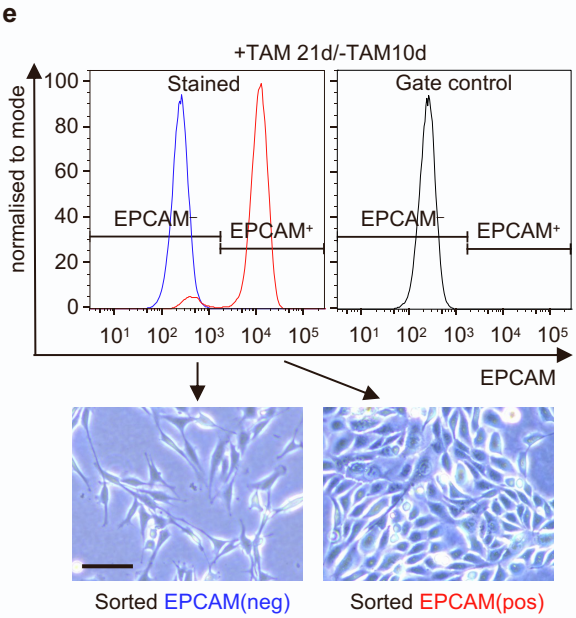
b

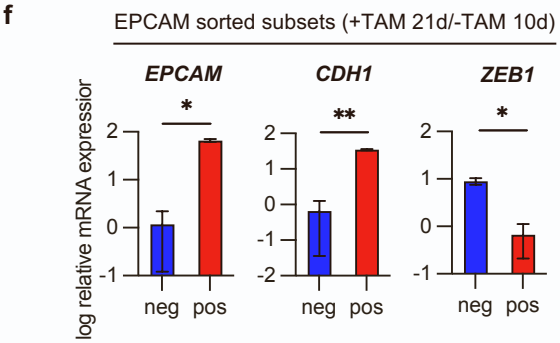
		SUM149PT		HCC1143		
		pos	neg		pos	neg
Epithelial cell lineage genes	EPCAM	1.45	0.94	EPCAM	1.56	1.47
	CDH1	1.17	0.19	CDH1	1.42	0.86
	ESRP1	1.61	0.64	ESRP1	1.58	1.01
	ESRP2	0.19	-0.63	ESRP2	0.64	1.39
	GRHL2	1.15	0.19	GRHL2	1.40	0.45
	CD24	2.32	1.76	CD24	2.29	1.41
	JUP	1.25	0.82	JUP	1.26	0.83
	OVOL2	0.41	-0.15	OVOL2	0.46	0.00
	KRT19	1.87	1.25	KRT19	1.99	1.15
	KRT15	0.78	-0.35	KRT15	0.76	0.79
	ALCAM	1.09	1.13	ALCAM	-0.13	0.16
Mesenchymal and stromal genes	CDH2	0.35	0.62	CDH2	0.68	1.44
	DCN	-1.03	0.28	DCN	-1.32	-1.75
	SPARC	-0.82	0.26	SPARC	-0.49	0.49
	ENG	0.04	1.50	ENG	0.16	0.22
	TNC	0.58	1.07	TNC	0.90	1.01
	SPOCK1	-1.18	0.33	SPOCK1	0.77	1.46
	ITGA5	-0.33	0.99	ITGA5	0.67	1.26
	LAMA5	0.27	1.58	LAMA5	-0.02	-0.52
	LAMB1	1.04	0.97	LAMB1	0.82	1.14
	LAMB2	0.85	1.23	LAMB2	1.07	1.28
	LAMC1	0.77	0.48	LAMC1	0.82	1.29
	COL1A1	-0.12	1.48	COL1A1	0.06	1.35
	COL4A1	-0.43	0.60	COL4A1	0.00	0.06
	COL12A1	0.50	0.98	COL12A1	0.27	0.64
	COL15A1	-1.25	0.93	COL15A1	-0.86	-0.27
	COL27A1	-0.40	0.77	COL27A1	-0.61	-0.33
	SERPINB4	-1.25	-0.23	SERPINB4	0.25	0.46
	SERPINB8	0.47	0.96	SERPINB8	0.72	1.24
	SPARC	-0.82	0.26	SPARC	-0.49	0.49
EMT markers and EMT-TFs	VIM	-0.44	1.71	VIM	1.85	3.04
	FN1	0.60	0.95	FN1	0.29	1.46
	SNAI1	0.59	1.44	SNAI1	1.11	1.55
	SNAI2	-0.11	1.09	SNAI2	0.06	0.87
	TWIST1	0.94	1.41	TWIST1	1.03	1.49
	TWIST2	-0.54	1.02	TWIST2	0.00	-1.28
	MME/CD10	0.24	1.03	MME/CD10	-0.50	0.76
	ZEB1	-1.40	0.79	ZEB1	0.12	1.29
	ZEB2	-0.94	0.45	ZEB2	-1.29	0.08
	TCF3	0.42	1.21	TCF3	0.17	0.42
	FOXC2	-0.70	1.63	FOXC2	-0.73	-0.18
	FOXC1	0.27	1.24	FOXC1	0.40	0.58
	GSC	-1.41	1.15	GSC	0.00	-1.49

Expression as RPKM, Log(10)
row min — row max





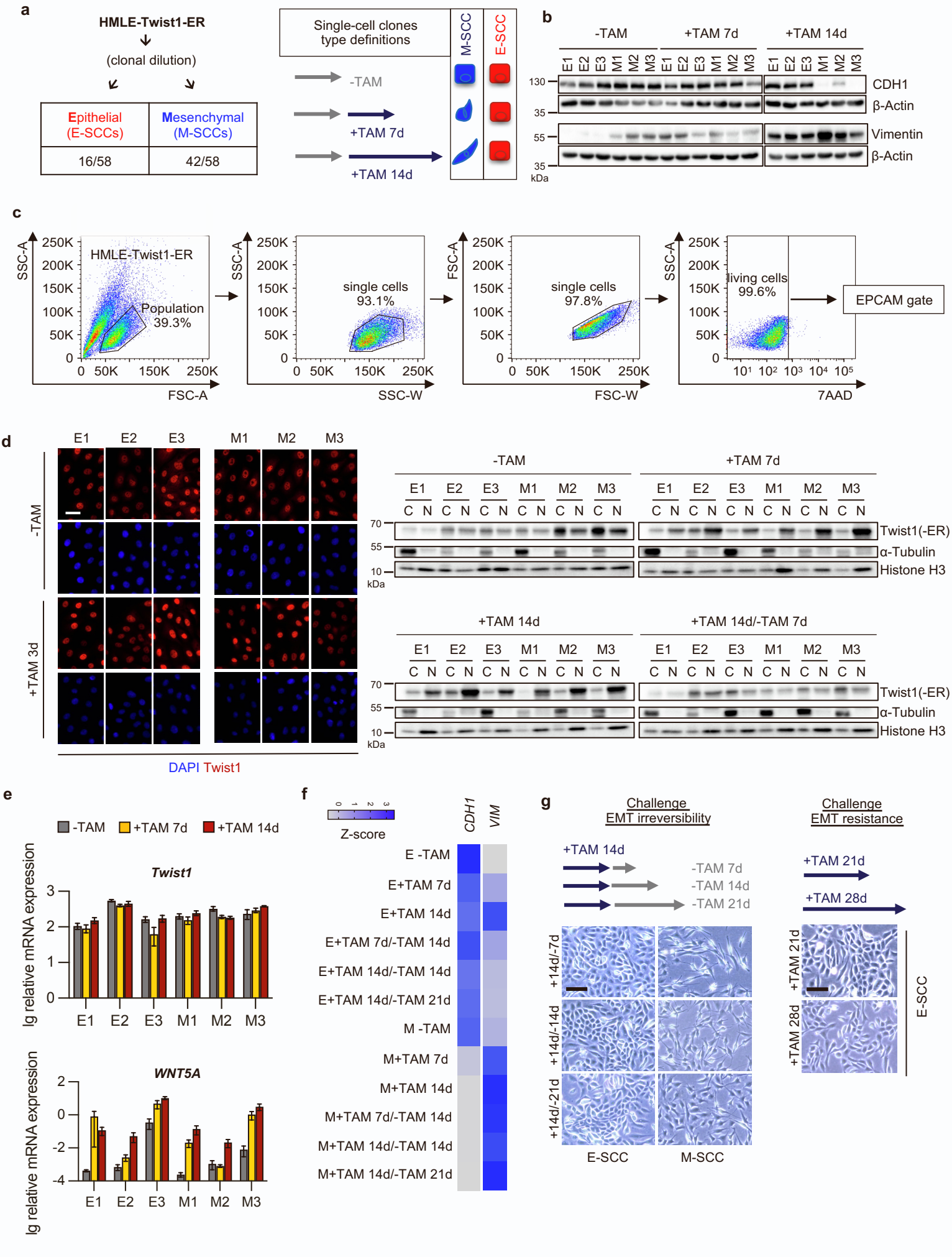




Supplementary Figure 2: Divergent response to EMT stimulation generates heterogeneity for EPCAM, related to Figure 2

(a) Flow cytometry for EPCAM and bright-field images of SUM-149PT in 2D culture. Scale bar: 100 µm. **(b)** Heatmaps highlighting row differences in mRNA expression of relevant genes from each indicated category; Log-transformed read counts (RPKM) of EPCAM-sorted fractions from the SUM-149PT cell line (left columns) or the HCC1143 cell line (right columns) **(c)** Schematic overview of the FACS sorting and transplantation experiments from the cell lines HCC1143 and SUM-229PE; pie charts show fraction of mice developing tumors, as assessed by MFP engraftment. **(d)** Bright-field images of HMLE-Twist1-ER bulk cellsScale bar: 100 µm. **(e)** FACS sorting of HMLE-Twist1-ER cells (+TAM21d/-TAM10d); histogram plots show flow cytometry (re-analysis) of EPCAM in the sorted fractions (blue and red plots)) for the experiment shown in Figure 2d; bright-field images of the cells after sorting and growth in 2D culture conditions. Scale bar: 100 µm. **(f)** Log relative mRNA expression of indicated genes of sorted EPCAM(neg) and EPCAM(pos) HMLE-Twist1-ER cells. n=2; mean ± SEM; multiple t-tests (Holm-Sidak correction); p-values: $*<0.05$, $**<0.005$.

Supplementary Figure 3: Single-cell clonal analysis reveals intrinsic resistance to EMT versus stable mesenchymal reprogramming



h

qRT-PCR statistical analysis

Individual clone comparisons **-TAM vs. +TAM**

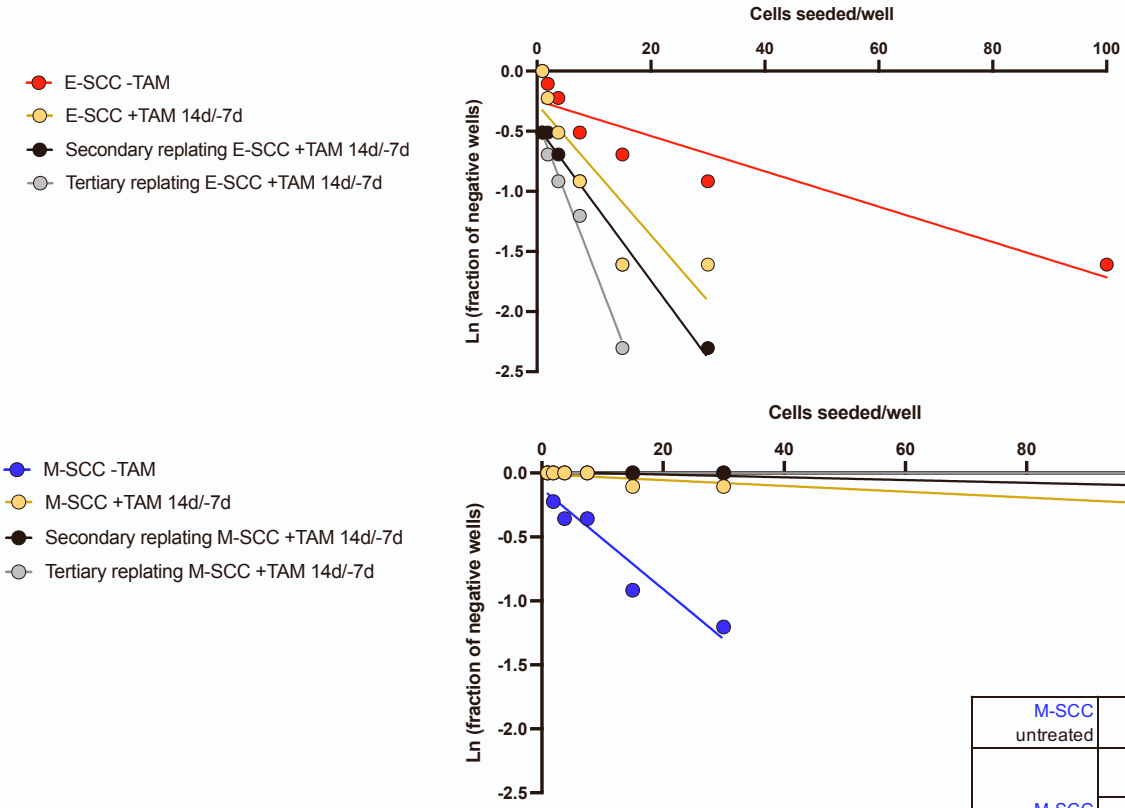
	Treatment	Mean of -TAM	Mean of +TAM 7d	P-Value	Significance
EPCAM	E1	115.40	91.58	0.513167	ns
	E2	108.50	74.68	0.528414	ns
	E3	92.34	47.26	0.328443	ns
	M1	77.09	25.56	0.030750	*
	M2	52.16	4.85	0.029734	*
	M3	40.16	1.29	0.103768	ns
CDH1	E1	19.60	14.70	0.465128	ns
	E2	21.28	11.66	0.448367	ns
	E3	11.94	7.47	0.321372	ns
	M1	16.79	4.42	0.170958	ns
	M2	12.60	0.59	0.068611	ns
	M3	3.44	0.10	0.337229	ns
VIM	E1	236.90	632.80	0.015611	*
	E2	145.30	470.60	0.012191	*
	E3	33.39	188.60	0.007277	*
	M1	230.30	1437.00	0.009408	*
	M2	389.50	1425.00	0.048090	*
	M3	1058.00	1794.00	0.189160	ns
FN1	E1	100.20	564.00	0.000026	***
	E2	375.40	661.90	0.358571	ns
	E3	146.70	510.90	0.009272	*
	M1	423.50	1172.00	0.107012	ns
	M2	623.70	1020.00	0.318882	ns
	M3	375.60	1217.00	0.069869	ns

Group comparisons **E-SCCs (E) vs. M-SCCs (M)**

	Treatment	Mean of E	Mean of M	Adjusted P Value	Significance
CDH1	-TAM	17.61	10.94	0.180804	ns
	+TAM 7d	11.27	1.70	0.011887	*
	+TAM 14d	9.97	0.08	0.011887	*
	+TAM 14d/-TAM 7d	12.48	0.04	0.002456	**
	+TAM 14d/-TAM 14d	9.57	0.05	0.041363	*
	+TAM 14d/-TAM 21d	10.26	0.02	0.001189	**
VIM	-TAM	138.50	559.40	0.067261	ns
	+TAM 7d	430.70	1552.00	0.000031	***
	+TAM 14d	2076.00	2548.00	0.567712	ns
	+TAM 14d/-TAM 7d	1059.00	2601.00	0.033260	*
	+TAM 14d/-TAM 14d	611.30	2118.00	0.000644	**
	+TAM 14d/-TAM 21d	481.40	2404.00	0.000644	**
FN1	-TAM	207.40	474.20	0.047906	*
	+TAM 7d	578.90	1136.00	0.047906	*
	+TAM 14d	1178.00	1474.00	0.464217	ns
	+TAM 14d/-TAM 7d	597.50	1390.00	0.134425	ns
	+TAM 14d/-TAM 14d	416.60	1206.00	0.134817	ns
	+TAM 14d/-TAM 21d	405.70	1565.00	0.047906	*

	Treatment	Mean of E	Mean of M	Adjusted P-Value	Significance
EPCAM	-TAM	105.40	56.47	0.029460	*
	+TAM 7d	71.17	10.57	0.000889	**
	+TAM 14d	61.04	1.01	0.000889	**

i

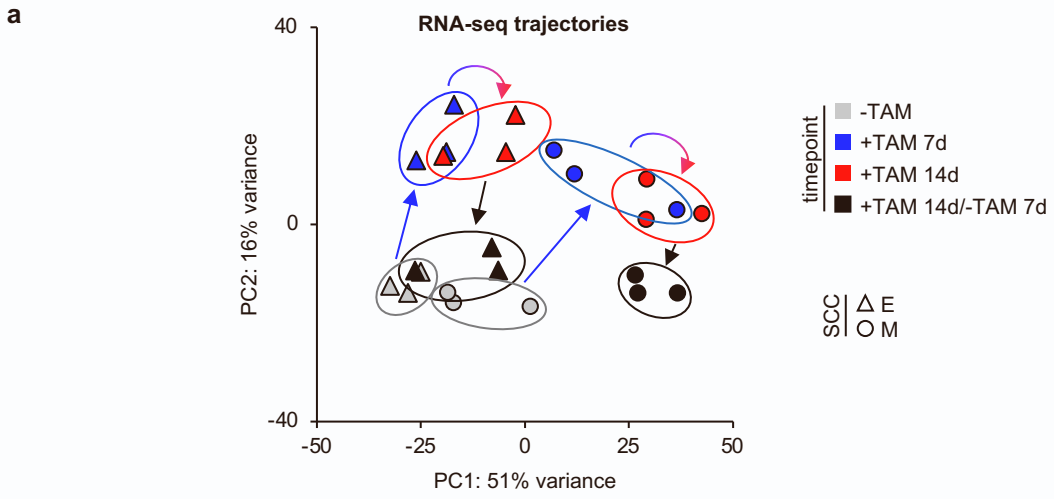


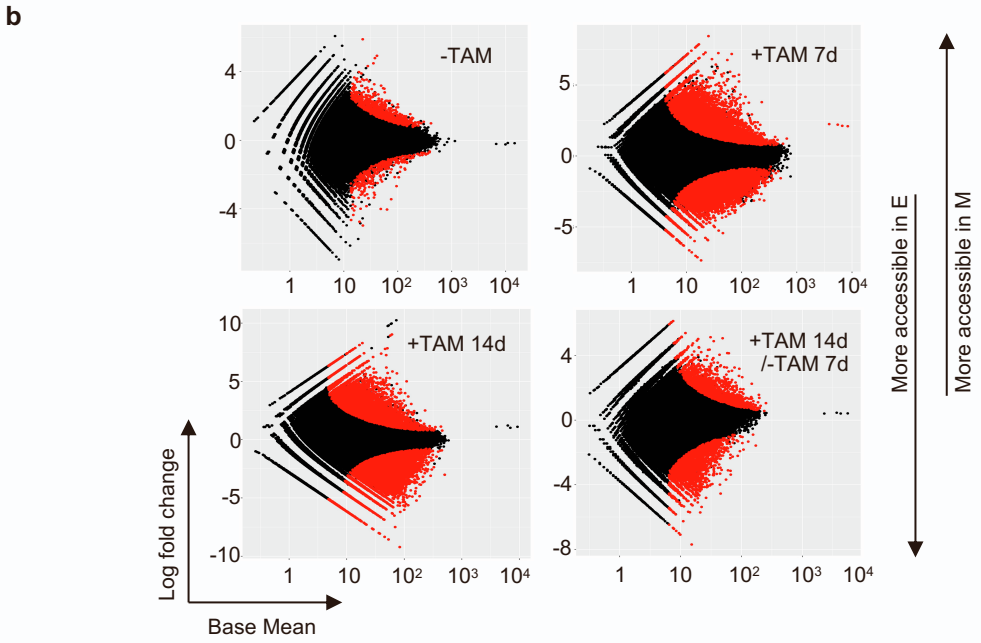
	Plating round	Frequency of MS-initiating cells	
M-SCC untreated	1st	1 / 23	***
M-SCC +TAM 14d/-7d	1st	1 / 366	*
	2nd	1 / 1540	ns
	3rd	n.d.	
E-SCC untreated	1st	1 / 32.6	**
E-SCC +TAM 14d/-7d	1st	1 / 11.8	ns
	2nd	1 / 7.21	ns
	3rd	1 / 4.43	

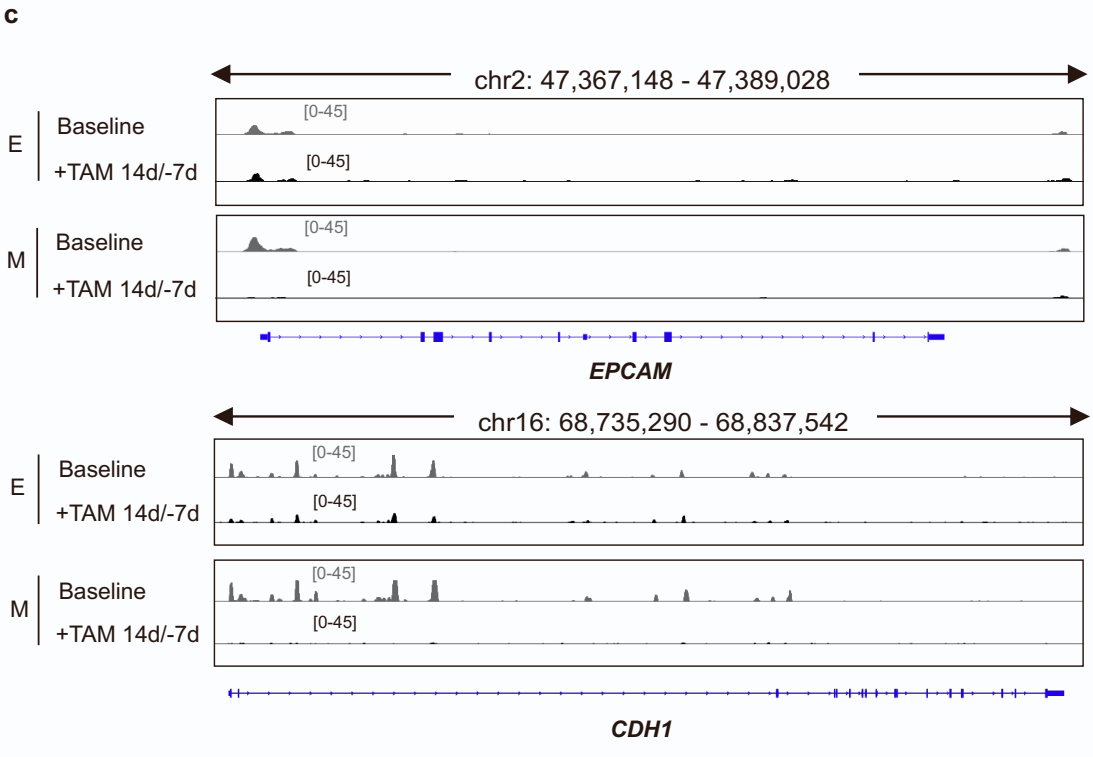
Supplementary Figure 3: Single-cell clonal analysis reveals intrinsic resistance to EMT versus stable mesenchymal reprogramming, related to Figure 3

(a) (left) schematic visualization of the yield in epithelial (E-) or mesenchymal (M-) single-cell clones (SCCs) isolation; (right) graphical summary of the assignment of the "E-SCC" or "M-SCC" definitions to each clone. **(b)** Immunoblot of EMT markers in E-SCCs (E1-E3) and M-SCCs (M1-M3). **(c)** Gating strategy utilized for flow cytometry shown in Figure 3b. **(d)** (left) IF staining for Twist1 in E-SCCs (E1-E3) and M-SCCs (M1-M3) ; scale bar: 20 μm; (right) Immunoblot of Twist1 in nuclear and cytoplasmic protein fractions of the same cells.. **(e)** Log relative mRNA expression of indicated genes in E-SCCs (E1-E3) and M-SCCs (M1-M3); n=3; mean ± SEM. **(f)** Heatmap of relative mRNA expression of indicated genes in E-SCCs (average of 3 clones) and M-SCCs (average of 3 clones); n=3 per each clone; mean values are shown. **(g)** (left) experiment to challenge EMT irreversibility; Bright-field images of a representative E-SCC (E1) and a representative M-SCC (M2; scale bar: 100 μm. (right) experiment to challenge EMT resistance; Bright-field images of a representative E-SCC (E1). Scale bar: 100 μm. **(h)** Statistical comparisons referring to Figure 3c and Supplementary Figure 3f; Student's t-test (left table) or multiple t-tests (Holm-Sidak correction, right tables); p-values: ∗<0.05, ∗∗<0.005, ∗∗∗<0.0005, ns=not significant. **(i)** Limiting dilution analysis of mammosphere (MS)-forming cells of a representative E-SCC and an M-SCC; n=10 per generation; MS-forming cells frequency values and statistical comparisons (bottom right table); p-values: ∗<0.05, ∗∗<0.005, ∗∗∗<0.0005, ns=not significant.

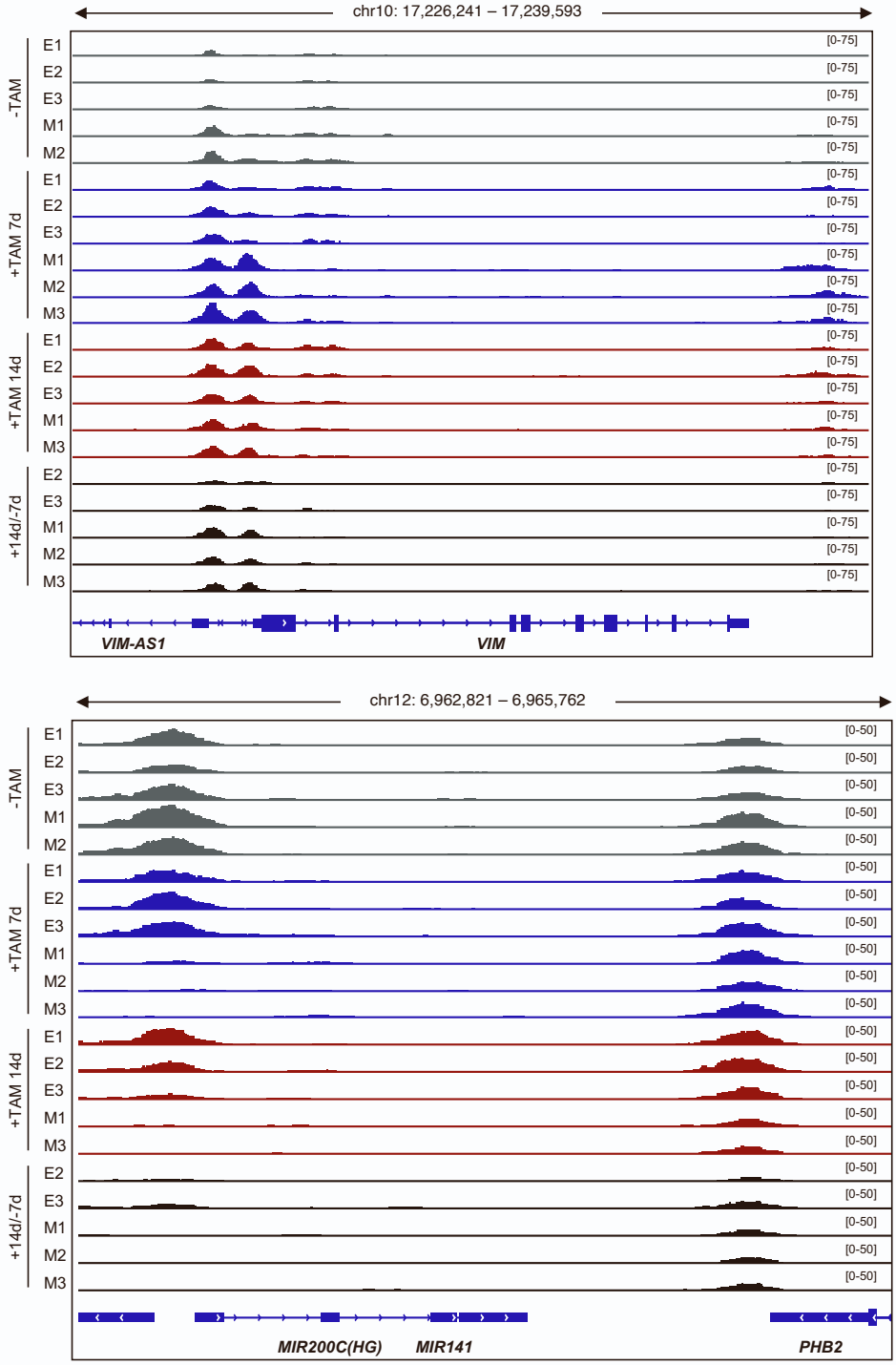
Supplementary Figure 4: Global chromatin changes upon transient stimulation define EMT resistance versus susceptibility







d

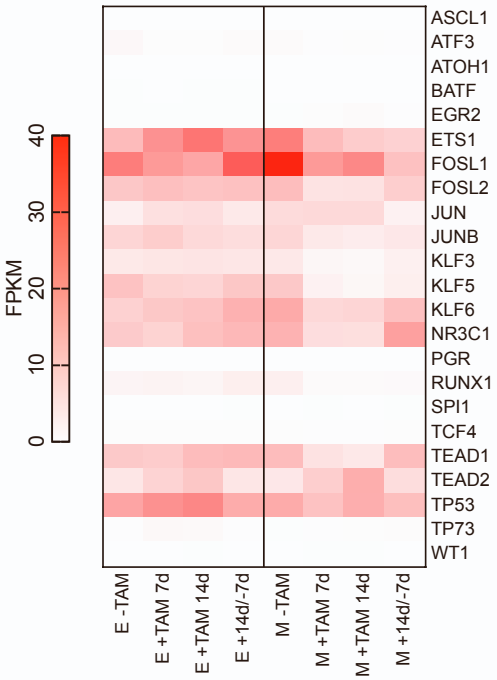


e

opening in M and E during +TAM (c3, c4)	Rank	Motif	Name	P-value	% of target sequences with motif	% of background sequences with motif
	1		TCF4	1E-370	55.74	0.2005
	2		Twist	1E-360	19.10	0.0174
	3		Fra1	1E-341	26.51	0.0438
	4		Atf3	1E-340	29.38	0.0558
	5		JunB	1E-323	26.14	0.0453

Opening exclusively in M during +TAM (c6, c7, c11, c9)	Rank	Motif	P-value	Best match rank1	Best match rank 2	% of target sequences with motif	% of background sequences with motif
	1		1E-2032	BATF	Atf3	44.97%	7.05%
	2		1E-580	TCF4	Ascl1	43.90%	19.39%
	3		1E-144	PGR	GRE	6.51%	1.78%
	4		1E-137	WT1	Egr2	19.13%	10.13%
	5		1E-121	ETS:RUNX	PU.1	10.27%	4.26%

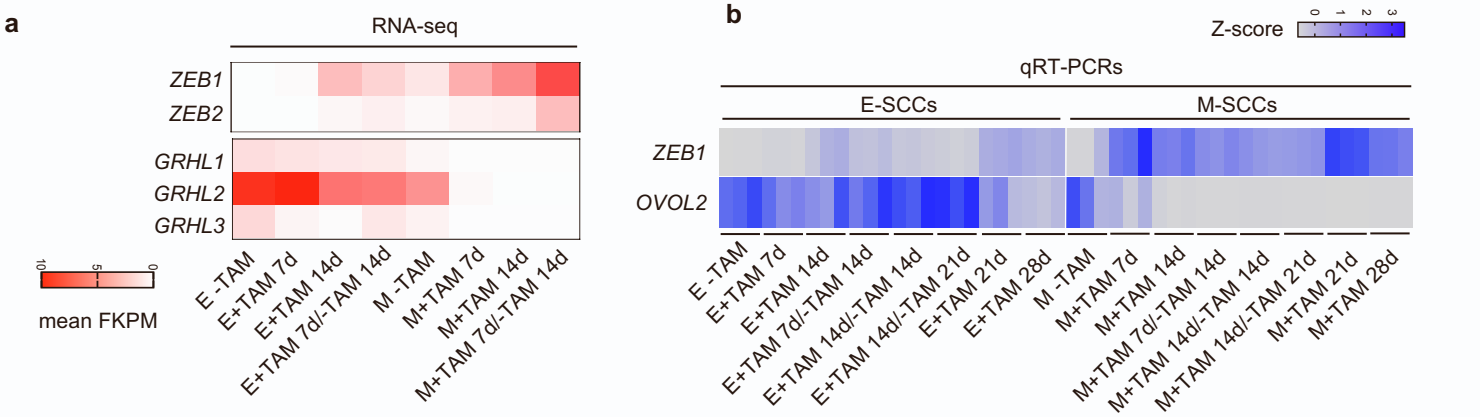
f



Supplementary Figure 4: Global chromatin changes upon transient stimulation define EMT resistance versus susceptibility, related to Figure 4

(a) PCA of RNA-seq data of E-SCCs (△) and M-SCCs (○)(n=3 E-SCCs, n=3 M-SCCs for all conditions). Each data point represents one SCC at the indicated time point. **(b)** Volcano plots of differential peak analysis referred to the ATAC-sequenced groups shown in Figure 4a; red-colored dots: *P*adj < 0.01 in M- *vs.* E-SCCs at each treatment; **(c)** ATAC-seq profile of the entire loci of *EPCAM* and *CDH1* in one representative M-SCC and one E-SCC. **(d)** ATAC-seq profile of the entire locus of indicated genesin M-SCCs and E-SCCs. **(e)** Top 5 hits of Homer transcription factor motif analysis of grouped clusters opening in M-SCCs and E-SCCs, or opening exclusively in M-SCCs during TAM treatment. **(f)** Heatmap of mean FPKM values of indicated genes of 3 E-SCCs and 3 M-SCCs.

Supplementary Figure 5: ZEB1-mediated mesenchymal re-programming is restrained by GRHL2 in E-SCCs



c

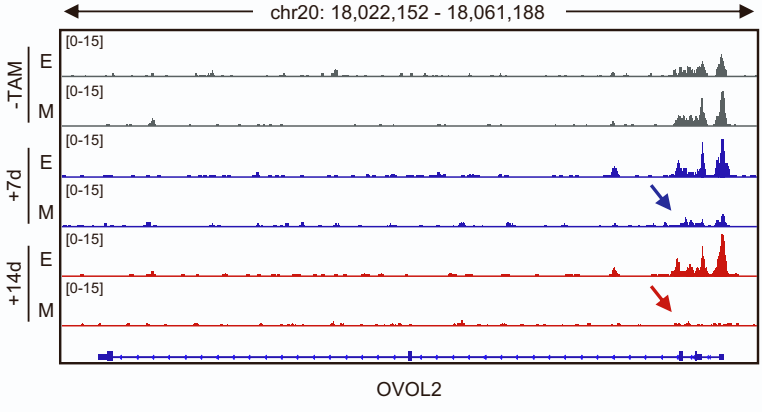
qRT-PCR statistical analysis

Group comparisons **E-SCCs (E) vs. M-SCCs (M)**

	Treatment	Mean of E	Mean of M	Adjusted P-Value	Significance
GRHL2	-TAM	2.13	1.14	0.002741	**
	+TAM 7d	2.40	0.43	0.000242	***
	+TAM 14d	2.97	0.05	0.000014	****
ESRP1	-TAM	1.89	1.06	0.001573	**
	+TAM 7d	1.75	0.31	0.000447	***
	+TAM 14d	1.92	0.03	0.000022	****
MIR 141	-TAM	38.95	23.12	0.081808	ns
	+TAM 7d	35.29	8.98	0.081808	ns
	+TAM 14d	18.19	1.13	0.013230	*
MIR 200A	-TAM	6.07	1.93	0.013197	*
	+TAM 7d	1.31	0.34	0.000809	**
	+TAM 14d	0.88	0.11	0.006563	*
MIR 200B	-TAM	931.00	983.00	0.919522	ns
	+TAM 7d	628.40	230.00	0.011765	*
	+TAM 14d	368.00	51.80	0.002827	**
MIR 200C	-TAM	561.30	651.00	0.790397	ns
	+TAM 7d	383.40	144.90	0.027436	*
	+TAM 14d	263.50	32.06	0.000205	***

	Treatment	Mean of E	Mean of M	Adjusted P-Value	Significance
ZEB1	-TAM	0.03	1.82	0.193799	ns
	+TAM 7d	0.50	17.94	0.075147	ns
	+TAM 14d	4.53	13.04	0.030886	*
	+TAM 14d/-TAM 7d	3.33	10.82	0.075147	ns
	+TAM 14d/-TAM 14d	2.17	9.31	0.034525	*
	+TAM 14d/-TAM 21d	1.29	9.27	0.057935	ns
	+TAM 21d	6.72	18.98	0.002704	**
	+TAM 28d	6.05	13.38	0.061117	ns
OVOL2	-TAM	2.31	1.73	0.289666	ns
	+TAM 7d	1.75	0.58	0.004929	**
	+TAM 14d	1.73	0.05	0.004214	**
	+TAM 14d/-TAM 7d	2.37	0.01	0.001423	**
	+TAM 14d/-TAM 14d	2.81	0.03	0.004929	**
	+TAM 14d/-TAM 21d	3.03	0.01	0.000116	***
	+TAM 21d	1.1270	0.0028	0.001136	**
	+TAM 28d	0.4970	0.0050	0.001258	**

e

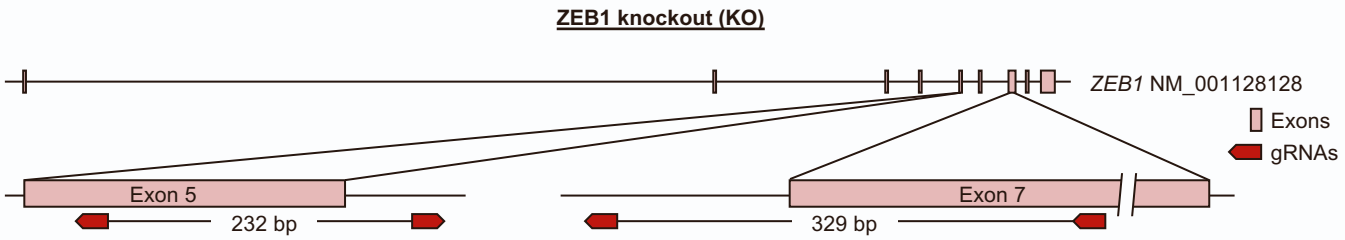


d

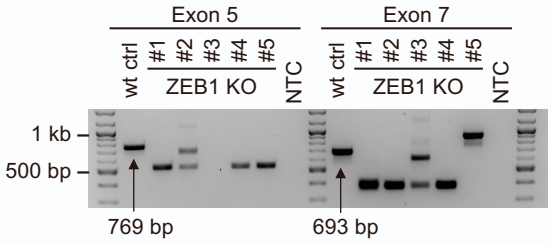
Individual clone comparisons **-TAM vs. +TAM**

	Clone	Mean of -TAM	Mean of +TAM 7d	P-Value	Significance
ZEB1	E1	0.02	0.62	0.348886	ns
	E2	0.03	0.55	0.320000	ns
	E3	0.03	0.33	0.350513	ns
	M1	0.33	13.33	0.293965	ns
	M2	0.28	14.61	0.265073	ns
	M3	4.86	25.87	0.362292	ns
OVOL2	E1	2.06	2.05	0.989301	ns
	E2	2.20	1.49	0.443035	ns
	E3	2.66	1.70	0.402908	ns
	M1	2.57	0.77	0.069714	ns
	M2	1.91	0.22	0.020098	*
	M3	0.71	0.77	0.939844	ns
MIR 141	M1	50.31	25.58	0.092874	ns
	M2	39.18	34.15	0.817601	ns
	M3	27.36	46.15	0.588773	ns
	M1	23.38	17.57	0.706404	ns
	M2	37.93	7.26	0.049538	*
	M3	8.05	2.12	0.228378	ns
MIR 200A	E1	8.96	1.62	0.081400	ns
	E2	5.68	1.18	0.160969	ns
	E3	3.58	1.13	0.006493	*
	M1	3.20	0.67	0.032790	*
	M2	1.74	0.21	0.007283	*
	M3	0.85	0.14	0.281506	ns
MIR 200B	E1	1164.00	686.30	0.486494	ns
	E2	867.70	553.40	0.543344	ns
	E3	761.40	645.40	0.598776	ns
	M1	1742.00	420.30	0.382930	ns
	M2	944.50	226.40	0.139996	ns
	M3	262.90	43.28	0.043840	*
MIR 200C	E1	646.00	424.90	0.560501	ns
	E2	524.70	334.60	0.565418	ns
	E3	513.00	390.80	0.544495	ns
	M1	1130.00	261.10	0.392072	ns
	M2	656.40	144.50	0.143916	ns
	M3	166.70	29.17	0.221568	ns
ESRP1	E1	1.84	3.14	0.031417	*
	E2	1.58	1.72	0.761870	ns
	E3	2.71	2.22	0.545828	ns
	M1	1.30	0.80	0.366251	ns
	M2	1.01	0.02	0.046849	*
	M3	0.78	0.53	0.497223	ns
GRHL2	E1	1.72	3.12	0.012621	*
	E2	1.95	1.62	0.621231	ns
	E3	2.71	2.48	0.774114	ns
	M1	1.52	1.20	0.552907	ns
	M2	0.94	0.05	0.010276	*
	M3	0.96	0.06	0.002929	**

f



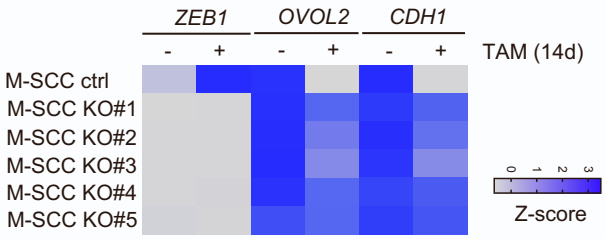
g



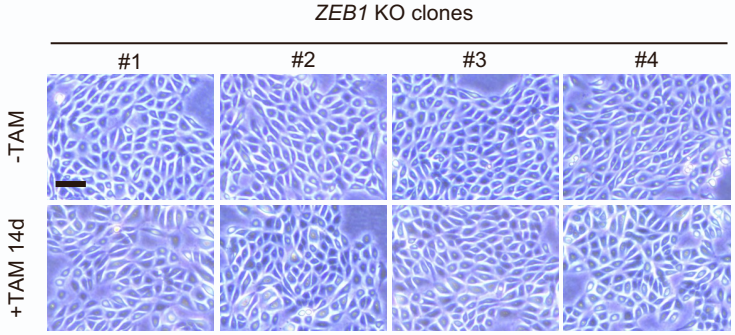
Expected PCR fragment sizes for *ZEB1* wildtype:

Exon 5: 769 bp Exon 7: 693 bp

i



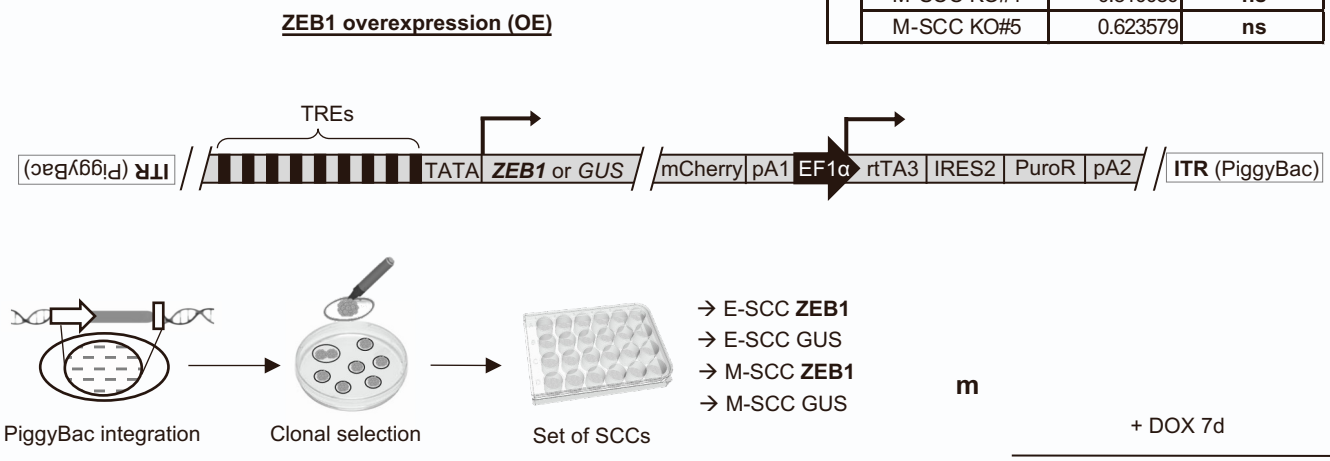
h



j

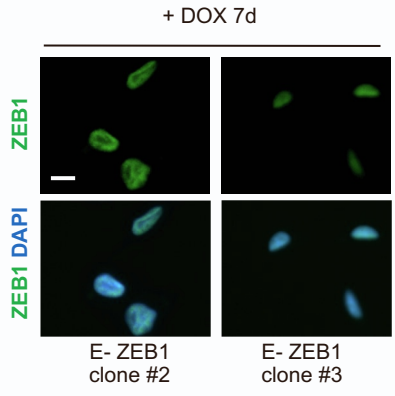
	Monoclonal culture	P-Value (-TAM vs. +TAM)	Significance
ZEB1	M-SCC ctrl	0.021734	*
	M-SCC KO#1	0.082183	ns
	M-SCC KO#2	0.354328	ns
	M-SCC KO#3	0.367749	ns
	M-SCC KO#4	0.329833	ns
	M-SCC KO#5	0.567278	ns
OVOL2	M-SCC ctrl	<0.000001	******
	M-SCC KO#1	0.059138	ns
	M-SCC KO#2	0.005860	**
	M-SCC KO#3	0.073351	ns
	M-SCC KO#4	0.272720	ns
	M-SCC KO#5	0.725486	ns
CDH1	M-SCC ctrl	0.000046	****
	M-SCC KO#1	0.490311	ns
	M-SCC KO#2	0.252772	ns
	M-SCC KO#3	0.012480	*
	M-SCC KO#4	0.810080	ns
	M-SCC KO#5	0.623579	ns

k

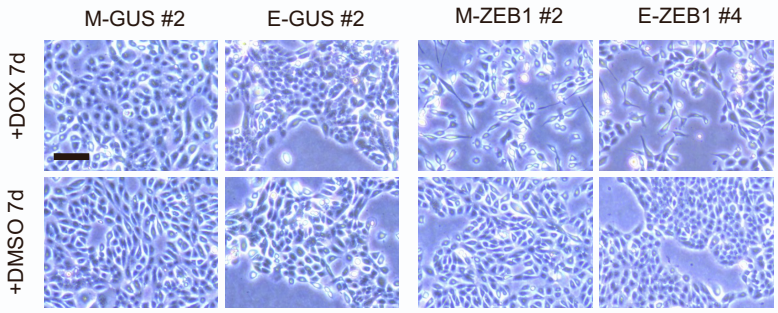


l

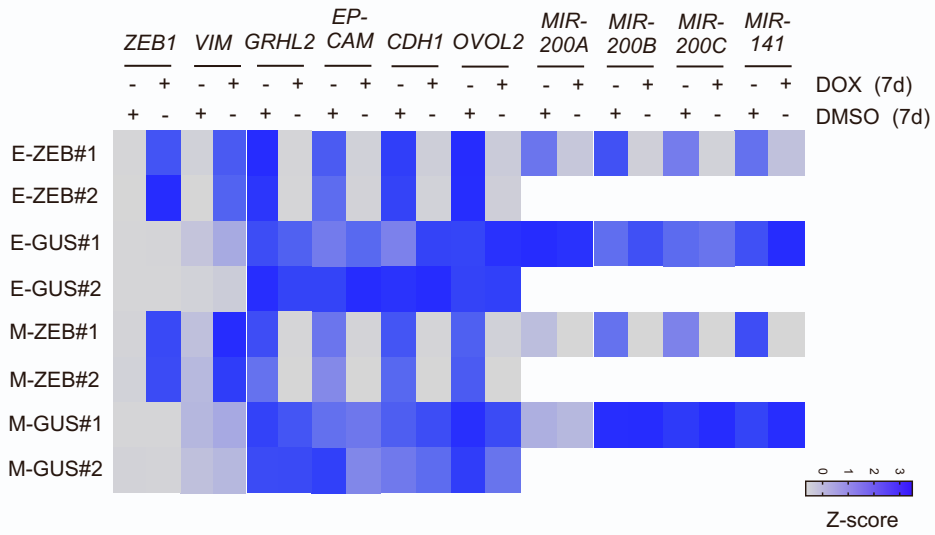
m



n



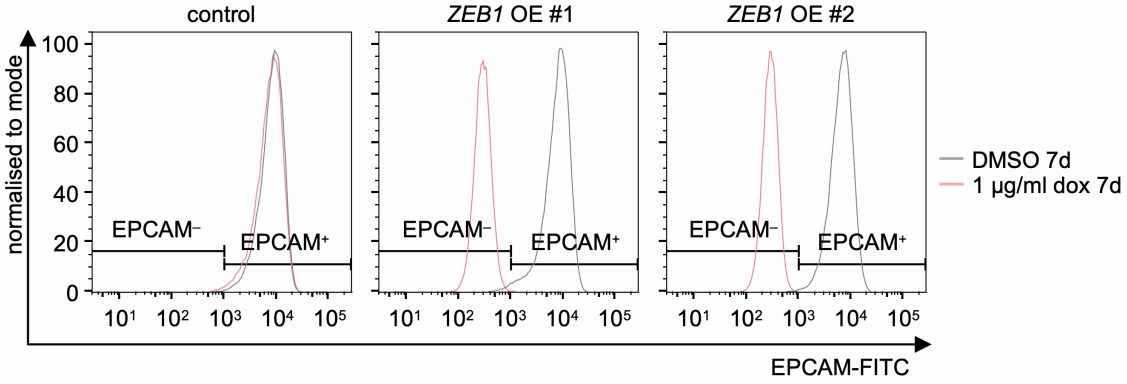
o



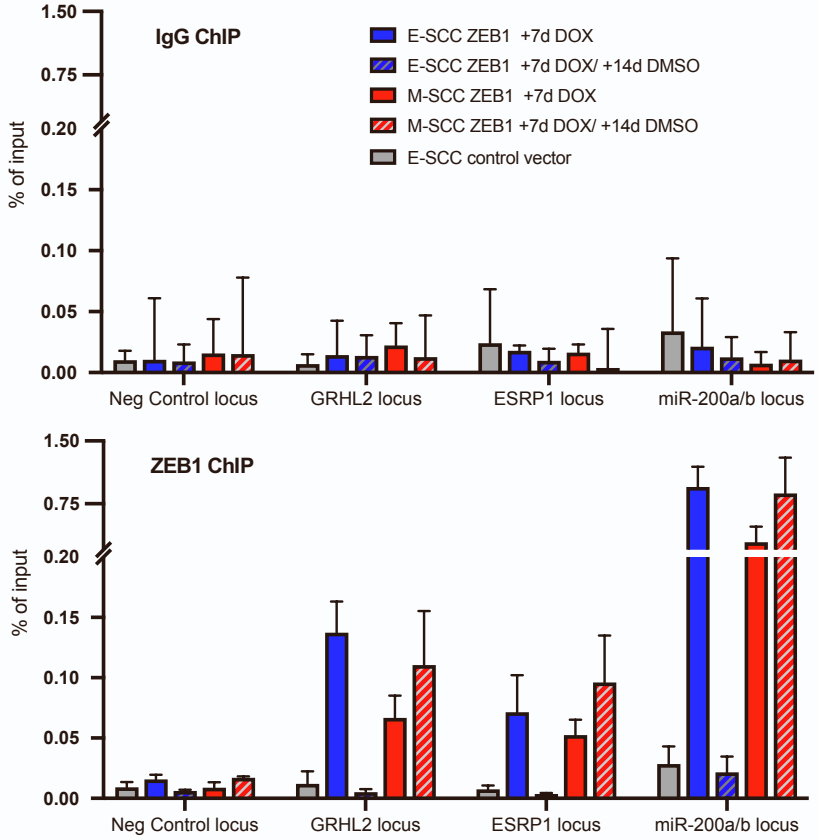
	Groups (2 to 4 monoclonal cultures each)	P-Value (DMSO vs. DOX)
ZEB1	E-SCC GUS	0.323674
	M-SCC GUS	0.822476
	E-SCC ZEB1	0.002247
	M-SCC ZEB1	0.055964
VIM	E-SCC GUS	0.223777
	M-SCC GUS	0.224942
	E-SCC ZEB1	0.000068
	M-SCC ZEB1	0.000178
GRHL2	E-SCC GUS	0.201344
	M-SCC GUS	0.603949
	E-SCC ZEB1	1.857E-10
	M-SCC ZEB1	0.000182
EPCAM	E-SCC GUS	0.634480
	M-SCC GUS	0.447814
	E-SCC ZEB1	0.000039
	M-SCC ZEB1	0.002387

	Groups (2 to 4 monoclonal cultures each)	P-Value (DMSO vs. DOX)
CDH1	E-SCC GUS	0.20134362
	M-SCC GUS	0.60394904
	E-SCC ZEB1	1.8574E-10
	M-SCC ZEB1	0.00018163
OVOL2	E-SCC GUS	0.40232975
	M-SCC GUS	0.09721229
	E-SCC ZEB1	9.9077E-09
	M-SCC ZEB1	0.00019806
MIR-200C	E-SCC GUS	0.68998081
	M-SCC GUS	0.83143974
	E-SCC ZEB1	0.04136063
	M-SCC ZEB1	0.18108969
MIR-141	E-SCC GUS	0.78791670
	M-SCC GUS	0.79218369
	E-SCC ZEB1	0.14416405
	M-SCC ZEB1	0.14104635

p



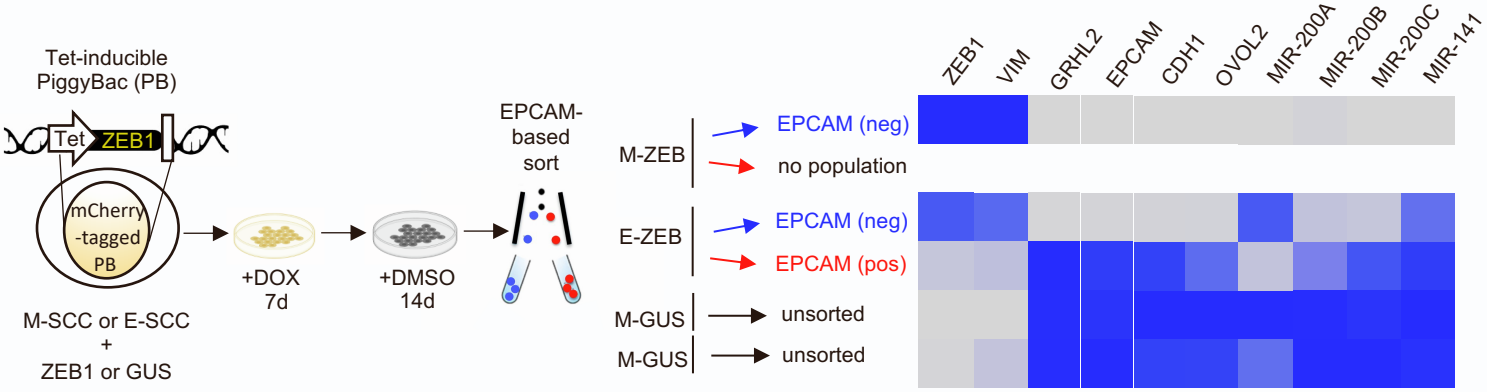
q



	Locus	Mean of E	Mean of M	P-Value	
+7d DOX-14d DMSO	GRHL2 (ZEB1 ChIP)	0.0047	0.1297	0.0244	*
	GRHL2 (IgG)	0.0145	0.0184	0.1345	ns
	ESRP1 (ZEB1 ChIP)	0.0050	0.0996	0.0012	**
	ESRP1 (IgG)	0.0097	0.0098	0.3954	ns
	miR-200 a/b (ZEB1 ChIP)	0.0161	0.8728	0.0014	**
	miR-200 a/b (IgG)	0.0144	0.0137	0.4678	ns
	Neg C. locus (ZEB1 ChIP)	0.0074	0.0187	0.0849	ns
	Neg.C. locus (IgG)	0.0104	0.0247	0.2887	ns

	Locus	Ctrl vector	+7d/-14d	P-Value	
E-SCC	GRHL2 (ZEB1 ChIP)	0.0175	0.0047	0.0663	ns
	GRHL2 (IgG ChIP)	0.0070	0.0145	0.0789	ns
	ESRP1 (ZEB1 ChIP)	0.0074	0.0050	0.2585	ns
	ESRP1 (IgG ChIP)	0.0242	0.0097	0.0532	ns
	miR-200 a/b (ZEB1 ChIP)	0.0285	0.0161	0.4517	ns
	miR-200 a/b (IgG)	0.0338	0.0144	0.1268	ns
	Neg. C. locus (ZEB1 ChIP)	0.0094	0.0074	0.4226	ns
	Neg. C. locus (IgG)	0.0102	0.0104	0.2697	ns
M-SCC	GRHL2 (ZEB1 ChIP)	0.0175	0.1297	0.0316	*
	GRHL2 (IgG)	0.0070	0.0184	0.0767	ns
	ESRP1 (ZEB1 ChIP)	0.0074	0.0996	0.0014	**
	ESRP1 (IgG)	0.0242	0.0098	0.1082	ns
	miR-200 a/b (ZEB1 ChIP)	0.0285	0.8728	0.0009	***
	miR-200 a/b (IgG)	0.0338	0.0137	0.0914	ns
	Neg. C. locus (ZEB1 ChIP)	0.0094	0.0187	0.1973	ns
	Neg. C. locus (IgG)	0.0102	0.0247	0.1736	ns

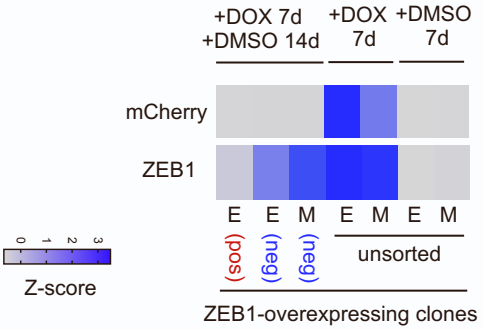
r



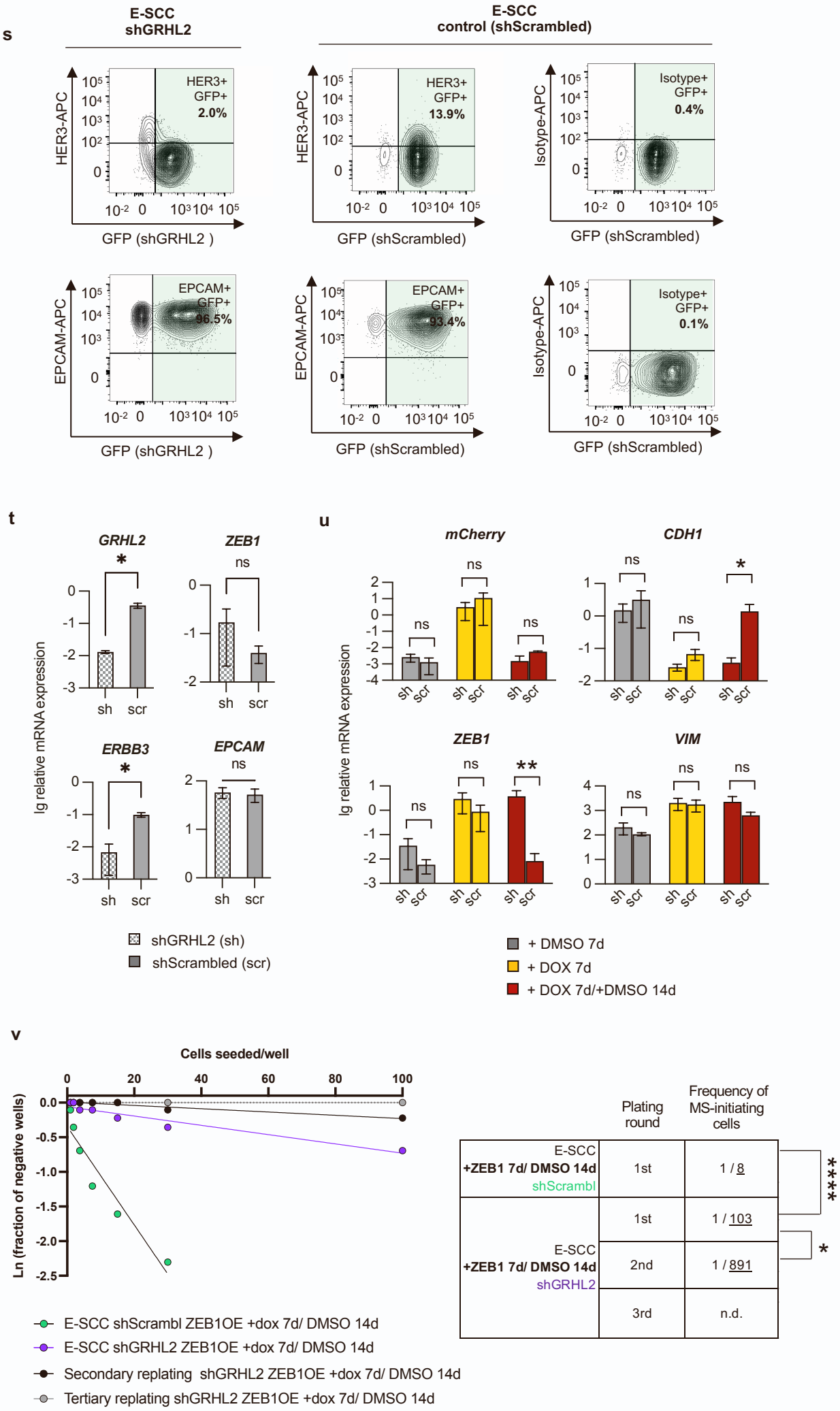
	Sorted EPCAM(neg)	Adjusted P-Value
M-GUS vs. M-ZEB	ZEB1	1.9957E-05
	VIM	0.00081742
	GRHL2	0.19533603
	CDH1	0.03932658

	Sorted EPCAM(neg)	Adjusted P-Value
E-GUS vs. E-ZEB	ZEB1	0.00120741
	VIM	0.000563
	GRHL2	0.00059827
	CDH1	0.00120741

	Sorted EPCAM(pos)	Adjusted P-Value
E-GUS vs. E-ZEB	ZEB1	0.17717128
	VIM	0.00295207
	GRHL2	0.85898063
	CDH1	0.65650587



+7d DOX vs. +7d/14d DMSO		
ZEB1	M (neg)	0.79145
mCherry	M (neg)	0.01474
ZEB1	E (neg)	0.21556
mCherry	E (neg)	0.01734
ZEB1	E (pos)	0.03796
mCherry	E (pos)	0.01681



	Plating round	Frequency of MS-initiating cells
E-SCC **+ZEB1 7d/ DMSO 14d** shScrambl	1st	1 / 8
	1st	1 / 103
E-SCC **+ZEB1 7d/ DMSO 14d** shGRHL2	2nd	1 / 891
	3rd	n.d.

****

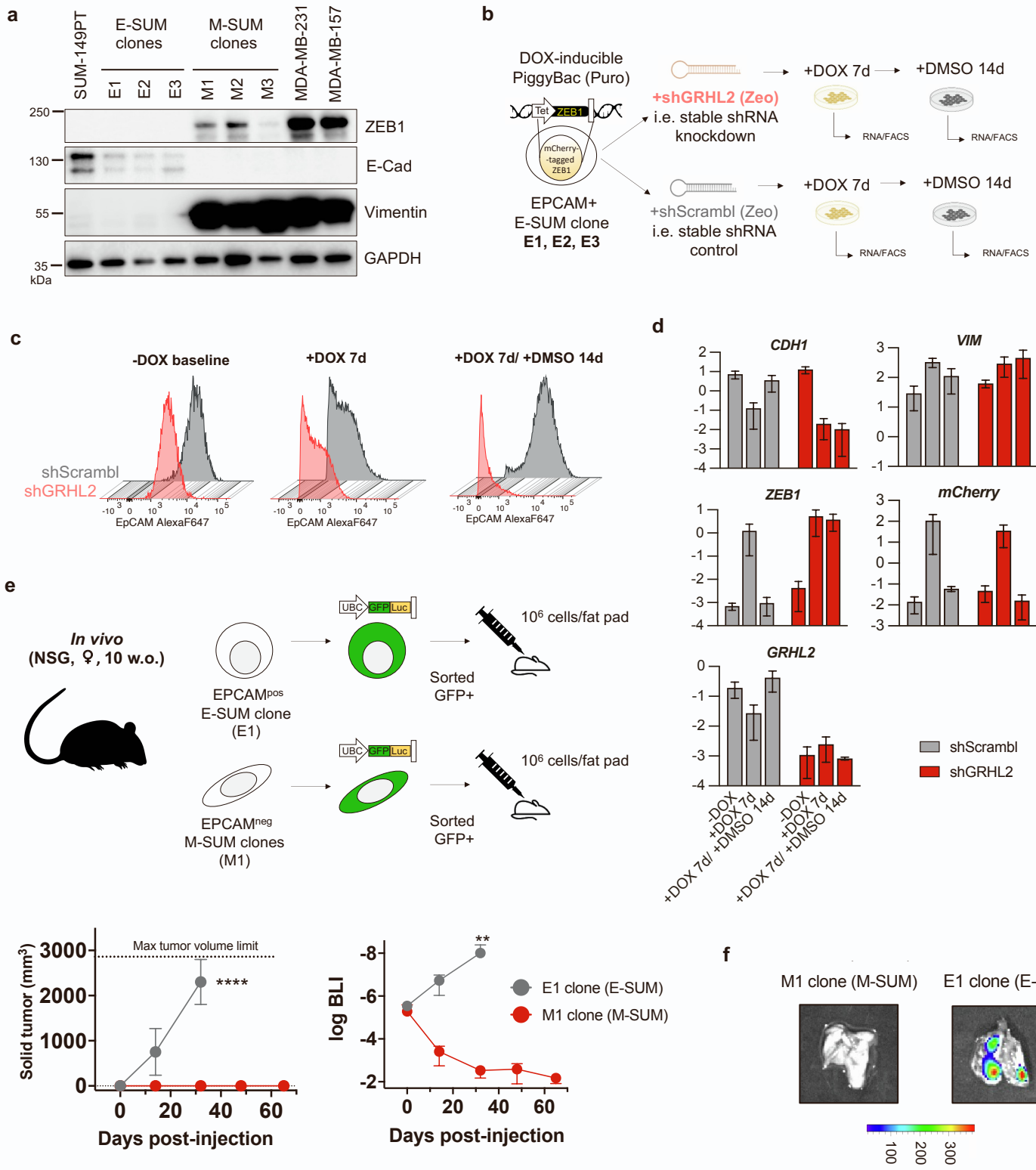
*

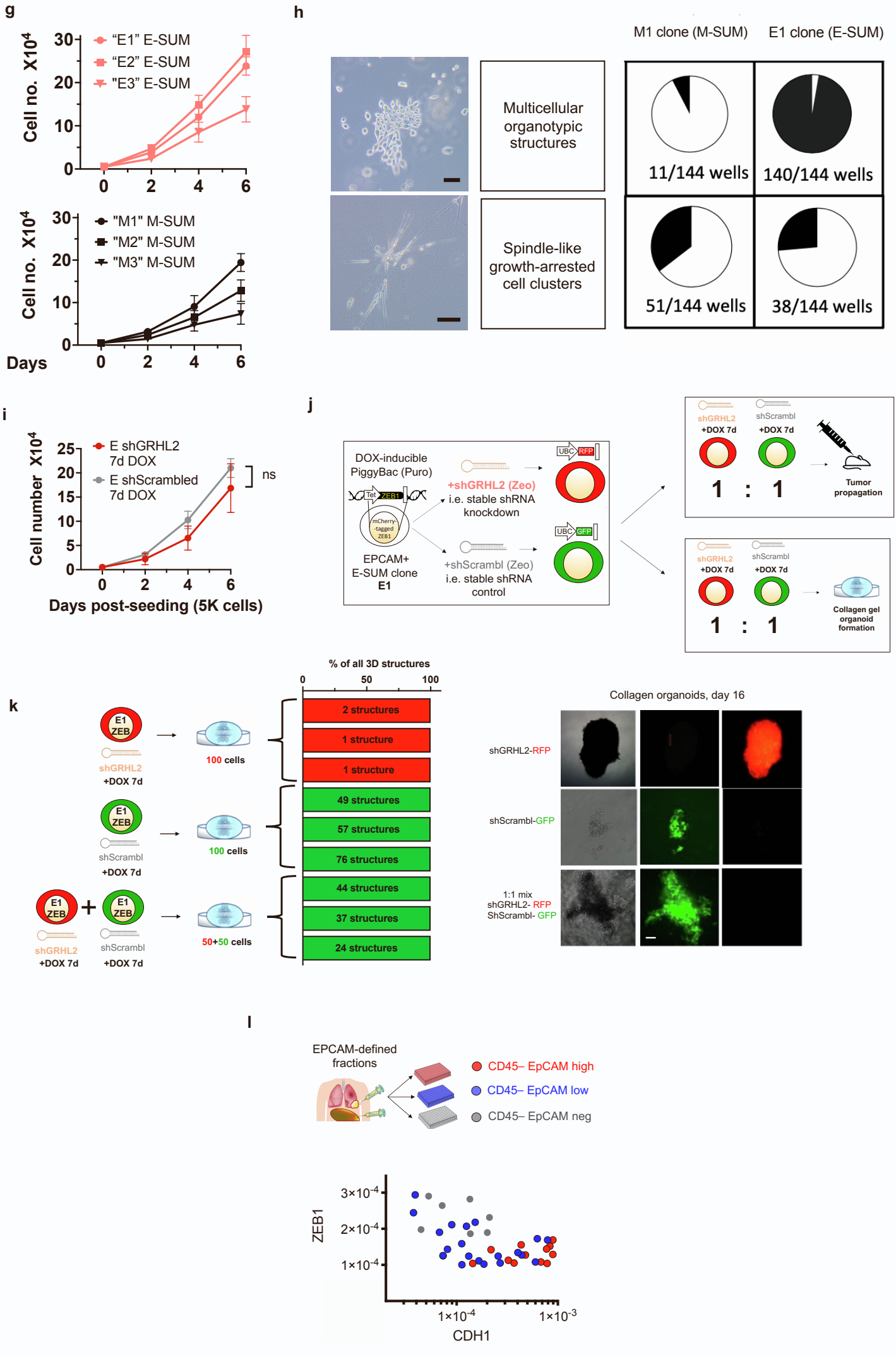
Supplementary Figure 5: ZEB1-mediated mesenchymal reprogramming is restrained by GRHL2 in E-SCCs, related to Figure 5

(a) Heatmaps of mean FPKM values of indicated genes of 3 E-SCCs and 3 M-SCCs.; **(b)** Heatmap of relative mRNA expression of indicated genes of 3 E-SCCs and 3 M-SCCs; n=3 replicates per clone; mean values are shown. **(c)** Statistical comparisons referring to Figure 5a (upper table) and to Supplementary Figure 5b (lower table); multiple t-tests (Holm-Sidak correction); p-values: ∗<0.05, ∗∗<0.005, ∗∗∗<0.0005, ns=not significant. **(d)** Statistical comparisons referring to Figure 5a; Student's t-tests; p-values: ∗<0.05, ∗∗<0.005, ∗∗∗<0.0005, ns=not significant. **(e)** ATAC-seq profile of the entire locus of *OVOL2* in one representative M-SCC and one E-SCC.**(f)** Schematic illustration of the genetic mapping of guide RNAs (gRNAs) for CRISPR/Cas9-mediated knock-out of the human *ZEB1* gene (NM00128128); expected deletions: 232 bp and 329 bp, respectively. **(g)** Evidence of successful knockout of the human *ZEB1* locus in M-SCCs; PCR products from the DNA regions targeted by gRNAs as in (f) are shown for an M-SCC control clone (wt ctrl) and for a set of 5 M-SCC *ZEB1* knockout clones; NTC, no template control. **(h)** Bright-field imagesof M-SCC *ZEB1* knockout clones (*ZEB1* KO #1-4); scale bar: 100 µm. **(i)** Heatmap of relative mRNA expression of indicated genesnof a M-SCC control clone (ctrl) and M-SCC *ZEB1* knockout clones (*ZEB1* KO #1-5). n=2 experiments; mean values are shown. **(j)** Statistical comparisons referring to Supplementary Figure 5i; n=3; Student's t-test. **(k)** Schematic depiction of the doxycycline inducible *ZEB1* PiggyBac construct. **(l)** Immunoblot for ZEB1 in E-SCCs and M-SCCs harboring inducible *GUS* (E-GUS, M-GUS) or *ZEB1* (E-ZEB, M-ZEB) Piggybac constructs. **(m)** Immunofluorescent staining for ZEB1 in two representative E-SCCs harboring an inducible ZEB1 construct; Scale bar: 5 µm. **(n)** Bright-field images of 4 representative clones in 2D culture, including one E-SCC and one M-SCC harboring an inducible *GUS* or *ZEB1* Piggybac construct. Scale bar: 100 µm. **(o)** (top) Heatmap of relative mRNA expression of indicated genes in 4 individual E-SCCs (E-ZEB-1, E-ZEB-2, E-GUS-1, E-

GUS-2) and 4 M-SCCs (M-ZEB-1, M-ZEB-2, M-GUS-1, M-GUS-2) harboring inducible *GUS* or *ZEB1* Piggybac; n=3 per clone; mean values are shown; (bottom) statistical comparisons referring to the top heatmap; Student's t-tests. **(p)** Flow cytometry for EPCAM of one representative SCC harboring inducible *ZEB1* Piggybac. **(q)** (left) ChIP-qPCR for ZEB1 recruitment in promoter regions of the indicated lociin a representative E-SCC and an M-SCC harboring an inducible *ZEB1* construct; n=2 anti-ZEB1 ChIPs; mean ± range; (right) statistical analysis performed on n=3 biological replicates for each condition; Student's t-tests. **(r)** (top right) Heatmap of relative mRNA expression of indicated genesin EPCAM-sorted fractions of 2 representative E-SCCs (E-ZEB-1, E-GUS-1) and 2 M-SCCs (M-ZEB-1, M-GUS-1) harboring inducible *GUS* or *ZEB1* Piggybac; mean value; (bottom) Heatmap of relative mRNA expression of indicated genes in EPCAM-sorted fractions of a representative E-SCCs and an M-SCC harboring inducible *ZEB1* Piggybac; n=3; mean values; Tables: statistical comparisons between the indicated groups; multiple t-tests (Holm-Sidak correction). **(s)** Flow cytometry for HER3, EPCAM, or matched isotype controls in a representative E-SCC transduced with GFP-expressing lentiviruses carrying shGRHL2 or shScrambled.. **(t)** Log relative mRNA expression of indicated genes in an E-SCC transduced with shGRHL2 (sh) or shScrambled (scr) and sorted for GFP; n=3; Mean ± SEM. **(u)** Log relative mRNA expression of indicated genes in an E-SCC harboring inducible *ZEB1* Piggybac and transduced with shGRHL2 (sh) or shScrambled (scr), sorted for GFP; n=3; Mean ± SEM. **(v)** Limiting dilution analysis of mammosphere (MS)-forming cells in an E-SCC harboring inducible *ZEB1* Piggybac and transduced with shGRHL2 or shScrambled; n=10 per generation; Table: MS-forming cells frequency values and statistical comparisons; p-values: $*<0.05$, $**<0.005$, $***<0.0005$, ns=not significant.

Supplementary figure 6: Clonal propagation *in vivo* requires resistance to EMT




g
Cell no. X10^4
"E1" E-SUM
"E2" E-SUM
"E3" E-SUM
"M1" M-SUM
"M2" M-SUM
"M3" M-SUM
Days
h
M1 clone (M-SUM)
E1 clone (E-SUM)
Multicellular organotypic structures
Spindle-like growth-arrested cell clusters
11/144 wells
140/144 wells
51/144 wells
38/144 wells
i
Cell number X10^4
E shGRHL2 7d DOX
E shScrambled 7d DOX
ns
Days post-seeding (5K cells)
j
DOX-inducible PiggyBac (Puro)
Tet
ZEB1
mCherry--tagged ZEB1
EPCAM+ E-SUM clone E1
+shGRHL2 (Zeo) i.e. stable shRNA knockdown
+shScrambl (Zeo) i.e. stable shRNA control
UBC
RFP
GFP
shGRHL2 +DOX 7d
shScrambl +DOX 7d
1 : 1
Tumor propagation
Collagen gel organoid formation
k
% of all 3D structures
2 structures
1 structure
1 structure
49 structures
57 structures
76 structures
44 structures
37 structures
24 structures
100 cells
100 cells
50+50 cells
Collagen organoids, day 16
shGRHL2-RFP
shScrambl-GFP
1:1 mix shGRHL2- RFP ShScrambl- GFP
l
EPCAM-defined fractions
CD45– EpCAM high
CD45– EpCAM low
CD45– EpCAM neg
ZEB1
CDH1

Supplementary Figure 6: Clonal propagation *in vivo* requires resistance to EMT, related to Figure 6. **(a)** Immunoblot for ZEB1 and EMT markers in E-SUM (E1-E3) and M-SUM (M1-M3) single-cell clones;reference cell lines: positive controls. **(b)** Experimental strategy for irreversible EMT induction in E-SUM clones. **(c)** Flow cytometry of EPCAM in a representative E-SUM. **(d)** Log relative mRNA expression of indicated genes in a representative E-SUM clone; n=2; Mean ± SEM. **(e)** Orthotopic tumorigenicity assay; solid tumor volume and BLI assessment over 9 weeks after MFP injection of E-SUM (*n*=3 clones) and M-SUM clones (*n*=3 clones) labeled as illustrated in the top panel; n=3 mice per group; mean ± SEM, unpaired Student's t-test with Welch's correction; p-values: ****<0.0001, **=0.0022. **(f)** BLI visualization of spontaneous lung metastases detected in representative mice orthotopically injected with E-SUM and M-SUM clones and sacrificed after 9 weeks; rainbow scale =photons/sec/cm^2/sr. **(g)** Growth curves of individual E-SUM and individual M-SUM clones in 2D cultures; n=4 replicates; mean ± SEM. **(h)** Representative bright-field images of one representative E-SUM clone (top image) and one M-SUM clone (bottom image) growing in collagen gels; Scale bar: 100 µm; Table: frequency distribution of the two recurrent growth patterns. **(i)** Growth curves of E-SUM clonesplated directly after 7 days of DOX stimulation; n=3 replicates per condition; mean ± SEM. **(j)** Experimental design referred to main Figure 6d and Supplementary Figure 6k. **(k)** (left panel) Organotypic assay of E-SUM clones in collagen gels, showing the number of organoids grown per each individual (n=3) replicate per condition; (right panel) representative images of organoids from the three conditions; Scale bar: 100 µm. **(l)** Transcript abundance of *ZEB1* and *CDH1* in single cells from two pleural effusions, including an ER-negative patient (Pat. 1) and an ER-positive patient (Pat. 5) assessed by single-cell qRT-PCR; values are presented as the multiplicative reciprocal of Ct-values.