Mesenchymal-epithelial transition in lymph node metastases of oral squamous cell carcinoma is accompanied by ZEB1 expression

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- and Maxillofacial 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

Competing Interests

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.

Abstract

Oral squamous cell carcinoma (OSCC) frequently metastasizes to the regional lymph nodes but only occasionally beyond. Initial phases of metastasis are associated with an epithelial-mesenchymal transition (EMT), the consolidation phase is associated with mesenchymal-epithelial transition (MET). This dynamic is referred to as epithelial-mesenchymal plasticity (EMP).

To capture EMP-associated heterogeneity in OSCC, we performed single-cell RNA sequencing of 5 primary tumors and 9 lymph node metastases and analyzed publicly available data of 9 additional primary tumors for a total of 7,263 carcinoma cells. Focusing initially on one metastasis to avoid inter-patient heterogeneity as confounder, we identified OSCC cells expressing genes characteristic of different epithelial and partial EMT stages. RNA velocity information together with the increase in inferred copy number variations indicated a progressive trajectory towards epithelial differentiation. Extension to all samples revealed a less stringent but essentially similar pattern. Interestingly, the epithelial differentiation in OSCC metastases is associated with higher activity of EMT-activator ZEB1, which was confirmed by co-expression of ZEB1 and cornifin-B by immunohistochemistry.

In summary, we show that EMP enables different EMT and epithelial phenotypes of OSCC cells, which are endowed with capabilities essential for the different stages of the metastatic process, including maintenance of cellular integrity.

41 **Frequently used abbreviations**

42 CNV	Copy number variation
43 DC	Dendritic cell
44 DEG	Differentially expressed genes
45 EC	Endothelial cell
46 ECM	Extracellular matrix
47 EMP	Epithelial-mesenchymal plasticity
48 EMT	Epithelial-mesenchymal transition
49 FFPE	Formalin-fixed, paraffin-embedded
50 GSEA	Gene set enrichment analysis
51 GSVA	Gene set variation analysis
52 HNSCC	Head and neck squamous cell carcinomas
53 HPV	Human papillomavirus
54 HTO	Hashtag oligo
55 IHC	Immunohistochemistry
56 MET	Mesenchymal-epithelial transition
57 MSigDB	Molecular signature database
58 OSCC	Oral cavity squamous cell carcinoma
59 PC	Principal component
60 pEMT	Partial EMT
61 scRNAseq	Single-cell RNA sequencing
62 SNN	Shared-nearest neighbor
63 UMAP	Uniform manifold approximation and projection

64

Introduction

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide, with 890,000 new cases and 450,000 deaths in 2018 [1]. The survival for HNSCC patients has improved modestly over the past decades; however, this improvement is partially attributable to the emergence of human papillomavirus (HPV)-associated HNSCC which has a better prognosis than HPV-negative tumors. One of the HPV-negative HNSCC subtypes is oral cavity squamous cell carcinoma (OSCC) which is mainly associated with tobacco and alcohol abuse [1]. OSCCs are often diagnosed at an early stage owing to the patient's self-identification of the mass lesion and symptoms, still regional lymph node metastases are frequent; thus, surgical removal of primary tumor is accompanied by neck dissection and radiotherapy [2]. Given the morbidity associated with this combined intervention, there is a need to identify molecular biomarkers to predict the presence of lymph node metastases and to prognosticate survival.

Invasion and metastasis becomes possible in many epithelial tumors through an epithelial-mesenchymal transition (EMT), i.e., a reactivation of an embryonic developmental program in which cells acquire migratory and invasive properties [3]. In EMT, epigenetic, transcriptional, and post-translational changes cause epithelial cells to break down the strong homotypic cell-cell junctions and adopt a mesenchymal morphology [4]. It is important to note that EMT should not be understood as a clearly defined process, but rather as many dynamic and complex processes, which may vary depending on tumor entity, stage, and microenvironment [4-6]. Thus, expression of EMT-related genes and their regulating transcription factors is highly heterogeneous even within one cancer entity, between patients, different lesions from one patient and between individual cancer cells within one lesion [4, 7]. Since it is a continuous, dynamic, and reversible process cancer cells can adopt a multitude of intermediate or partial states, e.g., epithelial to more mesenchymal or partial EMT (pEMT) states [5, 8-11]. Therefore, it has recently been recommended that this EMT continuum should rather referred to as epithelial-mesenchymal plasticity (EMP) [4, 10].

92 To capture the EMP-associated heterogeneity of cancers, single-cell analyses appear to be
93 the most appropriate approach, however, to date most EMP-related single-cell studies are
94 based on controlled *in vitro* and *in vivo* experiments [5, 6, 9, 12, 13]. Particularly in HNSCC
95 only few studies scrutinized EMP within freshly isolated tumor samples [8, 14, 15]. Especially
96 noteworthy is the landmark paper by Puram *et al.* They characterized 2,215 malignant cells
97 from 18 patients and discovered several pEMT states with high variability in EMP-related gene
98 expression [8].

99 In the work presented here, we investigated the phenotypic heterogeneity of 5 primary and 9
100 regionally metastatic OSCC lesions isolated from 7 patients using multiplexed single-cell RNA
101 sequencing (scRNAseq). In addition, we used a recently published series of scRNAseq data
102 from primary HNSCC that included 9 OSCC tumors to put our observations on an even broader
103 data base [15]. Our results not only confirm the EMP-associated heterogeneity of cancer cells
104 in primary and metastatic OSCC, but also demonstrate mesenchymal-epithelial transition
105 (MET) in established lymph node metastases. Surprisingly, we observed a high activity of
106 EMT-activator ZEB1 in metastatic OSCC cells with epithelial differentiation, a notion which was
107 confirmed by co-expression of ZEB1 and cornifin-B protein in individual tumor cells.

Results

Single-cell gene expression signatures of tumor cells from a single metastasis show several predominant, but not necessarily exclusive, functional phenotypes.

To avoid inter-patient heterogeneity as a confounding factor, we first focused on the analysis of a single OSCC metastasis to develop hypotheses which would be subsequently tested in the entire cohort. For this, we chose a metachronous lymph node metastasis that was removed one year after the primary tumor, because we assumed that consolidation processes are particularly pronounced in this longer existing metastasis. Multiplexed scRNAseq recovered 4,121 cells that could be assigned to the following cell types: 1,906 (46.8 %) tumor cells, 1,186 (29.1 %) fibroblasts, 507 (12.4 %) dendritic cells (DCs), 375 (9.2 %) macrophages and 102 (2.5 %) endothelial cells (ECs) (Figure 1A). The absence of T- or B-cells was in line with the histology showing completely disrupted lymph node structures and only occasional tumor-infiltrating lymphocytes (Supplementary Figure 1). OSCC cells were identified both by the presence of copy number variations (CNVs) inferred from scRNAseq data as well as the expression of epithelial markers including *S100A2, S100A14*, cytokeratins (*KRT5*, *KRT14*, *KRT17)* and stratifin (*SFN*) (Supplementary Figure 2).

Detailed phenotyping of the cancer cells identified several clusters to which we could assign the following predominant functional phenotypes that differ in their EMT state, immunomodulatory capacity, as well as their response to hypoxia, stress, and metabolic constraints (Figure 1B-D). However, the predominance of a functional phenotype does not preclude additional traits. Specifically, 515 cells exhibited a pEMT phenotype by expressing a mixture of epithelial and mesenchymal genes such as matrix metallopeptidases (*MMPs*), caveolin-1 (*CAV1*) and *CXCL14* that were previously described in pEMT signatures and are enriched in the EMT hallmark gene set from the molecular signatures database (MSigDB) (Supplementary Figure 3A) [8, 16]. In contrast, 385 cells showed higher expression of genes associated with epithelial differentiation such as *S100A7/A8/A9,* the keratinocyte envelope protein cornifin-B (*SPRR1B*), and various cytokeratins (e.g., *KRT6B* and *KRT16*). The EMP-

135 related gene expression patterns are correlating with established signatures from the EMTome
136 database (Supplementary Figure 3B-E) [7]. Interestingly, both gene expression signatures
137 were present in a mixed cluster with 184 cells characterized by the high expression of cell-
138 cycle related genes (despite the fact that we applied cell cycle regression).

139 With respect to cell clusters whose gene expression was not predominately associated with
140 EMP, 268 cells exhibited an immune-regulatory phenotype enriched for genes associated with
141 cytokine-mediated responses and higher expression of the chemokines *CXCL1/2/3/8* and
142 *CCL20* and for 554 cells the gene expression pattern suggests adaptation to environmental
143 factors. In detail, 51 cells had higher expression of transcription factors *FOS* and *JUN*
144 suggesting a stress response. 310 cells can be assumed to respond to hypoxic conditions in
145 the tumor based on the higher expression of *NDRG1* and *EGLN3* both of which are regulated
146 by oxygen levels [17, 18]: *NDRG1* regulates stress response and p53-mediate caspase
147 activation [17], and *EGLN3* has an important role in hypoxia-inducible factor 1 alpha (HIF1α)
148 regulation through prolyl hydroxylation [18]. 193 cells expressed genes associated with amino
149 acid metabolism, starvation response and mTORC1 signaling, i.e., a regulator of mitochondrial
150 metabolism [19]. The upregulated genes *ASNS*, *PSAT1* and *PHGDH* integrate the metabolites
151 of serine and glycine metabolism into glycolysis and therefore fuel glycolysis with amino acids
152 [20]; hence, these cells appear to be adapted to low-glucose conditions.

153 **OSCC cells in lymph node metastases undergo mesenchymal-epithelial transition**

154 We next focused on possible dynamics within the predominantly EMP-related cancer cell
155 clusters using a higher resolved shared-nearest neighbor (SNN) clustering. We defined 4
156 pEMT clusters (pEMT-1 to 4), 4 clusters of more epithelial differentiated cells (epi-1 to 4) and
157 one cluster with mixed phenotypes (mix) (Figure 2A). pEMT-1 is enriched for genes involved
158 in coagulation and play a role in angiogenesis as *THBS1*, *CYR61* and *F3* (Figure 2B,
159 Supplementary Figure 4A). pEMT-2 and 3 showed higher expression of extracellular matrix
160 (ECM) remodeling genes, but pEMT-2 also higher expressed cytokeratin *KRT15* and
161 chemokine *CXCL14* while pEMT-3 higher expressed the serine protease inhibitor *SERPINA1*

162 and podoplanin (*PDPN*) which mediates efficient ECM degradation by controlling invadopodia
163 [21]. Of the more epithelial differentiated cell clusters, epi-2's expression profile is closest to
164 the pEMT cluster: higher expression of *MMP1* and lower expression of *SPRR1B* and
165 *S100A8/A9* than epi-1, 3 and 4 (Figure 2B, E). Epi-3 showed higher expression of *S100A7* and
166 *KRTDAP*, whereas epi-4 has higher expression of kallikreins (*KLK6/7*), prostate stem cell
167 antigen (*PSCA*) and adipogenesis regulatory factor (*ADIRF*). Both *ADIRF* and *PSCA* play a
168 role in prostate cancer and *PSCA* is also reported as highly expressed in mucosal tissue, but
169 less in HNSCC [22-24].

170 To gain a better understanding on the gene expression dynamics we estimated RNA velocity,
171 which predicts the short-term future development in gene expression of individual cells using
172 the ratio of spliced and non-spliced mRNA counts (Figure 2C). This analysis revealed that
173 epithelial differentiated cells were strongly developing towards cluster epi-4, while most other
174 cells did not show a strong uniform developmental direction. Tracking the developmental
175 pathway within the metastasis by trajectory analysis across all EMP-related clusters revealed
176 a major developmental axis between pEMT and epithelial differentiated cells that is diversifying
177 within each end (Figure 2D, E, Supplementary Figure 4B). To confirm that we the found
178 progressive epithelial differentiation of the metastatic cells represents a MET, we inferred
179 CNVs from the scRNAseq data. This demonstrated an increased number of copy number gains
180 on chromosome 1, 8, 17, and 19 within epithelial differentiated cell clusters epi-1, epi-3, and
181 epi-4 (Figure 2F, Supplementary Figure 4C). It should be noted, however, that CNVs on
182 chromosomes 1 and 17 are associated with upregulated epithelial genes in close genomic
183 proximity; thus, these two copy number gains may not represent true genomic CNVs, but rather
184 reflect the high expression of these genes in epithelial differentiated cells (Supplementary
185 Figure 4D, E).

186 **Intra-tumoral heterogeneity of OSCC is driven by EMP**

187 To test whether our observation that OSCC cells in one lymph node metastasis undergo a
188 mesenchymal-epithelial transition is generally valid we extended our analyses by adding 5

189 primary tumors and 8 matched lymph node metastases from 6 patients (Table 1). From the
190 publicly available scRNAseq data set on primary HNSCC tumors published by Kürten *et al.*
191 [15] we chose to include the 9 HPV-negative primary tumors, which were all but one originating
192 from the oral cavity (HN07 originated from the larynx) in our analysis. In total, we analyzed
193 7,263 cancer cells from 16 different patients (Figure 3A). Importantly, the frequency of cancer
194 cells was unevenly distributed across samples, which could not be explained by differences in
195 tumor cell content across samples as determined by histopathology (Supplementary Figure 1).
196 Since the recovery of cells by scRNAseq is also low, the stability of epithelial tumor cell
197 assemblies, which may not be sufficiently broken up by dissociation protocols, likely interfered
198 with the generation of OSCC single-cell suspensions. In addition, as expected from the inter-
199 patient heterogeneity, cancer cells were clustered based on their gene expression by patient
200 rather than functional phenotype (Figure 3A, B). Thus, we accounted for the patient-specific
201 effects with batch-corrected principal components (PCs) using the harmony R package which
202 indeed resulted in a clustering by functional phenotypes [25] (Figure 3C, Supplementary Figure
203 5A-C). For initial annotation of the phenotype of each cluster, we were guided by the gene
204 signatures previously identified in the indicator sample, with additional predominantly
205 immunoregulatory phenotypes found. EMP-related phenotypes were present in all, but one
206 tumor sample with only one cell (Figure 3D). Nevertheless, direct comparison of the
207 frequencies of the predominant phenotypes in the pooled scRNAseq dataset proved difficult
208 to interpret due to differences in cell numbers, and patient-specific gene expression patterns.
209 To confirm the validity of our scRNAseq data, we examined the bulk transcriptome and
210 deconvoluted the respective cell types for our 18 OSCC samples (Supplementary Figure 6).
211 Using this approach, we observed a higher tumor and stroma content and less lymphocytes,
212 which was expected as scRNAseq is known to be biased towards leukocytes. Therefore, we
213 chose an alternative approach: to compare the EMP-related intrapatient heterogeneity of all
214 analyzed tumor cells between patients, we calculated the similarities of the differential gene
215 expression within each patient of all respective clusters of all patients. We first considered each
216 patient individually and performed clustering, annotation, inferCNV and differential expression

217 analysis. As exemplified for the primary OSCC of patient HN01, tumor cell clusters had the
218 same inferred CNVs and were annotated based on their phenotype using again the indicator
219 sample as a guide (Figure 3E, Supplementary Figure 5D, E). The cosine similarity between
220 patient-specific clusters demonstrated that within each patient, the heterogeneity in EMP is
221 most prominent and epithelial differentiated phenotypes are profoundly different from most
222 other, especially pEMT clusters (Figure 3F, Supplementary Figure 5F). Indeed, all pEMT
223 phenotypes are very similar to each other and show a large overlap of the gene expression
224 patterns with predominantly immune- and metabolic-related clusters. In epithelial differentiated
225 cells, the most upregulated genes are *S100A8* and *S100A9*, encoding calprotectin) and
226 *SPRR1B*, all members of the epidermal differentiation complex [27]. Of note, hypoxia- and
227 stress-related heterogeneity is similar between patients suggesting a reactive response rather
228 than an aspect of tumor evolution.

229 **EMT-related transcription factor ZEB1 is highly active in metastatic epithelial**
230 **differentiated OSCC cells**

231 The transcription factors ZEB1/2, TWIST1/2, Snail (SNAI1) and Slug (SNAI2) are key for
232 regulation of EMP [7, 27]. While mRNA expression of *SNAI2* within single OSCC cells was
233 recently reported, the others were not detected [8]. Here, we confirm this observation as we
234 detected *SNAI2* mRNA in almost half of the OSCC cells (Figure 4A). However, detection of
235 low expressed genes such as transcription factors by scRNAseq, especially in 10X genomics
236 technology, becomes unreliable due to dropout effects. Also, the activity of transcription factors
237 is often not reflected by the dynamics of their mRNA expression alone, as their activity depends
238 additionally on protein stability and posttranslational modifications; for example, ZEB1 protein
239 is more stable than Snail [28]. To circumvent this problem, we inferred the activity of these
240 transcription factors based on the mRNA expression profile of their target genes using the
241 algorithm VIPER with regulons defined by DoRothEA database [29-31]. Using this approach,
242 we were able to detect high activities of ZEB1, ZEB2, Snail and Slug in OSCC cells of different
243 patients with varying EMP phenotype (Figure 4B). Since, on the one hand, the observation that
244 epithelial differentiation in OSCC metastases is associated with higher activity of the EMT

245 activator ZEB1 was unexpected and, on the other hand, the method used to derive
246 transcription factor activities appears to maybe overestimate the activity for transcriptional
247 repressors, we addressed the plausibility of this observation (Figure 4C, Supplementary Figure
248 7). Consistent with the fact that one of the main functions of ZEB1 is the downregulation of E-
249 cadherin [32] (*CDH1*), OSCC cells generally showed low expression of *CDH1* (Figure 4B).
250 Next, we investigated the expression of ZEB1 protein by immunohistochemistry (IHC) in all 14
251 tumor lesions of our cohort. Since we observed nuclear ZEB1 expression in similar tumor areas
252 as high cytoplasmic cornifin-B expression, which served as a marker of epithelial differentiation
253 (Figure 4D, E), we validated the co-expression of ZEB-1 and cornifin-B in the same cell by
254 immunofluorescence double-staining (Figure 4F). In line with our scRNAseq data,
255 colocalization of both proteins was observed in a fraction of cancer cells in 9/14 (64 %)
256 samples, double positive cells were more frequently observed in lymph node metastases (7/9,
257 78 %; Table 1).

258 # Discussion

259 EMT represents the reactivation of an embryonic developmental program in which cells
260 acquire migratory and invasive properties, i.e., prerequisites for invasion and metastasis of
261 cancer [3, 33, 34]. Thus, in early stages of metastasis, tumor cells undergo EMT, whereas in
262 established metastases the reverse process *aka* MET is also observed [35, 36]. To assess the
263 EMP-associated heterogeneity among OSCC cells and gain some insight into the dynamics of
264 this process, we examined the transcriptomes of 7,263 individual carcinoma cells isolated from
265 primary or metastatic OSCC. Our results not only confirm the EMP-associated heterogeneity
266 of cancer cells in primary and metastatic OSCC, but also demonstrate progressive MET in
267 established lymph node metastases. Interestingly, the epithelial differentiation in OSCC
268 metastases is associated with higher activity of EMT-activator ZEB1, which was confirmed on
269 protein level by co-expression of ZEB1 and cornifin-B in individual tumor cells by
270 immunofluorescence staining.

271 EMP appears to be the main driver of cellular heterogeneity within OSCC: detailed phenotyping
272 of cancer cells identified several clusters whose predominant functional phenotypes
273 corresponded to different EMP states, ranging from a pEMT to an epithelial differentiated state.
274 However, particularly the pEMT phenotypes could be overlaid by features related to
275 angiogenesis, ECM remodeling, metabolic adaptations, stress, and interaction with the
276 immune system. Metabolic adaptations include response to environmental limitations as
277 hypoxia and low glucose. Low glucose conditions are counteracted with upregulation of genes
278 related to amino acid metabolism that fuel into glycolysis [20]. While previous studies
279 suggested that the activity of specific metabolic pathways in OSCC varies widely among
280 patients [37], we observed that hypoxia- and stress-related gene expression patterns are
281 rather similar between patients supporting the notion of an reactive response rather than an
282 aspect of individual tumor evolution.

283 In terms of EMP dynamics, it is assumed that cells may transit from one EMP state to another
284 along a continuous spectrum of changes. Currently, however, it is also discussed if long-lived
285 phenotypes representing discrete EMP states prevail [3, 5, 6, 8, 9, 11-13]. Most studies
286 supporting continuous transitions are based on *in vitro* or preclinical *in vivo* models that may
287 not fully reflect the complexity of the tumor and its microenvironments [5, 6, 9]. Indeed, human
288 *in situ* or *ex vivo* studies suggested distinct EMP states; however, these approaches do not
289 fully capture cellular dynamics [8, 12, 34]. Comparing the intratumoral heterogeneity we
290 demonstrated that within each patient, the EMP-driven heterogeneity is most prominent and
291 epithelial differentiated phenotypes are profoundly different from pEMT clusters suggesting
292 that these states may be rather static. Moreover, gene expression dynamics estimated by RNA
293 velocity demonstrated epithelial differentiated cells were strongly developing towards a more
294 pronounced epithelial differentiation with an increasing expression of genes of the epidermal
295 differentiation complex [27]. Of note, OSCC cells with a pEMT phenotype did not show such a
296 uniform developmental direction. The assumption that epithelial differentiated metastatic cells
297 developed later than pEMT cells, i.e., undergoing MET, is supported by an increasing number

298 of inferred copy number gains towards increasing epithelial differentiation even if accounting
299 for the limitations of this approach.

300 However, we cannot conclude whether MET happened within the metastasis or primary
301 malignancy, as our data reflects the tumor heterogeneity within a specific timepoint of tumor
302 evolution. Multiple disseminated tumor cells reflecting the observed EMP heterogeneity might
303 have migrated collectively from the primary tumor that could be crucial for metastatic
304 consolidation, as we observed a similar EMP heterogeneity also within primary tumors.

305 The high transcriptional activity of EMT-activator ZEB1 in epithelial differentiated OSCC cells
306 in both primary and metastatic tumor lesions was unexpected, even considering that the
307 scRNAseq data inferred transcription factor activities are biased towards transcriptional
308 repressors. In the case of primary tumors this may be interpreted as the incipient EMT, but this
309 would not work for the metastatic lesions in which ZEB1 activity was associated with a
310 progressive epithelial differentiation. Indeed, in previous studies depletion of ZEB1 has driven
311 tumor cells from pEMT towards an epithelial phenotype [38, 39]. However, depletion of Zeb1
312 also reduces phenotypic variability of cancer cells particular their phenotypic/metabolic
313 plasticity [38]. While it is well established that ZEB1 together with microRNAs stabilizes EMT
314 through a feedforward loop, this loop could also induce epithelial differentiation based on
315 environmental factors [40]. In addition to the transcriptional repressor activity, ZEB1 has been
316 demonstrated to induce the epithelial differentiation marker cornifin-B in response to IL-1β and
317 IFN-γ [41]. We not only demonstrate the coincidence of *SPRR1B* mRNA expression and ZEB1
318 activity, but also the co-localization of cornifin-B and ZEB1 protein expression in OSCC lymph
319 node metastases. Thus, although ZEB1 activity is crucial for the induction of the pEMT state,
320 it does not seem to completely prevent partial epithelial differentiation. Remarkably, no relevant
321 differences in *CDH1* expression were detected between the different EMP states in the
322 metastatic OSCC lesions. Therefore, the epithelial differentiated phenotypes we observed
323 most closely correspond to a partial epithelial differentiation analogous with the observed
324 pEMT phenotypes. We speculate that the driving force behind this EMP-associated

325 heterogeneity of OSCC cells is to maintain cellular integrity. For example, ZEB1 is an ATM-
326 substrate linking ATM and CHK1, promoting homologous recombination-dependent DNA
327 repair and thereby protecting cells from genotoxic stress whereas expression of keratin
328 intermediate filaments helps to protect cells from stress associated apoptosis [42, 43].

329 In summary, heterogeneity within OSCC is dominated by EMP differences resulting in distinct
330 pEMT and epithelial differentiated phenotypes, but particularly the pEMT phenotypes can be
331 accompanied by features related to metabolic adaptations, stress, and interaction with the
332 immune system. These phenotypes are likely to endow capabilities essential for the different
333 stages of the metastatic process, including maintenance of cellular integrity. Thus, further
334 studies are required and already arising to unravel the factors driving these evolutionary
335 processes in tumor cells [44].

Patients, Material and Methods

Tissue samples

Twenty tissue samples that included primary tumors, clinically apparent lymph node metastases, and potentially affected lymph nodes from 7 OSCC patients treated at the Department of Oral-, Maxillo- and Plastic Facial Surgery, University Hospital of the Heinrich-Heine-University Düsseldorf, Germany, were used for histology, IHC, immune fluorescence as well as bulk and single-cell transcriptome analyses. The clinical details are provided in table 1. 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 involving human participants are in accordance with the ethical standards of the institutional and research committee and with the Helsinki Declaration.

Histology and IHC

Hematoxylin and Eosin (H&E) and IHC were performed on 4 µm formalin-fixed, paraffin-embedded (FFPE) sections. H&E staining was performed using standard protocols (Supplementary Figure 1). Whole-slide imaging was performed using Zeiss Axioscan 7 and 10x magnification (Carl Zeiss Microscopy Deutschland GmbH, Oberkochen, Germany).

IHC using the rabbit polyclonal antibodies anti-SPRR1B (Cat. No.: SAB1301567-400UL, Sigma Aldrich, Darmstadt, Germany) anti-ZEB1 antibody (Cat. No. HPA027524-25UL, Sigma Aldrich) was performed as previously described [45]. Briefly, after sections were deparaffinized for 60 min at 60°C and rehydrated, sections were incubated for 15 min in an inverter microwave oven with antigen retrieval buffer pH 9 for anti-SPRR1B and pH 6 for anti-ZEB1. After 3x2 min washes with Tris-buffered saline with 0.1% Tween (TBST) sections were incubated for 8 min with 3% peroxidase and after an additional washing step for 30 min with 3% bovine serum albumin (BSA) in TBST. For single staining sections were subjected for 1h to anti-SPRR1B in 1:600 dilution or overnight to anti-ZEB1 in a 1:500 dilution, both at room temperature. Afterwards, the secondary anti-rabbit HRP Polymer was applied for 30 min, followed by 1:20 diluted 3,3'-Diaminobenzidin (DAB) for 10 min and 1:10 diluted hematoxylin for 3 min with

363 washing steps with TBST in between before a final washing with tap water for 3 min before
364 fixation. For multiplexed antigen detection, the OpalTM chemistry system (Akoya Biosciences,
365 Marlborough, MA, USA, Cat. No.: OP7TL4001KT) was used according to the manufacture's
366 description. Briefly, after deparaffinization and fixation, we processed the sections for 15 min
367 with retrieval buffers in an inverter microwave oven. Then, we incubated them with antibody
368 diluent for 10 min at room temperature, followed by incubation with the anti-SPRR1B antibody
369 for 30 min. Next, Opal Polymer horseradish peroxidase (HRP) secondary antibody solution
370 with the respective chromogen was applied for 10 min, antibodies were removed by microwave
371 treatment and the staining with anti-ZEB1 antibody was performed. Slides were lastly
372 incubated with 4′,6-diamidino-2-phenylindole (DAPI) for 5 min.

373 **Single-cell RNA sequencing**

374 Samples were processed immediately after surgery and temporarily stored for transport at 4°C
375 in tissue storage solution (Miltenyi Biotec, Bergisch Gladbach, Germany) as previously
376 described. Briefly, samples were dissociated into single-cell suspensions using the
377 gentleMACS Dissociator (Cat. No. 130-093-235, Miltenyi Biotec, Bergisch Gladbach,
378 Germany) with program "h_tumor_01", followed by 2x program "h_tumor_02" in 4.7 ml RPMI
379 1640 (Cat. No. P04-16500, PAN-Biotech) and an enzyme mix consisting of 200µl Enzyme H,
380 100 µl Enzyme R and 25 µl Enzyme A (Cat. No. 130-095-929 Miltenyi Biotec). Afterwards,
381 single-cell suspensions were reconstituted and washed thrice with 0.05 % BSA phosphate-
382 buffered saline (PBS) and filtered through a 100 µl cell strainer.

383 In cases multiple samples of one patient had to be analyzed (Table 1), antibody hashing for
384 multiplexing of samples was performed according to manufacturer's protocol. Briefly, 1 µg of
385 the respective TotalSeq anti-human hashtag antibody was used to incubate a maximum of ca.
386 2 million cells for 30 min at 4 °C (Cat. No. 394601, 394603, 394605 and 394661, 394663,
387 394665, respectively, Biolegend, San Diego, CA, USA). After 3 washes with PBS with 0.05 %
388 BSA respective cell suspensions were mixed prior to single-cell RNA library preparation. In
389 short, both unhashed and hashed single-cell suspensions were barcoded and processed with

390 the microfluidic system of 10x Genomics Chromium v2.0 platform as described in the
391 manufacturer's protocols (10x Genomics, Leiden, Netherlands). Due to a change of system,
392 both the 3' technology including Chromium Single Cell 3' Library & Gel Bead Kit version 2 (Cat.
393 No. 120237), Chromium Single Cell A Chip Kit (Cat. No. 120236), Chromium i7 Multiplex Kit
394 (Cat. No. 120262) as well as the 5' technology including Chromium Single Cell 5' Library & Gel
395 Bead Kits version 2 (Cat. No. 1000263), Chromium Next GEM Chip K Single Cell Kit (Cat. No.
396 1000286), Dual Index Kit TT set A (Cat. No. 1000215) were used; for library construction the
397 Chromium Single Cell 3'/5' Library Construction Kit (Cat. No. 1000020) was applied. After
398 library preparation, the library from patient one was sequenced with an Illumina HiSeq 4000
399 (Illumina, Berlin, Germany) at the DKFZ Genomics and Proteomics Core Facility in Heidelberg
400 and all other libraries were sequenced on an Illumina NovaSeq 6000 (Illumina, Berlin,
401 Germany) in three runs (Run 1: patient 2,4,5; Run2: patient 3, Run3: Patient 6 and 7) at the
402 West German Genome Center in Cologne.

403 **Bulk transcriptome analysis**

404 The bulk transcriptome was analyzed using a quantitative nuclease protection assay from the
405 HTG Transcriptome Panel (HTP) according to the manufacturers protocol (Cat. No. HTG-001-
406 224, HTG Molecular Diagnostics, Tucson, AZ, USA). Briefly, the tumor areas were macro-
407 dissected as depicted in Supplementary Figure 1 from 4 µm FFPE sections and subjected to
408 Proteinase K and DNase digestion. Next, the quantitative nuclease protection assay was
409 performed using the HTG EdgeSeq Processors before adapters and sample tags were added
410 in a PCR amplification. The resulting libraries were sequenced using on an Illumina NextSeq
411 500/550 High Output Kit v2.5 (75 cycles) (Cat. No. 20024906, Illumina, Berlin, Germany).

412 The resulting FASTQ files were processed towards a gene expression count matrix using the
413 HTG EdgeSeq Reveal Software version 4.0.1. Quality Control, normalization, and principal
414 component analysis (PCA) were performed using R version 4.0.5. Sample 5 failed QC due to
415 low number of sequenced reads and was removed from the analysis. Deconvolution was
416 performed with web application of CIBERSORTx (https://cibersortx.stanford.edu/) using a

417 signature matrix derived from the gene expression count matrix of combined scRNAseq data
418 of the samples analyzed with HTP [46]. We filtered for genes that are expressed less than 5%
419 within the given tumor phenotypes and randomly included only 75% of T- and B-Cells for better
420 performance. The resulting signature matrix was used for imputing cellular fractions from the
421 counts-per-million normalized HTP data without any batch correction or quantile normalization
422 and 500 permutations.

423 **Bioinformatic analysis of scRNAseq data**

424 ***Preprocessing***

425 Processing from FASTQ files towards the unfiltered count matrix (barcodes x genes) was
426 performed using Cellranger Software Suite version 3.1.0 and the human reference genome
427 build GRCh38, downloaded from 10x Genomics in version 3.0.0.

428 Cells were identified by evaluating quality criteria inspired by Luecken *et al.*, also see
429 Supplementary File 1 [47]. Cells were defined by having more than 500 UMIs, less than 10 %
430 mitochondrial gene expression and additionally for patient 3 and 5 having more than 30
431 housekeeper genes expressed. The filtered count matrices (cells x genes) were further
432 processed using Seurat version 4.0.1 and R version 4.0.5 [48]. Demultiplexing of hashed
433 libraries was performed choosing manual threshold of hashtag oligo (HTO) expression based
434 on quality assessments described in Supplementary File 1. We removed doublets that were
435 identified by demultiplexing of HTO expression matrices from all analyses.

436 ***Normalization and dimensionality reduction***

437 Every time we performed analysis of a specific set of cells, e.g., only tumor cells or only a
438 specific patient, the respective set of cells is normalized using the SCtransform algorithm and
439 the 3000 most variable genes were selected for PCA [49]. During normalization, we regressed
440 for cell cycle scores and percentage of mitochondrial gene expression. Cell cycle scores and
441 phases were determined in Seurat using log-normalized RNA counts and S- and G2M-Phase
442 genes defined by Tirosh *et al.* [50]. When generating a uniform manifold approximation and

443 projection (UMAP) without patient-specific batch effect we used corrected PCs derived with
444 the harmony R package version 0.1.0 [54]. Based on the variance explained by each PC and
445 the respective ranked elbow plot we choose an appropriate number of PCs for UMAP
446 visualization and SNN clustering as implemented in Seurat. For deriving patient-specific
447 clusters for calculating the intratumoral cosine similarity we used the same resolution
448 parameter for comparability. UMAPs colored by specific gene expression were ordered by
449 expression values.

450 ***Tumor cell identification and phenotyping***

451 We used a combination of SNN clustering, differential gene expression, gene set enrichment
452 analysis (GSEA) and expression of literature-based marker genes for identification and
453 description of tumor cell phenotypes. Malignant cells were first identified by high epithelial gene
454 expression as for example cytokeratins and inferred CNVs (Supplementary Figure 2). CNVs
455 were inferred using the R package inferCNV version 1.6.0 with the not normalized, filtered
456 count matrix including all cells as input and algorithm run in “samples” mode . Inferred CNVs
457 of mitochondrial genes were excluded. Further, we excluded cells with ambiguous cell type
458 marker expression. For example, within the tumor cells of the lymph node metastasis from
459 patient 1, we observed few cells expressing genes typical for fibroblasts and DCs that were
460 subsequently excluded from tumor-specific analysis. Similarly, we removed T-Cells, B-Cells,
461 mast cells, fibroblasts, muscle cells, melanocytes, and other Immune cells from the tumor cell
462 subsets. Differential gene expression was performed by calculating the log2 foldchanges
463 between one cluster and all other cells from the subset based on log-normalized data using
464 NormalizeData function and scale factor of 10,000. We filtered for genes with log2 foldchange
465 greater than 0.25 and a minimum percentual expression of at least 10 % within the cluster or
466 all other cells. For calculating the cosine similarity, we did not filter the log2 foldchanges. GSEA
467 was performed using the “fgsea” R package version 1.16.0, log2 foldchanges from differential
468 gene expression and gene ontology biological processes (GO:BP,C5 v7.1) as well as hallmark
469 gene sets (H, v7.1) downloaded from MSigDB database [16, 51, 52]. Gene sets were included

470 if they have at least 15 genes or at maximum 500 genes within the gene set using 10,000
471 permutations.

472 For deriving epithelial differentiating and pEMT gene signatures of patient 1 we calculate
473 foldchanges between the “epi” and “pEMT” cluster and included genes with log2 foldchanges
474 greater or lesser than 1, respectively, and with at least 10 % of either epi or pEMT cells
475 expressing that gene. Gene set variation analysis (GSVA) was performed using the R package
476 GSVA version 1.38.2 with default settings, i.e., gaussian kernel [53]. As input, EMTome
477 signatures, the pEMT and epithelial differentiation 1 and 2 signature from Puram *et al*. [8] and
478 the three EMT and the epithelial senescence signatures from Kinker *et al.* were used [12].

479 Trajectories were inferred using SlingShot version 1.8.0 with malignant cell clusters as shown
480 in Figure 2A of patient 1 and the first 20 PCs [54]. RNA velocity was inferred using the VeloCyto
481 python and R package (version 0.6) [55]. Creation of the loom file was done using default
482 options and gene annotations as used for Cellranger processing. Genes were filtered based
483 on the minimum average expression magnitude with a threshold of 0.05 for spliced and 0.02
484 for unspliced reads. Velocity estimates were calculated using as distance the inverse
485 correlation coefficient of the PC embedding correlation matrix, the top/bottom 2 % quantiles for
486 gamma fit, 50 neighboring cells projecting 1 deltaT into the future and projected on the UMAP
487 using 200 neighbors and 50 grid points.

488 Transcription factor activity was inferred using the VIPER algorithm (version 1.24.0) and
489 regulons from the DoRothEA database (version 1.2.2) [29-31]. Hierarchical clustering in the
490 heatmaps was performed using Euclidean distance and ward.D2 method unless otherwise
491 noted. Visualization was performed using ggplot2 version 3.3.3 and ComplexHeatmap version
492 2.6.2 [56, 57].

493 **Reanalysis of HNSCC dataset from Kürten *et al*.**

494 From the publicly available scRNAseq data set on primary HNSCC tumors published the
495 FASTQ files of CD45 negative and HPV-negative libraries from Kürten *et al.* were downloaded
496 from the sequencing read archive (SRA) accessed through gene expression omnibus (GEO)

497 accession ID GSE164690 in BioProject PRJNA691564 [15]. The data sets were analyzed the
498 same as described above. HPV-negative samples were chosen for comparability to the OSCC
499 dataset. However, the HN07 tumor originated from the larynx, while all other samples indeed
500 originated from the oral cavity. We adjusted the cell identification thresholds based on our
501 evaluation criteria pooled for all libraries: cells were defined by having more than 175 genes
502 expressed, less than 10 % mitochondrial gene expression and more than 60 housekeeper
503 genes expressed (see Supplementary File 1).

504 # Supplementary Information

505 **Supplementary Figure 1:** Histology of OSCC primary and metastatic tumors.

506 **Supplementary Figure 2:** | Cell type identification by differential expression and inferred
507 CNVs for a metachronous lymph node metastasis.

508 **Supplementary Figure 3:** PEMT and epithelial differentiating gene expression signatures are
509 comparable to previously published EMT signatures.

510 **Supplementary Figure 4:** Extended analysis of the tumor phenotype characterization for the
511 lymph node metastasis of patient 1.

512 **Supplementary Figure 5:** Malignant phenotypes characterized across all analyzed patients.

513 **Supplementary Figure 6:** Bulk transcriptomes reveal the cellular composition of OSCC.

514 **Supplementary Figure 7:** Inferred transcription factor activity might be biased by activator or
515 repressor function.

516 **Table 1:** Clinical and sequencing information from OSCC patients.

517 **File 1:** Supplementary Materials and Methods

518 # Data availability

519 The processed datasets generated and analyzed during the current study are available in the
520 GEO database with accession id GSE195655. The raw FASTQ files are available upon request
521 due to privacy reasons. The high-resolution H&E Images are available through figshare under
522 the DOI: https://doi.org/10.6084/m9.figshare.20905837.v1.

523 # Code availability

524 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 Yvonne Krause and Sophia Berger, both Department of Pathology, University Medicine Essen, Germany, for their help and guidance with HTG EdgeSeq. We further thank the DKFZ Genomics and Proteomics Core Facility and the West German Genome Center in Cologne for providing Illumina 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, scRNAseq and library preparation. FF established and performed histological, IHC and bulk transcriptome analysis. 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, immunofluorescence examination and manuscript 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 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.

3 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.

4 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.

5 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.

6 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.

7 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.

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

9 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.

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

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

599 12 Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM *et al*. Pan-cancer single-
600 cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet* 2020; 52: 1208-
601 1218; https://doi.org/10.1038/s41588-020-00726-6.

602

603 13 Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB *et al*. Mapping lung
604 cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution.
605 *Nat Commun* 2019; 10: 5587; https://doi.org/10.1038/s41467-019-13441-6.

606

607 14 Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM *et al*. Multimodal Analysis of
608 Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* 2020; 182: 497-
609 514.e422; https://doi.org/10.1016/j.cell.2020.05.039.

610

611 15 Kürten CHL, Kulkarni A, Cillo AR, Santos PM, Roble AK, Onkar S *et al*. Investigating immune and
612 non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nat*
613 *Commun* 2021; 12: 7338; https://doi.org/10.1038/s41467-021-27619-4.

614

615 16 Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
616 Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015; 1: 417-425;
617 https://doi.org/10.1016/j.cels.2015.12.004.

618

619 17 Guo D-D, Xie K-F, Luo X-J. Hypoxia-induced elevated NDRG1 mediates apoptosis through
620 reprograming mitochondrial fission in HCC. *Gene* 2020; 741: 144552;
621 https://doi.org/10.1016/j.gene.2020.144552.

622

623 18 Ivan M, Kaelin WG. The EGLN-HIF O2-Sensing System: Multiple Inputs and Feedbacks. *Mol Cell*
624 2017; 66: 772-779; https://doi.org/10.1016/j.molcel.2017.06.002.

625

626 19 Laplante M, Sabatini DM. mTOR signaling at a glance. *J Cell Sci* 2009; 122: 3589-3594;
627 https://doi.org/10.1242/jcs.051011.

628

629 20 DeBerardinis Ralph J. Serine Metabolism: Some Tumors Take the Road Less Traveled. *Cell*
630 *Metab* 2011; 14: 285-286; https://doi.org/10.1016/j.cmet.2011.08.004.

631

632 21 Martín-Villar E, Borda-d'Agua B, Carrasco-Ramirez P, Renart J, Parsons M, Quintanilla M *et al*.
633 Podoplanin mediates ECM degradation by squamous carcinoma cells through control of
634 invadopodia stability. *Oncogene* 2015; 34: 4531-4544; https://doi.org/10.1038/onc.2014.388.

635

636 22 de Nooij-van Dalen AG, van Dongen GAMS, Smeets SJ, Nieuwenhuis EJC, Stigter-van Walsum
637 M, Snow GB *et al*. Characterization of the human Ly-6 antigens, the newly annotated member
638 Ly-6K included, as molecular markers for head-and-neck squamous cell carcinoma. *Int J Cancer*
639 2003; 103: 768-774; https://doi.org/10.1002/ijc.10903.

640

641 23 Reiter RE, Gu Z, Watabe T, Thomas G, Szigeti K, Davis E *et al*. Prostate stem cell antigen: A cell
642 surface marker overexpressed in prostate cancer. *Proc Natl Acad Sci USA* 1998; 95: 1735-1740;
643 https://doi.org/10.1073/pnas.95.4.1735.

644

645 24 Zhou K, Arslanturk S, Craig DB, Heath E, Draghici S. Discovery of primary prostate cancer
646 biomarkers using cross cancer learning. *Sci Rep* 2021; 11: 10433;
647 https://doi.org/10.1038/s41598-021-89789-x.

648
649 25 Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K *et al*. Fast, sensitive and accurate
650 integration of single-cell data with Harmony. *Nat Methods* 2019; 16: 1289-1296;
651 https://doi.org/10.1038/s41592-019-0619-0.

652
653 26 Tyler M, Tirosh I. Decoupling epithelial-mesenchymal transitions from stromal profiles by
654 integrative expression analysis. *Nat Commun* 2021; 12: 2592; https://doi.org/10.1038/s41467-
655 021-22800-1.

656
657 27 Oh IY, de Guzman Strong C. The Molecular Revolution in Cutaneous Biology: EDC and Locus
658 Control. *J Invest Dermatol* 2017; 137: e101-e104; https://doi.org/10.1016/j.jid.2016.03.046.

659
660 28 Stemmler MP, Eccles RL, Brabletz S, Brabletz T. Non-redundant functions of EMT transcription
661 factors. *Nat Cell Biol* 2019; 21: 102-112; https://doi.org/10.1038/s41556-018-0196-y.

662
663 29 Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and
664 integration of resources for the estimation of human transcription factor activities. *Genome*
665 *Res* 2019; 29: 1363-1375; https://doi.org/10.1101/gr.240663.118.

666
667 30 Holland CH, Tanevski J, Perales-Patón J, Gleixner J, Kumar MP, Mereu E *et al*. Robustness and
668 applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data.
669 *Genome Biol* 2020; 21: 36; https://doi.org/10.1186/s13059-020-1949-z.

670
671 31 Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH *et al*. Functional characterization
672 of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet*
673 2016; 48: 838-847; https://doi.org/10.1038/ng.3593.

674
675 32 Sánchez-Tilló E, Lázaro A, Torrent R, Cuatrecasas M, Vaquero EC, Castells A *et al*. ZEB1
676 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling
677 protein BRG1. *Oncogene* 2010; 29: 3490-3500; https://doi.org/10.1038/onc.2010.102.

678
679 33 Williams ED, Gao D, Redfern A, Thompson EW. Controversies around epithelial–mesenchymal
680 plasticity in cancer metastasis. *Nat Rev Cancer* 2019; 19: 716-732;
681 https://doi.org/10.1038/s41568-019-0213-x.

682
683 34 Derynck R, Weinberg RA. EMT and Cancer: More Than Meets the Eye. *Dev Cell* 2019; 49: 313-
684 316; https://doi.org/10.1016/j.devcel.2019.04.026.

685
686 35 Dongre A, Weinberg RA. New insights into the mechanisms of epithelial–mesenchymal
687 transition and implications for cancer. *Nat Rev Mol Cell* 2019; 20: 69-84;
688 https://doi.org/10.1038/s41580-018-0080-4.

689

690 36 Tsai JH, Donaher JL, Murphy DA, Chau S, Yang J. Spatiotemporal regulation of epithelial-
691 mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* 2012;
692 22: 725-736; https://doi.org/10.1016/j.ccr.2012.09.022.

693
694 37 Xiao Z, Dai Z, Locasale JW. Metabolic landscape of the tumor microenvironment at single cell
695 resolution. *Nat Commun* 2019; 10: 3763; https://doi.org/10.1038/s41467-019-11738-0.

696
697 38 Krebs AM, Mitschke J, Lasierra Losada M, Schmalhofer O, Boerries M, Busch H *et al*. The EMT-
698 activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer.
699 *Nat Cell Biol* 2017; 19: 518-529; https://doi.org/10.1038/ncb3513.

700
701 39 Carstens JL, Yang S, Correa de Sampaio P, Zheng X, Barua S, McAndrews KM *et al*. Stabilized
702 epithelial phenotype of cancer cells in primary tumors leads to increased colonization of liver
703 metastasis in pancreatic cancer. *Cell Rep* 2021; 35: 108990;
704 https://doi.org/10.1016/j.celrep.2021.108990.

705
706 40 Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S *et al*. A reciprocal
707 repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in
708 cancer cells. *EMBO reports* 2008; 9: 582-589; https://doi.org/10.1038/embor.2008.74.

709
710 41 Li S, Gallup M, Chen YT, McNamara NA. Molecular mechanism of proinflammatory cytokine-
711 mediated squamous metaplasia in human corneal epithelial cells. *Invest Ophthalmol Vis Sci*
712 2010; 51: 2466-2475; https://doi.org/10.1167/iovs.09-4677.

713
714 42 Zhang P, Wei Y, Wang L, Debeb BG, Yuan Y, Zhang J *et al*. ATM-mediated stabilization of ZEB1
715 promotes DNA damage response and radioresistance through CHK1. *Nat Cell Biol* 2014; 16:
716 864-875; https://doi.org/10.1038/ncb3013.

717
718 43 Oshima RG. Apoptosis and keratin intermediate filaments. *Cell Death Differ* 2002; 9: 486-492;
719 https://doi.org/10.1038/sj.cdd.4400988.

720
721 44 Fennell KA, Vassiliadis D, Lam EYN, Martelotto LG, Balic JJ, Hollizeck S *et al*. Non-genetic
722 determinants of malignant clonal fitness at single-cell resolution. *Nature* 2022; 601: 125-131;
723 https://doi.org/10.1038/s41586-021-04206-7.

724
725 45 Spassova I, Ugurel S, Terheyden P, Sucker A, Hassel JC, Ritter C *et al*. Predominance of Central
726 Memory T Cells with High T-Cell Receptor Repertoire Diversity is Associated with Response to
727 PD-1/PD-L1 Inhibition in Merkel Cell Carcinoma. *Clinical Cancer Research* 2020; 26: 2257-2267;
728 https://doi.org/10.1158/1078-0432.Ccr-19-2244.

729
730 46 Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F *et al*. Determining cell type
731 abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019; 37:
732 773-782; https://doi.org/10.1038/s41587-019-0114-2.

733
734 47 Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst*
735 *Biol* 2019; 15: e8746; https://doi.org/10.15252/msb.20188746.

736
737 48 Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A *et al*. Integrated analysis of
738 multimodal single-cell data. *Cell* 2021; 184: 3573-3587.e3529;
739 https://doi.org/10.1016/j.cell.2021.04.048.

740
741 49 Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data
742 using regularized negative binomial regression. *Genome Biol* 2019; 20: 296;
743 https://doi.org/10.1186/s13059-019-1874-1.

744
745 50 Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ *et al*. Dissecting the
746 multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 2016; 352:
747 189; https://doi.org/10.1126/science.aad0501.

748
749 51 Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set
750 enrichment analysis. *bioRxiv* 2021: 060012; https://doi.org/10.1101/060012.

751
752 52 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al*. Gene set
753 enrichment analysis: A knowledge-based approach for interpreting genome-wide expression
754 profiles. *Proc Natl Acad Sci USA* 2005; 102: 15545; https://doi.org/10.1073/pnas.0506580102.

755
756 53 Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-
757 Seq data. *BMC Bioinform* 2013; 14: 7; https://doi.org/10.1186/1471-2105-14-7.

758
759 54 Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N *et al*. Slingshot: cell lineage and pseudotime
760 inference for single-cell transcriptomics. *BMC Genom* 2018; 19: 477;
761 https://doi.org/10.1186/s12864-018-4772-0.

762
763 55 La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V *et al*. RNA velocity of
764 single cells. *Nature* 2018; 560: 494-498; https://doi.org/10.1038/s41586-018-0414-6.

765
766 56 Wickham H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* 2016;

767
768 57 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
769 multidimensional genomic data. *Bioinformatics* 2016; 32: 2847-2849;
770 https://doi.org/10.1093/bioinformatics/btw313.

771

772

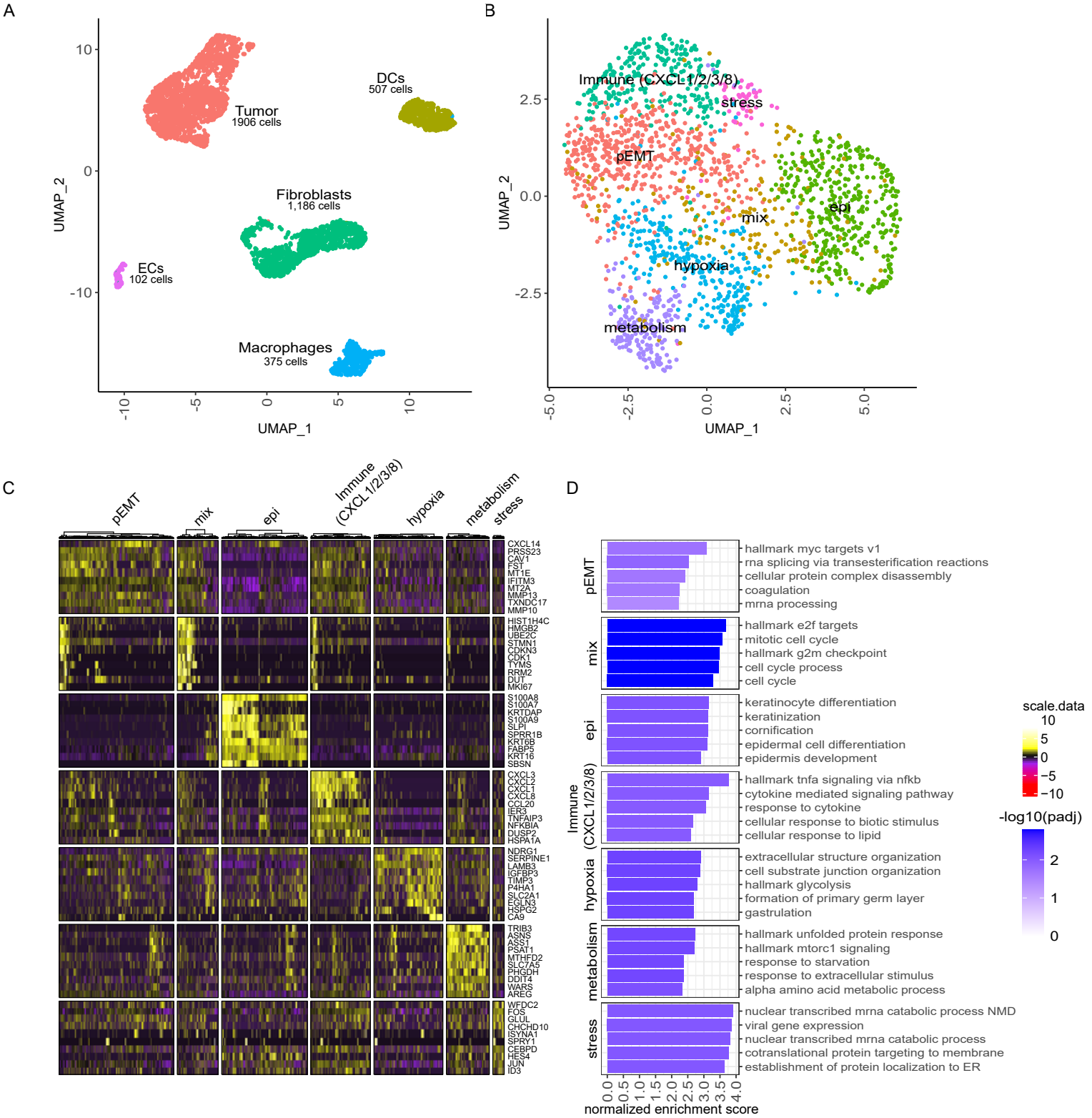

Figure 1 | Single-cell gene expression signatures in OSCC cells from a single metastasis reveal several predominant, but not necessarily exclusive, functional phenotypes. (A) UMAP based on scRNAseq data of 4,076 cells isolated from a metachronous lymph node metastasis. Cells are annotated and summarized according to the presumed cell type. **(B)** UMAP of 1906 OSCC cells depicted in A. Cells are annotated according to predominant functional phenotype. **(C)** Heatmap for scaled, log-normalized gene expression of tumor cells (columns) split by respective phenotype and the top 10 differentially expressed genes (DEGs) (rows) of the respective phenotype against all other tumor cells. DEGs are sorted from highest to lowest log2 foldchange. Row sections are ordered like column sections. **(D)** Top 5 enriched gene sets from log2 foldchanges of respective tumor phenotypes by normalized enrichment scores (x axis). Gene sets of respective phenotypes are sorted from highest to lowest enrichment. Bars are colored by the negative decadic logarithm of the Benjamini-Hochberg adjusted p-value (padj). DCs: dendritic cells. ECs: endothelial cells.

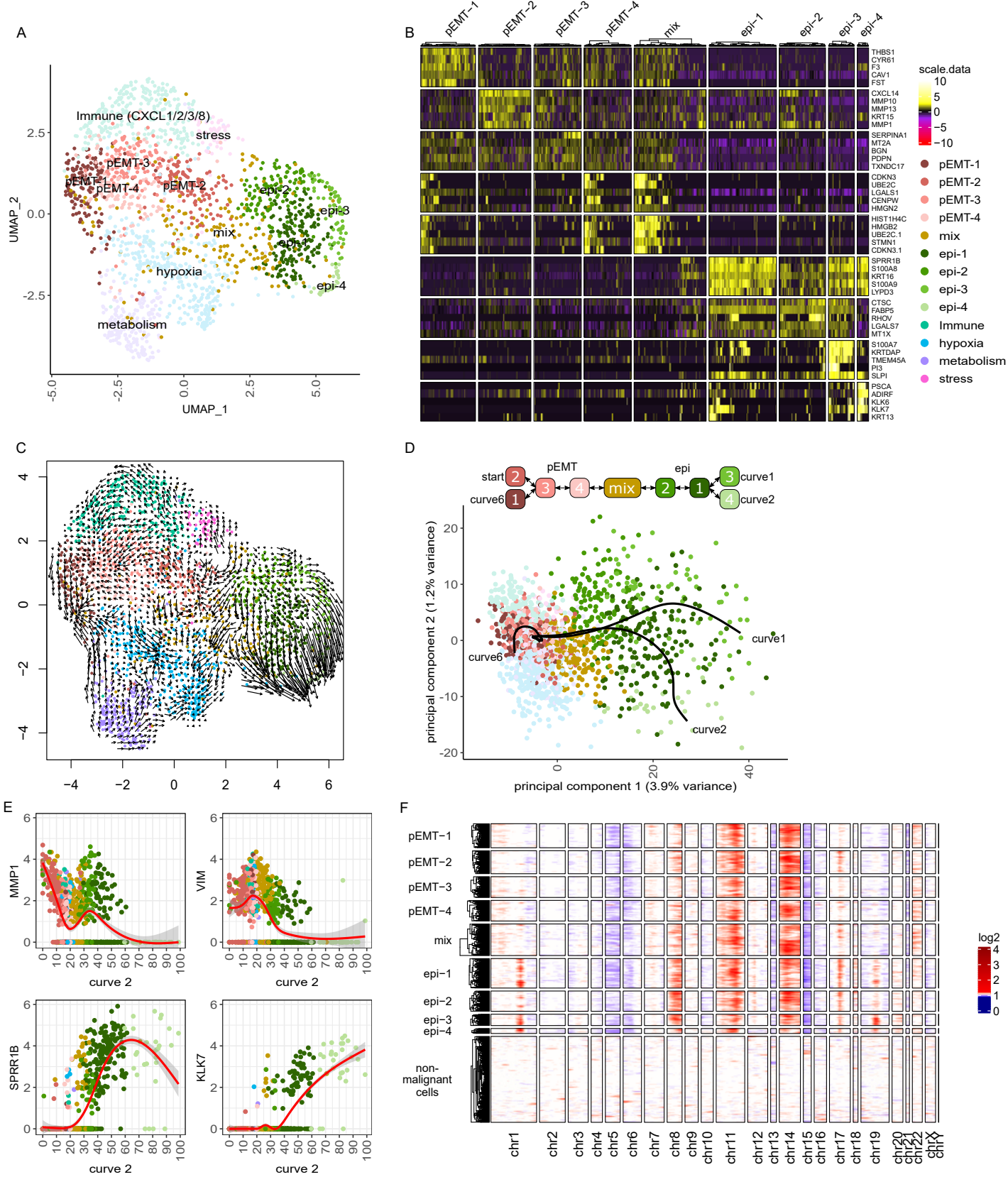

Figure 2 | Progressive epithelial differentiation, but no strong uniform direction of development in pEMT clusters. (A) UMAP of 1906 OSCC cells annotated based on SNN clustering, defining 4 pEMT (pEMT-1 to 4), 4 epithelial differentiated (epi-1 to 4) and one mixed (mix) cluster; clusters are numbered by size. **(B)** Heatmap for scaled, log-normalized gene expression in EMP-associated tumor cell phenotypes (columns) split by EMP cluster and their top 5 DEGs (rows) against all other EMP-related tumor cell phenotypes. DEGs are sorted from highest to lowest log2 foldchange. Row sections are ordered like column sections. **(C)** Projection of RNA velocity on the UMAP depicted in A. Arrows indicate the extrapolated direction of development and their length the strength of this future development. **(D)** First two principal components of OSCC cells with the three EMP-related principal curves that are derived from trajectory inference. Graph on top visualizes the relationship between EMP clusters described by the three principal curves forming a branching trajectory. **(E)** Log-normalized expression (y-axis) of MMP1, VIM, SPRR1B and KLK7 across pseudotime values (x-axis) of curve 2, color-coded by clusters. Red lines indicate smoothed expression values over the trajectory generated with a general additive model; 95 % confidence intervals are shaded gray. **(F)** Inferred CNVs across EMP-related tumor cells (rows) for all chromosomes (columns). Red indicates copy number gains, white diploid copy number and blue copy number loss. Columns show genes categorized in chromosomes and ordered by genome position; hence the size of the chromosome reflects the number of detected genes and not its nucleotide length. Mitochondrial genes were excluded.

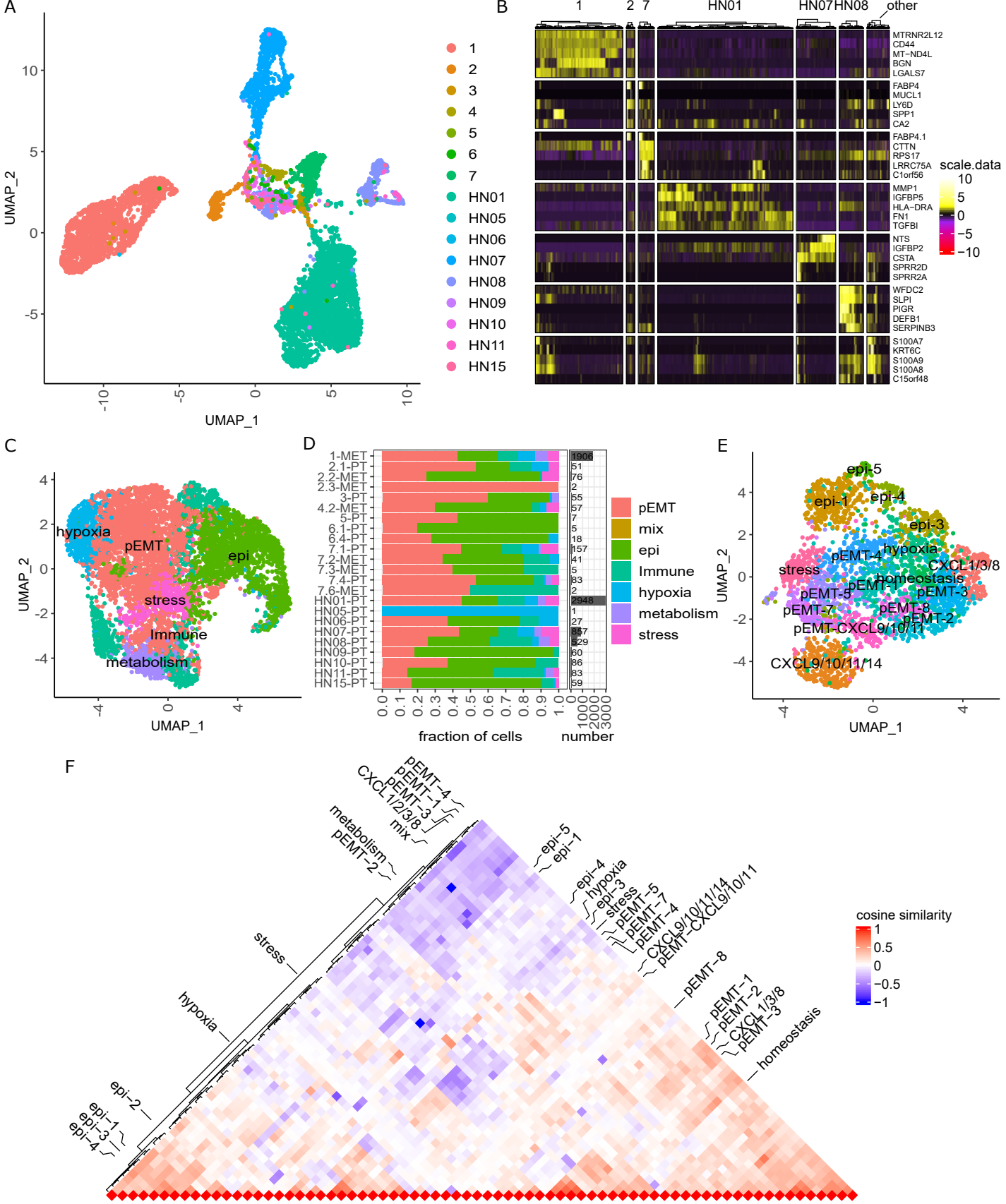

Figure 3 | Intra-tumoral heterogeneity of OSCC is driven by EMP. (A) UMAP based on scRNAseq data of 7,263 cancer cells from 16 different patients annotated by patient. **(B)** Heatmap for scaled, log-normalized gene expression of tumor cells split by patients and their top 5 DEGs (rows) against all other tumor cells. All patients with less than 100 cells are summarized in the 'other' column. DEGs are sorted from highest to lowest log2 foldchange. Row sections are ordered like column sections. **(C)** UMAP based on scRNAseq data depicted in A with PCs corrected for patient-specific effects using harmony. Cells are annotated according to their predominant phenotype. **(D)** Relative distribution of tumor cell phenotypes (left) and cancer cell abundance (right) across patients. The label on the y-axis shows the sample identification and tumor localization (primary tumor [PT] or lymph node metastasis [MET]). **(E)** UMAP based on scRNAseq data of 2,948 OSCC cells from patient HN01. Cells were annotated based on SNN clustering and the predominant phenotype. **(F)** Triangle heatmap of cosine similarity comparing the intratumoral heterogeneity across all patients. Cosine similarity is calculated between log2 fold changes from patient-specific clusters against all other tumor cells within the respective patient. Left side annotated are patient-specific clusters from patient 1 depicted in Figure 2A and right side from patient HN01 depicted in E. We included only patients with more than 50 tumor cells.

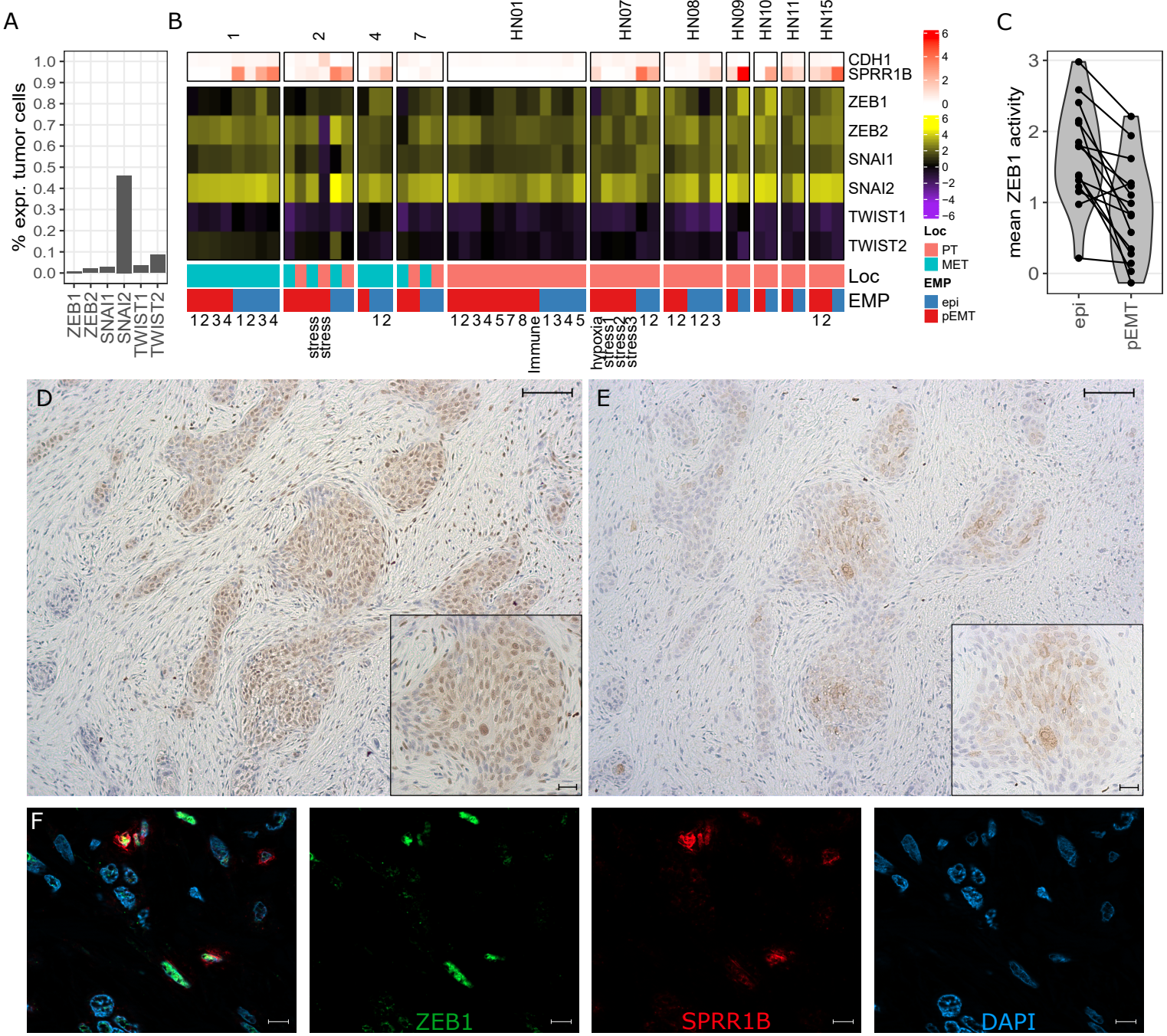

Figure 4 | ZEB1 is highly active in metastatic epithelial differentiated OSCC cells. (A) Percentage of tumor cells with detectable mRNA expression from scRNAseq (more than one UMI) encoding the indicted EMP-related transcription factors. **(B)** Mean inferred activity based on the target genes of the indicated transcription factors across tumor phenotypes from EMP-related patient-specific clusters. On top the log-normalized expression of *CDH1* and cornifin-B (*SPRR1B*) is shown, on bottom the localization (primary tumor [PT] or lymph node metastasis [MET]) and respective EMP phenotype of the cluster. **(C)** Mean activity of ZEB1 for epithelial differentiated and pEMT clusters of each patient, respectively. Connecting lines show dots belonging to the same patient. **(D, E)** ZEB1 (D) and cornifin-B (E) protein expression detected in serial sections by IHC of the primary tumor from patient 2; comparable areas are depicted. Scale bars equal 200 µm in overview and 100 µm in zoomed image. **(F)** Colocalized expression of ZEB1 (green) and cornifin-B (red) detected by double staining in the lymph node metastasis of patient 1 nuclei are stained in blue (DAPI), Scale bars equal 10 µm.