Single-cell transcriptomics reveals varying degrees of epithelial-mesenchymal plasticity in lymph node metastasis from oral cavity squamous cell carcinoma

Kai Horny[1,2], Christoph Sproll[3], Lukas Peiffer[1,2], Frauke Furtmann[1,2,4], Patricia Gerhardt[1,2], Jan Gravemeyer[1,2], Nikolas H. Stoecklein[5], Ivelina Spassova[1,2,4], Jürgen C. Becker[1,2,4,*]

[1] Translational Skin Cancer Research, German Cancer Consortium (DKTK), 45141 Essen, Germany

[2] German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

[3] Department of Oral-, Maxillo- and Plastic Facial Surgery, Medical Faculty, University Hospital of the Heinrich-Heine-University Düsseldorf, Germany

[4] Department of Dermatology, University Medicine Essen, 45141 Essen, Germany

[5] Department of General, Visceral and Pediatric Surgery, Medical Faculty, University Hospital of the Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

* Correspondence: j.becker@dkfz.de

Abstract

Oral cavity squamous cell carcinomas (OSCC) that spread to the regional lymph nodes are associated with a high degree of epithelial-mesenchymal plasticity (EMP). Here, RNA sequencing of 1,946 carcinoma and 2,175 stromal cells of an advanced metachronous OSCC lymph node metastasis revealed heterogeneity of the OSCC cells with respect to epithelial-mesenchymal transition (EMT) states, metabolic adaptations, and cell cycle. Seven cell clusters formed a branching trajectory with epithelial differentiation, partial EMT states and metabolic adaptations. Notably, partial EMT was associated with stress response, especially with HSPA upregulation, immune-modulation through *CXCL14* and a higher fraction of proliferating cells, whereas epithelial differentiation correlated with higher cytokeratin and kallikrein expression. While EMP-associated transcription factors SNAI2 and ZEB1 were active in all cancer cells, TWIST1, SNAI1 and ZEB1 activity were more pronounced in the epithelial branch. Epithelial differentiating cells have less receptor-ligand interactions with stromal cells than partial EMT cells indicating a more independent regulation. Interestingly, cells in and close to the trunk of the trajectory adapt to hypoxic and hypoglycemic conditions expressing genes related to glycolytic and amino acid catabolism. Our work provides in-depth insights into phenotypic heterogeneity within an OSCC lymph node metastasis, specifically how EMP drives locoregional progression.

Frequently used abbreviations

CAF	Cancer-associated fibroblast
CNV	Copy number variation
DEG	Differentially expressed genes
DGE	Differential gene expression
EMP	Epithelial-mesenchymal plasticity
EMT	Epithelial-mesenchymal transition
GSEA	Gene set enrichment analysis
GSVA	Gene set variation analysis
HNSCC	Head and neck squamous cell carcinoma
KNN	K-nearest neighbor
MET	Mesenchymal-epithelial transition
MSigDB	Molecular signature database
OSCC	Oral cavity squamous cell carcinoma
PC	Principal component
PCA	Principal component analysis
pEMT	partial EMT
scRNAseq	Single cell RNA sequencing
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier

Introduction

Oral cavity squamous cell carcinoma (OSCC) belongs to the class of head and neck squamous cell carcinomas (HNSCC) with an incidence of 8.8 new cases per 100,000 people and a mortality of 2.8 (32%) in Germany in 2020 [1, 2]. Most OSCCs are associated with tobacco and alcohol abuse [1]. The treatment for primarily diagnosed OSCCs is the surgical removal of tumor accompanied by neck dissection and radiotherapy [3]. Primary OSCC has a propensity to spread into the regional lymph nodes, but only few patients develop distant metastases. The dissemination of the primary tumor is associated with epithelial-mesenchymal transition (EMT) and metastasis consolidation is thought to be associated with the reverse process named mesenchymal-epithelial transition (MET) [4-6].

EMT is an adoption of a developmental program in which cells gain migratory and invasive properties. Cells lose epithelial cell-to-cell connections and acquire a mesenchymal morphology, accompanied by transcriptional, epigenetic, and post-translational changes [7]. EMT programs are involved in embryonal development and wound healing, but also in cancer metastasis [4, 5]. EMT is a multistep process with many highly variable alternative routes depending on the cancer type and conditions [8-10]. Thus, expression of EMT marker genes, transcription factors and gene expression signatures is highly heterogeneous between patients, different lesions from one patient and between individual cells within one lesion [7, 10]. Indeed, EMT is a continuous, dynamic, and reversible process resulting in cancer cells showing a multitude of intermediate or partial states [5, 8, 9, 11-15]. Therefore, it has recently been recommended that this EMT continuum is referred to as epithelial-mesenchymal plasticity (EMP) [7, 16]. During the metastatic cascade, cells evolve through this continuum, changing from epithelial to more mesenchymal or partial EMT states and back to epithelial phenotypes [17].

Most EMP-related single cell studies are based on controlled *in vitro* and *in vivo* experiments [8, 9, 11, 13] and only few studies observed EMP within freshly isolated tumor samples [12, 18, 19]. Notably for OSCC, Puram *et al.* characterized 2,215 malignant cells from 18 patients

and discovered a partial EMT state with high variability in EMP-related gene expression [12]. Here, we report comprehensive single-cell RNA sequencing data for 1,946 malignant cells isolated from an advanced metachronous metastasis of an OSCC from the lower lip mucosa. Our data demonstrates varying degrees of EMP within a single metastasis, driven in different directions and towards environmental adaptations such as hypoxia, supply of nutrients and possible cell-cell interactions.

Results

Patient and tumor characteristics

The patient initially suffered from an OSCC of the lower lip mucosa that was surgically removed by wedge excision with primary wound closure. Neither subsequent elective neck dissection nor irradiation was performed presumably because of the multimorbidity of the patient. One year later, the patient noticed a swelling under the left mandible, which was suspected to be multiple metastases on computed tomography. With a delay of an additional four months, a left sided neck dissection was performed confirming the diagnosis. One of these lymph node metastases was selected for single cell RNA sequencing (scRNAseq).

Histological examination of the analyzed tissue sample showed a completely disrupted lymph node structure by metastatic carcinoma (Figure 1A, Supplementary Figure 1). We additionally found extranodal spread with perineural and vascular wall invasion and a marked fibrosis with focal sclerotic areas around the tumor cells. Tumor infiltrating lymphocytes were rare. Multiplexed single cell RNA sequencing (scRNAseq) by 10x Genomics technology recovered 4,121 cells, which, based on their gene expression similarity clustered into five groups representing unique cell types: 1,946 (47.1%) malignant cells, 1,192 (29.1%) fibroblasts, 506 (12.2%) dendritic cells, 375 (9.2%) macrophages and 102 (2.5%) endothelial cells (Figure 1B). Cell types were identified by their respective marker genes expression (Figure 1C, Supplementary Figure 2A-D). In accordance with the histological evaluation, the dataset does not include any T- or B cells. Malignant cells were identified both by the presence of copy number variations (CNVs) inferred from scRNAseq data as well as the expression of *S100A2, S100A14*, cytokeratins (*KRT5*, *KRT14*, *KRT17)* and stratifin (*SFN*), the latter highlighting their epithelial origin (Figure 1C, D, Supplementary Figure 2E).

Phenotypic heterogeneity of tumor cell populations is characterized by specific gene expression patterns

Within tumor cells, we identified seven different cell populations based on gene expression profiles using clustering and uniform manifold approximation and projection (UMAP) (Figure

2A). The relationship between these populations is reflected by a trajectory branching into two curves that largely follow the first and second principal component (PC) (Figure 2B). The variation reflected in the first PC is mostly driven by EMP-related genes such as vimentin (*VIM*), cytokeratins (*KRT17*, *KRT6B*, *KRT6A*), S100 proteins (*S100A8*, *S100A9*) and matrix metallopeptidases (*MMP10*) (Figure 2C). The second curve varies in the expression of cell cycle (e.g., *HMGB2*, *TUBB*), metabolism (e.g., *ASNS*, *SLC2A1*) and hypoxia (e.g., *NDRG1*) related genes. The majority of malignant cells are accumulating closely on the center of the trajectory within clusters 2, 3 and 6, while the remaining cells showed larger variation in their expression patterns as indicated by larger distances in PCs and on the trajectory (Figure 2B, D).

Since the beginning and end of such trajectories cannot be defined with certainty, we estimated RNA velocity to predict the short-term future development of individual cells using the ratio of spliced and non-spliced mRNA counts (Figure 2E). While the cells in the center of the trajectory did not show a uniform developmental direction, the cells from cluster 4 were developing towards cluster 5, adopting a progressive epithelial differentiation, and cells from cluster 6 towards cluster 7 resembling cell cycle progression. Another possibility to determine the direction of a trajectory is the accumulation of genetic aberrations in the cancer cells. The inferred CNVs revealed an increasing frequency of copy number gains on chromosome 1, 8, 17, and 19 for cells developing towards cluster 4 and cluster 5 (Figure 2F). These gains suggest that these cells developed later during tumor evolution. However, it should be noted that two of the four identified CNVs on chromosome 1 and 17 are associated with upregulated epithelial genes in cluster 4 and 5 that are in close genomic proximity; thus, these two may not represent true genomic CNVs, but rather reflect the high expression of these genes (Supplementary Figure 3).

Epithelial-mesenchymal plasticity (EMP) drives heterogeneity of metastatic OSCC cells

The gene expression patterns of the different tumor cell populations reflect varying degrees of EMP, cell cycle, and metabolic adaptation (Figure 3A, B). Two clusters (cluster 3 & 6) express

genes associated with a partial EMT state, two clusters (cluster 4 & 5) express genes related to epithelial differentiation, two clusters (cluster 1 & 2) express genes linked to metabolic changes for hypoxia-response or in amino acid metabolism and one cluster (cluster 7) express genes related to cell cycle. A large number of differentially expressed genes (DEGs) of cluster 2, 3, 6 and 7 are within the EMT hallmark gene set from the molecular signatures database (MSigDB) and are known to contribute to previously reported partial EMT gene expression signatures [12, 20] (Figure 3A, Supplementary Figure 4, Supplementary Table 1). Cells of these clusters express a mixture of both epithelial and mesenchymal genes, thus reflecting intermediate states of EMT. For example, the mesenchymal marker *VIM* is highly expressed within those clusters and increases through the second curve towards cluster 6 and 7, while E-cadherin (*CDH1*) is generally expressed only at low levels (Figure 3C). Hence, curve 2 appears to reflect the development towards a more mesenchymal phenotype. Of note, mRNAs encoding the mesenchymal markers N-cadherin (*CDH2*) and fibronectin 1 (*FN1*) were not detected in the tumor cells (Supplementary Figure 4). The partial EMT-related cluster 3 is located in the center of the trajectory and highly expresses matrix metallopeptidases (*MMP1*, *MMP10*, *MMP13*), i.e., genes associated with remodeling of the extracellular matrix. Cells of this cluster also express genes associated with immune-regulatory functions such as *CXCL14* [21, 22]. Moreover, DEGs of these cells are enriched with genes involved in type I interferon response, such as *IFITM1/2/3,* and response to stress, e.g., heat shock proteins (HSPAs) or stress-related transcription factors *JUN*/*FOS*. (Figure 3A, D, Supplementary Figure 4, Supplementary Table 1). On the other hand, curve 1 moving towards cluster 4 and 5 reflects cells acquiring epithelial characteristics, i.e., DEGs enriched for keratinocyte differentiation. Specifically, cells of cluster 4 express classical epithelial markers such as *SPRR1B*, *KRT16*, *KRT6B*, *S100A7*, *S100A8* and *S100A9*. With advancing epithelial differentiation towards cluster 5 the expression of *KRT13*, *KRT15*, *PSCA* and kallikrein-related peptidases *KLK5/6/7* increases. We confirmed the gene expression-based observation by comparing our epithelial and partial EMT signatures to a broader set of gene expression signatures from the EMTome database (Supplementary Figure 5).

Gene expression patterns indicate differences in metabolic adaptation of OSCC cells

As already mentioned above, cluster 1 and 2 were characterized by DEGs that are essentially associated with changes in metabolism (Figure 3A, E, Supplementary Figure 4, Supplementary Table 1). Specifically, DEGs of cluster 2 are enriched for hallmark hypoxia genes such as *NDRG1* and *EGLN3* whose expression depends on the oxygen level [23, 24]. *NDRG1* regulates stress response and p53-mediated caspase activation [23]. *EGLN3* has an important role in hypoxia-inducible factor 1 alpha (HIF1α) regulation through prolyl hydroxylation [24]. Consistent with the indicated anaerobic conditions within the tumor, glycolysis-related genes are enriched and upregulated, e.g., *ENO1/2/3*, *ALDOA*, *PGK1* and *PFKP* [25]. Further, the percentage of mitochondrial gene expression is lower in cluster 2 cells, suggesting less oxidative stress by less mitochondrial activity (Figure 3F). However, other partial EMT related genes are also upregulated within the hypoxia-responding cluster, e.g., *SERPINE1*, *LAMB3* and *TIMP3* [12]. DEGs of cluster 1 are enriched for amino acid metabolism, response to starvation and mTORC1 signaling; the latter is a regulator of mitochondrial metabolism [26]. These genes include *ASNS*, *PSAT1* and *PHGDH* that integrate the metabolites of serine and glycine metabolism into glycolysis and therefore fuel glycolysis with amino acids [27]. Hence, cells appear to be adapted to low-glucose conditions.

Cell cycle drives partial EMT cells towards a more mesenchymal phenotype

Cells developing on curve 2 towards cluster 6 and 7 are progressing in cell cycle. Specifically, cluster 6 is enriched for cells entering S phase, while cells in G2M phase are predominant in cluster 7 (Figure 3A, F, Supplementary Figure 4). These cells of these clusters express *VIM*, *CAV1*, *THBS1* and *MT2A*, i.e., genes associated with a mesenchymal phenotype. For analyzing the fraction of cycling cells and to exclude cell cycle effects as possible confounders, we repeated the analyses after regressing for them (Supplementary Figure 6). Thereby, we identified similar cell populations and a trajectory that is independent of cell cycle progression. We still identified a cluster related to high expression of cell cycle related genes, but the cells do not form a uniform cluster on the UMAP. Dividing the cell populations by UMAP and

counting cells in specific cell cycle phases showed the highest percentage of proliferating cells in partial EMT cells and lowest in epithelial differentiating cells (Figure 3F).

Epithelial differentiating cells have a distinct active transcriptional program

Next, we characterized the regulation of EMP programs and hypoxia-response within the tumor populations by analyzing the activity of the EMP-related transcription factors Zinc finger E-box-binding homeobox 1 & 2 (ZEB1/2), Snail and Slug (SNAI1/2), Twist-related protein 1/2 (TWIST1/2), and HIF1α [7]. Only SNAI2 and HIF1α encoding mRNA was expressed at levels detectable by multiplexed scRNAseq; *SNAI2* expression is evenly distributed - with the exception of cluster 5 - across all tumor cell populations in 20% of the cells while HIF1α expression was detected in 40% of cells (Figure 4A, B). However, the activity of transcription factors is not well reflected by its mRNA expression alone, as it depends also on protein stability and posttranslational modifications; for example, ZEB1 is more stable than SNAI1 [28]. To address this notion, we inferred the activity of transcription factors based on the mRNA expression profile of their target genes using the algorithm VIPER with regulons defined by DoRothEA [29-31]. Surprisingly, despite the fact that we could only detect mRNA for *SNAI2*, we found similarly high ZEB1 and SNAI2 activity with some variation along the branching trajectory (Figure 4C). Surprisingly, the epithelial differentiating cells in cluster 4 and 5 showed higher ZEB1, SNAI1 and TWIST1 and less TWIST2 activities than the other cells. As some cells indicated a response to hypoxic conditions, we further investigated the possible link between hypoxia-response and EMP. Initially, both EMT and hypoxia were considered as separate events promoting invasion and metastasis. More recently, it was shown that the HIF pathway is interrelated with EMP, assuming a hypoxia-induced EMT [32, 33]. However, while we detected HIF1α mRNA expression, it appeared not to be truly active within the OSCC cells (Figure 4A, C).

Partial EMT OSCC cells have more interactions with stromal cells

In addition to OSCC cells, we identified fibroblasts, endothelial cells, macrophages, and dendritic cells (Figure 1B, Supplementary Figure 2A-D). At least three subpopulations of

cancer-associated fibroblasts (CAFs) as well as myoblast-like and endothelial cells were identified (Figure 5A-C). CAF1 showed high expression of fibroblast activation protein (*FAP*), which is low in CAF2 and 3. Instead, the latter both highly express melanoma cell adhesion molecule (*MCAM*). CAF3 fibroblasts showed pronounced gene expression of alpha smooth muscle actin (*ACTA2*), which characterizes myofibroblasts, and also Interleukin 6 (*IL-6*) [34]. *IL-6* does not only exert immune-regulatory functions but is also able to promote EMP processes [34, 35]. Interestingly, CAF3 clusters closely to myoblast-like cells, which while lacking expression of fibroblast markers such as collagens or *FN1*, express typical markers for myogenic differentiation such as *MYF5* and *MYF6* [36] (Figure 5B, C, Supplementary Figure 7A). The tumor infiltrating myeloid cells form three dendritic cell and three macrophage clusters shown on the UMAP (Figure 5D). Dendritic cell cluster 1 (dend1) and macrophage cell cluster 2 (macro2) showed high expression of stress-related genes like *HSPA1A*, *HSPA1B* and *DNAJB1* (Figure 5E and F). Macrophages clusters further showed differential expression of cytokines and chemokines, e.g., *SPP1* (macro1), *CXCL10* (macro2) and *CCL18* (macro3), indicating differences in their polarization [37].

Cell-cell interactions within the tumor microenvironment significantly influence the properties of the respective cells. Expression of ligands and receptors in the different cell types allows to infer probable cell-cell interactions using CellPhoneDB (see Methods) [38]. The majority of predicted interactions are between tumor and endothelial cells, tumor cells and fibroblasts as well as fibroblasts and endothelial cells (Figure 5G). Interestingly, epithelial differentiating OSCC cells (cluster 4 and 5) showed fewer ligand-receptor interactions with other cell types than those with a partial EMT phenotype (cluster 3 and 6) (Figure 5H). Specifically, *EGFR* mRNA was found to be highly expressed by the indicated tumor cells, which suggests interactions with cells expressing the respective ligands such as *MIF*, *HBEGF, and TGFB1* (Supplementary Figure 7B, C, Supplementary Table 2).

Discussion

In cancers arising from epithelial cells, some of the tumor cells adopt features of mesenchymal cells and lose their epithelial characteristics. The epithelial-mesenchymal plasticity is implicated in cancer metastasis, which in turn determines prognosis for most malignancies [39]. How these transitions occur is still not fully understood. During EMP, cells may go through a gradual transition from one state to the other along a continuous spectrum of change, but also distinct stable, long-lasting cell populations representing discrete EMT states have been postulated to exist [5, 8, 9, 11-15]. Most studies examining EMP showed continuous transitions, but these studies are based on *in vitro* systems or *in vivo* models that may not fully reflect the complexity of the tumor and its microenvironment [8, 9, 11]. Different EMT states are induced by a variety of microenvironmental factors [9]. Indeed, partial EMT states in which individual cancer cells express both epithelial and mesenchymal features have been described *in situ* or *ex vivo* [12, 13, 40]. Furthermore, distinct EMT states of cancer cells differ in their capacity to interact with stromal cells and these different interactions might actually stabilize respective cellular states [11]. Here, we report our in-depth *ex vivo* analysis of a metachronous lymph node metastasis from OSCC by scRNAseq. We detected a pronounced heterogeneity within the lesion with tumor cell populations characterized by specialized gene expression programs reflecting distinct states within the EMP continuum, metabolic adaptations, and cell cycle. Based on their mRNA expression profiles, these subpopulations form seven clusters that branch towards either epithelial differentiation, partial EMT, metabolic adaptation or cell cycle. Metastatic OSCC cells in or close to the trunk of the trajectory express genes indicating an adaptation to hypoxic and hypoglycemic conditions. While partial EMT is associated with stress response, immune-modulation and higher fraction of proliferating cells, epithelial differentiation correlate with higher cytokeratin and kallikrein expression. Puram *et al.* reported an epithelial differentiation expression program characterized by cytokeratin, kallikrein and S100A8/9 gene expression linked to a senescence-associated phenotype; thus, explaining our observations of less proliferating cells within the epithelial differentiating cluster [12, 13].

Gene expression signatures from the EMTome database reflect the EMP dynamics as they either correlate with the partial EMT state or epithelial differentiation observed in our study (Supplementary Figure 5C) [10]. Indeed, the observation of several EMT states within one lesion emphasizes the high degree of plasticity that might be needed for metastatic outgrowth [16]. The partial EMT state described by Puram *et al.* for OSCC tumors is similar to the partial EMT described here. However, their study may reflect a lower degree of cellular heterogeneity within samples compared to the presented sample probably due to lower numbers of sequenced cells per sample [12]. Common genes that characterize the partial EMT state include, for example, *MMP10*, *TGFBI*, *PDPN* and several cytokeratins; all of these may modulate the extracellular matrix and therefore allow for increased motility and invasiveness of cancer cells. In addition, the partial EMT phenotype is likely to be associated with tumor growth, as this population has a higher proportion of proliferating cells than the more epithelial differentiated cells.

We could demonstrate that besides the previously found SNAI2 expression also ZEB1 is active within OSCC cells although we could not detect ZEB1 mRNA expression [12]. High activity of ZEB1 is usually associated with EMT and hence downregulation of epithelial markers as E-Cadherin (*CDH1*) [41]. However, while ZEB1 acts mostly as transcriptional repressor, it can also induce the keratinocyte differentiation marker *SPRR1B* in response to IL-1β and IFN-γ [42]. *SPRR1B* is highly expressed within epithelial differentiating cells, but not in cells with partial EMT states, which may explain the inferred ZEB1 activity. Further, ZEB1, SNAI1 and TWIST1, which are involved in *TGF-β* induced EMT, are differentially active across EMP [28]. EMP within the analyzed sample is likely regulated by *TGF-β*, which is not only expressed in the microenvironment, but also may interact with highly expressed *EGFR* in tumor cells [43, 44]. However, hypoxia-associated mechanisms of EMT induction through HIF1α might also play a role [32, 33]. The hypoxia-responding cell population in cluster 2 showed upregulation of EMP-related genes as, e.g., *SERPINE1*, *LAMB3* and *TIMP3* [12]. Previous studies also showed that *LAMB3* is upregulated under hypoxic conditions; notably, *LAMB3* expression is associated with nodal metastasis and decreased survival [45, 46].

Furthermore, the adaptation of cancer cells to low glucose conditions is evidenced by gene expression patterns towards higher amino acid metabolism fueling into glycolysis. However, while previous studies also demonstrated that metabolism-related gene expression is upregulated in OSCC cells, the activity of specific metabolic pathways is highly heterogenous between patients [47]. Thus, further studies are needed to fully understand the heterogeneity of hypoxia-response and metabolic adaptations in OSCC, including direct measurements of metabolite concentrations.

A single metastasis sample represents a snapshot at a specific timepoint of tumor evolution. Therefore, the presented scRNAseq-based trajectory reflects the developmental relationships between tumor cell populations rather than passing through evolutional processes. Nevertheless, RNA velocity allows the extraction of short-term, directed dynamic information from scRNAseq data by linking the measurements to the underlying kinetics of gene expression [48]. Here, we found a clear direction from cluster 4 to cluster 5 and from cluster 6 to cluster 7, but no clear direction within the remaining populations. Hence, we observed ongoing epithelial differentiation and proliferation but cannot conclude whether malignant cells differentiated into heterogenous populations within the metastasis or earlier within the primary tumor. Still, the inferred CNV gains indicate that epithelial differentiating cells formed later than partial EMT cells and therefore reflect an undergone MET. However, partial EMT cells probably drive tumor progression as epithelial differentiating cells may reflect a senescent phenotype [13]. We further hypothesize that having both mesenchymal and epithelial features provides higher plasticity in partial EMT cells, supporting that these cells maybe differentiated into the observed epithelial or metabolically adapted phenotypes. Alternatively, if the cells constituting the respective clusters developed within the primary tumor, multiple individual disseminated tumor cell clones reflecting this heterogeneity would have been necessary to constitute the metastasis.

Since cancer evolution constitutes a formidable obstacle to successful treatment, further studies to unravel the evolutionary processes of tumor cells are required and already arising [49]. Integrating data from transcriptomes, genomes, and posttranslational modifications - best

on single-cell level and within the same individual – should provide a better overall picture of tumor evolution. Ideally, this would also include disseminated and circulating tumor cells enriched from blood and tumor draining lymph nodes, filling missing pieces in the metastatic cascade. Those insights would help in understanding the role of EMP in locoregional metastasis development and treatment response thereby supporting clinicians in their treatment decisions, e.g., with biomarkers indicating the risk of metastases and recurrences.

Methods

Patient information

The patient's tumor formula after primary tumor removal and histopathological work-up was the following: pT1 pNx pMx R0 L0 V0 Pn1, ICD-O 8070/3, horizontal tumor size 16 mm, infiltration depth: 6 mm, minimum resection margin: 5 mm. After neck dissection, the biggest of the lymph node metastases was divided with one half for histological analysis and the other was temporarily stored for transport from Düsseldorf to Essen in tissue storage solution (Miltenyi Biotec, Bergisch Gladbach, Germany) at 4°C and immediately processed for scRNAseq. Written informed consent was obtained and the Ethics Committee of the Medical Faculty of the Heinrich-Heine-University Düsseldorf (#3090) approved the study. All procedures performed in studies involving human participants are in accordance with the ethical standards of the institutional and research committee and with the Helsinki Declaration.

Histology

The FFPE tissue was cut into 4µm thick slices for staining. Masson Goldner Trichrom staining was performed according to the manufacturers protocol using the Kit from Morphisto (Catalog number 12043, Morphisto GmbH, Offenbach am Main, Germany) including Weigert's hematoxylin, acid-fuchsin-ponceau-azophloxin, phosphomolybdic acid Orange G, Light Green 0.2% and 10% acetic acid. Hematoxylin and Eosin staining was performed using Mayer's hematoxylin (catalog number 1.09249.0500, Merck) and Eosin Y solution (catalog number X883.2, Roth). The staining was performed as following: Eosin Y working solution was created by adding 50 µL acetic acid to 100ml Eosin Y solution and subsequent diluting 1:6 with water. Mayer's hematoxylin was filtrated before use. The staining protocol is as following: 20 min at 65°C, 2x7 min xylene, 2 min 100% ethanol, 2 min 96% ethanol, 2 min 70% ethanol, 2 min VE-H_2O, 3 min filtrated Mayer's hematoxylin, 3 sec 0.1% HCl, washing with tap water until solution is clear, 3-4 min bluing in tap water, 15 sec. fresh Eosin working solution, washing in VE water until solution is clear, shortly 70% ethanol, 2 min 96% ethanol, 2 min 100% ethanol, 2x5 min

xylene. Whole-slide imaging was performed using Zeiss Axioscan 7 and 10x magnification (Carl Zeiss Microscopy Deutschland GmbH, Oberkochen, Germany).

Single cell RNA library preparation and sequencing

The native lymph node metastasis was immediately dissociated into single cell suspension using the gentleMACS Dissociator (catalog nr.: 130-093-235, Miltenyi Biotec, Bergisch Gladbach, Germany) with program “h_tumor_01”, followed by 2x program “h_tumor_02” in 4.7 ml RPMI 1640 (catalog No. P04-16500 , PAN-Biotech) and an enzyme mix consisting of 200µl Enzyme H, 100 µl Enzyme R and 25µl Enzyme A (catalogue #130-095-929 Miltenyi Biotec). Cells were washed three times with 0.05% bovine serum albumin (BSA) phosphate-buffered saline (PBS) and filtered with a 100 µm cell strainer. 10,000 cells of the single cell suspension were then barcoded and processed with the 10x Genomics Chromium v2.0 platform (10x Genomics, Leiden, Netherlands) using the Chromium Single Cell 5’ Library & Gel Bead Kits (1000006) and Library Construction Kit (1000020) of version 2. Library preparation was performed according to the manufacturers protocol and subsequently sequenced on an Illumina HiSeq 4000 (Illumina, Berlin, Germany) at the Genomics and Proteomics Core Facility of the german cancer research center (DKFZ, Heidelberg, Germany). Sequencing was performed in paired-end with 26 cycles on the forward read and 74 cycles on the reverse read. A total of around 353 million reads were sequenced.

Single-cell RNA sequencing analysis

Preprocessing, cell filtering, normalization, clustering

Processing from FASTQ files with sequenced reads towards the unfiltered count matrix (barcodes x genes) was performed using Cellranger Software Suite version 3.1.0 and the human reference genome build GRCh38, downloaded from 10x Genomics in version 3.0.0. Cells were identified from the barcodes by evaluation of four quality criteria: (1) number of unique molecular identifiers (UMIs), (2) number of genes, (3) percentage of mitochondrial gene expression and (4) number of expressed housekeeper genes, derived from Tirosh et al. without the mouse gene PRPS1L3 [50]. Evaluating these quality criteria and their relationship was

inspired by the best practices of Theis *et al.* [51] (Supplementary Figure 8A, B). We defined cells by having a minimum number of 500 UMIs and excluded cells with mitochondrial gene expression above 10%. The filtered count matrix (cells x genes) was further processed using Seurat version 4.0.1 and R version 4.0.5. Visualization was performed using ggplot2 version 3.3.3 and ComplexHeatmap version 2.6.2 [52, 53]. Cells were normalized using SCtransform algorithm and the 3000 most variable genes were selected for principal component analysis (PCA) using a single value decomposition approximation implemented in Seurat [54]. Based on the variance explained by each PC we chose the first 20 PCs for UMAP visualization and clustering based on a K-nearest neighbor (KNN) graph as implemented in Seurat using resolution 0.8 for all subsets and resolution 1.2 for the subset with fibroblasts and endothelial cells (Supplementary Figure 8C). UMAPs colored by specific gene expression were ordered by expression values.

For downstream analysis, we subsetted the dataset only for specific cell types and repeated normalization, variable feature selection, PCA and UMAP. Inside the malignant cell cluster, we observed cells expressing typical fibroblast-associated genes and cells expressing typical macrophage-associated genes that were subsequently excluded from tumor-specific analysis (in total 40 cells). Similarly, we found 42 tumor cells within the fibroblast and endothelial subset and removed these from analysis.

Differential gene expression and gene set analysis

Cell types were identified using marker gene expression and differential gene expression (DGE) of clusters that were merged based on their proximity in the UMAP (Supplementary Figure 2). Malignant cells were additionally identified by inferred CNVs (Figure 1D, Supplementary Figure 2E). Cell cycle phases were determined using the CellCycleScoring function implemented in Seurat utilizing the S- and G2M-Phase genes defined by Tirosh et al. [50]. DGE was performed using the FindAllMarkers function implemented in Seurat, based on log-normalized data using NormalizeData function and scale factor of 10,000, using MAST version 1.16.0 and comparing each cell cluster with all other cells of the respective subset.

Only genes that were at least expressed in 10% of all cells within a group were considered. Genes were included in epithelial and partial EMT gene expression signatures if the log2 foldchange difference between cluster 3+6 and cluster 4+5 is greater or lesser than 1, respectively. Gene set enrichment analysis (GSEA) was performed using the “fgsea” R package version 1.16.0, log2 foldchanges from differential expression and gene ontology biological processes (GO:BP,C5 v7.1) as well as hallmark gene sets (H, v7.1) downloaded from MSigDB database [20, 55, 56]. Gene sets were included if they have at least 15 genes or at maximum 500 genes within the gene set using 10,000 permutations. Gene sets included in Figure 3A represent the following gene ontology or hallmark gene sets: GO_ALPHA_AMINO_ACID_METABOLIC_PROCESS, GO_RESPONSE_TO_STARVATION, GO_RESPONSE_TO_TYPE_I_INTERFERON, GO_REGULATION_OF_CELLULAR_RESPONSE_TO_HEAT, GO_KERATINOCYTE_DIFFERENTIATION, HALLMARK_HYPOXIA, HALLMARK_GLYCOLYSIS, HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION. Gene set variation analysis (GSVA) was performed using the R package GSVA version 1.38.2 with default settings, i.e., gaussian kernel [57]. As input, EMTome signatures, the partial EMT and epithelial differentiation 1 and 2 signature from Puram et al.[12] and the three EMT and the epithelial senescence signatures from Kinker *et al.* were used [13].

Trajectory, CNV, RNA velocity and receptor-ligand interaction inference

Trajectories were inferred using SlingShot version 1.8.0 with malignant cell clusters and the first 20 PCs [58]. Downstream analysis was repeated with additional cell cycle regression by including the S- and G2M-phase scores as regressable variables in SCTransform. CNVs were inferred using the R package inferCNV version 1.6.0 with the not normalized, filtered count matrix including all cells as input and algorithm run in “samples” mode. Inferred CNVs of mitochondrial genes were excluded. RNA velocity was inferred using the VeloCyto python and R package (version 0.6) [48]. First the Loom file that annotates spliced/unspliced reads was created using default options of the VeloCyto python package and gene annotations as used for Cellranger processing. Genes were filtered based on the minimum average expression

magnitude with a threshold of 0.05 for spliced and 0.02 for unspliced reads. Velocity estimates were calculated using as distance the inverse correlation coefficient of the PC embedding correlation matrix, the top/bottom 2% quantiles for gamma fit, 50 neighboring cells projecting 1 deltaT into the future and projected on the UMAP using 200 neighbors and 50 grid points. Transcription factor activity was inferred using the VIPER algorithm (version 1.24.0) and regulons from the DoRothEA database (version 1.2.2) [29-31]. VIPER infers the relative activity by calculating a normalized enrichment score (NES) based on the scaled gene expression of downstream targets of the respective transcription factor. If the NES is strongly negative, the transcription factor is inactive, a strong positive value indicates an active transcription factor and a value close to zero is ambiguous. The investigated transcription factors had the following confidence levels: HIF1A: A, SNAI1: D, SNAI2: C, TWIST1: A, TWIST2: D, ZEB1: B, ZEB2: C. Ligand-Receptor interactions between tumor and stroma were derived using log-normalized RNA counts and the cellphonedb python package (version 2.1.7) in a conda environment (version 4.,10.3) using default options of the statistical_analysis method [38]. CellphoneDB detects interactions by the mean expression values of receptors and ligands within clusters. The program generates a null distribution by permutating cluster annotations and derives a p-value from the correct cluster labels. The p-value can therefore be interpreted as the probability that a ligand-receptor interaction is specific to the two clusters. Only genes that are at least expressed in 10% of the cells were considered as well as the minimum expression of a member when representing a protein complex. Only ligand-receptor interactions with at a p-value below 0.05 between clusters were considered in quantification. Hierarchical clustering in the heatmaps was performed using euclidean distance and ward.D2 method unless otherwise noted.

Supplementary Information

Supplementary Figure 1: Histology of OSCC lymph node metastasis subjected to multiplexed scRNAseq.

Supplementary Figure 2: Cell type identification by differential gene expression and inferred CNVs.

Supplementary Figure 3: Inferred CNVs of OSCC cells with epithelial genes accumulate in close genomic proximity.

Supplementary Figure 4: Expression of EMP-related genes across the trajectory and selected, enriched gene sets.

Supplementary Figure 5: Partial EMT and epithelial differentiating gene expression signatures are comparable to previously published EMT signatures.

Supplementary Figure 6: Heterogenous tumor populations show different fractions of proliferating cells after cell cycle regression.

Supplementary Figure 7: Fibroblast marker expression and Interactions with the tumor microenvironment.

Supplementary Figure 8: Cell Identification from sequenced barcodes and selecting principal components from normalized gene expression matrices.

Supplementary Table 1: Results from gene set enrichment analysis of differentially expressed genes using GO:BP and HM annotations from MSigDB. (.xlsx)

Supplementary Table 2: Results from interaction analysis using cellphonedb. (.xlsx)

Data availability

The processed datasets generated and analysed during the current study are available in the Gene Expression Omnibus (GEO) database with accession id GSE195655. The raw FASTQ files are available upon request due to privacy reasons.

Code availability

Code used for analysis is available at https://github.com/sci-kai/single_cell_EMP.

Acknowledgements

We thank Önder Bozdogan, Translational Skin Cancer Research, German Cancer Consortium (DKTK), Essen, Germany, for his histological expertise and Linda Kubat, Department of Dermatology, University Medicine Essen, Germany, for helping with microscopy. We thank Christina H. Scheel, Skin Cancer Center, Department of Dermatology, Ruhr-University Bochum, Germany, for her helpful remarks to the manuscript. We further thank the DKFZ Genomics and Proteomics Core Facility for providing Illumina HiSeq Sequencing and related services.

Author Contributions

All authors approved the final version and agreed to be accountable for all aspects of the work. KH was responsible for data analysis, data curation, visualization, interpretation, conceptualization, drafting the manuscript and updating revisions. IS, LP and PG contributed by performing wet lab experiments including single cell dissociation, single cell sequencing and library preparation. FF established and performed histological staining. CS contributed with sample acquisition, interpretation and revising the manuscript. NS contributed by manuscript reviewing and interpretation. JG further contributed to data analysis, interpretation, and visualization. JCB was contributing by interpretation, analysis guidance and manuscript review and editing. Further, he was responsible for conceptualization and project administration including supervision and funding acquisition.

Competing Interests (COI)

JCB is receiving speaker's bureau honoraria from Amgen, Pfizer, MerckSerono, Recordati and Sanofi, is a paid consultant/advisory board member/DSMB member for Boehringer Ingelheim, InProTher, MerckSerono, Pfizer, 4SC, and Sanofi/Regeneron. His group receives research

grants from Bristol-Myers Squibb, Merck Serono, HTG, IQVIA, and Alcedis. None of these activities are related to the present manuscript. The other authors including K.H., C.S., L.P., F.F., P.G., J.G., N.S., I.S. declare no conflict of interest.

Funding

The project was funded by the German Cancer Consortium (DKTK) site budget OE 0460 ED003 and the federal ministry of education and research (03VP01062).

References

1 Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers* 2020; 6: 92; https://doi.org/10.1038/s41572-020-00224-3.

2 Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; 71: 209-249; https://doi.org/10.3322/caac.21660.

3 Cramer JD, Burtness B, Le QT, Ferris RL. The changing therapeutic landscape of head and neck cancer. *Nat Rev Clin Oncol* 2019; 16: 669-683; https://doi.org/10.1038/s41571-019-0227-z.

4 Dongre A, Weinberg RA. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nat Rev Mol Cell* 2019; 20: 69-84; https://doi.org/10.1038/s41580-018-0080-4.

5 Nieto MA, Huang Ruby Y-J, Jackson Rebecca A, Thiery Jean P. EMT: 2016. *Cell* 2016; 166: 21-45; https://doi.org/10.1016/j.cell.2016.06.028.

6 Tsai JH, Donaher JL, Murphy DA, Chau S, Yang J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* 2012; 22: 725-736; https://doi.org/10.1016/j.ccr.2012.09.022.

7 Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M *et al*. Guidelines and definitions for research on epithelial–mesenchymal transition. *Nat Rev Mol Cell* 2020; 21: 341-352; https://doi.org/10.1038/s41580-020-0237-9.

8 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: 1389-1398; https://doi.org/10.1038/s41588-019-0489-5.

9 Cook DP, Vanderhyden BC. Context specificity of the EMT transcriptional response. *Nat Commun* 2020; 11: 2142; https://doi.org/10.1038/s41467-020-16066-2.

10 Vasaikar SV, Deshmukh AP, den Hollander P, Addanki S, Kuburich NA, Kudaravalli S *et al*. EMTome: a resource for pan-cancer analysis of epithelial-mesenchymal transition genes and signatures. *Br J Cancer* 2021; 124: 259-269; https://doi.org/10.1038/s41416-020-01178-9.

11 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: 463-468; https://doi.org/10.1038/s41586-018-0040-3.

565 12 Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S *et al*. Single-Cell Transcriptomic
566 Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* 2017;
567 171: 1611-1624.e1624; https://doi.org/10.1016/j.cell.2017.10.044.

568
569 13 Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM *et al*. Pan-cancer single-
570 cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet* 2020; 52: 1208-
571 1218; https://doi.org/10.1038/s41588-020-00726-6.

572
573 14 Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB *et al*. Mapping lung
574 cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution.
575 *Nat Commun* 2019; 10: 5587; https://doi.org/10.1038/s41467-019-13441-6.

576
577 15 Jolly MK, Tripathi SC, Jia D, Mooney SM, Celiktas M, Hanash SM *et al*. Stability of the hybrid
578 epithelial/mesenchymal phenotype. *Oncotarget* 2016; 7;
579 https://doi.org/10.18632/oncotarget.8166.

580
581 16 Pal A, Barrett TF, Paolini R, Parikh A, Puram SV. Partial EMT in head and neck cancer biology: a
582 spectrum instead of a switch. *Oncogene* 2021; 40: 5049-5065;
583 https://doi.org/10.1038/s41388-021-01868-5.

584
585 17 Lambert AW, Pattabiraman DR, Weinberg RA. Emerging Biological Principles of Metastasis. *Cell*
586 2017; 168: 670-691; https://doi.org/10.1016/j.cell.2016.11.037.

587
588 18 Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM *et al*. Multimodal Analysis of
589 Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* 2020; 182: 497-
590 514.e422; https://doi.org/10.1016/j.cell.2020.05.039.

591
592 19 Kürten CHL, Kulkarni A, Cillo AR, Santos PM, Roble AK, Onkar S *et al*. Investigating immune and
593 non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nat*
594 *Commun* 2021; 12: 7338; https://doi.org/10.1038/s41467-021-27619-4.

595
596 20 Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
597 Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015; 1: 417-425;
598 https://doi.org/10.1016/j.cels.2015.12.004.

599
600 21 Parikh A, Shin J, Faquin W, Lin DT, Tirosh I, Sunwoo JB *et al*. Malignant cell-specific CXCL14
601 promotes tumor lymphocyte infiltration in oral cavity squamous cell carcinoma. *J Immunother*
602 *Cancer* 2020; 8: e001048; https://doi.org/10.1136/jitc-2020-001048.

603
604 22 Shurin GV, Ferris R, Tourkova IL, Perez L, Lokshin A, Balkir L *et al*. Loss of New Chemokine
605 CXCL14 in Tumor Tissue Is Associated with Low Infiltration by Dendritic Cells (DC), while
606 Restoration of Human CXCL14 Expression in Tumor Cells Causes Attraction of DC Both In Vitro
607 and In Vivo. *J Immunol* 2005; 174: 5490; https://doi.org/10.4049/jimmunol.174.9.5490.

608

609 23 Guo D-D, Xie K-F, Luo X-J. Hypoxia-induced elevated NDRG1 mediates apoptosis through
610 reprograming mitochondrial fission in HCC. *Gene* 2020; 741: 144552;
611 https://doi.org/10.1016/j.gene.2020.144552.

612
613 24 Ivan M, Kaelin WG. The EGLN-HIF O2-Sensing System: Multiple Inputs and Feedbacks. *Mol Cell*
614 2017; 66: 772-779; https://doi.org/10.1016/j.molcel.2017.06.002.

615
616 25 Gatenby RA, Gillies RJ. Why do cancers have high aerobic glycolysis? *Nat Rev Cancer* 2004; 4:
617 891-899; https://doi.org/10.1038/nrc1478.

618
619 26 Laplante M, Sabatini DM. mTOR signaling at a glance. *J Cell Sci* 2009; 122: 3589-3594;
620 https://doi.org/10.1242/jcs.051011.

621
622 27 DeBerardinis Ralph J. Serine Metabolism: Some Tumors Take the Road Less Traveled. *Cell*
623 *Metab* 2011; 14: 285-286; https://doi.org/10.1016/j.cmet.2011.08.004.

624
625 28 Stemmler MP, Eccles RL, Brabletz S, Brabletz T. Non-redundant functions of EMT transcription
626 factors. *Nat Cell Biol* 2019; 21: 102-112; https://doi.org/10.1038/s41556-018-0196-y.

627
628 29 Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and
629 integration of resources for the estimation of human transcription factor activities. *Genome*
630 *Res* 2019; 29: 1363-1375; https://doi.org/10.1101/gr.240663.118.

631
632 30 Holland CH, Tanevski J, Perales-Patón J, Gleixner J, Kumar MP, Mereu E *et al*. Robustness and
633 applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data.
634 *Genome Biol* 2020; 21: 36; https://doi.org/10.1186/s13059-020-1949-z.

635
636 31 Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH *et al*. Functional characterization
637 of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet*
638 2016; 48: 838-847; https://doi.org/10.1038/ng.3593.

639
640 32 Tam SY, Wu VWC, Law HKW. Hypoxia-Induced Epithelial-Mesenchymal Transition in Cancers:
641 HIF-1α and Beyond. *Front Oncol* (Review) 2020; 10; https://doi.org/10.3389/fonc.2020.00486.

642
643 33 Saxena K, Jolly MK, Balamurugan K. Hypoxia, partial EMT and collective migration: Emerging
644 culprits in metastasis. *Transl Oncol* 2020; 13: 100845-100845;
645 https://doi.org/10.1016/j.tranon.2020.100845.

646
647 34 Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM *et al*. A framework for
648 advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer* 2020; 20: 174-
649 186; https://doi.org/10.1038/s41568-019-0238-1.

650
651 35 Abaurrea A, Araujo AM, Caffarel MM. The Role of the IL-6 Cytokine Family in Epithelial–
652 Mesenchymal Plasticity in Cancer Progression. *Int J Mol Sci* 2021; 22: 8334;
653 https://doi.org/10.3390/ijms22158334.

36 Rudnicki MA, Schnegelsberg PNJ, Stead RH, Braun T, Arnold H-H, Jaenisch R. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* 1993; 75: 1351-1359; https://doi.org/10.1016/0092-8674(93)90621-V.

37 Murray PJ. Macrophage Polarization. *Annu Rev Physiol* 2017; 79: 541-566; https://doi.org/10.1146/annurev-physiol-022516-034339.

38 Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc* 2020; 15: 1484-1506; https://doi.org/10.1038/s41596-020-0292-x.

39 Williams ED, Gao D, Redfern A, Thompson EW. Controversies around epithelial–mesenchymal plasticity in cancer metastasis. *Nat Rev Cancer* 2019; 19: 716-732; https://doi.org/10.1038/s41568-019-0213-x.

40 Derynck R, Weinberg RA. EMT and Cancer: More Than Meets the Eye. *Dev Cell* 2019; 49: 313-316; https://doi.org/10.1016/j.devcel.2019.04.026.

41 Sánchez-Tilló E, Lázaro A, Torrent R, Cuatrecasas M, Vaquero EC, Castells A *et al*. ZEB1 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1. *Oncogene* 2010; 29: 3490-3500; https://doi.org/10.1038/onc.2010.102.

42 Li S, Gallup M, Chen YT, McNamara NA. Molecular mechanism of proinflammatory cytokine-mediated squamous metaplasia in human corneal epithelial cells. *Invest Ophthalmol Vis Sci* 2010; 51: 2466-2475; https://doi.org/10.1167/iovs.09-4677.

43 Dave N, Guaita-Esteruelas S, Gutarra S, Frias À, Beltran M, Peiró S *et al*. Functional Cooperation between Snail1 and Twist in the Regulation of ZEB1 Expression during Epithelial to Mesenchymal Transition*. *J Biol Chem* 2011; 286: 12024-12032; https://doi.org/10.1074/jbc.M110.168625.

44 Zhao Y, Ma J, Fan Y, Wang Z, Tian R, Ji W *et al*. TGF-β transactivates EGFR and facilitates breast cancer migration and invasion through canonical Smad3 and ERK/Sp1 signaling pathways. *Mol Oncol* 2018; 12: 305-321; https://doi.org/10.1002/1878-0261.12162.

45 Skowronki K, Andrews J, Rodenhiser DI, Coomber BL. Genome-Wide Analysis in Human Colorectal Cancer Cells Reveals Ischemia-Mediated Expression of Motility Genes via DNA Hypomethylation. *PLoS One* 2014; 9: e103243; https://doi.org/10.1371/journal.pone.0103243.

46 Parikh AS, Puram SV, Faquin WC, Richmon JD, Emerick KS, Deschler DG *et al*. Immunohistochemical quantification of partial-EMT in oral cavity squamous cell carcinoma primary tumors is associated with nodal metastasis. *Oral Oncol* 2019; 99: 104458-104458; https://doi.org/10.1016/j.oraloncology.2019.104458.

47 Xiao Z, Dai Z, Locasale JW. Metabolic landscape of the tumor microenvironment at single cell resolution. *Nat Commun* 2019; 10: 3763; https://doi.org/10.1038/s41467-019-11738-0.

48 La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V *et al*. RNA velocity of single cells. *Nature* 2018; 560: 494-498; https://doi.org/10.1038/s41586-018-0414-6.

49 Fennell KA, Vassiliadis D, Lam EYN, Martelotto LG, Balic JJ, Hollizeck S *et al*. Non-genetic determinants of malignant clonal fitness at single-cell resolution. *Nature* 2022; 601: 125-131; https://doi.org/10.1038/s41586-021-04206-7.

50 Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ *et al*. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 2016; 352: 189; https://doi.org/10.1126/science.aad0501.

51 Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* 2019; 15: e8746; https://doi.org/10.15252/msb.20188746.

52 Wickham H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* 2016;

53 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016; 32: 2847-2849; https://doi.org/10.1093/bioinformatics/btw313.

54 Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 2019; 20: 296; https://doi.org/10.1186/s13059-019-1874-1.

55 Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. *bioRxiv* 2021: 060012; https://doi.org/10.1101/060012.

56 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al*. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; 102: 15545; https://doi.org/10.1073/pnas.0506580102.

57 Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinform* 2013; 14: 7; https://doi.org/10.1186/1471-2105-14-7.

58 Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N *et al*. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genom* 2018; 19: 477; https://doi.org/10.1186/s12864-018-4772-0.

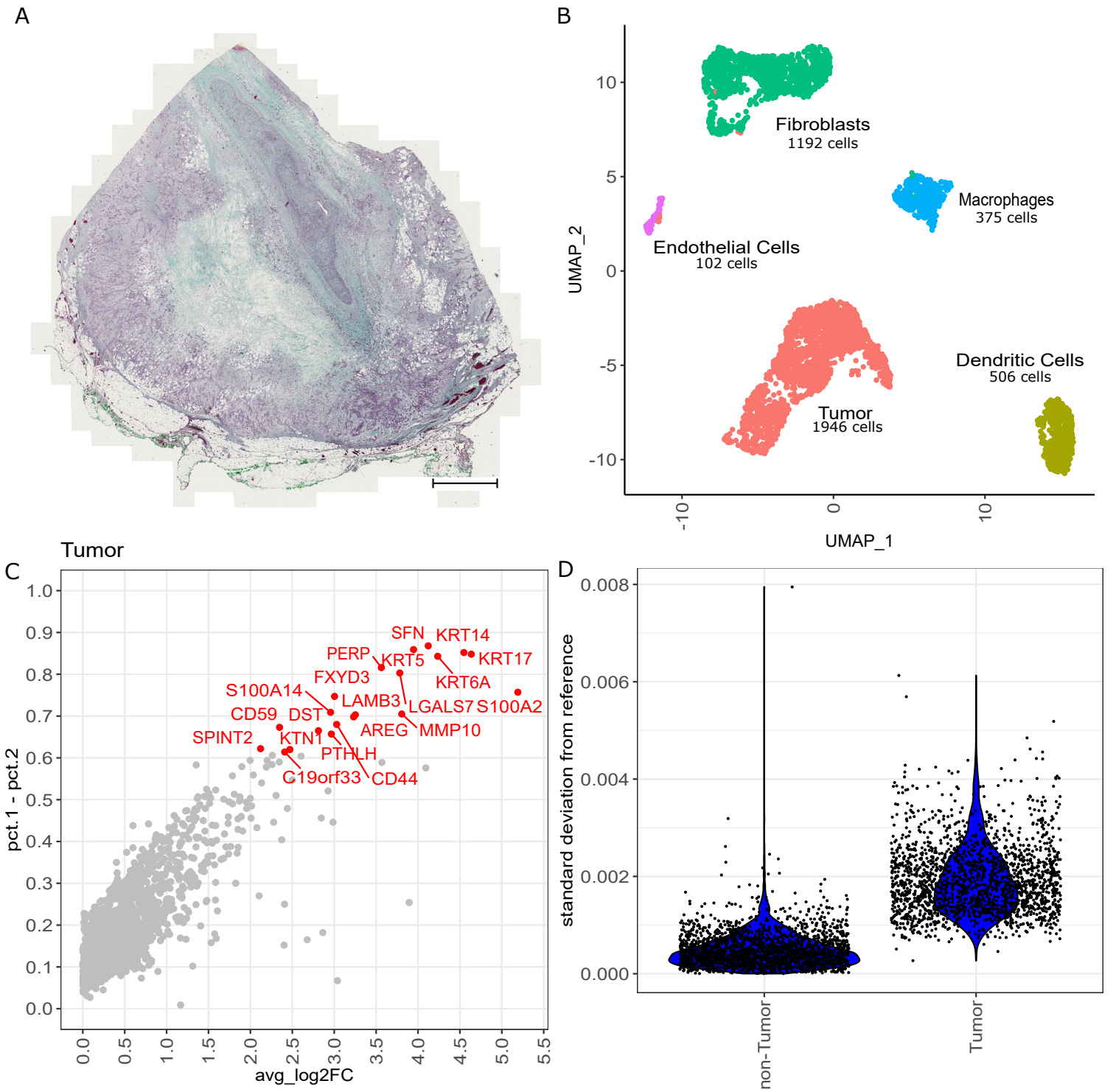

Figure 1 | Morphological and molecular characteristics of an OSCC derived metachronous lymph node metastasis.

(A) Masson-Goldner staining of an FFPE section of the OSCC derived lymph node metastasis subjected to multiplexed scRNAseq. Scale bar = 2 mm. **(B)** UMAP based on scRNAseq data of 4,121 cells isolated from the lesion depicted in A. Cells were clustered using KNN graphs and annotated according to the presumed cell type. **(C)** Average log2 fold change of gene expression (x-axis) and differences in cellular fractions expressing the respective gene (y-axis) between malignant and all other cells. Shown are only genes with positive log2 fold change. Labelled in red are the top 20 DEGs based on the differences in cellular fractions. **(D)** Standard deviation of the log2 values from inferred CNVs for each cell compared between non-malignant and malignant cells.

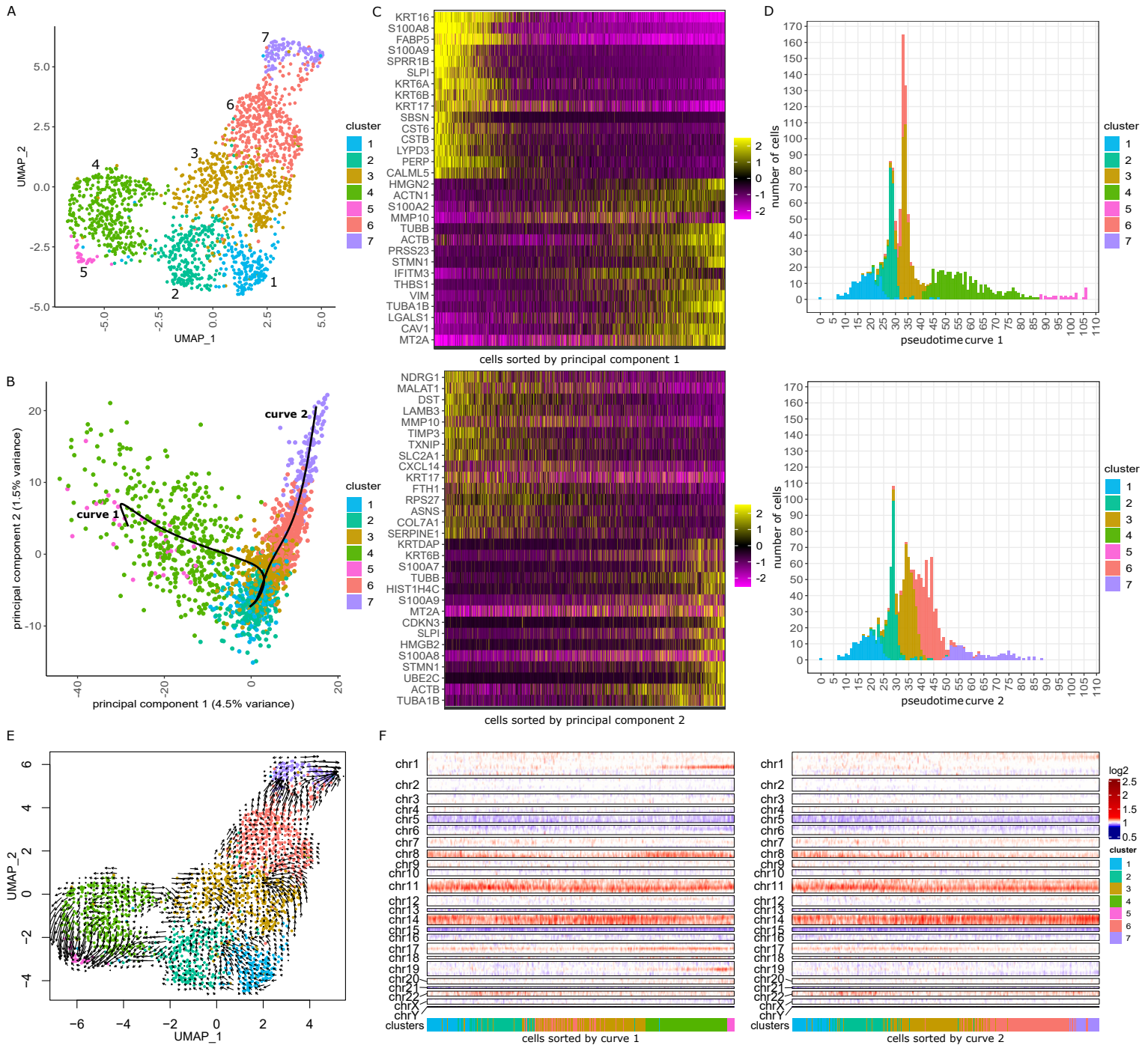

Figure 2 | Malignant cell populations are connected by a branching developmental trajectory.

(A) UMAP based on scRNAseq data of 1,899 malignant cells. Cells were clustered using KNN graphs and clusters are numbered by their appearance in the inferred trajectory. **(B)** PCA based on the same data as A with two principal curves from trajectory inference. Cells are annotated by clusters depicted in A. **(C)** Heatmap of scaled, log-normalized gene expression across cells sorted by principal component 1 (top) and 2 (bottom). Shown are the 15 genes with the highest or lowest PC loadings, respectively. **(D)** Distribution of malignant cells across curve 1 (top) and 2 (bottom). The histogram shows the number of cells with a respective pseudotime value using a bin width of 1, colored by clusters depicted in A. **(E)** Projection of RNA velocity on the UMAP depicted in A. Arrows indicate the extrapolated direction of development, their length the strength of this future development. **(F)** Inferred CNVs across cells (columns) sorted by curve 1 (left) and 2 (right); color code according to the clusters depicted in A. Rows shows genes categorized in chromosomes and ordered by genome position, hence the size of the chromosome reflects the number of detected genes and not the nucleotide length. Mitochondrial genes were excluded.

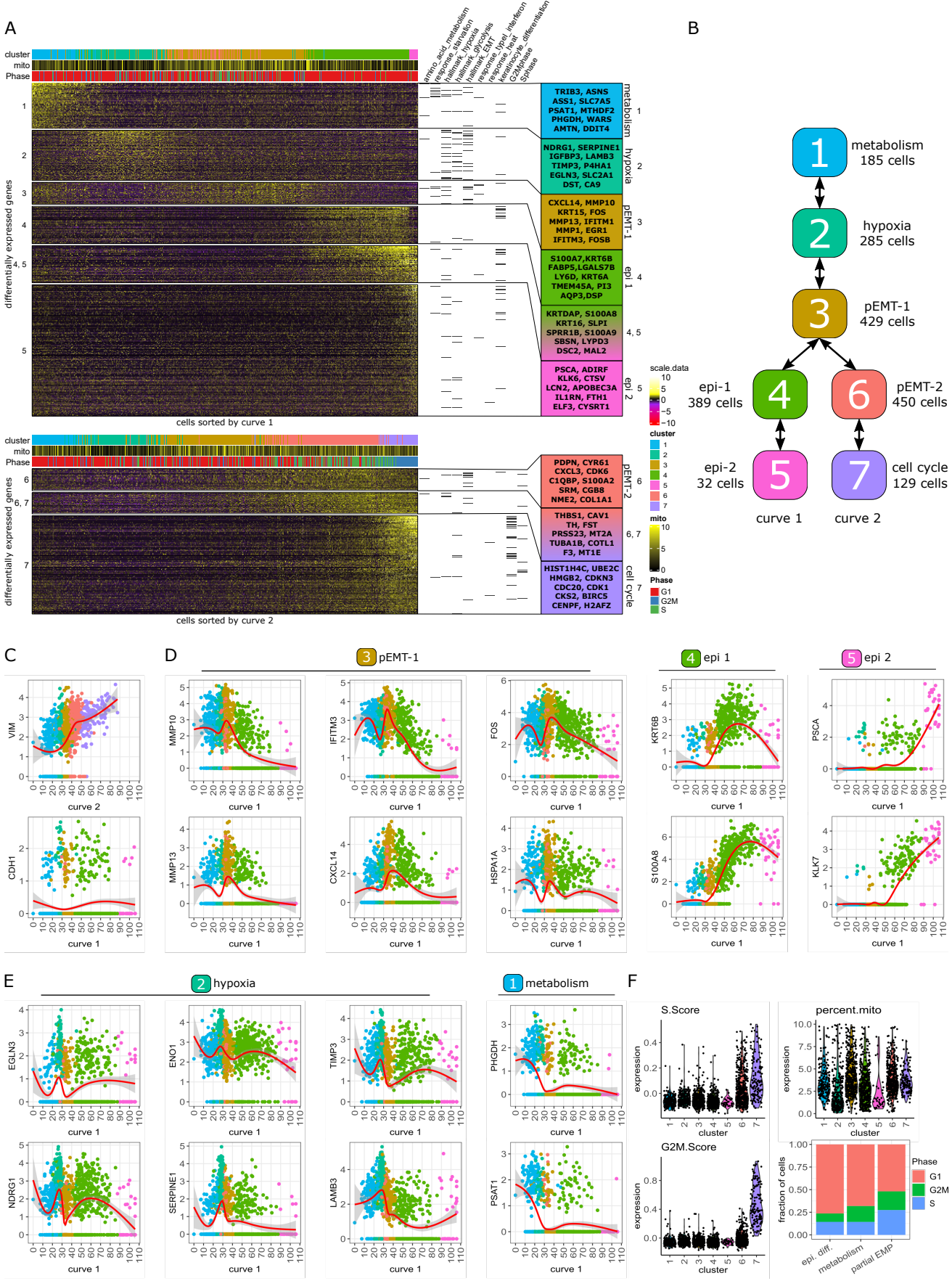

Figure 3 | Heterogeneity of OSCC cells within an individual lesion results from varying degrees of EMP, metabolic adaptations, and distinct cell cycle phases. (A) Heatmap for scaled, log-normalized gene expression of DEGs for all clusters of OSCC cells (see Figure 2A) against all other malignant cells. Shown are DEGs (rows) with a log2 fold change in gene expression above 0.5, ordered from highest log2 fold change (top) to lowest (bottom) for the cells (columns) sorted by pseudotime values for curve 1 or 2 (see Figure 2B). The annotation on top indicates the cluster, percentage of mitochondrial gene expression and cell cycle phase. On the right side of the heatmap is the occurrence of genes in a curated list of enriched gene sets, and in the rectangles the 10 DEGs with highest log2 fold change in the cluster are named. **(B)** Quantification and differential gene expression based functional annotation of OSCC clusters ordered by the branched trajectory. **(C)** Log-normalized expression of vimentin (VIM) and E-cadherin (CDH1) across pseudotime values, color-coded by clusters. Red line (as in D through F) indicates smoothed expression values over the trajectory generated with a general additive model; 95%-confidence intervals are shaded gray. **(D, E)** Log-normalized expression of the indicated DEGs in clusters 3 to 5 (partial EMT-1, epithelial differentiating 1 and 2) as well as 2 and 3 (hypoxia-response and amino acid metabolism) across the pseudotime values. **(F)** Fraction of cells S- and G2M scores (left panels) as well as percentage of UMIs for mitochondrial genes (upper right) within the clusters of OSCC cells. The bottom right panel summarizes the fraction of cells in G1, S and G2M phases for the 3 branches of the trajectory.

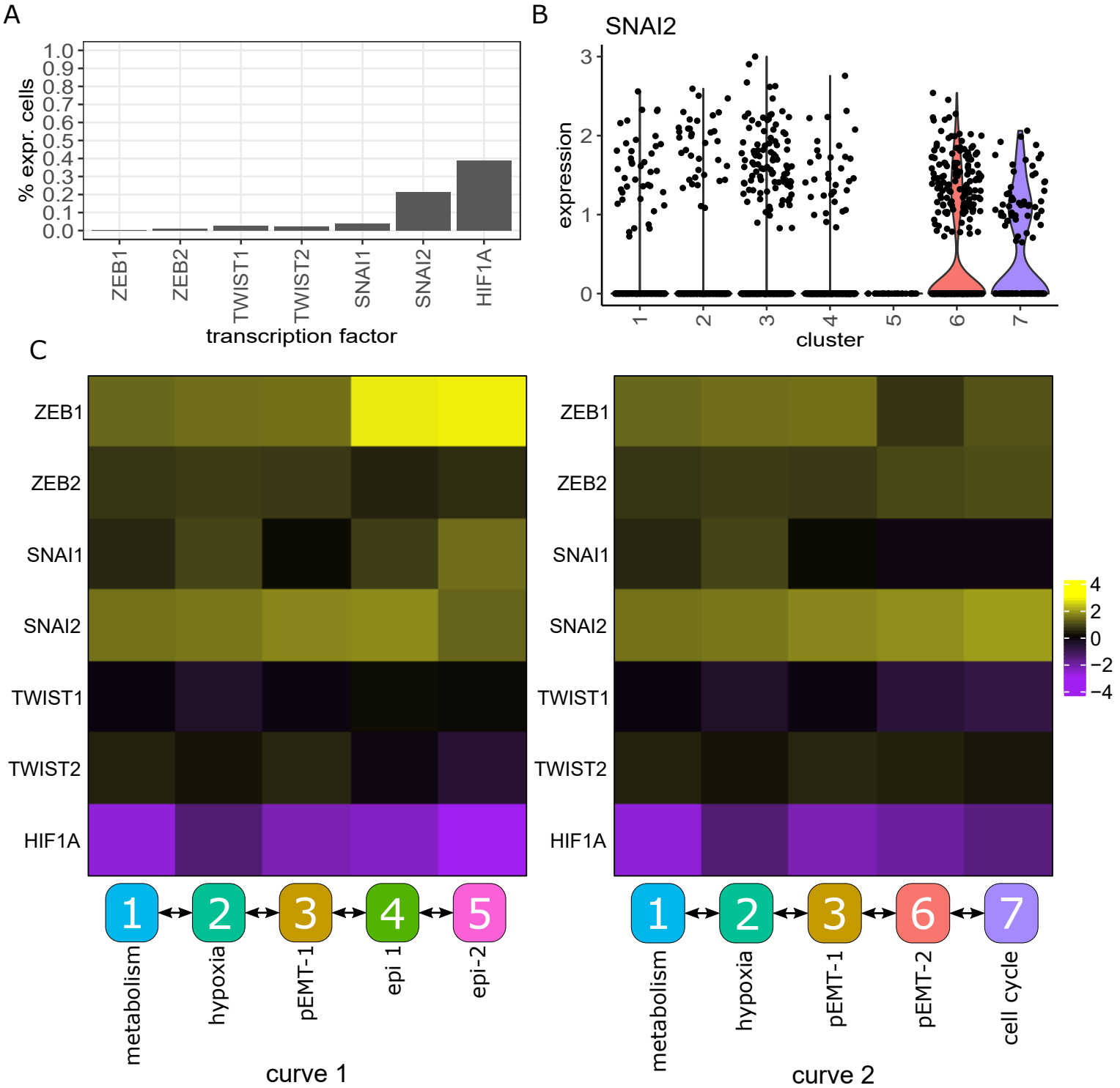

Figure 4 | Transcription factor regulons associated with EMP. (A) Percentage of tumor cells with detectable mRNA expression (more than one UMI) encoding the indicted EMP-related transcription factors and HIF1α. **(B)** Log-normalized mRNA expression of SNAI2 in the clusters of OSCC cells. **(C)** Mean of the inferred activity of the indicated transcription factors for the OSCC cell clusters sorted by curve 1 and 2.

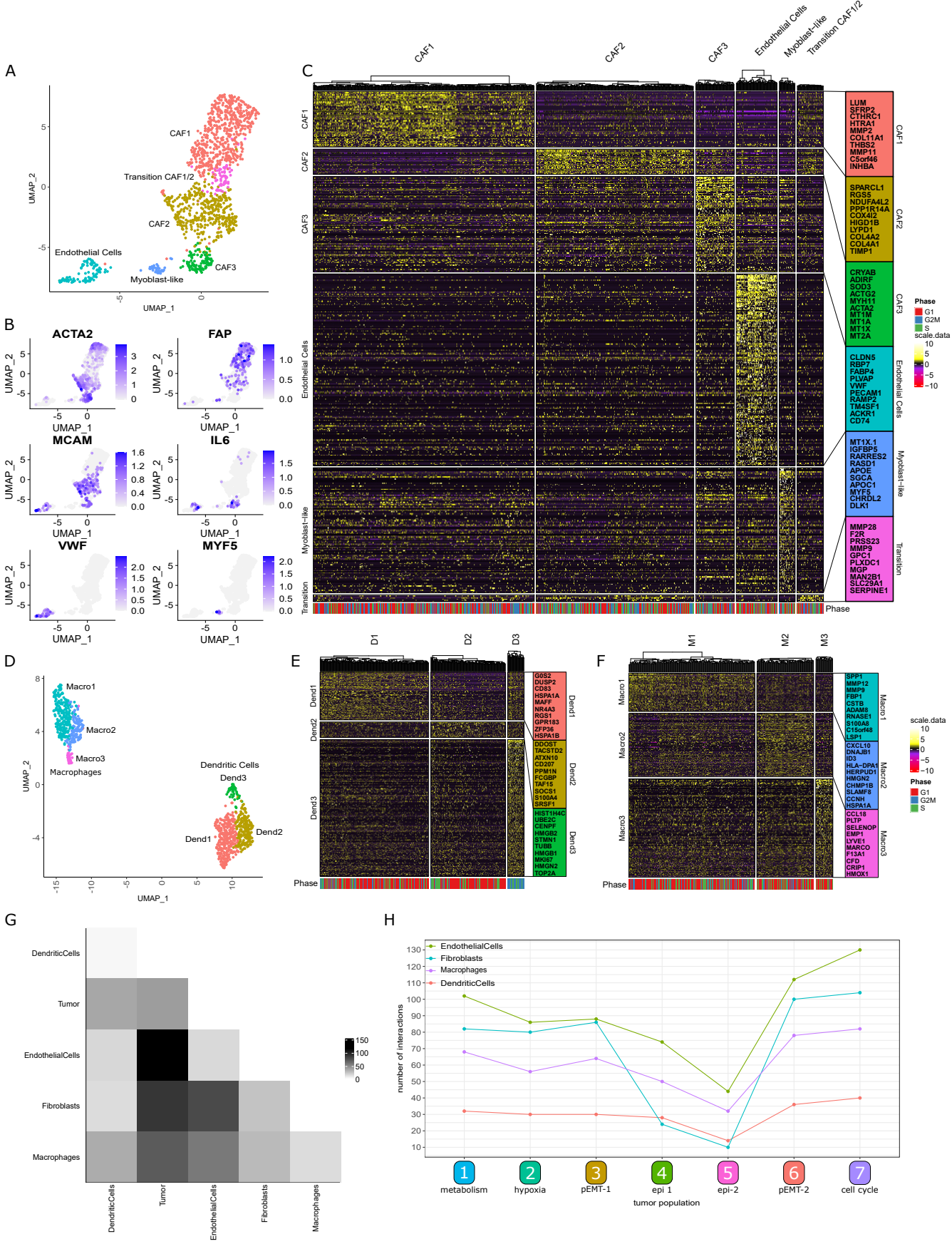

Figure 5 | Characteristics of stromal cells and their interaction with different OSCC cell populations. **(A)** UMAP based on scRNAseq data of 1154 fibroblasts and 101 endothelial cells isolated from the lesion depicted in Figure 1A. Cells were clustered using KNN graphs and annotated according to the presumed cell subtype. **(B)** Feature plots showing the log-normalized expression level of indicated fibroblast, endothelial and myoblast marker genes by an increasing blue shading. **(C)** Heatmap for scaled, log-normalized gene expression of DEGs in hierarchically clustered cells across all clusters depicted in A having a log2 foldchange above 1 with the highest change on the top. In the rectangles on the right the 10 DEGs with highest log2 foldchange in the cluster are named. The annotation at the bottom indicates the cell cycle phase. **(D)** UMAP based on scRNAseq data of 375 macrophages and 506 dendritic cells isolated from the same lesion as above. Clusters are annotated according to the presumed cell subtype. **(E, F)** Heatmap for scaled, log-normalized gene expression of DEGs in hierarchically clustered dendritic cells (E) and macrophages (F) across all clusters depicted in D presented in the same way as in C. **(G)** Quantification of ligand-receptor interactions between different cell types provided as lower triangle heatmap. **(H)** Number of ligand-receptor interactions of all clusters of OSCC cells with the indicated stromal cells. For G and H only interactions with ligand and receptor expression in at least 10% of cells in the respective cell type and with a p-value below 0.05 were considered.