

Identification of the tumour transition states occurring during EMT

Ievgenia Pastushenko¹, Audrey Brisebarre¹, Alejandro Sifrim^{2,3}, Marco Fioramonti¹, Tatiana Revenco¹, Soufiane Boumahdi¹, Alexandra Van Keymeulen¹, Daniel Brown^{2,4}, Virginie Moers¹, Sophie Lemaire¹, Sarah De Clercq⁵, Esmeralda Minguijón⁵, Cédric Balsat⁶, Youri Sokolow⁷, Christine Dubois¹, Florian De Cock¹, Samuel Scozzaro¹, Federico Sopena⁸, Angel Lanas⁹, Nicky D'Haene⁵, Isabelle Salmon^{5,6}, Jean-Christophe Marine^{4,10}, Thierry Voet^{2,3}, Panagiota A. Sotiropoulou^{1,12} & Cédric Blanpain^{1,11,12*}

In cancer, the epithelial-to-mesenchymal transition (EMT) is associated with tumour stemness, metastasis and resistance to therapy. It has recently been proposed that, rather than being a binary process, EMT occurs through distinct intermediate states. However, there is no direct in vivo evidence for this idea. Here we screen a large panel of cell surface markers in skin and mammary primary tumours, and identify the existence of multiple tumour subpopulations associated with different EMT stages: from epithelial to completely mesenchymal states, passing through intermediate hybrid states. Although all EMT subpopulations presented similar tumour-propagating cell capacity, they displayed differences in cellular plasticity, invasiveness and metastatic potential. Their transcriptional and epigenetic landscapes identify the underlying gene regulatory networks, transcription factors and signalling pathways that control these different EMT transition states. Finally, these tumour subpopulations are localized in different niches that differentially regulate EMT transition states.

EMT is a cellular process in which cells lose their epithelial characteristics and acquire mesenchymal features, which enable them to migrate more efficiently and invade the underlying mesenchyme. EMT is essential for gastrulation, somitogenesis and neural crest delamination during embryonic development and has been associated with various diseases. In cancer, EMT is associated with tumorigenesis, invasion, metastasis, tumour stemness and resistance to therapy^{1,2}. Although EMT has traditionally been viewed as a binary switch, some in vitro data (mainly co-expression of epithelial and mesenchymal markers within the same cells) have indicated that EMT may proceed in a step-wise manner through the generation of subpopulations that represent different intermediate states between the epithelial and mesenchymal states^{3–7}. However, it remains unclear whether EMT proceeds through these intermediate states in vivo, and if so how many intermediate steps exist, how plastic and reversible these intermediate states are, which mechanisms regulate the transition from one state to another and what the implications of these different EMT states are for tumour progression, stemness and metastasis^{1,2}.

Different tumour EMT transition states

To determine whether EMT in vivo occurs through a succession of different intermediate states, we used a genetic mouse model of skin squamous cell carcinoma (SCC) mediated by the conditional expression of *KRas*^{G12D} and *p53* deletion (*p53*^{KO}) in hair follicles. This model generates skin tumours that undergo spontaneous EMT, containing epithelial YFP⁺Epcam⁺ and mesenchymal-like YFP⁺Epcam[−] tumour cells (TCs)⁸. Using flow cytometry (fluorescence-activated cell sorting, FACS), we screened cells from these tumours for a large panel of cell surface markers and assessed whether these markers were

heterogeneously expressed in YFP⁺Epcam⁺ or YFP⁺Epcam[−] populations (Fig. 1a). The YFP⁺Epcam⁺ TCs were relatively homogenous, with only four markers being heterogeneously expressed (Fig. 1b). By contrast, half of the markers were heterogeneously expressed in YFP⁺Epcam[−] TCs (Fig. 1b, Extended Data Fig. 1a, b), suggesting that EMT is associated with important cellular heterogeneity. The markers that were most frequently heterogeneously expressed during EMT included CD61 (also known as Itgb3), CD51 (also known as Itgav) and CD106 (also known as Vcam1; Extended Data Fig. 1c), which mark subpopulations of TCs associated with tumour stemness, EMT or metastasis initiation in other tumour models^{9–12}. Other markers were not as frequently heterogeneously expressed when analysed in a larger cohort of tumours (Extended Data Fig. 1c). Combinatorial multicolour FACS analysis revealed that CD106, CD61, and CD51 discriminated six distinct populations within YFP⁺Epcam[−] TCs in most (75%) mixed tumours (Fig. 1c, e, h, Extended Data Fig. 2). About 10% of mixed tumours with a high proportion of Epcam⁺ cells presented only triple-negative (Epcam[−]CD51[−]CD61[−]CD106[−]) and Epcam[−]CD51⁺CD61⁺ populations, whereas highly mesenchymal tumours with only Epcam[−] TCs contained CD51⁺, CD51⁺CD61⁺, CD106⁺CD51⁺ and CD106⁺CD51⁺CD61⁺ triple-positive tumour subpopulations with almost no triple-negative and CD106⁺ subpopulations (Fig. 1c–g).

To define whether these different tumour populations correspond to distinct EMT transition states, we isolated the subpopulations by FACS and performed immunostaining on cytospin with epithelial (keratin 14, K14) and mesenchymal (vimentin) markers. Notably, loss of Epcam expression coincided with a gain in vimentin expression in all TCs, consistent with the first switch to the mesenchymal state (Fig. 1i, j). However,

¹Laboratory of Stem Cells and Cancer, Université Libre de Bruxelles, Brussels, Belgium. ²Department of Human Genetics, University of Leuven, Leuven, Belgium. ³Sanger Institute-EBI Single-Cell Genomics Centre, Wellcome Trust Sanger Institute, Hinxton, UK. ⁴Laboratory for Molecular Cancer Biology, VIB Center for Cancer Biology, Leuven, Belgium. ⁵Pathology Department, Erasme University Hospital, Université Libre de Bruxelles, Brussels, Belgium. ⁶DIAPath, The Center for Microscopy and Molecular Imaging, Université Libre de Bruxelles, Gosselies, Belgium. ⁷Thoracic Surgery, Erasme University Hospital, Université Libre de Bruxelles, Brussels, Belgium. ⁸Gastroenterology Department, Hospital Clínico Universitario “Lozano Blesa”, IIS Aragón, Zaragoza, Spain. ⁹CIBERehd, IIS Aragón, University of Zaragoza, Zaragoza, Spain. ¹⁰Laboratory for Molecular Cancer Biology, Department of Oncology, KU Leuven, Leuven, Belgium. ¹¹WELBIO, Université Libre de Bruxelles, Bruxelles, Belgium. ¹²These authors jointly supervised this work: Panagiota A. Sotiropoulou, Cédric Blanpain. *e-mail: cedric.blanpain@ulb.ac.be

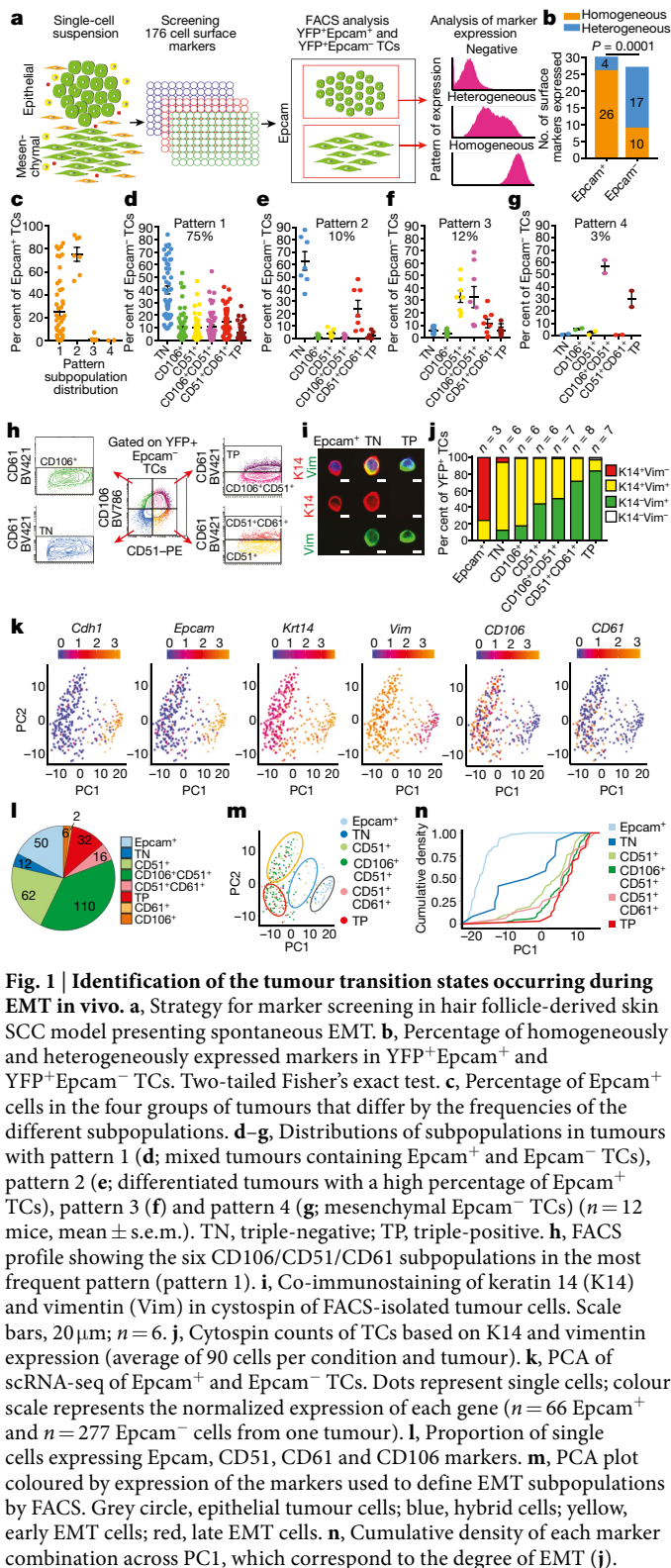


Fig. 1 | Identification of the tumour transition states occurring during EMT in vivo. **a**, Strategy for marker screening in hair follicle-derived skin SCC model presenting spontaneous EMT. **b**, Percentage of homogeneously and heterogeneously expressed markers in YFP⁺Epcam⁺ and YFP⁺Epcam⁻ TCs. Two-tailed Fisher's exact test. **c**, Percentage of Epcam⁺ cells in the four groups of tumours that differ by the frequencies of the different subpopulations. **d–g**, Distributions of subpopulations in tumours with pattern 1 (**d**; mixed tumours containing Epcam⁺ and Epcam⁻ TCs), pattern 2 (**e**; differentiated tumours with a high percentage of Epcam⁺ TCs), pattern 3 (**f**) and pattern 4 (**g**; mesenchymal Epcam⁻ TCs) ($n = 12$ mice, mean \pm s.e.m.). TN, triple-negative; TP, triple-positive. **h**, FACS profile showing the six CD106/CD51/CD61 subpopulations in the most frequent pattern (pattern 1). **i**, Co-immunostaining of keratin 14 (K14) and vimentin (Vim) in cytospin of FACS-isolated tumour cells. Scale bars, 20 μ m; $n = 6$. **j**, Cytospin counts of TCs based on K14 and vimentin expression (average of 90 cells per condition and tumour). **k**, PCA of scRNA-seq of Epcam⁺ and Epcam⁻ TCs. Dots represent single cells; colour scale represents the normalized expression of each gene ($n = 66$ Epcam⁺ and $n = 277$ Epcam⁻ cells from one tumour). **l**, Proportion of single cells expressing Epcam, CD51, CD61 and CD106 markers. **m**, PCA plot coloured by expression of the markers used to define EMT subpopulations by FACS. Grey circle, epithelial tumour cells; blue, hybrid cells; yellow, early EMT cells; red, late EMT cells. **n**, Cumulative density of each marker combination across PC1, which correspond to the degree of EMT (**j**).

some Epcam⁻ tumour subpopulations (triple-negative, CD106⁺ and CD51⁺) continued to express K14 and vimentin, whereas other Epcam⁻ subpopulations (CD51⁺CD61⁺ and triple-positive) were essentially K14⁻vimentin⁺, with rare TCs expressing low levels of K14 (Fig. 1i, j). These data indicate that the subpopulations identified during spontaneous EMT of primary skin tumours correspond to different tumour populations with different degrees of EMT, with some subpopulations corresponding to the hybrid tumour phenotypes with epithelial and mesenchymal features that have been described in cancer cell lines in vitro^{3–7}.

To further assess cellular heterogeneity during spontaneous EMT, we performed single cell RNA sequencing (scRNA-seq) of FACS-isolated YFP⁺Epcam⁺ and YFP⁺Epcam⁻ TCs. Dimensionality reduction using principal component analysis (PCA) revealed that the first principal component (PC1), which explained 21% of the variability, could be attributed to EMT state (Fig. 1k). Our scRNA-seq data confirmed that Epcam⁻ subpopulations showed greater transcriptional heterogeneity than Epcam⁺ subpopulations (Fig. 1k). The expression of epithelial and mesenchymal markers at the single-cell level confirmed the progressive acquisition of EMT features with epithelial, mesenchymal and hybrid states (Extended Data Fig. 3) and the distribution of CD51, CD61 and CD106 markers correlated with the degree of EMT along PC1, confirming the EMT gradient found across the different tumour subpopulations (Fig. 1l–n).

To assess whether these different subpopulations of EMT TCs reflect a more general mechanism occurring during EMT, we assessed the expression of these cell surface markers in metaplastic-like mammary tumours arising from oncogenic *Pik3ca* expression and *p53* deletion and in *MMTV-PyMT* mammary luminal tumours, which have been reported to present EMT features^{13–16}. Notably, a subset of mammary metaplastic-like and *MMTV-PyMT* luminal tumours also contained Epcam⁺ and Epcam⁻ TCs that could be subdivided into the same six subpopulations as found in *KRas*^{G12D}/*p53*^{KO} skin tumours (Extended Data Figs. 4a–d, 5a–c). Immunostaining on cytospin and real-time PCR with reverse transcriptase (RT-PCR) showed that the subpopulations isolated from the mammary tumours presented different degrees of EMT, similar to those identified in skin SCCs (Extended Data Figs. 4e, f, h, 5d, e, g), demonstrating that the different EMT transition states identified here represent a conserved mechanism during EMT.

To investigate whether these EMT transition states exist in human cancers, we assessed the expression of epithelial and mesenchymal markers in tumours derived from xenotransplantation (PDX) of poorly differentiated human breast cancers and SCCs. After several passages in immunodeficient mice, human stroma is entirely replaced by mouse cells¹⁷, making it possible to differentiate human TCs that underwent EMT and lost the expression of epithelial markers from mouse stroma using an antibody against human antigen (Ku-80). We detected areas expressing only epithelial markers, areas co-expressing epithelial and mesenchymal markers and areas expressing exclusively mesenchymal markers in poorly differentiated breast cancer, lung and oesophageal SCCs (Extended Data Fig. 6). These data demonstrate that EMT in human cancers is associated with different transition states including hybrid states, as was suggested by scRNA-seq of human SCCs¹⁸.

Stemness and plasticity of EMT states

EMT has been associated with cancer stemness, characterized by an increase in tumour-propagating cell (TPC) frequency^{1,8,19,20}. As previously described⁸, Epcam⁻ TCs contained five times as many TPCs as did Epcam⁺ TCs (Fig. 2a). Notably, all EMT subpopulations presented similar TPC frequencies (Fig. 2a). These data show that the earliest EMT state already exhibits increased TPC frequency, and tumour stemness does not increase further in later transition states. Whereas Epcam⁺ TCs showed higher proliferation than Epcam⁻ TCs, there was no difference in proliferation rate between the different EMT subpopulations (Fig. 2b). Thus, TPC frequency is inversely correlated to in vivo proliferation.

Cancer cells have been shown to be plastic in transplantation assays, with different tumour subpopulations able to transit back and forth between different states and to recapitulate primary tumour heterogeneity²¹. To determine whether the different subpopulations identified during EMT are similarly plastic, we analysed the tumour phenotypes of secondary tumours. As previously described⁸, secondary tumours arising from the transplantation of Epcam⁻ TCs comprise only Epcam⁻ TCs (Fig. 2c). Although all EMT subpopulations presented a certain degree of plasticity, at early time points following transplantation (3–4 weeks), the triple-negative subpopulation was relatively primed towards the epithelial phenotype and preferentially gave rise to

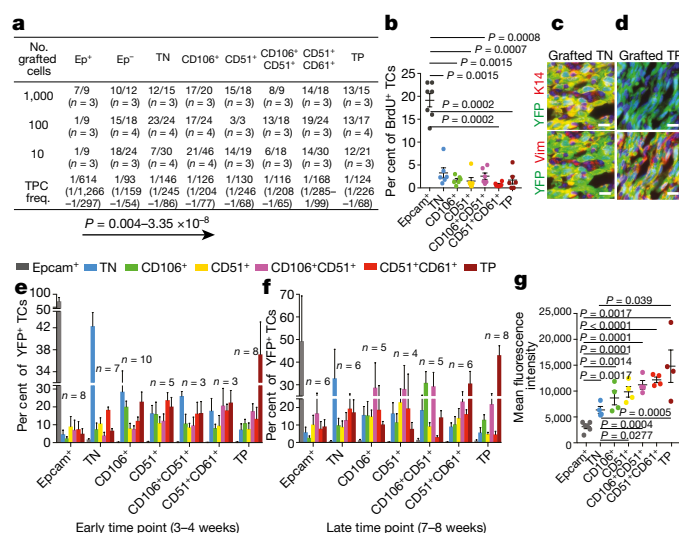


Fig. 2 | EMT transition states present similar TPC capacity but exhibit differing plasticity. a, Frequency of secondary tumours observed upon transplantation of limiting dilution of TC subpopulations and the estimation of TPC frequency (χ^2 test). **b**, FACS quantification of BrdU incorporation in TC subpopulations. Mean \pm s.e.m., two-tailed *t*-test. **c**, **d**, Immunofluorescence of YFP and K14 or vimentin in secondary tumours arising after subcutaneous transplantation of TN (**c**) or TP (**d**) cells ($n = 3$; scale bars, 20 μ m). **e**, **f**, Proportion of each subpopulation in secondary tumours 3–4 weeks (**e**) and 7–8 weeks (**f**) after subcutaneous transplantation of the different subpopulations (mean \pm s.e.m.). **f**, Invasion ability of different subpopulations measured by in vitro cell invasion assay. Mean fluorescence intensity correlates with the number of cells that migrate through the ECM. Two-tailed *t*-test, mean \pm s.e.m., $n = 3$.

triple-negative TCs. By contrast, the most mesenchymal triple-positive subpopulation preferentially gave rise to triple-positive TCs (Fig. 2c–e). The other EMT subpopulations were plastic, giving rise to each other in similar proportions (Fig. 2c). Notably, at later time points (7–8 weeks), the triple-negative subpopulation gave rise to different EMT subpopulations, whereas the triple-positive subpopulation gave rise preferentially to triple-positive cells (Fig. 2f). The different tumour subpopulations had different invasive capacities, which increased with the degree of EMT (Fig. 2f). Together, these data show that the different EMT tumour subpopulations are functionally distinct and have different clonogenic potentials, invasive properties, differentiation abilities and plasticity phenotypes.

Metastasis of EMT transition states

It remains unclear whether EMT is associated with increased metastatic potential^{11,14,22–25}. We have previously shown that Epcam⁻ cells are more able than Epcam⁺ TCs to metastasize to the lung⁸. To determine whether some EMT subpopulations showed increased ability to undergo vascular extravasation, lung colonization and metastasis, we injected these subpopulations of TCs intravenously into mice and assessed their ability to give rise to lung metastasis. Triple-negative and CD106⁺ hybrid subpopulations gave rise to substantially more metastases than did other EMT subpopulations (Fig. 3a). The increase in the metastatic potential of the hybrid states was also observed in mammary metaplastic-like and MMTV-PyMT luminal tumours (Extended Data Figs. 4g, 5f). Whereas CD106 has been shown to be associated with increased metastatic potential in human cancer cell lines^{9,11}, our data show that not all subpopulations of cells expressing CD106 are equally metastatic; their metastatic potential correlated more with the hybrid EMT state than the level of CD106 expression (Fig. 3a). As intravenous injection does not allow us to investigate the ability of cells to leave primary tumours and intravasate into blood vessels, we assessed the presence of YFP⁺ circulating tumour cells (CTCs) in the blood of *Lgr5Cre/ERKras*^{G12D}/*p53*^{KO} mice. The majority of CTCs in this mouse model underwent EMT, and the majority

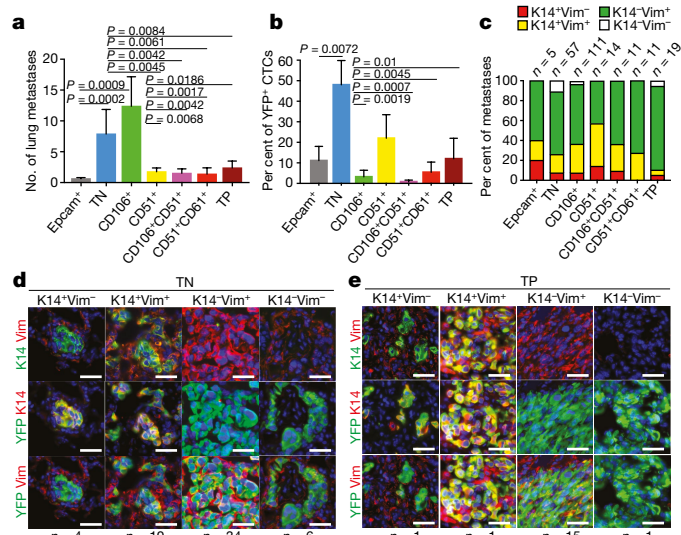


Fig. 3 | Different EMT transition states present distinct metastatic potential. a, Number of lung metastases arising from the injection of 500 YFP⁺ TCs with different degrees of EMT ($n = 8$). **b**, Phenotype of YFP⁺ CTCs based on expression of Epcam, CD106, CD51 and CD61 ($n = 10$). **a**, **b**, Two-tailed Mann–Whitney test, mean \pm s.e.m. **c**, Classification of metastasis based on K14 and vimentin expression (mean percentage). **d**, **e**, Co-immunostaining of YFP and K14 or vimentin in metastases arising from intravenously injected TN (**d**) or TP (**e**) TCs. Scale bars, 50 μ m.

of Epcam⁻ CTCs were triple-negative (Fig. 3b), further supporting the notion that the hybrid EMT state is associated with increased metastatic potential.

Mesenchymal-to-epithelial transition (MET) has been associated with increased metastatic potential in several mouse cancer models that are induced to undergo EMT by forced overexpression of EMT transcription factors such as Twist1 and Prrx1^{26,27}. Whereas subcutaneously injected Epcam⁻ cells did not give rise to Epcam⁺ TCs in secondary tumours, all Epcam⁻ subpopulations were able to revert back to Epcam⁺ TCs when colonizing the lung (Fig. 3c–e), showing that even the most extreme EMT states are not irreversibly locked into the mesenchymal state and can undergo MET within the lung microenvironment. However, the number of metastases did not correlate with the greater ability of these populations to undergo MET (Fig. 3c), suggesting that other mechanisms beside MET contribute to the higher metastatic potential of these hybrid populations.

Molecular characterization of EMT states

To unravel the molecular mechanisms associated with the different EMT tumour subpopulations, we performed RNA-seq and assay for transposase-accessible chromatin using sequencing (ATAC-seq) of the different FACS-isolated tumour subpopulations to define their transcriptional and chromatin landscapes. RNA-seq revealed that the level of expression of different epithelial markers, including cytoskeletal elements, adhesion molecules and epithelial transcription factors, progressively decreased from Epcam⁺ to triple-negative, CD106⁺ and CD51⁺ cells, and then remained very low in CD106⁺CD51⁺, CD51⁺CD61⁺ and triple-positive subpopulations (Fig. 4a). By contrast, some mesenchymal markers increased strongly with the loss of Epcam expression (in triple-negative cells), and remained stable, whereas other EMT markers such as *Fn1*, *Prrx1*, *Col3a1* and *Lox* presented a clear gradient of expression with maximal expression in triple-positive cells (Fig. 4b). This was further illustrated by unsupervised clustering analysis of the RNA-seq data (Extended Data Fig. 7a).

To further delineate the underlying molecular differences between the different EMT states, we performed unsupervised clustering analysis of the scRNA-seq data. Using SC3 clustering, we identified different tumour populations including epithelial, hybrid and mesenchymal populations (Extended Data Fig. 8a, b). Pseudotime ordering using

Monocle and branch expression analysis modelling (BEAM) showed that the transition from epithelial to mesenchymal states branches off into two different mesenchymal cell fates with distinct gene expression signatures (Extended Data Fig. 8c–e), demonstrating the extent of transcriptional heterogeneity in EMT.

To unravel the changes in chromatin landscape that underlie the modifications in gene expression observed in the different EMT subpopulations, we used ATAC-seq to identify chromatin regions that were specifically remodelled during the different transition stages in FACS-isolated subpopulations. We identified enhancers in epithelial genes, such as *Epcam* or *E-cadherin* (*Cdh1*) that were lost during the first EMT transition ($Epcam^+$ to triple-negative; Fig. 4c). The chromatin regions of epithelial genes that remained expressed in the hybrid populations, such as *Krt14* and *Krt17*, were still open, although at a reduced level, in the triple-negative subpopulation (Fig. 4d). The enhancer regions of common $Epcam^-$ EMT genes, such as *Vimentin* (*Vim*) and *Zeb1*, presented the same open chromatin regions as soon as TCs lost *Epcam* expression (Fig. 4e). Genes whose expression increases continuously during EMT, such as *Col24a1* and *Aspn*, have specific chromatin regions that become progressively more open as EMT progresses (Fig. 4f).

To define in a more quantitative manner, and at the genome-wide level, the chromatin remodelling that occurs within each tumour subpopulation and its global effect on gene expression, we compared the ATAC-seq peaks that are differentially regulated and associated with a change in gene expression between the different EMT subpopulations. These data showed stepwise and very specific chromatin remodelling associated with the different EMT states in vivo (Fig. 4g). Unbiased clustering analyses of the chromatin remodelling, normalized or unnormalized for gene expression, supported this notion and showed that TCs fell into three clusters: $Epcam^+$ and triple-negative, $CD106^+$ and $CD106^+CD51^+$, $CD51^+CD61^+$ and triple-positive (Fig. 4h), supporting the different degrees of EMT of these tumour subpopulations.

Motif discovery analyses of ATAC-seq peaks that are differentially regulated between the various EMT populations enabled us to predict which transcription factors are likely to regulate gene expression and cell fate transition during EMT. Common core motifs composed of AP1, Ets, Tead and Runx motifs were statistically highly enriched in the chromatin remodelling that occurred at every transition step during EMT or MET (Extended Data Fig. 7b–j), suggesting that at each transition during EMT or MET, the same transcription factors are required to induce chromatin remodelling irrespective of the degree of EMT. In addition to this common core of transcription factors, specific transcription factor motifs are associated with chromatin remodelling at different EMT transition states (Extended Data Fig. 7b–j).

To investigate the functional relevance of these in silico predictions and the gene regulatory networks that regulate EMT transition states, we studied the different EMT subpopulations in hair follicle-derived SCCs in which the expression of $\Delta Np63$ is sustained independently of EMT⁸ (Fig. 4i). Sustained $\Delta Np63$ expression inhibited the $Epcam^+$ to $Epcam^-$ transition and most $Epcam^-$ TCs were blocked in the early hybrid state, as demonstrated by the increase in triple-negative versus other EMT subpopulations (Fig. 4j, k, Extended Data Fig. 9a). Transcriptional profiling of FACS-isolated GFP^+Epcam^+ and GFP^+Epcam^- triple-negative TCs after sustained expression of $\Delta Np63$ induced the upregulation of 452 genes, 50% of which belonged to the tumour epithelial signature ($P < 2 \times 10^{-16}$), demonstrating the key role of $\Delta Np63$ in regulating early EMT hybrid states (Fig. 4l). To further challenge the in silico prediction of the role of TGF- β and Smad2, we administered pan-blocking antibodies recognizing TGF- β 1, TGF- β 2 and TGF- β 3 to *Lgr5Cre/ERKras^{G12D}/p53^{KO}* mice after inducing onco-gene expression. We found that anti-TGF- β antibodies accelerated tumour appearance, consistent with the cytostatic effect of TGF- β ²⁸. However, these tumours had a much more differentiated phenotype than tumours from control mice, with an increase in $Epcam^+$ and triple-negative populations and a decrease in the triple-positive

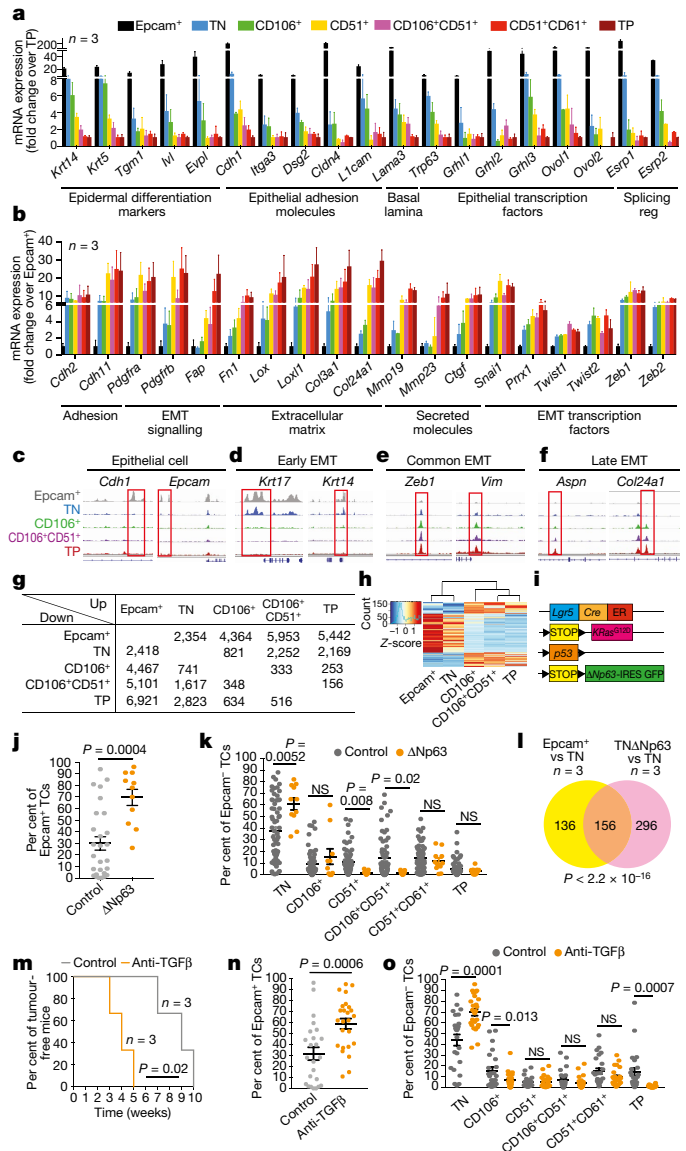


Fig. 4 | Transcriptional and chromatin landscape of EMT transition states.

a, b, mRNA expression of epithelial genes (**a**) and mesenchymal genes (**b**) in the different subpopulations as defined by RNA-seq ($n = 3$, mean \pm s.e.m.). **c–f**, ATAC-seq profiles showing increasing accessibility of chromatin regions that are specifically remodelled from $Epcam^+$ to TN cells (**c**), from TN to $CD106$ cells (**d**), in all $Epcam^-$ cells (**e**) and during the late stage of EMT (**f**). **g**, ATAC peaks up- or downregulated more than twofold and associated with genes up- or downregulated more than twofold in each tumour subpopulation. **h**, Unsupervised hierarchical clustering analysis of ATAC-seq peaks ($n = 1$) associated with genes differentially regulated more than twofold ($n = 3$). **i**, Genetic mouse model of hair follicle-derived SCCs allowing sustained $\Delta Np63$ expression. **j, k**, Percentage of $Epcam^+$ cells (**j**) and distribution of the different EMT subpopulations in control and $\Delta Np63$ -overexpressing TCs ($n = 27$ tumours from $n = 10$ control mice and $n = 11$ tumours from $n = 6$ $\Delta Np63$ mice). **l**, Venn diagram of the genes upregulated by $\Delta Np63$ in triple-negative TCs and naturally upregulated in $Epcam^+$ as compared to triple-negative TCs. Two-sided hypergeometric test. **m**, Kaplan–Meier plot of tumour appearance in control mice and mice treated with anti-TGF- β antibodies (log-rank Mantel–Cox). **n, o**, Proportion of $Epcam^+$ cells (**n**) and distribution of the different EMT subpopulations (**o**) in total YFP⁺ tumour cells in mice treated with anti-TGF- β and control antibodies ($n = 24$ tumours from three control mice and $n = 25$ tumours from three anti-TGF- β treated mice). **j, k, n, o**, Two-tailed t -test (mean \pm s.e.m.).

population (Fig. 4m–o, Extended Data Fig. 9b–d). These functional experiments validate our in silico predictions and show that $\Delta Np63$ promotes the epithelial and early EMT hybrid states, while TGF- β and

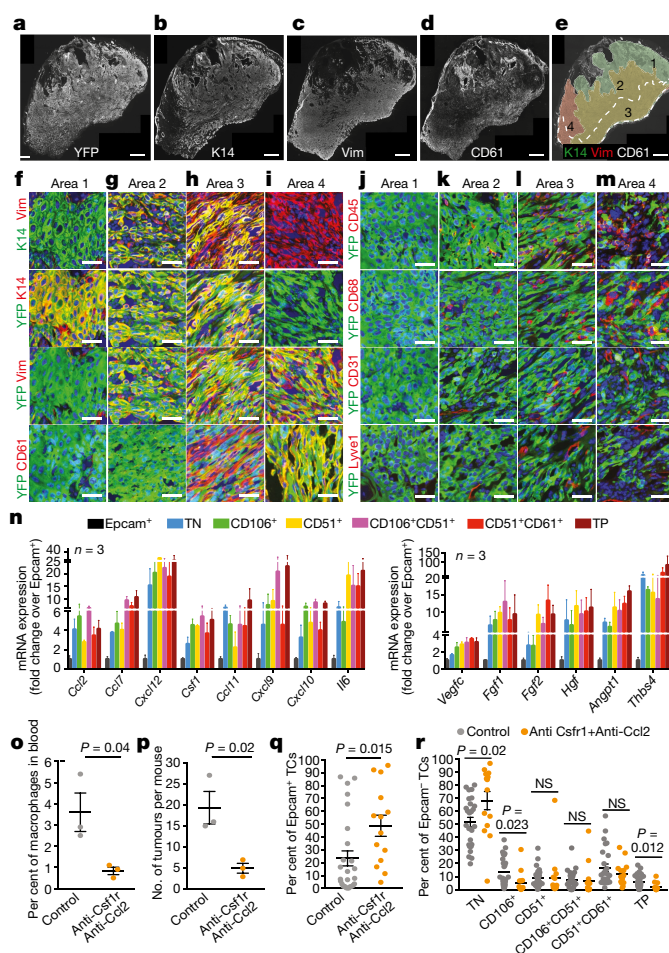


Fig. 5 | Different EMT transition states are localized in different niches. **a–d**, Mosaic images of immunostaining for YFP (**a**), K14 (**b**), vimentin (**c**) and CD61 (**d**) in a representative mixed SCC ($n = 5$). **e**, Definition of four areas based on the expression of K14, vimentin and CD61. Red, area expressing vimentin; green, area expressing K14; yellow, area expressing vimentin and K14; white outline, area expressing CD61. **f–i**, Expression of K14, vimentin, YFP and CD61 in area 1 (**f**), area 2 (**g**), area 3 (**h**) and area 4 (**i**). **j–m**, Expression of YFP and CD45, CD68, CD31 or Lyve1 in area 1 (**j**), area 2 (**k**), area 3 (**l**) and area 4 (**m**). Scale bars, 50 μm . **n**, mRNA expression of chemokine, pro-inflammatory and pro-angiogenic genes by TCs (mean \pm s.e.m.). **o**, **p**, Percentage of F4/80⁺CD11b⁺ macrophages in peripheral blood (**o**) and number of tumours (**p**) in control ($n = 3$) and anti-Csf1r/Ccl2-treated ($n = 3$) mice. **q**, **r**, Proportion of Epcam⁺ TCs (**q**) and distribution of different EMT subpopulations (**r**) in control and anti-Csf1r/Ccl2-treated mice ($n = 26$ tumours from $n = 3$ control mice and $n = 15$ tumours from $n = 3$ anti-Csf1r/Ccl2-treated mice). **o–r**, Two-tailed t -test, mean \pm s.e.m.

Smad2 promote the transition from the Epcam⁺ and triple-negative states towards more mesenchymal states.

Different niches of the EMT states

To determine whether these different tumour subpopulations are localized in particular areas within tumours, we performed immunostaining with markers characteristic of the distinct tumour subpopulations. YFP expression marked TCs irrespective of their degree of EMT. Mixed tumours could be subdivided into distinct areas corresponding to different parts of the tumour, with differing degrees of EMT based on the expression of K14, vimentin and CD61 (Fig. 5a–e). Area 1 corresponded to the pure epithelial part of the tumours. In area 2, the TCs were more elongated, although they were still cohesive. K14 and vimentin were co-expressed, marking the hybrid epithelial and mesenchymal states of the tumours. In area 3, TCs were further elongated with a more fibroblast-like appearance; some TCs did not show cell–cell

adhesion, expressed vimentin and had lost K14 expression. Most of the cells in this area expressed CD61. In area 4, YFP⁺ TCs were even more fibroblastic, did not form cell–cell junctions, did not express K14 and uniformly expressed vimentin and CD61 (Fig. 5f–i). These data reveal that the different EMT tumour subpopulations are spatially organized in specific regions rather than being randomly distributed throughout tumours.

To determine whether the TC subpopulations described here are localized in particular microenvironments, we investigated whether they were associated with specific stromal populations. In area 1, endothelial cells (CD31⁺), inflammatory cells (CD45⁺) and cancer-associated fibroblasts (YFP⁺ vimentin⁺) were relatively rare and surrounded the epithelial part of the tumour. By contrast, the composition and the localization of stromal cells changed markedly as TCs progressed towards EMT. Cells that were in close contact with YFP⁺ EMT TCs showed a large increase in the density of CD45⁺ immune cells, particularly monocytes and macrophages expressing CD68, and an increase in the density of endothelial and lymphatic cells (Fig. 5j–m). These different areas—composed of epithelial, hybrid and mesenchymal cells with increased vascularization and inflammatory cells—were also observed in *Pik3ca/p53*^{KO} and *MMTV-PyMT* mammary tumours (Extended Data Figs. 4i–p, 5h–m). Notably, the immune infiltration density correlated with the expression of chemokines and other pro-inflammatory and pro-angiogenic molecules by TCs (Fig. 5n), suggesting that TCs attract and regulate the stromal cells that comprise their niche and contribute directly to the spatial organization of tumour subpopulations.

To determine the functional relevance of the macrophage infiltration in regulating EMT progression, we assessed the effect of macrophage depletion on the proportions of EMT tumour subpopulations. Administration of anti-Csf1r and anti-Ccl2 blocking antibodies during tumorigenesis efficiently depleted circulating and tumour-infiltrating macrophages and decreased tumour formation; in addition, it increased the proportion of Epcam⁺ and triple-negative TCs, and decreased the proportion of triple-positive TCs (Fig. 5o–r, Extended Data Fig. 10). These results demonstrate the functional importance of the tumour microenvironment and macrophage infiltration in regulating the transition between EMT states.

Discussion

Our study demonstrates that spontaneous EMT in primary TCs in vivo proceeds through distinct intermediate states with different invasive, metastatic and differentiation characteristics. TCs with hybrid epithelial and mesenchymal phenotypes are more efficient in reaching the circulation, colonizing the lungs and forming metastases. Transcriptional, chromatin and scRNA profiling demonstrate that these different states have distinct cellular properties, chromatin landscapes and gene expression signatures that are regulated by common and distinct transcription factors and signalling pathways. In addition, these different EMT states are localized in different microenvironments and in contact with different stromal cells. The most mesenchymal subpopulations are localized close to endothelial and inflammatory cells. These TCs secrete high levels of chemokines and proteins that attract immune cells and promote angiogenesis, thereby organizing the formation of their own inflammatory and highly vascularized niche; this in turn further promotes EMT in these subpopulations. Our results identify different EMT transition states in vivo that are associated with different tumour functions, and have important implications for our understanding of tumour heterogeneity, growth, invasion, metastasis and resistance to therapy. The approaches used here to unravel EMT-related tumour heterogeneity can be used to define cellular heterogeneity in other biological systems.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0040-3>.

Received: 20 June 2017; Accepted: 5 March 2018;

Published online: 18 April 2018

- Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45 (2016).
- Puisieux, A. & Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. *Nat. Cell Biol.* **16**, 488–494 (2014).
- Huang, R. Y. et al. An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). *Cell Death Dis.* **4**, e915 (2013).
- Zhang, J. et al. TGF- β -induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci. Signal.* **7**, ra91 (2014).
- Hong, T. et al. An *Ovol2-Zeb1* mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput. Biol.* **11**, e1004569 (2015).
- Jolly, M. K. et al. Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* **7**, 27067–27084 (2016).
- Bierie, B. et al. Integrin- $\beta 4$ identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. *Proc. Natl Acad. Sci. USA* **114**, E2337–E2346 (2017).
- Latil, M. et al. Cell-type-specific chromatin states differentially prime squamous cell carcinoma tumor-initiating cells for epithelial to mesenchymal transition. *Cell Stem Cell* **20**, 191–204 (2017).
- Chen, Q., Zhang, X. H. & Massagué, J. Macrophage binding to receptor VCAM-1 transmits survival signals in breast cancer cells that invade the lungs. *Cancer Cell* **20**, 538–549 (2011).
- Seguin, L. et al. An integrin $\beta 3$ -KRAS-RalB complex drives tumour stemness and resistance to EGFR inhibition. *Nat. Cell Biol.* **16**, 457–468 (2014).
- Slack-Davis, J. K., Atkins, K. A., Harrer, C., Hershey, E. D. & Conaway, M. Vascular cell adhesion molecule-1 is a regulator of ovarian cancer peritoneal metastasis. *Cancer Res.* **69**, 1469–1476 (2009).
- Wang, J. et al. CD51 correlates with the TGF- β pathway and is a functional marker for colorectal cancer stem cells. *Oncogene* **36**, 1351–1363 (2017).
- Van Keymeulen, A. et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* **525**, 119–123 (2015).
- Ye, X. et al. Upholding a role for EMT in breast cancer metastasis. *Nature* **547**, E1–E3 (2017).
- Ye, X. et al. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256–260 (2015).
- DelPozo Martin, Y. et al. Mesenchymal cancer cell-stroma crosstalk promotes niche activation, epithelial reversion, and metastatic colonization. *Cell Reports* **13**, 2456–2469 (2015).
- Hidalgo, M. et al. Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov.* **4**, 998–1013 (2014).
- Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624 (2017).
- Mani, S. A. et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
- Ye, X. & Weinberg, R. A. Epithelial-mesenchymal plasticity: a central regulator of cancer progression. *Trends Cell Biol.* **25**, 675–686 (2015).
- Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337 (2013).
- Diepenbruck, M. & Christofori, G. Epithelial-mesenchymal transition (EMT) and metastasis: yes, no, maybe? *Curr. Opin. Cell Biol.* **43**, 7–13 (2016).
- Zheng, X. et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).
- Fischer, K. R. et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).
- Aiello, N. M. et al. Upholding a role for EMT in pancreatic cancer metastasis. *Nature* **547**, E7–E8 (2017).
- Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. & Yang, J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* **22**, 725–736 (2012).
- Ocaña, O. H. et al. Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer *Prrx1*. *Cancer Cell* **22**, 709–724 (2012).
- Siegel, P. M. & Massagué, J. Cytostatic and apoptotic actions of TGF- β in homeostasis and cancer. *Nat. Rev. Cancer* **3**, 807–821 (2003).

Acknowledgements We thank the ULB animal facility, the ULB genomic core facility (F. Libert and A. Lefort), G. Berx (Ghent University) and M. Mazzone (VIB-KUL) for MMTV-PyMT mice. I.P. is supported by TELEVE. P.A.S. is supported by FNRS. A.S., D.B. and T.V. are supported by KULeuven SymBioSys, Stichting Tegen Kanker, FWO postdoctoral fellowship #12W7318N and Marie Skłodowska-Curie fellowship #1205617N. The Department of Pathology acknowledges Fonds Yvonne Boel. The PDX platform is supported by Fonds Erasme. C.Bla. is supported by WELBIO, FNRS, Fondation Contre le Cancer, ULB Foundation, European Research Council, Worldwide Cancer Research and the Foundation Baillet Latour.

Reviewer information Nature thanks B. E. Bernstein, E. Thompson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions I.P. and C.Bla. designed the experiments and performed data analysis. I.P. performed most of the biological experiments. P.A.S. performed part of the FACS analysis and histological characterization. A.B. performed the bioinformatic analysis. A.S., D.B. and T.V. performed scRNA-seq and analysis. M.F., A.V.K. and J.-C.M. helped with breast cancer experiments. T.R. performed CTC experiments. S.B. helped with ATAC-seq. I.S., N.D'H., S.D.C., E.M., C.Bal., Y.S., A.L. and F.S. provided human samples. C.D. performed FACS sorting. V.M., S.L., F.D.C. and S.S. performed immunostaining, blocking antibody injections and follow-up of the mice. All authors read and approved the final manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0040-3>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0040-3>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to C.B. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.