Distinct EMT programs control normal mammary stem cells and tumour-initiating cells

Xin Ye1, Wai Leong Tam1,2,3, Tsukasa Shibue1, Yasemin Kaygusuz1, Ferenc Reinhardt1, Elinor Ng Eaton1 & Robert A. Weinberg1,4,5

Tumour-initiating cells (TICs) are responsible for metastatic dissemination and clinical relapse in a variety of cancers1,2. Analogies between TICs and normal tissue stem cells have led to the proposal that activation of the normal stem-cell program within a tissue serves as the major mechanism for generating TICs3–7. Supporting this notion, we and others previously established that the Slug epithelial-to-mesenchymal transition-inducing transcription factor (EMT-TF), a member of the Snail family, serves as a master regulator of the gland-reconstituting activity of normal mammary stem cells, and that forced expression of Slug in collaboration with Sox9 in breast cancer cells can efficiently induce entrance into the TIC state8. However, these earlier studies focused on xenograft models with cultured cell lines and invoked ectopic expression of EMT-TFs, often at non-physiological levels. Using genetically engineered knock-in reporter mouse lines, here we show that normal gland-reconstituting mammary stem cells9–11 residing in the basal layer of the mammary epithelium and breast TICs originating in the luminal layer exploit the paralogous EMT-TFs Slug and Snail, respectively, which induce distinct EMT programs. Broadly, our findings suggest that the seemingly similar stem-cell programs operating in TICs and normal stem cells of the corresponding normal tissue are likely to differ significantly in their details.

To define the functions of endogenously encoded, physiologically regulated Snail family EMT-TFs in breast cancer pathogenesis in vivo, we generated knock-in IRES (internal ribosomal entry site)–YFP (yellow fluorescent protein) reporters for Slug (also called Snai1) and Snail (also called Snai1) (Fig. 1a, b). These knock-in reporters faithfully reflected the expression of the endogenous genes (Extended Data Fig. 1a, b), and enabled the isolation of Slug+ or Snail+ cells by fluorescence-activated cell sorting (FACS) (Extended Data Fig. 6e–h).

Using these reporters, we found that Slug was expressed at higher levels in the normal mammary stem cell (MaSC)-enriched basal mammary epithelial cells (MECs) compared to the stromal fibroblasts surrounding the mammary ducts. In contrast, the EMT-TFs Snail, Twist and Zeb1 were expressed in stromal fibroblasts but not in either basal or luminal MECs (Fig. 1c–e, g and Extended Data Fig. 1c–f). In addition to the differential expression of EMT-TFs, the MaSC-enriched basal MECs displayed intermediate expression levels of both epithelial and mesenchymal markers (Fig. 1f, g and Extended Data Fig. 1g). Hence, Slug expression in the normal basal MECs was associated with only a partial conversion to the mesenchymal state.

Given the differential expression patterns of Slug and Snail, we undertook to analyse their expression during tumour development using the MMTV-PyMT transgenic model of mammary tumour formation, which mirrors the multi-step progression of human breast cancers beginning from hyperplastic lesions to high-grade carcinomas that spontaneously metastasize to the lungs12. In the initially formed hyperplastic lesions, we noted a marked reduction of the relative frequency of Slug+–YFP+ cells compared to normal glands, contrary to the hypothesis that activation of the Slug EMT-TF might be the preferred mechanism to generate TICs. These Slug–YFP+ cells were cytokeratin14+ (CK14) (Fig. 2a and Extended Data Fig. 2f), indicating that Slug expression was still confined to cells of the basal lineage, as was the case within the normal ducts. In these early-stage lesions, we detected for the first time Snail–YFP expression in a small fraction of the neoplastic cells displaying CK8+–Snail+ luminal characteristics (Fig. 2a, b and Extended Data Fig. 2a–c).

As these early-stage tumours progressed to high-grade carcinomas, the Snail+ cells remained largely confined to the basal sectors of each epithelial island, whereas the Snail+ cancer cells were sometimes fully detached from the epithelial islands and exhibited an elongated mesenchymal morphology (Fig. 2c). We found that virtually all Snail–YFP+ tumour cells had lost E-cadherin expression and activated expression of the Zeb1 EMT-TF; in contrast, most Slug–YFP+ tumour cells retained junctional E-cadherin and lacked Zeb1 expression.
Slug is associated with more complete expression of mesenchymal traits in mammary tumours. 

Figure 2 | Differential expression of Slug and Snail in mammary tumours. 

a, Hyperplastic mammary lesions of the indicated genotypes were stained for the indicated proteins. Arrow in a indicates Snail–YFP and CK8 double-positive cells. Arrows and arrowheads in b indicate Snail–YFP and cytokeratin double-positive cells, and Slug-positive cells, respectively. 

c, d, High-grade carcinomas of the indicated genotypes were stained for the indicated proteins. Arrows indicate Zeb1 and cytokeratin double-positive cells (Fig. 2f). 

d, e, Snail(YFP);MMTV-PyMT tumours were stained for the indicated proteins. Arrows indicate Snail–YFP-positive carcinoma cells. 

f, Tumour organoids of the indicated genotypes were stained for the indicated proteins. Arrows indicate Snail–YFP and cytokeratin double-positive cells (Fig. 2f).

To summarize, these results indicated that the differential expression of Slug and Snail is a common feature of mammary tumours despite their different subtypes, genetic backgrounds and oncogenic drivers. Although Snail is absent in normal MECs, it often becomes activated during breast cancer progression. Indeed, Snail expression was detected in ~80% microdissected human invasive ductal carcinomas.

Our observations raised the question of whether the Slug+, Snail+ or yet other cell subpopulations within mammary tumours were enriched in TICs. To address this issue, we developed a system that allowed us to isolate these various subpopulations. Thus, we FACS-purified premalignant EpCam+ MECs from 4–5-week-old Slug(YFP/+/);MMTV-PyMT;RFP (red fluorescent protein) or Snail(YFP/+/);MMTV-PyMT;RFP animals, and thereafter implanted these cells into cleared mammary fat pads of hosts that lacked these transgenes (Fig. 3a). The implanted premalignant MECs first grew as rudimentary ductal structures and then progressed over 6–7 months to form high-grade carcinomas that metastasized to the lungs (Extended Data Fig. 6a, b).

We FACS-resolved the RFP+ carcinoma cells based on the expression levels of the YFP (Slug or Snail) reporter and the EpCam epithelial marker. In the high-grade carcinomas and corresponding pulmonary metastases, EpCam expression was downregulated in 4–12% of the carcinoma cells. These EpCamlow cells had low Slug–YFP expression (Fig. 3b and Extended Data Fig. 5c) but high Snail–YFP expression (Fig. 3c and Extended Data Fig. 5d). In contrast, the EpCamhigh populations were Snail–YFPlow (Fig. 3c), and were composed of Slug–YFPlowEpCamhigh and Slug–YFPhighEpCamlow subpopulations (Fig. 3b). Hence, EpCam expression was inversely correlated only with Snail expression.

Using quantitative reverse transcription PCR (qRT–PCR) analyses, we confirmed that Snail expression was highest in the EpCamlow subpopulations, whereas Slug was enriched in EpCamhigh subpopulations (Fig. 3d, e). As expected, strong induction of mesenchymal markers and suppression of E-cadherin were only seen in the
EpCAM\textsuperscript{low} subpopulations (Fig. 3d, e). Using cell lines derived from Slug\textsuperscript{YFP}\textsuperscript{+}\textbar/MMTV-PyMT and Snail\textsuperscript{YFP}\textsuperscript{+}\textbar/MMTV-PyMT tumours, we also observed the segregation of Slug and Snail expression at the protein level and associated Snail but not Slug expression with strong induction of a mesenchymal phenotype (Extended Data Fig. 6e–h).

Together, these observations demonstrated our ability to resolve and isolate distinct tumour cell subpopulations with high levels of either Slug (Slug\textsuperscript{high}) or Snail (Snail\textsuperscript{high}) from the same primary tumours (Extended Data Fig. 7a), allowing us, in turn, to directly compare their respective tumour-initiating activities. To this end, we FACS-purified these subpopulations (Extended Data Fig. 6e–h) and implanted each at limiting dilutions to score tumour formation. Overall, Snail\textsuperscript{high} subpopulations exhibited more than two orders of magnitude higher proportions of TICs than did the other subpopulations. In contrast, the Slug–YFP\textsuperscript{high} cells were as deficient in tumour-initiating ability as the Slug–YFP\textsuperscript{low}EpCAM\textsuperscript{high} cells (Extended Data Fig. 7b, c). Hence, Snail but not Slug was tightly associated with a TIC phenotype.

To compare the TIC activities and metastatic powers of these tumour cell subpopulations coexisting in the same primary tumours in vivo, we FACS-purified each subpopulation from highly metastatic carcinomas generated by Slug\textsuperscript{YFP}\textbar/MMTV-PyMT\textbar;RFP cells and introduced them via the tail vein to gauge their respective abilities to seed pulmonary metastases. Notably, the Snail\textsuperscript{high} cells consistently gave rise to far more metastatic outgrowths relative to the EpCAM\textsuperscript{high} subpopulations. In particular, 40,000 Snail\textsuperscript{high} cells from a highly metastatic primary tumour (Extended Data Fig. 8a) seeded on average \(\times 90\) large metastases in each animal. In contrast, 40,000 cells of the Slug\textsuperscript{high} and Slug–YFP\textsuperscript{low}EpCAM\textsuperscript{high} subpopulations from the same tumour seeded an average of only 3.6 and 2.2 metastases per animal, respectively (Fig. 3f and Extended Data Fig. 8b). The Snail\textsuperscript{high} cells were also far more metastatic than the other two populations when implanted subcutaneously (Fig. 3g and Extended Data Fig. 8c–g). Interestingly, the metastatic outgrowths formed by the Snail\textsuperscript{high} cells harbourled gland-like structures composed of both Slug\textsuperscript{+} and Slug\textsuperscript{−} cells (Extended Data Fig. 8c, d). Hence, the Snail\textsuperscript{high} cells that seeded metastases were capable of differentiating within these outgrowths, thereby regenerating the complex cellular hierarchy present in the original primary tumours. These data also revealed that the TICs did not derive from basal MaSC-like cells (that is, the Slug\textsuperscript{high} cells) but instead arose in a different cell population.

We were curious whether correlates of these distinct behaviours of SLUG and SNAIL could be found in human clinical data sets, and therefore examined the prognostic powers of SLUG or SNAIL (ref. 24). Across various patient populations, we found that only elevated expression of SNAIL consistently associated with poor survival (Extended Data Fig. 9a).

These results strongly argue for divergent roles of Slug and Snail, and predicted that shutdown of Snail could selectively eliminate breast TICs. To test this notion, we knocked down either Slug or Snail in advanced MMTV-PyMT carcinoma-derived pBl.3G cells and MDA-MB-231 human breast cancer cells (Extended Data Fig. 9b, c). Strikingly, Snail knockdown but not Slug knockdown induced mesenchymal-to-epithelial transition (MET), leading to loss of Zeb1 and reactivation of E-cadherin (Fig. 4a, b). When these cells were injected orthotopically into mammary fat pads, Snail but not Slug knockdown attenuated primary tumour growth and strongly suppressed their metastatic spreading (Fig. 4c–e). Similarly, across a panel of human breast cancer cell lines, we found that SNAIL knockdown significantly suppressed tumour initiation in most of them, while SLUG knockdown failed to do so (Extended Data Fig. 9f–i). In contrast to the responses of breast cancer cells, the organoid-forming and gland-reconstituting activities of normal murine MaSCs were markedly affected by Slug knockdown but not by Snail knockdown (Fig. 4f and Extended Data Fig. 9g).

Given the distinct functions exerted by Slug and Snail, we analysed the transcription programs controlled by these paralogous EMT-TFs using ChIP-seq. We focused on two MMTV-PyMT tumour cell lines that differed in Slug and Snail expression and tumorigenic potential (Extended Data Fig. 10a, b). We recovered similar numbers of chromatin regions that were enriched for either Slug or Snail binding (Fig. 5a). Across the genome, Slug- and Snail-binding sites displayed similar fold enrichments and were both enriched for the known Snail family recognition CANNNTG E-box motif (Extended Data Fig. 10c, d). We found that Snail occupied 10,129 promoters, far more than that occupied by Slug (2,475 promoters) (Fig. 5b). Interestingly, the promoters of genes encoding key mesenchymal markers were only bound by Snail but not by Slug (Extended Data Fig. 10e). Gene-set enrichment analyses (GSEA) confirmed that Snail-bound but not Slug-bound genes were significantly enriched for EMT-related signatures (Fig. 5c). In particular, Snail, but not Slug, occupied the promoter of Zeb1 (Fig. 5d), a master regulator of
TICs under a variety of settings. The binding of Snail appears to activate Zeb1 expression in mammary tumour cells, because knockdown of Snail led to downregulation of Zeb1 (Fig. 4a, c), and ectopic expression of Snail in human MECs induced ZEB1 expression (Extended Data Fig. 10). Indeed, Snail had been reported to activate Zeb1 expression in non-mammary types of epithelial cells.

To investigate the possible differential abilities of SLUG and SNAIL in controlling ZEB1 expression in human breast cancer cells, we used ChIP-qPCR to examine SLUG and SNAIL binding at the ZEB1 promoter in MDA-MB-231 cells, which co-express SLUG and SNAIL.

Notably, although both SLUG and SNAIL appeared to occupy the ZEB1 promoter in these cells, SLUG binding (but not SNUG expression) was diminished in SNAIL knockdown cells (Fig. 5e–g). Since the knockdown of SNAIL, but not SLUG, resulted in downregulation of ZEB1 in MDA-MB-231 cells (Fig. 4c) and the binding of SLUG to the ZEB1 promoter is dependent on SNAIL expression, we concluded that ZEB1 expression was controlled by SNAIL but not SLUG in MDA-MB-231 cells as well.

Our data underscore profound differences in the transcription-regulating activities of the endogenously encoded Slug and Snail.
EMT-TFs, and provide indications that normal stem cells and TICs of the same tissue-of-origin could arise from different cellular compartments and exploit different molecular signalling circuits to activate related but distinct signalling pathways. We have previously correlated high levels of SLUG expression with poor prognosis in human breast cancer patients, and attributed such correlation to the experimentally observed EMT-inducing function of SLUG⁶. However, on the basis of the present work, we propose that the prognostic power of SLUG expression may be due in large part to its strong association with basal differentiation, which is, on its own, a well-known feature of aggressive breast cancers⁶. Nonetheless, since our in vivo analyses focused on MMTV promoter-driven tumours, which appear to derive primarily from luminal MECs, we suggest that the functions of Slug in basal MEC-derived tumours remain to be further characterized.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 December 2014; accepted 9 July 2015.

Published online 2 September 2015.

Acknowledgements The pBl.1 and pBl.3 murine PyMT tumour cell lines, from which pBl.1G and pBl.3G were derived, were gifts from the Harold L. Moses laboratory. We thank G. Bell for helping analysing the ChIP-seq data. We thank R. Bronson for help providing the ChIP-seq results. The pBl.1 and pBl.3 murine PyMT tumour cell lines, from which pBl.1 and pBl.3 were derived, were gifts from the Harold L. Moses laboratory. We thank R. Bronson for help assessing the histopathology of the murine tumours. We thank A. Lambert and S. Thiagalingam for providing the MCF10A-Ras and MCF10A-Ras-C cells. We thank W. Israelesen and M. Vander Heiden for providing the pBl.1 and pBl.3 murine mammary tumour samples. We thank the Keck Microscopy Facility at the Whitehead Institute for microscopy assistance and the Koch Institute Swanson Biotechnology Center (SBC) for technical support, especially the Histology and ES cell Transgenics Core. R.A.W. is an American Cancer Society and Ludwig Foundation professor. W.L.T. is supported by the National Research Foundation, Singapore (NRF-NRF2015-04). This research was supported by the Breast Cancer Research Foundation, the Samuel Waxman Cancer Research Foundation, the Ludwig Center for Molecular Oncology at MIT, National Cancer Institute Program P01-CA080111, R01-CA07461, U01-CA184937 (to R.A.W.), K99-CA194160 (to X.Y.), the Wistshire Charitable Foundation/Andria and Paul Healy Postdoctoral Fellowship (to X.Y.), the Mattina R. Proctor Foundation, and the Helen Hay Whitney Foundation (to X.Y.).

Author Contributions X.Y. and R.A.W. conceived the project and prepared the manuscript. X.Y. designed and performed the experiments and analysed the data. W.L.T. contributed to the ChIP-qseq experiments. T.S. performed the knock-in animals. Y.K. performed experiments and quantifications. F.R. performed the mammary fat pad injections. E.E. provided technical support.

Author Information The ChIP-qseq results have been deposited to Gene Expression Omnibus (GEO) under accession number GSE61198. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.Y. (xy@wi.mit.edu) or R.A.W. (weinberg@wi.mit.edu).
Tumour cell lines were all derived from high-grade carcinomas of 6–7 months old females.

**Cell lines and cell culture.** PyMT tumour cell lines were cultured in DMEM/F12 (1:1) supplemented with 5% adult bovine serum (Sigma B9433), non-essential amino acid (Invitrogen 11140), and Pen/Strep (Invitrogen 15070). MDA-MB-231, MDA-MB-361 and MDA-MB-468 cells were culture in DMEM/F12 (1:1) supplemented with 10% inactivated fetal bovine serum (Sigma F41135) and Pen/Strep. SKBR3, BT20 and MCF7/Ras cells were cultured in DMEM supplemented with 10% levamisole treated fetal bovine serum and Pen/Strep. SUM149 and SUM159 cells were cultured in F12 media supplemented with 5% inactivated fetal bovine serum, insulin (5 μg ml⁻¹), and hydrocortisone (1 μg ml⁻¹). T47D, Hs578T and BT549 cells were cultured in RPMI supplemented with 10% inactivated fetal bovine serum and Pen/Strep. MCF10A cells were cultured in DMEM/F12 supplemented with 5% horse serum (Sigma H2127), insulin (10 μg ml⁻¹) Sigma I9287), EGF (100 μg ml⁻¹) Sigma E9644), hydrocortisone (0.5 mg ml⁻¹) Sigma H8088), and cholera toxin (100 ng ml⁻¹ Sigma C8052). All human breast cancer cell lines were obtained from ATCC, and are free of mycoplasma contamination.

**METHODS**

**Animals.** The SlugYFP and SnailYFP alleles were generated by homologous recombination in mouse embryonic stem (ES) cells using standard gene-targeting methods. The MMTV-Neu animals and CAG-mRFP animals were obtained from the Jackson Laboratory (stock numbers 005038 and 005884). The MMTV-PyMT animals were originally obtained from the Jackson Laboratory (stock number 002374) and backcrossed for five generations to C57BL/6 background. Mice were housed and handled in accordance with protocols approved by the Animal Care and Use Committees of the Massachusetts Institute of Technology. Animals were randomized by age and weight. The investigators were blinded to allocation during experiments and outcome assessment for experiments shown in Fig. 3f, g and Extended Data Fig. 10b.

**Genotyping.** PCR primers (5’ to 3’ for SlugYFP genotyping were (sense strand, AACCTTCCAGAAATGCTGCTTCGTG; antisense strand, TTGAGGATGTATCTTATACACGTGGCC) and for SnailYFP genotyping were (sense strand, CTCCCGCATGTCCTGGTCCACCAAG; antisense strand, same as for SlugYFP). DNA extractions and subsequent PCR reactions were performed using the RedExtract-N-Amp Tissue PCR kit (Sigma XNAT). PCR was performed with 35 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C for 30 s, and elongation at 72 °C for 30 s.

**Tumour categorization.** MMTV-PyMT tumours were categorized into hyperplastic lesions, adenocarcinomas and high-grade carcinomas following the histological criteria described by Lin et al.². Briefly, in the genetic background of our animals, hyperplastic lesions usually occur at about 8–10 weeks of age, and consist of regionally packed lobules formed on the duct. We collectively refer to the adenoma/MIN and early carcinoma stages defined by Lin et al.² as adenocarcinomas, as the tumours from 2.5 to 4-month-old animals often have mixed characteristics of both kinds. High-grade carcinoma corresponds to the late carcinoma/advanced invasive carcinoma as defined by Lin et al.², and is always associated with metastatic disease.

**Immunofluorescence staining.** Tumours were fixed in 10% neutral buffered formalin overnight and embedded in paraffin for sectioning. Sections were cut at 5 μm. Tumour sections were deparaffinized in Histoclear II, and antigen retrieval was performed with Nuclear Decloaker (Biocare Medical) using a microwave. Sections were then blocked with 0.3% normal donkey serum (Jackson Immunoresearch Laboratories) in PBST (PBS 0.3% Triton X-100) for 1 h at room temperature. Sections were incubated with primary antibody at 4 °C overnight. After three washes with PBS, sections were incubated with secondary antibodies (Biotium) and DAPI for 2 h at room temperature, washed three times with PBS, and mounted in Prolong gold antifade reagent (Invitrogen P36930). For antibodies (Biotium) and DAPI for 2 h at room temperature, washed three times with PBS, and mounted in Prolong gold antifade reagent.

**Tumour organoids were fixed in collagen I gel with 4% paraformaldehyde for 1 h at room temperature, blocked with 0.5% normal donkey serum in PBST, and then incubated with primary antibodies at 4 °C overnight. After three washes with PBST, organoids were incubated with DAPI, secondary antibodies and Phalloidin at 4 °C overnight. After five washes with PBST, the collagen I gel containing the organoids were mounted in Prolong gold antifade reagent.

**Cultured tumour cells were fixed in 4% paraformaldehyde and blocked with 0.1% normal donkey serum in PBST for half an hour at room temperature. Cells were incubated in specific primary antibodies for 1–2 h, washed three times with PBS, then incubated with secondary antibodies for 1 h at room temperature. After three washes with PBS, stained cells were mounted in Prolong gold antifade reagent.

**Immunostained samples were imaged using Zeiss LSM710 and Zeiss LSM700 confocal microscopes and analysed with Zen software.**

**Antibodies used in this study are listed in the Supplementary Information.**

**Tumour dissociation, FACS fractionation and derivation of tumour cell lines.** Tumours were taken from the animals aseptically. At least one fragment from each tumour was saved for histological staging of the tumour. The remainder of each tumour was then minced with a razor blade, and the minced chunks were then washed twice with DMEM with 10% FBS, and filtered through a 70 μm cell strainer. The resulting cells were stained with DAPI, anti-EpCAM antibody, washed three times with PBS, and resuspended in PBS for flow cytometry analysis and FACS fractionation.

**To establish tumour cell lines, 1 × 10⁵** dissociated and filtered tumour cells were plated in a 10 cm dish in DMEM/F12 supplemented with 5% adult bovine serum, non-essential amino acids (Invitrogen), and penicillin/streptomycin. On the next day, dead cells were removed by medium change, and the attached cells were passaged at 1:2 to 1:3 for about five passages until each culture was established.
using http://kmplot.com (ref. 24). The patient samples were grouped as either high or low expressor for the genes of interest based on the expression level of the selected gene, and the upper tertile were used as the cut-off and median is computed over the entire data set.

**Chromatin immunoprecipitation sequencing.** ChIP assay was carried out as described previously31. The ChIP-seq results have been deposited to Gene Expression Omnibus (GEO) under accession number GSE61198. We performed ChIP-seq experiments using either an anti-Slug or anti-Snail antibody. In the Slug-high pBlG cells, as anticipated, we could only recover chromatin fragments from Slug ChIP but not Snail ChIP. Conversely, from the Snail-high pBl3G cells, chromatin fragments were only recoverable from Snail ChIP but not Slug ChIP, demonstrating the specificities of these antibodies. Slug- or Snail-bound DNA sites were determined with Model-based Analysis of ChIP-Seq (MACS) algorithm, and bound target genes were defined as containing Slug or Snail occupancy within 5 kb upstream and downstream relative to the transcription start site for each RefSeq transcript35. Fold enrichment of each MACS peak was calculated against the whole-cell extract. Enriched motifs were identified using 1,000 nucleotides centred at the peak summit of the top 1,000 Slug and Snail peaks (ranked by MACS peak scores). The sequences were processed through MemeChIP (http://meme.nbcr.net/meme/cgi-bin/meme-chip.cgi) using default settings.

Gene set enrichment analyses (GSEA) were performed with the GSEA platform of the Broad Institute (http://www.broadinstitute.org/gsea/index.jsp). Slug-bound genes and Snail-bound genes were ranked according to the fold enrichment of the correspondent MACS peaks.

For SLUG and SNAIL ChIP-qPCR in MDA-MB-231 cells, ChIP-enriched DNA was analysed by real-time PCR using the ABI PRISM 7900 sequence detection system and SYBR green master mix. Relative occupancy values were calculated by determining the apparent immunoprecipitation efficiency (ratios of the amount of immunoprecipitated DNA to that of the input sample) and normalized to the level observed. The primers used for the real-time PCR are ZEB1 target locus forward ACAAGCGAGGATCATGCG, reverse CACTCACCGTTATTGGCGCG; ZEB1 control locus forward TAATAATGGCGCGCAACGCC, reverse AGGAA CCAAGCGAGGCCCT.

**Statistical analysis.** Statistical analyses were carried out by two-tailed Student’s t-test unless otherwise specified. No statistical methods were used to predetermine sample size.

31. Tam, W. L. et al. Protein kinase Cz is a central signaling node and therapeutic target for breast cancer stem cells. *Cancer Cell* **24**, 347–364 (2013).
32. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).
33. Shibue, T., Brooks, M. W. & Weinberg, R. A. An integrin-linked machinery of cytoskeletal regulation that enables experimental tumor initiation and metastatic colonization. *Cancer Cell* **24**, 481–498 (2013).
34. Gupta, P. B. et al. The melanocyte differentiation program predisposes to metastasis after neoplastic transformation. *Nature Genet.* **37**, 1047–1054 (2005).
35. Zhang, Y. et al. Model-based analysis of ChiP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
Extended Data Figure 1 | Slug expression is associated with a partial EMT phenotype in normal MECs. a, Validation of the Slug–YFP knock-in reporter. Mammary tumour sections from Slug\textsuperscript{YFP/1};MMTV-PyMT female mice were stained for YFP (green), Slug (red), cytokeratin (grey) and DAPI (blue). b, Validation of the Snail–YFP knock-in reporter. Mammary tumour sections from Snail\textsuperscript{YFP/1};MMTV-PyMT female mice were stained for YFP (green), Snail (red), cytokeratin (grey) and DAPI (blue). c, Lin\textsuperscript{−} cells of normal mammary glands were separated into luminal MECs, basal MECs and stromal fibroblasts using CD24 and CD49f cell-surface markers. d, e, Representative FACS histogram showing relative expression levels of Slug–YFP and Snail–YFP reporters in the indicated cell populations in mammary glands during puberty (d) and during pregnancy (e). Note that luminal MECs from pregnant females exhibit higher levels of autofluorescence signals (grey dashed line in panel e). f, Normal human mammary tissue sections were stained for Slug or Zeb1 (green), CK14 (red), CK8 (grey), and DAPI (blue). Arrowheads indicate Slug\textsuperscript{+} CK14\textsuperscript{+} cells. g, Representative FACS histogram showing expression level of the epithelial cell-surface marker EpCAM in the indicated populations of the normal mammary gland. Panels d, e, g, are representative of three independent experiments. All scale bars indicate 10 μm.
Extended Data Figure 2 | Differential expression of Snail and Slug in mammary tumours. a, b, Quantifications of the frequencies of Slug–YFP+ and Snail–YFP+ tumour cells (a) and quantifications of Slug versus Snail expression (b) at different stages of mammary tumour development by immunofluorescence staining. For each stage, tumours from six animals were analysed for the quantifications. c, Individual channels of the stained image in Fig. 2e. d, Quantifications of E-cadherin and Zeb1 positivity (high-grade carcinomas from six animals were quantified). e, f, Quantification of CK8 and CK14 expression profile of Snail–YFP-positive and Slug–YFP-positive tumour cells. For each stage, tumours from six animals were analysed for the quantifications. b, d–f, n is number of cells.
Extended Data Figure 3 | Snail activation is associated with invasive changes in mammary tumour cells ex vivo. 

(a) Freshly isolated tumour organoids stained for YFP (green), CK14 (red), CK8 (grey) and DAPI (blue). Note that only background staining was detected for YFP and CK14. Scale bar, 20 μm.

(b) Tumour organoids from animals of the indicated genotypes were cultured in type I collagen gel for 48 h and stained for YFP (green), phalloidin (red) and DAPI (blue). Scale bar, 10 μm.

(c) Frequency of CK8⁺CK14⁺ leader cells expressing Slug–YFP and Snail–YFP (n, number of cells). Tumour organoids from five different animals were analysed for each genotype.

(d) Schematic diagram summarizing expression patterns of Snail and Slug in the normal mammary gland and at different stages of mammary tumour development in the MMTV-PyMT model.
Extended Data Figure 4 | Differential expression of Snail and Slug in MMTV-Neu and BRCA-1/p53-minus models of mammary tumours.

a–c, Representative immunofluorescence images of sections of aggressive MMTV-Neu tumours stained for DAPI (blue), cytokeratin (red) and Slug (green, a), Snail (green, b) or Zeb1 (green, c). Scale bar, 10 μm. d, H&E staining showing representative histology of differentiated area in MMTV-cre;p531/2;BRCA1fl/fl tumours. Scale bar, 50 μm. e, Representative immunofluorescence images of the differentiated areas in MMTV-cre;p531/2;BRCA1fl/fl tumours stained for the indicated proteins. Five tumours were analysed, and quantifications are shown in f (n, number of cells). Scale bar, 10 μm. g, H&E staining showing representative histology of differentiated area in MMTV-cre;p531/2;BRCA1fl/fl tumours. Scale bar, 50 μm. h, Representative immunofluorescence images of the invasive areas in MMTV-cre;p531/2;BRCA1fl/fl tumours stained for the indicated proteins. Five tumours were analysed, and quantifications are shown in i (n, number of cells).
Extended Data Figure 5 | Differential expression of Snail and Slug in human breast cancer cell lines.  

**a** | Representative immunofluorescence images of indicated human breast cancer cell lines stained for DAPI (blue), SNAIL (green) and SLUG (red). Scale bar, 10 μm.

**b** | Quantification of SLUG versus SNAIL expression in indicated human breast cancer cell lines (n, number of cells). Five fields were counted for each cell line.

**c** | Representative image showing the morphologies of the series of MCF10A cell lines in culture. Scale bar, 50 μm.

**d** | Western blot showing expression of SLUG and SNAIL in the indicated MCF10A cell lines. Panels a–d represent two independent experiments. Uncropped western blots are available in Supplementary Information.
Isolation of tumour cell subpopulations with differential Snail and Slug expression by FACS. 

**a, b**, Representative whole-mount images showing tumour progression in the transplantation model of mammary tumours illustrated in Fig. 3a. The implanted cells initially formed rudimentary gland-like structures (**a**) and eventually progressed to become high-grade carcinomas that spontaneously metastasize to the lungs. The RFP marker allows detection of pulmonary metastases as shown in **b**. Scale bars, 500 μm. Images represent five independent experiments.

**c, d**, FACS profiles of RFP tumor cells in the pulmonary metastases corresponding to the primary tumours shown in Fig. 3b, c. Major populations are outlined with dashed circles. **c**, Snail<sup>F<sup>RP</sup></sup>MMTV-PyMT tumour cells were separated into indicated populations by FACS. The morphologies of the unfractionated cells and the purified populations are shown. Scale bar, 50 μm.

**e, f**, Western blots showing expression of EMT markers in the indicated cell populations. Uncropped western blots are available in Supplementary Information. Data represent three independent experiments.
Extended Data Figure 7 | Fractionation of primary mammary tumours.

**a** Slug<sup>YFP<sup>+</sup></sup>; PyMT, RFP transplanted tumours associated with metastatic diseases.

Original metastases

Donor primary tumour

Dissociated and FACS sorted

Slug-YFP<sup>lo</sup>EpCAM<sup>hi</sup>

Slug-YFP<sup>hi</sup>EpCAM<sup>hi</sup> (Slugh<sup>high</sup>)

Slug-YFP<sup>h</sup>oEpCAM<sup>lo</sup> (Snail<sup>high</sup>)

Tail-vein injection
(Fig. 3f, Extended data figure 8a-d)

Subcutaneous implantation
(Fig. 3g, Extended data figure 8e-g)

**b** Tumour incidence of Slail<sup>YFP<sup>+</sup></sup>; PyMT cancer cell subpopulations (implanted with 10% Matrigel)

|                  | 1x10<sup>5</sup> | 1x10<sup>4</sup> | 1x10<sup>3</sup> | 1x10<sup>2</sup> | Estimated TIC Frequency (95% Confidence Interval) |
|------------------|-----------------|-----------------|-----------------|-----------------|-----------------------------------------------|
| Snail-YFP<sup>hi</sup>EpCAM<sup>lo</sup> | 6/6             | 6/6             | 6/6             | 3/6             | 1/143 (1/428~1/47)                                |
| Snail-YFP<sup>lo</sup>EpCAM<sup>hi</sup> | 5/6             | 2/6             | 0/6             | 0/6             | 1/46,014 (1/111,052~1/19,065)                     |

**c** Tumour incidence of Slug<sup>YFP<sup>+</sup></sup>; PyMT cancer cell subpopulations (implanted with 10% Matrigel)

|                  | 1x10<sup>6</sup> | 1x10<sup>5</sup> | 1x10<sup>4</sup> | 1x10<sup>3</sup> | 1x10<sup>2</sup> | Estimated TIC Frequency (95% Confidence Interval) |
|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------------------------------------|
| Slug-YFP<sup>lo</sup>EpCAM<sup>hi</sup> | 6/6             | 6/6             | 2/6             | 0/6             | 1/68,189 (1/167,066~1/27,832)                     |
| Slug-YFP<sup>hi</sup>EpCAM<sup>hi</sup> | 5/6             | 5/6             | 0/6             | 0/6             | 1/72,118 (1/177,908~1/29,234)                     |
| Slug-YFP<sup>lo</sup>EpCAM<sup>lo</sup> | N.D.            | 6/6             | 6/6             | 6/6             | 4/6             | 1/91 (1/254~1/33)                                |

Extended Data Figure 7 | Fractionation of primary mammary tumours.

a. Experimental scheme for Fig. 3f, 8a–g, b, c. Tumour cell subpopulations from Slail<sup>YFP<sup>+</sup></sup>;MMTV-PyMT tumour cell line (b) and Slug<sup>YFP<sup>+</sup></sup>;MMTV-PyMT tumour cell line (c) were injected subcutaneously at limiting dilutions to score primary tumour formation. Tumour-initiation cell frequencies were evaluated by ELDA. b, c. Tumour initiation was scored and presented as (number of tumour incidences/number of injections).
Extended Data Figure 8 | Breast TICs express Snail. a, H&E staining showing the histology of the donor primary tumour where cells used in Fig. 3f were isolated from. Scale bar, 200 μm. b, The original pulmonary metastases spawned by the primary tumour (left panel), and pulmonary metastases formed by the indicated tumour cell populations following tail-vein injection. Scale bar, 500 μm. c, Higher magnification images of H&E-stained lung sections showing histology of the original pulmonary metastases in the donor animal (left panel), and pulmonary metastases formed by the Slug-YFP<sup>lo</sup>EpCAM<sup>lo</sup> tumour cells following tail-vein injection. Scale bar, 200 μm. d, Representative immunofluorescence staining image of sections of pulmonary metastases formed by the Slug-YFP<sup>lo</sup>EpCAM<sup>lo</sup> tumour cells were stained for DAPI (blue), Slug (green), CK14 (red) and CK8 (grey). Arrowheads indicate Slug-positive cells. Scale bar, 20 μm. Images represent four independent experiments. e, H&E staining of the donor primary tumour where cells used in Fig. 3g were isolated from (left panel) and H&E staining of primary tumours formed by the indicated populations following subcutaneous implantation (with 25% Matrigel). Scale bar, 200 μm. f, Primary tumour burdens formed by the indicated populations after subcutaneous implantation (for EpCAM<sup>hi</sup>Slug<sup>lo</sup> cells 1 × 10<sup>4</sup> cells were injected, for the other two groups 1 × 10<sup>5</sup> cells were injected). Primary tumours and lungs were analysed 10 weeks post injection (n = 10 sites of injection for each group). Open circle indicates failure of tumour initiation. Source Data is associated with this figure. g, H&E staining of lung sections showing metastatic outgrowths spawned by the indicated cell populations following subcutaneous implantation. Scale bar, 500 μm.
Extended Data Figure 9 | Snail and Slug are differentially employed by normal MaSCs and breast TICs. a, Kaplan–Meier plots showing survival of patients with the indicated subtypes of breast cancers. Patient groups were separated based on SLUG (top row) or SNAIL (bottom row) mRNA expression. b, Western blot confirming Slug and Snail knockdown in established PyMT tumour cell line transduced with the indicated shRNA expression vectors. The shLuciferase (shLuc) shRNA was used as a control. c, Western blot confirming SLUG and SNAIL knockdown in MDA-MB-231 cells transduced with the indicated shRNA expression vectors. ShLuc was used as a control. Uncropped western blots are available in Supplementary Information. d, Tumour-sphere formation efficiencies (no. tumour spheres/1,000 cells for MDA-MB-361 cells, and no. tumour spheres/200 cells for all the other cell lines) of the indicated human breast cancer cells transduced with shSLUG#2, shSNAIL#2 and the shLuc control (mean ± s.d., n = 5 technical replicates per group). Data represent two independent experiments. e, f, SUM159 (e) and SUM149 (f) cells transduced with the indicated shRNAs were injected subcutaneously at limiting dilutions to score primary tumour formation. Tumour initiation was scored and presented as (no. of tumour incidences/no. of injections). Data represent two independent experiments. g, The organoid forming efficiencies of normal MECs transduced with the indicated shRNA expression vectors (mean ± s.d., n = 6 technical replicates per group, *P < 0.001, NS, not significant). Scale bar, 100 μm. Data represent three independent experiments.
Extended Data Figure 10 | Slug and Snail occupy different genomic regions.

a, Western blots showing expression of EMT-TFs and EMT markers in the PyMT tumour cell lines used for the ChIP-seq analyses. Uncropped western blots are available in Supplementary Information. Data represent three independent experiments.

b, Pulmonary metastases formed by 100,000 cells of the indicated cell lines following tail-vein injection (n = 9 animals per group). Source Data is associated with this figure.

c, Box plot showing distributions of fold enrichment of all peaks identified in Snail ChIP and Slug ChIP. Horizontal bar indicates the median and whiskers indicate the top and bottom tertiles.

d, Sample top motifs enriched around the summits of the anti-Snail and anti-Slug ChIP peaks.

e, Sample ChIP-seq signals for Slug and Snail are shown. Left column shows promoters bound by Slug only. Right column shows promoters bound by Snail only. Arrows indicate the directions of transcription.

f, MCF10A human mammary epithelial cells were transduced with rtTA and SNAIL driven by a tet-on promoter, untreated (left panel) or treated with 2 μg ml⁻¹ doxycycline (dox) for 48 h (right panel), and stained for E-cadherin (green) and ZEB1 (red). Scale bar, 20 μm. Data represent five independent experiments.