Tracing tumorigenesis in a solid tumor model at single-cell resolution

Samantha D. Praktiknjo1,6✉, Benedikt Obermayer1,2,6, Qionghua Zhu3,5,6, Liang Fang3,5, Haiyue Liu1, Hazel Quinn3, Marlon Stoeckius4, Christine Kocks1, Walter Birchmeier3✉ & Nikolaus Rajewsky1✉

Characterizing the complex composition of solid tumors is fundamental for understanding tumor initiation, progression and metastasis. While patient-derived samples provide valuable insight, they are heterogeneous on multiple molecular levels, and often originate from advanced tumor stages. Here, we use single-cell transcriptome and epitope profiling together with pathway and lineage analyses to study tumorigenesis from a developmental perspective in a mouse model of salivary gland squamous cell carcinoma. We provide a comprehensive cell atlas and characterize tumor-specific cells. We find that these cells are connected along a reproducible developmental trajectory: initiated in basal cells exhibiting an epithelial-to-mesenchymal transition signature, tumorigenesis proceeds through Wnt-differential cancer stem cell-like subpopulations before differentiating into luminal-like cells. Our work provides unbiased insights into tumor-specific cellular identities in a whole tissue environment, and emphasizes the power of using defined genetic model systems.

1Systems Biology of Gene Regulatory Elements, Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. 2Core Unit Bioinformatics, Berlin Institute of Health, Charité - Universitätsmedizin Berlin, Berlin, Germany. 3Signal Transduction in Development and Cancer, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. 4New York Genome Center, New York, NY, USA. 5Present address: Southern University of Science and Technology, Shenzhen, China. 6These authors contributed equally: Samantha D. Praktiknjo, Benedikt Obermayer, Qionghua Zhu. ✉email: samantha@praktiknjo.com; wbirch@mdc-berlin.de; rajewsky@mdc-berlin.de
Solid tumors represent one of the main causes of morbidity and mortality worldwide. The molecular understanding of what drives carcinogenesis and tumor progression remains elusive though. This is in part due to the great extent of intra- and interfumoral heterogeneity in human tumors which confounds a vast diversity of genetic and epigenetic factors that are subject to constant changes as a result of intrinsic and environmental cues. Additionally, clinical samples are often derived from advanced or mixed tumor stages where information related to the initial induction of the cancer is frequently lost.

Another challenge consists in the limited availability of methods, which allow to disentangle the highly complex composition of the diseased tissues where cancer cells coexist together with tumor-associated and non-tumoral elements of the tumor microenvironment. Advances in single-cell technologies have recently enabled several studies to elucidate the cellular complexity of a given tumor in more detail or to characterize specific tumor subtypes. These findings support hierarchical models of tumor initiation by cancer stem cells (CSCs) that proliferate and differentiate, and induce heterogeneities in cancer cell phenotypes. CSCs are defined by their inherent ability to initiate and drive tumor growth and resistance to conventional treatment strategies. Wnt signaling has been recognized to be a key driver for the initiation and maintenance of CSCs, and it appears that stemness can be reversibly acquired and lost via epigenetic or environmental triggers, such as metabolic reprogramming or epithelial-to-mesenchymal transition (EMT). CSCs have thus become a prime focus for investigating molecular mechanisms that control tumorigenesis and metastasis. Their study, however, has been difficult due to their low frequency, unclear cell surface immunophenotype and other variable biological properties during disease progression.

While emphasizing the great molecular diversity of human cancers, recent single-cell studies also revealed extensive coupling between different molecular levels. For instance, differences in genomic alteration patterns correlated with substantial transcriptional differences in head and neck cancer. In contrast, in gliomas and breast cancer, the transcriptomes of different tumors were found to share similar differentiation or stemness signatures that were independent of genetic lineage relationships. Moreover, complex relationships between tumor expression programs and tissue-of-origin transcriptomic signatures were found, underscoring the need to comprehensively profile not only the tumor and its microenvironment but also histologically normal tumor-adjacent tissues as well as healthy controls in order to gain a full understanding of the complex feedback between tumor and host tissue. Altogether, this accentuates the need to use controlled model systems to obtain a deeper understanding of general mechanisms that are intrinsic to tumorigenesis.

In a previous study, we found that high Wnt/β-catenin and low Bmp signaling were characteristic for aggressive forms of salivary gland and head and neck squamous cell carcinomas (SCCs) in humans. With the aim to better understand these Wnt-specific mechanisms, we thus created a mouse model with K14-cre-driven β-catenin gain-of-function (β-catGOF) and Bmpr1a loss-of-function (Bmpr1aLOF) mutations. We showed that these mice developed very specific salivary gland SCCs within 100 days after birth which contained highly self-renewing Wnt-dependent CD24+CD29+ CSCs which, upon isolation and injection into NOD/SCID mice, produced fast-growing tumors. These CSCs showed high activity of the stem cell-associated SSEA1 marker as well as nuclear β-catenin and Wnt-specific target genes such as Axin2 which were not found in other subpopulations within the tumor.

To gain a more basic understanding of tumorigenesis, we here used single-cell transcriptomics together with our Wnt-dependent double-mutant salivary gland SCC mouse model to systematically study CSCs in a controlled setting in vivo. Our setup enabled us to build a high-resolution salivary gland cell atlas, to dissect tumor heterogeneity in a whole tissue environment and to identify CSC-like cells de novo directly from solid tumor samples. We show that tumor-specific epithelial cells consist of luminal- and basal-like cells as well as a small, but distinct CSC-like population. Further molecular characterization together with pathway and lineage analyses allowed us to infer and reconstruct a robust trajectory of the tumor progression. We found that upon activation of β-catenin gain- and Bmpr1a loss-of-function mutations in basal cells, tumorigenesis is initiated by expression of an EMT signature and proceeds through heterogeneous populations of CSC-like cells driven by differential Wnt signaling, before differentiating into luminal-like cells. Our work reveals several genes and expression patterns that may be fundamental in the regulation of tumorigenesis, and provides a novel and unbiased approach to study CSCs from a developmental perspective.
and analyzed single-cell datasets from control salivary glands using Seurat \(^\text{21}\) (Fig. 1b). After inspection of marker genes, we assigned the cell type identity to clusters, for which the expression of specific genes has been reported in previous literature \(^\text{22–31}\), and molecularly characterized and validated additional cell types for which no or only ambiguous information was available (Fig. 1b, c; Supplementary Figs. 3, 4). In agreement with previous studies \(^\text{32–34}\), we observed that \(\text{Egf}\) and \(\text{Smgc}\) were expressed in cell populations with strongly sex-dependent representation (Supplementary Fig. 3) and identified several other marker genes with similar patterns (Supplementary Fig. 4a). Interestingly, we also noted that \(\text{Dcpp1}\) cells were more abundant in tissues from mice at the P40 than at the P90 stage and validated this finding by immunofluorescence in tissue sections (Supplementary Fig. 3b, c). Altogether, this indicates that we can reliably identify and characterize cell types at high resolution and provide evidence that the ductal composition of healthy salivary glands is sexually dimorphic and stage-dependent.

Our approach combined with imaging-based validations enabled us to chart all epithelial cells onto a consolidated anatomical sketch depicting female- (Fig. 1c) and male-specific (Supplementary Fig. 3e) features.

Identification of cancer stem cells and other tumor-specific cell populations. To systematically uncover cells that were specific to the tumor samples, we analyzed and clustered cells pooled from all single-cell datasets of control and double-mutant samples (Fig. 2a). This combined analysis enabled us to recapitulate all previously identified cell types without the need to resort to advanced sample alignment methods \(^\text{21}\) as cells from both control and double-mutant mice were distributed evenly on the tSNE for many clusters as shown in the local density plot (Fig. 2b). However, in line with our entropy-based analysis (Supplementary Fig. 2), we noted that a number of cell types, including stromal and immune-related cells were significantly more abundant in double-mutant (Supplementary Fig. 5a, Supplementary Fig. 6),
while acinar and some ductal cell types were accordingly more prevalent in control tissues (Supplementary Fig. 5b).

In particular, our analysis revealed clusters of epithelial cells with transcriptional profiles unique for the tumor context. This included luminal- and basal-like cells as well as a small but distinct cancer stem cell (CSC)-like population in which Wnt-specific genes were activated (Fig. 2b, c). Among others, the expression of several genes such as Axin2, Ptn, Wif1, Clu and Wfdc18 was particularly characteristic for these tumor-specific cell clusters (Fig. 2c). Using these genes as markers, we further confirmed that the tumor-specific epithelial cells identified in our data were truly located in tumor regions. Immunofluorescence analysis of submandibular gland tissue sections from double-mutant mice showed that antibodies against these gene products positively stained cells within evident tumor lesions (Supplementary Fig. 7). To further increase tumor-specific resolution, we included other epithelial markers where possible. In line with our transcriptome data, Clu and Wfdc18 distinctively stained K8-negative and greatly overlapped with high Axin2, but only partially with Ptn and Wif1 stainings. Taken together, we identified a small subset of tumor-specific epithelial cells and provide a marker set that can identify them both at the RNA and protein level.

Simultaneous quantification of mRNA and cell surface proteins resolves immune cell diversity. Unbiased transcriptional profiling identified four clusters of various adaptive and innate immune cells. T or NK cells ('T/NK'; Ccl5, Nkg7, Cd7) were already present in control salivary glands (Fig. 1b). In the tumor context, 'immune' cells were significantly more abundant in double-mutant compared to control samples (Supplementary Fig. 5a). Additionally, new populations of monocytic cells ('immune') and terminal differentiated B plasma cells ('Ig-producing'; Ly6c2, Slp1, Xbp1) emerged in double-mutant tissues showing an influx into or activation of immune cells in the tumor environment (Fig. 2a; Supplementary Fig. 5a). To investigate this tumor-specific immune compartment in more detail, we used 'CITE-seq' in conjunction with a panel of 63 oligonucleotide-coupled antibodies mainly directed against immune and some epithelial cell surface proteins (see Supplementary Table 1). This allowed us to simultaneously quantify mRNA transcripts and protein epitopes from single cells of freshly dissociated ~P70 control and tumor-bearing salivary glands (Fig. 3a). Implementation of CITE-seq did not introduce any obvious transcriptional biases as mRNA-based clustering with
our earlier obtained transcriptome datasets recapitulated all cell clusters with cells from the CITE-seq experiments evenly distributed in the tSNE (Fig. 3b). Integration of cell surface protein information enhanced the signals for specific markers such as CD172a and CD11b (Fig. 3c), confirming that ‘immune’ and ‘immune2’ clusters predominantly consist of myeloid cells from the monocytic lineage. Subclustering of the combined ‘immune’ clusters revealed four cell subsets that exhibited signatures of tumor-associated macrophages (TAMs) with diverse activation and functional states (Fig. 3d, e). These TAMs differed in expression of genes associated with tissue-remodeling (Vim, Mmp12, Fn1)41–43, proliferating (Hmgb2, Mki67)44,45, interferon-responsive (Isg15, Rsad2, Irf7)46,47 or inflammatory (Irg1, S100a9, S100a8)48–50 properties (Supplementary Fig. 8).
summary, our results suggest that the immune landscape in this Wnt-dependent salivary gland tumor model is dominated by myeloid cells and TAMs, accompanied by a small population of tumor-associated B plasma cells. Consistent with the aggressive nature of these tumors, infiltration by inflammatory T cells seemed low\(^1\) with T/NK mostly corresponding to tissue-resident cells. Our data support the newly emerging concept that TAMs exhibit a wide and continuous spectrum of functional and differentiation states\(^2,5\) rather than conforming to a defined polarizing tumor-supporting vs. tumor-suppressing model.

Subclustering reveals two subpopulations of cancer stem and basal cells. Since our data indicated heterogeneity within tumor-specific cells (Fig. 2b, c), we selected basal cells and CSC-like cells for subclustering (Fig. 4a), which showed that each of the two clusters consists of two distinct subpopulations. Inspection of the top marker genes (Fig. 4b; Supplementary Fig. 9a) revealed that the CSC 1 subpopulation exhibited more basal-like (K14 + /K5+) Wnt-high (Bmp2, Bmp4, Dkk4+) whereas CSC 2 rather luminal-like (K18+) features. Moreover, we were able to distinguish the tumor-specific basal cell subset (‘basal tumor’) from the ‘normal’ one (‘basal normal’), as illustrated by projecting the subclustered cell populations back onto the original tSNE coordinates from the clustering of all samples (Supplementary Fig. 9c).

To functionally characterize these differences, we performed differential expression and subsequent pathway analyses between the respective pairs of CSC-like and basal subpopulations (Fig. 4c). This confirmed our initial observation in CSC-like cells as terms related to Wnt/β-catenin signaling and basal cell carcinoma were most significantly enriched in the ‘CSC 1’ compared to the ‘CSC 2’ subcluster. When comparing the two basal subpopulations, we found that the main systematic differences were linked to extracellular matrix (ECM) proteins or epithelial-to-mesenchymal transition (EMT) signatures, which were strongly upregulated in the tumor-specific basal subset. Specifically, we observed a more than two-fold upregulation of the EMT master regulator Snail and strong induction of other characteristic genes such as Serpine2, Sparc, Acta2, S100a6 or the TGF-β modulator Bgn\(^3\) (Fig. 4c; Supplementary Fig. 9, Supplementary Data 1, Supplementary Fig. 10), while many other canonical EMT markers such as Zeb1/2, Twist1/2, E11 or Cdh2 were insufficiently detected in our data to be tested for differential expression. To make our data easily accessible, we created a resource, which allows to access and interrogate our single-cell data for any gene of interest interactively via a web-based online tool (https://shiny.mdc-berlin.de/sc_msga/).

Computational lineage analyses reconstruct tumorigenesis. We further investigated to what extent tumor-specific cells were connected to one another and contributed to the tumorigenesis in our genetic model. We chose a diffusion map approach, which embeds data in low-dimensional space where distances between cells represent a gradual but stochastic continuation such as during developmental processes\(^4\). Together with pseudotemporal ordering, this analysis allowed us to predict a differentiation trajectory for tumorigenesis (Fig. 5a, Supplementary Fig 11a). In agreement with our model, in which tumors are induced by activation of β-catenin and Bmpr1a mutations via the basal-specific K14-cre promoter\(^16,17\), we found that this trajectory initiates in basal tumor cells, further proceeds through the two CSC-like subpopulations, and ends in the luminal Clu+ cell cluster with continuous transitions in between (Fig. 5a, b). We show that this trajectory is robust and reproducible across individual double-mutant tumor samples (Supplementary Fig. 11b) and that removal of CSC-like cells from the analysis still predicts the tumorigenesis path although disrupting cell connectivity (Supplementary Fig. 11c).

To validate this trajectory, we made use of the different time points (P40 and P90) at which we collected samples and monitored the contributions of tumor-specific cells pooled by stage (Fig. 5d, e; Supplementary Fig. 12) instead of cell type. In agreement with our model, we found that the number of cells from the early P40 tumor stage was significantly increased at the initial phase of tumorigenesis compared to that from the late P90 one, while the reverse was true at a more advanced phase. Stage-specific quantification of tumor-specific cells showed that although relative proportions of both CSC-like populations were similar, basal tumor and luminal Clu+ cells were, in fact, significantly more and significantly less abundant in P40 cells than in P90 cells, respectively (Fig. 5e).

To study potential changes in global expression dynamics along the trajectory, we took the most differentially expressed genes between the four subpopulations and ordered their expression by pseudotime (Fig. 5f, Supplementary Fig 13). The results replicated the differential expression analyses summarized in Fig. 4c, and showed that we could additionally capture more subtle differences for individual genes for which the expression changed as a function of the tumor progression. We found that EMT-related genes such as Ctgf\(^5\) and Sparc\(^6,7\) were first activated in basal tumor cells, and that their expression extended and decreased within the cell population (Ctgf) itself or throughout both CSC-like subpopulations (Sparc) as tumorigenesis proceeded (Fig. 5f, Supplementary Fig. 13). A similar pattern was observed for Meg3, a long non-coding RNA previously found to regulate EMT\(^8,9,60\).

While several genes (e.g., Ptn, S100a6, S100a14) were expressed in a more ubiquitous manner in CSC-like cells, Wnt target genes such as Axin2 and Bmp4 were specifically switched on in the CSC 1 subpopulation and continued to be only mildly expressed and subsequently inactivated in CSC 2 (Supplementary Fig. 13). However, in the latter, increased expression of the transcriptional regulators Nupr1 and Elf3 could be detected, both genes with known functions in epithelial cell differentiation and tumorigenesis\(^61,62\). Finally, Clu, and Wfde18, which we identified as being specific to luminal Clu+ cells (Fig. 2), were expressed at a later stage towards the end of our inferred trajectory. Validating previously proposed associations between EMT, Wnt signaling and stemness\(^63\), we found that gradual loss of the EMT expression signature correlated with activation of
Fig. 4 Subclustering reveals additional heterogeneity in cancer stem cell-like and basal cells. a Subclustering of the tumor-specific epithelial subset identifies two cancer stem cell-like ('CSC 1' and 'CSC 2') and two basal cell ('basal normal' and 'basal tumor') subpopulations. b Expression of several genes in the subclustered tumor-specific tSNE representation as indicated by a color scale ranging from gray (no expression) to dark blue (high expression) with regions of high expression highlighted. c Gene set enrichment analysis on differentially expressed genes between CSC-like or basal subclusters, respectively. Shown are log2 fold changes for all genes per pathway expressed in at least 5% of cells, with selected genes highlighted. Orange dots indicate the pathway mean, dashed lines the cutoff set on the pathway mean.
Wnt signaling and induction of known CSC marker genes in head and neck squamous cell carcinoma (Fig. 5c). Moreover, by inspecting differential pathways with mean log2 fold changes below the cutoff imposed in Fig. 4c, we identified dysregulated metabolic signatures suggesting that activation of the EMT program in basal tumor cells is accompanied by metabolic reprogramming from oxidative phosphorylation towards glycolysis (Supplementary Fig. 14). Together, our results show that our approach can identify specific genes and expression patterns potentially regulating and driving tumorigenesis.
**Discussion**

In this study, we created a high-resolution salivary gland cell atlas and systematically dissected the cellular heterogeneity in a genetically-controlled, Wnt-dependent mouse model of a solid tumor. We identified, molecularly characterized and validated cell clusters that were specific to the tumor and established their lineage relationship by uncovering the progenitor and progeny populations of CSCs.

Different from other studies involving single-cell transcriptome profiling of human biopsy samples, we could not computationally pinpoint the tumor cells by global copy-number variation (not present in our data), nor by identifying β-catenin and Bmpr1a mutations (not detectable) or EYFP-positive cells (mRNA transcripts not sufficiently captured by our sequencing method). Instead, we leveraged the reproducibility and large sample size of our data to systematically compare cells from control and tumor-bearing tissues to reliably and robustly identify those that were specific to the tumor samples, and confirmed these results by immunostainings in tissue sections.

Although only representing <1% of cells in our data, we were able to detect and extensively characterize CSC-like cells without relying on pre-defined surface markers, which often also target unrelated cells. We found that Wnt target genes such as Axin2 and Bmp4 were exclusively activated in this small population, but also other genes with highly tumor-specific functions such as Ptn (promoting tumor angiogenesis) and S100a14 (involved in tumorigenesis). Further investigation revealed that these CSC-like cells consist of a Wnt-high/K14+ and a luminal-like/K8+ subpopulation, also characterized by differential expression of the transcriptional regulators Nupr1 and E13, which are thought to promote metastasis and the induction of chemoresistance. Further experiments will be required to functionally characterize these CSC subpopulations and to address their stemness potential. Nevertheless, it has been recognized that different stem cell types with tissue renewal capacity can reside within the same tissue, that a continuum of stem cell states may provide a higher degree of flexibility, and that Wnt signaling is essential for stem and progenitor cell formation and function throughout development. Our results indicate that the stabilizing β-catenin mutations in our mouse salivary gland SCC model trigger similar processes, as uncontrolled Wnt signaling can lead to aberrant expansion of stem cells or confer stem cell behavior, paving the way for malignant proliferation.

Since the tumor in our model was driven by a K14-cre, we argued that the CSCs most likely originated from the tumor-specific basal population, characterized by an EMT expression signature. The importance of the EMT program is well established as a major mechanism for the ‘invasion-metastasis cascade’ in cancer biology. With an ability to degrade and reorganize the ECM, it promotes the loss of cell-cell adhesion and the acquisition of migratory and invasive traits. In fact, EMT signatures have been observed in the tumor microenvironment, but also in malignant cells, and its critical role for the induction of a stem-like phenotype has only been recognized over the last years. It has emerged that epigenetic and environmental cues can control EMT without introducing new genetic alterations, that this transition is therefore reversible, rarely fully executed under physiological or malignant conditions, and that an intermediate state between the poles of fully epithelial and mesenchymal cell identity is particularly favorable for the induction of the CSC phenotype. Specifically, Snai2/Slug in cooperation with Sox9 have been identified as stemness- and EMT-inducing factors in the basal layer of the mammary epithelium. Moreover, the activity of these transcription factors can promote the existence of transient epithelial cell populations which can convert to stem cells with long-term tissue reconstituting ability, and we propose that similar mechanisms play an essential role in our system.

Further, while connections between Wnt signaling both downstream and upstream of the EMT have long been drawn, we find here that genetic Wnt activation and Bmp deactivation in our model triggers EMT (presumably through Snai2) and met abolitic reprogramming, before other Wnt targets, such as Bmp2/4 or Dkk2/4 can be detected.

Our data allowed us to put the different tumor-specific cell populations on a robust trajectory and to map the continuous shifts in cellular identities during tumorigenesis by means of diffusion maps, which have been widely used together with single-cell data to study differentiation dynamics during development. We note, however, that the directionality of the inferred trajectory is not conclusively fixed and requires further investigation. Nevertheless, based upon the K14-cre-dependence of mutations triggering the tumor in our model as well as the relative timing of early- and late-staged samples, we hypothesized that the tumorigenesis here is initiated in basal cells. This model further suggests that after EMT induction, Wnt signaling was upregulated in one CSC subpopulation in concert with the acquisition of a stemness phenotype that persisted throughout another CSC subpopulation already primed towards a luminal-like cell fate, towards which tumorigenesis finally transitioned. Comparable to epithelial differentiation processes, we noticed a general K14+ to K8+ gradient indicating that tumor-specific cells shifted from exhibiting basal-like to more luminal-like characteristics along the trajectory. This was accompanied by a gradual loss of EMT markers and therefore also suggests a reversal to an epithelial cell identity. Moreover, we found that the luminal-like cells exhibited high expression of Wifc18, a gene that we identified to be specific to the excretory and intercalated ducts in control tissues. This suggests that in our model, Wifc18-positive cells in the tumor may have undergone a similar differentiation path as normal cells in these ductal compartments.

Our work demonstrates the importance of using controlled models to robustly and reproducibly study essential mechanisms of carcinogenesis and tumor progression which would otherwise not be possible. Finally, our approach provides a blueprint to molecularly identify markers and characterize transcriptional events that are fundamental in the regulation of tumorigenesis, and ultimately facilitate further clinical studies in the design of appropriate treatment strategies.

**Methods**

**Mouse strains.** K14Cre(Δneo), β-cateninlox/lox, Bmpr1alox/lox alleles, and Cre-inducible R26YFP reporter mice have been described, and mutant mice were analyzed for genotype and recombination by PCR. To obtain the double mutants, homozygous mice carrying the β-cateninlox gain-of-function, the Bmpr1alox loss-
of-function allele and R26<sup>ERT</sup> were crossed with K14-Cre mice that were homozygous for the Bmpr1a<sup>−/−</sup> allele. Control mice were obtained by crossing K14-Cre mice with a Cre-inducible control (B6.129S1-Cre<sup>+</sup>Itm1, 120037), and this study had a C57BL6 background. Animal experiments were approved by LAGeSo Berlin and performed according to EU and national institutional regulations.

**Tissue dissociation and single-cell sample preparation.** Submandibular glands and primary tumor samples were collected, minced and dissociated with a Gen- tileMacs Dissociator (Miltenyi Biotec) in digestion buffer (DMEM/F12 1:1 (Invitro- gene), 1.67 mg/ml collagenase (Invitrogen), 1.33 mg/ml hyaluronidase (Sigma) and 1.67 mg/ml dispase (Invitrogen)). Cell suspensions were passed through a stainless steel filter (70 μm) and centrifuged at 900 x g for 5 min at 4 °C. Pellets were suspended in 10 ml Dulbecco’s modified Eagle/F12:1 medium and washed three times with PBS containing 10% fetal bovine serum (Invitrogen). Prior to sorting, cells were pelleted at 1000 x g, and then filtered through a 40 μm nylon mesh HIGH PACT Aria (BD Biosciences) instrument was used for sorting and dead cells excluded by elimination of DAPI-positive cells and gates set to exclude cell clusters. Cells were directly sorted and fixed in ice-cold 80% methanol and stored at −80 °C until further processing.

**Drop-seq procedure, single-cell and bulk library generation, and sequencing.** Monodisperse droplets of about 1 nl in size were generated using microfluidic DDS devices (Drop-Seq chips, FlowEMS, Toronto, Canada; pre-coated with Aquapel). Barcoded microparticles (Barcoded Beads SeqB; ChemGenes Corp., Wilmington, MA, USA) were prepared and flowed in using a self-built Drop-seq set up<sup>19</sup> (Online- Dropseq-Protocol-v2.1: http://mccarrelllab.com/dropseq/) as previously described<sup>16</sup>. Cell preparations and reagents were kept on ice and handled in the cold. Methanol-fixed cells<sup>16</sup> were centrifuged at 3000 × g for 5 min and then suspended in 0.5 ml PBS+0.01% BSA supplemented with RNase inhibitors (1 unit/ml RiboLock, ThermoFisher), pelleted and resuspended again in 0.5 ml PBS+0.01% BSA in the presence of RNase inhibitors. Cells were manually counted by means of a hemocytometer and diluted to a suspension of typically ~50–100 cells/μl in PBS+0.01% BSA. Droplets were collected in 50 ml Falcon tubes for ~13 min, corresponding to ~1 ml of combined aqueous flow volume (1 ml cells and 1 ml of beads). Droplets were broken immediately after collection and barcoded beads with captured transcriptomes were reverse transcribed and exonuclease-treated. First strand cDNA was amplified by equally distributing beads from one run to 24 PCR reactions (50 μl volume; 4–9 cycles). 20 μl fractions of each PCR reaction were pooled (total = 480 μl), then double-purified with 0.6 μl volumes of Ampure XP beads (Beckman Coulter). Amplified cDNA libraries were sized and quantified on a BioAnalyzer High Sensitivity Chip (Agilent) and the Qubit dsDNA HS Assay system (ThermoFisher). Libraries were fragmented, amplified and aliquoted (12 cycles) and indexed for sequencing with the Nextera XT v2 DNA sample preparation kit (Illumina) using custom primers enabling 3'-targeted amplification as described<sup>19</sup>. The libraries were double-purified with Ampure XP beads (0.6 ×, 1 ×), quantified and sequenced on Illumina NextSeq500 instrument in paired-end mode; 50 base pairs and 2 reads per sample (50 cycles, 24 PCR reactions, 0.5 μl volume; 4 × 9 cycles). Sequencing was performed on the NextSeq500 (Illumina) and a Leica TCS SP8.

**Processing and analysis of single-cell RNA-seq data.** Drop-seq data were processed using Drop-seq tools v1.12<sup>19</sup> based on the Gencode vM7 reference augmented by two pseudo-chromosomes containing the Cre and EYFP sequences, respectively. Selecting valid barcodes with the ‘knee’ method, resulting DGEs for each sample were combined and analyzed in R (version 3.4.4) using Seurat<sup>23</sup>. We additionally filtered for at least 100 genes and less than 15% mitochondrial content. Clustering and t-SNE were performed based on a PCA with significant components chosen using JackStraw. For clustering, we used a resolution of 0.7 for the control samples and of 1.0 for the combination of all samples, and merged several acinar subclusters which were not distinguishable from their marker genes. We controlled replicate-to-replicate consistency (see below) across across subclusters. We also performed an additional analysis where we mapped against a combined human and mouse genome to identify cells of human origin or doublets, and removed these cells from the standard pipeline output. We used CITEseq-count (v1.2) to quantitate ADT counts for the respective cells found in both pipeline runs. Specifically binding antibodies were identified by inspecting ADT count distributions for human and mouse cells.

**CITE-seq experiments.** Antibodies were covalently and irreversibly conjugated to DNA-barcoding oligonucleotides by IEDDA click chemistry as previously describ- ed<sup>80</sup>. The antibody panel (see Supplementary Table 1) was prepared by mixing equal quantities of each DNA-barcoated antibody and concentrating the panel on an Amicon Ultra 0.5 ml 30 kDa MWCO centrifugal filter (Millipore).

Live intact cells from control and double-mutant submandibular glands of ~P70 were fixed for 12 cycles together with TruSeq Small RNA primers for indexing and rRNA LT Sample Prep Kit, Illumina). Libraries were sequenced on Illumina NextSeq500 sequencers (library concentration 1.8 pM; NextSeq 500/550 High Output v2 kit (75 cycles) in paired-end mode; read 1 = 20 bp using the custom primer RAdCustSeqB<sup>19</sup>, read 2 = 64 bp).

For bulk sequencing, RNA was extracted with Trizol (Invitrogen) from freshly dissected submandibular glands. RNA integrity was assessed on a Bioanalyzer RNA Nano Chip (Agilent). Strand-specific cDNA libraries were generated from 500 ng total RNA according to the Illumina TruSeq protocol (TruSeq Stranded mRNA LT Sample Prep Kit, Illumina). Libraries were sequenced on an Illumina NextSeq 500 sequencer using the High Output v2 Kit (150 cycles), single read: 150 bp, index read: 6 bp.

**Immunostainings.** Immunofluorescence analyses were performed on formalin-fixed, paraffin-embedded tissue sections as described<sup>82</sup>. Antibody detection was accomplished by a biotinylated streptavidin conjugation method<sup>60</sup> followed by staining with either a secondary antibody or the secondary antibody and anti-CK14 antibodies (aka K14 or Krt14; ThermoFisher, MA5-11599), goat-anti-MA- PGR (aka Pgyp1r; ThermoFisher, MA1-41044), rabbit-anti-Hepacamar2 (Abcam, ab189943), goat-anti-Muc19 (Abcam, ab121014), rabbit-anti-MAF (Santa Cruz Biotechnology, sc-51422), rabbit-anti-Clusterin (Abcam, ab9695), guinea pig-anti-CK14 (aka Krt14; Proser, GP-K8), rabbit-anti-CK14 antibodies (aka K14 or Krt14; ThermoFisher, MA5-11599), rabbit-anti-Kl1 (Boster, PA1709), mouse-anti-Ptn (Santa Cruz Biotechnology, SC74443), goat-anti-Saprc (R&D Systems; AF942). Secondary antibodies were conjugated with Alexa-Fluor 488, 594, or 647 fluorochromes (Jackson Immunological Research Laboratories). Images were captured using an Axio imager Z1m and AxioCam MRM (Carl Zeiss) and a Leica TCS SP8.

Images were captured using an Axio imager Z1m and AxioCam MRM (Carl Zeiss) and a Leica TCS SP8.
and diffusion pseudovalues. We also ran destiny on each sample with more than 10 cells in the relevant subpopulations separately and compared sample-specific to global pseudotime estimates. Differential genes along the trajectory were identified on pooled counts using DESeq2 with a likelihood ratio test on all 4 clusters simultaneously, and gene expression values plotted along the trajectory after imputation using SAVER version 1.1.1. Gene set scores were computed supplying imputed expression values to Seurat's AddModuleScore function. An additional "stemness" gene set (Pou5f1, Nanog, Sox2, Prom1, Bmi1, Lgr5, Msi1, Tdgf1, Bmp4, Cgrip4, Cxcr4, Alcam, Slc2a13, Aldh1a1) was curated from the literature.8

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The sequencing data generated is available in GEO under the accession GSE124425. Metadata for each sequenced cell, from which figures based on the single-cell RNA sequencing data can be reproduced, are provided as a source data file.

Code availability
Custom R scripts used to analyze data and generate figures are available upon request.

Received: 8 July 2019; Accepted: 29 January 2020;
Published online: 20 February 2020

References
1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. Cell 144, 646–674 (2011).
2. Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell 21, 309–322 (2012).
3. Lawrence, D. A. et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature 526, 131–135 (2015).
4. Gao, R. et al. Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast cancer. Nat. Commun. 8, 228 (2017).
5. Filbin, M. G. et al. Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. Science 360, 331–335 (2018).
6. Tirio, L. et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Cell 171, 309–313 (2016).
7. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck. Cancer Cell 171, 1611–1624.e1624 (2017).
8. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science 361, 594–599 (2018).
9. Pattabiraman, D. R. & Weinberg, R. A. Tackling the cancer stem cells. Nat. Rev. Cancer 10, 59–63 (2010).
10. Reya, T. & Clevers, H. Wnt signalling in stem cells and cancer. Nat. Rev. Mol. Cell Biol. 9, 789–798 (2008).
11. Brabletz, T., Jung, A., Spaderna, S., Hlubek, F. & Kirchner, T. Opinion: cancer stem cells: controversial or just misunderstood? Cell Stem Cell 4, 195–197 (2009).
12. Mani, S. A. et al. The epithelial-mesenchymal transition generates cells with stemness characteristics. Nat. Rev. Mol. Cell Biol. 8, 39–54 (2007).
13. Jordan, C. T. Cancer stem cells: controversial or just misunderstood? Cell Stem Cell 4, 195–197 (2009).
14. Kusakabe, Y. et al. Sex-based differences in Smgc expression in the mammalian salivary glands. PLoS One 11, e0148973 (2016).
15. Galon, J. & Bruni, D. Approaches to treat immune hot, altered and cold tumors with combination immunotherapies. Nat. Rev. Drug Discov. 18, 197–218 (2019).
16. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. F1000Prime Rep. 6, 13 (2014).
17. Zhu, Q. et al. The Wnt-Driven Mll1 Epigenome Regulates Salivary Gland and Neural Tissue Development. Mol. Cell 74, 218–229 (2019).
18. Shi, W. et al. Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. Nat. Immunol. 16, 663–673 (2015).
19. Stockeius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. Nat. methods 14, 865–868 (2017).
20. Ao, R. et al. Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. Nat. Immunol. 16, 663–673 (2015).
21. Behrens, J. et al. Functional interaction of beta-catenin with the transcription factor LEF-1. Nature 382, 638–642 (1996).
22. Huber, O. et al. Nuclear localization of beta-catenin by interaction with transcription factor LEF-1. Mechanisms Dev. 59, 3–10 (1996).
23. Wynn, T. A. & Vannella, K. M. Macrophages in tissue repair, regeneration, and fibrosis. Immunity 44, 450–462 (2016).
24. Behmoraas, J. et al. Macrophage epoxide hydrolase determines a proinflammatory transcriptional signature. J. Immunol. 194, 4705–4716 (2015).
25. Pasterla, T., de Marchis, F., Palumbo, R. & Bianchi, M. E. High mobility group box 2 participates in the development of breast cancer. Cell Stem Cell 10, 1–8 (2017).
26. Shi, W. et al. Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. Nat. Immunol. 16, 663–673 (2015).
27. Stockeius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. Nat. methods 14, 865–868 (2017).
28. Ao, R. et al. Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. Nat. Immunol. 16, 663–673 (2015).
29. Sainz, B. J., Martin, B., Tatari, M., Heeschen, C. & Guerra, S. ISG15 is a critical microenvironmental factor for pancreatic cancer stem cells. Cancer Res. 74, 7309–7320 (2014).
30. Jiang, J. S. et al. Rsad2 is necessary for mouse dendritic cell maturation via the IRF7-mediated signaling pathway. Cell death Dis. 9, 823 (2018).
31. Michelacci, L. et al. Immune-responsive gene 1 protein links metabolism to immunity by catalyzing icatonic acid production. Proc. Natl Acad. Sci. USA 110, 7820–7825 (2013).
32. Lampropoulou, V. et al. Icatonin links inhibition of succinate dehydrogenase with macrophage metabolic remodeling and regulation of inflammation. Cell Metab. 24, 158–166 (2016).
33. Xia, C., Braunstein, Z., Toomey, A. C., Zhong, J. & Rao, X. S100 Proteins as an important regulator of macrophage inflammation. Front. Immunol. 8, 1908 (2017).
34. Galon, J. & Bruni, D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat. Rev. Drug Discov. 18, 97–218 (2019).
35. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. F1000Prime Rep. 6, 13 (2014).
53. Thakur, A. K. et al. TAP73 loss favors Smad-independent TGF-beta signaling that drives EMT in pancreatic ductal adenocarcinoma. Cell Death Differ. 23, 1358–1370 (2016).
54. Haghverdi, L., Buttnn, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. Nat. Methods 13, 845–848 (2016).
55. Sonnyal, S. et al. Connective tissue growth factor causes EMT-like cell fate changes in vivo and in vitro. J. Cell Sci. 126, 2164–2175 (2013).
56. Arnold, S. A. & Brekken, R. A. SPARC: a matricellular regulator of the Wnt/beta-catenin pathway. J. Cell Commun. Signal. 3, 255–273 (2009).
57. Framson, P. E. & Sage, E. H. SPARC and tumor growth: where the seed meets the soil? J. Cell. Biochem. 92, 679–690 (2004).
58. Terashima, M., Tange, S., Ishimura, A. & Suzuki, T. MEG3 Long noncoding RNA contributes to the epigenetic regulation of epithelial-mesenchymal transition in lung cancer cell lines. J. Biol. Chem. 292, 82–99 (2017).
59. Gao, Y. & Lu, X. Decreased expression of MEG3 contributes to retinoblastoma progression and affects retinoblastoma cell growth by regulating the activity of Wnt/beta-catenin pathway. Tumour Biol. 37, 1461–1469 (2016).
60. Zhou, Y. et al. Activation of p53 by MEG3 non-coding RNA. J. Biol. Chem. 282, 24731–24742 (2007).
61. Chowdhury, U. R., Samant, R. S., Fodstad, O. & Shevde, L. A. Emerging role of nuclear protein 1 (NUPR1) in cancer biology. Cancer Metastasis Rev. 28, 225–232 (2009).
62. Hou, J. et al. Transcriptional regulation of the murine Elf3 gene in embryonal carcinoma cells and their differentiated counterparts: requirement for a novel upstream regulatory region. Gene 340, 123–131 (2004).
63. Shibu, T. & Weinberg, R. A. SPARC: a matricellular regulator of the Wnt/beta-catenin pathway. Nat. Cell Biol. 15, 204–212 (2013).
64. Major, A. G., Pitty, L. P. & Farah, C. S. Cancer stem cell markers in head and neck squamous cell carcinoma. J. Cell Sci. 126, 1015–1028 (2013).
65. Shibue, T. & Weinberg, R. A. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nat. Rev. Clin. Oncol. 14, 611–629 (2017).
66. Major, A. G., Pitty, L. P. & Farah, C. S. Cancer stem cell markers in head and neck squamous cell carcinoma. Stem Cells Int. 2013, 319489 (2013).
67. Kadomatsu, K. & Muramatsu, T. Midkine and pleiotrophin in neural development and cancer. Cancer Lett. 204, 127–134 (2004).
68. Chen, H. et al. Involvement of S100A14 protein in cell invasion by affecting expression and function of matrix metalloproteinases (MMP)-2 via p53-dependent transcriptional regulation. J. Biol. Chem. 287, 17109–17119 (2012).
69. Visvader, J. E. & Clevers, H. Tissue-specific designs of stem cell hierarchies. Nat. Cell Biol. 18, 349–355 (2016).
70. Nusse, R. & Clevers, H. Wnt/beta-catenin signaling, disease, and emerging therapeutic modalities. Cell 169, 985–999 (2017).
71. Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor microenvironment. Nat. Med. 24, 1277–1289 (2018).
72. Li, H. et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nature 508, 1289–1298 (2014).
73. Pastushenko, I. et al. Identification of the tumour transition states occurring during EMT. Nature 556, 463–468 (2018).
74. Guo, W. et al. Slug and Sox9 cooperatively determine the mammary stem cell state. Cell 158, 1015–1028 (2014).
75. Moingard, V. et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat. Biotechnol. 33, 269–276 (2015).
76. Bach, K. et al. Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat. Commun. 8, 2128 (2017).
77. Harada, N. et al. Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. EMBO J. 18, 5931–5942 (1999).
78. Huelskens, J., Vogel, R., Erdmann, B., Cotsarelis, G. & Birchmeier, W. beta-Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. Cell 105, 533–545 (2001).
79. Mishina, Y., Hanks, M. C., Miura, S., Tallquist, M. D. & Behringer, R. R. Generation of Bmpr1a/Alk3 conditional knockout mice. Genesis 32, 69–72 (2002).
80. Thorey, I. S. et al. Selective disruption of genes transiently induced in differentiating mouse embryonic stem cells by using gene trap mutagenesis and site-specific recombination. Mol. Cell. Biol. 18, 3081–3088 (1998).
81. Srinivas, S. et al. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev. Biol. 1, 4 (2001).
82. Bates, D., Machler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. J. Stat. Soft. https://www.jstatsoft.org/article/view/v067i01 (2015).
83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
84. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 10, 161 (2009).
85. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics 31, 2989–2998 (2015).
86. Huang, M. et al. SAVER: gene expression recovery for single-cell RNA sequencing. Nat. Methods 15, 539–542 (2018).

Acknowledgements
We thank Margareta Herzog for help with single-cell experiments; Marion Müller for immunostaining experiments; Salah Ayoub and Anastasiya Bollengagen for technical support; Aristotleis Misios and Thomas Sommermann for help with the initial CITEx-seq implementation; Malin Praktiknjo for help with tissue illustrations; Veronica Rodilla, Eric Blanc, Damian Rieke and the Rajewsky lab for helpful discussions. This work was supported by an Early Independence Grant from the Berlin School of Integrative Oncology funded by the German Excellence Initiative (to S.D.P. and L.F.), by the Deutsche Forschungsgemeinschaft (to W.B.) and the Chan Zuckerberg Initiative (to N.R.).

Author contributions
N.R. and S.D.P. conceived the study. S.D.P. designed and performed single-cell experiments. Q.Z. and L.F. prepared samples from mice. Q.Z. and H.Q. collected imaging data. W.B. provided mouse model. M.S. provided advice and antibody panel for CITEx-seq experiments. B.O. designed and implemented computational strategy, and analyzed data with S.D.P. Data interrogation and interpretation were carried out mainly by S.D.P., B.O., C.K., and N.R. H.L. wrote the Shiny app. S.D.P., L.F., W.B., and N.R. procured funding. S.D.P., B.O. and N.R. wrote the paper with input from C.K., Q.Z., M.S., and W.B. N.R. and S.D.P. coordinated the study.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-14777-0.

Correspondence and requests for materials should be addressed to S.D.P., W.B. or N.R.

Peer review information Nature Communications thanks Samantha A. Morris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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.

© The Author(s) 2020