

ARTICLE OPEN



TGF- β signaling redirects Sox11 gene regulatory activity to promote partial EMT and collective invasion of oncogenically transformed intestinal organoids

Yu-Hsiang Teng^{1,2}, Bismark Appiah³, Geoffroy Andrieux³, Monika Schrempp¹, Katja Rose¹, Angelika Susanna Hofmann¹, Manching Ku⁴, Sven Beyes^{1,6}, Melanie Boerries^{3,5} and Andreas Hecht^{1,2}✉

© The Author(s) 2025

Cancer cells infiltrating surrounding tissue frequently undergo partial epithelial-mesenchymal transitions (pEMT) and employ a collective mode of invasion. How these phenotypic traits are regulated and interconnected remains underexplored. Here, we used intestinal organoids with colorectal cancer (CRC) driver mutations as model system to investigate the mechanistic basis of TGF- β 1-induced pEMT and collective invasion. By scRNA-seq we identified multiple cell subpopulations representing a broad pEMT spectrum, where the most advanced pEMT state correlated with the transcriptional profiles of leader cells in collective invasion and a poor prognosis mesenchymal subtype of human CRC. Bioinformatic analyses pinpointed *Sox11* as a transcription factor gene whose expression peaked in the potential leader/pEMT^{high} cells. Immunofluorescence staining confirmed Sox11 expression in cells at the invasive front of TGF- β 1-treated organoids. Loss-of-function and overexpression experiments showed that Sox11 is necessary, albeit not sufficient, for TGF- β 1-induced pEMT and collective invasion. In human CRC samples, elevated *SOX11* expression was associated with advanced tumor stages and worse prognosis. Unexpectedly, aside from orchestrating the organoid response to TGF- β 1, Sox11 controlled expression of genes related to normal gut function and tumor suppression. Apparently, Sox11 is embedded in several distinct gene regulatory circuits, contributing to intestinal tissue homeostasis, tumor suppression, and TGF- β -mediated cancer cell invasion.

Oncogenesis (2025)14:17; <https://doi.org/10.1038/s41389-025-00560-7>

INTRODUCTION

Distant organ metastasis inevitably requires that cancer cells attain the ability to breach normal tissue boundaries and emigrate from the primary tumor. These processes are thought to be facilitated by epithelial-mesenchymal transitions (EMT) in the course of which cancer cells coordinately lose apical-basal polarity and cell-adhesion systems typical of epithelial cells, while gaining front-rear polarity and other mesenchymal features which foster cellular motility and the capacity to invade surrounding stromal tissue [1]. Quite often, though, EMT appears to be only incompletely and reversibly executed, leading to highly dynamic hybrid states or pEMT whereby cancer cells retain to varying degrees epithelial features while additionally procuring mesenchymal characteristics [1, 2]. From a pathophysiological perspective, pEMT is highly disease-relevant, first, because in experimental models of metastasis, cancer cells in pEMT states were found to be much more potent at seeding secondary tumors when compared to cells in fully epithelial or mesenchymal states [3–6]. Second, pEMT has been associated with a collective mode of cancer cell invasion, which is by far the most frequently observed type of stromal

infiltration in human cancer [5–10]. Yet, how pEMT and collective invasion are mechanistically interconnected is not clear.

In prototypical instances of EMT, diverse extracellular signals and environmental conditions can trigger expression of members of the Snail, Zeb, and Twist families of transcription factors (TFs) [1]. These core EMT-TFs, in turn, orchestrate widespread gene expression changes that underlie the complementary losses and gains of epithelial and mesenchymal phenotypic traits, respectively. However, a growing number of studies showed that EMT processes may not necessarily unfold as a consequence of a uniform, largely invariant transcriptional program. Instead, large-scale transcriptome analyses of a broad spectrum of human cancers discovered a multiplicity of highly context-dependent and molecularly distinct gene expression programs that predominantly manifest as pEMT states [4, 9, 11–15]. As distinguishing characteristics, these newly described EMT programs often do not show downregulation of epithelial markers in conjunction with upregulation of mesenchymal features that is typically observed in prototypical EMT [11, 13]. Moreover, these new EMT variants apparently can be implemented even when expression of core

¹Institute of Molecular Medicine and Cell Research, Faculty of Medicine, University of Freiburg, Freiburg, Germany. ²Faculty of Biology, University of Freiburg, Freiburg, Germany.

³Institute of Medical Bioinformatics and Systems Medicine, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. ⁴Department of Pediatrics and Adolescent Medicine, Division of Pediatric Hematology and Oncology, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. ⁵German Cancer Consortium (DKTK), Partner site Freiburg, a partnership between DKFZ and Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. ⁶Present address: Robert Bosch Center for Tumor Diseases (RBCT), Stuttgart, Germany. ✉email: andreas.hecht@mol-med.uni-freiburg.de

Received: 2 December 2024 Revised: 28 April 2025 Accepted: 9 May 2025

Published online: 20 May 2025

EMT-TFs is very low or undetectable [4, 9, 11, 13, 14]. In fact, CRISPR/Cas9-mediated knock-out (KO) repeatedly showed that core EMT-TFs are not obligatory for EMT [6, 10, 16–18]. Altogether, evidence is accumulating that aside from prototypical, also mechanistically different, non-prototypical EMT programs exist, which are likely to be driven by TFs other than Snail, Zeb, and Twist family members. This raises the question precisely which TFs operate in different EMT settings.

Sox11 together with Sox4 and Sox12 forms the SoxC subfamily of Sry-related (Sox) TFs [19, 20]. Like all Sox proteins, Sox11 is evolutionary highly conserved and contains a high mobility group (HMG) box domain which facilitates sequence-specific binding to DNA [19, 20]. In addition to its pivotal functions in embryonic development and organogenesis [20], Sox11 emerged as a context-dependent tumor-promoting as well as tumor-suppressing factor with the capacity to both positively and negatively regulate cancer cell proliferation, migration, and invasion [20–26]. In breast cancer, Sox11 was tied to the induction of EMT and acquisition of a hybrid epithelial/mesenchymal state [24, 27]. Similar to other Sox family members with well-established roles in the control of EMT processes [19, 28–31], Sox11, therefore, could be an alternative regulator of EMT processes.

Mouse intestinal organoids carrying colorectal cancer driver mutations in *Apc*, *Kras*, and *Trp53* (TKA-organoids) represent versatile models to investigate various aspects of carcinogenesis. When treated with TGF- β 1, TKA-organoids undergo pEMT and form collectively invading cell masses independently of Snail1 and Zeb1 [10]. TKA-organoids thus provide an attractive experimental system highly suitable for the molecular analyses of an intestinal model for pEMT and its relationship to collective invasion. Here, we used single-cell RNA sequencing (scRNA-seq) to comprehensively capture transcriptional and cellular dynamics of TGF- β 1-treated TKA-organoids. We found that upon passing through a series of transitory states, cells at the periphery of collectively invading organoids adopted transcriptomic features of leader cells and expressed the highest levels of mesenchymal genes without downregulating epithelial markers. Sox11 emerged as a key driver of TKA-organoid pEMT with additional, TGF- β 1-independent gene regulatory activities related to normal gut functions and tumor suppression. In human CRC, expression of mesenchymal markers and *SOX11* turned out to be highly correlated, and elevated *SOX11* expression was paralleled by a worse prognosis. Accordingly, as in TKA-organoids, upregulation of *SOX11* might confer mesenchymal and more malignant traits to CRC cells and hence could be a worthwhile drug target.

MATERIALS AND METHODS

Organoid culture

Organoids were derived from C57BL/6-*Apc*^{S805/S805}; *Kras*^{L2D/+}; *Trp53*^{L2L-R172H/+}; *tg(Villin-CreER²)*/N mice as described [10]. Upon initial establishment, organoids were treated with 0.5 μ M 4-hydroxy-tamoxifen (4-OHT; #H7904, Sigma Aldrich, Taufkirchen, Germany) to generate the study-relevant TKA lines (Supplementary Fig. 1). Recombination was verified by PCR (primers listed in Supplementary Table 1).

Transwell invasion assays

Invasion experiments were conducted as described [10], except that 1×10^5 single organoid cells were dispensed into transwell inserts (#353097, Corning Life Sciences, Corning, New York, USA). After seeding, organoid cells were grown for 30–40 h before treatment with TGF- β 1 commenced.

Genome editing

Sox11 was inactivated by co-infecting organoids with viral expression vectors for Cas9 and sgRNAs (Supplementary Table 1), which had been designed and cloned into plasmids (listed in Supplementary Table 2) as described [10, 16]. Post infection/selection, single cell-derived organoid clones were screened by PCR (primers listed in Supplementary Table 1).

Sox11-HA and Prrx2-HA knock-ins were created following published procedures [32] by co-transfection of sgRNA and CAS9 expression vectors with pCRISPR-HOT-FS-FKBP12^{F36V}-HA₂-P2A-EGFP-loxP-PGK-NeoR-loxP-FS (Supplementary Table 2). After selection of transfected cells with geneticin (10131-027, ThermoFisher Scientific, Waltham, MA, USA), single cell-derived clones were isolated and genotyped by PCR. In-frame insertions of the FKBP12^{F36V}-HA₂-P2A-EGFP-loxP-PGK-NeoR-loxP cassette were confirmed by sequencing.

Viral transduction

Lentiviral and retroviral particles were produced following published procedures [10] using plasmids listed in Supplementary Table 2.

Immunofluorescence staining and confocal microscopy

A previously described protocol [10] was employed for whole mount immunofluorescence staining (antibodies listed in Supplementary Table 3) and imaging with a Leica Sp8 confocal microscope.

RNA isolation, cDNA synthesis, and qRT-PCR

RNA was isolated and further processed for qRT-PCR (primers listed in Supplementary Table 4) following published procedures [10, 16].

Bulk RNA-seq

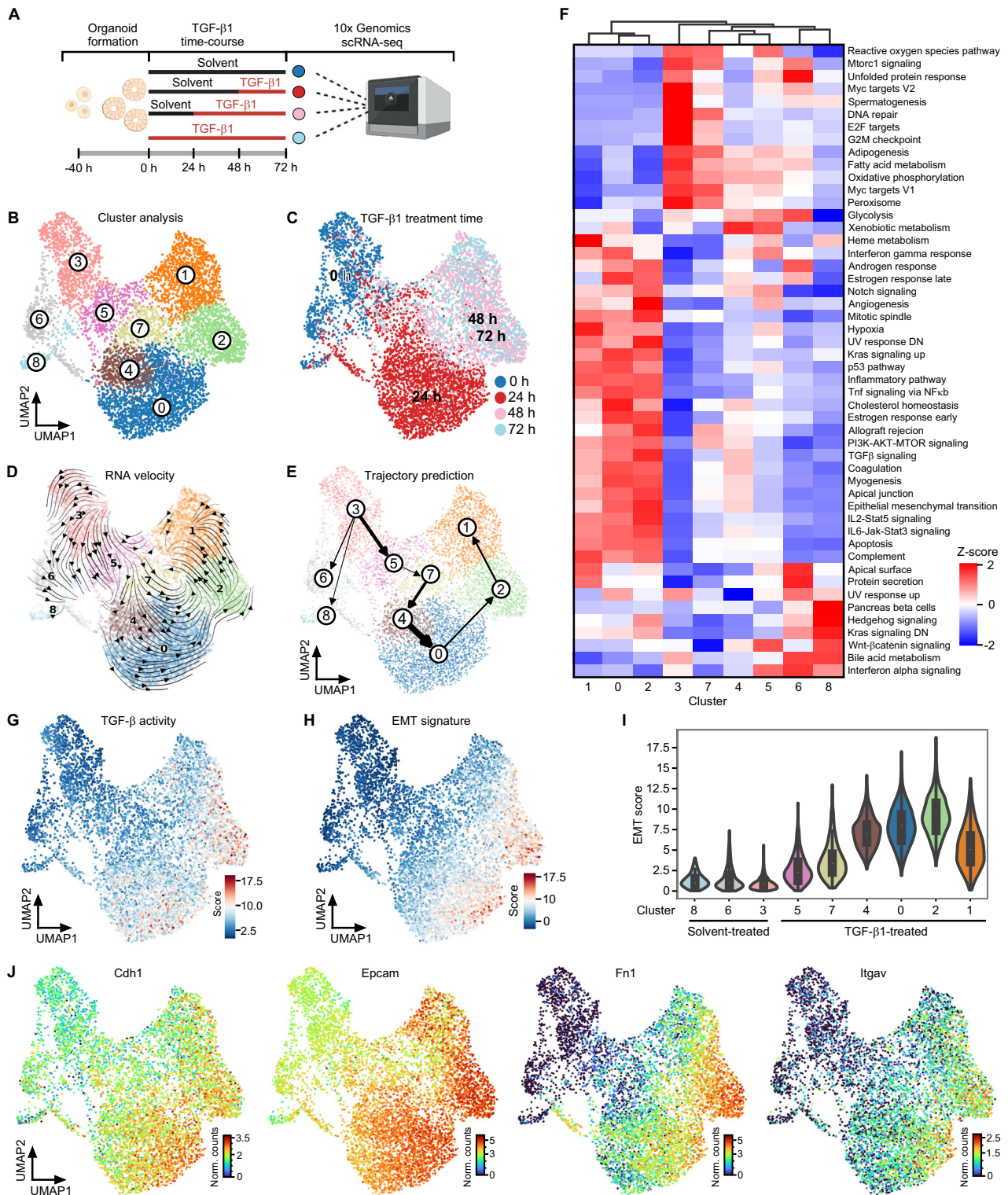
Total RNA was sent to Macrogen Inc. (Seoul, South Korea) for library preparation and sequencing. Raw sequencing results were processed on Galaxy EU (Version 0.23.4) (<https://usegalaxy.eu/>) [33]. Gene expression trajectories were visualized in a heatmap after z-score normalization and k-means clustering [34].

Pathway analysis

Murine gene names were converted to human identifiers using biomaRt [35]. Pathways and gene sets enriched among differentially expressed genes (DEGs) were identified using Enrichr [36].

Single-cell RNA-seq

Preparation of sequencing libraries. For each experimental condition, 1×10^4 single cells were isolated by flow cytometry and further processed for cDNA library construction using a Chromium chip and a 10x Chromium controller following the Chromium Next GEM Training Kit User Guide (v3.1, CG000204 Rev D; 10x Genomics B.V., Leiden, The Netherlands). Aliquots of the final sequencing libraries were pooled and sent to Macrogen Inc. for sequencing. **Single-cell data preprocessing and alignment.** Cell Ranger was used to align reads and quantify counts based on the reference genome “refdata-gex-mm10-2020-A” (10x Genomics) [37]. Spliced and unspliced reads were quantified with kallisto and bustools programs [38]. The final count matrices served as input for downstream analysis. **Quality control metrics** of count matrices were computed with SCANPY (version 1.9.1) [39]. Cells with fewer than 500 detected genes, more than 40,000 total UMI counts, or greater than 4% mitochondrial counts were discarded. **Normalization, dimensionality reduction, and clustering** Gene expression counts in each cell were log-normalized and scaled by a factor of 10,000 [40]. Principal component analysis (PCA) was performed using the top 5000 highly variable genes. The top 60 principal components were used for neighborhood graph construction. Clustering was performed using the Leiden algorithm [41], with a resolution parameter of 0.5. Uniform manifold approximation and projection (UMAP) [42] was employed for visualization. Highly expressed genes were identified with Wilcoxon rank-sum test. **Differential expression and pathway analysis** DEGs were identified using the “rank_genes_groups” function. Genes with an adjusted *p* value < 0.05 and log₂ fold change ≥ 1 were marked significant. Enrichment analysis was performed using MSigDB Hallmark gene sets and “decoupleR” (version 1.6.0) [43]. **Trajectory inference** Single-cell pseudotime trajectory analysis was done with scvelo (version 0.2.4) [44–46]. Initial and terminal cell states were estimated with CellRank 2 (version 1.5.1) [47]. Partition-based graph abstraction (PAGA) was used to establish connectivities among clusters [48]. **SCENIC** Single-cell regulatory network inference and clustering were performed with pySCENIC (version 0.12.1) [49]. **Consensus Molecular Subtype (CMS) classification** MmCMS (version 0.1.0) [50–52] was used for consensus molecular subtyping of pseudobulk data, which were obtained by pooling scRNA-seq clusters representing the same TGF- β 1 treatment times. **Leader cell analysis** For this, average expression of a custom list of 32 leader cell markers (Supplementary Table 5) was scored for each cell with the SCANPY function “tl.score_genes”.



Protein isolation and western blotting

Cytosolic and nuclear extracts were prepared and used for protein detection by Western blotting with antibodies listed in Supplementary Table 3 as described [10].

Analysis of TCGA colorectal cancer data

Gene expression data and survival information were retrieved from the UCSC Xena browser (accession date: 23 Jan 2024) [53]. Kaplan–Meier plots were generated using UCSC Xena browser tools. For CMS classification,

CMScaller was used [52]. Correlated gene expression and gene expression levels in CMS-stratified samples were computed and plotted by Matplotlib (version 3.8.4) [54] and Seaborn (Version 0.13.2) [55]. Statistical significance was analyzed by Mann–Whitney *U*-tests.

Statistics and software

Data were processed and visualized using RStudio [56]. Targeted gene expression experiments were analyzed as before [10]. When comparing two populations, statistical significance was assessed using the two-tailed

Fig. 1 **Detection of a broad spectrum of transitory states and a pEMT endpoint in TGF- β 1-treated TKA-organoids.** **A** Schematic overview of the scRNA-seq workflow. The figure was generated with BioRender.com. **B** Unsupervised clustering and UMAP visualization of 931-TKA organoid cells upon TGF- β 1 treatment for 0, 24, 48, and 72 h and scRNA-seq. **C** Individual organoid cells in the UMAP were annotated according to TGF- β 1 treatment times deduced by 10x Genomics sample indices. The color code represents TGF- β 1 treatment times. RNA velocity analyses (**D**) and PAGA-based trajectory prediction (**E**) of scRNA-seq results from 931-TKA organoids. The arrows in the RNA velocity plot predict the starting point of cell states and directions of future progression projected onto the UMAP space. The thickness of the lines denotes the strength of connectivity between clusters. **F** Heatmap showing the MSigDB hallmark gene sets enriched in gene expression profiles of organoid cell clusters. The red-blue color scale represents the z-score. **G** TGF- β pathway activity was evaluated by MSigDB with decoupleR. Enrichment was determined by over-representation analysis (ORA) using the SCANPY function: `score_gene` function and visualized in the UMAP. **H** Expression of an EMT signature was scored by MSigDB with decoupleR and projected onto the UMAP space. **I** UMAP cluster-specific visualization of EMT scores in violin plots. **J** UMAPs visualizing expression levels of epithelial and mesenchymal marker genes in organoid cells treated with TGF- β 1 for 0, 24, 48, and 72 h. The color code represents the normalized counts (Norm. counts) of the indicated genes.

ANOVA test with a confidence interval of 95%. Box plots were generated with ggplot2 [57]. Figures were assembled using Canvas X 2017 (Canvas GFX, Inc.). Schemes depicting experimental designs were generated with BioRender (<https://biorender.com/>).

Additional information about experimental procedures can be found in the Supplementary Information.

RESULTS

TGF- β 1 induces a broad spectrum of transition states and a pEMT endpoint in TKA-organoid cells

TGF- β 1-treated TKA-organoids form cohesive cell sheets that spread into the surrounding extracellular matrix (Supplementary Fig. 2A) [10]. Concomitantly, they upregulate mesenchymal markers but seemingly maintain epithelial gene expression (Supplementary Fig. 2B, C). Yet, immunofluorescence visualization of TGF- β pathway effectors, a cell polarity marker, and a mesenchymal marker demonstrated some intra-organoid cellular heterogeneity (Supplementary Fig. 2D, E), which might obscure the extent and type of pEMT in bulk analyses. To capture the full range of cellular heterogeneity and gauge the expression of epithelial and mesenchymal markers in individual cells, we performed scRNA-seq with organoids from line 931-TKA [10] which had been exposed to TGF- β 1 for different periods of time (Fig. 1A). Following initial processing and quality control of raw sequencing data, a total of 6863 cells allocated to the different treatment groups were available for further examination (Supplementary figure 3). After dimensionality reduction and visualization by UMAP, nine clusters of cells with similar gene expression profiles emerged (Fig. 1B). Comparison with TGF- β 1 treatment regimens showed that cells from each treatment group occupied multiple clusters, revealing even greater cellular heterogeneity than anticipated (Fig. 1B, C). Assessment of cell state transitions and trajectory inference using RNA velocity and PAGA identified TGF- β 1 treatment-naïve cluster 3 cells as origin of all other cell states (Fig. 1D, E and Supplementary Fig. 4A). Cluster 3 cells are distinguished by enrichment of transcriptomic features associated with cell proliferation and fatty acid metabolism (Fig. 1F), which does not apply to cluster 6 and 8 cells although most of these likewise had not been exposed to TGF- β 1. Further, cluster 6 and 8 cells represent developmental lineages distinct from the main trajectory (Fig. 1E, Supplementary Fig. 4A). Along the main trajectory, cells from the 24 h TGF- β 1 treatment group populated clusters 5, 7, 4, and 0, while the majority of cells from the 48 h and 72 h TGF- β 1 treatment groups were equally partitioned among clusters 2 and 1 (Fig. 1B–E). Clusters 5, 7, 4, and 0 exhibited a gradual decrease in expression of gene sets linked to cell proliferation, while enrichment of gene sets related to TGF- β and other signaling pathways continually increased (Fig. 1F). Enrichment of the TGF- β signaling gene set peaked in cluster 2, which is consistent with UMAP projections of TGF- β pathway activity and TGF- β target gene expression (Fig. 1F, G and Supplementary Fig. 4B). Expression of an EMT-related gene signature followed the same cluster 3→1 trajectory with

maximum expression in cluster 2 cells and a subset of cluster 0 cells which might represent precursors of the former (Fig. 1H, I). EMT dynamics of TKA-organoids were paralleled by a gradual transition of their transcriptomes from a profile resembling CMS2 of human CRC to that of CMS4 [58] (Supplementary Fig. 4C). Importantly, though, we could not identify a cell cluster which showed the complementary loss of epithelial and gain of mesenchymal markers commonly observed in prototypical EMT. Rather, UMAP projection of individual gene expression patterns proved co-expression of epithelial and mesenchymal markers at the single cell level, also in cluster 2 cells (Fig. 1J and Supplementary Fig. 4D). In fact, expression of epithelial markers even increased in TGF- β 1-treated cells. Altogether, scRNA-seq revealed that TKA-organoid cells, when stimulated with TGF- β 1, passed through a continuum of transition states and advanced to a pEMT endpoint reminiscent of non-prototypical pEMT observed in human cancer.

A leader cell gene expression profile in TKA-organoid cells with the most advanced pEMT state

TGF- β 1-treated TKA-organoids exhibit a collective mode of invasion in which cells typically segregate into leader and follower populations. Accordingly, scRNA-seq provided an opportunity to examine how the diverse pEMT states adopted by TGF- β 1-treated TKA-organoid cells aligned with phenotypically distinct subgroups of collectively invading cell cohorts. To this end, we created a custom gene signature characteristic of leader cells (Supplementary Table 5) and scored its expression in our scRNA-seq dataset. Notably, acquisition of the leader cell gene expression profile occurred progressively and mirrored the dynamics of TGF- β pathway activity and EMT signature expression, with maximum representation in cluster 2 (Fig. 2A, B). This was also evident when transcript levels of individual genes implicated in leader cell formation were examined (Fig. 2C) [4, 59–62]. Moreover, expression of fibronectin, a mesenchymal marker as well as component of the leader cell signature, peaked in the most peripheral cells of TGF- β 1-treated TKA organoids, which are likely to represent the leader cell population (Supplementary Fig. 2E). This hints that congruent expression dynamics of the EMT and leader cell signatures are not merely a temporal coincidence and supports the idea that a gradual gain of mesenchymal characteristics is paralleled by leader-follower cell fate decisions whereby the most advanced pEMT state (pEMT^{high}) appeared to be associated with leader cells.

High level Sox11 expression and regulon activity in pEMT^{high} cells located at the invasive front of TKA-organoids

Loss of individual core EMT-TFs does not abrogate TKA-organoid pEMT [10]. To investigate which other TFs might take over their function, we searched for differentially expressed TFs in a previously collected bulk RNA-seq dataset [10], which we expected to provide more complete representation and greater sensitivity than scRNA-seq. Downregulated TF genes included the

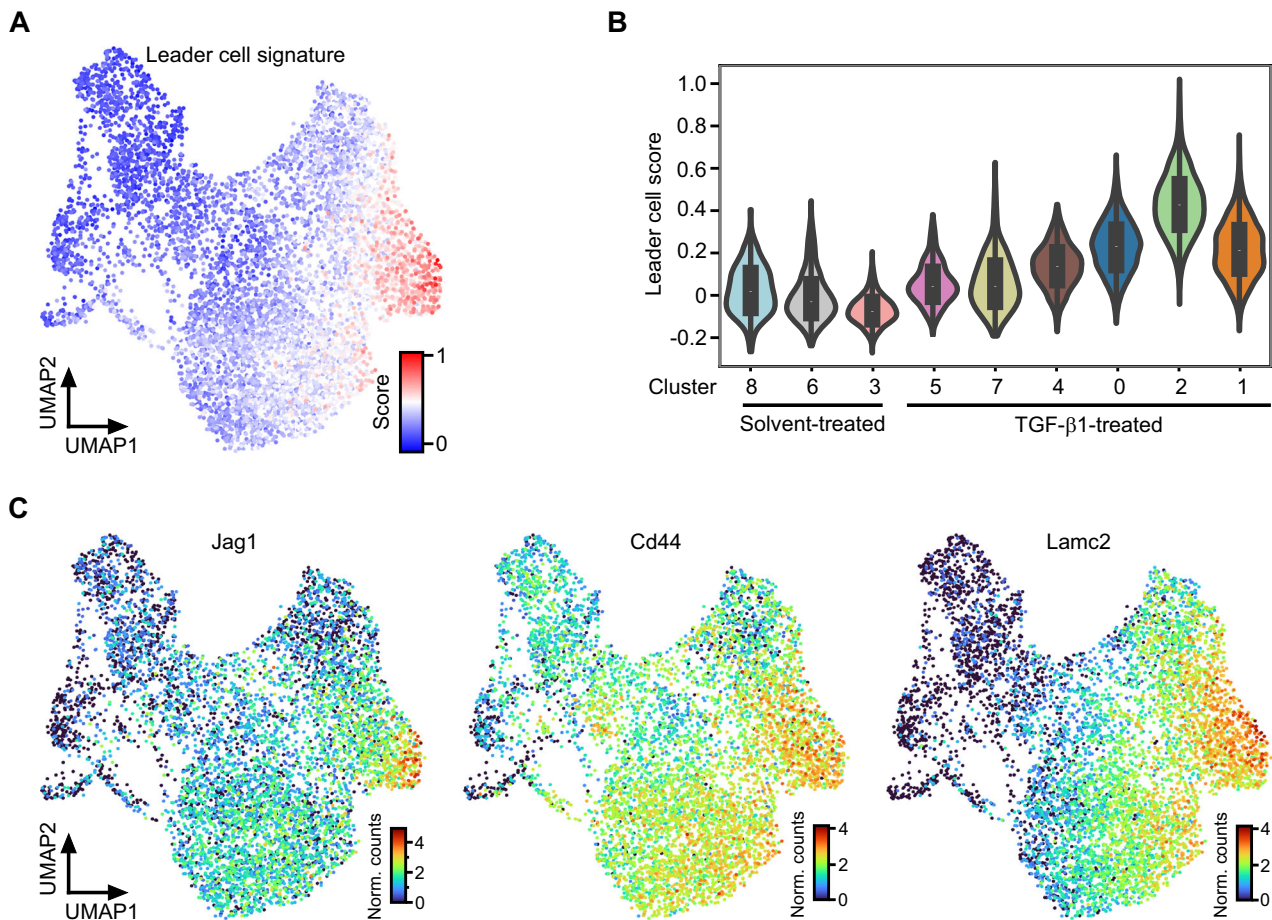
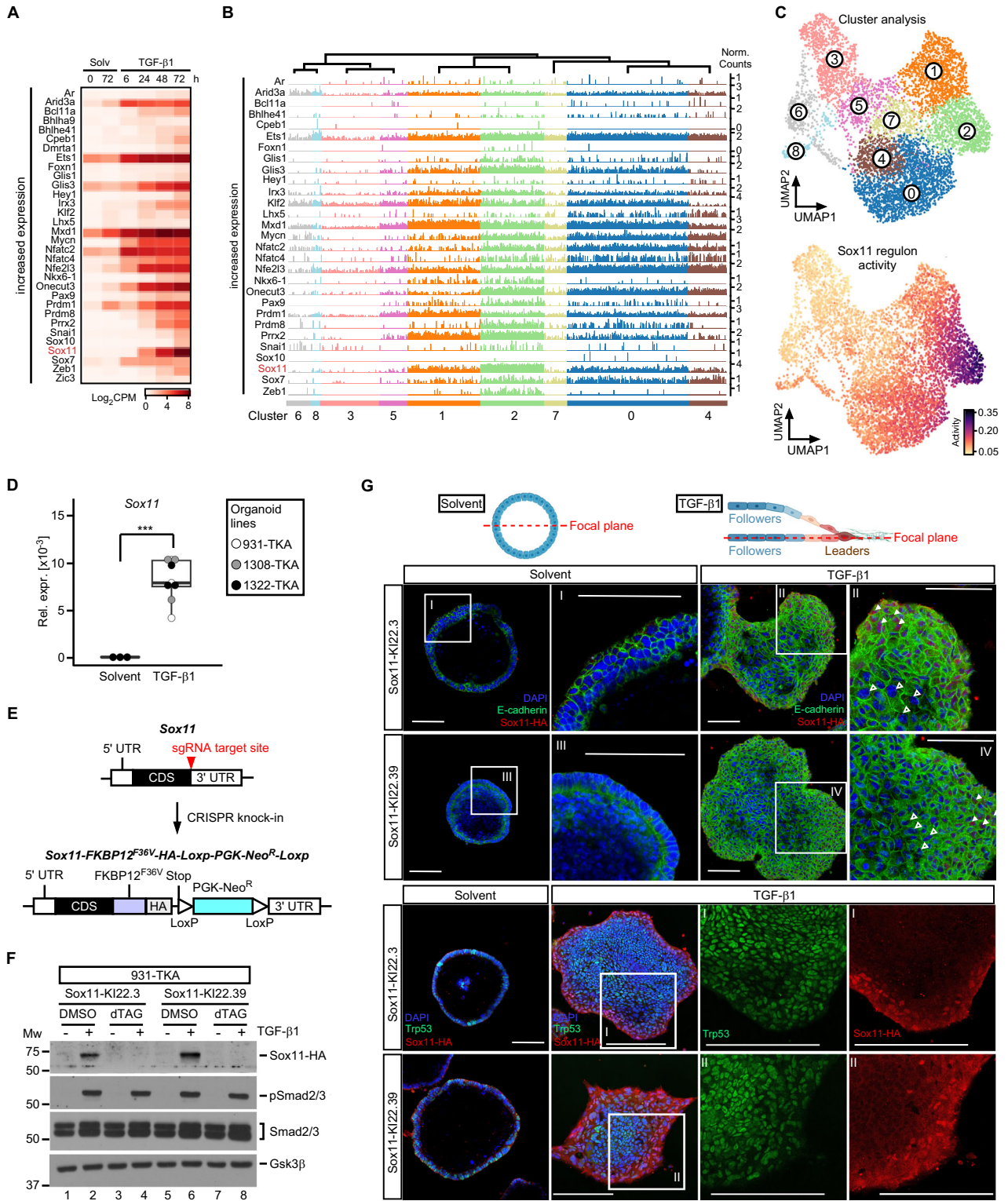


Fig. 2 Cells in the most advanced pEMT state express features of leader cells. **A** Expression of a custom leader cell gene expression profile was scored with SCANPY. The obtained leader cell scores were projected onto the UMAP. **B** UMAP cluster-specific visualization of leader cell scores in violin plots. **C** Expression of exemplary leader cell genes visualized in the UMAP space. The color scales represent the normalized counts (Norm. counts) of the indicated genes.

intestinal stem cell factor *Ascl2*, and the *E2f2* and *Myb* cell cycle regulators (Supplementary Fig. 5A) [63–65]. Diminished expression of these TFs is consistent with the loss of self-renewal capacity and proliferative potential of TGF- β 1-treated TKA-organoids (see above; [10]. Among the TF genes upregulated were core-EMT-TFs and several others reported to promote [66–70] or antagonize EMT [71–73] (Fig. 3A). Due to the sparsity of data, expression of some of the TFs could not be followed in scRNA-seq trackplots (Fig. 3B, Supplementary Fig. 5B). Among the TFs for which the trackplots were informative, *Sox11* stood out because *Sox11* transcripts showed the highest fold change in the bulk RNA-seq data, and were particularly abundant in scRNA-seq cluster 2 cells (Fig. 3A, B, and Supplementary Fig. 5C). Yet, since TF expression does not necessarily coincide with function, we additionally assessed TF activity across UMAP clusters by the SCENIC workflow [49]. Supporting validity of this approach, activity of regulons associated with E2f cell cycle regulators was found to be high in cluster 3 (Supplementary Fig. 5D), which had independently been identified as proliferative cell cluster by gene set enrichment analyses (see Fig. 1F). Strikingly, SCENIC assigned the highest activity score in cluster 2 to the *Sox11* regulon (Fig. 3C and Supplementary Fig. 5D). Therefore, expression and activity profile marked *Sox11* as a TF gene likely to be critical for the pEMT^{high} state of TGF- β 1 stimulated TKA-organoids.

To extend the findings from transcriptome analyses we confirmed that TGF- β 1 triggered *Sox11* expression in additional TKA-organoid lines (Fig. 3D). In the absence of commercially

available antibodies suitable for *Sox11* protein analyses we inserted a FKBP12^{F36V} cassette with a dual hemagglutinin (HA) epitope tag [74] into the *Sox11* gene (Fig. 3E). This enabled to monitor expression of the resulting *Sox11*-FKBP12^{F36V}-HA (*Sox11*-HA) fusion by high quality antibodies while dTAG^{V-1}-induced degradation of *Sox11*-HA allowed to control specificity [74, 75]. Indeed, TGF- β pathway activation, demonstrated by Smad2/3 phosphorylation, led to the appearance of an anti-HA-reactive and dTAG^{V-1}-sensitive band with the expected molecular weight in western blotting experiments (Fig. 3F). Hence, we proceeded to analyze *Sox11*-HA spatial distribution by immunofluorescence staining. While solvent-treated controls were uniformly negative, we detected nuclear *Sox11*-HA in the three to four outermost rows of cells of TGF- β 1-treated TKA-organoids (Fig. 3G). The *Sox11*-HA staining pattern was highly specific, as shown by the disappearance of the signal upon addition of dTAG^{V-1} (Supplementary Fig. 5E) and by comparison with Trp53, which was expressed in all nuclei of TKA-organoids (Fig. 3G). Staining artefacts could be further excluded because a Prrx2-HA fusion protein, also created by knock-in of the FKBP12^{F36V}-HA cassette, exhibited an intra-organoid distribution different from *Sox11*-HA (Supplementary Fig. 5E). Altogether, high-level *Sox11*-HA expression in a narrow band of nuclei at the organoid periphery confirmed that pEMT^{high} cells with a leader cell gene expression profile reside at the invasive fronts of TGF- β 1-treated TKA-organoids.



Sox11 is required for TGF-β1-induced collective invasion and pEMT

Next, we utilized a doxycycline (Dox-)inducible expression system to test whether overexpression of Sox11 carrying an HA-epitope tag (Sox11^{HA}), but not the FKBP12^{F36V} domain, was sufficient to elicit pEMT and collective invasion (Supplementary Fig. 6A). However, Sox11^{HA}-expressing TKA-organoids maintained the same cystic shapes as control organoids (Supplementary

Fig. 6B, C). Likewise, Sox11^{HA} overexpression neither enhanced organoid invasiveness by itself nor did it alter TGF-β1-induced invasion (Supplementary Fig. 6D–F). Consistent with this, expression of epithelial and mesenchymal markers did not change upon Dox stimulation (Supplementary Fig. 6G). Thus, Sox11^{HA} overexpression did not recapitulate TGF-β1-mediated phenotypic alterations. Nonetheless, Sox11 might still be necessary for pEMT of TKA-organoids. To investigate this, we knocked out Sox11 in

Fig. 3 Sox11 is the most highly expressed TGF- β 1-inducible transcription factor in cluster 2 cells and localizes to a layer of cells at the invasive front of TKA-organoids. **A** Heatmap displaying TF genes with increasing expression in combined, time-resolved bulk RNA-seq data from solvent (Solv) or TGF- β 1-treated 931-TKA and 947-TKA organoids [10]. All murine TFs with \log_2 fold changes > 4 at least at one time point of TGF- β 1 stimulation were selected for display. Samples treated with TGF- β 1 for 6 h and 24 h were compared to samples harvested at the beginning of the treatment (solvent 0 h). Samples treated with TGF- β 1 for 48 h and 72 h were compared to samples kept in solvent for 72 h. Normalized \log_2 CPM values were used as gene expression units. **B** Trackplots showing cluster-wise expression of the same TF genes as in (A) in scRNA-seq data from solvent or TGF- β 1-treated 931-TKA organoid cells. Each peak represents the normalized count for a given gene in a single organoid cell. **C** Unsupervised cell clustering and UMAP visualization of 931-TKA organoid cells upon TGF- β 1 treatment for 0, 24, 48, and 72 h (top) and projection of SCENIC-predicted Sox11 regulon activity onto the UMAP (bottom). **D** Expression of *Sox11* in TKA-organoid lines (931-TKA, 1308-TKA, and 1322-TKA). Organoids were embedded in 3 mg/ml Matrigel and treated with solvent or TGF- β 1 for 72 h, followed by RNA collection and cDNA synthesis. RNA levels of *Sox11* were evaluated by qRT-PCR and normalized by the expression levels of *Eef1a1*. The box plots show the 26th to 75th percentiles of the data and the median. Dots represent results of individual experiments ($n = 3$). One-Way ANOVA, ***: p value < 0.001 . **E** Scheme depicting the strategy for CRISPR-mediated in-frame knock-in (KI) of FKBP12^{F36V} coding sequences and the influenza virus hemagglutinin (HA)-epitope tag into the *Sox11* gene. The upper and lower schemes represent WT and modified *Sox11*, respectively. The red arrowhead indicates the sgRNA target site. FKBP12^{F36V} served as a degradation tag. The PGK-promoter driven neomycin resistance gene (Neo^R) flanked by two loxP sites was used as selection marker. **F** Sox11 expression in Sox11-KI clones. Sox11-KI clones KI22.3 and KI22.39 were treated with solvent or TGF- β 1 for 72 h. Two hours before harvest organoids additionally received DMSO or 500 nM dTAG^{V-1}. Organoids were collected for western blot analysis to monitor expression of HA-tagged Sox11, phospho-Smad2/3 (pSmad2/3), and total Smad2/3. Gsk3 β was detected to control equal loading. Representative results from one out of three experiments are shown ($n = 3$). MW: molecular weight standard in kDa. **G** 931-TKA Sox11-KI organoids were seeded in 3 mg/ml Matrigel and treated with solvent or TGF- β 1 for 72 h, followed by whole-mount immunofluorescence staining of the antigens indicated and confocal microscopy. Nuclei were counterstained with DAPI. Positions of focal planes examined are shown in the schemes above the micrographs. Boxed areas with Roman numerals are shown at higher magnification. White arrowheads indicate nuclear Sox11-HA staining at organoid invasive fronts; empty arrowheads indicate absence of Sox11-HA staining in the organoid center. Scale bars: 100 μ m. Representative images from one out of three experiments are shown ($n = 3$). The schemes were generated with BioRender.com.

three different TKA-organoid lines by deleting most of the translated region, including the part encoding the HMG-box (Fig. 4A, Supplementary Fig. 7A, B). This deletion abrogated TGF- β 1-inducible upregulation of *Sox11* except in one case, although the mutant mRNA of this clone would not be able to give rise to functional protein (Supplementary Fig. 7C). Using for comparison parental lines and *Sox11* wild-type (WT) organoid clones that had been exposed to non-targeting sgRNAs, we examined the TGF- β 1 response of the *Sox11* KO organoids. While we observed some line-specific differences in severity of the defects, *Sox11* KO organoids did not flatten and spread as much as control organoids, and in some cases showed signs of disintegration (Fig. 4B and Supplementary Fig. 8A). *Sox11* inactivation also impaired invasive capacities (Fig. 4C, D and Supplementary Fig. 8B, C), and diminished TGF- β 1-mediated upregulation of epithelial and mesenchymal markers (Fig. 4E, F and Supplementary Fig. 8D, E). Accordingly, while it did not qualify as an autonomous EMT inducer, Sox11 proved to be a necessary effector of TGF- β signaling to promote pEMT and collective invasion of TKA-organoids.

Dual impact of Sox11 on basal and TGF- β 1-stimulated gene expression of TKA-organoids

To comprehensively characterize the *Sox11*-dependent transcriptome in TKA-organoids, we performed RNA-seq (Fig. 5A). PCA showed that corresponding samples from two independent replicates clustered together, demonstrating experimental reproducibility (Fig. 5B). Separation of samples along PC1 most likely reflects the TGF- β 1 response, while separation of *Sox11* WT and KO samples along PC2 hinted at additional, TGF- β 1-independent impact of *Sox11* on gene expression. Overall, *Sox11* KO resulted in aberrantly reduced, as well as increased gene expression, arguing that Sox11 acts both as transcriptional activator and repressor in TKA-organoids (Supplementary Fig. 9A). Inter-organoid line comparison additionally demonstrated that TGF- β 1-mediated gene expression changes were highly comparable in *Sox11* WT 931-TKA and 1308-TKA organoids whereas transcriptomes of *Sox11*-deficient cells diverged more between the two organoid backgrounds (Supplementary Fig. 9A).

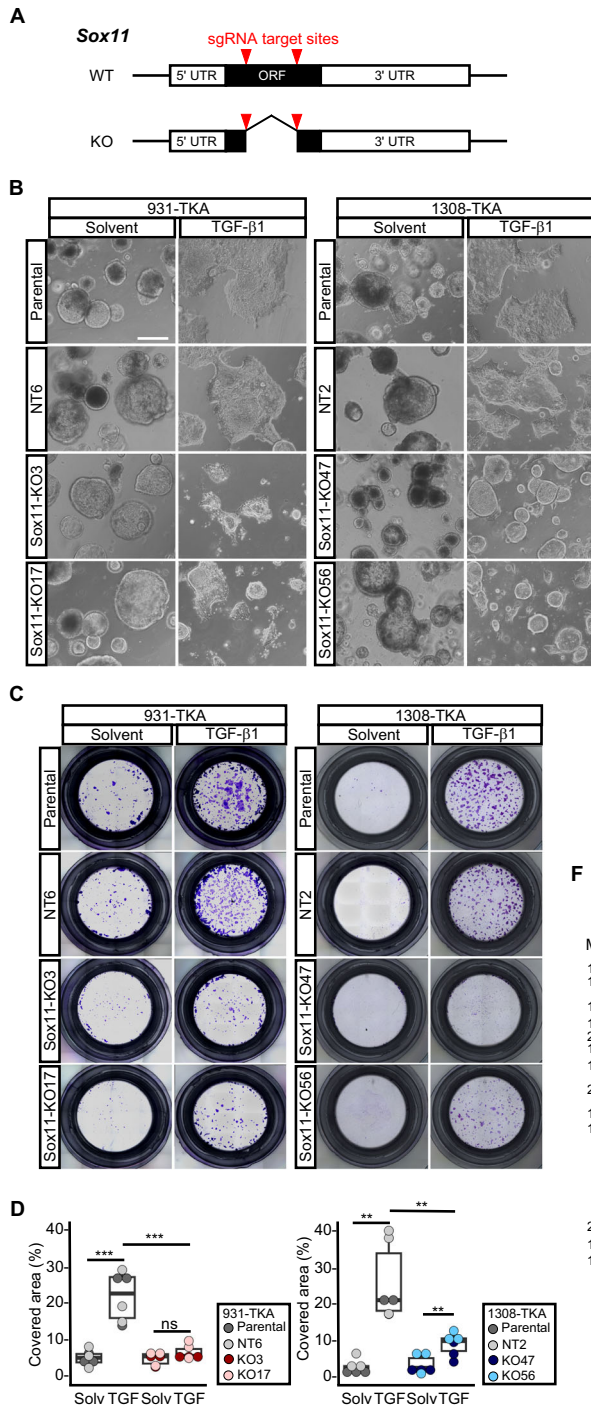
To further examine *Sox11*-regulated transcription, we collected all DEGs in 931-TKA and 1308-TKA organoids and displayed them in a heatmap (Fig. 5C and Supplementary Table 6). This again

revealed similarities but also differences (clusters 3, 6, 7, 10) among *Sox11* WT and KO transcriptional profiles of 931-TKA and 1308-TKA organoids (Fig. 5C). Focusing on the TGF- β 1 response, clusters 8 and 9 comprised genes with diminished expression in TGF- β 1-treated *Sox11*-mutant TKA-organoids. Among these were several genes associated with EMT, TGF- β and Pdgf signaling, actin-myosin filament dynamics, and components of the leader cell signature, some of which are also top-ranking members of the *Sox11* regulon. Pathway analyses likewise indicated that gene sets related to EMT, embryonic development, as well as interaction with and organization of extracellular matrix were the most highly enriched among TGF- β 1-regulated genes with diminished expression in *Sox11* KO organoids (Fig. 5D, Supplementary Fig. 10, and Supplementary Table 7). Moreover, bulk RNA-seq cluster 8 contained a sizeable fraction of genes showing comparatively higher expression in scRNA-seq cluster 2 cells (Supplementary Fig. 9B, C). All of this aligns well with the idea of *Sox11* being a key regulator of TGF- β 1-induced pEMT and collective invasion of TKA-organoids.

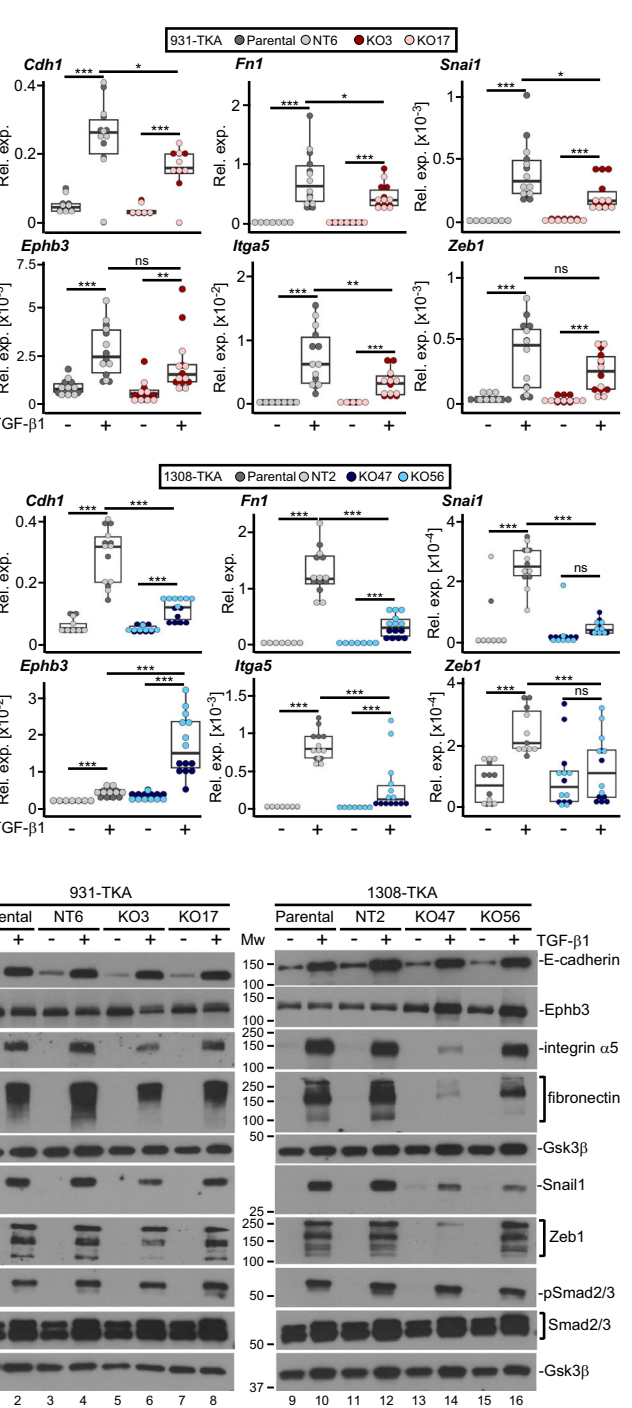
When inspecting *Sox11*-dependent gene expression in TGF- β 1-naïve organoids (clusters 1, 2, 4, 5), *Sox11* deficiency was found to diminish expression of several tumor suppressors and genes involved in intestinal function and metabolism, whereas expression of genes related to hypoxia, and interferon signaling was elevated (Fig. 5C, D, Supplementary Fig. 10; Supplementary Table 7) [76–79]. Importantly, *Sox11*-dependent expression of a selection of genes from bulk RNA-seq clusters 1, 2, and 8, and of the tumor suppressor *Gata5*, a paralog of *Gata4* [76], had already been proven by targeted by qRT-PCR analyses (Fig. 4 and Supplementary Fig. 8), or was newly confirmed (Supplementary Fig. 11). In sum, the bulk RNA-seq analyses suggested that *Sox11* might play a hitherto unknown role in intestinal tissue homeostasis and undergoes a TGF- β 1-induced switch from potential tumor suppressor to tumor-promoting factor.

Elevated SOX11 expression correlates with reduced survival of CRC patients

To determine whether the findings from TKA-organoids were relevant for human disease, we retrieved gene expression data from healthy tissue and CRC samples from The Cancer Genome Atlas cohorts (TCGA). Although SOX11 expression in unstratified tumor samples was not significantly higher than in normal tissue,



SOX11 expression steadily increased during tumor progression (Fig. 6A, B). Higher *SOX11* expression was paralleled by reduced overall survival and shorter disease-free intervals (Fig. 6C). Notably, human orthologs of several genes implicated in EMT, TGF- β , and PDGF signaling, which had shown *Sox11*-dependent expression in TKA-organoids, exhibited highly correlated expression with *SOX11* in TCGA samples (Fig. 6D), indicative of some regulatory relationship among these genes also in human cancer. Finally, we examined expression of *SOX11* and the human counterparts of a selection of *Sox11*-dependent genes in colon and rectum adenocarcinoma samples upon CMS stratification [58]. Indeed, especially when compared to CMS2 samples, expression of *SOX11* and several other genes was elevated in CMS4 (Fig. 6E and



Supplementary Fig. 12), the subtype with evidence for TGF- β 1 pathway activation and mesenchymal characteristics [58], which is consistent with TGF- β 1-mediated upregulation of their mouse orthologs in TKA-organoids, and the TGF- β 1-induced CMS2 to CMS4 transition of TKA-organoid transcriptional profiles. In summary, the results of these analyses strongly argue in favor of a disease-promoting role for *SOX11* in human CRC.

DISCUSSION

While cellular plasticity and pEMT provide attractive conceptual frameworks to understand cancer cell metastasis, mechanistic aspects concerning the induction and implementation of pEMT in

Fig. 4 Sox11 is necessary for TGF- β 1 induced pEMT and collective invasion. **A** Structures of *Sox11* WT and KO loci. The red arrowheads mark the target sites for two sgRNAs used to delete a large part of the *Sox11* open reading frame (ORF). UTR: untranslated region. **B** Whole-mount phase contrast microscopy of 931-TKA and 1308-TKA organoid lines treated with TGF- β 1 for 72 h. The parental lines and clonally derived organoids treated with non-targeting (NT) sgRNAs (931-TKA: NT6; 1308-TKA: NT2) have WT *Sox11* genes. KO lines carry biallelic deletions in *Sox11* (931-TKA: KO3 and KO17; 1308-TKA: KO47 and KO56). Scale bar: 100 μ m. Representative pictures from one of three independent biological replicates are shown. **C** Representative images of transwell invasion assays with *Sox11* WT and mutant organoid lines as indicated. Organoids were seeded in 3 mg/ml Matrigel in the upper chambers of transwell inserts and treated with TGF- β 1 for 72 h. Invasive organoid cells that had passed through the Matrigel layer and crossed the transwell bottom membrane were visualized by crystal violet staining. **D** Quantification of transwell assays with *Sox11* WT and mutant organoid lines. Areas of transwell membranes covered by invaded organoid cells as exemplarily shown in **(C)** were measured by ImageJ ($n = 3$). Statistical significance was assessed by one-way ANOVA, ns: not significant; **: p value < 0.01; ***: p value < 0.001. **E** RNA expression levels of epithelial (*Cdh1*, *Ephb3*) and mesenchymal markers (*Itga5*, *Fn1*, *Snai1*, *Zeb1*) in the indicated *Sox11* WT and mutant organoid lines treated with solvent or TGF- β 1 for 72 h were determined by qRT-PCR and normalized to those of *Eef1a1* ($n = 3$). The box plots show the 26th to 75th percentiles of the data and the median. Dots represent results of individual measurements. Rel. exp.: relative expression. One-way ANOVA; ns: not significant; *: p value < 0.05; **: p value < 0.01; ***: p value < 0.001. **F** Protein expression levels of epithelial markers (E-cadherin, Ephb3) and mesenchymal markers (integrin α 5, fibronectin, Snai1, Zeb1) in the indicated *Sox11* WT and mutant organoid lines treated with solvent or TGF- β 1 for 72 h were determined by western blot analyses. Phosphorylated Smad2/3 (pSmad2/3) and total Smad2/3 amounts were analyzed to show TGF- β pathway activation. Detection of Gsk3 β served to control equal loading. MW: molecular weight standards in kDa. Images are representative examples from one of three independent biological replicates.

carcinogenesis remain unclear. This is because gene expression patterns indicative of pEMT in human cancer cells are highly context-dependent and heterogeneous. Often, classical EMT markers are not among DEGs, and core EMT-TFs appear dispensable [11–14]. The preponderance of pEMT over complete EMT further suggests that progression along the epithelial/mesenchymal continuum is limited in vivo [11–14], whereas in vitro EMT models typically advance to fully mesenchymal states [6, 17, 18, 80]. Clearly, novel model systems are needed which accurately reflect recent findings from cancer transcriptomics and allow to generate mechanistic insights required for making the phenomenon of cancer cell pEMT clinically exploitable. Notably, TKA-organoids precisely replicate several aspects of non-prototypical in vivo instances of pEMT, namely persistent expression of epithelial markers, independence at least from single core EMT-TFs, and incomplete execution of EMT. Accordingly, we think that TKA organoids could represent a versatile experimental alternative to complement existing prototypic EMT models.

Despite a broad spectrum of phenotypic manifestations and striking mechanistic differences, there are some features which appear to be shared by the diversity of prototypic and newly emerging variant EMT programs and which are also displayed by the TKA-organoid model. These commonalities include inducibility by TGF- β signaling, gradual progression through a spectrum of transition states towards more mesenchymal phenotypes [3, 5, 6], attenuated proliferation, and changes in metabolic features [16, 81]. A peculiarity of TGF- β 1-induced pEMT of TKA-organoids, though, concerns expression of epithelial markers, which is largely maintained and in some cases even increased. While this aligns well with results from gene expression profiling across multiple cancer entities [11–14], it constitutes a noteworthy difference to cell lines models of prototypic EMT and mouse models of pancreatic ductal adenocarcinoma, skin squamous cell carcinoma, and mammary carcinoma where exit from fully epithelial states is accompanied by immediate downregulation for example of *Epcam* and *Cdh1* expression [3, 5].

Currently, it remains unknown why TGF- β 1-treated TKA-organoids do not execute a full EMT program and advance to a pEMT state only. Possibly, TGF- β pathway activity does not reach a critical threshold sufficient to enforce completion of EMT due to rapid induction of negative feedback regulators as implied by the temporal profiles of TGF- β -regulated genes. An alternative explanation is suggested by findings that experimental inactivation of core EMT-TFs in pancreatic cancer partly phenocopies TKA-organoid pEMT [6] and that overexpression of *Snai1* and *Zeb1* cannot trigger EMT in the intestinal organoid model [10]. Hence, functionality of core EMT-TFs may be intrinsically impaired in TKA-

organoids. For example, deficiencies in the repressor functions of core EMT-TFs could account for persistent and even elevated expression of epithelial gene expression, including that of epithelial gatekeeper TFs *Foxa1* and *Grhl2* [10, 82, 83], and prevent complete EMT of TGF- β 1-treated TKA-organoids. As an ancillary notion, loss of epithelial gene expression seemingly is not a prerequisite for gain of mesenchymal gene expression, questioning whether double-negative feedback loops are inherent regulatory motifs across all EMT variants [80]. Obviously, epithelial and mesenchymal transcriptional programs can co-exist and appear to be driven by mechanistically separable regulatory pathways.

The prevailing type of stromal infiltration in human cancer is collective invasion, whereby invading cell masses segregate into leader and follower cells [7]. The TKA-organoid model corroborated observations from head and neck cancer, pancreatic, lung, and breast cancer that pEMT and collective invasion are interconnected [4, 6, 8, 9]. It also revealed that pEMT^{high} cells are located at the invasive front of TKA-organoids and possess leader cell features. This matches observations in head and neck cancer [4] but differs from a breast cancer model [84], attesting again to context-dependency and variable impact of distinct pEMT states on cancer cell behavior. Notably, TKA-organoid cells gradually acquired leader cell features while advancing towards the pEMT^{high} state without a clear dichotomy between leader and follower cell transcriptomes. However, comparatively minor transcriptional differences between the two cell populations align very well with leader/follower cell plasticity and may facilitate dynamic exchanges of cellular positions and functions in collectively invading cell cohorts [85].

Sox11 was identified as an essential mediator of TGF- β 1-induced pEMT. Beyond this, *Sox11* strongly influenced gene expression in TGF- β 1 treatment-naïve TKA-organoids. *Sox11* is transiently expressed in the developing mouse gut [86, 87], but expression and functional importance of *Sox11* in postnatal intestine would be a novel finding extending the list of Sox TFs operating in adult intestinal homeostasis [88]. However, we inactivated *Sox11* in oncogenically transformed cells. Therefore, it needs to be clarified whether the observed transcriptional changes in non-stimulated TKA-organoids reflect a regulatory role of *Sox11* in normal gut rather than a function of *Sox11* acquired during tumorigenesis.

The necessity for *Sox11* for pEMT in a model for intestinal carcinogenesis fits very well with the pro-invasive properties of human *SOX11* in head and neck and breast cancer [23–25, 27]. Interestingly, EMT of murine mammary epithelial cells can also be mediated by *Sox4*, another SoxC family member [30]. SoxC TFs are often co-expressed and may act redundantly [86, 87, 89]. Indeed, *Sox4* and *Sox12* are expressed in TKA-organoids, and both genes respond

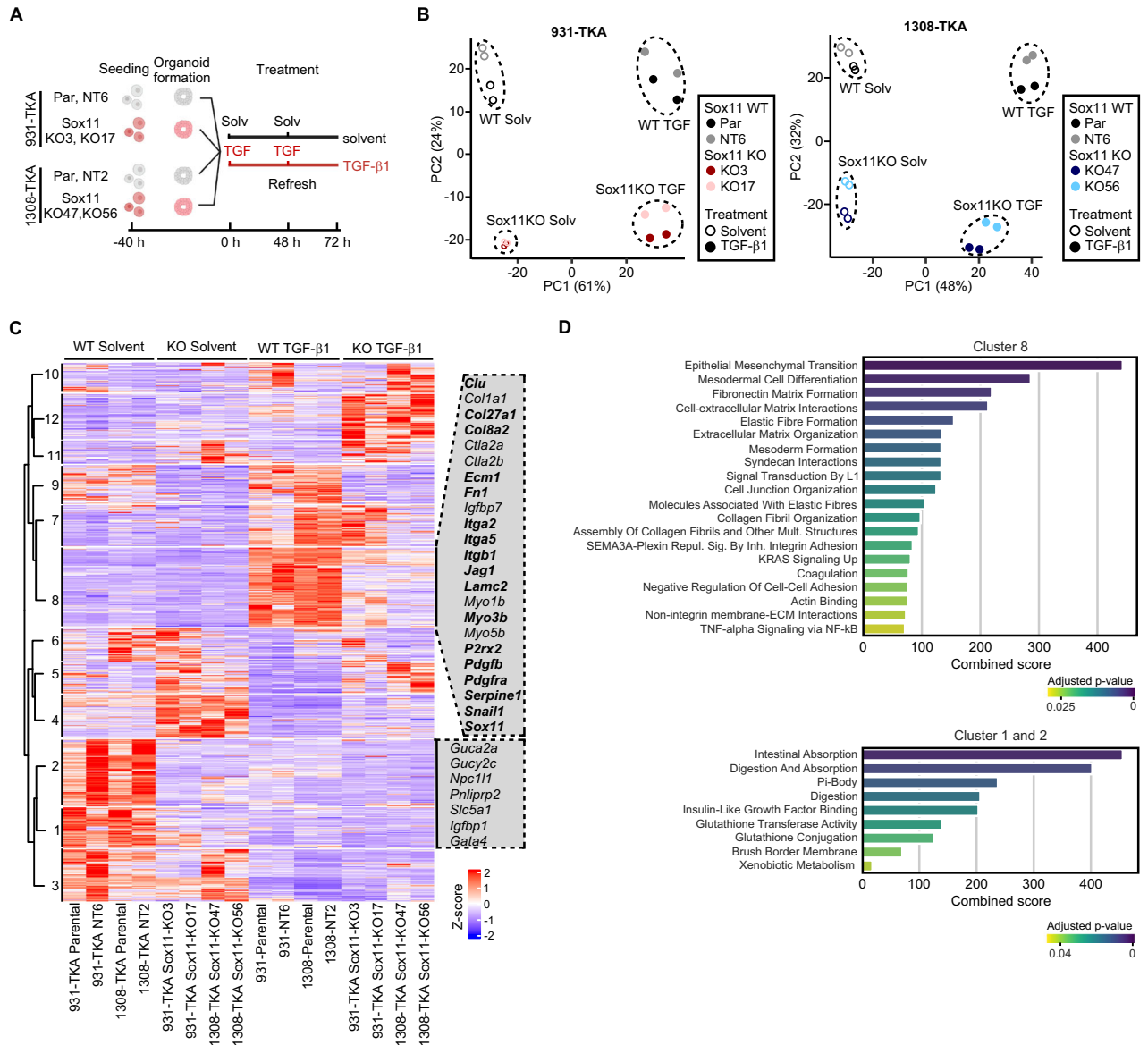


Fig. 5 Bulk RNA-seq reveals multiple functions of Sox11 in resting and TGF- β 1-stimulated organoid cells. **A** Experimental set-up. Sox11 WT and mutant organoid lines, as indicated, were seeded in 3 mg/ml Matrigel and treated with solvent or TGF- β 1 for 72 h, followed by RNA collection and next generation sequencing. The figure was generated with BioRender.com. **B** Principal component analyses of RNA-seq results from the indicated Sox11 WT and mutant organoid lines cultured in conditions as shown. The color codes specify Sox11 genotypes. Empty and filled symbols distinguish samples treated with solvent or TGF- β 1 for 72 h. **C** Heatmap showing differentially expressed genes in the indicated Sox11 WT and mutant organoid lines treated with solvent or TGF- β 1 for 72 h. Genes with absolute values of \log_2 fold changes of expression > 1 and adjusted p values < 0.05 in any condition when comparing Sox11 WT and mutant organoid lines were extracted from the bulk RNA-seq results. Gene expression levels in counts per million (CPM) were then normalized by z-score transformation, sorted with a pre-determined number of 12 clusters, and visualized in the heatmap. Selected genes from clusters 1, 2, and 8 are shown next to the heatmap. Bold type identifies members of the Sox11 regulon with an importance score < 300 . The color code indicates z-scores of the genes. **D** Pathway analyses. Differentially expressed genes from clusters 1, 2, and 8 in (C) were examined for enrichment of gene expression signatures from the MSigDB Hallmark and Reactomes datasets and the GO collections Biological Process, Cellular Component, and Molecular Function. Results of pathway analyses are given as combined scores based on adjusted p -values and odds ratios, which reflect the fraction of genes from a given gene set showing enrichment among DEGs versus the total number of genes in the respective set. Color codes indicate adjusted p values with Fisher's exact test.

to TGF- β 1 stimulation, albeit in opposite directions (Supplementary Fig. 13). This makes it unlikely that Sox12 compensates loss of Sox11. Still, despite clear-cut effects resulting from Sox11 inactivation, we cannot rule out that Sox4 buffers Sox11 deficiency, explaining the somewhat variable Sox11 KO phenotype. Nonetheless, our study corroborates the functional importance of SoxC family members in EMT processes, whereby the cellular background may dictate which of them is of greater relative importance.

Sox4 may also serve as paradigm to mechanistically understand the context-dependency of Sox11 gene regulatory functions. Engagement of Sox4 in alternative transcriptional programs involves Sox4 upregulation and reorganization of TF landscapes by TGF- β signaling in pancreatic and mammary cancer cells [28, 31]. Moreover, Sox4 interacts with Smad3, and both TFs are co-bound to cis-regulatory elements in TGF- β -treated cells [28, 31]. If applicable to Sox11 and intestinal cells, dosage effects and tight

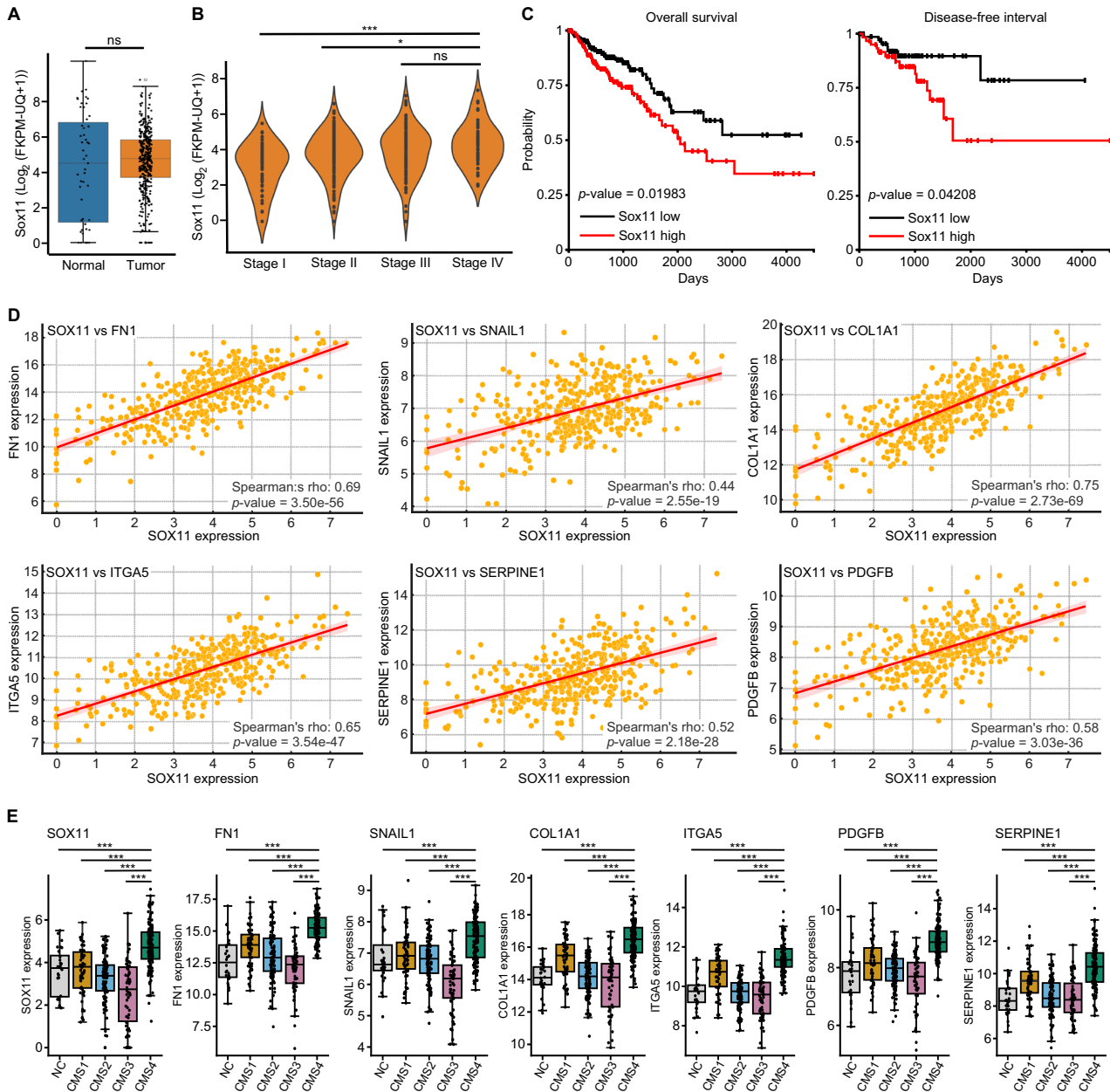


Fig. 6 Higher $SOX11$ expression correlates with shorter survival times in colon and rectum adenocarcinomas. **A** The gene expression matrices of colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ) samples from the TCGA database were retrieved from the UCSC Xena browser, and $SOX11$ RNA expression values in combined COAD and READ cohorts were visualized as box plots. Black dots show $SOX11$ expression in individual samples. The box plots show the 26th to 75th percentiles of the data and the median. Statistical analyses were performed using Mann-Whitney U -test, ns: not significant. **B** COAD and READ gene expression matrices from the TCGA database were stratified according to tumor stages, and $SOX11$ expression levels in combined COAD and READ stage I, II, III, and IV primary tumors are displayed. Black dots show $SOX11$ expression in individual samples. Mann-Whitney U test, ns: not significant; *: p value < 0.05 ; **: p value < 0.01 ; ***: p value < 0.001 . **C** Kaplan-Meier plots showing overall survival and disease-free interval associated with high [$\text{Log}_2(\text{FPKM-UQ} + 1) > 3.842$] and low [$\text{Log}_2(\text{FPKM-UQ} + 1) < 3.842$] $SOX11$ expression in TCGA COAD and READ cohorts. The p values were calculated by the log-rank method. **D** Correlation analyses of the expression of the genes indicated in combined transcriptomes of COAD and READ primary tumors. Each orange dot shows an individual sample. A linear regression line (red line), rank-based Spearman correlation coefficients, and p values are annotated in each plot. The pink shaded regions indicate 95% confidence intervals. Gene expression levels are given as $\text{Log}_2(\text{FPKM-UQ} + 1)$ values. **E** Box plots visualizing expression of the genes indicated in CMS-stratified transcriptomes of COAD and READ primary tumors from the TCGA database. Each black dot shows an individual sample. The box plots show the 26th to 75th percentiles of the data and the median. Statistical analyses were performed using Mann-Whitney U -test; ***: p value < 0.001 . NC not classifiable.

cooperation with TGF- β pathway components would readily explain how Sox11 transcriptional activity might be redirected to promote pEMT. Furthermore, dependency on TGF- β signaling could provide a simple explanation for the inability of Sox11 to autonomously trigger pEMT.

Until now, little is known about $SOX11$ in human CRC. Prior studies ascribed both growth-inhibiting and growth-promoting functions to $SOX11$ [26, 90]. $SOX11$ was also reported to suppress rather than promote CRC cell invasion [21]. Possibly, differences in TGF- β pathway activity and CMS subtype of the experimental

models used account for these conflicting results and should be taken into account when testing *SOX11* function. Nonetheless, we believe that the contribution of *Sox11* to TGF- β 1-induced pEMT and collective invasion in a mouse organoid model for intestinal carcinogenesis reflects cancer-promoting activities of *SOX11* in human colorectal tumors. Future work should aim to substantiate this hypothesis and identify critical targets of *SOX11*, which may provide new perspectives for CRC treatment.

DATA AVAILABILITY

RNA sequencing datasets were deposited in the Gene Expression Omnibus (accession codes GSE278332 and GSE278490). All other data are presented in the article and its associated Supplementary information, are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The code used will be made available upon request.

REFERENCES

- Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol*. 2019;20:69–84.
- Pastushenko I, Blanpain C. EMT transition states during tumor progression and metastasis. *Trends Cell Biol*. 2019;29:212–26.
- Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. *Nature*. 2018;556:463–8.
- Puram SV, Tirosh I, Parkir AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*. 2017;171:1611–24.e24.
- Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell*. 2021;39:1150–62.e9.
- Carstens JL, Yang S, Correa de Sampaio P, Zheng X, Barua S, McAndrews KM, et al. Stabilized epithelial phenotype of cancer cells in primary tumors leads to increased colonization of liver metastasis in pancreatic cancer. *Cell Rep*. 2021;35:108990.
- Friedl P, Locker J, Sahai E, Segall JE. Classifying collective cancer cell invasion. *Nat Cell Biol*. 2012;14:777–83.
- Bronsert P, Enderle-Ammour K, Bader M, Timme S, Kuehs M, Csanadi A, et al. Cancer cell invasion and EMT marker expression: a three-dimensional study of the human cancer-host interface. *J Pathol*. 2014;234:410–22.
- Aiello NM, Maddipati R, Norgard RJ, Balli D, Li J, Yuan S, et al. EMT subtype influences epithelial plasticity and mode of cell migration. *Dev Cell*. 2018;45:681–95.e4.
- Flum M, Dicks S, Teng Y-H, Schrempf M, Nyström A, Boerries M, et al. Canonical TGF β signaling induces collective invasion in colorectal carcinogenesis through a Snail1- and Zeb1-independent partial EMT. *Oncogene*. 2022;41:1492–506.
- Cook DP, Vanderhyden BC. Transcriptional census of epithelial-mesenchymal plasticity in cancer. *Sci Adv*. 2022;8:eabi7640.
- Cook DP, Vanderhyden BC. Context specificity of the EMT transcriptional response. *Nat Commun*. 2020;11:2142.
- Tyler M, Tirosh I. Decoupling epithelial-mesenchymal transitions from stromal profiles by integrative expression analysis. *Nat Commun*. 2021;12:2592.
- Malagoli Tagliacucchi G, Wiecek AJ, Withnell E, Secrier M. Genomic and micro-environmental heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer. *Nat Commun*. 2023;14:789.
- Zhang Y, Donaher JL, Das S, Li X, Reinhardt F, Krall JA, et al. Genome-wide CRISPR screen identifies PRC2 and KMT2D-COMPASS as regulators of distinct EMT trajectories that contribute differentially to metastasis. *Nat Cell Biol*. 2022;24:554–64.
- Antón-García P, Haghghi EB, Rose K, Vladimirov G, Boerries M, Hecht A. TGF β 1-induced EMT in the MCF10A mammary epithelial cell line model is executed independently of SNAIL1 and ZEB1 but relies on JUNB-coordinated transcriptional regulation. *Cancers* 2023;15:558.
- Kröger C, Afeyan A, Mraz J, Eaton EN, Reinhardt F, Khodor YL, et al. Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proc Natl Acad Sci USA*. 2019;116:7353–62.
- McFaline-Figueroa JL, Hill AJ, Qiu X, Jackson D, Shendure J, Trapnell C. A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat Genet*. 2019;51:1389–98.
- Jiang J, Wang Y, Sun M, Luo X, Zhang Z, Wang Y, et al. SOX on tumors, a comfort or a constraint?. *Cell Death Discov*. 2024;10:67.
- Yang Z, Jiang S, Lu C, Ji T, Yang W, Li T, et al. SOX11: friend or foe in tumor prevention and carcinogenesis?. *Ther Adv Med Oncol*. 2019;11:1758835919853449.
- Wu H, Dai Y, Zhang D, Zhang X, He Z, Xie X, et al. LINC00961 inhibits the migration and invasion of colon cancer cells by sponging miR-223-3p and targeting SOX11. *Cancer Med*. 2020;9:2514–23.
- Trekitkarnmongkol W, Shah V, Kai K, Katayama H, Wong J, Ladha FA, et al. Epigenetic activation of SOX11 is associated with recurrence and progression of ductal carcinoma in situ to invasive breast cancer. *Br J Cancer*. 2024;131:171–83.
- Shepherd JH, Uray IP, Mazumdar A, Tsimelzon A, Savage M, Hilsenbeck SG, et al. The SOX11 transcription factor is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression. *Oncotarget*. 2016;7:13106–21.
- Oliemuller E, Newman R, Tsang SM, Foo S, Muirhead G, Noor F, et al. SOX11 promotes epithelial/mesenchymal hybrid state and alters tropism of invasive breast cancer cells. *Elife*. 2020;9:e58374.
- Huang J, Ji EH, Zhao X, Cui L, Misuno K, Guo M, et al. Sox11 promotes head and neck cancer progression via the regulation of SDCCAG8. *J Exp Clin Cancer Res*. 2019;38:138.
- Huang Y, Wang Y, Wu Z, Li T, Li S, Wang C, et al. SOX11-dependent CATSPER1 expression controls colon cancer cell growth through regulation the PI3K/AKT signaling pathway. *Genes Genom*. 2022;44:1415–24.
- Xiao Y, Xie Q, Qin Q, Liang Y, Lin H, de Zeng. Upregulation of SOX11 enhances tamoxifen resistance and promotes epithelial-to-mesenchymal transition via slug in MCF-7 breast cancer cells. *J Cell Physiol*. 2020;235:7295–308.
- David CJ, Huang Y-H, Chen M, Su J, Zou Y, Bardeesy N, et al. TGF- β Tumor Suppression through a Lethal EMT. *Cell*. 2016;164:1015–30.
- Guo W, Keckesova Z, Donaher JL, Shibue T, Tischler V, Reinhardt F, et al. Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell*. 2012;148:1015–28.
- Tiwari N, Tiwari VK, Waldmeier L, Balwierc PJ, Arnold P, Pachkov M, et al. Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming. *Cancer Cell*. 2013;23:768–83.
- Vervoort SJ, Lourenço AR, Tufegdžić Vidaković A, Mocholi E, Sandoval JL, Rueda OM, et al. SOX4 can redirect TGF- β -mediated SMAD3-transcriptional output in a context-dependent manner to promote tumorigenesis. *Nucleic Acids Res*. 2018;46:9578–90.
- Schmid-Burgk JL, Höning K, Ebert TS, Hornung V. CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat Commun*. 2016;7:12338.
- Galaxy Community The. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res*. 2024;52:W83–W94.
- Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847–9.
- Durink S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protoc*. 2009;4:1184–91.
- Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrich. *Curr Protoc*. 2021;1:e90.
- Zheng GXY, Terry JM, Belgrader P, Rykin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017;8:14049.
- Melsted P, Boeshaghi AS, Liu L, Gao F, Lu L, Min KHJ, et al. Modular, efficient and constant-memory single-cell RNA-seq preprocessing. *Nat Biotechnol*. 2021;39:813–8.
- Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*. 2018;19:15.
- Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*. 2015;33:495–502.
- Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep*. 2019;9:5233.
- McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection. *JOSS*. 2018;3:861.
- Badia-I-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv*. 2022;2:vbac016.
- Bergen V, Lange M, Peidl S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol*. 2020;38:1408–14.
- Gayoso A, Weiler P, Lotfollahi M, Klein D, Hong J, Streets A, et al. Deep generative modeling of transcriptional dynamics for RNA velocity analysis in single cells. *Nat Methods*. 2024;21:50–9.
- La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature*. 2018;560:494–8.
- Weiler P, Lange M, Klein M, Pe'er D, Theis FJ. CellRank 2: unified fate mapping in multiview single-cell data. *Nat Methods*. 2024;21:1196–205.

48. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* 2019;20:59.
49. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods.* 2017;14:1083–6.
50. Amirkhah R, Gilroy K, Malla SB, Lannagan TRM, Byrne RM, Fisher NC, et al. MmCMS: mouse models' consensus molecular subtypes of colorectal cancer. *Br J Cancer.* 2023;128:1333–43.
51. Hoshida Y. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. *PLoS ONE.* 2010;5:e15543.
52. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci Rep.* 2017;7:16618.
53. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol.* 2020;38:675–8.
54. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng.* 2007;9:90–5.
55. Waskom M. seaborn: statistical data visualization. *JOSS.* 2021;6:3021.
56. RStudio Team. RStudio: Integrated Development for R. Boston, MA; 2015. Available from: URL: <http://www.rstudio.com/>.
57. Wickham H. ggplot2: elegant graphics for data analysis. 2nd edn. Cham: Springer International Publishing; 2016. Available from: URL: <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=4546676>.
58. Guinney, Dienstmann J, Wang R, Reyniès X, de A, Schlicker A, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med.* 2015;21:1350–6.
59. Sumnerbell ER, Mouw JK, Bell JSK, Knippler CM, Pedro B, Arnst JL, et al. Epigenetically heterogeneous tumor cells direct collective invasion through filopodia-driven fibronectin micropatterning. *Sci Adv.* 2020;6:eaa26197.
60. Gao F, Zhang G, Liu Y, He Y, Sheng Y, Sun X, et al. Activation of CD44 signaling in leader cells induced by tumor-associated macrophages drives collective detachment in luminal breast carcinomas. *Cell Death Dis.* 2022;13:540.
61. Yang C, Cao M, Liu Y, He Y, Du Y, Zhang G, et al. Inducible formation of leader cells driven by CD44 switching gives rise to collective invasion and metastases in luminal breast carcinomas. *Oncogene.* 2019;38:7113–32.
62. Yoon SB, Chen L, Robinson IE, Khatib TO, Arthur RA, Claussen H, et al. Subpopulation commensalism promotes Rac1-dependent invasion of single cells via laminin-332. *J Cell Biol.* 2024;223:e202308080.
63. Schuijers J, Junker JP, Mokry M, Hatzis P, Koo B-K, Sasselli V, et al. Ascl2 acts as an R-spondin/Wnt-responsive switch to control stemness in intestinal crypts. *Cell Stem Cell.* 2015;16:158–70.
64. Beyes S, Andrieux G, Schrempp M, Aicher D, Wenzel J, Antón-García P, et al. Genome-wide mapping of DNA-binding sites identifies stemness-related genes as directly repressed targets of SNAIL1 in colorectal cancer cells. *Oncogene.* 2019;38:6647–61.
65. Gao Y, Qiao X, Liu Z, Zhang W. The role of E2F2 in cancer progression and its value as a therapeutic target. *Front Immunol.* 2024;15:1397303.
66. Dittmer J. The role of the transcription factor Ets1 in carcinoma. *Semin Cancer Biol.* 2015;35:20–38.
67. Li C, Wang Z, Chen Y, Zhou M, Zhang H, Chen R, et al. Transcriptional silencing of ETS-1 abrogates epithelial-mesenchymal transition resulting in reduced motility of pancreatic cancer cells. *Oncol Rep.* 2015;33:559–65.
68. Gawronska-Kozak B, Grabowska A, Kur-Piotrowska A, Kopcewicz M. Foxn1 transcription factor regulates wound healing of skin through promoting epithelial-mesenchymal transition. *PLoS ONE.* 2016;11:e0150635.
69. Ma L, Young J, Prabhala H, Pan E, Messtagh P, Muth D, et al. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nat Cell Biol.* 2010;12:247–56.
70. Juang Y-L, Jeng Y-M, Chen C-L, Lien H-C. PRRX2 as a novel TGF- β -induced factor enhances invasion and migration in mammary epithelial cell and correlates with poor prognosis in breast cancer. *Mol Carcinog.* 2016;55:2247–59.
71. Grudzien-Nogalska E, Reed BC, Rhoads RE. CPEB1 promotes differentiation and suppresses EMT in mammary epithelial cells. *J Cell Sci.* 2014;127:2326–38.
72. Li H-J, Yu P-N, Huang K-Y, Su H-Y, Hsiao T-H, Chang C-P, et al. NKX6.1 functions as a metastatic suppressor through epigenetic regulation of the epithelial-mesenchymal transition. *Oncogene.* 2016;35:2266–78.
73. Chung H-H, Lee C-T, Hu J-M, Chou Y-C, Lin Y-W, Shih Y-L. NKX6.1 Represses tumorigenesis, metastasis, and chemoresistance in colorectal cancer. *Int J Mol Sci.* 2020;21:5106.
74. Stengel KR, Ellis JD, Spielman CL, Bomber ML, Hiebert SW. Definition of a small core transcriptional circuit regulated by AML1-ETO. *Molecular Cell.* 2021;81:530–45.e5.
75. Nabet B, Ferguson FM, Seong BKA, Kuljanin M, Leggett AL, Mohardt ML, et al. Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. *Nat Commun.* 2020;11:4687.
76. Akiyama Y, Watkins N, Suzuki H, Jair K-W, van Engeland M, Esteller M, et al. GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. *Mol Cell Biol.* 2003;23:8429–39.
77. Bashir B, Merlino DJ, Rappaport JA, Gnass E, Palazzo JP, Feng Y, et al. Silencing the GUCY2A-GUCY2C tumor suppressor axis in CIN, serrated, and MSI colorectal neoplasia. *Human Pathol.* 2019;87:103–14.
78. Gyimesi G, Pujol-Giménez J, Kanai Y, Hediger MA. Sodium-coupled glucose transport, the SLC5 family, and therapeutically relevant inhibitors: from molecular discovery to clinical application. *Pflügers Arch Eur J Physiol.* 2020;472:1177–206.
79. Ansari I, Solé-Boldo L, Ridnik M, Gutekunst J, Gilliam O, Korshko M, et al. TET2 and TET3 loss disrupts small intestine differentiation and homeostasis. *Nat Commun.* 2023;14:4005.
80. Zhang J, Tian X-J, Zhang H, Teng Y, Li R, Bai F, et al. TGF- β -induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci Signal.* 2014;7:ra91.
81. Vega S, Morales AV, Ocaña OH, Valdés F, Fabregat I, Nieto MA. Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev.* 2004;18:1131–43.
82. Frisch SM, Farris JC, Pifer PM. Roles of Grainyhead-like transcription factors in cancer. *Oncogene.* 2017;36:6067–73.
83. Jäggle S, Busch H, Freihen V, Beyes S, Schrempp M, Boerries M, et al. SNAIL1-mediated downregulation of FOXA proteins facilitates the inactivation of transcriptional enhancer elements at key epithelial genes in colorectal cancer cells. *PLoS Genet.* 2017;13:e1007109.
84. Westcott JM, Camacho S, Nasir A, Huysman ME, Rahhal R, Dang TT, et al. Δ Np63-Regulated epithelial-to-mesenchymal transition state heterogeneity confers a leader-follower relationship that drives collective invasion. *Cancer Res.* 2020;80:3933–44.
85. Campanale JP, Montell DJ. Who's really in charge: diverse follower cell behaviors in collective cell migration. *Curr Opin Cell Biol.* 2023;81:102160.
86. Dy P, Penzo-Méndez A, Wang H, Pedraza CE, Macklin WB, Lefebvre V. The three Sox proteins-Sox4, Sox11 and Sox12-exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res.* 2008;36:3101–17.
87. Hoser M, Potzner MR, Koch JMC, Bösl MR, Wegner M, Sock E. Sox12 deletion in the mouse reveals nonreciprocal redundancy with the related Sox4 and Sox11 transcription factors. *Mol Cell Biol.* 2008;28:4675–87.
88. Gracz AD, Magness ST. Sry-box (Sox) transcription factors in gastrointestinal physiology and disease. *Am J Physiol Gastrointest Liver Physiol.* 2011;300:G503–15.
89. Miao Q, Hill MC, Chen F, Mo Q, Ku AT, Ramos C, et al. SOX11 and SOX4 drive the reactivation of an embryonic gene program during murine wound repair. *Nat Commun.* 2019;10:4042.
90. Elkhamisy FAA, Eesa AN, Elnady OM, Elnaghi KAEA, Foda AAM. Reduced expression of SOX11 in colorectal adenocarcinoma is associated with mucinous and signet ring cell types, poor survival, and lower ALK expression. *Pathol Res Pr.* 2024;260:155450.

ACKNOWLEDGEMENTS

We are grateful to Kerstin Bruder for help with scRNA-seq sample preparation and staff members of the Life Imaging Center (LIC) at the Center for Biological Systems Analysis (ZBSA), Albert-Ludwigs-University Freiburg, for assistance with confocal microscopy. Financial support for the study was obtained from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (Grant number 322977937/GRK2344 to AH) and the Wilhelm Sander-Stiftung (Grant number 2023.125.1 to AH). MB acknowledges funding from the DFG within CRC1160 (Project ID 256073931-Z02), CRC/TRR167 (Project ID 259373024-Z01), CRC1453 (Project ID 431984000-S1), CRC1479 (Project ID: 441891347- S1), TRR 359 (Project ID 491676693-Z01), TRR 353/1 (Project ID 471011418-SP02), and FOR 5476 UcarE (Project ID 493802833-P7). Additional funding was received from the German Federal Ministry of Education and Research (BMBF) within the Medical Informatics Funding Scheme, PM4Onco-FKZ 01ZZ2322A (MB) and EkoEstMed-FKZ 01ZZ2015 (GA).

AUTHOR CONTRIBUTIONS

YHT: design of experiments, data generation and interpretation, phenotypic characterization of organoids, scRNA-seq and bulk RNA-seq experiments, CRISPR/Cas9 genome editing, figure preparation, manuscript writing; BA and GA: bioinformatic analyses of scRNA-seq results, data visualization and interpretation; MS: data generation for phenotypic characterization of WT and Sox11-deficient organoids, CRISPR/Cas9 genome editing experiments; KR and ASH: CRISPR/Cas9-mediated knock-in, data generation and interpretation; MK: instruction of YHT and implementation of 10x Genomics technology; SB: bioinformatic analyses, visualization and interpretation of Sox11 KO RNA-seq data; MB: bioinformatic analyses of

scRNA-seq data, data interpretation, manuscript writing; AH: design of experiments, data interpretation, figure preparation, manuscript writing. All authors critically read and commented on the final version of the manuscript.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

All methods were performed in accordance with the relevant guidelines and regulations. Mice were handled in compliance with legal directives at the Center for Experimental Models and Transgenic Service of the University of Freiburg Medical Center (project registration number: X-21/02F). No research involving humans was performed in this study.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41389-025-00560-7>.

Correspondence and requests for materials should be addressed to Andreas Hecht.

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025