

Supplementary Material for:

Defining the role of ZEB1 in the pathogenesis of lung cancer

Jill E. Larsen^{1,5,6}, Vaishnavi Nathan⁵, Jihan K. Osborne², Rebecca K. Farrow⁵, Dhruba Deb¹,
James P. Sullivan¹, Patrick D. Dospoy¹, Alexander Augustyn¹, Suzie K. Hight¹, Mitsuo Sato⁷,
Luc Girard^{1,2}, Carmen Behrens⁸, Ignacio Wistuba⁸, Adi F. Gazdar^{1,4}, Nicholas K. Hayward⁵ and
John D. Minna^{1,2,3}

¹Hamon Center for Therapeutic Oncology Research and the Simmons Comprehensive Cancer Center, Departments of ²Pharmacology, ³Internal Medicine, ⁴Pathology, The University of Texas Southwestern Medical Center, Dallas, Texas, USA; ⁵QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; ⁶School of Medicine, The University of Queensland, Brisbane, Queensland, Australia; ⁷Department of Respiratory Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan; ⁸Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Present affiliation for:

J. P. Sullivan, Verily Life Sciences, Mountain View, CA, USA 94043

J. K. Osborne, Children's Hospital Boston, Boston, MA, USA 02115

*Extended Materials and Methods**Reagents*

Recombinant active human TGF β 1 (PeproTech, Rocky Hill, NJ) was used at 5 ng/mL. The TGF β inhibitor SB431542 (Tocris Bioscience, Bristol, UK) was used at a final concentration of 10 μ M. TGF β monoclonal antibody (clone 1D11) and control antibody 13C4 monoclonal antibody were kind gifts from Rolf A. Brekken (University of Texas Southwestern Medical Center).

Viral vector construction and viral transduction

pRETRO-SUPER retroviral vector containing short hairpin sequences for ZEB1 (pSRP-shZEB1) (Table S9) was a kind gift from Dr. Thomas Brabletz, University of Freiburg, Germany (1). The ZEB1 ORF was digested from human ZEB1-pCIneo (a kind gift from Dr. Jennifer Richer, University of Colorado Health Sciences Center, USA) (2) with *XhoI* and *SmaI* and cloned into pMSCV-hyg (Clontech, Mountain View, CA) using *XhoI* and *HpaI* sites (designated pMSCV-ZEB1). Retrovirus containing medium were produced as described previously (3) and transduced cells selected with puromycin (1.0-1.5 μ g/mL) or hygromycin (20 μ g/mL) for 5-7 days. Stable knockdown of p53, and over-expression of mutant KRAS^{V12}, and MYC has been described previously (4).

Immunoblotting and ELISA

Preparation of total cell lysates and Western blotting were performed as described previously (4). Primary antibodies are listed in Table S8. Levels of secreted TGF β 1 were measured with the TGF β 1 EMAX Immunoassay ELISA kit (Promega, Madison, WI) according to the manufacturer's instructions.

siRNA assays

siRNA reverse transfections were performed as described previously (5) using 25 nM siRNA with Dharmafect 1 lipid (Dharmacon/Thermo Scientific, Waltham, MA). Cells were harvested 48 h post-transfection to seed in *in vitro* tumorigenicity assays. siRNA oligos were commercially validated and included positive (AllStars Hs Cell Death siRNA) and negative (AllStars Negative Control siRNA) controls (Qiagen, Hilden, Germany) (Table S9).

In vitro transformation assays

For anchorage-dependent colony formation assays, 200-600 viable cells were seeded per 100 mm plate in triplicate and cultured for two weeks before staining colonies with 0.5% methylene blue. Anchorage-independent (soft agar) growth assays were performed by seeding 200-1,000 viable cells in 12-well plates. Migration (scratch) assays were performed by either creating a wound with a sterile 200 μ L pipette tip or using an Essen IncuCyte Zoom in 96-well format. Cells were treated for 3 h with 1 μ g/mL Mitomycin C (Sigma-Aldrich, St Louis, MO, USA) before a wound was created in confluent cells, washed three times with 37°C PBS, then replaced with growth media (with and without supplements/FBS) and a phase photomicrograph image taken every 2 h until wound closed. Invasion assays were performed by either 24-well transwell assays (BD BioCoat Matrigel Invasion Chambers, BD Biosciences, Franklin Lakes, NJ) or using an Essen IncuCyte Zoom. BD BioCoat assays were performed as per manufacturer's guidelines seeding cells in serum- or supplement-free media (RPMI 1640 or KSFM without EGF and BPE supplements) in the upper chamber with serum- or supplement-containing media (RPMI 1640 with 5% FBS or KSFM with EGF and BPE supplements) in the lower chamber as the chemoattractant. Cells were allowed to migrate for 24-72 h then migrated cells were stained with 1% methylene blue/1% borax as per manufacturer's instructions. The mean number of

migrated cells was calculated from counting five microscopic fields at 20X magnification. Invasion assays using the Essen IncuCyte Zoom were performed in 96-well format as described above for the migration assays, seeding cells on 10% Matrigel® Growth Factor Reduced Basement Membrane Matrix (GFR-BMM) (Corning, Corning, NY) coated wells, and layering wounded cells with 30% GFR-BMM. Each condition was performed in triplicate.

CD24/CD44 and Aldefluor flow cytometric analysis

Fluorescence-activated cell sorting (FACS) was performed as previously described (6) (adapted from (7)) using dual-staining for CD24 and CD44 (Table S8) with propidium iodide exclusion of non-viable cells. ALDH activity was measured using the Aldefluor kit (Stem Cell Technologies Inc., Vancouver, Canada) as described previously (6).

ChIP Assay

Confluent cells were fixed at room temperature in 1% (v/v) formaldehyde for 8 min with gentle agitation. Fixation was stopped with 125 mM glycine with gentle agitation for 2 min at room temperature. Fixed cells were washed three times in ice cold PBS containing fresh PMSF (Thermo Scientific, 1 µL/mL), then resuspended in ChIP lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0 and protease cocktail). Chromatin was sheared by sonication using Bioruptor 200 then diluted 10-fold in dilution buffer (0.1% Triton, 20 mM Tris-HCl, 2 mM EDTA). Lysates were pre-cleared for one hour. Antibodies against ZEB1 (Table S8) and lysates were incubated overnight at 4°C with 35 µL of protein A/G beads. Beads were washed and protein-DNA complexes eluted (1% SDS, 100 mM NaHCO₃), then cross-links were reversed. DNA was extracted by phenol-CHCl₃. Primers listed in Table S9.

In vivo tumorigenicity assays and histologic analysis

Subcutaneous xenograft growth was evaluated by a subcutaneous injection of $3-5 \times 10^6$ viable cells in 0.2 mL of PBS into the flank of female 5- to 6-week-old NOD/SCID mice as previously described (4). Intra-venous xenograft growth was evaluated by tail vein injection of 1×10^6 viable cells in 0.2 mL of PBS as previously described (8). All animal care was in accord with institutional guidelines and approved Institutional Animal Care and Use Committee (IACUC) protocols. Formalin-fixed, paraffin-embedded (FFPE) xenograft tumor tissue was sectioned and stained with hematoxylin and eosin (H&E).

RNA isolation and qRT-PCR of mRNA expression

RNA isolation and quantitative Reverse Transcription PCR (qRT-PCR) were performed as previously described (6) using validated Taqman primers and probes (Applied Biosystems, Foster City, CA). Relative expression was calculated using the $2^{-\Delta\Delta CT}$ method by comparing to *GAPDH* or, for the analysis of 275 primary lung adenocarcinoma and squamous cell carcinomas (SPORE-275), using a global normalization method (DataAssist v3.01, Applied Biosystems).

Microarray analysis

mRNA microarray analysis was performed as described previously (4). HCC-NSCLC (Hamon Cancer Center 78 NSCLC cell lines) cell lines were profiled with Illumina® HumanWG6 v3 Expression BeadChips (Illumina Inc.). All HBEC cell lines were profiled with Illumina® HumanHT-12 v4 Expression BeadChips (Illumina Inc.). Probe intensity and detection data were obtained using Illumina® GenomeStudio Data Analysis Software 2011.1, and further processed with GeneSpring™ GX 13.1.1 software (Agilent Technologies, Santa Clara, CA, USA). Functional analysis of differentially expressed genes was conducted using Ingenuity Pathway Analysis (IPA; Ingenuity Systems, Inc.). The data discussed in this publication have been made

available in the National Center for Biotechnology Information's Gene Expression Omnibus (GEO) public repository (<http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO Series accession number GSE77925. To identify candidate genes that demonstrate mRNA expression correlates with ZEB1 we compared ZEB1-associated genes across seven independent mRNA datasets comprising isogenic HBEC cell lines, lung cancer cell lines and primary lung adenocarcinomas to identify commonly up- or down-regulated genes. The seven datasets are outlined in Table S5.

References

1. Spaderna S, Schmalhofer O, Wahlbuhl M, Dimmler A, Bauer K, Sultan A, Hlubek F, Jung A, Strand D, Eger A, et al. The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. *Cancer Res.* 2008;68(2):537-44.
2. Singh M, Spoelstra NS, Jean A, Howe E, Torkko KC, Clark HR, Darling DS, Shroyer KR, Horwitz KB, Broaddus RR, et al. ZEB1 expression in type I vs type II endometrial cancers: a marker of aggressive disease. *Mod Pathol.* 2008;21(7):912-23.
3. Sato M, Vaughan MB, Girard L, Peyton M, Lee W, Shames DS, Ramirez RD, Sunaga N, Gazdar AF, Shay JW, et al. Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer Res.* 2006;66(4):2116-28.
4. Sato M, Larsen JE, Lee W, Sun H, Shames DS, Dalvi MP, Ramirez RD, Tang H, Dimaio JM, Gao B, et al. Human Lung Epithelial Cells Progressed to Malignancy through Specific Oncogenic Manipulations. *Molecular cancer research.* 2013;11(6):638-50.
5. Greer RM, Peyton M, Larsen JE, Girard L, Xie Y, Gazdar AF, Harran P, Wang L, Brekken RA, Wang X, et al. SMAC mimetic (JP1201) sensitizes non-small cell lung cancers to multiple chemotherapy agents in an IAP-dependent but TNF-alpha-independent manner. *Cancer Res.* 2011;71(24):7640-8.
6. Sullivan JP, Spinola M, Dodge M, Raso MG, Behrens C, Gao B, Schuster K, Shao C, Larsen JE, Sullivan LA, et al. Aldehyde dehydrogenase activity selects for lung adenocarcinoma stem cells dependent on notch signaling. *Cancer Res.* 2010;70(23):9937-48.
7. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, and Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A.* 2003;100(7):3983-8.
8. Osborne JK, Larsen JE, Shields MD, Gonzales JX, Shames DS, Sato M, Kulkarni A, Wistuba II, Girard L, Minna JD, et al. NeuroD1 regulates survival and migration of neuroendocrine lung carcinomas via signaling molecules TrkB and NCAM. *Proc Natl Acad Sci U S A.* 2013;110(16):6524-9.
9. Gemmill RM, Roche J, Potiron VA, Nasarre P, Mitas M, Coldren CD, Helfrich BA, Garrett-Mayer E, Bunn PA, and Drabkin HA. ZEB1-responsive genes in non-small cell lung cancer. *Cancer Lett.* 2011;300(1):66-78.
10. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, Misek DE, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med.* 2008;14(8):822-7.
11. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014;511(7511):543-50.
12. Hammerman PS, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, Stojanov P, McKenna A, Lander ES, Gabriel S, et al. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012;489(7417):519-25.

Supplementary Figures

Figure S1: Expression of ZEB1 and EMT markers in MYC-transformed derivatives of HBEC2^{p53,KRAS} and HBEC17^{p53,KRAS}. Mean±SD from three replicate experiments.

Supplementary Figure 1

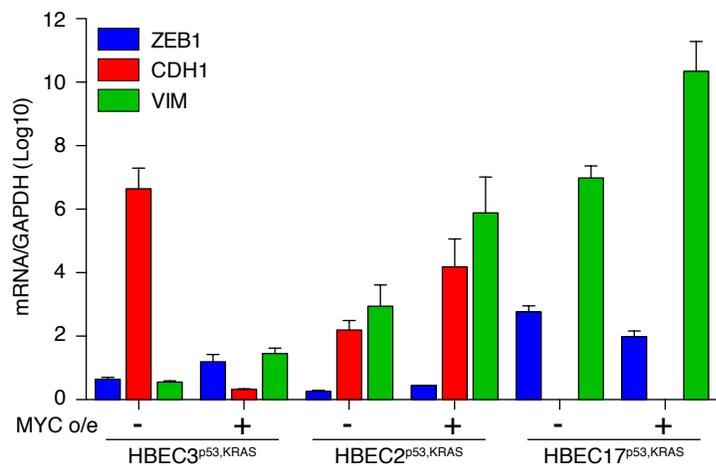


Figure S2: Anchorage-dependent (liquid) colony forming ability showing acute response of HBEC2 and HBEC30 manipulated with p53 knockdown and mutant KRAS^{V12} following TGF β treatment. Representative image from three replicate experiments.

Supplementary Figure 2

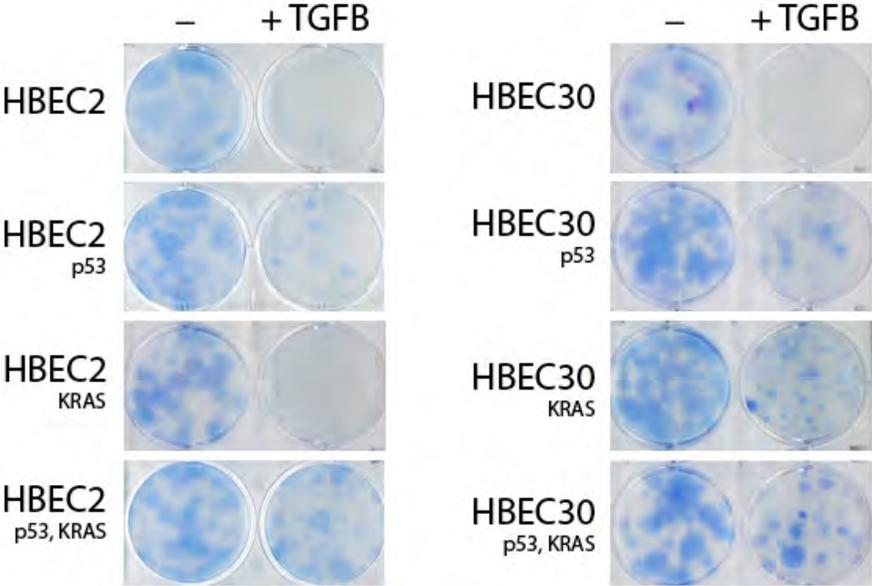


Figure S3: A) Representative phase micrographs of HBEC3^{pMSCV} and HBEC3^{ZEB1} in a scratch assay. B) Representative phase micrographs (20X magnification) of invaded HBEC3^{pMSCV} and HBEC3^{ZEB1} cells following Matrigel invasion assay. Panels are representative data of at least three independent experiments.

Supplementary Figure 3

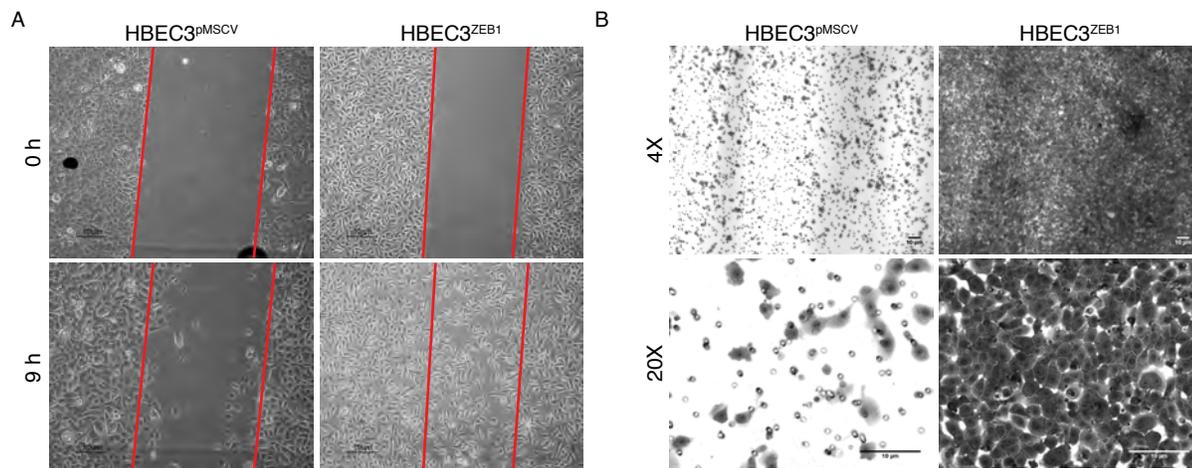


Figure S4: A) Phase micrographs of HBEC3^{p53,KRAS}+TGFβ and HBEC3^{p53,KRAS,MYC} following 14 day treatment with SB431542. B) Phase micrographs of HBEC3^{p53,KRAS}+TGFβ and HBEC3^{p53,KRAS,MYC} following treatment with TGFβ monoclonal antibody (1D11) compared with an isotype control antibody (13C4). C) Immunoblot of EMT markers in HBEC3^{p53} HBEC3^{p53,KRAS}+TGFβ and HBEC3^{p53,KRAS,MYC} following treatment with TGFβ monoclonal antibody (1D11) or an isotype control antibody (13C4). β-tubulin was used as loading control. D) Significantly over-represented canonical pathways based upon mRNA expression of HBEC3^{p53,KRAS,MYC} compared with HBEC3^{p53,KRAS}. UnTx, Untreated. Panels A-C are representative data of at least three independent experiments.

Supplementary Figure 4

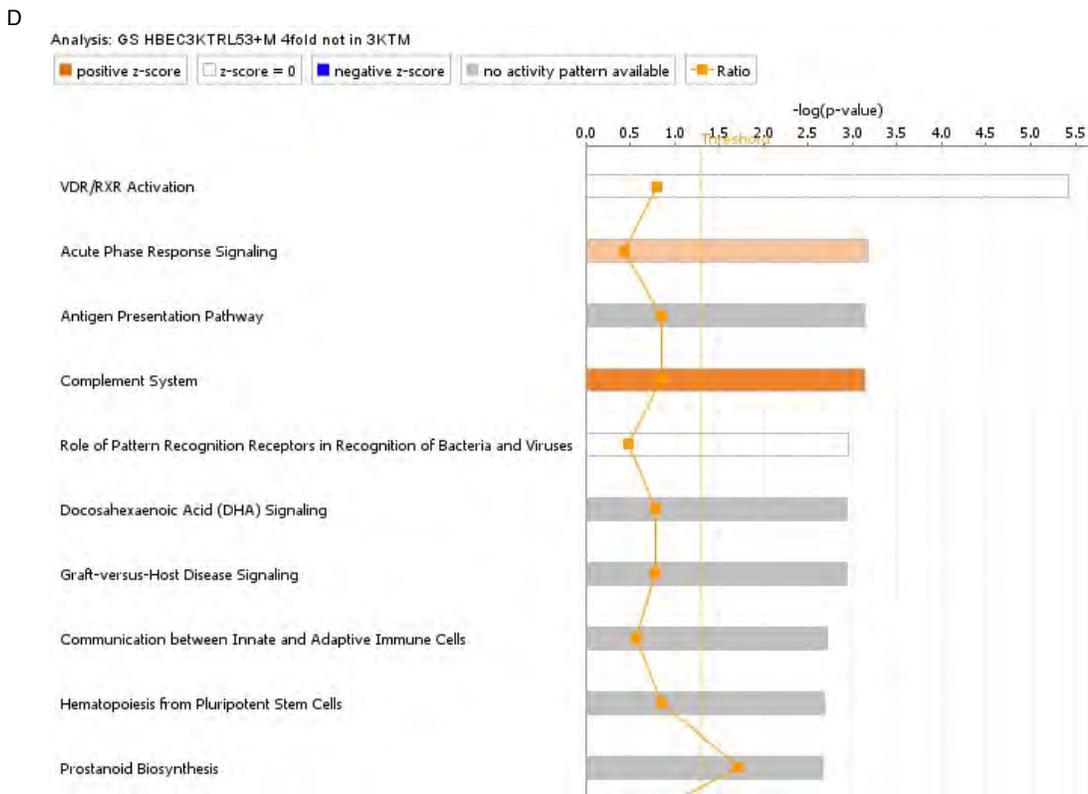
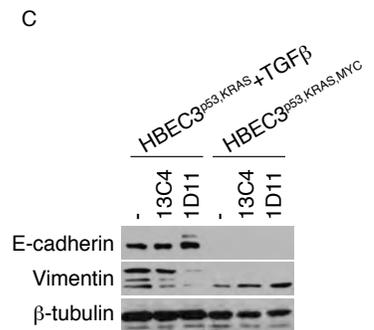
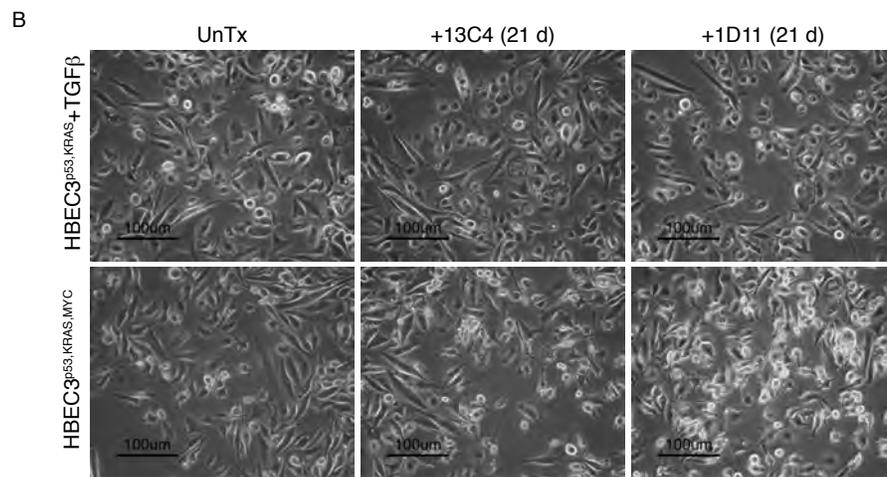
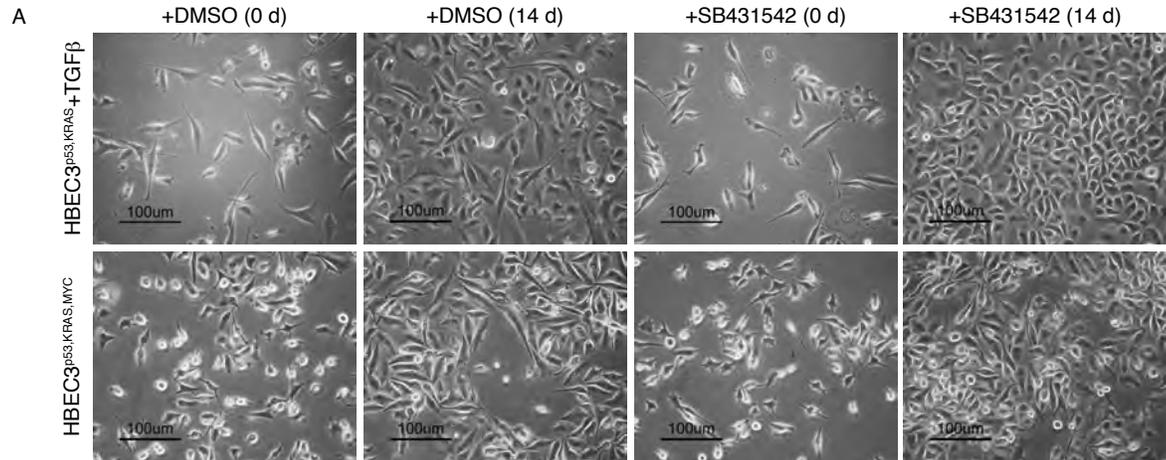


Figure S5: A-B) mRNA expression (A) and immunoblot (B) of HBEC3^{p53,KRAS}+TGF β following siRNA-mediated knockdown of ZEB1. C) Matrigel invasion in HBEC3^{p53,KRAS}+FBS-c1 and representative phase micrographs (20X magnification) following transient siRNA-mediated knockdown with siNTC (non-targeting) or ZEB1 following Matrigel invasion assay. pMSCV, vector control. HSP90 was used as loading control. UnTx, Untreated; siNTC, non-targeting control siRNA. *P* values were obtained by one-way ANOVA (panel C). Data presented as mean \pm SD, *n* = 3. Panels are representative data of at least three independent experiments. **P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P* < 0.0001.

Supplementary Figure 5

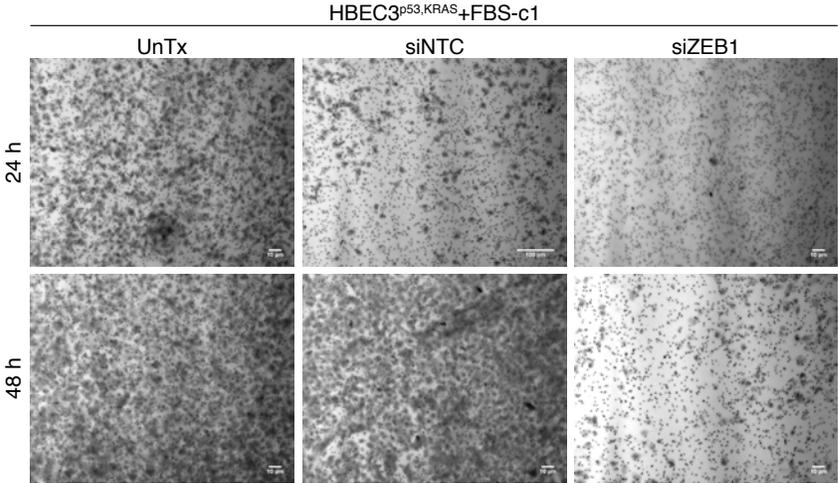
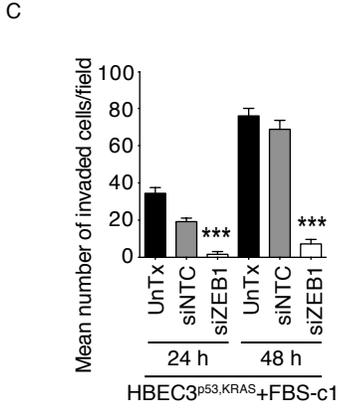
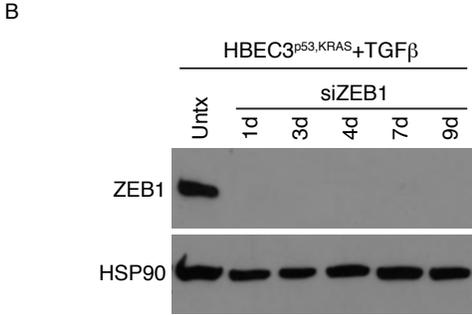
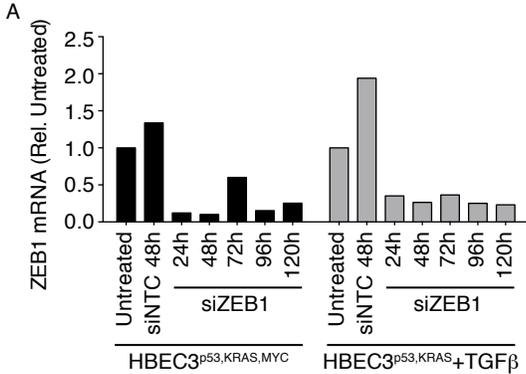


Figure S6: A-B) mRNA expression of ZEB1 in 79 NSCLC and 19 SCLC cell lines stratified by A) derived tumor site and histological subtype and B) mutation status of genes relevant to lung cancer. C) Immunoblot of lung cancer cell lines following stable knockdown of ZEB1 using shRNA. D) qRT-PCR analysis of *ZEB1* and *CDH1* mRNA expression in lung cancer cell lines following stable knockdown of ZEB1. E) Cellular proliferation growth rates of lung cancer cell lines following stable knockdown of ZEB1 (mean±SD). F) Subcutaneous xenograft growth in NOD/SCID mice of NCI-H1792 following ZEB1 knockdown (mean±SD). G) Bioluminescence imaging of colonizing ability of Calu1-Luc following i.v. injection into NOD/SCID mice. pSRP, vector control; RLU, relative light units. HSP90 was used as loading control. *P* values were obtained by one-way ANOVA (panels A, B, D) and a non-linear regression model (panels E, F). Data presented as mean ± SD, *n* = 3. Panels are representative data of at least three independent experiments. **P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P* < 0.0001.

Supplementary Figure 6

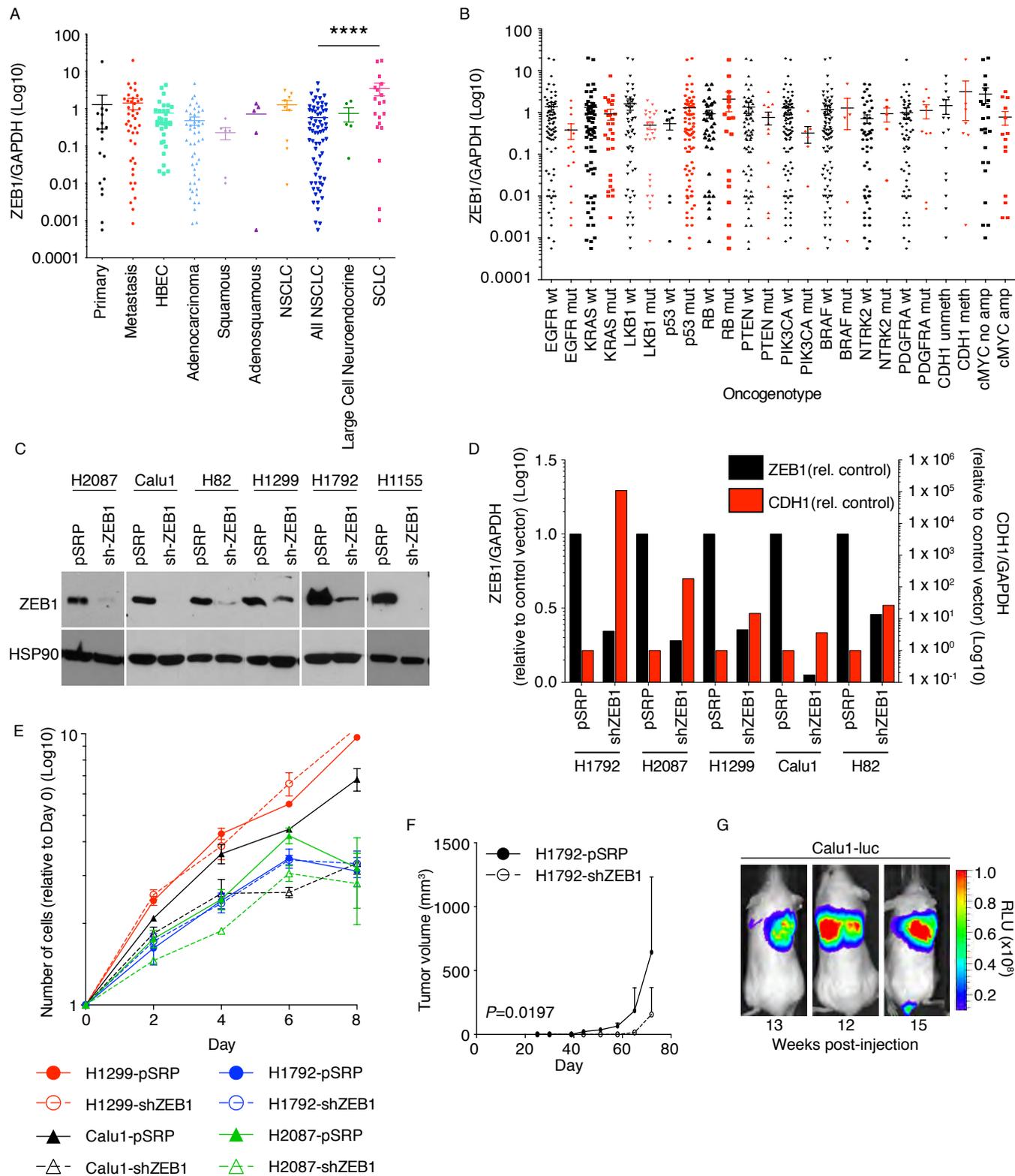
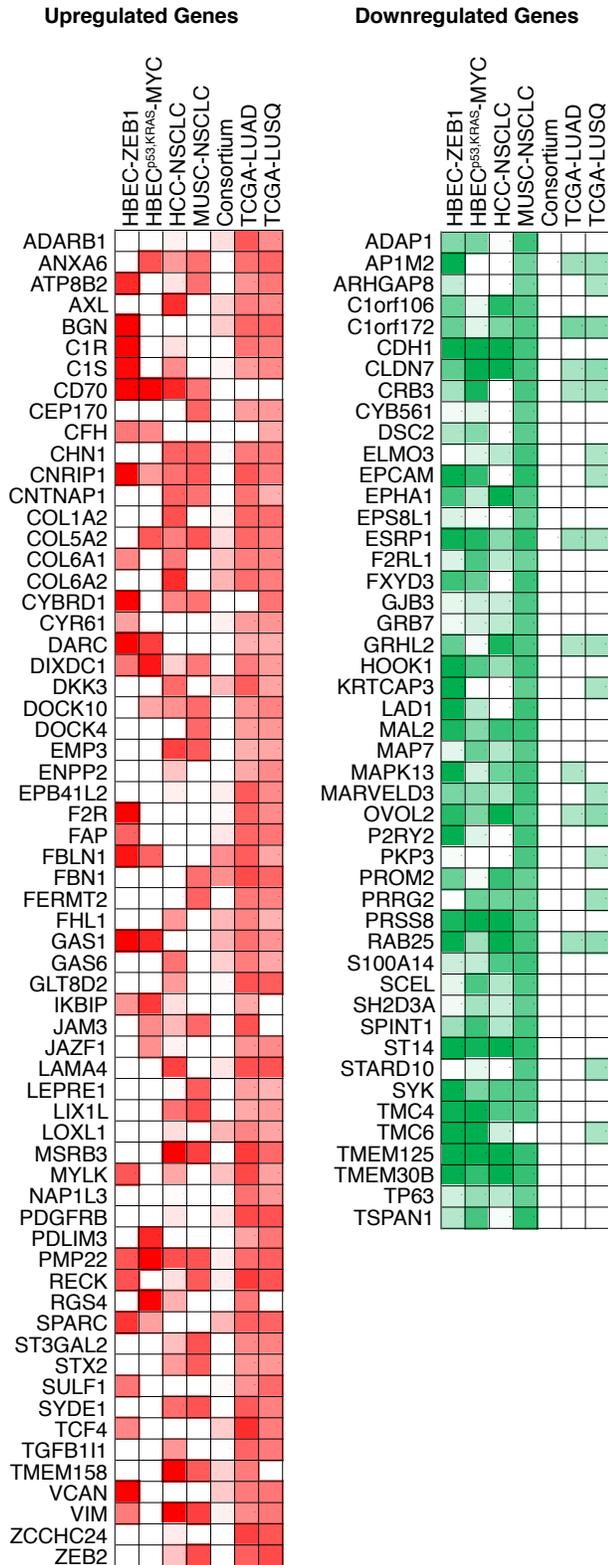


Figure S7: A) ZEB1-associated genes identified from comparing seven independent mRNA microarray datasets. Red and green indicate an increase or decrease, respectively, in gene expression in relation with high *ZEB1* expression where color intensity indicates the level of expression. B) Supervised clustering of HBEC3 derivatives with the 110 ZEB1-associated genes. Supervised hierarchical two-dimensional clustering of 10 HBEC3 cell lines (Euclidean similarity, Wards linkage analysis) with a filtered list of xxx probes representing the 110 ZEB1-associated genes. Each column is a sample, and each row is a gene. Heat map indicates level of gene expression; red, high expression; blue, low expression. EMT status of cell lines indicated beneath heat map; green, epithelial-like; yellow, mesenchymal-like.

Supplementary Figure 7

A



B

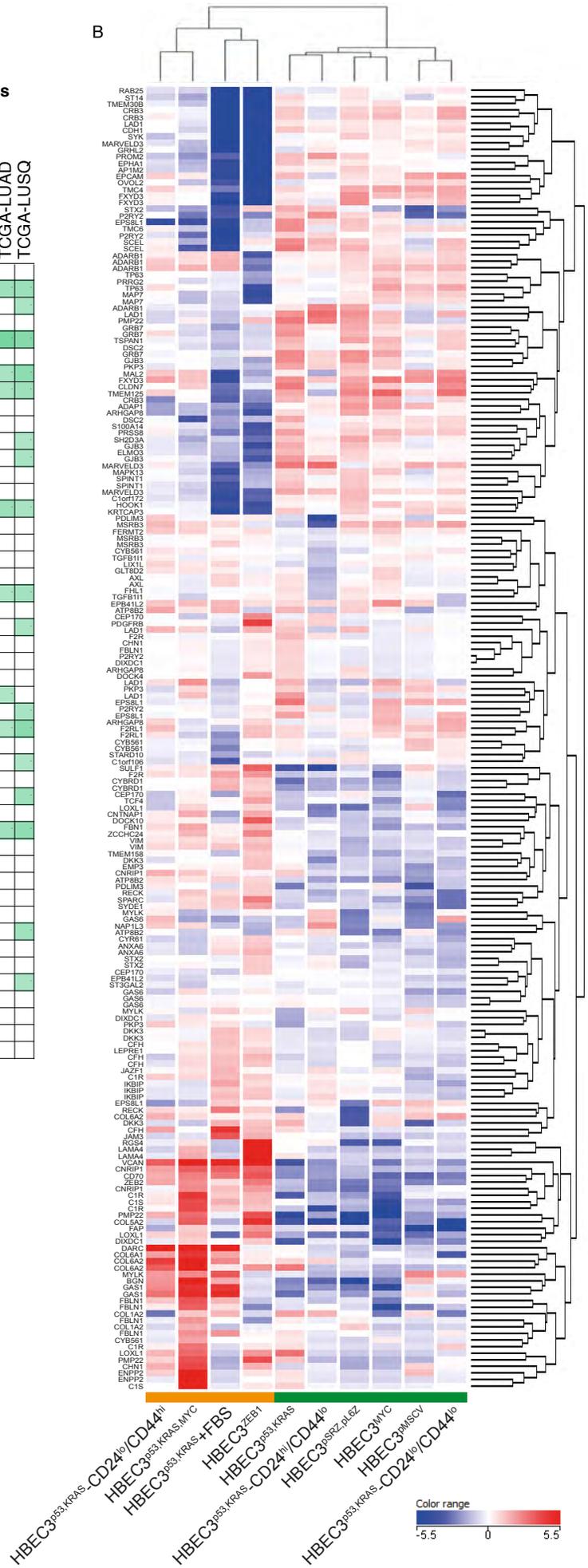


Figure S8: A) *ZEB1* mRNA expression in 267 primary, resected NSCLC tumors with a smoking history, stratified by TNM stage. B-I) Kaplan-Meier analysis (log-rank) of 267 primary, resected NSCLC tumors stratified by median *ZEB1* (B-C), *ZEB2* (D-E), *CDH1* (F-G), and *ESRP1* (H-I) expression in relation to overall survival (B,D,F,H) and cancer-free survival (C,E,G,I). J-K) Kaplan-Meier analysis of primary, resected NSCLC tumors stratified by median *SNAI1* expression in relation to overall survival in adenocarcinomas (n=183) (J) and squamous cell carcinomas (n=80) (K). Tick marks, patients whose data were censored at last follow-up. OS(yr), overall survival (years); CaFS(yr), cancer-free survival (years). *P* values were obtained by one-way ANOVA (panel A) and Kaplan-Meier analysis (log-rank) (panels B-K). Data presented as mean \pm SD (panel A), *n* = 267. **P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P* < 0.0001.

Supplementary Figure 8

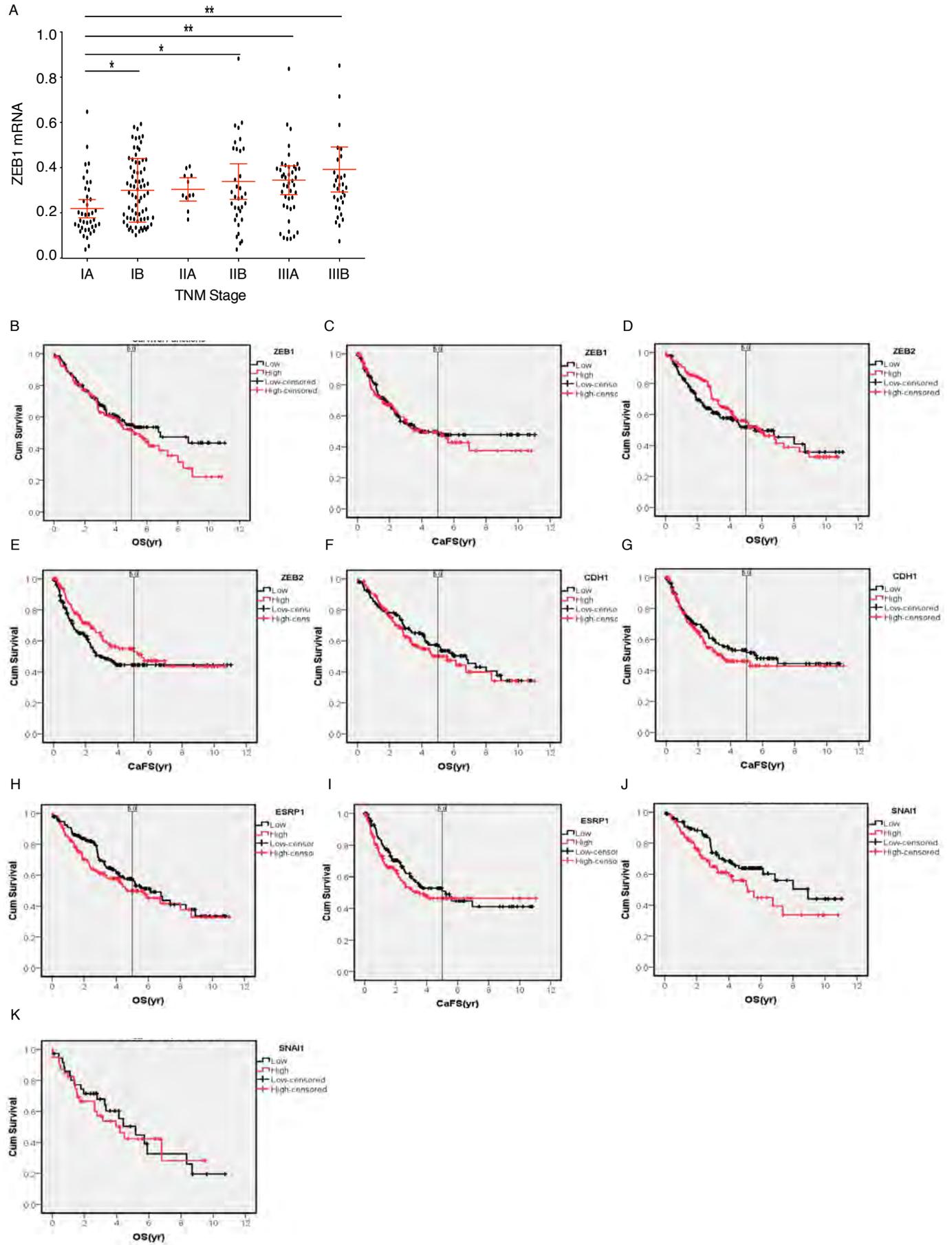


Figure S9: A) Schematic of ESRP1 promoter region indicating locations of E-box (*red*) and Z-box (*blue*) sequences that represent putative ZEB1 binding sites. B) Schematic of CD44 exons showing standard exons (exons 1-5 and 16-20) (*light grey*) and variant exons (exons 6-15) (*dark grey*) and the design of RT-PCR primers (*arrows*) to amplify CD44s and CD44v isoforms.

Supplementary Figure 9

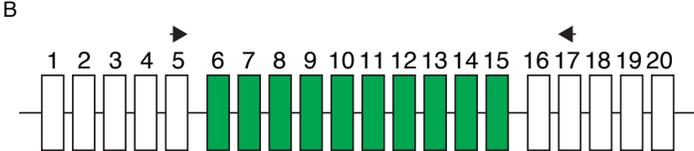
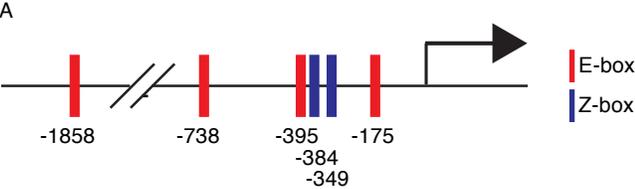


Figure S10: FACS analysis of cell surface expression of CD24 and CD44 following over-expression of ZEB1 in HBEC3. pMSCV, vector control. Representative data of at least three independent experiments.

Supplementary Figure 10

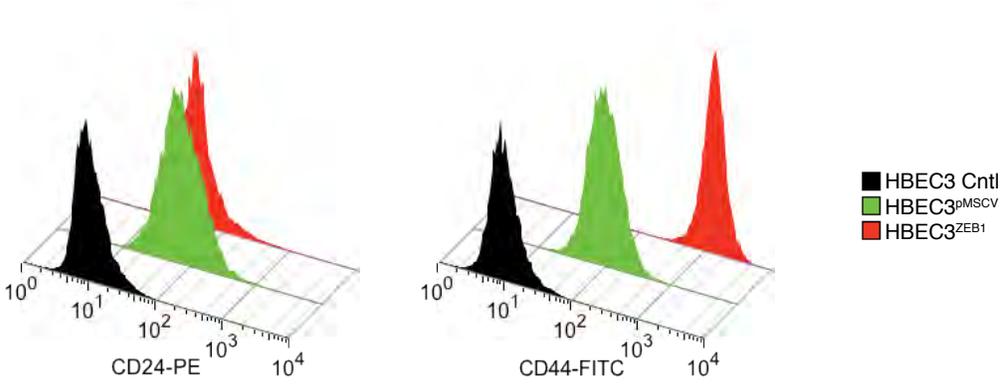
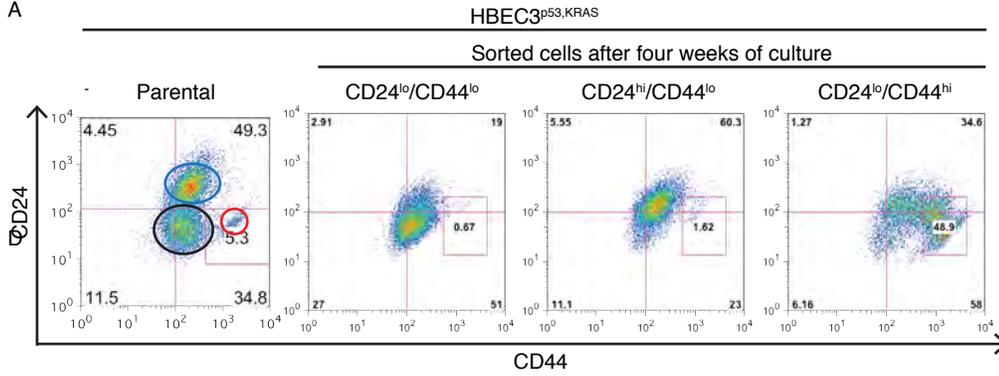


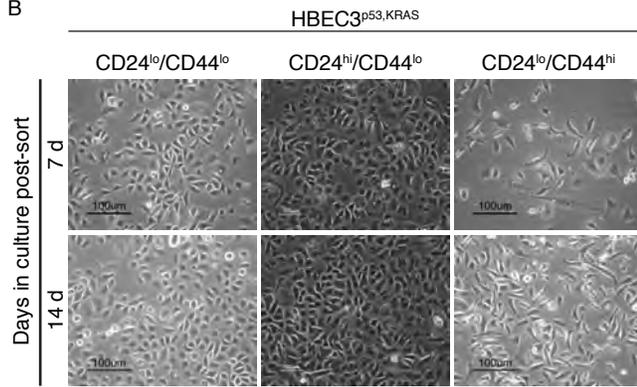
Figure S11: A) Long-term stability of CD24/CD44 profiles in CD24/CD44-sorted populations of HBEC^{p53,KRAS} (using FACS following four weeks of culturing). B) Phase micrographs CD24/CD44-sorted populations of HBEC^{p53,KRAS} following 7 and 14 days of culturing. C) Liquid colony formation assays of CD24/CD44-sorted populations of HBEC^{p53,KRAS} seeded either immediately following cell sorting, or after four weeks of culturing. D) Proportion of aldehyde dehydrogenase (ALDH) positive cells relative to CD24^{lo}/CD44^{hi} cells in HBEC3 series (mean±SD). In panel A, numbers in each corner represent the percentage of cells within that quarter. Gates were drawn in control cells to represent CD24^{lo}/CD44^{hi} cells, numbers within the boxed regions represent percent CD24^{lo}/CD44^{hi} cells. Panels are representative data of at least three independent experiments.

Supplementary Figure 11

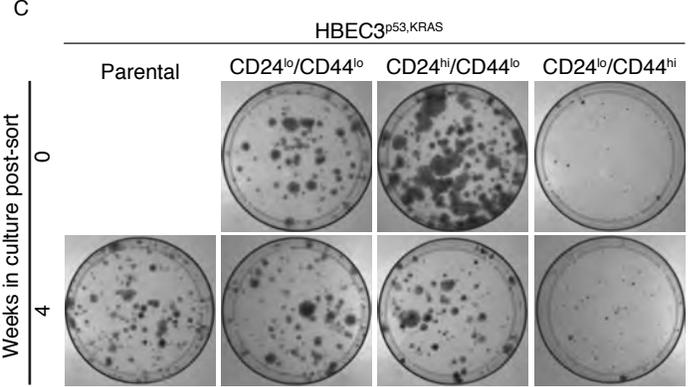
A



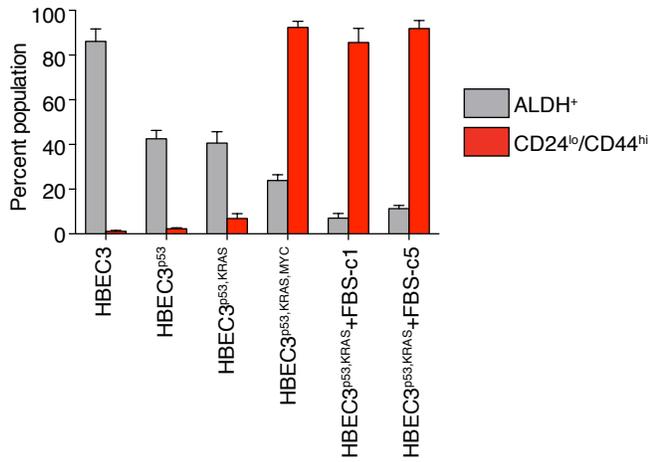
B



C



D



Supplementary Tables

Table S1. Spearman correlation of mRNA expression between EMT-TFs and VIM in the isogenic, oncogenically manipulated series of HBEC3 (16 unique derivatives of HBEC3).

	CDH1	ZEB1	ZEB2	SNAI1	SNAI2	TWIST1
Spearman r value	-0.18	0.82	0.74	0.53	-0.03	0.34
Two-tailed P value	0.4934	0.0001	0.0009	0.0362	0.9133	0.2039

Table S2. Spearman correlation of mRNA expression between EMT-TFs and VIM in a panel of 10 NSCLC cell lines.

	CDH1	ZEB1	SNAI1	SNAI2	TWIST1
Spearman r value	-0.503	0.770	0.115	-0.087	0.667
Two-tailed P value	0.1964	0.0126	0.7589	0.8444	0.0831

Table S3. Predicted transcription factor binding in ZEB1 promoter using ENCODE ChIP-Seq Significance Tool^a

Factor	Total Genes with Factor
Ap2gamma	4170
Atf2	3589
Batf	854
Bcl11a	761
Bcl3	2548
Bhlhe40	6189
Ccnt2	6986
Cfos	3726
CfosTam14h	1265
Chd1	5485
Chd2	8737
CmycSerumstim	7241
CtcfSerumstim	5377
E2f6	8658
Ebf1	3943
Egr1	9491
Elf1	9647
Elk1	4638
Ets1	6329
Foxm1	3276
Gata2	2187
Ikzf1	193
Ini1	2962
Irf4	1671
Max	11417
Maz	10288
Mef2a	2116
Mta3	1822
Mxi1	9383
Nfatc1	1108
Nfic	3904
NfkbTnfa	6987
Nrsf	9119
P300	7444
Pax5	6665
Pbx3	2314
Phf8	9308

Plu1	7419
Pml	7152
Pol2	13720
Pol2Etoh01	8593
Pol2Serumstim	9256
Pol2Serumstvd	9027
Pol2b	5310
Pol2s2	7266
Pou2f2	5971
Pu1	4392
Rad21	5777
Rbbp5	8723
Rfx5	4979
Runx3	7192
Sin3a	10934
Sp1	7864
Srf	4208
Stat3	1170
Taf1	11637
Tblr1	4411
Tbp	10552
Tcf12	5918
Tcf3	3858
Ubf	2664
Ubf	4905
Whip	2546
Yy1	10448
Zeb1	1883

^a <http://encodeqt.simple-encode.org/>. Search terms: Regulatory element type = Protein-coding genes; Analysis Window Parameters Feature Type = TSS / 5' end; Upstream Pad = 500bp; Downstream Pad = 500bp; Cell lines = all.

Table S4. Characteristics of lung cancer cell lines used in sh-ZEB1 studies

Cell Line	Histology	P53	RB	KRAS	LKB1	EGFR	BRAF	NRAS	PTEN	PIK3CA	MYC
Calu-1	Muco-epidermoid	HD	WT	M	WT	WT	WT	WT	WT	WT	NA
H82	SCLC	M	M	WT	WT	WT	WT	WT	WT	WT	A
H1155	Large Cell Neuroendocrine	M	WT	M	WT	WT	WT	WT	M	WT	NA
H1299	Large Cell Neuroendocrine	HD	WT	WT	WT	WT	WT	M	WT	WT	NA
H1792	Adenocarcinoma	M	WT	M	WT	WT	WT	WT	WT	WT	A
H2087	Adenocarcinoma	M	WT	WT	WT	WT	M	M	WT	WT	NA

A, amplification; HD, homozygous deletion; NA, no amplification; M, mutant; WT, wildtype.

Table S5. Comparison of ZEB1-associated genes across seven mRNA whole-genome datasets

Dataset Name	Sample description	No. samples	Analysis outline ^a	Number of ZEB1 correlated genes
HBEC-ZEB1	HBEC3 vs. HBEC3 ^{ZEB1}	2	4-fold change	705 (404 up, 301 down)
HBEC-MYC	HBEC3 ^{p53,KRAS} vs. HBEC3 ^{p53,KRAS,MYC}	2	4-fold change	1,537 (737 up, 800 down)
HCC-NSCLC	NSCLC cell lines	119	20 ZEB1-high tumors vs. 20 ZEB1-low NSCLC cell lines (SAM, FDR<5%, \geq 2-fold)	1,185 (480 up, 666 down)
MUSC-NSCLC (9)	NSCLC cell lines	38	Spearman rank correlation (\geq 0.5 or \leq -0.5)	613 (142 up, 472 down)
Director's Challenge Consortium (10)	Primary lung adenocarcinomas	442	40 ZEB1-high tumors vs. 40 ZEB1-low tumors (SAM, FDR<5%, \geq 2-fold)	618 (551 up, 67 down)
TCGA-LUAD (11)	Primary lung adenocarcinomas	556	Pearson correlation coefficient (\geq 0.3 or \leq -0.3)	1,755 (1,233 up, 522 down)
TCGA-LUSC (12)	Primary lung squamous cell carcinomas	482	Pearson correlation coefficient (\geq 0.3 or \leq -0.3)	1,448 (1,226 up, 222 down)

^a HCC-NSCLC: NSCLC cell lines were stratified by high (n=20) or low (n=20) ZEB1 expression as determined by qRT-PCR. Director's Challenge Consortium: a supervised analysis comparing the top 10th percentile (n = 40 samples) and the bottom 10th percentile (n = 40 samples) was performed using probe 212764_a

Table S6: Spearman correlation of *ZEB1* mRNA expression with its associated genes in mRNA extracted from frozen specimens of 267 primary, resected NSCLC tumors.

	ZEB1	ZEB2	CDH1	ESRP1	SNAI1
Number of samples	267	267	267	267	267
Spearman r value	1.000	0.6688	-0.3463	-0.5349	-0.0304
Two-tailed P value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	n.s.

Table S7: Culture media for cell lines used in this study

Cell Line	Culturing Medium	Tumorigenicity <i>in vivo</i>	<i>In vitro</i> transformation	ZEB1 expression level
HBECs (immortalized, non-transformed)	KSFM (Life Technologies Inc., Carlsbad, CA) containing 50 µg/mL of Bovine Pituitary Extract (BPE) (Life Technologies Inc.) and 5 ng/mL of EGF (epidermal growth factor) (Life Technologies Inc.)	Non-tumorigenic	-	Low
HBEC3 ^{pMSCV}	KSFM+EGF+BPE	Non-tumorigenic	-	Low
HBEC3 ^{MYC}	KSFM+EGF+BPE	Non-tumorigenic	+	Low
HBEC3 ^{ZEB1}	KSFM+EGF+BPE	Not determined	++	High
HBEC3 ^{p53}	KSFM+EGF+BPE	Non-tumorigenic	+	Low
HBEC3 ^{p53,KRAS}	KSFM+EGF+BPE	Non-tumorigenic	++	Low
HBEC3 ^{p53,KRAS,MYC}	KSFM+EGF+BPE	Tumorigenic	+++	High
HBEC3 ^{p53,KRAS} -CD24 ^{lo} /CD44 ^{lo} population	KSFM+EGF+BPE	Non-tumorigenic	++	Low
HBEC3 ^{p53,KRAS} -CD24 ^{hi} /CD44 ^{lo} population	KSFM+EGF+BPE	Non-tumorigenic	++	Low
HBEC3 ^{p53,KRAS} -CD24 ^{lo} /CD44 ^{hi} population	KSFM+EGF+BPE	Non-tumorigenic	+++	High
HBEC3 ^{p53,KRAS} +TGFβ	KSFM+EGF+BPE + TGFβ (5 ng/mL)	Tumorigenic	+++	High
HBEC3 ^{p53,KRAS} +FBS	RPMI1640 + 10% fetal bovine serum	Tumorigenic	+++	High
HBEC3 ^{p53,KRAS,MYC} +FBS	RPMI1640 + 10% fetal bovine serum	Tumorigenic	+++	High
HBEC3 ^{p53,KRAS} +FBS-c1, -c5, -c11	RPMI1640 + 10 % fetal bovine serum	Tumorigenic	+++	High
Lung cancer cell lines	RPMI1640 + 5% fetal bovine serum	Tumorigenic	+++	Varied

Table S8. Primary antibodies used for immunoblotting, ChIP, immunofluorescence and/or FACS.

Antibody	Catalogue	Manufacturer	Assay
anti- β -catenin	9587	Cell Signaling Technology, Inc.	IB
anti-cleaved PARP	9541	Cell Signaling Technology, Inc.	IB
anti-E-cadherin	610181	BD Biosciences	IB
anti-HSP90	sc-13119	Santa Cruz Biotechnology, Inc.	IB
anti-vimentin	550513	BD Biosciences	IB
anti-ZEB1 (E-20)	sc-10572	Santa Cruz Biotechnology, Inc.	IB
anti-ZEB1 (H-102)	sc-25388	Santa Cruz Biotechnology, Inc.	IB, C
anti-ESRP1	H00054845-B01P	Abnova	IF
anti-CD44s	MAB7045	Fisher	IF
anti-CD24-PE	Clone ML5	BD Biosciences	F
anti-CD44-FITC	Clone G44-26	BD Biosciences	F

IB, immunoblotting; C, ChIP; IF, immunofluorescence; F, FACS.

Table S9. Oligo sequences for siRNAs, shRNAs, qRT-PCR and RT-PCR

Target	Sequence
CD44 759-781 F	5-GCACAGACAGAATCCCTGCTACC-3
CD44 2063-2085 R	5-TTTGCTCCACCTTCTTGACTCCC-3
shZEB1 s	GATCCCCAGATGATGAATGCGAGTCGTTCAAGAGATGACTCGCATTTCATCATCTT TTTTGGAAA
shZEB1 as	AGCTTTTCCAAAAAAGATGATGAATGCGAGTCATCTCTTGAACGACTCGCATTCA TCATCTGGG
siZEB1	SI04272492 (Qiagen)