Pre-cancerous Niche Remodelling Dictates Nascent Tumour Persistence

G. Skrupskelyte[1,2,#*], J. E. Rojo Arias[1,2,3#], H. Ajith[1,2], Y. Dang[4,5,6], D. Rossetti [1,2], S. Han[7], M. K. S. Tang[1,2], M. T. Bejar[1,2], B. Colom[8,9], J. C. Fowler[8], K. Murai[8], W. Knight[8], D. Aust[10,11,], M. H. H. Schmidt[12], J. Jászai[12], S. Zeki[13], A. Noorani[8,14], P. H. Jones[8,15], S. Rulands[5,16], B. D. Simons[1,7,17], M. P. Alcolea[1,2,*]

Affiliations:

[1]Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, Puddicombe Way, Cambridge, CB2 0AW, UK.

[2]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK.

[3]Currently at RhyGaze AG, Mittlere Strasse 91, 4031 Basel, Switzerland.

[4]Max Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307, Dresden, Germany

[5]Max Planck Institute for the Physics of Complex Systems, Noethnitzer Str. 38, 01187 Dresden, Germany

[6]Center for Systems Biology, Pfotenhauerstrasse 108, 01307, Dresden, Germany

[7]Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK.

[8]Wellcome Sanger Institute, Hinxton CB10 1SA, UK

[9]Currently at Cambridge Institute of Science, Altos Labs, Cambridge CB21 6GP, UK

[10]Institut für Pathologie/Tumor and Normal Tissue Bank (TNTB), Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, Germany

[11]Institute of Pathology, University Hospital CGC Dresden, TU Dresden, Dresden, Germany

[12]Institut für Anatomie, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Medizinische Fakultät, Fetscherstraße 74, 01307 Dresden, Germany

[13]Guy's and St Thomas' NHS Foundation Trust, London, UK

[14]Addenbrooke's Hospital, Cambridge University Hospital NHS Trust, Cambridge, UK.

[15]Department of Oncology, University of Cambridge, Hutchison Research Centre, Cambridge Biomedical Campus, Cambridge CB2 0XZ, UK

[16]Ludwigs-Maximilians-Universität München, Arnold Sommerfeld Center for Theoretical Physics, Theresienstr. 37, 80333 München, Germany

[17]Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Science, University of Cambridge, Cambridge, CB3 0WA, UK.

[#]These authors contributed equally to this work.

*Corresponding authors: gs463@cam.ac.uk (G.S.); mpa28@cam.ac.uk (M.P.A.).

Title

Pre-cancerous Niche Remodelling Dictates Nascent Tumour Persistence

Summary

Interactions between mutant cells and their environment play a key role in determining cancer susceptibility[1-3]. However, our understanding of how the pre-cancerous microenvironment contributes to early tumorigenesis remains limited. Here, we show that newly emerging tumours at their most incipient stages shape their microenvironment in a critical process that determines their survival. Analysis of nascent squamous tumours in the upper gastrointestinal tract of the mouse reveals that the stress response of early tumour cells instructs the underlying mesenchyme to form a supportive "pre-cancerous niche", which dictates the long-term outcome of epithelial lesions. Stimulated fibroblasts beneath emerging tumours activate a wound healing response that triggers a dramatic remodelling of the underlying extracellular matrix, resulting in the formation of a fibronectin-rich stromal scaffold that promotes tumour growth. Functional heterotypic 3D culture assays and *in vivo* grafting experiments, combining carcinogen-free healthy epithelium and tumour derived stroma, demonstrate that the pre-cancerous niche alone is sufficient to confer tumour properties to normal epithelial cells. We propose a model where both mutations and the stromal response to genetic stress together define the likelihood of early tumours to persist and progress towards more advanced disease stages.

Background

Groundbreaking studies in human genomics over the past decade have revealed that our healthy tissues accumulate cancer-associated mutations with age[4-8]. These observations highlight new levels of complexity in the early pathophysiology of cancer, questioning what additional factors, beyond cancer mutations, may be playing a role during early carcinogenesis.

Early tumour models spanning a range of epithelial tissues, including oesophagus, skin, and intestine, have started to offer a clearer understanding of what drives tumour initiation[1-3,9-11]. Work in this area has shown that tumour formation represents more than the mere accumulation of genetic alterations, highlighting the important role that environmental cues and non-genetic mechanisms play in this process[3,12-15]. Indeed, mounting evidence indicates that the predisposition of a mutated epithelium to develop tumoral lesions depends on the complex interactions between mutant cells and their dynamic surroundings. Coexisting mutant clones may either synergise or compete, contributing to early tumour initiation[16,17]. And, even after tumours are formed, the presence of neighbouring mutant clones may continue to influence tumorigenesis[1]. Alternative environmental cues such as ECM stiffness[3,13], as well as direct cell-cell communication between mutant and non-mutant cells[2,9-11,14], have also been shown to impact mutant clone expansion, susceptibility to tumour initiation, and invasion[18-20]. Despite this, our understanding of the mechanisms by which environmental factors determine the formation and long-term persistence of emerging tumours, remains limited.

Previous studies using an oesophageal early-tumour model demonstrated that not all nascent tumours have the same chance of survival. Most tumours are cleared from the tissue soon after formation through competition with neighbouring mutant clones. Surviving tumours, instead, persist long-term, becoming susceptible to cancer progression[1]. Yet, a key question remains: how are pre-cancerous tumours able to withstand the competitive mutant environment that surrounds them? Understanding the processes underlying early tumour persistence and the relevance of the microenvironment at pre-neoplastic stages provides a critical opportunity to dissect the mechanisms driving pre-cancer progression, opening new avenues to halt cancer in its tracks. Here, we combine single-cell RNA sequencing with lineage tracing and 3D heterotypic cultures to study the unique features of the few nascent tumours that escape existing protective barriers preventing tumorigenesis. We demonstrate that, during

the earliest stages of tumour development, fibroblasts react to the pre-neoplastic epithelium by promoting the formation of a fibrotic pre-cancer niche that, in turn, feeds back on the epithelium favouring early tumour growth and survival.

1. Pre-cancerous tumour persistence

To study the processes underlying the persistence of pre-neoplastic nascent tumours, we used a well-established, clinically relevant mouse model of upper gastrointestinal tract (including oesophagus and forestomach) tumorigenesis driven by a mutagen found in tobacco smoke (DEN; diethyl-nitrosamine)[1,21] (**Extended Data Fig. 1a-j**). Following DEN treatment, the tissue becomes an evolving patchwork of mutant clones competing for space and survival, recapitulating the complex mutational landscape of the normal human ageing oesophagus[22]. This results in the emergence of pre-neoplastic squamous tumours with the potential to persist long-term (**Extended Data Fig. 1a-c**).

Nascent epithelial tumours, marked by KRT6A (Keratin 6A) and KRT17 (Keratin 17)[1], can be detected in tissue whole-mounts from their most incipient stages, as early as 10 days (d) post-DEN administration (**Extended Data Fig. 1d-e**). Emerging tumours are microscopic, containing as few as 10 cells, and are characterised by their distinctive rosette-like structure[1,21] (**Extended Data Fig. 1d,f**). This brief window of formation is followed by a tumour clearing process, in which more than a third of initial tumours are progressively eliminated[1]. Surviving tumours may persist in the tissue for over a year, largely as low-grade dysplasia (pre-neoplastic/pre-cancer stages), with sporadic progression to invasive squamous cell carcinomas (**Extended Data Fig. 1c,g,h**)[1], mimicking human carcinogenesis[23]. As a result, only a subset of the original tumours survives long-term, allowing us to study the mechanisms modulating pre-cancerous tumour persistence.

To understand what drives early tumour survival, we first set out to compare the phenotypic traits of **nascent tumours** and those persisting **long-term** (10d and >8m post-DEN, respectively; **Fig. 1a**). Histological analysis showed that persistent dysplastic tumours (**Extended Data Fig. 1h**) were characterised by a prominent stromal remodelling (**Fig. 1b**. These nest-like structures were formed by stromal fibroblasts (PDGFRα+) that protruded

towards the epithelial compartment, seemingly enclosing early tumours to create a supportive scaffold or a "pre-cancerous niche" (**Fig. 1b-c**). Unlike in persisting tumours, at nascent stages most epithelial lesions (~70%, 199 out of 296) showed no apparent stromal reorganisation (**Fig. 1c-d)**, denoting the existence of two phenotypically different nascent tumour subtypes, hereafter referred to as **Niche+** and **Niche-** (**Fig. 1c**).

Next, we assessed the dynamic nature of these two nascent tumour subtypes. We found that the number of Niche+ tumours, despite constituting the minority of all initial tumours, remained constant over time, while the number of Niche- tumours decreased drastically (**Extended Data Fig. 2a**). As a result, the tissue became progressively enriched in Niche+ tumours, with the majority (~82%, 65 out of 79) showing a supportive stromal scaffold by 8m post-DEN (**Fig. 1d**). This enrichment in Niche+ tumours prompted us to explore whether stromal remodelling was associated with nascent tumour persistence. We found that Niche+ lesions were hyperproliferative, and more likely to persist and enlarge compared to Niche- tumours (**Fig. 1e-g; Extended Data Fig. 2b-f**). Close analysis revealed that keratinocytes in contact with the niche showed a particularly high proliferative activity (**Extended Data Fig. 2e**), suggesting an active epithelial-stromal communication at pre-cancerous stages. These results were further reinforced by observations in the squamous forestomach, where long-term pre-neoplastic tumours also exhibited a remodelled stromal niche (**Extended Data Fig. 2g,h**).

Collectively, our data support a model in which the remodelled stromal scaffold acts as a **"pre-cancerous niche**" promoting tumour growth and survival. These observations link stromal remodelling in nascent tumours with pre-neoplastic tumour progression.

2. Niche signals drive tumour traits

The importance of the pre-cancerous niche during early tumorigenesis became evident in 3D heterotypic cultures. These co-cultures revealed that signals from the early tumour stroma are sufficient to confer tumour features to epithelial cells never exposed to carcinogens[24] (**Extended Data Fig. 2i-l**). Using reporter mouse lines to track the tissue origin, we found that the untreated phenotypically normal epithelium directly exposed to the denuded tumour niche (lacking the epithelial compartment) acquired a tumour like morphology and became highly proliferative, reaching levels similar to early tumours *in vivo* (**Extended Data Fig. 2k,l**).

Moreover, heterotypic tissue constructs grafted into immune-deficient NOD scid gamma mice (NSG; NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ) showed that the pro-survival phenotype conferred by the tumour niche was also observed *in vivo*. Normal epithelial cells were more likely to engraft long-term when exposed to early tumour stromal signals (**Extended Data Fig. 2m-o**).

Overall, these results demonstrate that the early tumour microenvironment promotes epithelial cell growth, favouring pre-cancerous tumour survival and, ultimately, disease progression.

3. Local fibroblasts form nascent tumour niche

Given the key role of the niche in nascent tumour survival, we next explored its cellular composition.

Under normal conditions, the squamous upper GI tract is characterised by three distinct layers of stromal tissue. The lamina propria, a thin loose connective tissue directly beneath the epithelium; the muscularis mucosae, a layer of smooth muscle cells; and the submucosae, a dense irregular lower stromal compartment[25] (**Extended Data Fig. 3a-d**). In line with observations in other epithelial tissues[26,27], immunofluorescence analysis revealed two fibroblast populations that showed distinctive expression levels of the pan-fibroblast marker PDGFRα, tissue compartmentalisation and morphology (**Extended Data Fig. 3b-d**). PDGFRα^{low} fibroblasts resided in the upper stroma (the lamina propria), while PDGFRα^{high} fibroblasts populated the deepest stromal layer (the submucosae; **Extended Data Fig. 3c,d**). Histological analysis of emerging tumours revealed that the major cellular component of the pre-cancerous niche was PDGFRα^{low} fibroblasts, phenotypically indistinguishable from neighbouring lamina propria fibroblasts (**Fig. 2a**). Further characterisation showed that endothelial and immune cells were largely absent from the niche in nascent tumours (10d post-DEN; **Fig. 2b,c**).

Since the role of cancer-associated fibroblasts (CAFs) in tumorigenesis, drug resistance and disease progression is well-recognised[28,29], we next assessed whether niche forming fibroblasts exhibited CAF features. Except for nuclear localisation of YAP (active YAP; aYAP; **Extended Data Fig. 3e**), the expression of CAF markers such as FAP (Fibroblast activation protein) and α-SMA (alpha-smooth muscle actin) was not detectable in 10d tumour fibroblasts

(**Extended Data Fig. 3f**). We conclude that, while Niche+ fibroblasts in incipient tumours lack a full CAF phenotype, their aYAP status is consistent with a pre-CAF transitional state within the nascent tumour niche[30].

Another important stromal contributor to tumorigenesis and cancer progression is the immune compartment[31]. Immune cell characterisation of Niche+ and Niche- tumours, however, did not show significant differences (**Fig. 2b,c; Extended Data Fig. 3g**). Accordingly, the emergence and persistence of Niche+ and Niche- tumours remained unaltered in immune-deficient mice (NSG; **Extended Data Fig. 3h-l**). These results suggest that immune cells do not discriminate between nascent Niche+ and Niche- tumours, acting as bystanders in early tumour persistence.

4. Pre-cancerous stromal reorganisation

To better understand the contribution of stromal fibroblast to tumour niche formation, we used an unbiased genetic lineage tracing approach to target fibroblasts. Sporadic confetti labelling of fibroblasts, across the lamina propria and submucosae compartments, was induced in *Col1a2-Cre^{ER}; R26$^{FlConfetti/WT}$* mice followed by DEN administration (**Fig. 2d**; **Extended Data Fig. 4a-d**). To trace the origin of the tumour "niche", recombination of the confetti cassette was induced prior DEN administration (**Fig. 2d).** Analysis of confetti clones 6 months post-DEN showed that fibroblasts within the niche underwent clonal expansion (**Fig. 2e; Extended Data Fig. 4e-i; Supplementary Table 1**). Immunostaining further revealed that clones within the niche were formed by lamina propria PDGFRα^{low} fibroblasts (**Fig. 3e,f; Extended Data Fig. 4e**).

To confirm this, we traced the PDGFRα^{low} and PDGFRα^{high} fibroblast populations separately in *Pdgfra-Cre^{ER}; R26$^{FlConfetti/WT}$* mice (**Extended Data Fig. 4j-o**). We reasoned that differential PDGFRα expression levels would allow us to control the level of recombination in the lamina propria and submucosae. Paradoxically, the PDGFRα^{low} fibroblast population showed a markedly higher recombination efficiency, with negligible recombination detected in the lower PDGFRα^{high} compartment (**Extended Data Fig. 4j-m**), likely due to different tamoxifen accessibility between stromal layers[32].

The distinctive recombination efficiency allowed us to trace PDGFRα^{low} fibroblasts at early tumour stages (i.e., 6 weeks post-DEN) to interrogate their contribution to the niche formation process (**Extended Data Fig. 4n,o**). We observed that the early tumour niche was composed of PDGFRα^{low} derived fibroblast clones that expanded locally within the upper stromal compartment. No clonal expansion events were found in the lower stroma (PDGFRα^{high} compartment) or spanning across stromal compartments (**Extended Data Fig. 4o**).

Overall, these observations demonstrate that local PDGFRα^{low} fibroblasts in the lamina propria not only maintain but also contribute to the formation of the pre-neoplastic tumour niche.

5. Pro-fibrotic PDGFRα^{low} niche fibroblasts

To determine the mechanisms underlying the formation of the pre-cancerous niche, we set out to study epithelial-mesenchymal communication. Single-cell RNA sequencing (scRNA-seq) was performed in individually micro-dissected dysplastic tumours of the squamous upper GI tract 8m post-DEN (**Fig. 3a,b; Extended Data Fig. 5a-f; Supplementary Table 2,3**).

The transcriptional profile of pre-neoplastic tumour cells (**Tumour**) was compared to that of cells from adjacent tumour free areas in DEN-treated mice (**DEN**; internal control), and to that of cells from healthy untreated animals (Control, **Ctrl**; **Extended Data Fig. 5g**).

In line with our histological and lineage tracing characterisation (**Fig. 2; Extended Data Fig. 3c,d; 4**), the scRNA-seq analysis revealed a marked degree of heterogeneity in the expression of the fibroblast marker *Pdgfra* (**Fig. 3c**). Distinctive *Pdgfra^{low}* and *Pdgfra^{high}* fibroblast populations were present across conditions (**Supplementary Table 5**). *Pdgfra^{low}* fibroblasts expressed higher levels of genes encoding structural (scaffolding matrix) ECM components (such as *Fn1, Fbn1, Has1, Has2, Loxl2, Mfap5 Cd248, Col1a1*)[33], common in loose connective tissue. *Pdgfra^{high}* fibroblasts, instead, showed an enrichment of genes encoding ECM fibrillar (underlying matrix) collagens (*Col6a3, Col5a3*), vascular support collagens (such as *Col4a1/2, Col8a1, Col15a1, Col13a1*) and other basement membrane components (such as *Lama1, Thbs1/2*; **Fig 3d; Extended Data Fig. 6a-c; Supplementary Table 6**)[33]. Accordingly, immunolabelling of *Pdgfra* derived clones (from **Extended Data Fig. 4j**) in the lamina propria

and submucosae, respectively, evidenced increased fibronectin production in the *Pdgfra*^{low} fibroblast compartment (**Extended Data Fig. 6d**).

We went on to interrogate the tumour specific transcriptional signature of *Pdgfra*^{low} niche forming fibroblasts. Fibroblasts in Cluster 19 (C19), enriched in pre-neoplastic tumour stroma (**Extended Data Fig. 6e**), showed a significant upregulation of matrisome genes associated with wound healing/fibrosis (matrix deposition: *Fn1, Fbln5, Tnxb, Cd248, Vim, Plaur, Mfap5;* thrombospondins*: Thbs1, Thbs3;* collagens*: Col3a1, Col1a1/2, Col5a3,* remodelling*: Timp1, Adam9, Loxl1;* other factors*: Fgfr2, Bmp1, Bmp6, Cx3cr1*; **Extended Data Fig. 6f**; **Supplementary Table 7,8**)[9,34]. The pro-fibrotic nature of these niche fibroblasts was supported by immunolabelling of both nascent and surviving tumours (**Fig. 3e; Extended Data Fig. 6g**), as well as by second-harmonic generation (SHG) imaging, which showed a marked ECM remodelling in the tumour niche (**Extended Data Fig. 6h-j**). Accordingly, alterations in integrin α6 (CD49f) expression in nascent epithelial tumour cells hinted at 'wound healing-like' changes in epithelial–ECM interactions[35] (**Extended Data Fig. 6k**). Together, these findings pointed at the activation of a tissue repair response in the stroma of early tumours, in line with the long-standing notion that tumours are "wounds that do not heal"[36].

We next explored whether niche forming fibroblasts (C19) in long-term surviving tumours presented a CAF signature. Despite a subtle upregulation of a reduced subset of CAF-associated genes[12,28,37-39] (*Vim, S100a4, Mfap5, Col1a2;* **Extended Data Fig. 6l-n**), their expression at the protein level (VIM, FSP, FAP) remained largely unaltered or non-detectable (**Extended Data Fig. 6o**). Accordingly, fibroblast proliferation remained unaltered[28] (**Extended Data Fig. 6p**). These observations indicate that Niche+ fibroblasts in surviving dysplastic tumours lack a fully established CAF phenotype. However, the presence of nuclear YAP and the pro-fibrotic nature of Niche+ fibroblasts point to a pre-CAF signature (**Extended Data Fig. 3e**)[30], transitioning to myCAFs at more advanced stages (invasive squamous cell carcinomas 14 months post-treatment; **Extended Data Fig. 6q)**. Label transfer analysis from *Buechler et al*. fibroblast atlas[40] indicated that tumour niche fibroblasts likely derive from universal, rather than tissue-specific, fibroblast populations. *Pdgfra*^{low} fibroblasts, including tumour enriched cluster 19, partially recapitulated the transcriptional signature of the *Pi16*+ universal population, whereas *Pdgfra*^{high} fibroblasts aligned with the *Col15a1*+ universal fibroblast subset (**Extended Data Fig. 6r,s)**.

To assess whether stromal genetic alterations drive the tumour niche phenotype, we performed deep-targeted sequencing of 192 cancer-related genes (**Supplementary Table 9**)[22]. Results argued against somatic mutations in fibroblasts being responsible for pre-cancerous niche formation. The data revealed that DEN administration induces gene perturbations primarily in the epithelial compartment, showing a minimal mutational burden in the tumour stroma that matched the level of untreated or internal DEN control samples (**Extended Data Fig. 7a,b**).

In contrast to nascent stages (**Fig. 2b,c**), we found that as tumours progress, they showed a marked remodelling of the vascular network and an increased immune infiltrate (**Extended Data Fig. 7c,d**). In line with previous findings, transcriptional analysis revealed notable changes in the immune cell composition in long-term surviving tumours, indicative of active immune cell recruitment with progression towards an immunosuppressive microenvironment at later stages (**Extended Data Fig. 7e-h, Supplementary Table 10**)[41].

Taken together, our data revealed a significant stromal reorganisation in surviving tumours. Analysis of lamina propria derived *Pdgfra*^{low} fibroblasts was consistent with a notable fibrotic ECM remodelling in the pre-cancerous niche, prior to the emergence of a fully established CAF phenotype.

6. Nascent tumour heterogeneity

Next, we explored the epithelial transition from healthy/normal to pre-neoplastic states. Pseudotime analysis revealed two distinctive basal cell trajectories denoting tumour and non-tumour states. These trajectories largely converged in committed and differentiating cells (**Fig. 3f; Extended Data Fig. 8a-f**). Gene score enrichment analysis of tumour specific gene modules identified by pseudotime analysis revealed further heterogeneity within epithelial tumour states, referred to as **Tumour 1** and **Tumour 12** (**Fig. 3f; Extended Data Fig. 8d-h; Supplementary Table 11-13**), that expressed increased levels of early tumour markers *Krt6a* and *Krt17* [1,21] (**Extended Data Fig. 8i**).

Gene set enrichment analysis unveiled a unique signature in **Tumour 12** cells (**Extended Data Fig. 8j,k**), enriched for genes related to cancer-associated processes such as i) tissue

development/morphogenesis (*Cdh1, Cdh4, Sox4, Klf4, Sox9, Foxa1*), ii) hypoxia (*Hmox1, Vegfa, Edn1, Cited2, Sirt1*), as well as iii) stress induced pathways, including EGFR (*Areg, Hbegf, Nrg1, Nrg2, Egfr, Epha2*[42], *Lamc2*[43]), Hippo, TNF, p53 and TGF-β (*Tead1, Tnfaip3, Nfkbia, Ccng2, Rela, Nfkb1, Ptgs2, Gadd45a, Cdkn1a, Id1-4*) (**Extended Data Fig. 8k, Supplementary Table 14**). This was reinforced by increased levels of genes encoding transcription factors associated with a tumour stress response (*Jun, Fos, Fosb, Runx1, Atf3, Egr1, Egr3, Myc*)[41,44-48] (**Extended Data Fig. 8l**). Crucially, validation at the protein level further supported the heterogenous nature of the epithelial cells populating nascent early tumours, with **Tumour 12** markers staining only a subset of tumour cells (**Fig. 3g; Extended Data Fig. 8m,n**).

Additional expression changes in **Tumour 12** comprised the upregulation of genes associated with stromal communication, including cell adhesion (*Col12a1, Itgav, Itga2, Itgb6, Lama3, Vcl, Cadm1, Icam1, Runx1*[49]), as well as ECM break-down (*Adamst1*[50]). Increased expression of cell adhesion genes *Ccn1, Ccn2*[51] and *Thbs1*[52] was of particular interest due to their recognised role in communication with fibroblasts (**Extended Data Fig. 8l**).

Since the data so far suggested a close communication between **Tumour 12** epithelial and stromal cells, we reasoned that **Tumour 12** cells may be linked to the emergence of the tumour niche, and thereby, to tumour persistence/survival. Indeed, SOX9 expression, used as a proxy to mark the **Tumour 12** state (alongside KRT6A and/or KRT17[1,21]), was expressed primarily in early tumour cells in direct contact with the niche (**Extended Data Fig. 8o**), indicative of their close interaction. Accordingly, nascent Niche+ tumours showed significantly higher SOX9 expression than Niche- tumours (**Fig. 3h; Extended Data Fig. 9a-c**).

To explore whether SOX9+ cells were associated with the formation of the pre-neoplastic tumour niche, we took advantage of sporadic clusters of KRT6A+/SOX9+ cells in phenotypically normal (non-tumour) regions of DEN-treated tissue, potentially marking prospective tumour cells prior to lesion formation. Isolated SOX9+ cell clusters showed signs of fibroblast attraction, presenting fibroblasts in closer proximity and at a higher density than in the surrounding tissue (**Fig. 3i; Extended Data Fig. 9d-f**). Overall, this data establishes the **Tumour 12** state as a relevant player in early tumour stromal remodelling and niche formation. Accordingly, SOX9 depletion in *Krt14-Cre*^{ER}*; Sox9*$^{flox/flox}$ DEN treated mice (**Fig. 3j; Extended Data Fig. 9g**) led not only to a significant reduction in early tumour survival (1m post-DEN; **Fig.**

3j), but also to a reduction in nascent Niche+ tumour size comparable to Niche- tumours (**Extended Data Fig. 9h**).

7. Cell-cell communication in early tumours

To study epithelial-mesenchymal communication in surviving tumours, we assessed ligand-receptor interactions[53] enriched across co-existing cell populations in DEN-treated tissue (**Fig. 4a,b; Extended Data Fig. 10a**; Fibroblasts: **Fig 3c; Extended Data Fig. 6e;** Epithelia: **Fig. 3f; Extended Data Fig. 8g-k**.). Our analysis included interactions between non-tumour cells, i.e., control keratinocytes (**Non-tumour Krt**) and control fibroblasts (*Pdgfra*^{low}; C19 **DEN Fbr** as internal control), as well as between tumour associated cells, i.e., tumour keratinocytes (**Tumour 1**), tumour niche keratinocytes (**Tumour 12**), and tumour niche fibroblasts (*Pdgfra*^{low}; C19 **Tumour Fbr**). Pro-fibrotic ECM-related pathways were among the top outgoing interactions predicted to preferentially signal from **tumour niche fibroblasts** to **tumour niche keratinocytes**. These pathways included LAMININ, COLLAGEN, FN1, THBS and TENASCIN (**Fig. 4c**). Given its well-established role in ECM assembly and association with fibrosis and advanced cancer progression[54], we reasoned that fibronectin (FN1) may represent a central player modulating ECM interactions across tumour cell compartments. The relevance of FN1 interactions was supported both by the specific upregulation of FN1 receptor genes in **tumour niche keratinocytes** (**receiver cells**), and by the increased expression of FN1, at both mRNA and protein level, in **tumour niche fibroblasts** (**sender cells**; **Fig. 3d,e, Extended Data Fig. 6f,g; 10b**). Overall, our analyses predicted robust pro-fibrotic/wound healing epithelial-mesenchymal interactions in surviving tumours.

Next, we explored outgoing signals from **tumour niche keratinocytes**. Here, EGF was identified as one of the strongest incoming signals for **tumour niche fibroblasts** (**Fig. 4c**). EGF ligands (including *Areg* and *Hbegf*) were enriched in **tumour niche keratinocytes** (**sender cells**) at both the mRNA and protein levels (**Fig. 3g; Extended Data Fig. 8l,m; 10b**), with its receptor (*Egfr*) being markedly expressed in the underlying **tumour niche fibroblast** population (**receiver cells**; **Extended Data Fig. 10b**).

The upregulation of both AREG (Amphiregulin) and FN1 in nascent tumours (10d post-DEN) (**Fig. 3e,g; Extended Data Fig. 6g; 8m**) highlighted the significance of the epithelial-mesenchymal communication from nascent tumour stages.

8. EGF-SOX9-FN1 supports tumour survival

Given the well-known role of EGF and FN1 signalling in epithelial-stromal communication during wound healing and tissue damage[55,56], we hypothesised that these factors may be playing a similar role in response to DEN-induced genetic stress.

To determine whether tumour niche keratinocytes (**Tumour 12**) exert their mesenchymal remodelling effect (**Fig. 3h,i; Extended Data Fig. 8k,o; 9a-f**) via the EGFR pathway (**Fig. 4c**), we performed a chemoattractant assay (**Fig. 4d**). We found that the EGFR ligand AREG positively stimulated fibroblast migration, confirming that keratinocytes in nascent tumours may promote fibroblast chemotaxis and mesenchymal remodelling through paracrine EGF stimulation **(Fig. 4d,e; Extended Data Fig. 10c)**.

Further insights into the dynamic nature of epithelial-mesenchymal communication were gained by co-culturing regenerative 3D oesophageal cultures (Epithelioids[57]) with primary fibroblasts (**Fig. 4f**). We found that expanding keratinocytes, which exhibit increased levels of stress markers (including SOX9; **Extended Data Fig. 10d**), prompted fibroblasts to segregate spatially into two distinct populations, mirroring the *in vivo* scenario. PDGFRα^{low} fibroblasts localised immediately adjacent to the growing epithelium, while PDGFRα^{high} fibroblasts were found in distant areas (**Fig. 4g; Extended Data Fig. 10e**). The interaction between expanding epithelial cells and fibroblasts also led to FN1 deposition and Vimentin upregulation in the PDGFRα^{low} population (**Fig. 4g**; **Extended Data Fig. 10e**). Gefitinib-mediated inhibition of EGFR signalling in epithelioids further confirmed the role of EGFR signalling in epithelial-mesenchymal communication in a regenerative/stress context, showing i) reduced epithelial SOX9 expression, ii) diminished fibroblast segregation/compartmentalisation, and iii) hindered fibroblast ECM remodelling (**Fig. 4f-h; Extended Data Fig. 10d,e**). These data directly link EGFR signalling and SOX9 expression in the epithelium with mesenchymal remodelling.

FN1, a major component of the fibrotic niche in early tumours (**Fig. 3d,e**; **Extended Data Fig. 6b-g**), also represents a critical and well-established regulator of the stromal wound healing response[58]. To determine whether the newly formed fibronectin-rich tumour niche plays a critical role in promoting early tumour growth and survival, we treated established 3D Epithelioids (exhibiting steady-state levels of proliferation)[57] with soluble FN1 for 24 hours. Indeed, FN1 promoted epithelial proliferation (**Extended Data Fig. 10f**). In contrast, *in vivo* Bleomycin treatment, which promotes tissue fibrosis, resulted in a negligible increase in SOX9 expression in DEN-untreated mice (**Extended Data Fig. 10g**). These data indicate that, although fibrosis promotes tumour growth, a fibrotic environment alone is not sufficient to drive a pre-neoplastic response.

These *ex vivo* experiments revealed that the EGF-SOX9-FN1 axis governs epithelial-mesenchymal communication and subsequent tissue reorganisation in response to epithelial perturbations. Accordingly, *in vivo* experiments showed that inhibition of either fibronectin fibrillogenesis (FUD peptide)[59] or EGFR signalling with Gefitinib (GFT) led to a significant reduction in the number of Niche+ tumours (**Fig. 4i,j**).

Altogether, our results demonstrate the central role of the EGF-SOX9-FN1 axis in early tumour niche formation. In response to genetic stress, SOX9+ epithelial cells stimulate fibroblast migration and ECM remodelling through EGF signalling. This, in turn, promotes the formation of a pro-fibrotic, fibronectin-rich tumour scaffold, which favours tumour persistence and progression by perpetuating the pro-tumorigenic phenotype.

9. Human tumour niche remodelling

To validate the relevance of our observations in a human context, we analysed chemo-naïve early-stage human oesophageal squamous cell carcinomas (SCCs; **T1a, T1b, Tis**; **Fig. 6a-d**) and residual dysplastic tissues post-chemotherapy (**ypT0; Fig. 5**).

Immunolabelling showed that patient tumours, unlike adjacent tissue, recapitulate observations in the DEN mouse model. T1a and T1b tumours displayed homogeneous expression of KRT6A and KRT17 (a pan-tumour marker in mice) and heterogeneous expression of SOX9 (marking tumour cells associated with stromal remodelling in mice; **Tumour 12**; **Fig.**

5b,c). Accordingly, SOX9 expressing cells exhibited high AREG expression levels (**Fig. 5b**) and were found next to areas with increased fibroblast density (PDGFRα; **Fig. 5c**). Analysis of tissue whole-mounts reinforced the link between SOX9 expression and stromal remodelling, with marked FN1 deposition in the vicinity of SOX9+ tumour cells (**Fig. 5d**).

These features were also observed in residual dysplastic tissue post-chemotherapy (**Fig. 5e,f**), which showed marked ECM remodelling (fibronectin deposition) in the proximity of AREG+/SOX9+ tumour cells.

Our observations support the presence of a heterogeneous AREG+/SOX9+ population in early-stage squamous tumours of the human oesophagus. The data further reinforces the association between this population, mesenchymal changes and ECM remodelling in early human oesophageal tumorigenesis, unveiling the potential clinical relevance of our study.

Discussion

Studies over the past decade have shown that mutations, traditionally thought to be the sole cause of cancer, can also be found in healthy ageing tissues[4-8], where they form part of normal tissue physiology. This has redirected the interest of the cancer community to fill the knowledge gap around the earliest disease stages, and particularly to understand how mutant cells interact with adjacent tissue compartments[1-3,13,21,22]. This study provides mechanistic insights into the processes that determine whether tumours emerging within complex mutant landscapes persist long-term or are outcompeted and eliminated from the tissue[1]. We show that early tumour survival and subsequent progression rely on intricate interactions between nascent tumour cells and their dynamic niche.

This work reveals that exposure to mutagens activates a heterogeneous “tissue stress” response whereby incipient epithelial and mesenchymal tumour cells signal and feedback onto each other through the EGF-SOX9-FN1 axis. Tumours failing to activate this communication axis are less likely to persist and grow. In particular, a tumour-specific stress state, defined by high SOX9 expression, promotes the recruitment of fibroblasts to the nascent tumour through EGF signalling. This, in turn, facilitates the formation of a pre-cancerous niche, rich in fibronectin, that perpetuates a pro-tumorigenic phenotype, favouring tumour growth

and persistence. Interfering with fibronectin fibrillogenesis *in vivo* significantly impaired niche formation, prevented tumour survival, and reduced the overall tumour burden. These findings support a self-sustaining process where the reciprocal communication between niche mesenchymal cells and SOX9^{high} epithelial cells supports tumour survival, favouring disease progression over time **(Fig. 5g)**.

These data are relevant to early human carcinogenesis. The heterogeneous expression of SOX9, EGFR ligands, and associated deposition of FN1 are recapitulated in early stage human oesophageal SCCs, consistent with active epithelial-stromal communication in nascent human tumours. Whether interfering with ECM assembly represents a valid approach to prevent cancer progression in patients, and whether analogous mechanisms operate in other tumour types, requires further investigations.

Overall, our data demonstrate the unprecedented capacity of the early tumour niche to perpetuate tumour survival signals beyond intrinsic changes driven by genetic alterations, enabling nascent tumours to persist in highly competitive mutant landscapes. This offers the new perspective that not only mutations but also the environmental response to genetic stress defines the likelihood of tumours to progress towards more advanced disease stages. Our findings indicate that strategies targeting both tumour cells as well as supporting neighbouring cells may open new avenues for cancer prevention and improve long-term outcomes.

References

1 Colom, B. *et al.* Mutant clones in normal epithelium outcompete and eliminate emerging tumours. *Nature* **598**, 510-514 (2021). https://doi.org/10.1038/s41586-021-03965-7

2 Yum, M. K. *et al.* Tracing oncogene-driven remodelling of the intestinal stem cell niche. *Nature* **594**, 442-447 (2021). https://doi.org/10.1038/s41586-021-03605-0

3 Bansaccal, N. *et al.* The extracellular matrix dictates regional competence for tumour initiation. *Nature* **623**, 828-835 (2023). https://doi.org/10.1038/s41586-023-06740-y

4 Kakiuchi, N. & Ogawa, S. Clonal expansion in non-cancer tissues. *Nat Rev Cancer* **21**, 239-256 (2021). https://doi.org/10.1038/s41568-021-00335-3

5 Fowler, J. C. & Jones, P. H. Somatic Mutation: What Shapes the Mutational Landscape of Normal Epithelia? *Cancer Discov* **12**, 1642-1655 (2022). https://doi.org/10.1158/2159-8290.CD-22-0145

6 Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880-886 (2015). https://doi.org/10.1126/science.aaa6806

7 Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911-917 (2018). https://doi.org/10.1126/science.aau3879

8 Martincorena, I. Somatic mutation and clonal expansions in human tissues. *Genome Med* **11**, 35 (2019). https://doi.org/10.1186/s13073-019-0648-4

9 Gallini, S. *et al.* Injury prevents Ras mutant cell expansion in mosaic skin. *Nature* **619**, 167-175 (2023). https://doi.org/10.1038/s41586-023-06198-y

10 van Neerven, S. M. *et al.* Apc-mutant cells act as supercompetitors in intestinal tumour initiation. *Nature* **594**, 436-441 (2021). https://doi.org/10.1038/s41586-021-03558-4

11 Flanagan, D. J. *et al.* NOTUM from Apc-mutant cells biases clonal competition to initiate cancer. *Nature* **594**, 430-435 (2021). https://doi.org/10.1038/s41586-021-03525-z

12 Chen, Y. *et al.* Epithelial cells activate fibroblasts to promote esophageal cancer development. *Cancer Cell* **41**, 903-918 e908 (2023). https://doi.org/10.1016/j.ccell.2023.03.001

13 Fiore, V. F. *et al.* Mechanics of a multilayer epithelium instruct tumour architecture and function. *Nature* **585**, 433-439 (2020). https://doi.org/10.1038/s41586-020-2695-9

14 Tagore, M. *et al.* GABA Regulates Electrical Activity and Tumor Initiation in Melanoma. *Cancer Discov* **13**, 2270-2291 (2023). https://doi.org/10.1158/2159-8290.CD-23-0389

15 Parreno, V. *et al.* Transient loss of Polycomb components induces an epigenetic cancer fate. *Nature* **629**, 688-696 (2024). https://doi.org/10.1038/s41586-024-07328-w

16 Marusyk, A. *et al.* Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* **514**, 54-58 (2014). https://doi.org/10.1038/nature13556

17 Reeves, M. Q., Kandyba, E., Harris, S., Del Rosario, R. & Balmain, A. Multicolour lineage tracing reveals clonal dynamics of squamous carcinoma evolution from initiation to metastasis. *Nat Cell Biol* **20**, 699-709 (2018). https://doi.org/10.1038/s41556-018-0109-0

18 Polyak, K., Haviv, I. & Campbell, I. G. Co-evolution of tumor cells and their microenvironment. *Trends Genet* **25**, 30-38 (2009). https://doi.org/10.1016/j.tig.2008.10.012

19 Karras, P., Black, J. R. M., McGranahan, N. & Marine, J. C. Decoding the interplay between genetic and non-genetic drivers of metastasis. *Nature* **629**, 543-554 (2024). https://doi.org/10.1038/s41586-024-07302-6

20 Janiszewska, M. *et al.* Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments. *Nat Cell Biol* **21**, 879-888 (2019). https://doi.org/10.1038/s41556-019-0346-x

21 Alcolea, M. P. *et al.* Differentiation imbalance in single oesophageal progenitor cells causes clonal immortalization and field change. *Nat Cell Biol* **16**, 615-622 (2014). https://doi.org/10.1038/ncb2963

22 Colom, B. *et al.* Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. *Nat Genet* **52**, 604-614 (2020). https://doi.org/10.1038/s41588-020-0624-3

23 Chang, J. *et al.* Genomic alterations driving precancerous to cancerous lesions in esophageal cancer development. *Cancer Cell* **41**, 2038-2050 e2035 (2023). https://doi.org/10.1016/j.ccell.2023.11.003

24 Bejar, M. T. *et al.* Defining the transcriptional signature of esophageal-to-skin lineage conversion. *bioRxiv*, 2021.2002.2019.431899 (2021). https://doi.org/10.1101/2021.02.19.431899

25 Tetreault, M. P. & Kahrilas, P. GI Manifestations With a Focus on the Esophagus: Recent Progress in Understanding Pathogenesis. *Curr Rheumatol Rep* **21**, 42 (2019). https://doi.org/10.1007/s11926-019-0841-x

26 Huycke, T. R. *et al.* Patterning and folding of intestinal villi by active mesenchymal dewetting. *Cell* **187**, 3072-3089 e3020 (2024). https://doi.org/10.1016/j.cell.2024.04.039

27 Manieri, E. *et al.* Role of PDGFRA(+) cells and a CD55(+) PDGFRA(Lo) fraction in the gastric mesenchymal niche. *Nat Commun* **14**, 7978 (2023). https://doi.org/10.1038/s41467-023-43619-y

28 Sahai, E. *et al.* A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer* **20**, 174-186 (2020). https://doi.org/10.1038/s41568-019-0238-1

29 de Visser, K. E. & Joyce, J. A. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer Cell* **41**, 374-403 (2023). https://doi.org/10.1016/j.ccell.2023.02.016

30 Calvo, F. *et al.* Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nat Cell Biol* **15**, 637-646 (2013). https://doi.org/10.1038/ncb2756

31 Gonzalez, H., Hagerling, C. & Werb, Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev* **32**, 1267-1284 (2018). https://doi.org/10.1101/gad.314617.118

32 Kumar, R., Mao, Y., Patial, S. & Saini, Y. Induction of whole-body gene deletion via R26-regulated tamoxifen-inducible Cre recombinase activity. *Front Pharmacol* **13**, 1018798 (2022). https://doi.org/10.3389/fphar.2022.1018798

33 Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *J Cell Sci* **123**, 4195-4200 (2010). https://doi.org/10.1242/jcs.023820

34 Guerrero-Juarez, C. F. *et al.* Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. *Nat Commun* **10**, 650 (2019). https://doi.org/10.1038/s41467-018-08247-x

35 He, W., Xu, H., Zhang, Q. & Zheng, Y. Dynamic expression of alpha6 integrin indicates epidermal cell behaviors. *Biochem Biophys Res Commun* **515**, 119-124 (2019). https://doi.org/10.1016/j.bbrc.2019.04.185

36 Dvorak, H. F. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N Engl J Med* **315**, 1650-1659 (1986). https://doi.org/10.1056/NEJM198612253152606

37 Nurmik, M., Ullmann, P., Rodriguez, F., Haan, S. & Letellier, E. In search of definitions: Cancer-associated fibroblasts and their markers. *Int J Cancer* **146**, 895-905 (2020). https://doi.org/10.1002/ijc.32193

38 Dinh, H. Q. *et al.* Integrated single-cell transcriptome analysis reveals heterogeneity of esophageal squamous cell carcinoma microenvironment. *Nat Commun* **12**, 7335 (2021). https://doi.org/10.1038/s41467-021-27599-5

39 Davidson, S. *et al.* Single-Cell RNA Sequencing Reveals a Dynamic Stromal Niche That Supports Tumor Growth. *Cell Rep* **31**, 107628 (2020). https://doi.org/10.1016/j.celrep.2020.107628

40 Buechler, M. B. *et al.* Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575-579 (2021). https://doi.org/10.1038/s41586-021-03549-5

41 Yao, J. *et al.* Single-cell transcriptomic analysis in a mouse model deciphers cell transition states in the multistep development of esophageal cancer. *Nat Commun* **11**, 3715 (2020). https://doi.org/10.1038/s41467-020-17492-y

42 Tandon, M., Vemula, S. V. & Mittal, S. K. Emerging strategies for EphA2 receptor targeting for cancer therapeutics. *Expert Opin Ther Targets* **15**, 31-51 (2011). https://doi.org/10.1517/14728222.2011.538682

43 Garg, M. *et al.* Laminin-5gamma-2 (LAMC2) is highly expressed in anaplastic thyroid carcinoma and is associated with tumor progression, migration, and invasion by modulating signaling of EGFR. *J Clin Endocrinol Metab* **99**, E62-72 (2014). https://doi.org/10.1210/jc.2013-2994

44 Ge, Y. *et al.* Stem Cell Lineage Infidelity Drives Wound Repair and Cancer. *Cell* **169**, 636-650 e614 (2017). https://doi.org/10.1016/j.cell.2017.03.042

45 Larsimont, J. C. *et al.* Sox9 Controls Self-Renewal of Oncogene Targeted Cells and Links Tumor Initiation and Invasion. *Cell Stem Cell* **17**, 60-73 (2015). https://doi.org/10.1016/j.stem.2015.05.008

46 Yang, Y. *et al.* The pioneer factor SOX9 competes for epigenetic factors to switch stem cell fates. *Nat Cell Biol* **25**, 1185-1195 (2023). https://doi.org/10.1038/s41556-023-01184-y

47 Zhang, X. *et al.* Dissecting esophageal squamous-cell carcinoma ecosystem by single-cell transcriptomic analysis. *Nat Commun* **12**, 5291 (2021). https://doi.org/10.1038/s41467-021-25539-x

48 Vercauteren Drubbel, A. & Beck, B. Dedifferentiation of esophageal progenitors in metaplasia and cancer. *Mol Cell Oncol* **8**, 1991758 (2021). https://doi.org/10.1080/23723556.2021.1991758

49 Osorio, K. M., Lilja, K. C. & Tumbar, T. Runx1 modulates adult hair follicle stem cell emergence and maintenance from distinct embryonic skin compartments. *J Cell Biol* **193**, 235-250 (2011). https://doi.org/10.1083/jcb.201006068

50 Canals, F., Colome, N., Ferrer, C., Plaza-Calonge Mdel, C. & Rodriguez-Manzaneque, J. C. Identification of substrates of the extracellular protease ADAMTS1 by DIGE proteomic analysis. *Proteomics* **6 Suppl 1**, S28-35 (2006). https://doi.org/10.1002/pmic.200500446

51 Chen, C. C., Chen, N. & Lau, L. F. The angiogenic factors Cyr61 and connective tissue growth factor induce adhesive signaling in primary human skin fibroblasts. *J Biol Chem* **276**, 10443-10452 (2001). https://doi.org/10.1074/jbc.M008087200

52 Sakai, K. *et al.* Thrombospondin-1 promotes fibroblast-mediated collagen gel contraction caused by activation of latent transforming growth factor beta-1. *J Dermatol Sci* **31**, 99-109 (2003). https://doi.org/10.1016/s0923-1811(02)00150-0

53 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021). https://doi.org/10.1038/s41467-021-21246-9

54 Singh, P., Carraher, C. & Schwarzbauer, J. E. Assembly of fibronectin extracellular matrix. *Annu Rev Cell Dev Biol* **26**, 397-419 (2010). https://doi.org/10.1146/annurev-cellbio-100109-104020

55 Mimura, Y. *et al.* Epidermal growth factor induces fibronectin expression in human dermal fibroblasts via protein kinase C delta signaling pathway. *J Invest Dermatol* **122**, 1390-1398 (2004). https://doi.org/10.1111/j.0022-202X.2004.22618.x

56 Shiraha, H., Gupta, K., Drabik, K. & Wells, A. Aging fibroblasts present reduced epidermal growth factor (EGF) responsiveness due to preferential loss of EGF receptors. *J Biol Chem* **275**, 19343-19351 (2000). https://doi.org/10.1074/jbc.M000008200

57 Herms, A. *et al.* Self-sustaining long-term 3D epithelioid cultures reveal drivers of clonal expansion in esophageal epithelium. *Nat Genet* (2024). https://doi.org/10.1038/s41588-024-01875-8

58 Arwert, E. N., Hoste, E. & Watt, F. M. Epithelial stem cells, wound healing and cancer. *Nat Rev Cancer* **12**, 170-180 (2012). https://doi.org/10.1038/nrc3217

59 Tomasini-Johansson, B. R. *et al.* A 49-residue peptide from adhesin F1 of Streptococcus pyogenes inhibits fibronectin matrix assembly. *J Biol Chem* **276**, 23430-23439 (2001). https://doi.org/10.1074/jbc.M103467200

Fig. 1: Pre-cancerous niche remodelling is linked to long-term tumour persistence. a, Experimental DEN (diethyl-nitrosamine) carcinogen protocol. Wild-type mice were exposed to DEN in the drinking water for 2 months. Tissues were collected at 10d, 2m, 8m and 1y. **b, c** Representative confocal images of long-term persisting tumours 8m post-DEN (**b**) and nascent tumours 10d post-DEN (**c**), stained for DAPI (blue), KRT6A (tumour marker, red) and PDGFRα (fibroblast marker, greyscale). Scale bars, 50 µm (**b**), 10 µm (**c**). Image settings adjusted to upper stromal layer. **d**, Percentage of Niche+ and Niche- tumours at the indicated time points post-DEN from three mice per time point; statistical significance was determined by a one-sided Chi-squared test. **e**, Confocal images showing EdU incorporation (green) in KRT6A+ nascent tumours (red, dashed line) 10d post-DEN. Scale bars, 10 µm. Images were generated omitting upper most suprabasal layer. **f,** Diameter (x100 µm) of Niche+ (red) and Niche- (blue) tumours at the indicated time points. Tumours were quantified in three mice per time point; at 10d n=128 tumours; at 2m n=84 tumours; at 8m n=53 tumours and at 1y n=49 tumours. Data expressed as mean±s.e.m. Two-tailed Welch's t-test comparing Niche- *versus* Niche+ tumours. **g,** Cartoon illustrating the association between tumour niche remodelling and long-term tumour survival. Red, surviving tumours; blue, disappearing tumours. Days (d), months (m), year (y). Panels (**a**,**g**) created in BioRender https://BioRender.com/0g5wodl. Source data (**d,f**).

Fig. 2: The early tumour niche is formed by PDGFRα^{low} fibroblast population. a, Representative confocal image of a nascent Niche+ tumour 10d post-DEN withdrawal, side-view shows the niche is comprised of lamina propria (Lp) not submucosae (Sb) fibroblasts. Muscularis mucosae (Mm), Epithelium (Ep). Dashed lines label the layers. White arrow-head points at the nascent tumour niche arising from lamina propria. DAPI (blue); KRT6A (red); PDGFRα (greyscale). Scale bar, 25 µm. **b**, Number of stromal cells in tumour free DEN tissue, Niche- and Niche+ tumours per surface area, 10d post-DEN. n=27 (Niche-), 22 (Niche+), 15 (DEN) areas, from 3 mice, dots represent each area. PDGFRα+ fibroblasts, CD45+ immune cells, CD31+ endothelial cells. Data expressed as mean±s.e.m. Statistical significance assessed by one way Welch's ANOVA with multiple comparisons. **c**, 3D-rendered confocal side-views of Niche- and Niche+ tumours 10d post-DEN. CD45 (green), CD31 (orange; absent), KRT6A (red), PDGFRα (greyscale). Dashed lines label the basal membrane. Scale bar, 10 µm. Image settings adjusted to upper stromal layer. **d**, Experimental protocol for fibroblast lineage tracing: *Col1a2^{CreER}; R26$^{FlConfetti/wt}$* mice received a dose of Tamoxifen (TAM) followed by DEN treatment. Samples were collected 6m post-DEN. **e**, Representative top-down (top) and side-views (bottom) of control or tumour tissue from (**d**). PDGFRα greyscale; lineage traced confetti+ cells (yellow and red). Scale bars, 50 µm. Dashed lines separate stromal compartments. Control n=3; DEN n=4 animals. **f**, Single channel from **e** shows marked difference in PDGFRα expression across the two stromal compartments both in control and tumour samples; lamina propria, PDGFRα^{low} (underlying epithelium); submucosae, PDGFRα^{high} (deeper stromal layer). Days (d), months (m). Panel (**d**) created in BioRender https://BioRender.com/hwbs32m. Source data (**b**).

Fig. 3: Nascent tumour heterogeneity in epithelial compartment is linked to stromal remodelling. **a**, Microdissection of squamous upper-GI 8m post-DEN for single-cell RNA sequencing. **b**, UMAP showing cell type annotation. Krt, keratinocytes. **c**, Heterogenous expression of *Pdgfra* in fibroblasts in the UMAP space. Inset showing fibroblast clusters. **d**, Violin plots showing levels of *Pdgfra* and *Fn1* expression in fibroblast clusters. Black line, mean. **e,** Representative images from 6 animals of DEN area and nascent tumour stroma 10d post-DEN showing the accumulation (white arrowhead) of Fibronectin (FN1, green) in the niche. DAPI (blue). Scale bars, 10 μm. **f**, Heatmap (left) of the top 1500 differentially expressed genes along the basal keratinocyte pseudotime trajectory. Pseudotime trajectories, top right (committed, blue; basal, yellow; tumour, magenta). Populations representing two tumour transcriptional modules (Tumour 1, red and Tumour 12, cyan; bottom right UMAP). **g,** Representative images from 6 mice, showing Tumour 12 markers in nascent tumours 10d post-DEN (homogenous KRT6A (red) and KRT17 (yellow); heterogenous SOX9, EGR1 (cyan) and AREG, RUNX1 (magenta); DAPI (blue). Scale bars 10 μm. **h,** Representative images showing SOX9 distribution in Niche- and Niche+ tumours (10d post-DEN). White arrowheads highlight SOX9+ keratinocytes (cyan). KRT6A, (red); PDGFRα, (greyscale). Scale bars, 10 μm. **i,** Images of SOX9+ clusters in DEN-treated tumour free areas 10d post-DEN. Control was adjacent area negative for SOX9. DAPI, (blue); SOX9, (cyan); PDGFRα, (greyscale). Scale bars, 10 μm. White arrowheads highlight keratinocyte to fibroblast proximity, white dashed line labels epithelia to stroma border. **j,** Left, experimental protocol: *Krt14*^{CreER}*; Sox9*$^{flox/flox}$ mice received Tamoxifen (TAM) followed by DEN. Tissues collected 1m post-DEN and compared to DEN-treated un-induced controls. Right, quantification of tumour burden. n=3 mice per condition; data presented as mean±s.e.m. One-tailed Mann-Whitney test. Images captured by confocal microscopy. Days (d), months (m). Panels (**a,j**) created in BioRender https://BioRender.com/xjxwb1m. Source data (**j**).

Fig. 4: EGF-SOX9-FN1 axis promotes tumour survival. a, Cartoon showing tumour cell composition. **b**, Cell-cell communication network. Circles represent cells; circle sizes are cell numbers; thickness of the connections represents number of significant interactions. **c**, Heatmap showing top 10 signalling predictions. Interactions in Tumour (Tmr) 12 keratinocytes and Tmr fibroblast highlighted in yellow. Signals with the strongest interactions in Tmr fibroblasts highlighted in pink; in Tmr 12 keratinocytes highlighted in cyan. **d**, Top: schematics of chemoattractant assay. Bottom: representative images. Scale bars, 25 μm; inset, 10 μm. FBS, foetal bovine serum. **e**, Percentage of fibroblasts crossing the membrane. Data expressed as mean±s.e.m. Dots represent replicate cultures from n=4 mice. Significance assessed by one-way Welch's ANOVA with multiple comparisons (other groups in EDF 10c). **f**, Schematic representation of epithelioid and fibroblasts co-culture treated with gefitinib (GFT) or vehicle (DMSO). **g**, Representative images from (**f**) showing fibroblast interaction with keratinocytes in DMSO and EGFR inhibition (GFT) conditions. Top insets show PDGFRα+ fibroblast (red) heterogeneity: PDGFRα^{low} fibroblasts assemble adjacent to keratinocytes, while PDGFRα^{high} fibroblasts position further away in controls. The fibroblast activation at the border labelled by VIM (white) and FN1 (green). This was inhibited in GFT (bottom). CDH1 (green); nuclear Tomato, nT; keratinocytes, krt, (red), DAPI (blue). Scale bar, 500 μm; inset 50 μm. **h,** Schematic representation of keratinocytes-fibroblasts interaction under DMSO/GFT conditions. **i**, Experimental protocol of drug intervention with a fibronectin-assembly inhibiting peptide (FUD) or GFT 20-day regime in DEN. **j**, Tumour burden decreased in GFT and FUD groups compared to vehicle (VHC) or scrambled (SCR) control, respectively. Tissues were collected 10 days post-DEN. Data from n=3 (VHC), 4 (GFT), 5 (SCR), 5 (FUD) mice. Data represents mean±s.e.m. Significance was assessed by one-tailed Mann-Whitney test. Images captured by confocal microscopy. Panels (**a,d,f,h,i**) created in BioRender https://BioRender.com/eghet5p. Source data (**e,j**).

Fig. 5: Early tumour niche remodelling in human oesophageal squamous cell carcinoma. a, Schematic representation of tumour sample origin in (**b-d**); tumour resections from human patients performed prior to chemotherapy (chemo-naïve). **b**, Representative confocal image (from n=4) of a T1B tumour section showing widespread KRT6A (red), and heterogenous SOX9 (cyan) and AREG (magenta) expression. DAPI (blue). CTRL sample is normal area identified by a pathologist in the tumour section. Scale bar, 200 µm; inset, 50 µm. **c**, Representative confocal image (from n=5) of a T1A tumour section showing widespread KRT17 (yellow), heterogenous SOX9 (cyan) expression and fibroblast attraction to areas marked by PDGFRα (greyscale) expression. DAPI, blue. Scale bars, 200 µm; inset, 10 µm. **d**, Representative confocal image (from n=1) of a Carcinoma *in situ* (Tis) whole mount showing heterogenous SOX9 (cyan) expression and fibronectin (FN1, green) accumulation underneath (white arrowhead). DAPI, blue. Scale bars, 50 µm. **e**, Schematic representation of tumour sample origin in (**f**), resections were performed after chemotherapy. **f**, Representative confocal image (from n=1) of a post-treatment, pathological staging T0 (dysplasia), whole-mount, showing heterogenous expression of SOX9 (cyan) and AREG (magenta) and fibronectin (FN1, green) accumulation underneath. DAPI (blue). Control sample is a biopsy distal to the tumour. Scale bar, 500 µm; inset, 50 µm. **g**, Schematic of proposed model whereby epithelial cells respond to genetic perturbation by activating a stress gene signature (Tumour 12), denoted by SOX9 and EFG ligand overexpression. This drives the migration of underlying fibroblasts towards the nascent tumour, followed by the formation of a fibrotic scaffold. The establishment of this EGF-SOX9-FN1 signalling axis between epithelial and mesenchymal of nascent tumour cells results in the formation of an early tumour niche that favours tumour growth and promotes long-term persistence. Panels (**a**, **e**, **g**) created in BioRender https://BioRender.com/e94fdb9.

Methods

Clinical samples

High grade dysplasia, squamous cell carcinoma and macroscopically normal clinical samples, as well as the corresponding clinical information, were collected following research ethics approval and individual informed consent from patients who underwent oesophagectomy for oesophageal cancer.

The T1A and T1B stage chemo-naive surgical tumour samples were donated by patients who had undergone surgery at the Clinic for Visceral, Thoracic and Vascular Surgery at TU Dresden or at the Medical Department I, of the Carl Gustav Carus University Hospital. Macroscopically normal samples adjacent to the proximal resection margin were sampled from cancer resection specimens. The corresponding FFPE material (tumour and normal tissue) from a total of ten characterised oesophageal squamous cell carcinomas was selected from the archive of the Institute of Pathology of the University Hospital Carl Gustav Carus (EK 59032007) by the Tumour and Normal Tissue Bank (TNTB) Dresden. Studies presented in the manuscript involving early chemo-naïve human oesophageal tumour samples from Dresden were approved by the Ethics Committee of TU Dresden, Germany (Ref.: SR+BO-ff (Mono)-EK-161042025). Studies presenting chemo-naïve or post-chemo human oesophageal tumour samples from Guy's and St Thomas' (London, UK) and Addenbrooke's Hospital (Cambridge, UK), respectively were approved by East of Scotland Research Ethics approval committee (REC 18/ES/133). Histological sections of the tissue samples and haematoxylin-eosin staining of reference slide series for determining tumour cell content of the individual patient samples were carried out at the Institute of Pathology, University Hospital CGC Dresden, TU Dresden.

Mice strains

All animal experiments were approved by the local ethical review committees of the University of Cambridge and conducted according to Home Office project licenses PPL70/8866 and PP7037913 at the Cambridge Stem Cell Institute, University of Cambridge.

Unless otherwise specified, C57BL/6J mice (ordered from Charles River, UK; strain code 632) were used. Other mouse strains includes: cell cycle reporter line *R26*^{Fucci2aR} (*Fucci2a*)[60] kindly provided by Ian J. Jackson; *Pdgfra*^{EGFP} mice (stock #007669, Jackson Laboratory); *Sox9*$^{flox/flox}$ mice[61] (obtained from MRC-Harwell, on behalf of the European Mouse Mutant Archive (https://www.infrafrontier.eu/)); *K14*^{CreER} mice (stock #005107, Jackson laboratory); *R26*$^{mT\text{-}mG}$ mice (*mTmG*; stock #007676, Jackson Laboratory); *Col1a2*^{CreER} mice (strain #029567, Jackson Laboratory); *R26*^{FIConfetti} mice[62] (strain #017492, Jackson laboratory, kindly provided by Hans Clevers)[62]; *R26*$^{nT\text{-}nG}$ mice (*nTnG*, strain #023537, Jackson Laboratory); *H2B-EGFP* mice (*CAG::H2B-EGFP*; strain #006069, Jackson Laboratory); NOD scid gamma

mice (NSG; NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ; strain #005557, Jackson Laboratory), *Pdgfra*^{CreER} mice (stock#018280, Jackson Laboratory[63]). Additional information about experimental mouse lines can be found in Supplementary Methods section '**Experimental mouse lines**'.

Recombination of *Col1a2*^{CreER}*R26*$^{FIConfetti/WT}$ mice was induced by a single intraperitoneal (i.p.) tamoxifen injection (3 mg per 20 g of body weight). *Col1a2*^{CreER}*R26*$^{FIConfetti/WT}$ mice were induced by a single i.p. tamoxifen injection (0.5mg or 5mg per 20 g of body weight). *K14*^{CreER}*Sox9*$^{flox/flox}$ received two subcutaneous tamoxifen injections (5 mg per 20 g of body weight) 48 hours apart. Tamoxifen was prepared by dissolving in ethanol (<10% total volume) and diluting in sunflower seed oil.

All strains were maintained in a C57BL/6 background. All experiments comprised a mixture of male and female mice with no gender-specific differences observed (unless specified otherwise). For RNA sequencing experiments, only male animals were used in order to avoid cofounding effects due to estrous cycle. All animals exposed to the carcinogen and their respective controls were adults between 8 and 14 weeks of age (see chemically induced mutagenesis below). Mice were bred and maintained under specific-pathogen-free conditions at the Gurdon Institute and The Anne McLaren Building, University of Cambridge. All animals were housed between 20-24°C, 45-65 % humidity and a 12 hour light-dark cycle.

Chemical tumorigenesis model

Mice were treated with Di-ethyl-nitrosamine (DEN, Sigma-Aldrich; N0756) at 40 mg per litre in Ribena-flavoured water for 24 hours, 3 days a week (Monday, Wednesday and Friday) for eight weeks[1,21]. Mice received sweetened water between DEN dosages, and normal water upon DEN treatment completion. Control mice received sweetened water as vehicle control for the length of the treatment. Animals exposed to DEN were monitored as per adverse effects stated in our Home Office project licenses (PPL70/8866 and PP7037913) regulated procedures on protected animals. In summary, animals were weighed daily on weekdays for the first week, weekly for the next month and then monthly thereafter. Animals were also checked every day for any clinical signs or abnormal behaviour. Any concerning animals were weighed every other day or daily, if necessary, until the weight was stable again. If the weight loss approached 10 %, animals were weighed daily until stable and received wet mash or palatable diet. Animals showing 15 % weight loss measured for 2 consecutive days were sacrificed immediately.

EdU tracing

For EdU (5-ethynyl-2'-deoxyuridine) labelling experiments, animals received 100 µg of EdU in PBS (Life Technologies; A10044) intraperitoneally 2 hours prior to tissue collection. *In vitro* 3D cultures (see

above) received media supplemented with 10 μM of EdU and incubated for 2 hours at 37 °C and 5 % CO_2 before fixation. EdU incorporation in tissue whole-mount (see above) was detected using Click-iT EdU kit according to the manufacturer's instructions (Invitrogen, C10337). EdU^+ cells were quantified using confocal microscopy.

Inhibitor treatment *in vivo*

Mice were treated with Gefitinib for 20 days at 80 mg/kg three times a week (or vehicle control) to inhibit Epidermal Growth Factor Receptor (EGFR) pathway. Treatment started 10 days prior the end of DEN treatment and ended 10 days after. Gefitinib was prepared concentrated by dissolving in DMSO and further diluted in corn oil.

Pharmacological inhibition of Fibronectin (FN1) fibrillogenesis was achieved by treating mice with the Functional Upstream Domain (FUD) peptide[64] intraperitoneally for 20 days at a concentration of 12.5 mg/kg body weight. Control animals were treated with scrambled (SCR) control peptide. Treatment started 10 days prior the end of DEN treatment and ended 10 days after. Peptides were synthesised at >95% purity (WatsonBio; Peptide sequence below). Lyophilised peptides were reconstituted in PBS.

Peptide sequences:

FUD - Cys-GSKDQSPLAGESGETEYITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKLN

SCR - Cys-QGQTGPVNSKVKIDNYELESNPEKIEANDLQVEGTTTYESKFMGDLTGSGNPED

Whole-mount preparation

The upper gastrointestinal tract (oesophagus and forestomach) from control and DEN-treated mice was dissected at the indicated time points. Oesophagi were excised and cut open longitudinally. The muscle layer was then removed, and tissue was flattened under a dissecting microscope using fine forceps. Stomachs were cut open longitudinally and rinsed twice with PBS to remove any food remains. The glandular stomach was excised away and the forestomach kept and flattened for downstream analysis. For epithelial-only and stromal-only whole-mounts, tissues were incubated in 5 mM Ethylene-diamine-tetraacetic acid (EDTA) (Life Technologies; 15575020) in PBS for 3 hours at 37°C. Following incubation, the epithelium was gently peeled from the stroma using fine forceps. Subsequently, each of these layers was individually flattened.

Oesophageal and forestomach whole-mounts (either peeled or not peeled) were fixed in 4% paraformaldehyde (PFA) (Alfa Aesar; 043368) in PBS for 30 minutes at room temperature (RT).

Histology

For histology, tissues were fixed in 10 % formalin in PBS overnight at RT prior to storage at 4°C in absolute ethanol. Hematoxylin and Eosin (H&E) staining was performed in 7 µm paraffin-embedded sections by the Histology Core Service at Cambridge Stem Cell Institute and imaged using Zeiss AxioScan Z1 microscope. We thank B. Mahler-Araujo and the MRC Metabolic Diseases Unit (MC_UU_00014/5) for the histological analysis of murine tumour samples.

Ex vivo tissue recombination assay

Under a dissecting microscope in a laminar flow hood, oesophagi were dissected and epithelial-stromal layers isolated as described in section '**Whole-mount preparation**'. Thereafter, tissues were rinsed in 1 % P/S in PBS three times to remove residual EDTA and flattened. Combination of epithelium and stroma from different experimental conditions (DEN treated and/or control) were prepared by carefully placing the epithelial layer over the relevant stroma (referred to as "tissue recombination" composites). The remaining epithelium was trimmed to match the size of the stroma, and the resulting construct was cut in half. Flattened epithelium-stromal constructs were cultured in 6-well plate inserts (ThinCert Greiner Bio-One; 657641). Size-matched Polydimethylsiloxane (PDMS) stencil frames were placed around the tissue construct to prevent cell expansion (see "**Stencil production**"). The tissue was settled for 10 min prior to adding 2 ml of minimal medium (mFAD), containing one-part DMEM (Fisher Scientific; 41966029) and one-part DMEM/F12 (Fisher Scientific; 11320033) supplemented with 5 µg/ml insulin (Sigma-Aldrich; 15500), 5 % foetal calf serum (Fisher Scientific; 26140079), 1 % P/S and 5 µg/ml Apo-Transferrin (Sigma-Aldrich; T2036), as previously published[24,65]. 3D heterotypic cultures were maintained in standard humidified cell culture incubators at 37°C with 5 % CO_2 for up to 7 days. At the end point samples were fixed in 4 % PFA in PBS for 30 minutes at RT and stored for downstream confocal analysis.

Stencil production

Silicone elastomer (PDMS) was mixed with a curing agent (Avantor VWR; Sylgard 184 Elastomer Kit 634165S) at a 10:1 ratio and centrifuged at 300×g for 10 min to remove bubbles. The resulting mix was poured on a dish at around 70 mg/cm^2 and left on an even surface to polymerise overnight at 37°C. The following day the resulting polymer was cut into 2x5 mm rectangle-shaped frames, sterilised in 70 % ethanol overnight, and treated with 1 % pluronic acid (Sigma-Aldrich; P2443-250g) in PBS for 1 hour at 37°C. The frames were then left to air dry prior to use.

In vivo tissue recombination grafting

Tissue recombination composites (as described in section "***Ex vivo* tissue recombination assay**") of DEN-treated oesophageal stroma and untreated (control) oesophageal epithelium were prepared for *in vivo* grafting adapting the strategy described in "***Ex vivo* tissue recombination assay**". Prior to separating the epithelium from the stroma, all visible tumours were marked with a partial incision using a punch biopsy tool (1 mm diameter, Merk; WHAWB100040). After separating the tissue layers, all stromal compartments were assessed for peeling efficiency under a fluorescence dissecting microscope (Leica M165 FC), and any remaining epithelium, identified by the dense epithelial nuclei clusters were excised from the tissue using a 1 mm biopsy punch. For heterotypic tissue constructs, 2 mm biopsies (Selles Medical; instrument BP20F) of tumour or controls stroma were excised and a 2 mm healthy untreated epithelium biopsy placed above. Composites were cultured overnight as described above and grafted in the back skin of anesthetised shaved NSG female mice (2 constructs per incision, 2 incisions per animal). Longitudinal incisions for grafting were approximately 5 mm in length. The wounds were closed with GLUture glue (Fisher Scientific; NC0632797) and the mice were left to recover. 3 to 6 months later, mice were culled and the back skin fixed with 4% PFA in PBS for 30 minutes at RT and stored for downstream confocal analysis.

Primary mouse fibroblast isolation and migration assay

Oesophagi were dissected as described above and cut in half. Tissue was incubated in 0.5 mg/ml Dispase (Sigma-Aldrich; D4818) for 10 min at 37°C rotating. After incubation, the epithelium was peeled away, and the stroma was minced finely and incubated in Trypsin-EDTA (0.25%) (ThermoFisher; 25200056) for 15 min at 37°C rotating. The resulting suspension was mixed by pipetting and DMEM supplemented with 10 % FBS and 1 % P/S was added (1:1 vol:vol). The suspension was passed through a 70 µm filter (PluriSelect; 43-10070-40) and cells pelleted by centrifugation at 300×g for 5 min at 4°C. Pellets were resuspended in 0.5 % FBS, 1 % P/S in DMEM, and seeded on 8.0 µm pore transwell insert (24-well plates; ThinCert Greiner Bio-One; 662638). The lower compartment of the transwell contained 1 % P/S DMEM supplemented either with 0.5 % FBS, 2 % FBS, 10 % FBS or 2 % FBS with 1 µg/ml Amphiregulin (AREG) (R&D; 989-AR-100/CF). Primary fibroblasts were cultured for 48 hours prior to fixation in 4 % PFA in PBS for 10 min. Membranes were incubated with 1 µg/ml DAPI (Sigma-Aldrich; D9542) in PBS for 30 min at RT, cut and mounted in 1.52 Rapiclear mounting medium (SUNJin Lab; RC152001) keeping their original orientation, followed by confocal analysis. Further information on quantification can be found in Supplementary Methods section '**Analysis of fibroblast migration assay**'.

Keratinocyte cultures and fibronectin treatment *in vitro*

Oesophagi were cultured using the 3D epithelioid organ culture approach[65]. Briefly, tissues were dissected, cut into 3x5 mm rectangles, and placed on a transwell insert with the epithelium side up. Tissue was left to settle for about 5 min. Explants were expanded in complete medium (cFAD) containing mFAD supplemented with $1x10^{-10}$ M cholera toxin (Sigma-Aldrich; C8052), 10 ng/ml Epidermal Growth Factor (EGF, Fisher Scientific PeproTech^{R}; AF-100-15), 0.5 µg/ml hydrocortisone (Calbiochem; 386698). Tissue explants were removed by aspiration five days after culture set up and maintained in mFAD for 2 weeks to confluence. Soluble fibronectin (Fisher Scientific Corning™; 356008) was added to the medium for 24 hours at 100 µg/ml after diluting it in mFAD with 25mM Hepes (Fisher Scientific; 15630056). Samples were fixed with 4 % PFA in PBS for 30 minutes at RT, after a 2h EdU chase, and kept for downstream confocal analysis.

Keratinocyte and fibroblast interaction assay

Epithelioids were set up as described above and maintained in mFAD until keratinocyte migration started. The original tissue was then removed and explant left overnight prior to adding freshly isolated oesophageal fibroblasts (as described above). DMSO as vehicle control or 2 µM of Gefitinib prepared in DMSO were added together with the fibroblast suspension. 3 days after fibroblasts were introduced to culture, samples were fixed with 4 % PFA in PBS for 30 minutes at RT and kept for downstream confocal analysis.

Immunostaining

After fixation, epithelial-stromal composites or human tissue whole-mounts were incubated for 30 min in permeabilisation buffer (PB1; 0.5 % bovine serum albumin (VWR International, 126575-10), 0.25 % fish-skin gelatin (Sigma-Aldrich; G7765), 1 % Triton X-100 (Fisher Scientific Ltd, 10102913) in PBS), then blocked for 2 hours in PB1 containing 10 % donkey serum (DS) (Scientific Laboratory Supplies, D9663). Next, tissues were incubated with primary antibodies diluted in 10% DS in PB1 for 3 days at 4°C followed by four washes 30 minutes each with 0.2 % Tween-20 (Promega UK Ltd, H5151) in PBS. Thereafter, tissues were incubated overnight with secondary antibodies diluted 1:500 in 10 % DS in PB1 at RT. Unbound antibody was removed by four washes with 0.2 % Tween-20 in PBS throughout the next day. Antibody details are provided in **Supplementary Table 15**. To stain cell nuclei, tissues were incubated with 1 µg/ml DAPI in PBS at 4°C overnight. Afterwards, samples were rinsed three times in PBS and mounted in 1.52 RapiClear mounting media for imaging.

Immunolabeling of individual tissue layers (epithelium or stroma) or sections consisted of an incubation for 30 min in permeabilisation buffer (PB2; 0.5 % Bovine serum albumin, 0.25% Fish skin

gelatin, 0.5 % Triton X-100 in PBS). Tissues were then blocked for 2 hours in PB2 containing 10% DS. Next, samples were incubated overnight at RT with primary antibodies diluted in 10% DS in PB2 followed by three washes with 0.2 % Tween-20 in PBS 30 minutes each. Secondary antibodies were diluted 1:500 in 10 % DS in PB2 and incubated with tissues overnight at 4°C after which unbound antibody was removed by three washes with 0.2 % Tween-20 in PBS and staining continued as above. Thick cryosections of fixed tissues embedded in optimal cutting temperature compound (OCT; Thermo Scientific; 12678646), cut with a thickness of 50 µm onto glass slides, were immunolabelled using the exact same protocol. 7 µm paraffin-embedded sections were immunolabelled following the above protocol after antigen retrieval performed by heating tissue sections in either 1 mM EDTA buffer (pH 8.0) or 10 mM sodium citrate buffer (pH 6.0) for 10 min at 95 °C.

When staining with primary antibodies raised in the same host, one of these antibodies was acquired as pre-conjugated with a fluorophore or conjugated in-house following manufacturer's instructions (Invitrogen; A20186/A20187). The staining proceeds as described above with the unconjugated primary antibodies. After incubation with the corresponding secondary antibodies, the samples were blocked for 3 hours at RT with 10 % DS in PB with the IgG from the relevant host species (1:500). Afterwards, samples were incubated with conjugated antibodies diluted in PB containing 10 % DS and the relevant host IgG (1:500) overnight at RT. At this point staining proceeds as above. Immuno-stained samples were analysed by confocal imaging.

Confocal Imaging

Confocal images were acquired using either an inverted Leica SP5 microscope with standard laser configuration, or a Stellaris 8 FALCON FLIM microscope with a white light laser using LAS X 4.7.0.28176 or 3.5.5.19976 software. Typical confocal settings used included: bidirectional scanning, a 40× immersion objective, an optimal pinhole size (as defined by the software), a scan speed of 400-600 Hz with 2-3x line averaging, optimal Z-step size (as defined by the software) and a resolution of 512 × 512 or 1,024 × 1,024 pixels unless stated otherwise. Three-dimensional reconstructions from optical sections and their corresponding image renders were generated using Volocity 5.5.5 (PerkinElmer) and Volocity 7 (Quorum), Zen 3.2 and Arivis 3.5.1. Further information about specific image analysis such as '**Second-harmonic generation imaging**' can be found in **Supplementary Methods** sections.

Transcriptomics

Library preparation and scRNA-Seq

Sample preparation for libraries can be found in **Supplementary Methods** in **'Single-cell and RNA isolation for single-cell RNA-Sequencing (scRNA-Seq)**' section.

scRNA-seq libraries were generated using the 10x Genomics Chromium Next GEM Single Cell 3' Reagent Kit (v3) and sequenced at the Genomics Core Facility of Cancer Research UK (CRUK), Cambridge Institute. Libraries were generated in two different batches. The information about library batches can be found in **Supplementary Table 2**. Control samples were included in both batches to provide a reference to assess potential batch effects. The cells for each biological replicate were loaded into a 10x Chromium microfluidics chip channel to generate one library from each. In total, 17 libraries were sequenced on either an Illumina HiSeqx4000 or a NovaSeq6000 system using one SP, two S1, and two S2 flow cells. Note, given the punch biopsy approach used, DEN samples may contain sporadic tumour cells from tumours not visible under dissection scope.

scRNA-seq data pre-processing, dimensionality reduction, and visualisation

The raw scRNA-seq data were processed with CellRanger (version 7.0.1). Reads were aligned to the mouse reference genome (mm10 2020-A), empty droplets were filtered, and unique molecular identifiers (UMIs) were counted to generate gene expression matrices. Doublets were identified with Scrublet[66] (version 0.2.3) and removed together with low-quality cells based on per-sample quality control metrics (**Supplementary Table 2**); cells with >15 % mitochondrial reads or genes expressed in fewer than three cells were excluded, resulting in 91,347 high-quality cells. Count matrices were processed using a standard Seurat workflow[67] (version 5.0.3) up to dimensionality reduction. Data were integrated by tissue of origin (oesophagus or forestomach) using Harmony[68] (version 1.2.3), and the integrated embeddings were projected in two dimensions using Seurat's RunUMAP function. Further details on data clustering, annotation, as well as trajectory and communication analysis can be found in **Supplementary methods** and on the AlcoleaLab GitHub page.

Low input targeted DNA sequencing

Oesophagi from control and DEN-treated animals were dissected as described above. Tissues were flattened with the epithelial side up and visible tumour lesions were marked with a partial incision using a 1mm diameter punch tool under a dissecting microscope. Tissues were then incubated in 5mM EDTA for 3 hours at 37°C rotating. After incubation, the epithelium was removed, the stroma was flattened, and tissues were fixed as described above.

Immunofluorescent labelling against KRT14 was performed as described above. Only tumour stroma footprints negative for KRT14 (i.e. lacking epithelial cells) were considered for DNA sequencing in order to avoid the identification of genetic mutations present in epithelial cells. Tumour stroma matching the criteria was dissected under a fluorescence dissecting microscope (Leica M165 FC) using a 1 mm punch biopsy tool. Punch biopsies of equivalent size were collected from untreated healthy tissues as

control. The DNA from individual biopsies was extracted using the Arcturus®PicoPure® DNA Extraction Kit (Fisher Scientific Ltd, KIT0103) following the manufacturer's instructions. Briefly, Proteinase K was reconstituted in 155 µl and the sample was lysed in 20 µl at 65°C overnight. Thereafter, Proteinase K was inactivated by incubation at 95°C for 10 min.

Samples were then sheared, libraries prepped using the NEBNext®Ultra™ II Fragmentase System, and index tags applied (Sanger 168 tag set). Material was subjected to 12 PCR cycles (Initial denaturation; 95°C, 5 min; 12 cycles: 98°C, 30 sec; 65°C, 30 sec; 72°C, 2 min; final elongation: 72°C, 10 min) and quantified (Accuclear dsDNA Quantitation Solution, Biotium). 500 ng of pooled material was taken forward for hybridisation capture and enrichment (SureSelect Target enrichment system, Agilent technologies) using a previously designed bait panel of 192 genes (**Supplementary Table 9**), including those commonly mutated in squamous cancers[22]. After cleanup, libraries were normalised to ~6nM and submitted to cluster formation for sequencing on a Novaseq6000 (Illumina) to generate 100 bp paired-end reads.

Aligned reads were mapped to the mouse GRCm38 reference genome using BWA-mem (version 0.7.17)[69]. Duplicate reads were marked using SAMtools[70] (v1.11). Depth of coverage was also calculated using SAMtools to exclude reads which were: unmapped, not in the primary alignment, failing platform/vendor quality checks, or PCR/Optical duplicates. Bedtools (version 2.23.0)[71] was then used to calculate the depth of coverage per base across samples.

Variant calling was performed using the deepSNV R package (also commonly referred to as ShearwaterML; version 1.21.3; https://github.com/gerstung-lab/deepSNV). Variants were annotated using VAGrENT. R version 3.3.0 was used (from 2016-05-03)[7,72].

Mutations called by deepSNV ShearwaterML were filtered using the following criteria: 1) positions of called SNVs have a coverage of at least 100 reads, 2) germline variants called from the same individual are omitted from the list of called variants, 3) adjustment for FDR and mutations use support from at least one read from both strands for the mutations identified, 4) pairs of SNVs on adjacent nucleotides within the same sample are merged into a dinucleotide variant if at least 90% of the mapped DNA reads containing at least one of the SNV pairs also contained the other one. DeepSNV ShearwaterML was run with a normal panel of approximately 12k reads. The sequencing dataset was deposited at the European nucleotide Archive (ENA) under dataset accession number ERP134942.

Statistics and reproducibility

The numbers of biological replicates and animals are indicated in figure legends (n refers to the number of independent replicates per time point and/or condition). A minimum of three independent mice or *ex*

vivo cultures were used in all cases. All experiments were performed independently at least three times with similar results, unless otherwise stated. The reproducibility of all key findings was confirmed in independent experiments conducted on different days and using independent biological samples. For image analysis, a minimum of three independent samples were inspected in all cases. All figures show representative images. The data are expressed as mean values ± s.e.m. unless otherwise indicated.

All statistical tests were performed comparing biological replicates. Differences in tumour burden were assessed by one-tailed unpaired non-parametric Mann-Whitney U tests. Differences between Niche- and Niche+ tumour distribution in aging animals was assessed by one-sided Chi squared test. For large datasets normality was assessed by Kolmogorov Smirnov test; for normally distributed data, differences between two groups were assessed by two-tailed Welch's t-tests; for not normally distributed data two-tailed Mann Whitney U test was used. Differences between more than two groups were calculated using either one-way Welch's ANOVA, followed by a Dunnett's T3 multiple comparison test or Kruskal-Wallis one-way ANOVA, followed by Dunn's multiple comparison test, for normally distributed or not normally distributed data, respectively, unless specified otherwise. Exact P-values are indicated in the relevant figures with a precision of up to four decimal places. Statistical tests were conducted in GraphPad Prism (10.5.0) with 95% confidence interval. No statistical method was used to predetermine sample size. The experiments were performed without randomisation. Blinding was performed for tumour count per condition and *in vitro* sample analyses by confocal microscopy. In cases where quantification was performed in tumours and the morphologically normal areas blinding was not possible due to differences in physical sample appearance.

Data availability

Mouse reference genomes GRCm38 and mm10 2020-A were used. The single cell RNA sequencing data generated in this study have been deposited in the Gene Expression Omnibus (GEO) repository under accession code GSE271962. The DNA sequencing dataset was deposited at the European nucleotide Archive (ENA) under dataset accession number ERP134942. Source data are provided with this paper.

Code Availability

No new algorithms were developed for this paper. The analysis code is available on the AlcoleaLab GitHub page.

References

60 Mort, R. L. *et al.* Fucci2a: a bicistronic cell cycle reporter that allows Cre mediated tissue specific expression in mice. *Cell Cycle* **13**, 2681–2696 (2014). https://doi.org/10.4161/15384101.2015.945381

61 Kist, R., Schrewe, H., Balling, R. & Scherer, G. Conditional inactivation of Sox9: a mouse model for campomelic dysplasia. *Genesis* **32**, 121–123 (2002). https://doi.org/10.1002/gene.10050

62 Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010). https://doi.org/10.1016/j.cell.2010.09.016

63 Kang, S. H., Fukaya, M., Yang, J. K., Rothstein, J. D. & Bergles, D. E. NG2+ CNS glial progenitors remain committed to the oligodendrocyte lineage in postnatal life and following neurodegeneration. *Neuron* **68**, 668–681 (2010). https://doi.org/10.1016/j.neuron.2010.09.009

64 Sottile, J. *et al.* Fibronectin-dependent collagen I deposition modulates the cell response to fibronectin. *Am J Physiol Cell Physiol* **293**, C1934–1946 (2007). https://doi.org/10.1152/ajpcell.00130.2007

65 Herms, A. *et al.* Epithelioids: Self-sustaining 3D epithelial cultures to study long-term processes. *bioRxiv*, 2023.2001.2003.522589 (2023). https://doi.org/10.1101/2023.01.03.522589

66 Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291 e289 (2019). https://doi.org/10.1016/j.cels.2018.11.005

67 Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293–304 (2024). https://doi.org/10.1038/s41587-023-01767-y

68 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019). https://doi.org/10.1038/s41592-019-0619-0

69 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 (2013). <https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L>.

70 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021). https://doi.org/10.1093/gigascience/giab008

71 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). https://doi.org/10.1093/bioinformatics/btq033

72 Menzies, A. *et al.* VAGrENT: Variation Annotation Generator. *Curr Protoc Bioinformatics* **52**, 15 18 11–15 18 11 (2015). https://doi.org/10.1002/0471250953.bi1508s52

73 Grommisch, D. *et al.* Defining the contribution of Troy-positive progenitor cells to the mouse esophageal epithelium. *Dev Cell* **59**, 1269–1283 e1266 (2024). https://doi.org/10.1016/j.devcel.2024.03.011

74 Rezakhaniha, R. *et al.* Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy. *Biomech Model Mechanobiol* **11**, 461–473 (2012). https://doi.org/10.1007/s10237-011-0325-z

75 Oliver, A. J. *et al.* Single-cell integration reveals metaplasia in inflammatory gut diseases. *Nature* **635**, 699–707 (2024). https://doi.org/10.1038/s41586-024-07571-1

76 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019). https://doi.org/10.1038/s41586-019-0969-x

77 Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408–1414 (2020). https://doi.org/10.1038/s41587-020-0591-3

78 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018). https://doi.org/10.1038/s41586-018-0414-6

79 Yang, D., Liu, J., Qian, H. & Zhuang, Q. Cancer-associated fibroblasts: from basic science to anticancer therapy. *Exp Mol Med* **55**, 1322–1332 (2023). https://doi.org/10.1038/s12276-023-01013-0

Acknowledgements

We thank members of the Alcolea's lab for comments and suggestions; Dr. rer. nat. Irene Michalk from the Tumour and Normal Tissue Bank (TNTB) Dresden for coordinating human sample collection and documentation, Jessica Cordle, Clinical Research Nurse from Guy's and St Thomas' NHS Foundation Trust for coordinating human sample collection and delivery. We are also grateful to the Cancer Aging and Somatic Mutation (CASM) teams for accommodating work related to human samples; especially Elizabeth Anderson from the Adams group and to Laura-Anne Gilbey alongside the rest of the CASM research management team. Peter Humphreys and Darran Clements in imaging core facilities at Jeffrey Cheah Biomedical Centre (JCBC); Nicola Lawrence at the Gurdon Institute Imaging Facility; Irina Pshenichaya in Histology facility at JCBC for technical support; Betania Mahler-Araujo and James Warner from Histopathology at Institute of Metabolic Science-Metabolic Research Laboratories Medical Research Council Metabolic Diseases Unit for histopathology analysis; Doreen Streichert from the Institute of Anatomy at TU Dresden for technical support with human sample staining. Maike Paramor and Vicky Murray in the Next Generation Sequencing Core Facility at JCBC; Katarzyna Kania in the Single Cell Team at Cancer Research UK Cambridge Centre Cancer Institute; Irina Mohorianu's team at the bioinformatics facility at Cambridge Stem Cell for their contribution to the scRNA-seq data processing; the staff of the University Biomedical Services Gurdon Institute and The Anne McLaren Building, as well as the other JCBC core facilities. We are also grateful to Prof. Ian J. Jackson (Fucci2a) and Prof. Hans Clevers (Rosa26Confetti) for donation of mouse lines. Parts of the figures were Created with BioRender. This work was mainly supported by grants to M.P.A. from the Wellcome and The Royal Society (105942/Z/14/Z and 105942/Z/14/A), Medical Research Council (MR/P019013/1), Worldwide Cancer Research and Guts UK (19-0192 and 23-0063). This research was funded in whole, or in part, by the Wellcome [203151/Z/16/Z, 203151/A/16/Z] and the UKRI Medical Research Council [MC_PC_17230] and core support grant for Cambridge Stem Cell Institute Discovery Research Wellcome Platform "Discovery Research Platform for Tissue Scale Biology" [226795/Z/22/Z]. This work also received funding from the European Research Council (ERC) Executive Agency under HORIZON ERC Synergy Grant Programme (Grant agreement No: 101167202 — ClonEScape — ERC-2024-SyG; M.P.A.). G.S. was funded by the Isaac Newton Trust (21.07(a)), Medical Research Council (MR/P019013/1) and Worldwide Cancer Research (19-0192 and 23-0063). J.E.R.A. was supported by a University of Cambridge / Wellcome Junior Interdisciplinary Fellowship (ISSF 11/2/2020), and the Medical Research Council (MR/P019013/1). Y.D. was funded by an ELBE Postdoctoral Fellowship from the Center for Systems Biology Dresden (CSBD). M.T.B. received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement

No 794664 (OESOPHAGEAL FATE). H.A. and M.T.B was also supported by the Isaac Newton Trust (Research Grants 16.24(e)) and Leverhulme Trust (RPG-2023-136). D.R. was funded by Cancer Research UK Cambridge Centre (CANCTA-2023/100003). S.H. acknowledges funding from the Human Frontier Science Program (LT000092/2016-L). P.H.J., K.M. and J.C.F. are supported by the Wellcome (108413/A/15/D) and a Cancer Research UK programme grant (C609/A27326). A.N. and W.K. were funded by CRUK (RCCCSF-Nov22/100005). D.A. M.H.H.S. and J.J were funded by intramural sources of TU Dresden allocated to Institute of Pathology and Institute of Anatomy, respectively. S.R. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 950349). B.D.S. acknowledges funding from the Royal Society (E.P. Abraham Research Professorship, RSRP\R\231004) and Wellcome (219478/Z/19/Z).

Author contributions

G.S., J.E.R.A. and M.P.A. designed, validated and conducted experiments; S.H. guided the experimental design for single-cell RNA sequencing and together with H.A. and J.E.R.A. performed data processing. H.A. and J.E.R.A. conducted the trajectory inference analysis. H.A. and Y.D. carried out cell-to-cell communication analysis and advised and supported single-cell RNA sequencing analysis overall. S.R. analysed lineage tracing data and provided expert advice and guidance on single-cell RNA sequencing data analysis. J.C.F. processed DNA sequencing samples and performed data analysis. B.C. provided insights and technical expertise in targeted DNA sequencing. M.T.B. carried out *in vivo* transplantation assays. A.N., S. Z. were involved in human tissue recruitment and D.R., K.M., W.K assisted with human tissue experiments. D.A. performed histopathological staging of human tumour samples. M.H.H.S. and J.J. advised comparative anatomical aspects of human and rodent oesophagus and sample processing M.K.S.T. assisted with *in vitro* validation experiments. M.T.B., B.C., P.H.J. and B.D.S advised on various parts of the study and provided expertise in the epithelial stem cell and tumour biology fields and assisted with manuscript writing. G.S. and M.P.A conceived the project, supervised/performed experiments, and wrote the manuscript with input from all authors. Review and editing of final manuscript was carried out by all authors.

For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Competing Interests Statement

The authors declare no competing interests.

Extended Data Figure 1: Early tumorigenesis model in squamous upper-gastrointestinal tract of a mouse; (Corresponds to Fig. 1). a, H&E image of transversal cross-sections from the murine upper gastro-intestinal tract (including oesophagus and forestomach) displaying multiple tumours (black arrowheads) 6m post-DEN. Scale bars, 100 µm. **b**, Number of tumours per oesophagus (OE) at the indicated time points post-DEN, n=3 mice per time point. A line is drawn across the means of each time point. One way Welch's ANOVA with multiple comparison was used to assess significance, indicated p values show two consecutive timepoint comparison indicated by the black line. **c**, Cartoon (edges) illustrating tumour clearance over time, with examples of confocal images (middle) of growing persistent tumours stained for DAPI (blue) and KRT6A (red). Tissues were collected at the indicated time points post-DEN: 10d, 2m and 1y from 5 mice per timepoint. Scale bars, 25 µm. **d-e,** Confocal images showing early tumour marker expression in nascent tumours 10d post-DEN from 18 mice. DAPI (blue) (**d** & **e**); KRT6A (red) (**d**); KRT17 (yellow) (**e**). Scale bars: 25 µm. **f**, H&E image of a transversal cross-section depicting a nascent tumour 10d post-DEN (black triangle), representative image from 4 mice. **g**, H&E images of tumours transversally cross-sectioned at different time points post-DEN at different stages of progression in the squamous upper gastro-intestinal tract. Scale bars, 50 µm. n=4 (1m), 4 (9m), 10 (12m), 2 (14m) mice per timepoint. **h**, Bar plot (left) showing distribution of DEN induced pathology before (<) and after (>) 9m post-DEN. Note, 10d tumours are not identifiable to pathologists due to their size. **i**, Representative whole mount confocal images from control (ctrl) or 10d post-DEN oesophagi showing localised KRT17 expression (yellow) in tumours, otherwise absent in surrounding or age-matched control tissues. Scale bar, 500 µm. **j**, Percentage of proliferating (KI67+) fibroblasts in DEN-treated oesophagi (0d post-DEN) and in normal age-matched controls from *Pdgfra*^{EGFP} mice. n=17 and 9 areas per condition respectively, each from 3 animals. Data is expressed as mean±s.e.m. Statistical significance was assessed by a two-tailed Mann-Whitney U test. Days (d), months (m), year (y). Panels (**c,h**) created in BioRender https://BioRender.com/z50odrm. Source data (**b,h,j**).

Extended Data Figure 2: Pre-cancerous niche promotes epithelia proliferation and tumour growth; (Corresponds to Fig. 1). a, Density of Niche+ and Niche- tumours at the indicated time points post-DEN. n=3 mice per time point; solid lines represent means. One-tailed Mann-Whitney test Niche- versus Niche+ tumours per time-point. **b,** Confocal images showing EdU incorporation (green) in KRT6A+ nascent tumours (red, dashed line) 10d post-DEN. DAPI (blue). Scale bars, 10 μm. **c**, Number of EdU+ basal keratinocytes normalised per basal cell density at the indicated time points post-DEN. N, number of normal areas or Niche- or Niche+ tumours; n=84 (10d), 20 (2m) Niche- tumours, n=35 (10d), 38 (2m) Niche+ tumours, and n=20 (10d), 14 (2m) DEN areas from 3 mice per condition. Data expressed as violin plots with mean (solid lines) and quartiles (dashed lines). One-way Welch's ANOVA with multiple comparison was used to assess significance. **d**, Quantification of basal keratinocyte cells per area in DEN, Niche- or Niche+ early tumours. N is number of DEN or tumour areas from 3 mice. Data expressed as mean+s.e.m. One-way Welch's ANOVA with multiple comparisons was used to assess significance. **e**, 3D-rendered confocal side-views of a tumour 3m post-DEN and its respective age-matched control showing *R26*^{Fucci2aR} tissue (mCherry, G1 cells; red mVenus, S/G2/M cells; green). PDGFRα (fibroblasts, greyscale). Scale bar, 50 μm. White arrows mark proliferating cells in the basal layer. **f**, 3D-rendered confocal side-views showing growing Niche+ tumours at the indicated time points post-DEN; DAPI (blue); KRT6A (red); PDGFRα (greyscale). **g**, Representative confocal images of forestomach epithelia 10d and 3m post-DEN. DAPI (blue); KRT6A (red). White arrowheads highlight tumours at indicated time points. Scale bars, 1.5mm. Data form n=3 animals per timepoint. **h**, Representative confocal images of a transversal cross-section of a forestomach tumour showing the presence of the stromal niche 8m post-DEN. DAPI (blue); KRT6A (red); KRT17 (yellow); PDGFRα (greyscale). Scale bar, 25 μm. Data from n=4 animals. **i**, Schematic illustration of *ex vivo* 3D heterotypic organ cultures. After separating epithelial and stromal layers, control epithelia were combined with stroma of surviving tumours 3m post-DEN (dashed lines). This was compared to tissue constructs combining control epithelia and control stroma. **j,** Schematic representation of the heterotypic construct approach used (left). Ep, epithelium; WT, wild-type. Images (middle and right) show the emerging tumour like structure in control epithelium, side and top-down views, respectively. Grafted epithelium expressed nuclear tdTomato (nT, red, from *nTnG* mice). DAPI (blue); β-catenin (CTNNB1, green), PDGFRα (greyscale). Scale bars, 25μm. Data from 10 biological replicates. **k**, Experimental protocol (left) and representative side-views of 3D-rendered confocal images (right) of 3D heterotypic tissue constructs 7d post-culture and *in vivo* control or tumour tissue for comparison. PDGFRα labels fibroblasts (greyscale); in *R26*^{Fucci2aR} tissue mCherry, G1 cells (red); mVenus, S/G2/M cells (green). Scale bars, 25 μm. Ep, epithelium; WT, wild-type; Ctrl, control. Tumour bearing tissue was collected 3m post-DEN. **l**, Quantification of cycling (mVenus+) basal keratinocytes (Krt) from **k**, expressed as percentage

of total number of basal cells. Dots are data from an individual tumour (Tmr, *in vivo*), tumour-like structure (Tmr-like, heterotypic culture), and the respective *in vivo* and *in vitro* control areas. Dots represent areas assessed across different biological replicates n=9 control areas, from 3 mice *in vivo*, 23 control areas from 3 mice *in vitro* and n=10 tumours from 6 mice *in vivo* and 7 tumour like structures from 5 mice *in vitro*. Data expressed as mean±s.e.m. One-way Welch's ANOVA with multiple comparisons were used to assess significance. **m,** Schematic representation of the subcutaneous heterotypic grafting approach to NOD SCID gamma (NSG) immunodeficient mice. Grafted epithelium expressed nuclear tdTomato (nT, red, *nTnG* mice), while accompanying stroma was EGFP (green, from *H2B-EGFP* mice) to distinguish the origin of cells in the graft. DEN treated tissue was collected 3m post-DEN. **n**, Representative confocal images (from **m**) showing the long-term survival of a graft combining healthy untreated nT epithelium with EGFP tumour or control stroma. Scale bars 500 μm. **o**, Quantification of surviving and lost graft constructs from **n,** 3-6m post-transplantation expressed as percentage. 19 control and 23 DEN constructs were transplanted across 10 animals, respectively. Statistical significance was determined by a one-sided Chi-squared test. Days (d), months (m), control (ctrl). Panels (**i,j,k,m**) created in BioRender https://BioRender.com/kvs2ti9. Source data (**a,c,d,l,o**).

Extended Data Figure 3: Nascent tumour niche is mainly comprised of fibroblast; (Corresponds to Fig. 2). a, Schematic representation of tissue layers of the murine upper-gastrointestinal tract. **b**, Schematic representation of the imaging angle used according to the sample preparation method. **c**, Image of tissue whole-mounts showing differential PDGFRα (greyscale) expression between the fibroblast layers. α-SMA smooth muscle layer (green), DAPI (blue). Scale bar, 25 μm. Lamina propria- Lp, Muscularis mucosae- Mm, Submucosae- Sb. **d**, Image of a tissue section showing reduced PDGFRα expression in upper stromal layers (i.e. lamina propria). Dashed lines (**c,d**) separate the tissue layers as listed in schematic. **e**, Confocal images of nascent tumour stroma 10d post-DEN. Images show active YAP (aYAP) expression in PDGFRα labelled fibroblasts (white arrowheads). aYAP (magenta); CD31 and CD45 (green); PDGFRα (greyscale); DAPI (blue). Scale bars, 25μm. **f**, Representative confocal images of Niche+ tumours 10d post-DEN, showing the absence of typical cancer associate fibroblast (CAF) marker in the niche forming fibroblasts. **g,** Number of CD45+ (immune) cells in epithelia of Niche-/Niche+ tumours and adjacent DEN area normalised to surface area, 10d post-DEN. n=27 (Niche-), 22 (Niche+) tumours, from 3 mice each, dots represent a tumour. Data expressed as mean±s.e.m. Statistical significance assessed by one-way Welch's ANOVA with multiple comparisons. **h**, Experimental DEN carcinogen protocol in wild-type (wt) and immunocompromised NOD SCID gamma (NSG) mice. **i, j,** Niche+ (**i**) and Niche- (**j**) tumours per area in wt and NSG mice at 10d and 2m post-DEN withdrawal. N=3 wt and 4 NSG mice at 10d and n=3 wt and 3 NSG at 2m post-DEN time point; expressed as mean±s.e.m. One-tailed Mann-Whitney test was used to assess statistical significance. **k**, Representative confocal images of Niche+ nascent tumours from 10d post-DEN showing fibroblast (PDGRα, greyscale) recruitment (white arrow) to a forming epithelial tumour (KRT6A, red) in wt and NSG mice. CD45 (green), DAPI (blue). Scale bar, 10 μm, **l**, Diameter (μm) of Niche+ and Niche- tumours, n=126 (tumours in wt from 3 mice) and n=189 (tumours in NSG from 4 mice) at 10d post DEN and n= 57 (tumours in wt from 3 mice) and n=36 tumours in NSG from 3 mice) at 2m post-DEN. Data expressed as mean±s.e.m. Two-tailed Welch's t-test was used to assess statistical significance. Days (d), months (m). Panels (**a, b**, **h**) created in BioRender https://BioRender.com/8klr068. Source data (**g, i, j, l**).

Extended Data Figure 4: Expansion of PDGRα^{low} fibroblast population in the early tumour niche. (Corresponds to Fig. 2) a, Experimental protocol of fibroblast lineage tracing in *Col1a2^{CreER}Rs26^{FConfetti}* mice. **b**, Confocal images of confetti-labelled cells 24h after induction in different stromal layers. Scale bars, 25 µm. **c, d**, Efficiency of confetti construct recombination after a 72h chase in the PDGFRα+ fibroblast population. Total percentage of recombined cells in the PDGFRα+ fibroblast population (**c**), and split colour percentage within the recombined population (**d**). N represents number of fields from 3 animals, n=36; each replicate shown with a different shade of colour. **e**, Confocal 3D top-down or side-view images of a DEN area. Scale bars, 50 µm. Corresponds to Fig. 2d-f. **f**, Distance to the nearest neighbour fibroblast (fbr) labelled in the same colour in external control (Ctrl), internal control (DEN) and tumours. The horizontal line denotes the average, and the box signifies 95% confidence intervals obtained via bootstrapping. Pairwise Wilcoxon signed-rank tests were performed, and P-values were adjusted using the Holm-Bonferroni correction. **g,** Cartoon illustrating fibroblast expansion in the early tumour niche: lineage tracing results revealed that stroma between tumour lobes show increased fibroblast density and decreased distance between the same colour fibroblasts (black arrowheads) if compared to the tissue outside of lobes or control. **h**, Histogram showing the percentage of cells in a given bin of the distance to the nearest lobe contour. **i**, The average distance to the nearest neighbour of the same colour as a function of the distance to the lobe boundary. Results show that labelled cells are closer together on the lobe contour than elsewhere. Data expressed as mean and error bars denote 95% confidence intervals obtained from boot strapping. The vertical dashed line in **h** and **i** denote 10 µm, the estimated width of the tumour lobe contour. **f,h** and **i** data from n=18 areas in control from 3 animals, n=18 areas in DEN from 3 animals and n=16 tumours from 4 animals. **j**, Schematic representation of the confetti cassette and experimental protocol of fibroblast lineage tracing in *Pdgfra^{CreERT}Rs26^{FConfetti}* mice. **k,** Images of labelled fibroblasts (from **j**) in lamina propria and submucosae; data from 5 animals. **l,** Percentage of fibroblasts labelled in lamina propria (Lp) and submucosae (Sb) (from **j**). **m**, Representative confocal images of tissue from **j** showing top-down projections and confetti labelled fibroblast; 5 and 3 animals for low and high TAM dose, respectively. Quantified induction efficiency per field of view shown as an average above the images. **n**, Experimental DEN protocol in TAM induced *Pdgfra^{CreERT}Rs26^{FConfetti}* mice. **o**, Representative confocal images of confetti labelled fibroblasts in early tumours 6w post-DEN withdrawal. Dashed lines mark tumour and stromal layer margins. 0.5mg/20g body weight is low; 5mg/20g body weight is high tamoxifen dose. DEN treated animals represent n=4 (high TAM) and 3 (low TAM). PDGFRα (greyscale), α-SMA (in **b** greyscale); CONFETTI (cyan, yellow and red) in (**b**). Dashed lines in (**e,k,o**) label lamina propria and submucosae layers in side views. Hours (h), weeks (w), days (d), months (m), tamoxifen (TAM). Scale bars (**k, m, o**) 50 µm. Panels (**a**,**g**,**j**,**n**) created in BioRender https://BioRender.com/7zgcsff. Source data (**c, d, l**).

Extended Data Figure 5: Single cell RNA sequencing annotation; (Corresponds to Fig. 3) a, Schematic visualisation of representative DEN-induced tumours under the dissection microscope, before and after microdissection using punch biopsy tool. **b**, Image composite depicting sample collection strategy for scRNA-seq. Squamous oesophagus and forestomach (outlined in red) first underwent tumour marking and were then peeled to separate epithelial (Ep) and Stromal (Str) compartments. This was followed by sample biopsying, where tumours and size-matched control biopsies (adjacent tissue as internal control, DEN; untreated tissue as external control, Control) were micro-dissected for scRNA-seq. **c**, UMAP representing cell cluster distribution as defined by Seurat. **d**, UMAP cell distribution of cell-types identified using label transfer from (Grommisch et al, 2024)[73]. **e**, Heatmap showing expression of representative marker genes used for curated cell type annotation across the 22 clusters. Log-transformed normalised expression levels were averaged by cluster for each gene and scaled across all cells belonging to each group. Scale bar denotes expression range (scale: -4 to 6). **f**, UMAPs showing expression of representative genes for each cell type identified. Colour bars of UMAPs indicate log-transformed normalised expression levels. **g**, UMAP cell distribution representing identified cell types in each condition. Panel (**a**) created in BioRender https://BioRender.com/l73dcie.

Extended Data Figure 6: Tumour niche fibroblasts transition towards Cancer Associated Fibroblast (CAF) identity. (Corresponds to Fig. 3) a, Heatmap showing the log_2 fold-change (FC) of differentially expressed genes in *Pdgfra*^{low} relative to *Pdgfra*^{high} fibroblasts. Genes characterising *Pdgfra*^{low} shown in red, *Pdgfra*^{high} in blue. DAPI (blue). Scale bars, 10 μm. Dashed line marks fibroblast contour. **b,** UMAP projection denoting the heterogenous expression of *Fn1* (like *Pdgfra*) in fibroblasts; inset shows fibroblast cluster distribution. **c**, Violin plots showing the levels of *Pdgfra* and *Fn1* expression in fibroblast clusters i.e., 3, 2, 4, and 19 in control conditions. *Fn1* expression levels are inversely proportional to those seen of *Pdgfra*. Black line across violin plots shows is mean. **d**, Representative images of stromal whole mount showing fibronectin (FN1, green) expression in lamina propria PDGFRα^{low} (white arrowhead) and submucosae PDGFRα^{high} fibroblasts, labelled in yellow (from induced *Pdgfra*^{CreER}*Rs26*^{FConfetti} mice; EDF 4j; n=3 mice). **e,** Stacked bar plot (left) showing fibroblast cluster enrichment across conditions. Significance was assessed by one-sided Chi-squared test, standardised residual (Std Res) values are shown in the heatmap on the right, indicating Cluster 19 (C19) to be the most enriched cluster in tumour conditions **f**, Volcano plot showing differential gene expression between Tumour and DEN conditions in Cluster 19, *Pdgfra*^{low} fibroblast cluster enriched in tumours. Non-significant genes ($p \geq 0.05$), green and grey. Significant genes with FC>1, red; FC≤1, blue. Genes defining the profibrotic nature of cluster 19 in tumours are labelled on the right. **g** Representative confocal images of a DEN adjacent area and nascent tumours 10d post-DEN or persisting tumour 6m post-DEN showing the accumulation of Fibronectin (FN1, green). DAPI (blue). Scale bars, 25 μm (10d), 100 μm (6m). **h**, Second harmonic generation (SHG) image of lamina propria (directly underneath the epithelium) in a control (Ctrl) and tumour areas 3m post-DEN. Dashed lines mark insets. Extracellular matrix (ECM) fibres detected by SHG, (cyan); nuclei (red). Scale bars, 25 μm. The yellow diamonds represent the longitudinal orientation of the oesophagus. **i**, ECM fibre density was scored as SHG signal intensity 3m post-DEN. N is the number of regions or tumours assessed from 6 animals; n=36 (Ctrl), 20 (DEN), 31 (Tmr). Bars indicate mean±s.e.m. One-way Welch's ANOVA with multiple comparison was used to assess significance. **j**, Plot depicting the orientation of ECM fibres in control regions and tumours 9m post-DEN. N is the number of fields of view assessed in 3 animals per condition, n=12 (Ctrl) and 8 (Tmr). Data expressed as average ±s.d. **k**, Representative confocal image of a nascent tumour from 6 animals, showing integrin α6 (CD49f, magenta) sequestering in the early tumour niche. Samples were also labelled for KRT6A (red), PDGFRα (greyscale) and DAPI (blue). Image settings adjusted to upper stromal layer. **l,** Venn diagram displaying overlapping genes between known CAF markers (top left, red) and differentially expressed genes (DEGs) in DEN tumour (Tmr) fibroblasts (Cluster 19) relative to external control (Ctrl; top right, yellow) or to internal control cells (DEN; bottom, green). Overlapping genes are listed. Known subtypes of CAFs: Myofibrotic (my), immune (i), and antigen-presenting (ap)

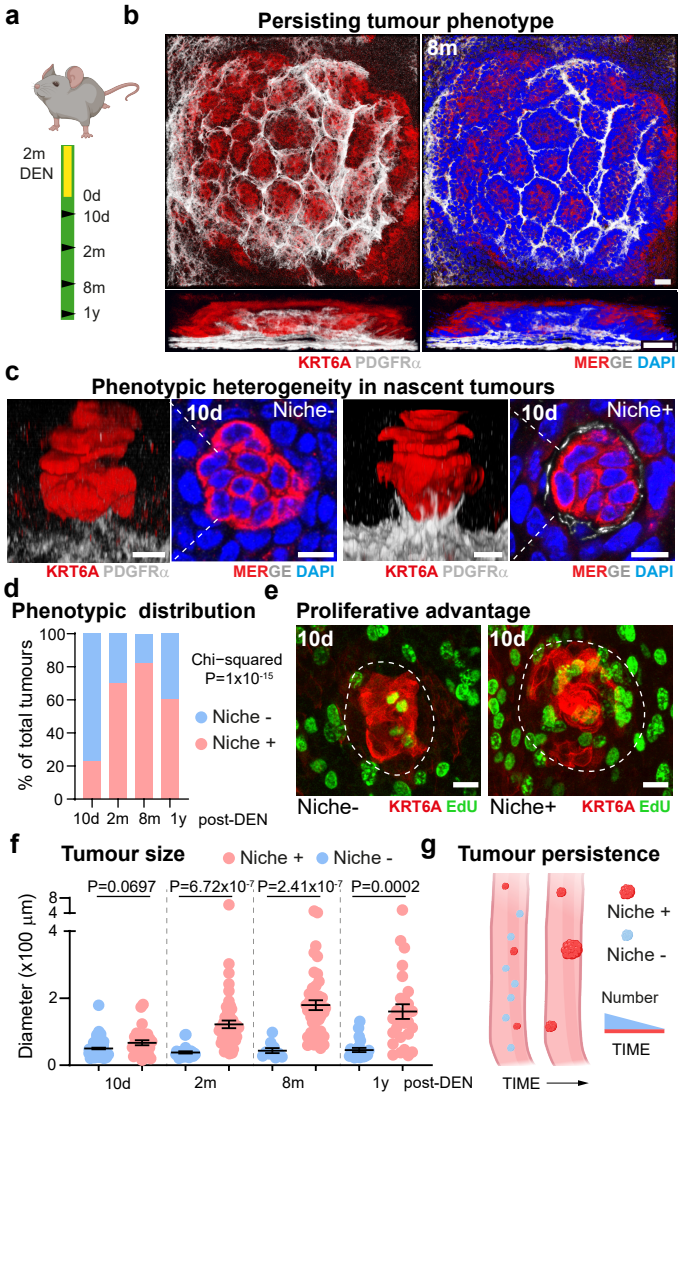
were considered. **m,n,** Violin plots showing expression of signature genes found to be enriched in Cluster 19 tumour fibroblasts (**m**) or other canonical CAF markers (**n**); external control (ctrl, green), internal control (DEN, blue), and tumours (Tmr) (red). Black line is mean. **o,** Representative 3D-rendered side-view images of control tissue and tumours 6-7m post-DEN. Expression of CAF markers (FAP, VIM, FSP) was not detected in niche fibroblasts. KRT6A (red); CD45 (immune), and CD31 (endothelia) (cyan); Vimentin (VIM) (green); PDGFRα (top left and right) and FSP (bottom), (greyscale); FAP (green), DAPI, blue. Scale bars, 50μm. Dashed line outlines tumour area. **p**, Quantification of Ki67+ EGFP+ fibroblasts from *Pdgfra*^{EGFP} mice in control and Niche+ tumours 6-12m post-DEN. N= 25 control areas (from 7 animals) and 50 tumours (from 8 animals). Data is expressed as mean±s.e.m. Two-tailed Mann Whitney test. **q,** Representative bright field (H&E) and confocal (IF) images of invasive carcinoma from 14m, post-DEN withdrawal tissue. H&E was used to diagnose the pathology. In different panels αSMA, FAP, VIM (green) co-expression with PDGFRα (greyscale) show CAF emergence in DEN model. DAPI (blue). Scale bars, 100 μm (black) 50 μm (white); representative data from 5 animals >12m post DEN. **r,** Violin plots showing the prediction score from label transfer for *Pi16*+ and *Col15a1*+ cross-tissue universal fibroblast population (Buechler et al, 2021)[40] in upper GI fibroblast clusters. **s,** upper GI fibroblast clusters (top left). Remaining UMAPS show *Pdgfra* expression in UMAP space of upper GI and cross-tissue fibroblasts (Buechler et al, 2021). Days (d), months (m). Source data (**i,p**).

Extended Data Figure 7*:* Niche of growing tumours recruits other stromal cells. (Corresponds to Fig. 3) a, Experimental protocol for targeted DNA sequencing of stromal niche 7m post-DEN. **b**, Substitutions per megabase (Sb/Mb) were calculated and used as indicative of the mutation burden in stroma across conditions. n=6 control (from 4 animals), 13 DEN (from 6 animals) and 18 tumour (from 6 animals). Epithelial tissue (12m post-DEN treatment) was used as benchmark and includes both tumour and tumour-free tissues (DEN+Tmr; from 3 mice). Data is expressed as mean±s.e.m. Two-tailed Welch's t-test comparing tumour stroma with internal or external control stroma or with epithelium. **c,** Stacked bar plot showing scRNA-seq captured cell type distribution across conditions. **d**, Representative 3D-rendered confocal image of an early tumour (KRT6A, red) 4m (left) and 7m (right) post-DEN withdrawal showing the recruitment of immune cells (CD45, green) and blood vessels (CD31, orange) to the persisting tumour niche. DAPI (blue). Scale bar, 50 µm. **e**, Heatmap showing expression of representative marker genes across the 14 immune cell types identified in the scRNA-seq data. Log-transformed normalised expression levels were averaged by cluster for each gene and scaled across all cells belonging to each group. Scale bar denotes expression range (scale: -2 to 2). **f**, UMAP of identified immune cell types split by condition. **g**, Top, stacked bar plot showing immune cell type enrichment across conditions. Significance was assessed by a one-sided Chi-square test. (Bottom) heatmap showing standardised residuals values from Chi-square test. Asterisk (*****) marks cell types used in CellChat analysis. **h**, Heatmaps representing outgoing (left) and incoming (right) interaction strengths as identified by CellChat and rescaled to their row maxima. Cumulative interaction strength is depicted by bars on the right. Months (m). Panel (**a**) created in BioRender https://BioRender.com/udeggcf. Source data (**b**).

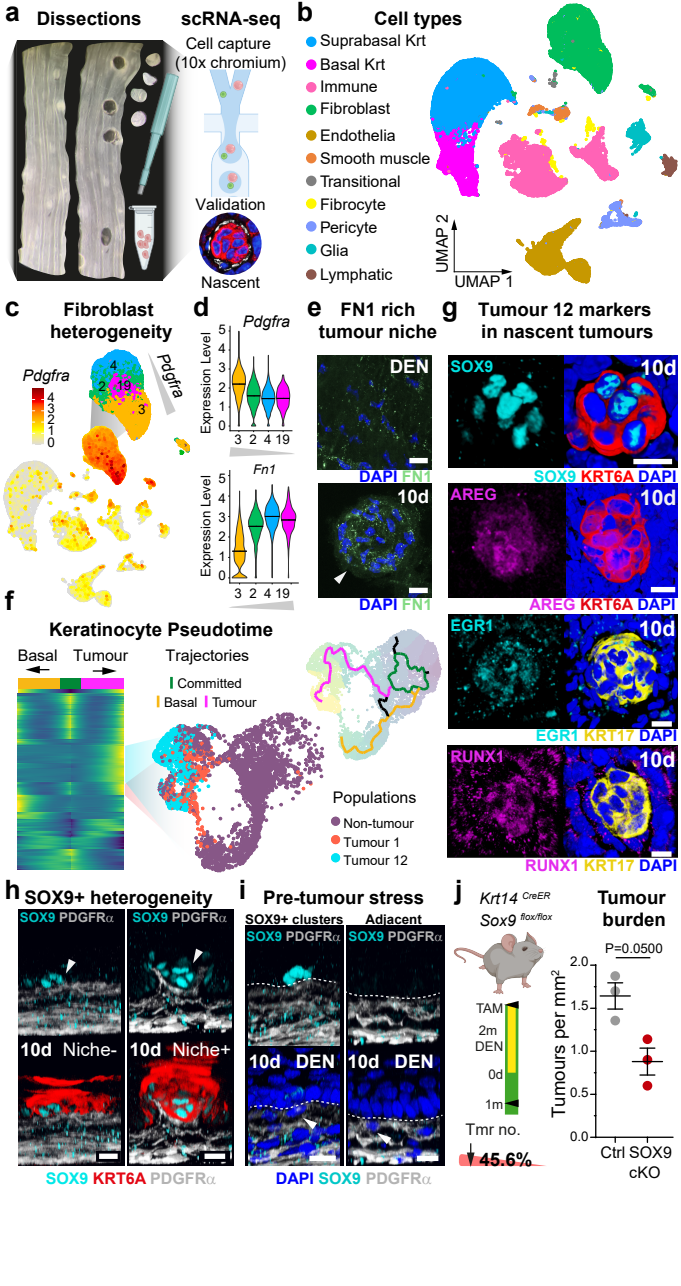
Extended Data Figure 8: Pseudotime analysis of basal keratinocytes denotes distinctive early tumour state; (Corresponds to Fig. 3) a, Left (inset), basal keratinocyte cell distribution in the UMAP space annotated by condition (right, UMAP consisting of all cell types). **b**, UMAP of basal keratinocyte cell distribution representing cells in each condition. **c**, UMAPs showing expression of representative basal keratinocytes genes (*Col17a1,* basal; *Top2a,* cycling basal; *Krt4*, committed). Colour bars of UMAPs indicate log-transformed normalised expression levels. **d**, UMAP showing cell clusters as defined by Seurat. **e**, Mapped RNA velocity of basal keratinocytes colour coded by condition. Black arrows are velocity vectors that show predicted cellular state change overtime. **f**, Basal keratinocyte two-dimensional pseudotime (PST) UMAP (top) and trajectories (black line) (bottom) inferred by Monocle 3. Legend shows the range of pseudotime values **g**, Heatmap (left) of the top 1500 genes identified as differentially expressed between cells in the tumour and basal trajectories. Log-transformed expression levels were scaled in a gene-wise fashion from -4 to 4 (scale) and are shown after hierarchical clustering along trajectories from committed to basal (blue to orange) and from committed to tumour (blue to magenta). The average expression patterns of the genes contained in each cluster are depicted on the right, with two modules (1 and 12) identified as tumour unique based on their preferential upregulation at the end of the tumour trajectory (Δ_tumour > 2, Δ_basal < 1). (Blue, committed; Magenta, tumour; Orange, basal). Dashed vertical lines label the highest PST score of committed. **h**, Tumour cell populations representing Tumour 1 and Tumour 12 clustering in UMAP space. **i,** Violin plots highlighting *Krt6a* and *Krt17* expression between Tumour 12 and Tumour 1 states. **j**, Volcano plot showing differential gene expression between Tumour 12 and Tumour 1 states. Non-significant genes (p value higher than p=0.05), green. Significant genes with FC≤1, red. Genes driving stress and extracellular matrix related processes are labelled. **k,** Gene Set Enrichment Analysis of Tumour 12 versus Tumour 1 differentially expressed genes. Terms ranked by FDR (false discovery rate); representative genes listed below the bar plot. GO:BP (Gene Ontology Biological Processes); KEGG (Kyoto Encyclopedia of Genes and Genomes). Cartoon (top right) representing Tumour 12 keratinocyte communicates with stroma. **l,** Violin plots highlighting differences between Tumour 12 and Tumour 1 states. **m**, Representative confocal images from 6 mice showing expression of tumour markers: SOX9 and EGR1 (cyan); AREG and RUNX1, (magenta); KRT6A, red; KRT17 (yellow). DEN images were acquired 10d post-DEN in regions adjacent to tumours. DAPI (blue). Scale bars, 10 μm. **n**, Relative frequency (%) (top) of tumours expressing cells positive for SOX9 (Tumour 12 state marker) per 1,000 μm^2, tumours 10d post-DEN, n- number of tumours=119 from 3 biological replicates. **o**, Representative confocal image from 10 mice demonstrating SOX9+ (cyan) keratinocytes are situated in direct contact to stroma (PDGFRα, greyscale). KRT6A (red), DAPI (blue). Scale bar 50 μm. Days (d), months (m). Panels (**k**) created in BioRender https://BioRender.com/mhoqqbw. Source data (**n**).

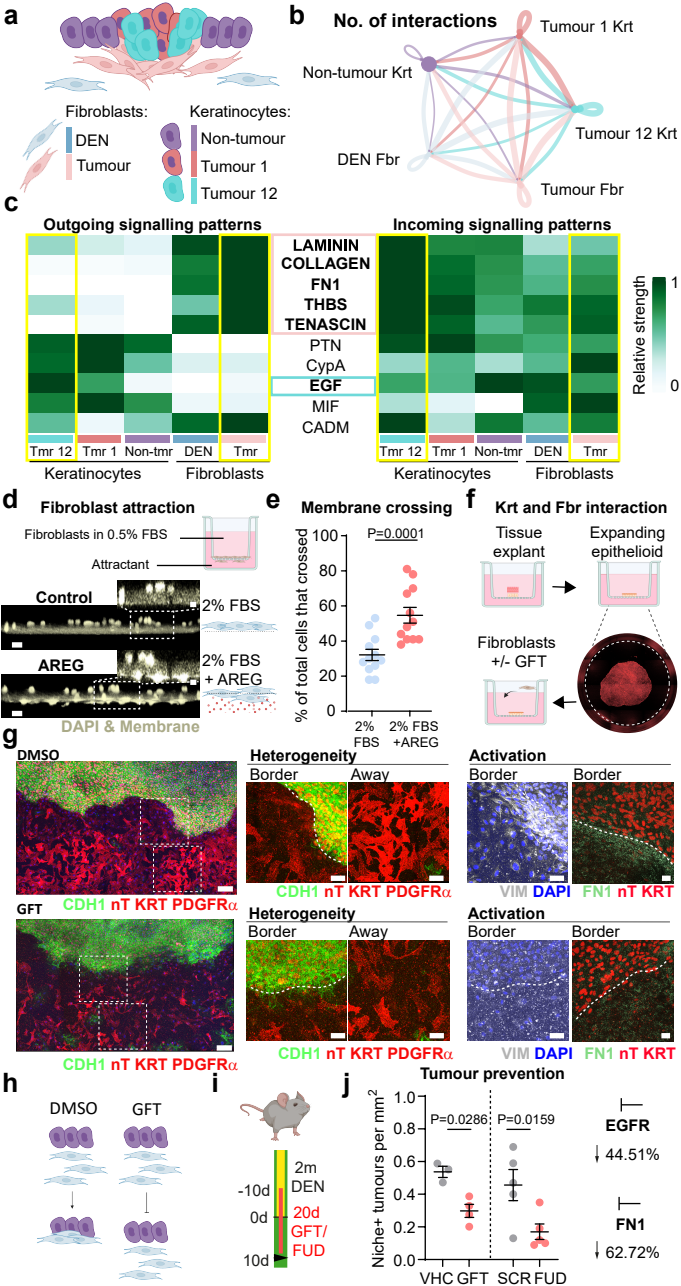
Extended Data Figure 9: SOX9+ early tumour keratinocytes are associated with stromal reorganisation; (Corresponds to Fig. 3) a, 3D-rendered confocal images of Niche+ and Niche- tumours showing KRT6A (red) and SOX9 (cyan) 10d post-DEN. PDGFRα (greyscale), DAPI (blue). Scale bars, 10 µm. Corresponds to Fig. 3h**. b, c** Quantification of $SOX9^{+}$ keratinocytes (krt) (**b**) and SOX9 intensity (**c**) in **a,** normalised to surface area. DEN, equivalent morphologically normal areas from DEN-treated tissues. 3 mice per condition with n=20 DEN areas, 84 Niche- and 35 Niche+ tumours. Statistical significance assessed by one-way Welch's ANOVA with multiple comparisons. **d,** 3D-rendered confocal images (basal view, left; side-view, right) of SOX9+ keratinocyte clusters exclusively found in DEN condition. SOX9 (cyan), β catenin (green), PDGFRα (greyscale), DAPI (blue). Scale bar, 10 µm. Corresponds to Fig. 3i. **e,** Quantification of fibroblast proximity to basal layer from **d**, expressed as the inverse of the distance between basal keratinocytes and underlying fibroblasts. 1/distance (μm^{-1}). **f**, Number of fibroblasts directly underneath the basal layer corrected per area. The number of control (n=27) or SOX9 expressing regions (n=42) assessed in tissues from 4 mice. Two-tailed Welch's t-test. **g,** Strategy to validate SOX9 knock-out in basal keratinocytes (left). SOX9 expression was induced by culturing oesophageal tissue using the 3D epithelioid approach. Established cultures from tamoxifen-induced $Krt14^{CreER}$ $R26^{mTmG}Sox9^{flox/flox}$ mice were confocal imaged (right), confirming the presence of areas lacking SOX9 (cyan) expression. n=3, biological replicates. White arrowheads mark the SOX9 loss. Membrane Tomato (purple), DAPI (blue). Scale bar, 25 µm. **h**, Top, experimental protocol: $Krt14^{CreER}Sox9^{flox/flox}$ mice received a dose of Tamoxifen (TAM) followed by the DEN treatment. Tissues were collected 1m post-DEN (from n=3 mice) and tumour diameter was compared to DEN-treated un-induced controls (n=3 mice). Bottom, tumour diameter (µm) of Niche+ and Niche- tumours, from the above. Data expressed as mean±s.e.m. Significance was assessed using two-tailed Welch's t-test. Days (d), months (m). Corresponds to Fig 4j. Panels (**g,h**) created in BioRender https://BioRender.com/ip2ihnq. Source data (**b,c,e,f,h**).

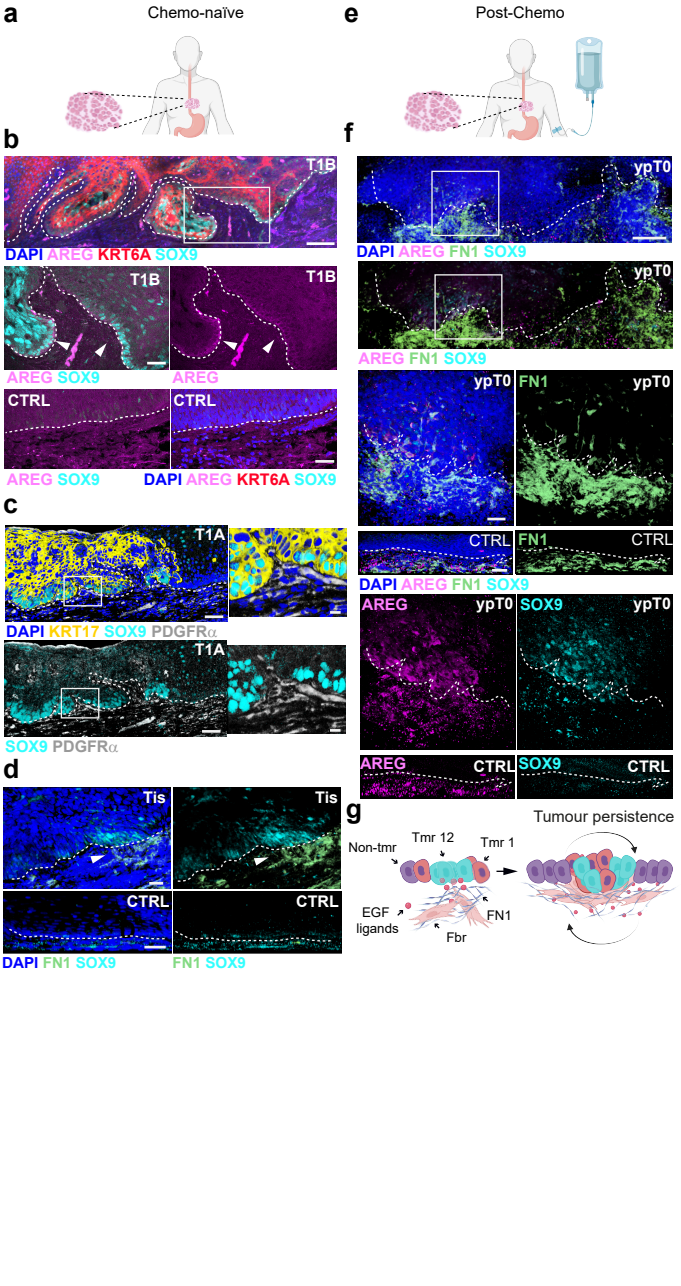
Extended Data Figure 10: Tumour epithelial cells are associated with early tumour stromal remodelling via the EGF-SOX9-Fibronectin axis (Corresponds to Fig. 5) a, Scatterplot displaying the dominant senders and receivers as identified by CellChat. Circle size represents 'communication probabilities'. **b,** Heatmaps (top) showing the relative importance of each cell group based on the computed network centrality scores for the Fibronectin (FN1) and EGF signalling network, respectively. Violin plots of relevant ligands and receptors across different cell types (bottom). These include expression of fibronectin (*Fn1*) and fibronectin-binding receptors (*Itga3, Itgav, Itgb6*), as well as EGF ligands (*Areg, Hbegf*) and binding receptors (*Egfr, Erbb2*). **c**, Migration of fibroblast through cell culture insert membrane (cartoon, top) in low (0.5%) or high (10%) serum (FBS) conditions was quantified (bottom) and expressed as a percentage of the total number of cells. Data expressed as mean±s.e.m. Each dot represents a technical replicate; data 4 biological replicates. One way Welch's ANOVA test with multiple comparison was used to assess significance between groups presented in the main Fig 4e. **d**, schematic representation of regions imaged in epithelioid cultures from (top) and representative confocal images showing SOX9 (cyan) downregulation in keratinocytes in central epithelioid areas as well as on the border of GFT-treated epithelioids. Representative images of the DMSO (ctrl) condition shown for comparison. Nuclear tdTomato (nT) keratinocytes (krt) (red); CDH1 (green); DAPI (blue). Scale bar, 50 µm. **e,** Schematic representation of an experiment (left) and representative confocal images (right) of expanding epithelioid cultures exposed to fibroblasts under DMSO (control) or Gefitinib (GFT) conditions. Vimentin, VIM (greyscale); PDGFRα (red); nuclear Tomato (nT) keratinocytes (krt) (red); DAPI (blue), E-Cadherin, CDH1 (green); Fibronectin, FN1 (green). Scale bar 500 µm (inset 50 µm). Corresponds to Fig 4g. **f**, Top, experimental schematics, bottom, quantification of EdU incorporation (2h chase) in confluent 3D epithelial cultures (Epithelioids) exposed to Fibronectin for 24 hours. Data expressed as a percentage of the total number of cells (mean±s.e.m). Each dot represents a technical replicate; data from 3 biological replicates. Two-tailed Welch's t-test comparing EdU incorporation in control (Ctrl) and fibronectin-treated cells. **g**, Schematic representation of Bleomycin (BLEO) administration to mice (top) and representative confocal images (bottom) of SOX9 (cyan) expression in epithelia of bleomycin treated animals. DAPI (blue). Scale bars, 50 µm. N=5 (control) 4 (BLEO) mice. Panels (**c,d,e,f,g**) created in BioRender https://BioRender.com/nm3lx8y. Source data (**c,f**).


a
b
Persisting tumour phenotype
2m
DEN
0d
10d
2m
8m
1y
8m
KRT6A PDGFRα
MERGE DAPI
c
Phenotypic heterogeneity in nascent tumours
10d
Niche-
10d
Niche+
KRT6A PDGFRα
MERGE DAPI
KRT6A PDGFRα
MERGE DAPI
d
Phenotypic distribution
% of total tumours
100
80
60
40
20
0
Chi−squared
P=1x10-15
Niche -
Niche +
10d 2m 8m 1y post-DEN
e
Proliferative advantage
10d
10d
Niche-
KRT6A EdU
Niche+
KRT6A EdU
f
Tumour size
Niche +
Niche -
P=0.0697
P=6.72x10-7
P=2.41x10-7
P=0.0002
Diameter (x100 μm)
8
4
4
2
0
10d
2m
8m
1y
post-DEN
g
Tumour persistence
Niche +
Niche -
Number
TIME
TIME

a
Niche formed by lamina propria fibroblasts
10d
Ep
Lp
Mm
Sb
KRT6A PDGFRα DAPI
b
Stromal cell abundance in nascent tumours
Fibroblast
Immune
Endothelia
Cells / 10^3 µm^2
P=0.0190
P=0.0005
P=0.0028
P=0.4206
P=0.9909
P=0.0289
P=0.5368
P=0.1452
P=0.0032
Niche-
Niche+
DEN
c
Fibroblasts as a major niche component
10d Niche-
10d Niche+
CD45 CD31 KRT6A
MERGE PDGFRα DAPI
d
Col1a2CreER
R26fConfetti/wt
Col1a2 Cre
Unbiased fibroblast lineage tracing
STOP
GFP
YFP
RFP
CFP
2m DEN
TAM
0d
6m
e
Tumour
6m Tumour
Top-down view
Side view
CONFETTI
CONFETTI PDGFRα
MERGE DAPI
Control
6m Control
f
Tumour
Control
Lamina propria PDGFRα^low
Submucosae PDGFRα^high
PDGFRα


a
Dissections
scRNA-seq
Cell capture
(10x chromium)
Validation
Nascent
b
Cell types
Suprabasal Krt
Basal Krt
Immune
Fibroblast
Endothelia
Smooth muscle
Transitional
Fibrocyte
Pericyte
Glia
Lymphatic
UMAP 2
UMAP 1
c
Fibroblast
heterogeneity
Pdgfra
Pdgfra
d
Pdgfra
Expression Level
Fn1
Expression Level
e
FN1 rich
tumour niche
DEN
DAPI FN1
10d
DAPI FN1
g
Tumour 12 markers
in nascent tumours
SOX9
10d
SOX9 KRT6A DAPI
AREG
10d
AREG KRT6A DAPI
EGR1
10d
EGR1 KRT17 DAPI
RUNX1
10d
RUNX1 KRT17 DAPI
f
Keratinocyte Pseudotime
Basal
Tumour
Trajectories
Committed
Basal
Tumour
Populations
Non-tumour
Tumour 1
Tumour 12
h
SOX9+ heterogeneity
SOX9 PDGFRα
SOX9 PDGFRα
10d Niche-
10d Niche+
SOX9 KRT6A PDGFRα
i
Pre-tumour stress
SOX9+ clusters
Adjacent
SOX9 PDGFRα
SOX9 PDGFRα
10d DEN
10d DEN
DAPI SOX9 PDGFRα
j
Krt14 CreER
Sox9 flox/flox
TAM
2m
DEN
0d
1m
Tmr no.
45.6%
Tumour
burden
P=0.0500
Tumours per mm2
2.0
1.5
1.0
0.5
0.0
Ctrl
SOX9
cKO


a
Fibroblasts:
DEN
Tumour
Keratinocytes:
Non-tumour
Tumour 1
Tumour 12
b
No. of interactions
Tumour 1 Krt
Non-tumour Krt
Tumour 12 Krt
DEN Fbr
Tumour Fbr
c
Outgoing signalling patterns
Incoming signalling patterns
LAMININ
COLLAGEN
FN1
THBS
TENASCIN
PTN
CypA
EGF
MIF
CADM
Relative strength
1
0
Tmr 12
Tmr 1
Non-tmr
DEN
Tmr
Keratinocytes
Fibroblasts
d
Fibroblast attraction
Fibroblasts in 0.5% FBS
Attractant
Control
2% FBS
AREG
2% FBS
+ AREG
DAPI & Membrane
e
Membrane crossing
P=0.0001
% of total cells that crossed
100
80
60
40
20
0
2% FBS
2% FBS +AREG
f
Krt and Fbr interaction
Tissue explant
Expanding epithelioid
Fibroblasts +/- GFT
g
DMSO
GFT
CDH1 nT KRT PDGFRα
Heterogeneity
Activation
Border
Away
VIM DAPI
FN1 nT KRT
h
DMSO
GFT
i
2m DEN
-10d
0d
20d GFT/ FUD
10d
j
Tumour prevention
Niche+ tumours per mm2
1.0
0.8
0.6
0.4
0.2
0.0
P=0.0286
P=0.0159
VHC
GFT
SCR
FUD
EGFR
↓ 44.51%
FN1
↓ 62.72%


a
Chemo-naïve
e
Post-Chemo
b
T1B
DAPI AREG KRT6A SOX9
T1B
T1B
AREG SOX9
AREG
CTRL
CTRL
AREG SOX9
DAPI AREG KRT6A SOX9
f
ypT0
DAPI AREG FN1 SOX9
ypT0
AREG FN1 SOX9
ypT0
FN1
ypT0
CTRL
FN1
CTRL
DAPI AREG FN1 SOX9
AREG
ypT0
SOX9
ypT0
AREG
CTRL
SOX9
CTRL
c
T1A
DAPI KRT17 SOX9 PDGFRα
T1A
SOX9 PDGFRα
d
Tis
Tis
CTRL
CTRL
DAPI FN1 SOX9
FN1 SOX9
g
Tumour persistence
Non-tmr
Tmr 12
Tmr 1
EGF ligands
FN1
Fbr