Genomic and Phenotypic Heterogeneity of Malignant Tumors and their Microenvironment

Dissertation

zur

Erlangung des Doktorgrades

Dr. rer. nat.

der Fakultät für

Biologie

an der

Universität Duisburg-Essen

vorgelegt von

Kai Horny

aus Dorsten

Datum der Abgabe: April 2023

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden in der Abteilung für Translationale Hautkrebsforschung der Universität Duisburg-Essen/Universitätsklinikum Essen und des Deutschen Konsortiums für Translationale Krebsforschung (DKTK)/Deutsches Krebsforschungszentrum (DKFZ) durchgeführt.

1. Gutachter: Prof. Dr. Dr. Jürgen C. Becker

2. Gutachter: Prof. Dr. Daniel Hoffmann

3. Gutachter: Prof. Dr. Ralf Gutzmer

Vorsitzender des Prüfungsausschusses: Prof. Dr. Ralf Küppers

Tag der mündlichen Prüfung: 11. September 2023

DOI: 10.17185/duepublico/81426
URN: urn:nbn:de:hbz:465-20240202-084300-6

Table of Contents

Abbreviations I
Abstract II
Zusammenfassung IV
1 Introduction 1
1.1 Genomic heterogeneity in Merkel cell carcinoma 1
1.1.1 Merkel cell carcinoma 1
1.1.2 Merkel cell polyomavirus is oncogenic in Merkel cell carcinoma 2
1.1.3 Genomic variations in Merkel cell carcinoma 3
1.2 Phenotypic heterogeneity in oral squamous cell carcinoma 4
1.2.1 Oral squamous cell carcinoma 4
1.2.2 Metastatic spread of malignant cells 4
1.2.3 Epithelial-mesenchymal plasticity 5
1.2.4 Cancer-associated fibroblasts 7
2 Aim 9
3 Contributed Articles 10
3.1 Mutational landscape of virus- and UV-associated Merkel cell carcinoma cell lines is comparable to tumor tissue. 11
3.2 Mesenchymal-epithelial transition in lymph node metastases of oral squamous cell carcinoma is accompanied by ZEB1 expression. 30
4 Discussion 99
4.1 Genomic diversity in Merkel cell carcinoma 99
4.2 Heterogeneity of EMP states in metastatic OSCC 101
4.3 Conclusions and outlook 103
5 References 105
6 Acknowledgements 117
7 Eidesstattliche Erklärung A1

Abbreviations

CAF	Cancer-associated fibroblast
CNV	Copy number variation
ECM	Extracellular matrix
EMP	Epithelial-mesenchymal plasticity
EMT	Epithelial-mesenchymal transition
HNSCC	Head and neck squamous cell carcinoma
HPV	Human papillomavirus
ICI	Immune checkpoint inhibitor
LTA	Large T antigen
MCC	Merkel cell carcinoma
MCPyV	Merkel cell Polyomavirus
MET	Mesenchymal-epithelial transition
OSCC	Oral cavity squamous cell carcinoma
SBS	Single base substitution signature
SCC	Squamous cell carcinoma
scRNAseq	Single cell RNA sequencing
SNV	Single nucleotide variation
STA	Small T antigen
TMB	Tumor mutational burden
TME	Tumor microenvironment
UV	Ultraviolet

Abstract

The advance of next-generation sequencing approaches facilitates the exploration of molecular characteristics across cancer genomes, transcriptomes, methylomes, and many other levels. The increasing number of sequenced tumor samples reveals the molecular heterogeneity between patients, primary and metastatic lesions and cells as well as molecular mechanisms associated with metastatic dissemination or treatment resistance. These potentially can influence tumor progression and therapeutic intervention.

The first part of this thesis investigates the genomic landscape of Merkel cell carcinoma (MCC) cell lines. MCC is a rare and aggressive neuroendocrine skin cancer associated either with ultraviolet (UV)-light induced DNA damage or clonal integration of the Merkel cell polyomavirus (MCPyV). The associated genomic variations are well studied in tumor tissue, but not in MCC cell lines that are frequently used as preclinical models. Therefore, I analyzed whole-exome sequencing data of 3 UV- and 6 MCPyV-associated MCC cell lines demonstrating that their mutational landscape is congruent with MCC tumor tissue. UV-associated cell lines have a high tumor mutational burden (TMB), single-base substitution signatures (SBS) reflecting UV-light induced DNA damage, pathogenic variants in *TP53* and *RB1* genes and a high number of copy number variations (CNVs). In contrast, MCPyV-associated cell lines show only few mutations with no specific signature, but characteristic CNVs. We further found additional genomic heterogeneity with an amplified and overexpressed c-*MYC* in the UV-associated cell line UM-MCC34, indicating the presence of potential further MCC subtypes.

The second part of this thesis analyzes the transcriptomic and cellular heterogeneity of oral cavity squamous cell carcinoma (OSCC). OSCC is caused by an abuse of tobacco- and alcohol consumption and frequently develops locoregional lymph node metastases. The metastatic spread is often mediated by epithelial-mesenchymal transition (EMT), in which cells change from a more adherent towards a migratory phenotype, and the reverse process of mesenchymal-epithelial transition (MET). Hence, I analyzed single cell mRNA sequencing (scRNAseq) data of 14 primary tumors, 9 metastatic and 5 tumor-free lymph nodes of in total 16 OSCC patients. First, the analysis focused on one metastatic lymph node from which the malignant cell phenotypes and their developmental relationship are inferred. Besides a partial EMT state, we found another intermediate state possibly resembling a partial MET. This state characterizes a progressive epithelial differentiation marked by higher expression of the keratinocyte-envelope protein cornifin B, increasing number of CNVs and high activity of the EMT-inducing transcription factor ZEB1. This population was also validated in the other malignant samples using scRNAseq data and by detecting co-expressed ZEB1 and cornifin B protein. Additionally, we could link variations in partial EMT phenotypes with tumor-promoting

properties such as angiogenesis and investigated the tumor microenvironment (TME) utilizing scRNAseq and additional bulk transcriptomic data. The TME of OSCC is mainly composed of fibroblasts and their subtypes, most prominently an immunomodulatory phenotype, that were similarly found in primary and metastatic lesions. In summary, the second part revealed the complexity of the epithelial-mesenchymal plasticity that includes additional intermediate manifestations linked with ZEB1 and cornifin B expression. This indicates a role of ZEB1 beyond EMT-induction that could be associated with epithelial differentiation.

The results of this thesis add insights into the genomic and transcriptomic heterogeneity in cancer, which could be important for metastatic disease and malignant progression. The identified phenotypes may provide the basis for generating molecular markers with predictive or prognostic relevance supporting future therapy decision making.

Zusammenfassung

Durch den Fortschritt in der *next-generation*-Sequenzierung ist es möglich, die molekularen Charakteristika von Krebserkrankungen in Genomen, Transkriptomen, Methylomen und vielen anderen Ebenen zu erforschen. Die steigende Anzahl an sequenzierten Tumoren offenbart die molekulare Heterogenität zwischen erkrankten Personen, primären und metastatischen Läsionen und einzelnen Zellen sowie die molekularen Mechanismen, die mit Metastasierung oder Therapieresistenzen assoziiert sind. Diese können großen Einfluss auf die weitere Entwicklung des Tumors und den therapeutischen Erfolg haben.

Der erste Teil dieser Dissertation untersucht die genomische Vielfalt von Merkelzellkarzinom (MZK)-Zelllinien. Das MZK ist ein seltener, aggressiver und neuroendokriner Hauttumor, der entweder mit DNS-Schäden durch Ultraviolettes (UV)-Licht oder der klonalen Integration des Merkelzellpolyomavirus (MZPyV) einhergeht. Die damit verbundenen genomischen Veränderungen sind im Tumorgewebe gut erforscht, aber nicht in oft als präklinische Modelle genutzten Zelllinien. Deswegen habe ich *whole-exome*-Sequenzdaten von 3 UV- und 6 MZPyV-assoziierten MZK-Zelllinien untersucht, die zeigen, dass die Mutationslandschaft von MZK-Zelllinien denen von Tumorgeweben entspricht. UV-assoziierte Zelllinien haben eine hohe Mutationslast, Signaturen von Einzelnukleotidveränderungen, die den durch UV-Licht verursachten DNS-Schaden widerspiegeln, pathogene Varianten in den Genen *TP53* und *RB1* und eine hohe Anzahl an Kopienzahlveränderungen (KZV). Im Gegensatz dazu haben MZPyV-assoziierte Zelllinien nur wenige Mutationen ohne spezifische Signatur, aber mit charakteristischen KZV. Wir fanden außerdem weitere genomische Heterogenität zwischen den MZK-Zelllinien, die auf die Existenz weiterer MZK Subtypen hinweist: Das *c-MYC* Gen ist in der UV-assoziierten Zelllinie UM-MCC34 amplifiziert und überexprimiert.

Der zweite Teil dieser Doktorarbeit untersucht die transkriptomische und zelluläre Heterogenität von oralen Plattenepithelkarzinomen (oPEK). oPEK werden durch exzessives Rauchen oder Alkoholmissbrauch verursacht und entwickeln häufig lokoregionäre Lymphknotenmetastasen. Die Metastasierung ist dabei oft durch die Epithelial-Mesenchymale Transition (EMT) getrieben, in den Zellen von einem mehr adhärenten zu einem migratorischen Phänotyp wechseln, und dem umgekehrten Prozess der Mesenchymal-Epithelialen Transition (MET). Daher habe ich *single-cell* mRNS-Sequenzierungsdaten (scRNSseq) von 14 Primärtumoren, 9 metastasierten und 5 tumorfreien Lymphknoten in insgesamt 16 oPEK-Patient:innen analysiert. Zuerst hat sich die Analyse auf einen metastasierten Lymphknoten und die dort enthaltenen malignen Zellphänotypen sowie deren evolutionäre Zusammenhänge fokussiert. Neben einem partiellen EMT-Zustand haben wir außerdem einen anderen, intermediären Zustand entdeckt, der auf eine partielle MET

hindeutet. Dieser Zustand ist charakterisiert durch eine progressive, epitheliale Differenzierung, gekennzeichnet durch höhere Expression des Keratinozyt-Hüllprotein Cornifin B, steigende Anzahl an KZV und hohe Aktivität des EMT-induzierenden Transkriptionsfaktors ZEB1. Diese Zellen wurden ebenso in anderen Proben gefunden, darunter in den scRNS Datensätzen und mittels Immunhistochemie als Koexpression von ZEB1 und Cornifin B Protein. Des Weiteren konnten wir Variationen im partiellen EMT-Phänotyp mit tumorfördernden Eigenschaften, wie zum Beispiel der Angiogenese, verlinken sowie die Tumor Mikroumgebung (TMU) mithilfe von scRNSseq und konventionellen RNS-Daten untersuchen. Diese haben gezeigt, dass die TMU von oPEK hauptsächlich aus Fibroblasten und deren Subtypen besteht, wovon Immunmodulierende Fibroblasten am meisten präsent sind und die sowohl in Primärtumoren als auch in Metastasen vorkommen. Zusammenfassend zeigt der zweite Teil die Komplexität der Epithelial-Mesenchymalen Plastizität, die weitere intermediäre Ausprägungen assoziiert mit ZEB1 und Cornifin B Expression inkludiert. Damit deutet diese Arbeit auch auf eine weitergehende Rolle von ZEB1 hin, die über die EMT-Induktion hinausgeht und mit epithelialer Differenzierung verbunden sein könnte.

Die Resultate dieser Dissertation haben weitere Erkenntnisse über die genomische und transkriptomische Heterogenität in Krebserkrankungen hervorgebracht, die potenziell wichtig für die Metastasierung und weitere Tumorentwicklung sein könnten. Die identifizierten Phänotypen könnten als Basis für die Entwicklung von molekularen Markern dienen, die eine prädiktive oder prognostische Relevanz haben und zukünftige Therapieentscheidungen unterstützen könnten.

1 Introduction

1.1 Genomic heterogeneity in Merkel cell carcinoma

1.1.1 Merkel cell carcinoma

Merkel cell carcinoma (MCC) is a rare, neuroendocrine skin cancer with high aggressiveness and mortality (Becker et al., 2017). It has a generally low incidence in the population reported as 0.3-0.4 cases per 100,000 persons per year in Schleswig-Holstein, Germany (Eisemann et al., 2014). In other countries of the northern hemisphere, the incidence rate ranges between 0.1-0.8 cases, for example, 0.66 in the USA (Jacobs et al., 2021), 0.35 in the Netherlands (Reichgelt & Visser, 2011), 0.25 in France (Fondain et al., 2018), 0.31 in Spain (Rubió-Casadevall et al., 2016) and 0.3 in Scotland (Samuel et al., 2015), as reviewed by Gauci *et al.* (Gauci et al., 2022). Incidence rates generally tend to increase, for example, in the USA and Germany (Eisemann et al., 2014; Paulson et al., 2018; D. Schadendorf et al., 2017; Stang et al., 2018) and is generally higher in countries closer to the equator (Stang et al., 2018), indicating its correlation with high UV-light exposure.

The risk for developing MCC is increased in individuals that are of old age, male sex, have chronic sun-exposed skin or a suppressed immune system, for example due to hematological diseases, human immunodeficiency virus infection, or organ transplantation (Becker et al., 2017; Gauci et al., 2022). The five-year survival rate ranges from 48 % to 72 % and decreases with progressing metastatic spread from high rates in cases of local disease (51 %) towards lower rates in cases developing regional (35.4 %) and distant (13.5 %) metastases (Gauci et al., 2022; Harms et al., 2016; Song et al., 2021). Indeed, upon diagnosis nearly half of the patients (42 %, 56/133) already present with a metastasized tumor in stage III or IV, and 59 % (78/133) of all patients developed metastases from which 13 % (10/78) are distant metastases (Song et al., 2021).

Today, MCC can be treated using immune checkpoint therapy. To avoid an excess adaptive immune response, for example by chronic infection, the programmed-death receptor PD-1 and its ligand PD-L1 can induce an exhausted T cell phenotype (Wherry & Kurachi, 2015). MCC cells can hijack this mechanism for immune escape, which is disrupted by Immune-checkpoint inhibitors (ICIs) and leads to durable responses in MCC patients (Dirk Schadendorf et al., 2017). Two ICIs are currently used in treatment of MCC: The anti-PDL1 antibody avelumab and the anti-PD1 antibody pembrolizumab, the latter being included in the 2017 NCCN guidelines for systemic treatment options in MCC (Bichakjian et al., 2018; Chan et al., 2018). Other ICIs such as, for example, the anti-PD1 antibody nivolumab are currently investigated in

neoadjuvant or adjuvant treatment settings (Becker et al., 2022; Topalian et al., 2020). However, clinical and biological markers for predicting ICI therapy response are needed as only around 65 % of patients show a clinical response or stable disease and around one third of patients died during the first 12 months of follow-up time (Spassova et al., 2022).

1.1.2 Merkel cell polyomavirus is oncogenic in Merkel cell carcinoma

In general, the formation of a malignant tumor can be attributed to individual exposure to risk factors, for example tobacco- and alcohol consumption, environmental carcinogens such as UV-light, and also infectious pathogens such as the human papillomavirus (HPV) (Narayanan et al., 2010; Tran et al., 2022; zur Hausen, 1991). MCC is distinguished in two different types, which show different cancerigenesis and prevalence in Europe: Virus-associated MCC tumors represent 80 % of MCC cases and integrated the Merkel cell polyomavirus (MCPyV) genome into the genome of the malignant cells, while the other 20 % include MCC driven by UV-light induced DNA damages without an integrated MCPyV genome (Becker et al., 2017; von der Grün et al., 2019).

In 2008 a study by Feng *et al.* found the association between Merkel cell carcinoma and the integration of the MCPyV genome using digital transcriptome subtraction on high-throughput mRNA sequencing data (Feng et al., 2008). The MCPyV has a circular, double-stranded DNA genome with 5,387 base pairs and contains the viral genes small T antigen (STA), large T antigen (LTA) and the viral envelope proteins VP1 and VP2 (Becker et al., 2017). In MCC, both the wildtype STA and a truncated version of the LTA are expressed (Becker et al., 2017; Feng et al., 2008; Starrett et al., 2020). The truncated form of LTA removes the helicase and DNA-binding domains and hence inactivates the viral capabilities for replication, p53 binding and induction of associated DNA damage responses (Cheng et al., 2013; Li et al., 2013; Shuda et al., 2008; Siebels et al., 2020). Both LTA and STA can transform human fibroblasts into malignant cells *in-vitro* and STA alone is sufficient for tumor formation in mice (Cheng et al., 2013; Verhaegen et al., 2015). Both viral proteins drive tumorigenesis through interactions with host proteins: LTA can bind and inhibit RB, which is important for cell cycle control and STA can form a complex with MYCL and E400 that transcriptionally activates a large number of genes, such as the oncogene *MYC* (Cheng et al., 2017; Hesbacher et al., 2016; Schrama et al., 2016). STA further activates proteasomal degradation of p53 through upregulation of ubiquitin ligases MDM2 and MDM4 (Park et al., 2019) and is involved in several other cellular processes such as translation through interaction with the phosphorylated 4E-BP1 (Shuda et al., 2011), cell motility through interaction with microtubules and the actin cytoskeleton (Knight et al., 2015; Stakaitytė et al., 2018), inhibition of protein degradation through FBXW7 (Kwun et al., 2013), and modulating glycolysis (Berrios et al., 2016). Especially, the inhibition of protein

degradation is linked to accumulation of cell-cycle related genes such as cyclin E and prevention of LTA degradation (Kwun et al., 2013; Verhaegen et al., 2015).

1.1.3 Genomic variations in Merkel cell carcinoma

Endogenous and environmental factors generate somatic DNA damage within cells, which can lead to pathogenic mutations that induce a malignant transformation (Stratton et al., 2009). The increasing amount of next-generation sequencing data enabled the resolving of mutational patterns that are associated with those mutagens (Alexandrov et al., 2020; Alexandrov et al., 2013). These patterns are reflected by single base substitution signatures (SBS) that have characteristic frequencies of trinucleotide sequences, which include the two surrounding base pairs from a single nucleotide variation (SNV). SBS can reflect common mutagenic processes: SBS1 results from the spontaneous deamination process at CpG-islands converting cytosine to thymine (C>T) that accumulate during aging (Nik-Zainal et al., 2012). Another example is SBS3, which reflects defective homologous recombination caused by *BRCA1/2* mutations, and SBS6, which reflects defective DNA mismatch repair and microsatellite instability (Alexandrov et al., 2013; Nik-Zainal et al., 2012; Zámborszky et al., 2017). Some SBS can be directly linked to environmental mutagens: SBS4 is associated with tobacco smoke and has prominent C>A mutations, while SBS7a-d are associated with UV-light induced DNA damage, having characteristic C>T substitutions, mainly discovered in malignant melanoma and skin fibroblasts samples (Alexandrov et al., 2013; Hayward et al., 2017; Nik-Zainal et al., 2015; Saini et al., 2016). Such UV-associated C>T substitutions arise from cyclobutane pyrimidine dimers that cause the double base substitutions CC>TT (Pfeifer et al., 2005).

MCPyV-negative MCC are characterized by the presence of UV-light induced DNA damage signatures, a high tumor mutational burden (TMB) and a high number of copy number variations (CNVs) as well as pathogenic mutations in tumor-suppressor genes and oncogenes such as *TP53* and *RB1* (Goh et al., 2015; Harms et al., 2015; Knepper et al., 2019; Starrett et al., 2020; Wong et al., 2015). Indeed, abrogation of p53 and RB function either via genomic aberrations or MCPyV proteins is a common hallmark across MCC types and also found within other neuroendocrine cancers such as small cell lung cancer (George et al., 2015). MCPyV-positive MCC lacks a high TMB and pathogenic genomic variations but show a low number of characteristic CNV patterns and additional variations in the integrated MCPyV genome (Paulson et al., 2009; Starrett et al., 2020). Smaller subsets of MCC samples harbor additional pathogenic or coding mutations in genes such as *KMT2D*, *PIK3CA*, *TERT, NOTCH* family genes, or other genomic aberrations such as a copy number amplification of the *L-MYC* gene or deletion of *RB1* or *PTEN* (Knepper et al., 2019; Paulson et al., 2009).

1.2 Phenotypic heterogeneity in oral squamous cell carcinoma

1.2.1 Oral squamous cell carcinoma

Head and neck squamous cell carcinoma (HNSCC) is a class of cancer types covering a specific anatomical area including larynx, pharynx and oral cavity (Johnson et al., 2020). HNSCC is the sixth most common cancer worldwide with an incidence of 15.98 cases per 100,000 people per year as reported for Thuringia in Germany (Dittberner et al., 2020). While laryngeal and oral cavity squamous cell carcinoma (OSCC) are mainly induced by tobacco- and alcohol consumption, oropharyngeal cancer is more commonly induced by persistent infection with HPV (Chaturvedi et al., 2011; Johnson et al., 2020). In general, the main cause of cancer-related deaths is attributed to metastatic disease representing up to 90 % of these deaths (Lambert et al., 2017). About 40 % of OSCC patients develop regional lymph node metastases that are often occult during initial diagnosis and can worsen prognosis (Noguti et al., 2012). Those locoregional metastases are mostly treated with elective neck dissection surgery during primary tumor removal as well as systemic treatments such as radiotherapy, chemotherapy, targeted anti-EGFR therapy with cetuximab and immunotherapy (Cramer et al., 2019; D'Cruz et al., 2015).

1.2.2 Metastatic spread of malignant cells

During metastatic dissemination, primary malignant cells face several, dynamically changing obstacles to effectively disseminate into regional or distant tissues (Lambert et al., 2017). First, cells have to overcome cell-cell adhesion, especially in epithelial tumors, break through tissue barriers and invade into and travel through the lymphatic or vascular system (Lambert et al., 2017). Within the circulatory system, cells can travel alone as single cells or in collective cell clusters that may enhance metastatic dissemination (Cheung et al., 2016). During that journey, malignant cells are surrounded by immune cells that are threatening their survival. To avoid them, malignant cells feature immune escape mechanisms, for example the inhibition of clearance by natural killer cells through interactions with platelets (Kopp et al., 2009). In addition, malignant cells can mediate immune evasion in premetastatic lymph node tissue, for example via extracellular vesicles released from the primary tumor that bind to macrophages and promote immunosuppressive B cells (Jones et al., 2018; Pucci et al., 2016). From the circulatory system, malignant cells or cell clusters need to extravasate into regional or distant tissues by attaching to endothelial cells and undergoing transendothelial migration (Reymond et al., 2013). Once within the metastatic site, cells facilitate metastatic colonization by exhibiting stemness features for differentiation and self-renewal, proliferation, metabolic reprogramming, and other, often tissue-specific adaptations (Jehanno et al., 2022). Heterogenous cancer cells can adapt to the different conditions during the metastatic cascade

by reversibly switching their cellular phenotype, i.e., having a high plasticity, and by the emergence of less abundant but better adapted tumor cell clones (Jehanno et al., 2022; Lawson et al., 2018). Several biological programs can equip malignant cells with abilities for overcoming the obstacles during their journey. For example, if cells experience hypoxic conditions, the hypoxia-associated transcription factor HIF1α is activated, which induces angiogenesis through expression of vascular endothelial growth factors (Majmundar et al., 2010). One of the most important phenotypic switches promoting metastatic dissemination in carcinomas is the epithelial-mesenchymal transition (EMT) in which epithelial tumor cells change towards a more migrative and invasive phenotype (Nieto et al., 2016). The reverse process of mesenchymal-epithelial transition (MET) can hence favor metastatic outgrowth (Nieto et al., 2016; Yao et al., 2011).

1.2.3 Epithelial-mesenchymal plasticity

EMT is a developmental process in which cells change from an adherent to a more migrative phenotype, i.e., from the epithelial phenotype with stronger cell-cell adhesion and apical-basal polarity towards a more mesenchymal phenotype with elongated, spindle-like cell shape and higher motility (Brabletz et al., 2021). During embryonal development and wound healing, cells undergo EMT to migrate through tissue barriers and reach, for example, the wound site (Nieto et al., 2016). Hence, malignant cells that underwent EMT can more easily invade surrounding tissues, while the reverse process MET involves epithelial differentiation and can likewise support metastatic outgrowth (Brabletz et al., 2021; Nieto et al., 2016; Yao et al., 2011).

EMT is a dynamic, reversible, and controlled process which is composed of a large spectrum of phenotypic changes, that can be better represented by the term epithelial-mesenchymal plasticity (EMP) as recently proposed (Yang et al., 2020). While EMP is a change in cellular properties, most studies aim to characterize EMP through molecular markers such as gene expression changes (Yang et al., 2020). These changes involve the transcriptional regulation through EMP-related transcription factors ZEB1/2, Snail and Slug (SNAI1/2), and TWIST1/2 which have in common the downregulation of epithelial and upregulation of mesenchymal genes (Brabletz et al., 2021; Dongre & Weinberg, 2019; Stemmler et al., 2019). For example, ZEB1 or Snail expression is associated with decrease of E-cadherin, an important protein for cell-cell adhesion (Batlle et al., 2000; Cano et al., 2000; Sánchez-Tilló et al., 2010). Also, Twist1 enhances expression of the mesenchymal gene vimentin by inducing expression of a circular RNA that diminishes vimentin-repressing microRNAs (Meng et al., 2018). EMP can also be governed through other mechanisms: Within a pancreatic ductal adenocarcinoma mouse model, a protein internalization mechanism represses the epithelial phenotype (Aiello et al., 2018). Changes in EMP are transcriptionally heterogenous and depend on the molecular and cellular context: Recent meta-analyses of single cell mRNA sequencing (scRNAseq)

studies across multiple cancer types revealed their high transcriptional diversity while having common regulatory networks and conserved genes such as the mesenchymal genes vimentin, fibronectin and integrins (Cook & Vanderhyden, 2022; Kinker et al., 2020). Signaling molecules such as TGF-β, EGF and TNF can initiate these diverse EMP changes within malignant cells, indicating the high relevance of the tumor microenvironment in regulating EMP (Cook & Vanderhyden, 2020). The signaling molecule TGF-β is frequently expressed in the metastatic environment of OSCC and induces EMT by upregulation of SNAI1/2, TWIST1/2 and ZEB1 and downregulation of miR-200 (Dongre & Weinberg, 2019; Gregory et al., 2008; Puram et al., 2017). TGF-β stabilizes EMT by a positive feedback loop, in which for example, SNAI1 and SNAI2 expression induces expression of TGF-β pathway genes (Dhasarathy et al., 2011). In addition, TGF-β also plays a major role in immune evasion in colorectal cancer mouse models (Tauriello et al., 2018).

In controlled *in-vitro* experiments, TGF-β-induced and spontaneous EMT in epithelial cells is accompanied by a continuous trajectory of cells changing their transcriptome towards less epithelial and more mesenchymal genes expressed (McFaline-Figueroa et al., 2019). Disturbing specific pathways in this model, for example KRAS/MAPK and TGF-β signaling, led to cells accumulating in intermediate states in this continuous trajectory (McFaline-Figueroa et al., 2019). Those intermediate EMT states were previously observed within tumor samples and are described as partial EMT phenotypes as they display both epithelial and mesenchymal traits (Kinker et al., 2020; Pal et al., 2021; Pastushenko et al., 2018; Puram et al., 2017; Yang et al., 2020). Partial EMT cells were found throughout the metastatic cascade, in primary and metastatic HNSCC lesions as well as in circulating tumor cells of breast cancer patients (Puram et al., 2017; Yu et al., 2013). Indeed, the mixture of mesenchymal and epithelial characteristics can be advantageous for metastasis formation: within a murine model of prostate cancer partial EMT not only enhances motility but also plasticity and stemness for tumor-initiation at the new metastatic site (Pal et al., 2021; Ruscetti et al., 2015). Moreover, partial EMT states are more likely to support metastatic spread through collective clusters, while complete EMT supports single cell dissemination in a pancreatic ductal adenocarcinoma mouse model (Aiello et al., 2018). EMP is interlinked with a multitude of other signaling pathways and cellular processes that might be relevant for metastatic dissemination, including senescence, metabolic reprogramming, hypoxia response, immune evasion and therapy resistance (Brabletz et al., 2021; Cook & Vanderhyden, 2020; Kinker et al., 2020). For example, hypoxia may contribute to controlling EMP, as HIF1α is described to mediate hypoxia-associated EMT induction (Saxena et al., 2020; Tam et al., 2020).

1.2.4 Cancer-associated fibroblasts

The tumor tissue is not only composed of malignant cells but also inhabits a multitude of stromal and immune cells in the tumor microenvironment (TME) (Anderson & Simon, 2020). The advance of single cell technologies enabled the detailed characterization of TME components, revealing heterogenous phenotypes within each cell type (Christensen et al., 2022). In the stromal TME, fibroblasts are the major cellular component and have important homeostatic functions within the tissue (Kalluri, 2016). Malignant tumors corrupt these cells and modify them towards cancer-associated fibroblasts (CAFs) which can promote tumor growth, angiogenesis, metastatic invasion and immune evasion (Joshi et al., 2021). For example, specific fibroblast subsets in breast cancer cell lines secrete *CXCL12* (aka SDF-1) and TGF-β, which induces EMT and increases the proliferation and migration of malignant cells (Pelon et al., 2020; Yu et al., 2014). Hence, the transcriptional heterogeneity of fibroblasts in the TME is studied extensively *ex-vivo* to avoid the risk of *in-vitro* artefacts (Biffi & Tuveson, 2021; Elyada et al., 2019; Galbo et al., 2021; Kürten et al., 2021; Puram et al., 2017; Qian et al., 2020; Solé-Boldo et al., 2020).

The phenotypes of fibroblasts reflect their diverse functions within the tissue: One type is composed of activated fibroblasts that maintain the extracellular matrix (ECM) by secreting collagens and matrix metallopeptidases (MMPs) (Kalluri, 2016). These ECM-producing fibroblasts were also named secretory, desmoplastic or matrix CAFs and were found in healthy skin, breast cancer, HNSCC, melanoma and lung cancer (Bartoschek et al., 2018; Galbo et al., 2021; Pelon et al., 2020; Solé-Boldo et al., 2020). Depending on the tissue or cancer type, the phenotype of ECM-producing fibroblasts can slightly deviate. For example, mesenchymal fibroblasts in healthy skin were also associated with cartilage and bone development (Solé-Boldo et al., 2020). A study investigating lung, colorectum and ovary cancer reports two clusters that might represent an ECM-producing phenotype with high activity in glycolysis, hypoxia and EMT through expression of matrix metalloproteinases and the TGF-β signaling co-activator *COMP* (Qian et al., 2020). In HNSCC, another ECM-producing phenotype is also associated with several hypoxia and EMT-related transcription factors such as *HIF1A*, *SNAI2* and *TWIST2* (Puram et al., 2017). Another fibroblast phenotype is represented by myofibroblasts or vascular CAFs which modulate tissue structures by expressing, for example, alpha smooth muscle actin. These vascular CAFs were found in HNSCC, melanoma, breast, ovary, colorectum and lung cancer (Bartoschek et al., 2018; Costa et al., 2018; Galbo et al., 2021; Pelon et al., 2020; Puram et al., 2017; Qian et al., 2020). Also, in HNSCC samples these cells can express the cytokine *IL6* which likely influences the immune response (Puram et al., 2017). Importantly, fibroblasts can influence inflammatory responses by stimulating and attracting immune cells (Kalluri, 2016). A common feature of these immunomodulatory

fibroblasts, as found in melanoma, HNSCC and lung cancer, is the expression of chemokines including *CXCL12/14*, cytokines such as *IL6/7* and complement factors such as *C3* and *C7* together with low collagen expression (Galbo et al., 2021). Galbo *et al.* suggested that these fibroblasts are immunosuppressive; indeed, in breast cancer, immunomodulating fibroblasts secreted *CXCL12* which attracts immunosuppressive $CD4^{+}CD25^{+}$ T lymphocytes (Costa et al., 2018; Galbo et al., 2021; Pelon et al., 2020). Previous studies described a large variety of immunomodulatory fibroblasts, for example, expressing *CCL10*, *CXCL2* and *CXCL3* in pro-inflammatory healthy skin fibroblasts (Solé-Boldo et al., 2020).

Within HNSCC, the fibroblast phenotypes are comprehensively studied using scRNAseq (Chen et al., 2020; Galbo et al., 2021; Ji et al., 2020; Kürten et al., 2021; Puram et al., 2017). A recent study by Kürten *et al.* provided a comprehensive dataset of HPV-positive and negative HNSCC samples with high numbers of malignant, stromal and immune cells, distinguishing between normal activated CAFs low in *FAP*, *PDGFRA*, *LOX* and *MMPs,* elastic CAFs highly expressing *CXCL12/14* and *ELN*, and undefined CAFs (Kürten et al., 2021).

2 Aim

Malignant tumors have an intrinsic molecular heterogeneity with individual biological characteristics that can contribute to cancer progression or treatment resistance (Dagogo-Jack & Shaw, 2018). Molecular biomarkers inferred from these tumor-specific attributes could help to translate biological knowledge into clinically relevant information such as therapy response or treatment resistance predictions. The overall aim of this thesis is to unravel the genomic and phenotypic heterogeneity of malignant cells and their microenvironment, thereby creating the foundation for predictive biomarker discovery.

The first part of this thesis focuses on the genomic heterogeneity of MCC cell lines (Horny et al., 2021). MCC tumor tissue harbors several genomic alterations that drive the malignant transformation, especially in UV-associated MCC (Goh et al., 2015; Harms et al., 2015; Knepper et al., 2019; Starrett et al., 2020; Wong et al., 2015). However, most genomic studies lack description of MCC cell lines that are frequently used as preclinical models to study oncogenic mechanism. In addition, we previously demonstrated that often used MCPyV-negative variant MCC cell lines exhibit epigenetic features closely resembling squamous cell carcinoma (SCC) rather than MCC, questioning their usage as MCC cell line models (Gravemeyer et al., 2021). The presented work investigated the mutational landscape of 3 UV- and 6 MCPyV-associated MCC cell lines using whole-exome sequencing (Horny et al., 2021). Here, genomic characteristics were explored by bioinformatic analysis of coding mutations, significantly mutated genes, mutational signatures, and CNVs. The functional relevance of different mutated genes was verified using quantitative real-time PCR. Overall, this analysis aimed to validate if MCC cell lines reflect similar genomic properties as MCC tumor tissue. Moreover, comparing the genomic differences between MCC cell lines also exposed genomic subtypes of MCC with potential translational relevance.

The second part of this thesis investigates the cellular and transcriptional heterogeneity of OSCC using scRNAseq. Malignant cells can develop diverse phenotypes with different biological abilities, which may support metastatic dissemination (Jehanno et al., 2022). Hence, this work aimed on studying the transcriptional heterogeneity across primary and locoregional OSCC lesions and their potential relevance for malignant processes (Horny et al., 2023). Malignant and microenvironmental phenotypes were inferred from single cell and bulk transcriptomic data using modern bioinformatic methods and are verified by immunohistochemistry. A particular focus of this study was on EMP as it represents an important biological program for metastatic spread and colonization (Nieto et al., 2016). The insights into the transcriptional heterogeneity can guide the derivation of potential molecular biomarkers, for example for predicting the metastatic risk in OSCC.

3 Contributed Articles

The presented thesis consists of two original articles:

I. Kai Horny, Patricia Gerhardt, Angela Hebel-Cherouny, Corinna Wülbeck, Jochen Utikal, Jürgen C. Becker (2021). "Mutational Landscape of Virus- and UV-Associated Merkel Cell Carcinoma Cell Lines Is Comparable to Tumor Tissue." *Cancers*, 13(4), 649. doi:10.3390/cancers13040649

In the context of this doctoral work, the following article was accepted for publication at the Journal of Translational Medicine and was previously published as preprint in *bioRxiv*:

II. Kai Horny, Christoph Sproll, Lukas Peiffer, Frauke Furtmann, Patricia Gerhardt, Jan Gravemeyer, Nikolas H. Stoecklein, Ivelina Spassova, Jürgen C. Becker (2023). “Mesenchymal-epithelial transition in lymph node metastases of oral squamous cell carcinoma is accompanied by ZEB1 expression.” *Journal of Translational Medicine*, 21(1), 267. doi: 10.1186/s12967-023-04102-w*, accepted version.*

Additional supplementary tables from the first publication can be found online.

3.1 Mutational landscape of virus- and UV-associated Merkel cell carcinoma cell lines is comparable to tumor tissue.

Cumulative Thesis/Extent of Contribution

Cumulative thesis of Mr. Kai Horny

Author contributions

Title of the publication:

Mutational Landscape of Virus- and UV-Associated Merkel Cell Carcinoma Cell Lines Is Comparable to Tumor Tissue.

Authors:
Kai Horny, Patricia Gerhardt, Angela Hebel-Cherouny, Corinna Wülbeck, Jochen Utikal, Jürgen C. Becker

Contributions:

- Conception – 10 %: Planning the analysis workflow for DNA sequencing data.
- Conduction of experimental work – 10 %: Planning of qPCR experiment for MYC expression analysis.
- Data analysis – 80 %: Processing, analysis and visualization of mutational data including mutational burden, signatures, and pathogenic mutations.
- Statistical analysis – 80 %: Conducting statistical significance of mutated genes.
- Writing the manuscript – 70 %: Writing of Abstract, Introduction, Results, Discussion, Material and Methods, Conclusions and Figures.
- Revision of the manuscript – 70 %: Revising Abstract, Introduction, Results, Discussion, Material and Methods, Conclusions and Figures.

Signature of the Doctoral Candidate Signature of the Doctoral Supervisor

cancers

IMPACT FACTOR 6.575

Indexed in: PubMed

Article

Mutational Landscape of Virus- and UV-Associated Merkel Cell Carcinoma Cell Lines Is Comparable to Tumor Tissue

Kai Horny, Patricia Gerhardt, Angela Hebel-Cherouny, Corinna Wülbeck, Jochen Utikal and Jürgen C. Becker

Special Issue

The Biological and Clinical Aspects of Merkel Cell Carcinoma

Edited by

Dr. Virve Koljonen, Dr. Weng-Onn Lui and Prof. Dr. Jürgen C. Becker

MDPI

https://doi.org/10.3390/cancers13040649

cancers

MDPI

Article

Mutational Landscape of Virus- and UV-Associated Merkel Cell Carcinoma Cell Lines Is Comparable to Tumor Tissue

Kai Horny [1,2], Patricia Gerhardt [3], Angela Hebel-Cherouny [3], Corinna Wülbeck [3], Jochen Utikal [4,5] and Jürgen C. Becker [1,2,3,*]

1 Translational Skin Cancer Research, German Cancer Consortium (DKTK), 45141 Essen, Germany; k.horny@dkfz.de
2 German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
3 Department of Dermatology, University Medicine Essen, 45141 Essen, Germany; patricia.gerhardt@uni-due.de (P.G.); angela.cherouny@uk-essen.de (A.H.-C.); c.wuelbeck@dkfz.de (C.W.)
4 Skin Cancer Unit, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; j.utikal@dkfz.de
5 Department of Dermatology, Venerology and Allergology, University Medical Center Mannheim, 68167 Mannheim, Germany
* Correspondence: j.becker@dkfz.de; Tel.: +49-201-1836727

Simple Summary: Merkel cell carcinoma (MCC) is an aggressive, rare skin cancer which is caused either by a virus or chronic UV exposure. For both forms, distinct genetic alterations have been described; however, these observations were mostly made in tumor tissue. Since cancer cell lines are frequently used as preclinical models to investigate biological function, we considered it necessary to establish the genomic landscape of MCC cell lines by whole-exome sequencing. We confirmed the presence of UV-induced DNA damage, a high number of mutations and several coding mutations in virus-negative cell lines which were absent in virus-positive cell lines; these, however, harbored characteristic copy number variations, suggesting some virally caused genetic instability. Knowing the genomic features of MCC cell lines validates previous, and facilitates upcoming, experimental studies to discover their biological and translational relevance.

Abstract: Merkel cell carcinoma (MCC) is a rare, highly aggressive cutaneous malignancy that is either associated with the integration of the Merkel cell polyomavirus or chronic UV exposure. These two types of carcinogenesis are reflected in characteristic mutational features present in MCC tumor lesions. However, the genomic characteristics of MCC cell lines used as preclinical models are not well established. Thus, we analyzed the exomes of three virus-negative and six virus-positive MCC cell lines, all showing a classical neuroendocrine growth pattern. Virus-negative cell lines are characterized by a high tumor mutational burden (TMB), UV-light-induced DNA damage, functionally relevant coding mutations, e.g., in *RB1* and *TP53*, and large amounts of copy number variations (CNVs). In contrast, virus-positive cell lines have a low TMB with few coding mutations and lack prominent mutational signatures, but harbor characteristic CNVs. One of the virus-negative cell lines has a local *MYC* amplification associated with high *MYC* mRNA expression. In conclusion, virus-positive and -negative MCC cell lines with a neuroendocrine growth pattern resemble mutational features observed in MCC tissue samples, which strengthens their utility for functional studies.

Keywords: merkel cell carcinoma; merkel cell polyoma virus; UV; cell line; MYC; TP53; RB1; whole-exome; significantly mutated genes; copy number variation

Citation: Horny, K.; Gerhardt, P.; Hebel-Cherouny, A.; Wülbeck, C.; Utikal, J.; Becker, J.C. Mutational Landscape of Virus- and UV-Associated Merkel Cell Carcinoma Cell Lines Is Comparable to Tumor Tissue. *Cancers* **2021**, *13*, 649. https://doi.org/10.3390/cancers13040649

Academic Editor: Virve Koljonen
Received: 23 December 2020
Accepted: 2 February 2021
Published: 5 February 2021

1. Introduction

Merkel cell carcinoma (MCC) is a rare, highly aggressive neuroendocrine skin cancer. It is either associated with chronic Ultraviolet (UV)-light exposure or the genomic integration of the Merkel cell polyomavirus (MCPyV) [1,2]. Virus-associated MCCs are highly

prevalent in countries with high latitude, while UV-associated MCCs are more frequent in regions close to the equator [1–3].

The different forms of carcinogenesis of MCC are represented in various genomic features, as demonstrated by targeted [4–10], whole-exome [11–14] and whole-genome sequencing [15], as well as comparative genomic hybridization (CGH) [16–21]. Virus-negative MCCs are characterized by a high tumor mutational burden (TMB), the presence of UV-light-induced DNA damage, functional driver mutations, and high numbers of copy number variations (CNVs). Virus-positive MCCs have a very low TMB and lack known cancer-driving mutations and prominent mutational signatures. Characteristic CNV patterns have repeatedly been reported for virus-associated MCC.

The majority of genomic studies analyzed MCC tissue samples; only a few studies addressed MCC cell lines. CNV patterns of six virus-positive MCC cell lines were previously characterized using CGH [17]. Three virus-negative MCC cell lines, previously characterized by targeted sequencing, have variant growth characteristics [6]. Notably, the origin of "variant" MCC cell lines is controversial, since these have different growth and gene expression patterns to other "classical" virus-negative cell lines [22,23], which share neuroendocrine growth features—i.e., growing in suspension as spheroids—with virus-positive cell lines. Thus, since comprehensive mutational characterization of the MCC cell lines is missing, we analyzed the mutational landscape of cell lines that are frequently used in MCC research by whole-exome sequencing (WES).

2. Results

Whole-exome sequencing of the virus-positive cell lines WaGa, MKL-1, UKE-MCC3b, UM-MCC13, UM-MCC29 and PeTa, as well as the virus-negative cell lines UM-MCC9, UM-MCC32 and UM-MCC34, was performed using the SureSelect Exon V6 Kit on a HiSeq 4000 with, on average, 118 million reads per sample. Moreover, we directly compared the cell line PeTa with cryopreserved tissue from which PeTa has been established to assess possible differences between the cell line and tissue.

2.1. Mutational Burden and Signatures of MCC Cell Lines Are in Accordance with MCC Tissue Characteristics

Virus-negative MCC cell lines have a higher mutational burden with, on average, 44.5 mutations per megabasepairs (mut/Mbp), constituting, on average, 2.693 absolute mutations per cell line than virus-positive MCC cell lines, which contain, on average, 10.5 mut/Mbp (an average of 637 mutations) (Figure 1A, Table S1). Similarly, the number of coding mutations is higher in virus-negative MCC cell lines with, on average, 15.1 mut/Mbp (an average of 913 mutations per cell line, i.e., 33.6% of respective mutations) compared to 2.31 mut/Mbp in virus-positive MCC cell lines (on average, 140 mutations, i.e., 21.9% of mutations). The average fraction of missense (29.3%) and silent (15.6%) mutations in virus-negative cell lines is also higher than in virus-positive cell lines (16.4%/6.8%) (Figure 1A).

Since mutations are called between the respective cell line and the human reference genome hg19, the observed somatic TMB strongly depends on the filtering strategy for potential polymorphisms (Figures 1A and S1). Polymorphisms are identified using the variant allele frequency (VAF) reported in databases covering nonmalignant exomes and genomes. In general, exome databases cover 71.6% of all MCC cell line mutations while genome databases cover either 91.4% in the 1000 genomes database or 97.7% in the "genome aggregation database" (gnomAD) genome database (Figure S1B,C). Therefore, we filtered for VAFs greater than 0.001% with the comprehensive gnomAD genome database. This reduced the presented TMB of virus-positive MCC cell lines by 98.8% from, on average, 54,850 to 637 mutations per cell line. The TMB of virus-negative MCC cell lines shows a smaller reduction by 94.8% from, on average, 52,614, to 913 mutations (Figures 1B and S1A).

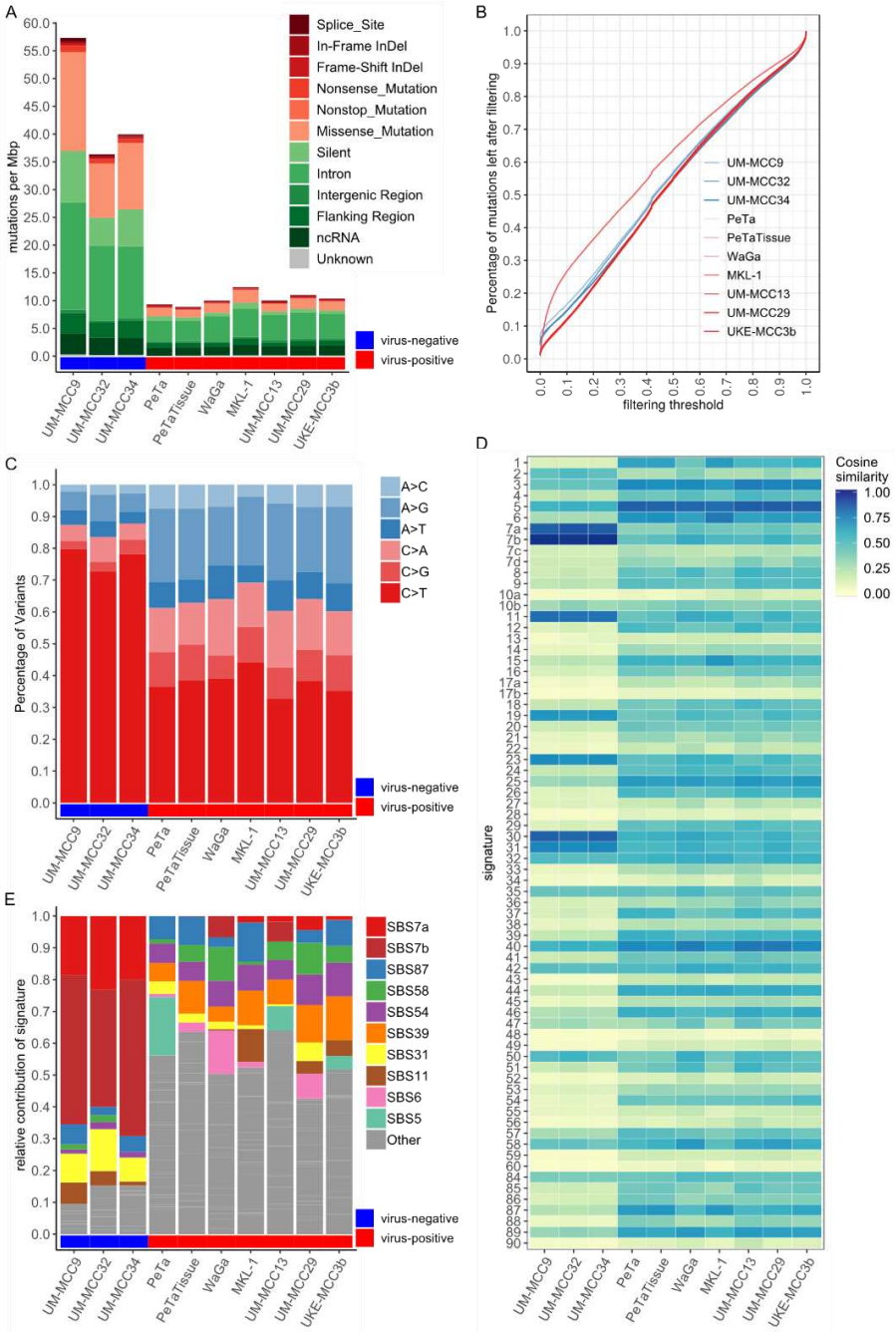


Figure 1. Virus-negative Merkel cell carcinoma (MCC) cell lines show high tumor mutational burden (TMB) and presence of UV-light-induced DNA damage, while virus-positive cell lines have low TMB and lack prominent signatures; (**A**) Mutational burden in mut/Mbp, color-coded by variant classification; (**B**) Filtering of polymorphisms in MCC cell lines showing the relative decrease in TMB (y-axis) with increasing variant allele frequency (VAF) threshold (x-axis) from gnomAD genome database; Figure S1A depicts the same plot with log-transformation of x-axis; (**C**) Contributions of base-pair transitions for single nucleotide variants (SNVs), normalized by total number of SNVs. Complementary transitions are merged in one category (e.g., G > A and C > T as C > T); (**D**) Cosine similarity between trinucleotide context frequencies (TCFs) of MCC cell lines and reference signatures reveals two distinct patterns for virus-positive and -negative cell lines; (**E**) Signature contribution of MCC cell lines after fitting to reference signatures. Signature contributions are normalized to total number of SNVs in the respective cell line. Signatures not reaching at least 10% contribution in at least one sample are summarized as "Other". Abbreviations: mut/Mbp: Mutations per Megabasepair, SBS: Single Base Substitution.

Virus-negative MCC cell lines are characterized by a high fraction of, on average, 77% C > T single-nucleotide variations (SNVs), as compared to 38% in virus-positive MCC cell lines (Figure 1C). This observation already suggests different forms of mutagenesis. Hints regarding the underlying mutagenic process can be retrieved from the first preceding and following basepair of an SNV, i.e., the trinucleotide context frequency (TCF). TCFs for virus-negative MCC cell lines show characteristic C > T transition patterns known to be caused by UV-induced mutagenesis (Figure S2) [24]. In contrast, virus-positive MCC cell lines have a "flat" TCF distribution, i.e., low frequencies for most categories, with only slightly elevated C > T and T > C transitions. For MKL-1 and UKE-MCC3b, there is a higher presence of C > T transitions with guanine as the following basepair; a pattern which often originates from spontaneous deamination of CpGs correlating with progressing age. The systematic comparison of the TCFs of MCC cell lines with reference mutational signatures reflecting defined mutagenic processes reveals distinct patterns for virus-negative and -positive cell lines (Figure 1D). Notably, the aging signature 1 and defective DNA mismatch repair signatures 6 and 15 were very similar to the TCF of MKL-1. Fitting reference signatures to the TCFs demonstrates a high contribution of signatures 7a and 7b for virus-negative MCC cell lines (on average, 67.2%), which are both associated with UV-light-induced DNA damage (Figure 1E). Virus-positive MCC cell lines generally have low individual signature contributions, with no prominent mutational signature present: approximately 50% of the total signature contribution for virus-positive MCC cell lines originates from signatures with less than 10% contribution (Figure 1E). Most of the absolute differences in the mutational burden between virus-negative and -positive MCC cell lines are due to signatures 7a and 7b. However, some mutational signatures have slightly higher signature contributions relative to others, namely signature 31 in virus-negative and signatures 5, 6, 11, 39, 54, 58 and 87 in virus-positive MCC cell lines. The reconstruction efficiency after signature fitting is, on average, higher in virus-negative (99.57%) than in virus-positive (96.95%) cell lines.

To test if the mutational landscape of the MCC cell lines indeed represents that of the original tumor, we compared the MCC cell line PeTa with cryopreserved tissue from which the cell line was derived (Figure 2, Table S2). The respective exomes share almost 80% of mutations, with 21% (120/565) being unique in the cell line and 17% (92/537) unique in the tumor tissue. Somatic variant calling for the cell line using the tissue as reference retrieved 124 variants, of which 38 (31%) were already among the germline-called variants in PeTa. Vice versa, somatic variant calling for the tissue using the cell line as reference resulted in 480 mutations, of which only five (1%) were present in germline-called variants of the tissue (Figure 2).

2.2. Mutations Altering Protein Structure

Next, we investigated mutations predicted to change the amino acid code and likely have an effect on protein function (Figure 3A). Virus-negative MCC cell lines harbor a higher number of nonsense mutations, i.e., mutations introducing a stopcodon, (on average, 53 mutations per cell line, corresponding to 2% of respective mutations) than virus-positive MCC cell lines (on average, six mutations, 0.9% of respective mutations) (Table S3). A total of 21% (≈12 mutations) and 33% (≈2 mutations) of nonsense mutations for virus-negative and -positive cell lines, respectively, are within genes of Hallmark Gene Sets, representing specific biological processes from the molecular signatures database (MSigDB) [25]. Nonsense mutations that are predicted to be pathogenic and cancer-related in ClinVar are in *RB1* in UM-MCC9 (rs794727481) and UM-MCC34 (rs121913304), in *BAP1* in UM-MCC32 (chr3.52437267.G > A), and in the tumor-suppressor gene *CHEK2* in MKL-1 (chr22.29091725.C > T).

Frameshift Insertions and deletions (InDels) are also enriched in absolute numbers in virus-negative MCC cell lines with, on average, 21 mutations (0.8% of respective mutations) compared to 12 mutations (2% of respective mutations) in virus-positive MCC cell lines (Table S3). A total of 18% (≈4 mutations) of virus-negative and 16% (≈2 mutations) of

virus-positive frameshift InDels are within Hallmark Gene Sets, among those annotated as pathogenic and cancer-related in ClinVar are *ERCC2* in UM-MCC9 (chr19.45855805.T > -), and *BRCA2* in UM-MCC29 (chr13.32911298.AAAC > -). UM-MCC9 and UM-MCC34 both harbor frameshift deletions in *TP53* (UM-MCC9: chr17.7577070.G > -, UM-MCC34: chr17.7579518.CTTCA > -), and UM-MCC32 a frame-shift deletion in *RB1* (chr13.48919254. CCAGTACCAAAGTTGATAAT > -); the inhibition of both tumor suppressors plays an important role in MCC carcinogenesis [4–9,11–14,26–28]. Besides the frameshift Indels and nonsense mutations, there are missense mutations of *TP53* in UM-MCC29 (rs1057520000) and UM-MCC32 (rs121912651). In UM-MCC9, following the frameshift deletion of *TP53*, are a missense (rs786201059) and a silent (chr17.7577558.G > A) mutation, while UM-MCC34 harbors only a silent mutation (chr17.7579516.G > A) before the frameshift deletion. For *RB1*, UM-MCC9 has a missense mutation (rs137853294) following the nonsense mutation, which, therefore, has no effect on the amino acid sequence. There are several other frameshift InDels that likely contribute to MCC carcinogenesis, for example, a frameshift deletion in *NOTCH1* in UM-MCC9 (chr9.139399867.AG > -). Moreover, only two nonstop mutations are found in this study, the first in the transcription repressor *GMNN* [29] in UM-MCC29 (rs757538616) and the other in chaperonin *TCP1* in WaGa (rs779397332) (Table S3).

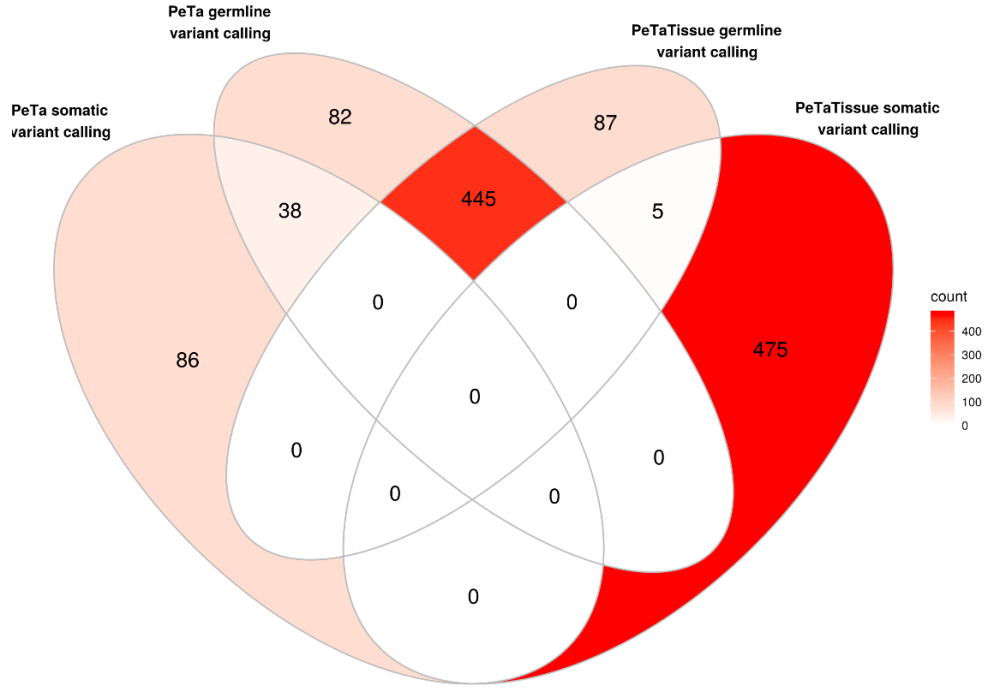


Figure 2. Comparison of unique mutations between PeTa and the respective tumor tissue. Venn-diagram showing how many of the same mutations are shared between the tumor cell line and tissue using either the tissue or the cell line as normal reference for somatic variant calling.

2.3. Significantly Mutated Genes

Next, we tested for genes with a significantly higher mutational burden as expected by chance, aka significantly mutated genes (SMGs) (Figure 3B–F, Tables S4–S6) [30]. In this approach, the mutations of several samples are aggregated and compared with a local background model of silent mutations for each respective gene [30]. This analysis was performed separately for all virus-negative (Figure 3B,D) and for all virus-positive (Figure 3C,E) cell lines. Downstream analysis was restricted to Hallmark Gene sets to focus on genes possibly relevant to MCC carcinogenesis. We evaluated the significance of the respective mutational burden by visualizing the distribution of *p*-values (Figure 3B–E). When correcting all *p*-values for multiple testing using Benjamini–Hochberg procedure, the genes *KRT4*, *MDK* and *CACNA1B* remained the only SMGs in the Hallmark Gene Sets present in both MCC types. Virus-negative MCC cell lines harbor more SMGs with a *p*-value lower than 0.01 within Hallmark Gene Sets

compared to virus-positive MCC cell lines, i.e., 16 vs. 5 genes, respectively (Figure 3D–F). Among the SMGs for virus-negative cell lines are *TP53* and *RB1*, which are frequently mutated tumor-suppressor genes in virus-negative MCC (Figure 3F) [4–9,11–14]. Of the SMGs found in virus-positive MCC cell lines, UM-MCC29 has a frame-shift deletion in the chromatin modifier *CBX3* (chr7.26248161.A > -), and UKE-MCC3b a falsely annotated nonstop mutation in *NAPA*, the latter composed of an in-frame insertion (chr19.47998837.- > ATTAAA) and deletion (chr19.47998843.GTT > -), resulting in the addition of two and deletion of one amino acid without introducing a stopcodon (Figure 3F).

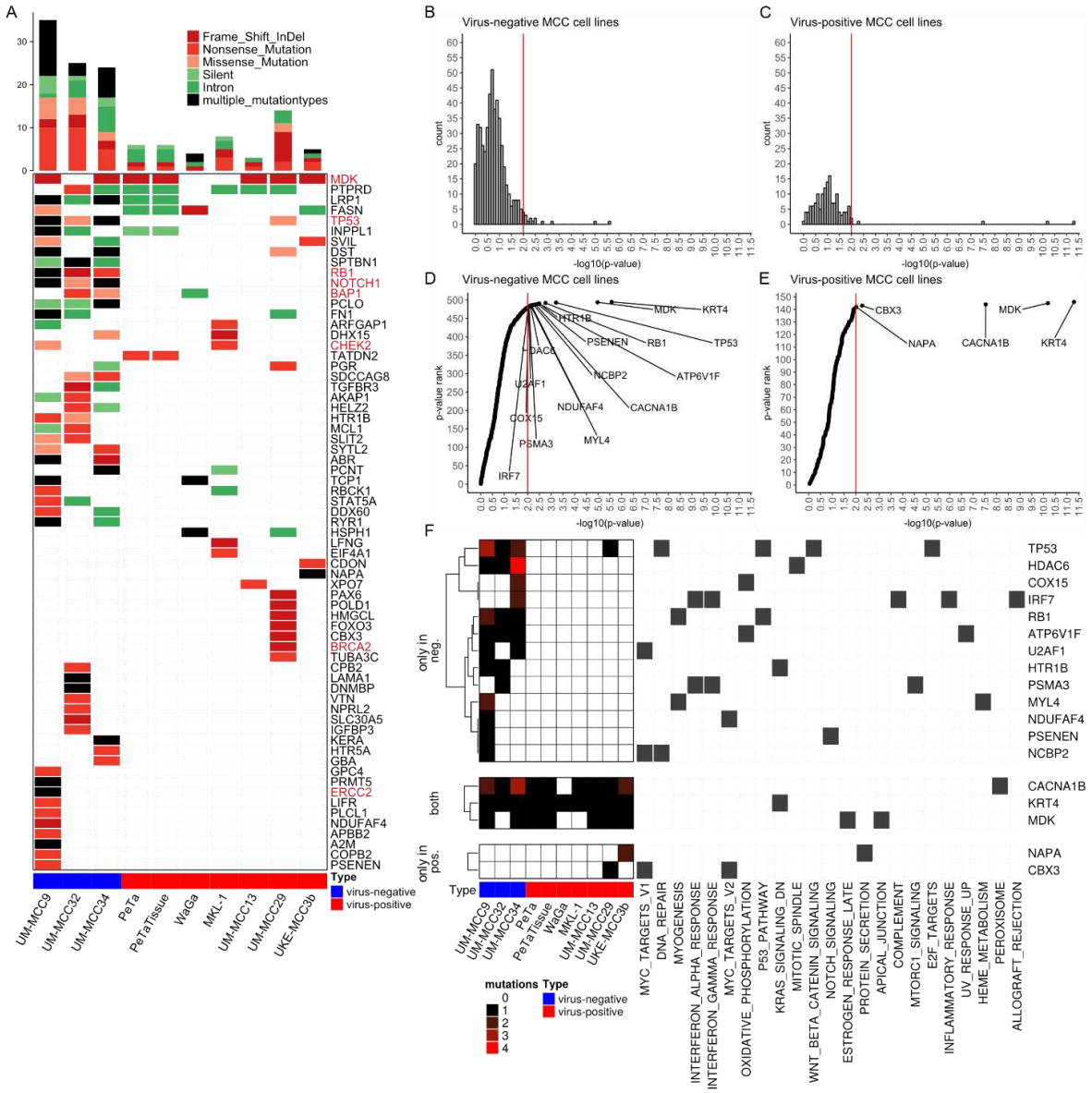


Figure 3. Virus-negative MCC cell lines have high number of coding mutations altering protein structure and significantly mutated genes. (**A**) Oncoplot showing genes selected by the following criteria: (i) containing either a frameshift InDel, nonsense or nonstop mutation and (ii) it is either within a Hallmark Gene Set or its mutation is annotated as pathogenic in ClinVar database. The number of mutations within the selected genes are depicted as bar chart. Both plots are colored by variant classification. Genes emphasized in red are discussed in the results section; (**B**–**E**) Distribution of *p*-values for identification of SMGs as histogram (**B**,**C**) and ranked by *p*-value (**D**,**E**) for virus-negative (**B**,**D**) and -positive (**C**,**E**) MCC cell lines; only genes present in Hallmark Gene Sets were taken into account, red lines indicate a *p*-value of 0.01, genes with a *p*-value of exactly 1 are not shown; (**F**) Mutational burden and involvement in biological processes of SMGs with *p*-value below 0.01 and presence in Hallmark Gene Sets. Abbreviations: MCC: Merkel cell carcinoma, InDel: Insertion and deletion, SMG: Significantly mutated genes.

We identified three SMGs (*KRT4*, *MDK* and *CACNA1B*) with extraordinary low *p*-values (Figure 3D–F) in all MCC cell lines. *KRT4* contains the exact same large in-frame insertion in all samples (rs11267392), which has a VAF of 87% in the 1000 genomes database. *MDK* comprise the exact same frameshift deletion in a cytosine-rich repeat in six samples, which is actually a mixture of a single-cytosine (chr11.46404342.C > -) and double-cytosine (chr11.46404342.CC > -) deletion. *CACNA1B* has the same large-scale insertion at a splice site in seven samples (chr9.140773612.- > ACGACACGGAGCCC-TATTTCATCGGGATCTTTTGCTTCGAGG CAGGGA, rs370237172).

2.4. Characteristic Copy Number Variation Patterns in Virus-Positive MCC Cell Lines

CNVs were determined from the exome sequencing data (Figure 4A, Table S7). The virus-negative cell lines UM-MCC9 and UM-MCC34, but not UM-MCC32, are characterized by numerous, varying CNVs covering most of the genome. Virus-positive MCC cell lines have less, but more characteristic CNVs, which include whole-chromosome gains of chr1 (UM-MCC29), chr5, chr7, chr8 (UM-MCC29), chr6 (WaGa, PeTa), chr11 (UM-MCC29), chr13 (UM-MCC29, PeTa), chr19, chr20 (UM-MCC29, WaGa) and a complete loss of chr10 (UM-MCC13). Several chromosomes are partially amplified, e.g., chr1q (UM-MCC13, PeTa), chr3q (MKL-1, PeTa) and chr11 (WaGa, PeTa), while others show partial losses, such as chr3p (UM-MCC13), chr8p (UM-MCC13, MKL-1, WaGa) and chr10q (MKL-1, WaGa, PeTa). Only UKE-MCC3b lacks any substantial copy number changes.

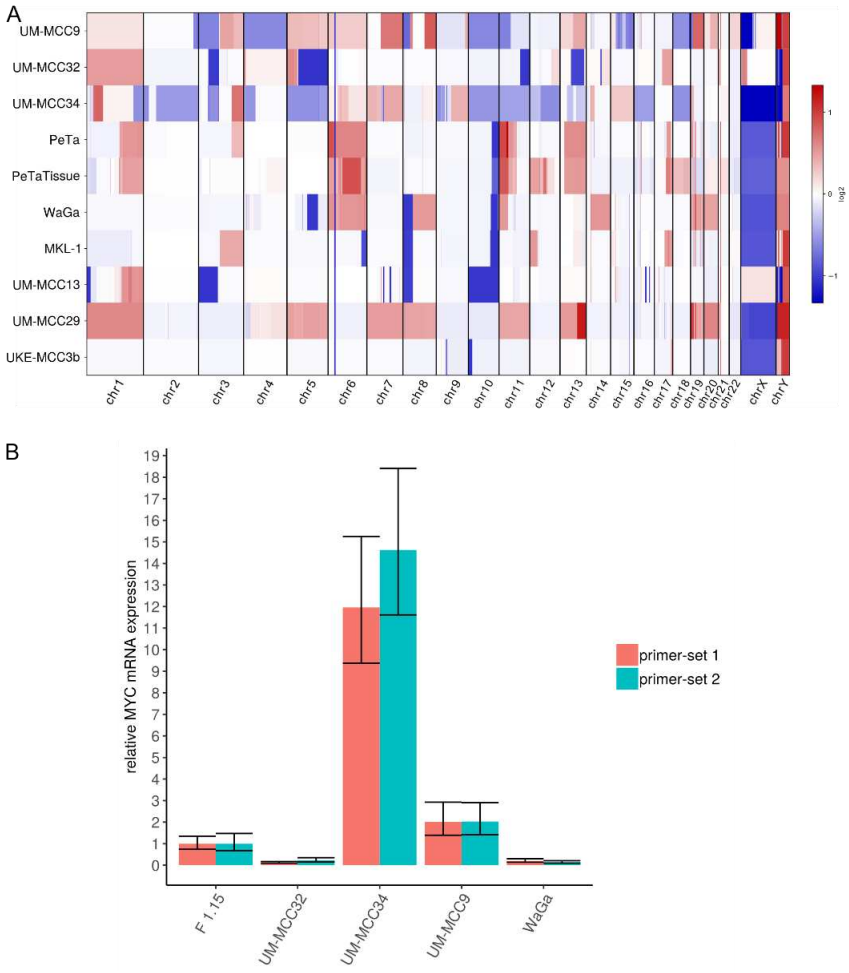


Figure 4. CNVs in MCC cell lines. (**A**) Graphical display of derived CNVs using CNVkit with sex chromosomes relative to haploid reference; (**B**) Expression of *MYC* mRNA in MCC cell lines was determined by qRT-PCR. Cq values were normalized to GAPDH expression and compared to ΔCq value of fibroblasts (F 1.15).

Previous studies reported copy number losses covering *RB1* on chromosome 13 [4,5,11,12,14,16,31]. We also observe large single-copy deletions on chromosome 13 including the loss of *RB1* in virus-negative cell lines (UM-MCC32 and UM-MCC34); in contrast, there are large single-copy gains that include *RB1* in one virus-negative (UM-MCC9) and two virus-positive cell lines (PeTa, UM-MCC29).

Local amplifications of *MYCL* on chromosome 1 have previously been reported for both MCC types [4,5,16,31]. Here, *MYCL* is included in the whole-chromosome gains of UM-MCC29 and UM-MCC32 as well as the partial chromosome gains in UM-MCC34. Interestingly, UM-MCC34 has an extraordinarily high, localized amplification of *MYC* (aka *c-MYC*), with 106 copies covering ~530,000 basepairs on chromosome 8. *MYC* is also included in larger whole- or partial-chromosome gains in UM-MCC29, WaGa and UM-MCC9. These amplifications are associated with a higher *MYC* mRNA expression, which is most pronounced in UM-MCC34 (Figure 4B).

3. Discussion

Due to the lack of suitable genetically engineered mouse models (GEMMs), preclinical functional studies rely on MCC cell lines. However, the detailed genomic characteristics of the applied cell lines are not fully established. Indeed, most studies investigating genomic features of MCC by targeted or WES are based on fresh frozen or formalin-fixed paraffin-embedded (FFPE) tissue samples [4–8,10–14]. Only Wong et al. included three virus-negative cell lines [6] that may be not representative for MCC [22,23]. Here, we present the mutational landscape of three classical virus-negative and six virus-positive MCC cell lines (characteristics are summarized in Table 1). The ratio of virus-positive to -negative cell lines recapitulates the ratio of MCC tumors in countries with high latitude [1]. The genomic features of the MCC cell line cohorts are very similar to those previously reported for the respective MCC tumors. Furthermore, direct comparison of one matched cell line-tissue pair confirmed that genomic alterations accumulated during cell culture only caused minor differences in their mutational landscape. However, expectedly, the cell line did not capture the complete tumor heterogeneity, as many somatic mutations were specific to the tissue.

Table 1. Genomic features of MCC cell lines for both MCC types.

MCC Cell Line Type	Virus-Negative	Virus-Positive
Tumor Mutational Burden	high (on average, 44.5 mut/Mbp)	low (on average, 10.5 mut/Mbp)
Mutagenic processes detected	UV-light-associated DNA damage (SBS7a, SBS7b)	flat mutation profile without prominent signatures.
Coding mutations	many mutations with potential functional effect; many mutated genes	few mutations with potential functional effect; few mutated genes
Copy Number Variations	many widespread CNVs	few, characteristic CNVs

The bold is used to emphasize the row names for the subsequent summary.

Virus-positive MCCs are characterized by very low TMB, a lack of prominent mutational signatures and the absence of functional mutations (Table 1) [4–6,11–13]. Previously reported TMBs for virus-positive MCC, however, show large differences and are inconsistently specified, e.g., regarding normalization. For the WES studies, TMB was reported either as a median of 12.5 SNVs [12], an average of 0.4 mut/Mbp [11] or a median of 1.57 mut/Mbp [13]. We observed, on average, 11 mut/Mbp, which is comparable with studies using targeted sequencing approaches (i.e., an average of 5–10 mut/Mbp [6], a median of 1.2 coding mut/Mbp [5] or up to 16 mut/Mbp [4]). All studies with higher TMB lacked individual normal tissues as a reference for somatic variant calling, hence databases

reporting common polymorphisms (e.g., 1000 genomes, exome aggregation consortium (ExAC), gnomAD databases) had to be used for filtering non-somatic variants. Thus, the observed higher TMBs are likely caused by polymorphisms not represented in common databases. This notion is supported by the absence of any prominent mutational signature in virus-positive MCC samples. No single mutational signature has a relevant contribution to the TMB; only "flat" TCF distributions were detected for virus-positive MCC cell lines, which likely represent randomly distributed, unfiltered polymorphisms that may impair the detection of other mutagenic processes. The absence of functional, cancer-related mutations and low signature reconstruction efficiency is in line with this assumption. In contrast, in virus-negative MCC cell lines, TMB is high (on average, 44.5 mut/Mbp), mutational patterns are strongly associated with UV-light-induced DNA damage, and many coding mutations of cancer-related genes exist (Table 1). The primary origin of virus-negative MCC cell lines is associated with UV-exposed areas. UM-MCC9 and UM-MCC32 were derived from primary tumors localized on the scalp, and UM-MCC34 was derived from axillary metastasis presumably originating from a primary tumor on the upper extremity (Table 2) [28]. Some of the virus-positive MCC cell lines were generated from tumors without a clear association with chronic UV-exposure, e.g., PeTa and UKE-MCC3b originated from tumors of the trunk (Table 2) [17]. Interestingly, we did not observe major differences in TMB between cell lines derived from primary tumors (UM-MCC9, UM-MCC32, PeTa) and metastases (UM-MCC34, WaGa, MKL-1, UM-MCC13, UM-MCC29, UKE-MCC3b), which would have been expected from more general observations in cancer (Table 2) [32].

Table 2. Overview of analyzed Merkel cell carcinoma cell lines.

Cell Line	MCPyV Status	Established From	Localization of Primary	Time in Culture	Reference
UM-MCC9	negative	primary, scalp	scalp	>6 years	[28]
UM-MCC32	negative	primary, scalp	scalp	>6 years	[28]
UM-MCC34	negative	axillary lymph node metastasis	presumably arm	>6 years	[28]
PeTa	positive	primary, trunk	trunk	>7 years	[17,26]
WaGa	positive	malignant ascites	head	>10 years	[17,33]
MKL-1	positive	nodal metastasis	unknown	>30 years	[17,34]
UM-MCC13	positive	metastasis, leg	presumably leg	>6 years	[28]
UM-MCC29	positive	inguinal lymph node metastasis	presumably leg	>6 years	[28]
UKE-MCC3b	positive	metastasis, trunk	trunk	>3 years	-

All virus-negative MCC cell lines show *RB1* and *TP53* disruption, either by frameshift deletion, nonsense, missense mutation or, for *RB1*, possibly copy number losses. Alterations in both genes are recurrent mutational features in virus-negative MCC [4–9,11–14,26]. Notably, the exact same nonsense mutations in *RB1* were previously reported for UM-MCC9 (rs794727481 [9,14]) and UM-MCC34 (rs121913304 [4,6]). *RB1* and *TP53* abrogation is also common in other neuroendocrine carcinomas, e.g., in small cell lung, neuroendocrine prostate and pancreatic carcinoma [35]. In this context, it is interesting to note that *MYC* binding motifs are enriched in neuroendocrine genes; thus, it has been proposed that *MYC* overexpression drives the temporal tumor cell evolution [36]. We detected an extraordinarily high *MYC* amplification associated with equally high mRNA expression in UM-MCC34. *MYC* family gene amplification, i.e., 6% for *MYCL* and 4% for *MYC* in virus-negative MCCs [5,16], as well as high *MYC* protein expression, was previously reported [13,37].

The biological importance of SMGs relies on the fact that these may be more prone for mutations due to open chromatin regions, i.e., reflecting the functional state of a cell during mutagenesis, or being positively selected during tumor evolution. The SMGs

with extraordinary low *p*-values were *KRT4*, *MDK* and *CACNA1B*, suggesting that these genes may be relevant for MCC carcinogenesis. However, critical examination of these mutations demonstrate that this is very unlikely. The mutations in *KRT4* are present in all cell lines and have been previously identified as a common polymorphism with 87% VAF in the 1000 genomes database. Thus, the *KRT4* mutation is actually the major allele of a single nucleotide polymorphism (SNP) not reflected in the hg19 reference genome. For *MDK*, the detected cytosine deletion is embedded in a sequence of 15 cytosines in close proximity to a stopcodon and is therefore in a region prone to sequencing artifacts. Actually, variations in cytosine counts of this region have already been reported in dbSNP as polymorphisms (rs74916763). Finally, the large-scale insertion in *CACNA1B* is localized at the last basepair of an exon and the inserted sequence is identical to the beginning of the following exon, hence we assume a deletion of an intronic region in between, which has already been reported with 0.3% VAF in the Allele Frequency Aggregator (ALFA) database. Consequently, these three variants likely reflect limitations in the representation and annotation of polymorphisms, which emphasizes the importance of variant filtering and evaluation. In virus-negative MCC cell lines, *TP53* and *RB1* have a relatively low *p*-value compared to other genes (Figure 3D) and, due to their recurrency in MCC, these genes are likely associated with tumorigenesis of virus-negative MCC.

CNVs were previously characterized in MCC using CGH [16–21], genome-wide microarrays [31] or next-generation sequencing [4–7,11,12,15]. We observe higher CNV numbers in virus-negative, and fewer, but characteristic, CNV patterns in virus-positive MCC, indicating a common alteration mechanism for the latter. Notably, the MCPyV-encoded small T antigen was reported to induce centrosome overproduction and to increase the frequency of micronuclei by interaction with E3-ligases, causing chromosome instability [38]. The virus-positive MCC-specific losses and gains may actually affect the tumor suppressor *RB1* and oncogene *MYC*.

APOBEC-mediated mutagenesis is a known feature of viral oncogenesis, e.g., in human-papilloma-virus-associated cancer [39]. In our and previously reported studies, *APOBEC* mutations seem to be absent in MCC [4,5]. However, *APOBEC*-related mutagenesis is restricted to localized, hypermutated regions, aka kataegis, that are difficult to detect by WES and even more so by targeted sequencing. Indeed, in whole-genome analysis, an *APOBEC*-related kataegis was reported in a virus-positive MCC [15]. Thus, to detect *APOBEC*-related mutagenesis with enhanced sensitivity, signature analysis should be restricted to such hypermutated regions [40].

In summary, WES of virus-positive and -negative MCC cell lines with a neuroendocrine growth pattern revealed mutational features resembling those previously observed in MCC tissue samples; hence, our report strengthens the utility of these classical MCC cell lines for functional studies.

4. Materials and Methods

4.1. Cell Lines and Tissues

The MCC cell lines WaGa [33], PeTa [26], MKL-1 [34], UM-MCC13, UM-MCC29, UM-MCC9, UM-MCC32, UM-MCC34 [28] were described before (Table 2). UM-MCC13, UM-MCC29, UM-MCC9, UM-MCC32 and UM-MCC34 were provided by Monique E. Verhaegen, University of Michigan, Ann Arbor, MI, USA. UKE-MCC3b was established at the department of dermatology at the University Medicine Essen, Essen, Germany and the patient gave informed consent (ethics committee approval: 11–4715; 17-7538-BO). WaGa, PeTa, MKL-1, UKE-MCC3b and UM-MCC34 were maintained at 37 °C with 5% CO_2 in RPMI-1640 with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin (PAN Biotech, Aidenbach, Germany), while UM-MCC13, UM-MCC29 and UM-MCC32 were maintained as described previously [28] in self-renewal media [41] including low-glucose DMEM, Neurobasal-A medium, 2-mercaptoethanol, N-2 Supplement (100× (times)), B-27™ Supplement (50×, minus vitamin A), MEM non-essential amino acids solution (100×), Gibco™ Amphotericin (all Thermo Fischer, Dreieich, Germany), retinoic acid (Sigma Aldrich, Darm-

stadt, Germany), basic fibroblast growth factor, recombinant human IGF-I (Peprotech, Hamburg, Germany), 1% penicillin/streptomycin (PAN Biotech). The self-renewal medium was further supplemented with chicken embryo extract containing HBSS, PBS (PAN Biotech), MEM with Earle's salts and L-glutamine (Thermo Fischer) and Hyaluronidase specs (Sigma Aldrich) [42]. Primary cutaneous fibroblasts (F 1.15) were generated and maintained as previously described [43].

4.2. Library Preparation and Sequencing

DNA was purified using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). Library preparation and sequencing were performed by DKFZ Genomics and Proteomics Core Facility. WES libraries were prepared using SureSelect All Exon V6 Kit (Agilent Technologies, Santa Clara, CA, USA) and subsequently sequenced on HiSeq 4000 (Illumina) paired-end 100bp reads with, on average, 118 million reads per sample.

4.3. Alignment and Variant Calling

Processing of reads in FASTQ format to genomic variations in variant call format (VCF) was performed according to genome analysis toolkit (GATK) best practices of germline short variant discovery for all MCC cell lines. Additionally, for PeTa and PetaTissue, GATK best practices of somatic short variant discovery were used. Paired-end reads in FASTQ Format were aligned to the human reference genome hg19 (GRCh37) using Burrows–Wheeler aligner (BWA) mem v0.7.17 [44]; duplicates were marked using Picard MarkDuplicates and aligned reads sorted using samtools v1.7. GATK Toolkits of version 4.0.12.0 were used. For germline short variant discovery, GATK BaseRecalibrator and ApplyBQSR were applied and, subsequently, variants were called using GATK HaploTypeCaller without normal tissue reference data. For somatic short-variant discovery, the panel of normal (PoN) for PeTa and PeTaTissue were created and variants were called with GATK Mutect2, once with Peta cell line and once with PeTaTissue as normal reference. Variants were annotated using ANNOVAR (Version from 8 June 2020) and databases of Ensembl Gencode v31 (29 September 2019), dbSNP with allelic splitting and left-normalization v150 (29 September 2017) ClinVar (05 March 2015), ExAC (29 November 2015), gnomAD exome and genome collection (v2.1.1, 18 March 2019), 1000 genomes dataset (24 August 2015) and Kaviar database (03 December 2015) were used.

4.4. Variant Filtering

The Maftools R package v2.0.05 was used for VCF to Mutation Annotation Format (MAF) conversion using ensemble genes as gene column, and used for manipulation of MAF files in R [45]. Variants that are not within the probe region of the SureSelect All Exon V6 Kit were removed from analysis. Variants from germline variant calling were filtered and removed from analysis if one of the following criteria was met: SNVs with QD < 2.0, MQ < 50.0, FS > 60.0, SOR > 5.0, MQRankSum < −12.5 or ReadPosRankSum < −8.0 and InDels with QD < 2.0, FS > 200.0, SOR > 10.0, InbreedingCoeff < −0.8 or ReadPosRankSum < −20.0. For evaluation of subsequent filtering of possible polymorphisms, we compared different databases reporting VAFs (Figure S1B,C). Based on this analysis, we filtered a variant as germline polymorphism if it reported a VAF of more than 0.001% in gnomAD v2.1.1 genome. Variants from somatic variant calling of Peta/PeTaTissue were filtered using GATK FilterMutectCalls and not filtered for germline polymorphisms.

4.5. Quantitative Real-Time PCR (qRT-PCR)

RNA of cell lines was extracted using peqlabGold Micro RNA Kit (PEQLAB Biotechnologie GmbH, Erlangen, Germany) and transcribed to cDNA using SuperScript IV Reverse Transcriptase (1000u, Life Technologies GmbH, Darmstadt, Germany). qRT-PCR was performed on the CFX Real-Time PCR system (Bio-Rad Laboratories, Hercules, CA, USA) using LuminoCT SYBR Green qPCR ready Mix (Sigma Aldrich). For *MYC* following primers were used: primer-set 1 forward: GGCTCCTGGCAAAAGGTCA, reverse: CTGCG-

TAGTTGTGCTGATGT; primer-set 2 forward: GTCAAGAGGCGAACACACAAC, reverse: TTGGACGGACAGGATGTATGC. *GAPDH* primer set: forward ACCACAGTCCATGCCATCAC, reverse TCCACCACCCTGTTGCTGTA. Annealing was performed at 60 °C for 15 s. Relative quantification was performed using the $2^{-\Delta\Delta Cq}$ method implemented in the R package "pcr" [46].

4.6. Bioinformatic Processing

The R Markdown script for analysis of MAF format files is attached in File S1. Normalization to mutations per Megabasepair was done through dividing the number of mutations by the sum of the length of all regions covered by probes (60,456,963 basepairs). Signature analysis was performed using only SNVs and MutationalPatterns R package v1.8.0 [47]. Reference mutational signatures of version 3.1 [48] were downloaded from the Cataloque of Somatic Mutations in Cancer (COSMIC, https://cancer.sanger.ac.uk/cosmic/signatures, (accessed date 1 October 2020)). SMGs were determined using MutSigCV v1.41 [30] and Hallmark Gene Sets v7.1 downloaded from MSigDB was used [25]. Heatmaps were created using ComplexHeatmap R package v2.1.0 [49], other plots with ggplot v3.1.0 [50] and ggVennDiagram using R programming language v3.5.2. CNVs were derived using CNVkit v0.9.6 with default settings.

5. Conclusions

Virus-negative MCC cell lines show high TMB, UV-light DNA damage and several functional coding mutations, while virus-positive MCC cell lines harbor few mutations. Thus, the mutational landscape of MCC cell lines that are frequently used in preclinical research reflect the observations from tumor tissue and confirm their suitability for functional studies.

Supplementary Materials: The following are available online at https://www.mdpi.com/2072-6694/13/4/649/s1, Figure S1: Filtering of polymorphisms in MCC cell lines, Figure S2: TCF for each MCC cell line, Table S1: Mutations found in MCC cell lines, Table S2: Mutations from somatic variant calling of PeTa and tissue of PeTa, Table S3: All nonsense, frameshift and nonstop mutations found in MCC cell lines with respective Hallmark Gene Set, Table S4: Mutations found in SMGs with $p < 0.01$ and within a Hallmark Gene Set, Table S5: SMGs from MutSigCV of virus-negative MCC cell lines, Table S6: SMGs from MutSigCV of virus-positive MCC cell lines, Table S7: CNVs found in MCC cell lines, File S1: R Markdown script used for analysis of MAF files.

Author Contributions: Data curation: K.H.; formal analysis: K.H.; funding acquisition: J.C.B.; investigation: P.G., A.H.-C., C.W.; methodology: K.H.; project administration: J.C.B.; resources: P.G., A.H.-C., C.W., J.U.; supervision: J.C.B.; visualization: K.H.; conceptualization: J.C.B.; writing—original draft preparation: K.H.; writing—review and editing: J.C.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the DKTK site budget OE 0460 ED03.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the University Duisburg-Essen (11-4715, 14 September 2011).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study is available in Tables S1 and S2. FASTQ Files are available on request from the corresponding author. The FASTQ Files are not publicly available due to privacy reasons.

Acknowledgments: For technical support and sequencing services we thank the German Cancer Research Center (DKFZ) Genomics & Proteomics Core Facility, Heidelberg, Germany. We thank Monique E. Verhaegen, University of Michigan, USA, for providing MCC cell lines. Further, we thank Jan Gravemeyer, Translational Skin Cancer Research, German Cancer Consortium (DKTK), Essen, Germany, for his excellent advice and help in bioinformatic analysis.

Conflicts of Interest: J.C.B. is receiving speaker's bureau honoraria from Amgen, Pfizer, Merck-Serono, Recordati and Sanofi, is a paid consultant/advisory board member/DSMB member for Boehringer Ingelheim, eTheRNA, 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. J.U. is on the advisory board or has received honoraria and travel support from Amgen, Bristol Myers Squibb, GSK, LeoPharma, Merck Sharp and Dohme, Novartis, Pierre Fabre, Roche, Sanofi outside the submitted work. All of the other authors declare no conflict of interest.

References

1. Becker, J.C.; Stang, A.; DeCaprio, J.A.; Cerroni, L.; Lebbé, C.; Veness, M.; Nghiem, P. Merkel cell carcinoma. *Nat. Rev. Dis. Primers* **2017**, *3*, 17077. [CrossRef]
2. Feng, H.; Shuda, M.; Chang, Y.; Moore, P.S. Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. *Science* **2008**, *319*, 1096–1100. [CrossRef] [PubMed]
3. Garneski, K.M.; Warcola, A.H.; Feng, Q.; Kiviat, N.B.; Helen Leonard, J.; Nghiem, P. Merkel Cell Polyomavirus Is More Frequently Present in North American than Australian Merkel Cell Carcinoma Tumors. *J. Investig. Dermatol.* **2009**, *129*, 246–248. [CrossRef] [PubMed]
4. Starrett, G.J.; Thakuria, M.; Chen, T.; Marcelus, C.; Cheng, J.; Nomburg, J.; Thorner, A.R.; Slevin, M.K.; Powers, W.; Burns, R.T.; et al. Clinical and molecular characterization of virus-positive and virus-negative Merkel cell carcinoma. *Genome Med.* **2020**, *12*, 30. [CrossRef] [PubMed]
5. Knepper, T.C.; Montesion, M.; Russell, J.S.; Sokol, E.S.; Frampton, G.M.; Miller, V.A.; Albacker, L.A.; McLeod, H.L.; Eroglu, Z.; Khushalani, N.I.; et al. The Genomic Landscape of Merkel Cell Carcinoma and Clinicogenomic Biomarkers of Response to Immune Checkpoint Inhibitor Therapy. *Clin. Cancer Res.* **2019**, *25*, 5961. [CrossRef]
6. Wong, S.Q.; Waldeck, K.; Vergara, I.A.; Schröder, J.; Madore, J.; Wilmott, J.S.; Colebatch, A.J.; De Paoli-Iseppi, R.; Li, J.; Lupat, R.; et al. UV-Associated Mutations Underlie the Etiology of MCV-Negative Merkel Cell Carcinomas. *Cancer Res.* **2015**, *75*, 5228. [CrossRef] [PubMed]
7. Harms, P.W.; Collie, A.M.B.; Hovelson, D.H.; Cani, A.K.; Verhaegen, M.E.; Patel, R.M.; Fullen, D.R.; Omata, K.; Dlugosz, A.A.; Tomlins, S.A.; et al. Next generation sequencing of Cytokeratin 20-negative Merkel cell carcinoma reveals ultraviolet-signature mutations and recurrent TP53 and RB1 inactivation. *Mod. Pathol.* **2016**, *29*, 240–248. [CrossRef]
8. Veija, T.; Sarhadi, V.K.; Koljonen, V.; Bohling, T.; Knuutila, S. Hotspot mutations in polyomavirus positive and negative Merkel cell carcinomas. *Cancer Genet.* **2016**, *209*, 30–35. [CrossRef] [PubMed]
9. Cohen, P.R.; Tomson, B.N.; Elkin, S.K.; Marchlik, E.; Carter, J.L.; Kurzrock, R. Genomic portfolio of Merkel cell carcinoma as determined by comprehensive genomic profiling: Implications for targeted therapeutics. *Oncotarget* **2016**, *7*, 23454. [CrossRef] [PubMed]
10. Graves, C.A.; Jones, A.; Reynolds, J.; Stuart, J.; Pirisi, L.; Botrous, P.; Wells, J. Neuroendocrine Merkel Cell Carcinoma Is Associated with Mutations in Key DNA Repair, Epigenetic and Apoptosis Pathways: A Case-Based Study Using Targeted Massively Parallel Sequencing. *Neuroendocrinology* **2015**, *101*, 112–119. [CrossRef]
11. Harms, P.W.; Vats, P.; Verhaegen, M.E.; Robinson, D.R.; Wu, Y.-M.; Dhanasekaran, S.M.; Palanisamy, N.; Siddiqui, J.; Cao, X.; Su, F.; et al. The Distinctive Mutational Spectra of Polyomavirus-Negative Merkel Cell Carcinoma. *Cancer Res.* **2015**, *75*, 3720–3727. [CrossRef]
12. Goh, G.; Walradt, T.; Markarov, V.; Blom, A.; Riaz, N.; Doumani, R.; Stafstrom, K.; Moshiri, A.; Yelistratova, L.; Levinsohn, J.; et al. Mutational landscape of MCPyV-positive and MCPyV-negative Merkel cell carcinomas with implications for immunotherapy. *Oncotarget* **2015**, *7*, 3403. [CrossRef] [PubMed]
13. González-Vela, M.d.C.; Curiel-Olmo, S.; Derdak, S.; Beltran, S.; Santibañez, M.; Martínez, N.; Castillo-Trujillo, A.; Gut, M.; Sánchez-Pacheco, R.; Almaraz, C.; et al. Shared Oncogenic Pathways Implicated in Both Virus-Positive and UV-Induced Merkel Cell Carcinomas. *J. Investig. Dermatol.* **2017**, *137*, 197–206. [CrossRef]
14. Cimino, P.J.; Robirds, D.H.; Tripp, S.R.; Pfeifer, J.D.; Abel, H.J.; Duncavage, E.J. Retinoblastoma gene mutations detected by whole exome sequencing of Merkel cell carcinoma. *Mod. Pathol.* **2014**, *27*, 1073–1087. [CrossRef]
15. Starrett, G.J.; Marcelus, C.; Cantalupo, P.G.; Katz, J.P.; Cheng, J.; Akagi, K.; Thakuria, M.; Rabinowits, G.; Wang, L.C.; Symer, D.E.; et al. Merkel Cell Polyomavirus Exhibits Dominant Control of the Tumor Genome and Transcriptome in Virus-Associated Merkel Cell Carcinoma. *mBio* **2017**, *8*, e02079-02016. [CrossRef] [PubMed]
16. Paulson, K.G.; Lemos, B.D.; Feng, B.; Jaimes, N.; Peñas, P.F.; Bi, X.; Maher, E.; Cohen, L.; Helen Leonard, J.; Granter, S.R.; et al. Array-CGH Reveals Recurrent Genomic Changes in Merkel Cell Carcinoma Including Amplification of L-Myc. *J. Investig. Dermatol.* **2009**, *129*, 1547–1555. [CrossRef] [PubMed]
17. Schrama, D.; Sarosi, E.-M.; Adam, C.; Ritter, C.; Kaemmerer, U.; Klopocki, E.; König, E.-M.; Utikal, J.; Becker, J.C.; Houben, R. Characterization of six Merkel cell polyomavirus-positive Merkel cell carcinoma cell lines: Integration pattern suggest that large T antigen truncating events occur before or during integration. *Int. J. Cancer* **2019**, *145*, 1020–1032. [CrossRef]

18. Van Gele, M.; Leonard, J.H.; Van Roy, N.; Van Limbergen, H.; Van Belle, S.; Cocquyt, V.; Salwen, H.; De Paepe, A.; Speleman, F. Combined karyotyping, CGH and M-FISH analysis allows detailed characterization of unidentified chromosomal rearrangements in Merkel cell carcinoma. *Int. J. Cancer* **2002**, *101*, 137–145. [CrossRef]
19. Van Gele, M.; Speleman, F.; Vandesompele, J.; Van Roy, N.; Leonard, J.H. Characteristic pattern of chromosomal gains and losses in Merkel cell carcinoma detected by comparative genomic hybridization. *Cancer Res.* **1998**, *58*, 1503–1508.
20. Larramendy, M.L.; Koljonen, V.; Böhling, T.; Tukiainen, E.; Knuutila, S. Recurrent DNA copy number changes revealed by comparative genomic hybridization in primary Merkel cell carcinomas. *Mod. Pathol.* **2004**, *17*, 561–567. [CrossRef]
21. Härle, M.; Arens, N.; Moll, I.; Back, W.; Schulz, T.; Scherthan, H. Comparative genomic hybridization (CGH) discloses chromosomal and subchromosomal copy number changes in Merkel cell carcinomas. *J. Cutan. Pathol.* **1996**, *23*, 391–397. [CrossRef] [PubMed]
22. Gele, M.V.; Boyle, G.M.; Cook, A.L.; Vandesompele, J.; Boonefaes, T.; Rottiers, P.; Roy, N.V.; De Paepe, A.; Parsons, P.G.; Leonard, J.H.; et al. Gene-expression profiling reveals distinct expression patterns for Classic versus Variant Merkel cell phenotypes and new classifier genes to distinguish Merkel cell from small-cell lung carcinoma. *Oncogene* **2004**, *23*, 2732–2742. [CrossRef] [PubMed]
23. Daily, K.; Coxon, A.; Williams, J.S.; Lee, C.-C.R.; Coit, D.G.; Busam, K.J.; Brownell, I. Assessment of Cancer Cell Line Representativeness Using Microarrays for Merkel Cell Carcinoma. *J. Investig. Dermatol.* **2015**, *135*, 1138–1146. [CrossRef] [PubMed]
24. Alexandrov, L.B.; Nik-Zainal, S.; Wedge, D.C.; Aparicio, S.A.J.R.; Behjati, S.; Biankin, A.V.; Bignell, G.R.; Bolli, N.; Borg, A.; Børresen-Dale, A.-L.; et al. Signatures of mutational processes in human cancer. *Nature* **2013**, *500*, 415–421. [CrossRef]
25. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [CrossRef] [PubMed]
26. Houben, R.; Dreher, C.; Angermeyer, S.; Borst, A.; Utikal, J.; Haferkamp, S.; Peitsch, W.K.; Schrama, D.; Hesbacher, S. Mechanisms of p53 restriction in Merkel cell carcinoma cells are independent of the Merkel cell polyoma virus T antigens. *J. Investig. Dermatol.* **2013**, *133*, 2453–2460. [CrossRef]
27. Hesbacher, S.; Pfitzer, L.; Wiedorfer, K.; Angermeyer, S.; Borst, A.; Haferkamp, S.; Scholz, C.-J.; Wobser, M.; Schrama, D.; Houben, R. RB1 is the crucial target of the Merkel cell polyomavirus Large T antigen in Merkel cell carcinoma cells. *Oncotarget* **2016**, *7*, 32956. [CrossRef]
28. Verhaegen, M.E.; Mangelberger, D.; Weick, J.W.; Vozheiko, T.D.; Harms, P.W.; Nash, K.T.; Quintana, E.; Baciu, P.; Johnson, T.M.; Bichakjian, C.K.; et al. Merkel cell carcinoma dependence on bcl-2 family members for survival. *J. Investig. Dermatol.* **2014**, *134*, 2241–2250. [CrossRef] [PubMed]
29. Dhar, S.K.; Yoshida, K.; Machida, Y.; Khaira, P.; Chaudhuri, B.; Wohlschlegel, J.A.; Leffak, M.; Yates, J.; Dutta, A. Replication from oriP of Epstein-Barr Virus Requires Human ORC and Is Inhibited by Geminin. *Cell* **2001**, *106*, 287–296. [CrossRef]
30. Lawrence, M.S.; Stojanov, P.; Polak, P.; Kryukov, G.V.; Cibulskis, K.; Sivachenko, A.; Carter, S.L.; Stewart, C.; Mermel, C.H.; Roberts, S.A.; et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **2013**, *499*, 214–218. [CrossRef]
31. Carter, M.D.; Gaston, D.; Huang, W.-Y.; Greer, W.L.; Pasternak, S.; Ly, T.Y.; Walsh, N.M. Genetic profiles of different subsets of Merkel cell carcinoma show links between combined and pure MCPyV-negative tumors. *Hum. Pathol.* **2018**, *71*, 117–125. [CrossRef]
32. Patel, S.A.; Rodrigues, P.; Wesolowski, L.; Vanharanta, S. Genomic control of metastasis. *Br. J. Cancer* **2021**, *124*, 3–12. [CrossRef]
33. Houben, R.; Shuda, M.; Weinkam, R.; Schrama, D.; Feng, H.; Chang, Y.; Moore, P.S.; Becker, J.C. Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens. *J. Virol.* **2010**, *84*, 7064–7072. [CrossRef] [PubMed]
34. Rosen, S.T.; Gould, V.E.; Salwen, H.R.; Herst, C.V.; Le Beau, M.M.; Lee, I.; Bauer, K.; Marder, R.J.; Andersen, R.; Kies, M.S.; et al. Establishment and characterization of a neuroendocrine skin carcinoma cell line. *Lab. Investig.* **1987**, *56*, 302–312.
35. Rickman, D.S.; Beltran, H.; Demichelis, F.; Rubin, M.A. Biology and evolution of poorly differentiated neuroendocrine tumors. *Nat. Med.* **2017**, *23*, 664–673. [CrossRef] [PubMed]
36. Ireland, A.S.; Micinski, A.M.; Kastner, D.W.; Guo, B.; Wait, S.J.; Spainhower, K.B.; Conley, C.C.; Chen, O.S.; Guthrie, M.R.; Soltero, D.; et al. MYC Drives Temporal Evolution of Small Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate. *Cancer Cell* **2020**, *38*, 60–78.e12. [CrossRef] [PubMed]
37. Shao, Q.; Kannan, A.; Lin, Z.; Stack, B.C.; Suen, J.Y.; Gao, L. BET Protein Inhibitor JQ1 Attenuates Myc-Amplified MCC Tumor Growth In Vivo. *Cancer Res.* **2014**, *74*, 7090. [CrossRef]
38. Kwun, H.J.; Wendzicki, J.A.; Shuda, Y.; Moore, P.S.; Chang, Y. Merkel cell polyomavirus small T antigen induces genome instability by E3 ubiquitin ligase targeting. *Oncogene* **2017**, *36*, 6784–6792. [CrossRef] [PubMed]
39. Burns, M.B.; Temiz, N.A.; Harris, R.S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* **2013**, *45*, 977–983. [CrossRef]
40. Maura, F.; Degasperi, A.; Nadeu, F.; Leongamornlert, D.; Davies, H.; Moore, L.; Royo, R.; Ziccheddu, B.; Puente, X.S.; Avet-Loiseau, H.; et al. A practical guide for mutational signature analysis in hematological malignancies. *Nat. Commun.* **2019**, *10*, 2969. [CrossRef]
41. Molofsky, A.V.; He, S.; Bydon, M.; Morrison, S.J.; Pardal, R. Bmi-1 promotes neural stem cell self-renewal and neural development but not mouse growth and survival by repressing the p16Ink4a and p19Arf senescence pathways. *Genes Dev.* **2005**, *19*, 1432–1437. [CrossRef]

42. Stemple, D.L.; Anderson, D.J. Isolation of a stem cell for neurons and glia from the mammalian neural crest. *Cell* **1992**, *71*, 973–985. [CrossRef]
43. Fan, K.; Spassova, I.; Gravemeyer, J.; Ritter, C.; Horny, K.; Lange, A.; Gambichler, T.; Ødum, N.; Schrama, D.; Schadendorf, D.; et al. Merkel cell carcinoma-derived exosome-shuttle miR-375 induces fibroblast polarization by inhibition of RBPJ and p53. *Oncogene* **2020**. [CrossRef]
44. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]
45. Mayakonda, A.; Lin, D.C.; Assenov, Y.; Plass, C.; Koeffler, H.P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **2018**, *28*, 1747–1756. [CrossRef]
46. Ahmed, M.; Kim, D.R. pcr: An R package for quality assessment, analysis and testing of qPCR data. *PeerJ* **2018**, *6*, e4473. [CrossRef] [PubMed]
47. Blokzijl, F.; Janssen, R.; van Boxtel, R.; Cuppen, E. MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. *Genome Med.* **2018**, *10*, 33. [CrossRef] [PubMed]
48. Alexandrov, L.B.; Kim, J.; Haradhvala, N.J.; Huang, M.N.; Tian Ng, A.W.; Wu, Y.; Boot, A.; Covington, K.R.; Gordenin, D.A.; Bergstrom, E.N.; et al. The repertoire of mutational signatures in human cancer. *Nature* **2020**, *578*, 94–101. [CrossRef]
49. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, *32*, 2847–2849. [CrossRef] [PubMed]
50. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.

Cancers **2020**, *12*,

Supplementary Material: Mutational Landscape of Virus- and UV-Associated Merkel Cell Carcinoma Cell Lines Is Comparable to Tumor Tissue

Kai Horny, Patricia Gerhardt, Angela Hebel-Cherouny, Corinna Wülbeck, Jochen Utikal and Jürgen C. Becker

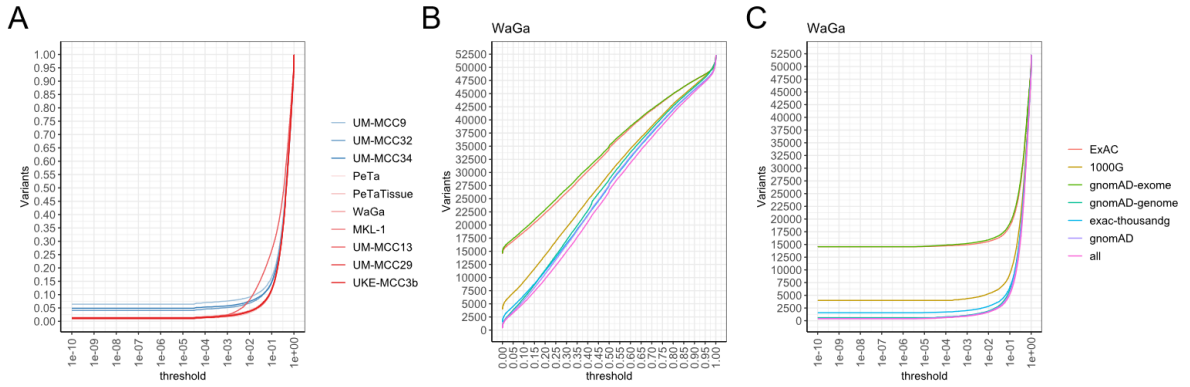


Figure S1. Filtering of polymorphisms in MCC cell lines.

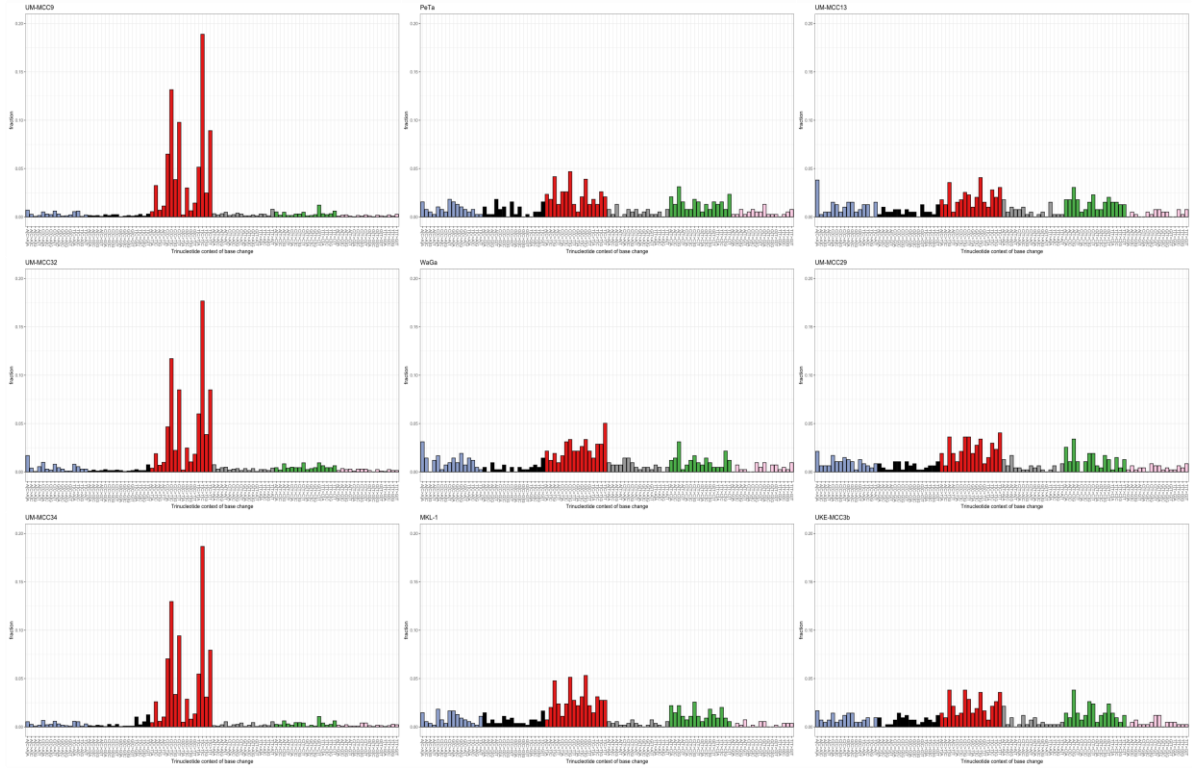


Figure S2. TCFs for each MCC cell line.

Table S1 (available in separate CSV file). Mutations found in MCC cell lines.

Table S2 (available in separate CSV file). Mutations from somatic variant calling of PeTa and tissue of PeTa.

Table S3 (available in separate CSV file). All nonsense, frameshift and nonstop mutations found in MCC cell lines with respective Hallmark Gene Set.

Table S4 (available in separate CSV file). Mutations found in SMGs with $p < 0.01$ and within a Hallmark Gene Set.

Table S5 (available in separate CSV file). SMGs from MutSigCV of virus-negative MCC cell lines.

Table S6 (available in separate CSV file). SMGs from MutSigCV of virus-positive MCC cell lines.

Table S7 (available in separate CSV file). CNVs found in MCC cell lines, File S1: R Markdown script used for analysis of MAF files.

File S1 (available in separate R Markdown file). R Markdown script used for analysis of MAF files.

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

Cumulative Thesis/Extent of Contribution

Cumulative thesis of Mr. Kai Horny

Author contributions

Title of the publication:

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

Authors:
Kai Horny, Christoph Sproll, Lukas Peiffer, Frauke Furtmann, Patricia Gerhardt, Jan Gravemeyer, Nikolas H. Stoecklein, Ivelina Spassova, Jürgen C. Becker

Contributions:

- Conception – 50 %: Planning and conceptualization of single cell RNA data analysis.
- Conduction of experimental work – 30 %: Contributing to planning of single cell RNA library preparation and sequencing, planning of bulk transcriptome sequencing and evaluation of histological images.
- Data analysis – 80 %: Processing, analysis, integration, and visualization of single cell RNA sequencing, bulk transcriptome, and imaging data.
- Statistical analysis – 80 %: Differential gene, gene set enrichment and variation analysis.
- Writing the manuscript – 80 %: Writing of Abstract, Introduction, Results, Discussion, Material and Methods, Conclusions and Figures.
- Revision of the manuscript – 70 %: Revising Abstract, Introduction, Results, Discussion, Material and Methods, Conclusions and Figures.

Signature of the Doctoral Candidate Signature of the Doctoral Supervisor

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

Abstract

Background

Oral squamous cell carcinoma (OSCC), an HPV-negative head and neck cancer, frequently metastasizes to the regional lymph nodes but only occasionally beyond. Initial phases of metastasis are associated with an epithelial-mesenchymal transition (EMT), while the consolidation phase is associated with mesenchymal-epithelial transition (MET). This dynamic is referred to as epithelial-mesenchymal plasticity (EMP). While it is known that EMP is essential for cancer cell invasion and metastatic spread, less is known about the heterogeneity of EMP states and even less about the heterogeneity between primary and metastatic lesions.

Methods

To assess both the heterogeneity of EMP states in OSCC cells and their effects on stromal cells, we performed single-cell RNA sequencing (scRNAseq) of 5 primary tumors, 9 matching metastatic and 5 tumor-free lymph nodes and re-analyzed publicly available scRNAseq data of 9 additional primary tumors. For examining the cell type composition, we performed bulk transcriptome sequencing. Protein expression of selected genes were confirmed by immunohistochemistry.

Results

From the 23 OSCC lesions, the single cell transcriptomes of a total of 7,263 carcinoma cells were available for in-depth analyses. We initially focused on one lesion to avoid confounding inter-patient heterogeneity and identified OSCC cells expressing genes characteristic of different epithelial and partial EMT stages. RNA velocity and the increase in inferred copy number variations indicated a progressive trajectory towards epithelial differentiation in this metastatic lesion, i.e., cells likely underwent MET. Extension to all samples revealed a less stringent but essentially similar pattern. Interestingly, MET cells show increased activity of the EMT-activator ZEB1. Immunohistochemistry confirmed that ZEB1 was co-expressed with the epithelial marker cornifin B in individual tumor cells. The lack of E-cadherin mRNA expression

suggests this is a partial MET. Within the tumor microenvironment we found immunomodulating fibroblasts that were maintained in primary and metastatic OSCC.

Conclusions

This study reveals that EMP enables different partial 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. During MET, ZEB1 appears to be functionally active, indicating a more complex role of ZEB1 than mere induction of EMT.

Keywords: single cell RNA; oral cavity; squamous cell carcinoma; Epithelial-mesenchymal plasticity; EMT; MET; ZEB1; heterogeneity; partial EMT

Background

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 that has a better prognosis than HPV-negative tumors (1). 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 and 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.

In many epithelial tumors, invasion and metastasis becomes possible 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). EMT has also been shown to impact the characteristics of mesenchymal cells in the tumor stroma either by cell polarization or as a direct contributor to the cancer-associated fibroblast (CAF) population (5, 6). Conversely, CAFs also modify the EMT status of tumor cells.

Importantly, 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, 7, 8). Thus, expression of EMT-related genes and their regulating transcription factors is highly heterogeneous, even within one cancer entity, between patients, in different lesions from one patient, and between individual cancer cells within one lesion (4, 9). 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 (7, 10-13). Therefore, it has recently been recommended that this EMT continuum should rather referred to as epithelial-mesenchymal plasticity (EMP) (4, 12).

Single-cell analyses are a powerful tool to capture the EMP-associated heterogeneity of cancer cells and their impact on stromal cells. However, to date most EMP-related single-cell studies are based on controlled *in vitro* and *in vivo* experiments (7, 8, 11, 14, 15). Particularly in HNSCC only few studies scrutinized EMP within freshly isolated tumor samples (10, 16, 17). Of particular note is the seminal work of Puram *et al*. in which 2,215 malignant cells from 18 patients were characterized, revealing multiple pEMT states with high variability in EMP-related gene expression (10).

In the work presented here, we investigated the cellular heterogeneity of 5 primary, 9 regionally metastatic OSCC lesions and 5 tumor-free lymph nodes isolated from 7 patients using multiplexed single-cell RNA sequencing (scRNAseq). In addition, we re-analyzed a recently published series of scRNAseq data from primary HNSCC that included 9 OSCC tumors to put our observations on an even broader data base (17). Our results not only confirm the EMP-associated heterogeneity of cancer cells in primary and metastatic OSCC, but also demonstrate that immunomodulating CAFs are preserved in primary and metastatic OSCC. Furthermore, we demonstrated a mesenchymal-epithelial transition (MET) of OSCC cells in established lymph node metastases. Surprisingly, we observed a high activity of the EMT-activator ZEB1 in metastatic OSCC cells with epithelial differentiation, which was confirmed by co-expression of ZEB1 and cornifin-B protein in individual tumor cells.

Methods

Tissue samples

From 7 OSCC patients treated at the Department of Oral and Maxillofacial Plastic Surgery of the University Hospital of Heinrich Heine University Düsseldorf, we examined a total of 19 tissue samples - 5 primary tumors and 14 potentially affected lymph nodes - by histopathological examination and bulk and single cell RNA sequencing. Of the 14 lymph nodes, 9 represented lymph node metastases as indicated by detection of carcinoma cells in histopathological examination and scRNAseq. The clinical details are provided in table 1. Due to the large size of the excised lesion, we were able to analyze two different areas of the primary tumors of patients #6 and #7 as separate samples to better capture any heterogeneity that may exist; these samples are designated #6.1 and #6.4 as well as #7.1 and #7.4, respectively.

Histology and Immunohistochemistry (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) and anti-ZEB1 (Cat. No. HPA027524-25UL, Sigma Aldrich) was performed as previously described (18). 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. Following an additional washing step, slides were incubated for 30 min with 3% bovine serum albumin (BSA) in TBST. For single stainings, sections were incubated at room temperature for 1h with anti-SPRR1B at a dilution of 1:600, or overnight with anti-ZEB1 at a dilution of 1:500 dilution. 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. Samples were washed with TBST in between incubations and, with tap water for 3 min before fixation. For multiplexed antigen detection, the OpalTM chemistry system (Akoya Biosciences, Marlborough, MA, USA, Cat. No.: OP7TL4001KT) was used according to the manufacturer's description. Briefly, after deparaffinization and fixation, we processed the sections for 15 min with retrieval buffers in an inverter microwave oven. Then, we incubated them with antibody diluent for 10 min at room temperature, followed by incubation with the anti-SPRR1B antibody for 30 min. Next, Opal Polymer horseradish peroxidase (HRP) secondary antibody solution with the respective chromogen was applied for 10 min, antibodies were removed by microwave treatment and the staining with anti-ZEB1 antibody was performed. Finally, slides were incubated with 4′,6-diamidino-2-phenylindole (DAPI) for 5 min.

Single-cell RNA sequencing

Samples were processed immediately after surgery and temporarily stored for transport at 4°C in tissue storage solution before processing (Miltenyi Biotec, Bergisch Gladbach, Germany). Briefly, samples were dissociated into single-cell suspensions using the gentleMACS Dissociator (Cat. No. 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 (Cat. No. P04-16500, PAN-Biotech) and an enzyme mix consisting of 200µl Enzyme H, 100 µl Enzyme R and 25 µl Enzyme A (Cat. No. 130-095-929 Miltenyi Biotec). Afterwards, single-cell suspensions were reconstituted and washed thrice with 0.05 % BSA phosphate-buffered saline (PBS) and filtered through a 100 µl cell strainer.

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

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

Bulk transcriptome analysis

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

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

signature matrix derived from the gene expression count matrix of combined scRNAseq data of the samples analyzed with HTP (19). We filtered for genes that are expressed less than 5% within the given tumor phenotypes and randomly included only 75% of T- and B cells for better performance. The resulting signature matrix was used for imputing cellular fractions from the counts-per-million normalized HTP data without any batch correction or quantile normalization and 500 permutations.

Bioinformatic analysis of scRNAseq data

Preprocessing

Processing from FASTQ files 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 by evaluating quality criteria inspired by Luecken *et al.* (see additional file 1) (20). Cells were defined by having more than 500 unique molecular identifiers (UMIs), less than 10 % mitochondrial gene expression and additionally for patient 3 and 5 having more than 30 housekeeper genes expressed. The filtered count matrices (cells x genes) were further processed using Seurat version 4.0.1 and R version 4.0.5 (21). Demultiplexing of hashed libraries was performed choosing manual threshold of hashtag oligo (HTO) expression based on quality assessments described in Additional file 1. We removed doublets that were identified by demultiplexing of HTO expression matrices from all analyses.

Normalization and dimensionality reduction

When performing analysis of a specific set of cells, e.g., only tumor cells or only cells of a specific patient, the respective set of cells was normalized using the SCtransform algorithm and the 3000 most variable genes were selected for PCA (22). During normalization, we regressed for cell cycle scores and percentage of mitochondrial gene expression. Cell cycle scores and phases were determined in Seurat using log-normalized RNA counts and S- and G2M-Phase genes defined by Tirosh *et al*. (23). When generating a uniform manifold approximation and projection (UMAP) without patient-specific batch effect, we used corrected

PCs derived with the harmony R package version 0.1.0 [54]. Based on the variance explained by each PC and the respective ranked elbow plot we choose an appropriate number of PCs for UMAP visualization and SNN clustering as implemented in Seurat. For deriving patient-specific clusters for calculating the intratumoral cosine similarity, we used the same resolution parameter for comparability. UMAPs colored by specific gene expression were ordered by expression values.

Tumor cell identification and phenotyping

We annotated all cell types and identified tumor and fibroblast phenotypes by using a combination of methods: SNN clustering, differential gene expression, gene set enrichment analysis (GSEA), expression of literature-based marker genes and automated reference-based annotation with SingleR using the Monaco bulk-RNA Immune dataset (Supplementary Figure 2) (24, 25). Automated, reference-based annotation with SingleR version 1.4.1 was run on SNN clusters with a resolution of 100, yielding extremely small clusters including only few cells but increasing performance. Further, we excluded cells with ambiguous cell type marker expression. For example, within the tumor cells of the lymph node metastasis from patient 1, we observed few cells expressing genes typical for fibroblasts and DCs that were subsequently excluded from tumor-specific analysis. Similarly, we removed T cells, B cells, mast cells, fibroblasts, muscle cells, melanocytes, and other immune cells from the tumor cell subsets, as well as T cells highly expressing CD3 genes from the fibroblast subset. Malignant cells were first identified by high epithelial gene expression, e.g., high cytokeratin expression, and inferred CNVs (Supplementary Figure 2C-E). 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. Differential gene expression was performed by calculating the log2 foldchanges between one cluster and all other cells from the subset based on log-normalized data using NormalizeData function and a scale factor of 10,000. We filtered for genes with log2 foldchange greater than 0.25 and a minimum percentual expression of at least 10 % within the cluster or all other cells. For calculating the cosine similarity, we did not filter the log2 foldchanges. GSEA was

performed using the “fgsea” R package version 1.16.0, log2 foldchanges from differential gene expression and gene ontology biological processes (GO:BP,C5 v7.1) as well as hallmark gene sets (H, v7.1) downloaded from MSigDB database (26-28). Gene sets were included if they had at least 15 genes or at maximum 500 genes within the gene set using 10,000 permutations.

For deriving epithelial differentiating and pEMT gene signatures of patient 1 we calculated foldchanges between the “epi” and “pEMT” cluster and included genes with log2 foldchanges greater or lesser than 1, respectively, and with at least 10 % of either epi or pEMT cells expressing that gene. Gene set variation analysis (GSVA) was performed using the R package GSVA version 1.38.2 with default settings, i.e., gaussian kernel (29). As input, EMTome signatures, the pEMT and epithelial differentiation 1 and 2 signature from Puram *et al*. (10) and the three EMT and the epithelial senescence signatures from Kinker *et al.* were used (14).

Trajectories were inferred using SlingShot version 1.8.0 with malignant cell clusters as shown in Figure 2A of patient 1 and the first 20 PCs (30). RNA velocity was inferred using the VeloCyto python and R package (version 0.6) (31). Creation of the loom file was done using default options 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 the inverse correlation coefficient of the PC embedding correlation matrix as distance, 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 30 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) (32-34). Hierarchical clustering in the heatmaps was performed using the Euclidean distance and ward.D2 method unless otherwise noted. Visualization was performed using ggplot2 version 3.3.3 and ComplexHeatmap version 2.6.2 (35, 36).

Reanalysis of HNSCC dataset from Kürten *et al*.

From the publicly available scRNAseq data set on primary HNSCC tumors we downloaded the FASTQ files of CD45-negative and HPV-negative libraries from the sequencing read archive (SRA) under accession ID SRP301444 (17). The data sets were analyzed the same as described above. HPV-negative samples were chosen for comparability to the OSCC dataset. However, the HN07 tumor originated from the larynx, while all other samples originated from the oral cavity. We adjusted the cell identification thresholds based on our evaluation criteria pooled for all libraries: cells were defined by having more than 175 genes expressed, less than 10 % mitochondrial gene expression and more than 60 housekeeper genes expressed (see Additional file 1).

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 (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,* cytokeratins (*KRT5*, *KRT14*, *KRT17)* and stratifin (*SFN*) (Supplementary Figure 2C-E).

Detailed phenotyping of the cancer cells identified several clusters to which we could assign 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 exclude additional traits. Specifically, 515 cells exhibited a pEMT phenotype characterized by expression of a mixture of epithelial and mesenchymal genes such as matrix metallopeptidases (*MMPs*), caveolin-1 (*CAV1*) and C-X-C motif chemokine ligand 14 (*CXCL14*). Those genes were previously described in pEMT signatures and are enriched in the EMT hallmark gene set from the molecular signatures database (MSigDB) (Supplementary Figure 3A) (10, 27). 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 specific cytokeratins (e.g., *KRT6B* and *KRT16*). The EMP-related gene expression patterns correlate with established signatures

from the EMTome database (Supplementary Figure 3B-E) (9). Interestingly, 184 cells from both EMP-phenotypes were present in a mixed cluster characterized by the high expression of cell-cycle related genes (despite the fact that we applied cell cycle regression).

With respect to cell clusters whose gene expression was not predominately associated with EMP, 268 cells exhibited an immune-regulatory phenotype enriched for genes associated with cytokine-mediated responses and higher expression of the chemokines *CXCL1/2/3/8* and *CCL20*. For 554 cells, the gene expression pattern suggests adaptation to environmental factors. Specifically, 51 cells had higher expression of transcription factors *FOS* and *JUN,* suggesting a stress response, and 310 cells can be assumed to respond to hypoxic conditions in the tumor based on the higher expression of *NDRG1* and *EGLN3*, which are both regulated by oxygen levels (37, 38): *NDRG1* regulates stress response and p53-mediated caspase activation (37) and *EGLN3* has an important role in regulation of hypoxia-inducible factor 1 alpha (HIF1α) through prolyl hydroxylation (38). 193 cells expressed genes associated with amino acid metabolism, starvation response and mTORC1 signaling, i.e., a regulator of mitochondrial metabolism (39). The upregulated genes *ASNS*, *PSAT1* and *PHGDH* integrate the metabolites of serine and glycine metabolism into glycolysis and therefore fuel glycolysis with amino acids (40); hence, these cells appear to be adapted to low-glucose conditions.

OSCC cells in lymph node metastases undergo mesenchymal-epithelial transition

We next focused on possible dynamics within the predominantly EMP-related cancer cell clusters using a higher resolved shared-nearest neighbor (SNN) clustering. Based on this, we defined 4 pEMT clusters (pEMT-1 to 4), 4 clusters of more epithelial differentiated cells (epi-1 to 4) and one cluster with mixed phenotypes (mix) (Figure 2A). pEMT-1 is enriched for genes involved in coagulation such as *THBS1*, *CYR61* and *F3,* and may play a role in angiogenesis (Figure 2B, Supplementary Figure 4A). pEMT-2 and 3 both showed higher expression of extracellular matrix (ECM) remodeling genes, but in addition pEMT-2 was characterized by higher expression of cytokeratin *KRT15* and chemokine *CXCL14* while pEMT-3 showed higher expression of the serine protease inhibitor *SERPINA1* and podoplanin (*PDPN*) which mediates efficient ECM degradation by controlling invadopodia (41). Of the more epithelial differentiated

cell clusters, epi-2's expression profile is closest to the pEMT cluster, having higher expression of *MMP1* and lower expression of *SPRR1B* and *S100A8/A9* than epi-1, epi-3, or epi-4 (Figure 2B, E). Epi-3 showed increased expression of *S100A7* and *KRTDAP*, whereas epi-4 showed higher expression of kallikreins (*KLK6/7*), prostate stem cell antigen (*PSCA*) and adipogenesis regulatory factor (*ADIRF*). Both *ADIRF* and *PSCA* play a role in prostate cancer and *PSCA* is also reported as highly expressed in mucosal tissue, but less in HNSCC (42-44).

To gain a better understanding on the gene expression dynamics in this metastasis, we estimated RNA velocity, which predicts the short-term future development in gene expression of individual cells using the ratio of spliced and non-spliced mRNA counts (Figure 2C) (31). This analysis revealed that epithelial differentiated cells were strongly developing towards cluster epi-4, while most other cells show more or less random patterns of developmental directions, hence could not be interpreted. Tracking the developmental pathway within the metastasis by trajectory analysis across all EMP-related clusters revealed a major developmental axis between pEMT and epithelial differentiated cells that is diversifying within each end (Figure 2D, E, Supplementary Figure 4B). To confirm that the found progressive epithelial differentiation of the metastatic cells represents an MET, we inferred CNVs from the scRNAseq data. This demonstrated an increased number of copy number gains on chromosome 1, 8, 17, and 19 within epithelial differentiated cell clusters epi-1, epi-3, and epi-4 (Figure 2F, Supplementary Figure 4C). It should be noted, however, that CNVs on chromosomes 1 and 17 are associated with upregulated epithelial genes in close genomic proximity; thus, these two copy number gains may not represent true genomic CNVs, but rather reflect the high expression of these genes in epithelial differentiated cells (Supplementary Figure 4D, E).

Intra-tumoral heterogeneity of OSCC is driven by EMP

To test whether our observation that OSCC cells in one lymph node metastasis undergo a mesenchymal-epithelial transition is generally valid, we extended our analyses by adding 5 primary tumors and 8 matched lymph node metastases from 6 patients (Table 1). The patients presented with a history of tobacco smoking and alcohol abuse except the female patients #4

and #5, both of whom, however, also lack HPV positivity. From the publicly available scRNAseq data set on primary HNSCC tumors published by Kürten et al. (17) we chose to include the 9 HPV-negative primary tumors in our analysis, of which all but one originated from the oral cavity (HN07 originated from the larynx). In total, we analyzed 7,263 cancer cells from 16 different patients (Figure 3A). Importantly, the frequency of cancer cells was unevenly distributed across samples, which could not be explained by differences in tumor cell content across samples as determined for our cohort by histopathology (Supplementary Figure 1). The stability of epithelial tumor cell assemblies, which may not be sufficiently broken up by dissociation protocols, likely interfered with the generation of OSCC single-cell suspensions (Supplementary Figure 5A). In addition, as expected from the inter-patient heterogeneity, cancer cells were clustered based on their gene expression by patient rather than functional phenotype (Figure 3A, B). Thus, we accounted for the patient-specific effects with batch-corrected principal components (PCs) using the harmony R package which indeed resulted in a clustering by functional phenotypes (45) (Figure 3C, Supplementary Figure 5B-D). For annotation of the phenotype of the clusters we were guided by the gene signatures previously identified in the indicator sample, but also found several additional, predominantly immunoregulatory phenotypes. EMP-related phenotypes were present in all but one tumor sample with only one EMP cell (Figure 3D). To compare the EMP-related intrapatient heterogeneity of all analyzed tumor cells between patients, we performed differential gene expression within each patient and calculated the similarities between the resulting clusters of all patients. We first considered each patient individually and performed clustering, annotation, inferCNV and differential expression analysis. As exemplified for the primary OSCC of patient HN01, tumor cell clusters had the same inferred CNVs and were annotated based on their phenotype again using the indicator sample as a guide (Figure 3E, Supplementary Figure 5E, F). The cosine similarity between patient-specific clusters demonstrated that within each patient, the heterogeneity in EMP is most prominent and epithelial differentiated phenotypes are profoundly different from most other clusters, especially pEMT (Figure 3F, Supplementary Figure 5G). Indeed, all pEMT phenotypes are very similar to each other and show a large overlap of the gene expression patterns with predominantly immune- and metabolic-related

clusters. In epithelial differentiated cells, the most upregulated genes are *S100A8* and *S100A9*, encoding calprotectin, and *SPRR1B*, which are all members of the epidermal differentiation complex (46). Of note, hypoxia- and stress-related heterogeneity is similar between patients suggesting a reactive response rather than an aspect of tumor evolution.

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

The transcription factors ZEB1/2, TWIST1/2, Snail (*SNAI1*) and Slug (*SNAI2*) are key for regulation of EMP [7, 27]. While mRNA expression of *SNAI2* within single OSCC cells was reported by Puram *et al.*, the other transcription factors were not detected (10). Here, we confirm this observation as we detected *SNAI2* mRNA in almost half of the OSCC cells, but none of the other transcription factors (Figure 4A). However, detection of lowly expressed genes such as transcription factors by scRNAseq, especially in 10X genomics technology, becomes unreliable due to dropout effects (47). Also, the activity of transcription factors is often not reflected by the dynamics of their mRNA expression alone, as their activity additionally depends on protein stability and posttranslational modifications; for example, the ZEB1 protein is more stable than Snail (48). To circumvent this problem, we inferred the activity of these transcription factors based on the mRNA expression profile of their target genes using the algorithm VIPER with regulons defined by DoRothEA database (32-34). Using this approach, we were able to detect high activities of ZEB1, ZEB2, Snail and Slug in OSCC cells of different patients with varying EMP phenotype (Figure 4B). This shows that epithelial differentiation in OSCC metastases is associated with higher activity of the EMT activator ZEB1 (Figure 4C). Since, on the one hand, this was unexpected and, on the other hand, the method used to derive transcription factor activities potentially overestimates the activity for transcriptional repressors, we addressed the plausibility of this observation (Supplementary Figure 6). Consistent with the fact that one of the main functions of ZEB1 is the downregulation of E-cadherin (49) (*CDH1*), OSCC cells generally showed low expression of *CDH1* (Figure 4B). Next, we investigated the expression of the ZEB1 protein by immunohistochemistry (IHC) in all 14 tumor lesions of our cohort. We observed nuclear ZEB1 expression in similar tumor

areas as cytoplasmic cornifin-B expression, which served as a marker of epithelial differentiation (Figure 4D, E). Consequently, we validated the co-expression of ZEB-1 and cornifin-B in the same cell by immunofluorescence double-staining (Figure 4F). In line with our scRNAseq data, colocalization of both proteins was observed in a fraction of cancer cells in 9/14 (64 %) samples and double positive cells were more frequently observed in lymph node metastases (7/9, 78 %; Table 1) compared to primary tumors (2/5, 40 %).

Immunomodulating CAFs are present in primary tumors and tumor-involved lymph nodes.

Next, we investigated the OSCC tumor microenvironment (TME) and derived its potential impact on metastatic dissemination. For this, we additionally analyzed the scRNAseq data of 5 tumor-free lymph nodes from patients #4, #6 and #7 (Figure 5A, B). In this expanded cohort, most of the 41,284 cells were derived from the tumor-involved or tumor-free lymph nodes (34,599 cells, 84%), which as expected were predominantly immune cells, (35,856 cells, 87%, Figure 5C, Supplementary Figure 7A). The other non-malignant cells were fibroblasts (1,595 cells, 4%), pericytes (551 cells, 1 %), endothelial cells (399 cells, 1 %) and muscle cells (55 cells, 0.1 %). In comparison, due to the negative selection of CD45+ leukocytes, the data set of Kürten *et al*. shows a higher proportion of stromal cells, including endothelial cells (5,972 cells, 28 %), fibroblasts (3,067 cells, 15 %) and pericytes (673 cells, 3 %, Figure 5D, Supplementary Figure 7B). Quantification of cell type composition in scRNAseq datasets is difficult to interpret because of technical biases in sample preparation (e.g., larger and stiffer cell types are generally underrepresented) that results in cell number and patient-specific differences (Supplementary Figure 7C, D). Hence, we examined the bulk transcriptome and deconvoluted the respective cell types for our samples, revealing higher tumor and stroma cell content compared to cell type proportions derived from scRNAseq data (Supplementary Figure 8A). Still, the tumor-free lymph nodes contained a high number of lymphocytes, whereas the metastatic samples had a composition similar to primary tumors despite the relevant differences between samples (Supplementary Figure 8A, B).

Since the EMP status of tumor cells affects the properties of CAFs and vice versa, we focused on the transcriptional phenotypes of CAFs. Three main phenotypes were identified within our dataset: ECM-producing and -modifying fibroblasts (1,071 cells, 67 %), immunomodulating fibroblasts (311 cells, 19 %) and contractile myofibroblasts (144 cells, 9 %, Figure 5E, Supplementary Figure 9A-D). Additionally, there was a small population of fibroblast reticular cells (FRCs, 41 cells, 3 %) and myoblasts (28 cells, 2 %). These phenotypes were also present in the Kürten *et al.* dataset and due to the higher number of available cells, we were also able to differentiate the immunomodulating fibroblasts into further subtypes varying in their expression of chemokines and cytokines and to identify a fibroblast population associated with cell stress (Figure 5F, Supplementary Figure 9E-H).

The ECM-producing and -modifying phenotype is characterized by higher expression of MMPs, collagens (i.e., I, III, V and VI) and is enriched for gene sets related to formation and organization of the ECM (Supplementary Figure 9A-C, E-G). Cells with contractile functions include pericytes identified by expression of the regulator of G-protein signaling 5 (*RGS5*), myofibroblasts identified by cytoskeleton genes such as alpha smooth muscle actin 2 (*ACTA2*), actin gamma smooth muscle 2 (*ACTG2*) and myosin heavy chain 11 (*MYH11*), and myoblasts identified by desmin (*DES*), chordin like 2 (*CHRDL2*) and transcription factors associated to myogenic differentiation (*MYF5/6*, Supplementary Figure 9A, D, E, H). Moreover, myofibroblasts have enriched gene sets related to muscle contraction and similar to pericytes only express collagens IV and XVIII. Stress-associated cells express heat shock proteins (HSPA's), AP-1 related genes *JUN* and *FOS* and also ECM-producing genes, indicating they are ECM-producing fibroblasts impregnated with a transcriptional stress response signature as the predominant phenotype (Supplementary Figure 9G). Immunomodulating fibroblasts exhibit higher expression of chemokines such as *CXCL12* or *CXCL14*, cytokines such as Interleukin 6 (*IL6*), complement factors such as *C3* and *CFD,* and phospholipases such as *PLA2G2A* and *APOD,* with most enriched gene sets being related to immune response mechanisms (Supplementary Figure 9A, D, E, H). Hence, these cells probably exert an immune-modulatory effect within the TME. FRCs cluster closely to the immunomodulating cells

and highly express chemokines *CCL2*, *CCL8*, *CXCL2*, *CXCL12* as well as *CCL19* and *CCL21*. The latter two chemokines regulate lymphocyte homing and are characteristic of lymph node FRCs (50-52). FRCs which are usually present in mucosal, skin and lymph node tissue were accordingly most abundant in tumor-free lymph nodes (on average 18 % vs 6 % in metastatically affected lymph nodes, Supplementary Figure 9I) (53-55). Interestingly, we also detected FRCs within the primary tumors, suggesting they are functioning in mucosa-associated lymphoid tissue (MALT, Figure 9I, J).

Discussion

EMT represents the reactivation of an embryonic developmental program in which cells acquire migratory and invasive properties, i.e., prerequisites for invasion and metastasis of cancer (3, 56, 57). Thus, in early stages of metastasis, tumor cells undergo EMT, whereas in established metastases the reverse process *aka* MET is also observed (58, 59). To assess the EMP-associated heterogeneity among OSCC cells and gain some insight into the dynamics of this process, we examined the transcriptomes of 7,263 individual carcinoma cells isolated from primary and metastatic OSCC. Although we collected a high number of carcinoma cells in total, one of the limitations of this study is the often low number of malignant cells examined per tumor lesion. Despite this, we were able to demonstrate a progressive MET within a single, established lymph node metastases and confirm the EMP-associated heterogeneity in primary and metastatic OSCC. Interestingly, the epithelial differentiation in OSCC metastases is associated with higher activity of the EMT-activator ZEB1, which was confirmed on protein level by detection of co-expression of ZEB1 and cornifin-B in individual tumor cells using immunofluorescence staining. Consistent with previous reports showing that the EMP status of tumor cells influences the properties of CAFs and vice versa, we also detected distinct CAF phenotypes in primary tumors and tumor-involved lymph nodes; interestingly, immunomodulating fibroblasts were found throughout the metastatic cascade (5, 6).

EMP appears to be the main driver of cellular heterogeneity within OSCC: detailed phenotyping of cancer cells identified several clusters whose predominant functional phenotypes corresponded to different EMP states, ranging from a pEMT to a more epithelial

differentiated state. Moreover, pEMT phenotypes in particular might superimpose with traits related to angiogenesis, ECM remodeling, metabolic adaptations, stress, and interactions with the immune system. Metabolic adaptations include response to environmental limitations such as hypoxia and low glucose. Low glucose conditions are counteracted with upregulation of genes related to amino acid metabolism that fuel into glycolysis (40). While previous studies suggested that the activity of specific metabolic pathways in OSCC varies widely among patients (60), we observed that hypoxia- and stress-related gene expression patterns are similar between patients, supporting the notion of a reactive response rather than an aspect of individual tumor evolution.

In terms of EMP dynamics, it is assumed that cells may transit from one EMP state to another along a continuous spectrum of changes. Currently, however, it is also discussed if long-lived phenotypes representing discrete EMP states prevail (3, 7, 8, 10, 11, 13-15). Most studies supporting continuous transitions are based on *in vitro* or preclinical *in vivo* models that may not fully reflect the complexity of the tumor and its microenvironments (7, 8, 11). Indeed, human *in situ* or *ex vivo* studies suggested distinct EMP states; however, these approaches do not fully capture cellular dynamics (10, 14, 57). We demonstrated that within each patient, the EMP-driven differences are most prominent and epithelial differentiated phenotypes are profoundly different from pEMT clusters, suggesting that these states may be more static. Moreover, gene expression dynamics estimated by RNA velocity demonstrated epithelial differentiated cells were strongly developing towards a more pronounced epithelial differentiation with an increasing expression of genes of the epidermal differentiation complex (46). Of note, OSCC cells with a pEMT phenotype did not show such a uniform developmental direction. The assumption that epithelial differentiated metastatic cells developed later than pEMT cells, i.e., underwent MET, is supported by an increasing number of inferred copy number gains towards increasing epithelial differentiation even if accounting for the limitations of this approach.

However, we cannot conclude whether MET happened within the metastasis or primary malignancy, as our data reflects the tumor heterogeneity within a specific timepoint of tumor

evolution. As we observed a similar EMP heterogeneity in primary tumors, multiple disseminated tumor cells reflecting this heterogeneity might have migrated collectively, which could be crucial for metastatic consolidation (61).

Unexpectedly, we found high transcriptional activity of the EMT-activator ZEB1 in epithelial differentiated OSCC cells in both primary and metastatic tumor lesions, even considering that the scRNAseq data inferred transcription factor activities are biased towards transcriptional repressors. In the case of primary tumors this may be interpreted as the incipient EMT, but this hypothesis would not work for metastatic lesions where ZEB1 activity was associated with a progressive epithelial differentiation. Indeed, in previous studies, depletion of ZEB1 was reported to drive tumor cells from pEMT towards an epithelial phenotype (62, 63). However, depletion of Zeb1 in a mouse model also reduces phenotypic variability of cancer cells, particular their phenotypic/metabolic plasticity (62). While it is well established that ZEB1 together with microRNAs stabilizes EMT through a feedforward loop, this loop could also induce epithelial differentiation based on environmental factors (64). In addition to the transcriptional repressor activity, ZEB1 has been demonstrated to induce the epithelial differentiation marker cornifin-B in response to IL-1β and IFN-γ (65). We not only demonstrate the simultaneous occurrence of *SPRR1B* mRNA expression and ZEB1 activity, but also the co-localization of cornifin-B and the ZEB1 protein expression in OSCC lymph node metastases. Thus, although ZEB1 activity is crucial for the induction of the pEMT state, it does not seem to completely prevent partial epithelial differentiation. Remarkably, no relevant differences in *CDH1* expression were detected between the different EMP states in the metastatic OSCC lesions. Therefore, the more epithelial differentiated phenotypes we observed most closely correspond to a partial epithelial differentiation analogous with the observed pEMT phenotypes. We speculate that the driving force behind this EMP-associated heterogeneity of OSCC cells is to maintain cellular integrity. For example, ZEB1 is an ATM-substrate linking ATM and CHK1, promoting homologous recombination-dependent DNA repair and thereby protecting cells from genotoxic stress whereas expression of keratin intermediate filaments helps to protect cells from stress associated apoptosis (66, 67).

Similar to previous reports, we detected various fibroblast phenotypes in OSCC lesions, of which, remarkably, the immunomodulatory *CXCL14*-expressing fibroblasts were found in both primary tumors and lymph node metastases (10, 17, 55, 68). This indicates the special importance of this subgroup, as they may enable tumor cells to escape from the immune system. Using single-cell mRNAseq data, *CXCL14*-expressing fibroblasts have previously been detected in HNSCC, melanoma, and lung cancer lesions and are presumed to have immunosuppressive effects; the latter explained the association of their presence with poorer prognosis (68). However, CXCL14 is also constitutively expressed and secreted by fibroblasts and keratinocytes in healthy skin and mucosa (69, 70). Indeed, the effects of the chemokine CXCL14 seem to depend strongly on the cellular context (71). For example, restored *CXCL14* expression in HPV-positive oropharyngeal carcinoma is associated with better survival in immunocompetent syngeneic mice (72) but *CXCL14*-producing CAFs promoted tumor growth in a prostate cancer model (73). Similarly, ECM-modifying and contractile fibroblasts can promote or suppress tumor progression by consolidating or disrupting tissue structure, as ECM remodeling can affect both tumor and immune cell migratory ability (74).

Our study demonstrates that the comprehensive molecular characterization of tumor lesions captures both their complexity, as well as the heterogeneity between manifestations and the dynamics of their cellular composition (75). In particular, the heterogeneity of EMP status in HNSCC appears to be of translational importance as it provides further insight into tumor aggressiveness and treatment resistance. Similarly, it may help to assess the impact of systemic therapies on the microenvironment and correlate different EMP phenotypes with further clinical progression (76). This also applies to the effect of perioperative drugs, which are designed to counteract the spread of cancer by inhibiting stress-inflammatory responses such as the release of catecholamines and prostaglandins (77). Patients would therefore benefit from translational molecular companion programs by differentiating early effective from ineffective interventions.

In summary, the data presented here indicates that the interplay between tumor and stromal cell interactions is a highly complex process and that the EMP status of tumor cells and the polarization of stromal cells may influence each other. Our observations suggest that tumor cells and CAFs behave similarly in primary and metastatic OSCC samples. These findings may help to unravel the role of fibroblasts in predicting metastasis risk, which in turn may influence treatment decisions in OSCC.

Conclusions

Single cell transcriptomics reveals that heterogeneity within OSCC cells is dominated by EMP differences resulting in distinct partial EMT and epithelial differentiated phenotypes. Particularly, the partial EMT phenotypes can be accompanied by features related to metabolic adaptations, stress, and interaction with the immune system. In addition, CAFs were shown to be a major component of the TME, with immunomodulating *CXCL14*-expressing fibroblasts in both primary OSCC tumors and lymph node metastases indicating their relevance during immune escape. The EMP phenotypes likely endow capabilities that are essential for the different stages of the metastatic process, including maintenance, cellular integrity and polarization of stromal cells. This could be a possible additional function of ZEB1, as it is also expressed during progressive epithelial differentiation in OSCC metastases.

List of abbreviations

CAF	Cancer-associated fibroblast
CNV	Copy number variation
DC	Dendritic cell
DEG	Differentially expressed genes
EC	Endothelial cell
ECM	Extracellular matrix
EMP	Epithelial-mesenchymal plasticity
EMT	Epithelial-mesenchymal transition
FFPE	Formalin-fixed, paraffin-embedded
GSEA	Gene set enrichment analysis
GSVA	Gene set variation analysis
HNSCC	Head and neck squamous cell carcinomas
HPV	Human papillomavirus
HTO	Hashtag oligo
IHC	Immunohistochemistry
MET	Mesenchymal-epithelial transition
MSigDB	Molecular signature database
OSCC	Oral cavity squamous cell carcinoma
PC	Principal component
pEMT	Partial EMT
scRNAseq	Single-cell RNA sequencing
SNN	Shared-nearest neighbor
TME	Tumor microenvironment
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier

Supplementary Information

Supplementary Figure 1 [.pdf]: Histology of OSCC primary and metastatic tumors.

Supplementary Figure 2 [.pdf]: Cell type identification by marker genes, automated reference-based annotation, differential expression and inferred CNVs.

Supplementary Figure 3 [.pdf]: PEMT and epithelial differentiating gene expression signatures are comparable to previously published EMT signatures.

Supplementary Figure 4 [.pdf]: Extended analysis of the tumor phenotype characterization for the lymph node metastasis of patient 1.

Supplementary Figure 5 [.pdf]: Malignant phenotypes characterized across all analyzed patients.

Supplementary Figure 6 [.pdf]: Inferred transcription factor activity might be biased by activator or repressor function.

Supplementary Figure 7 [.pdf]: Cell type abundances across patients and tissue types.

Supplementary Figure 8 [.pdf]: Bulk transcriptomes reveal the cellular composition of OSCC across tissue types.

Supplementary Figure 9 [.pdf]: Characterization of OSCC-derived fibroblasts.

Table 1 [.xlsx]: Clinical and sequencing information from OSCC patients.

Additional file 1 [.pdf]: Supplementary Materials and Methods.

Declarations

Data availability

The processed datasets generated and analyzed during the current study are available in the GEO database with accession id GSE195655. The raw FASTQ files are available upon request due to privacy reasons. The high-resolution H&E Images are available through figshare under the DOI: https://doi.org/10.6084/m9.figshare.20905837.v1. The publicly available FASTQ files from Kürten *et al.* were downloaded from the sequencing read archive (SRA) accession id SRP301444.

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 and Severina Klaus, Center for Infectious Diseases, Heidelberg University Hospital, Germany, for carefully reading and improving 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.

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).

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.

Ethics approval and consent to participate

Written informed consent was obtained from each patient 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.

Consent for publication

Not applicable.

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.

References

1. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. Nature Reviews Disease Primers. 2020;6(1):92.

2. Cramer JD, Burtness B, Le QT, Ferris RL. The changing therapeutic landscape of head and neck cancer. Nature Reviews Clinical Oncology. 2019;16(11):669-83.

3. Nieto MA, Huang Ruby Y-J, Jackson Rebecca A, Thiery Jean P. EMT: 2016. Cell. 2016;166(1):21-45.

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(6):341-52.

5. Brabletz S, Schuhwerk H, Brabletz T, Stemmler MP. Dynamic EMT: a multi-tool for tumor progression. The EMBO Journal. 2021;40(18):e108647.

6. del Pozo Martin Y, Park D, Ramachandran A, Ombrato L, Calvo F, Chakravarty P, et al. Mesenchymal Cancer Cell-Stroma Crosstalk Promotes Niche Activation, Epithelial Reversion, and Metastatic Colonization. Cell Reports. 2015;13(11):2456-69.

7. 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. Nature Genetics. 2019;51(9):1389-98.

8. Cook DP, Vanderhyden BC. Context specificity of the EMT transcriptional response. Nature Communications. 2020;11(1):2142.

9. 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(1):259-69.

10. 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(7):1611-24.e24.

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(7702):463-8.

12. 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(32):5049-65.

13. Jolly MK, Tripathi SC, Jia D, Mooney SM, Celiktas M, Hanash SM, et al. Stability of the hybrid epithelial/mesenchymal phenotype. Oncotarget. 2016;7(19).

14. Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM, et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. Nature Genetics. 2020;52(11):1208-18.

15. Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB, et al. Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. Nat Commun. 2019;10(1):5587.

16. Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, et al. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. Cell. 2020;182(2):497-514.e22.

17. Kürten CHL, Kulkarni A, Cillo AR, Santos PM, Roble AK, Onkar S, et al. Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. Nature Communications. 2021;12(1):7338.

18. Spassova I, Ugurel S, Terheyden P, Sucker A, Hassel JC, Ritter C, et al. Predominance of Central Memory T Cells with High T-Cell Receptor Repertoire Diversity is Associated with Response to PD-1/PD-L1 Inhibition in Merkel Cell Carcinoma. Clinical Cancer Research. 2020;26(9):2257-67.

19. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773-82.

20. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019;15(6):e8746.

21. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87.e29.

22. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):296.

23. 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(6282):189.

24. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163-72.

25. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carré C, et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. Cell Rep. 2019;26(6):1627-40.e7.

26. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2021:060012.

27. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.

28. 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(43):15545.

29. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinform. 2013;14(1):7.

30. 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(1):477.

31. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560(7719):494-8.

32. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019;29(8):1363-75.

33. Holland CH, Tanevski J, Perales-Patón J, Gleixner J, Kumar MP, Mereu E, et al. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. Genome Biology. 2020;21(1):36.

34. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat Genet. 2016;48(8):838-47.

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

36. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9.

37. Guo D-D, Xie K-F, Luo X-J. Hypoxia-induced elevated NDRG1 mediates apoptosis through reprograming mitochondrial fission in HCC. Gene. 2020;741:144552.

38. Ivan M, Kaelin WG. The EGLN-HIF O2-Sensing System: Multiple Inputs and Feedbacks. Mol Cell. 2017;66(6):772-9.

39. Laplante M, Sabatini DM. mTOR signaling at a glance. J Cell Sci. 2009;122(20):3589-94.

40. DeBerardinis Ralph J. Serine Metabolism: Some Tumors Take the Road Less Traveled. Cell Metab. 2011;14(3):285-6.

41. Martín-Villar E, Borda-d'Agua B, Carrasco-Ramirez P, Renart J, Parsons M, Quintanilla M, et al. Podoplanin mediates ECM degradation by squamous carcinoma cells through control of invadopodia stability. Oncogene. 2015;34(34):4531-44.

42. de Nooij-van Dalen AG, van Dongen GAMS, Smeets SJ, Nieuwenhuis EJC, Stigter-van Walsum M, Snow GB, et al. Characterization of the human Ly-6 antigens, the newly annotated member Ly-6K included, as molecular markers for head-and-neck squamous cell carcinoma. Int J Cancer. 2003;103(6):768-74.

43. Reiter RE, Gu Z, Watabe T, Thomas G, Szigeti K, Davis E, et al. Prostate stem cell antigen: A cell surface marker overexpressed in prostate cancer. Proc Natl Acad Sci USA. 1998;95(4):1735-40.

44. Zhou K, Arslanturk S, Craig DB, Heath E, Draghici S. Discovery of primary prostate cancer biomarkers using cross cancer learning. Sci Rep. 2021;11(1):10433.

45. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289-96.

46. Oh IY, de Guzman Strong C. The Molecular Revolution in Cutaneous Biology: EDC and Locus Control. J Invest Dermatol. 2017;137(5):e101-e4.

47. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. Nature Methods. 2014;11(7):740-2.

48. Stemmler MP, Eccles RL, Brabletz S, Brabletz T. Non-redundant functions of EMT transcription factors. Nature Cell Biology. 2019;21(1):102-12.

49. 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(24):3490-500.

50. Link A, Vogt TK, Favre S, Britschgi MR, Acha-Orbea H, Hinz B, et al. Fibroblastic reticular cells in lymph nodes regulate the homeostasis of naive T cells. Nature Immunology. 2007;8(11):1255-65.

51. Rodda LB, Lu E, Bennett ML, Sokol CL, Wang X, Luther SA, et al. Single-Cell RNA Sequencing of Lymph Node Stromal Cells Reveals Niche-Associated Heterogeneity. Immunity. 2018;48(5):1014-28.e6.

52. Brown FD, Turley SJ. Fibroblastic Reticular Cells: Organization and Regulation of the T Lymphocyte Life Cycle. The Journal of Immunology. 2015;194(4):1389.

53. Oliveira-Neto HH, de Souza PPC, da Silva MRB, Mendonça EF, Silva TA, Batista AC. The expression of chemokines CCL19, CCL21 and their receptor CCR7 in oral squamous cell carcinoma and its relevance to cervical lymph node metastasis. Tumor Biology. 2013;34(1):65-70.

54. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. Science Advances. 2021;7(31):eabh2169.

55. Solé-Boldo L, Raddatz G, Schütz S, Mallm J-P, Rippe K, Lonsdorf AS, et al. Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. Communications Biology. 2020;3(1):188.

56. Williams ED, Gao D, Redfern A, Thompson EW. Controversies around epithelial–mesenchymal plasticity in cancer metastasis. Nat Rev Cancer. 2019;19(12):716-32.

57. Derynck R, Weinberg RA. EMT and Cancer: More Than Meets the Eye. Dev Cell. 2019;49(3):313-6.

58. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. Nature Reviews Molecular Cell Biology. 2019;20(2):69-84.

59. 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(6):725-36.

60. Xiao Z, Dai Z, Locasale JW. Metabolic landscape of the tumor microenvironment at single cell resolution. Nat Commun. 2019;10(1):3763.

61. Aiello NM, Maddipati R, Norgard RJ, Balli D, Li J, Yuan S, et al. EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. Developmental Cell. 2018;45(6):681-95.e4.

62. Krebs AM, Mitschke J, Lasierra Losada M, Schmalhofer O, Boerries M, Busch H, et al. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. Nature Cell Biology. 2017;19(5):518-29.

63. Carstens JL, Yang S, Correa de Sampaio P, Zheng X, Barua S, McAndrews KM, et al. Stabilized epithelial phenotype of cancer cells in primary tumors leads to increased colonization of liver metastasis in pancreatic cancer. Cell Rep. 2021;35(2):108990.

64. Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S, et al. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO reports. 2008;9(6):582-9.

65. 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(5):2466-75.

66. Zhang P, Wei Y, Wang L, Debeb BG, Yuan Y, Zhang J, et al. ATM-mediated stabilization of ZEB1 promotes DNA damage response and radioresistance through CHK1. Nat Cell Biol. 2014;16(9):864-75.

67. Oshima RG. Apoptosis and keratin intermediate filaments. Cell Death Differ. 2002;9(5):486-92.

68. Galbo PM, Jr, Zang X, Zheng D. Molecular Features of Cancer-associated Fibroblast Subtypes and their Implication on Cancer Pathogenesis, Prognosis, and Immunotherapy Resistance. Clinical Cancer Research. 2021;27(9):2636-47.

69. Kurth I, Willimann K, Schaerli P, Hunziker T, Clark-Lewis I, Moser B. Monocyte selectivity and tissue localization suggests a role for breast and kidney-expressed chemokine (BRAK) in macrophage development. J Exp Med. 2001;194(6):855-61.

70. Shellenberger TD, Wang M, Gujrati M, Jayakumar A, Strieter RM, Burdick MD, et al. BRAK/CXCL14 Is a Potent Inhibitor of Angiogenesis and a Chemotactic Factor for Immature Dendritic Cells. Cancer Research. 2004;64(22):8262-70.

71. Gowhari Shabgah A, Haleem Al-qaim Z, Markov A, Valerievich Yumashev A, Ezzatifar F, Ahmadi M, et al. Chemokine CXCL14; a double-edged sword in cancer development. International Immunopharmacology. 2021;97:107681.

72. Ozawa S, Kato Y, Ito S, Komori R, Shiiki N, Tsukinoki K, et al. Restoration of BRAK/CXCL14 gene expression by gefitinib is associated with antitumor efficacy of the drug in head and neck squamous cell carcinoma. Cancer Science. 2009;100(11):2202-9.

73. Augsten M, Sjöberg E, Frings O, Vorrink SU, Frijhoff J, Olsson E, et al. Cancer-associated fibroblasts expressing CXCL14 rely upon NOS1-derived nitric oxide signaling for their tumor-supporting properties. Cancer Res. 2014;74(11):2999-3010.

74. Kai F, Drain AP, Weaver VM. The Extracellular Matrix Modulates the Metastatic Journey. Developmental Cell. 2019;49(3):332-46.

75. Yin J, Zheng S, He X, Huang Y, Hu L, Qin F, et al. Identification of molecular classification and gene signature for predicting prognosis and immunotherapy response in HNSCC using cell differentiation trajectories. Scientific Reports. 2022;12(1):20404.

76. Wang D, Pei P, Shea FF, Bissonnette C, Nieto K, Din C, et al. Fenretinide combines perturbation of signaling kinases, cell–extracellular matrix interactions and matrix metalloproteinase activation to inhibit invasion in oral squamous cell carcinoma cells. Carcinogenesis. 2022;43(9):851-64.

77. Shaashua L, Shabat-Simon M, Haldar R, Matzner P, Zmora O, Shabtai M, et al. Perioperative COX-2 and β-Adrenergic Blockade Improves Metastatic Biomarkers in Breast Cancer Patients in a Phase-II Randomized Trial. Clinical Cancer Research. 2017;23(16):4651-61.

Figure 1

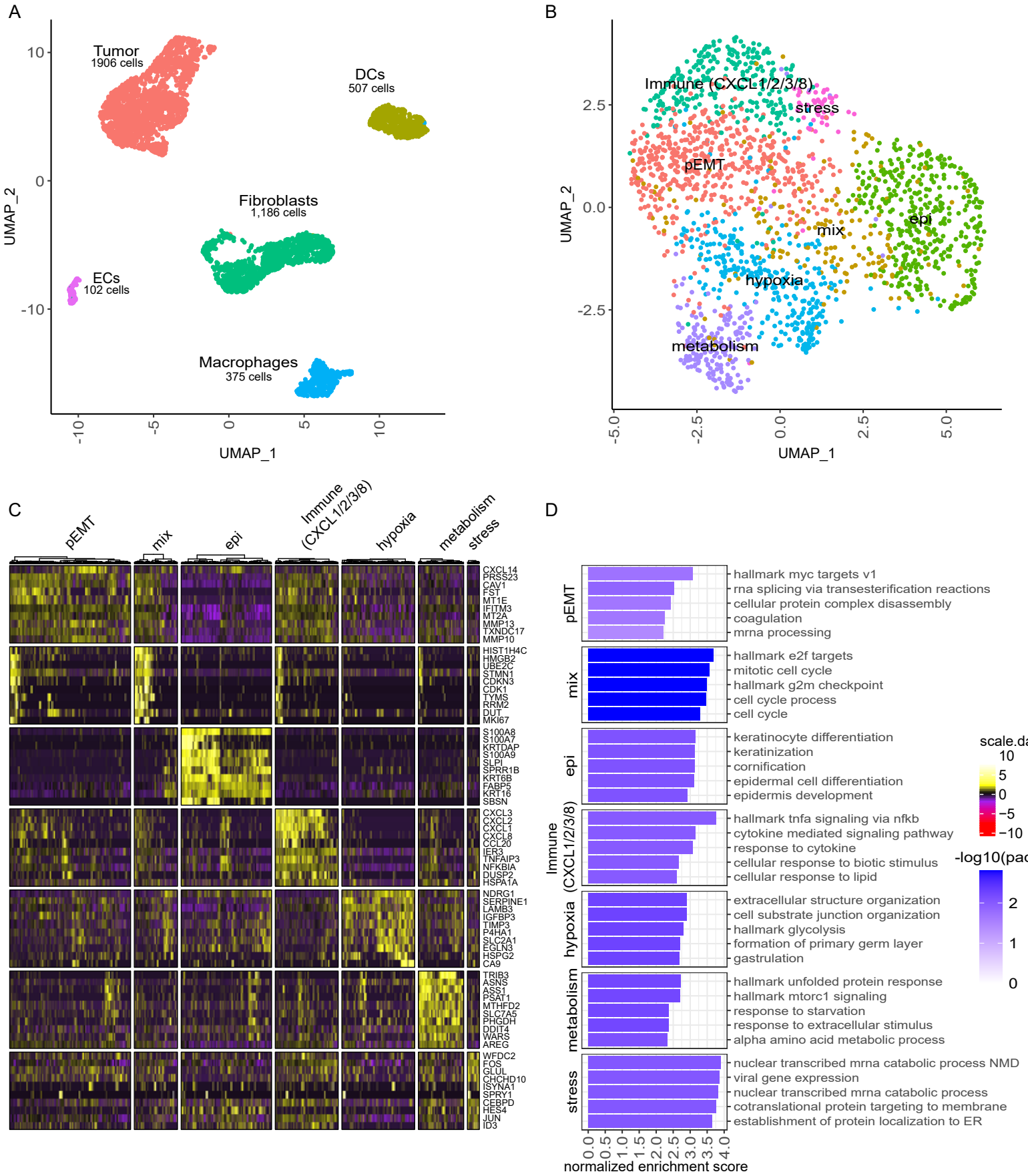

A
Tumor
1906 cells
DCs
507 cells
Fibroblasts
1,186 cells
ECs
102 cells
Macrophages
375 cells
UMAP_2
UMAP_1
B
Immune (CXCL1/2/3/8)
stress
pEMT
mix
epi
hypoxia
metabolism
UMAP_2
UMAP_1
C
pEMT
mix
epi
Immune (CXCL1/2/3/8)
hypoxia
metabolism
stress
CXCL14
PRSS23
CAV1
FST
MT1E
IFITM3
MT2A
MMP13
TXNDC17
MMP10
HIST1H4C
HMGB2
UBE2C
STMN1
CDKN3
CDK1
TYMS
RRM2
DUT
MKI67
S100A8
S100A7
KRTDAP
S100A9
SLPI
SPRR1B
KRT6B
FABP5
KRT16
SBSN
CXCL3
CXCL2
CXCL1
CXCL8
CCL20
IER3
TNFAIP3
NFKBIA
DUSP2
HSPA1A
NDRG1
SERPINE1
LAMB3
IGFBP3
TIMP3
P4HA1
SLC2A1
EGLN3
HSPG2
CA9
TRIB3
ASNS
ASS1
PSAT1
MTHFD2
SLC7A5
PHGDH
DDIT4
WARS
AREG
WFDC2
FOS
GLUL
CHCHD10
ISYNA1
SPRY1
CEBPD
HES4
JUN
ID3
D
pEMT
hallmark myc targets v1
rna splicing via transesterification reactions
cellular protein complex disassembly
coagulation
mrna processing
mix
hallmark e2f targets
mitotic cell cycle
hallmark g2m checkpoint
cell cycle process
cell cycle
epi
keratinocyte differentiation
keratinization
cornification
epidermal cell differentiation
epidermis development
Immune (CXCL1/2/3/8)
hallmark tnfa signaling via nfkb
cytokine mediated signaling pathway
response to cytokine
cellular response to biotic stimulus
cellular response to lipid
hypoxia
extracellular structure organization
cell substrate junction organization
hallmark glycolysis
formation of primary germ layer
gastrulation
metabolism
hallmark unfolded protein response
hallmark mtorc1 signaling
response to starvation
response to extracellular stimulus
alpha amino acid metabolic process
stress
nuclear transcribed mrna catabolic process NMD
viral gene expression
nuclear transcribed mrna catabolic process
cotranslational protein targeting to membrane
establishment of protein localization to ER
normalized enrichment score
scale.data
-log10(padj)

Figure 1 | Single-cell gene expression signatures in OSCC cells from a single metastasis reveal predominant 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 1,906 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

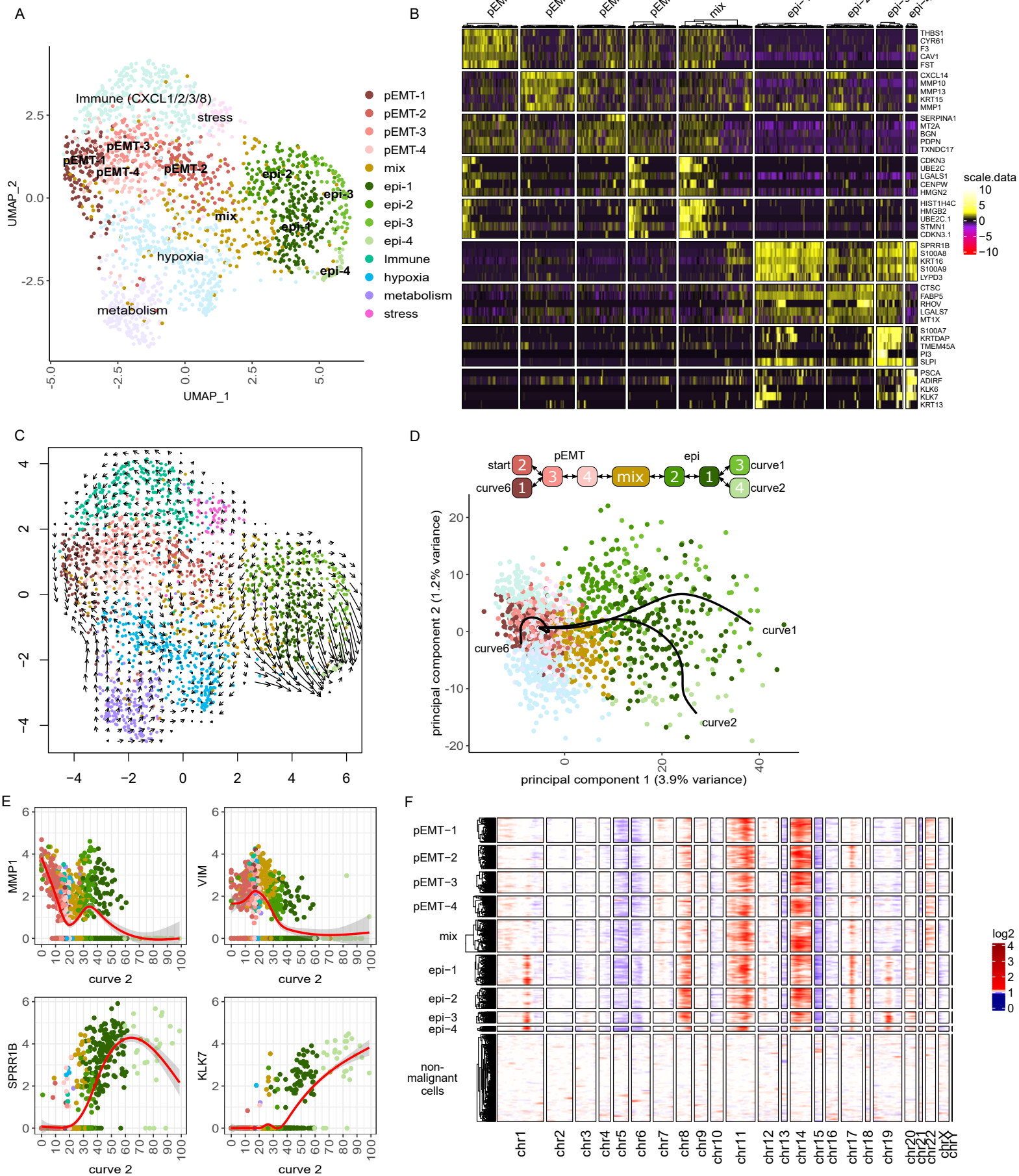

A
Immune (CXCL1/2/3/8)
stress
pEMT-3
pEMT-1
pEMT-4
pEMT-2
epi-2
epi-3
mix
epi-1
hypoxia
epi-4
metabolism
UMAP_2
UMAP_1
pEMT-1
pEMT-2
pEMT-3
pEMT-4
mix
epi-1
epi-2
epi-3
epi-4
Immune
hypoxia
metabolism
stress
B
THBS1
CYR61
F3
CAV1
FST
CXCL14
MMP10
MMP13
KRT15
MMP1
SERPINA1
MT2A
BGN
PDPN
TXNDC17
CDKN3
UBE2C
LGALS1
CENPW
HMGN2
HIST1H4C
HMGB2
UBE2C.1
STMN1
CDKN3.1
SPRR1B
S100A8
KRT16
S100A9
LYPD3
CTSC
FABP5
RHOV
LGALS7
MT1X
S100A7
KRTDAP
TMEM45A
PI3
SLPI
PSCA
ADIRF
KLK6
KLK7
KRT13
scale.data
C
D
start
pEMT
mix
epi
curve1
curve2
curve6
principal component 2 (1.2% variance)
principal component 1 (3.9% variance)
E
MMP1
VIM
SPRR1B
KLK7
curve 2
F
non-malignant cells
log2

Figure 2 | A progressive epithelial differentiation, but no strong uniform direction of development in pEMT clusters. (A) UMAP of 1,906 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; arrow length indicates strength of 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

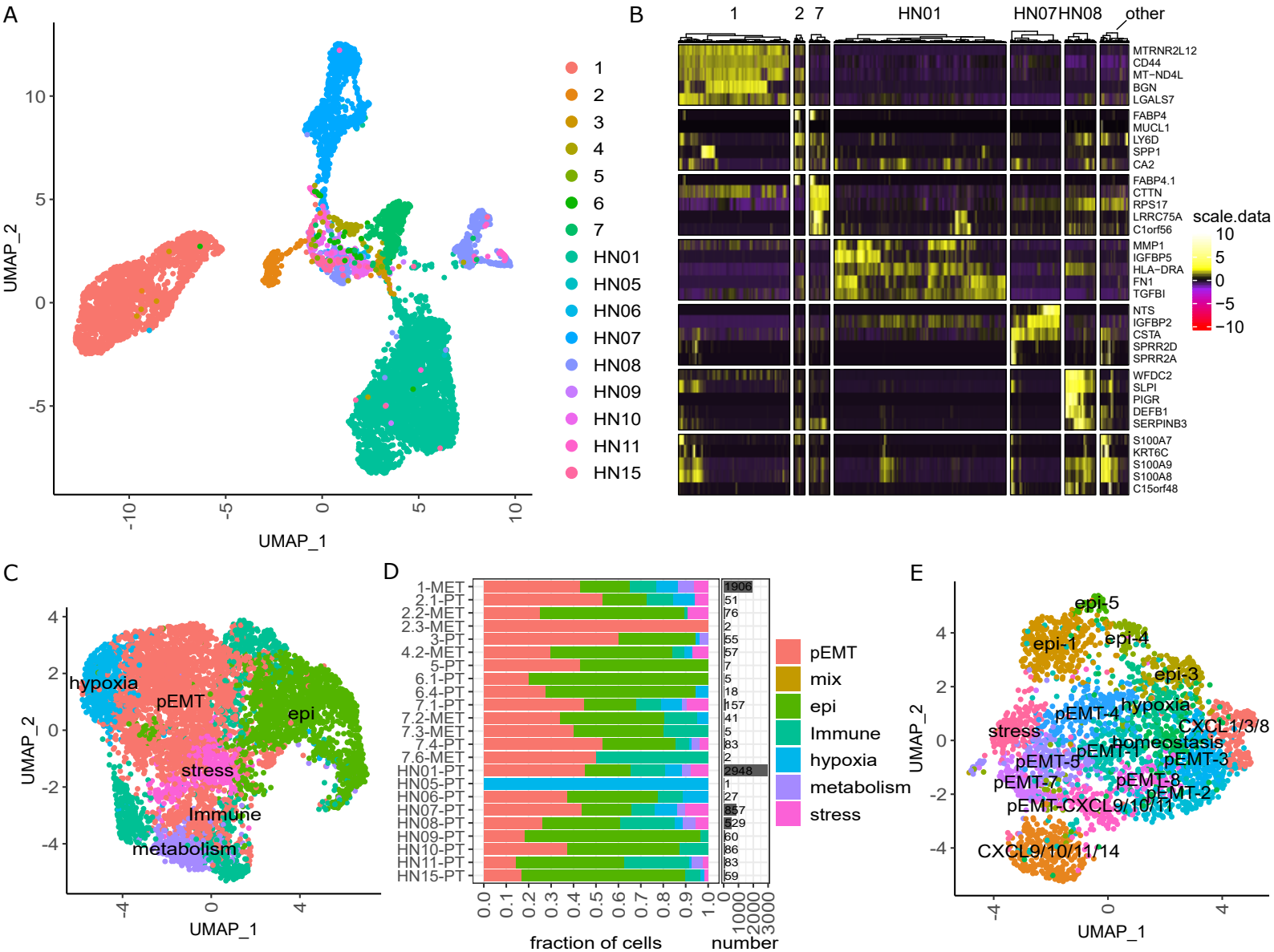

A
UMAP_2
UMAP_1
1
2
3
4
5
6
7
HN01
HN05
HN06
HN07
HN08
HN09
HN10
HN11
HN15
B
1
2 7
HN01
HN07HN08
other
MTRNR2L12
CD44
MT−ND4L
BGN
LGALS7
FABP4
MUCL1
LY6D
SPP1
CA2
FABP4.1
CTTN
RPS17
LRRC75A
C1orf56
MMP1
IGFBP5
HLA−DRA
FN1
TGFBI
NTS
IGFBP2
CSTA
SPRR2D
SPRR2A
WFDC2
SLPI
PIGR
DEFB1
SERPINB3
S100A7
KRT6C
S100A9
S100A8
C15orf48
scale.data
C
hypoxia
pEMT
epi
stress
Immune
metabolism
D
1-MET
2.1-PT
2.2-MET
2.3-MET
3-PT
4.2-MET
5-PT
6.1-PT
6.4-PT
7.1-PT
7.2-MET
7.3-MET
7.4-PT
7.6-MET
HN01-PT
HN05-PT
HN06-PT
HN07-PT
HN08-PT
HN09-PT
HN10-PT
HN11-PT
HN15-PT
fraction of cells
number
pEMT
mix
epi
Immune
hypoxia
metabolism
stress
E
epi-5
epi-1
epi-4
epi-3
hypoxia
pEMT-4
stress
CXCL1/3/8
homeostasis
pEMT-1
pEMT-5
pEMT-3
pEMT-7
pEMT-8
pEMT-2
pEMT-CXCL9/10/11
CXCL9/10/11/14
F
pEMT-4
pEMT-1
pEMT-3
CXCL1/2/3/8
mix
metabolism
pEMT-2
stress
hypoxia
epi-2
epi-1
epi-3
epi-4
epi-5
epi-1
epi-4
hypoxia
epi-3
stress
pEMT-5
pEMT-7
pEMT-4
CXCL9/10/11/14
pEMT-CXCL9/10/11
pEMT-8
pEMT-1
pEMT-2
CXCL1/3/8
pEMT-3
homeostasis
cosine similarity

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

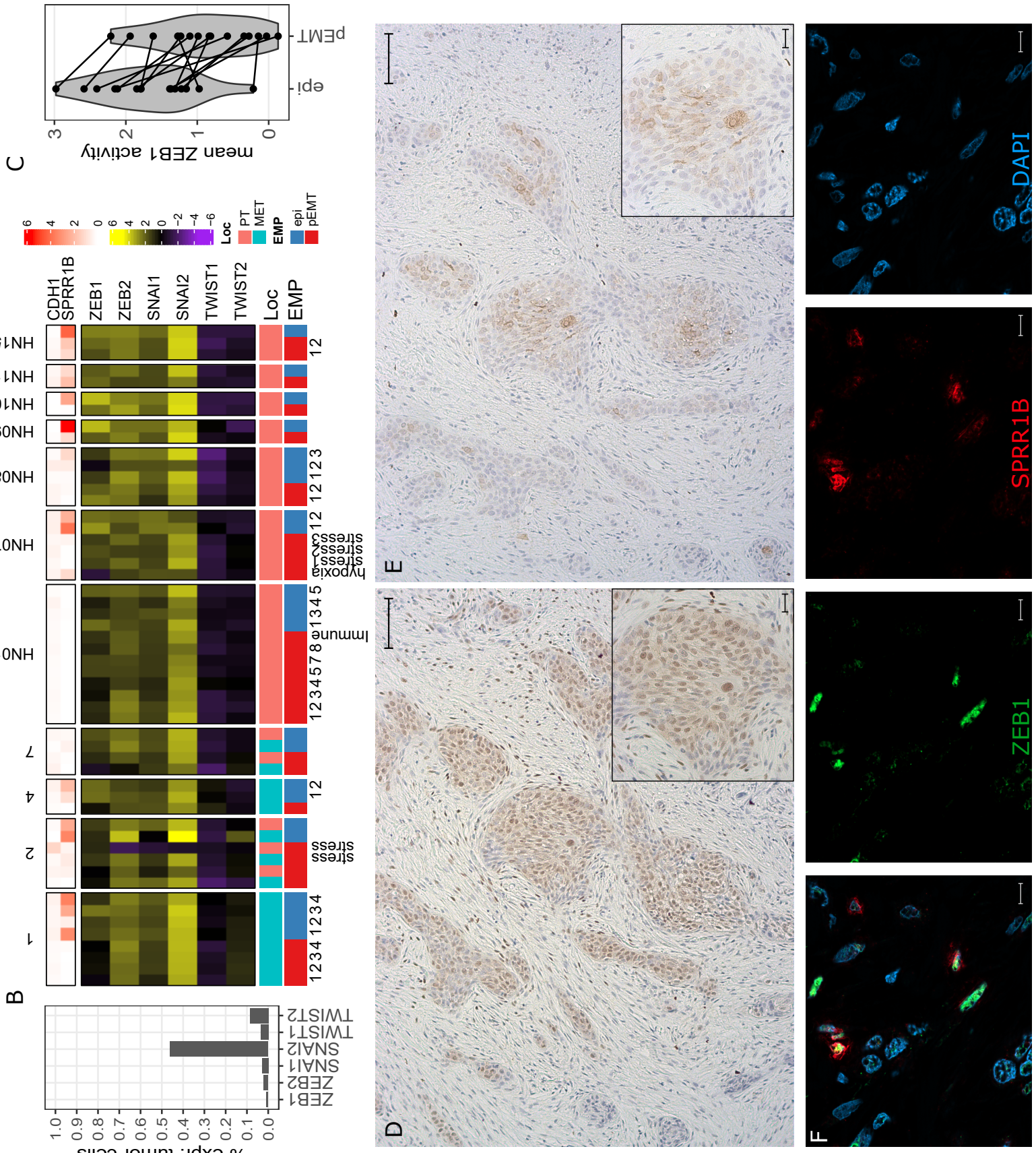

A
% expr. tumor cells
ZEB1
ZEB2
SNAI1
SNAI2
TWIST1
TWIST2
B
CDH1
SPRR1B
ZEB1
ZEB2
SNAI1
SNAI2
TWIST1
TWIST2
Loc
EMP
PT
MET
epi
pEMT
HN01
HN07
HN08
HN09
HN10
HN11
HN15
hypoxia
stress1
stress2
immune
C
mean ZEB1 activity
epi
pEMT
D
E
F
DAPI
SPRR1B
ZEB1

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. **(E)** 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.

Figure 5

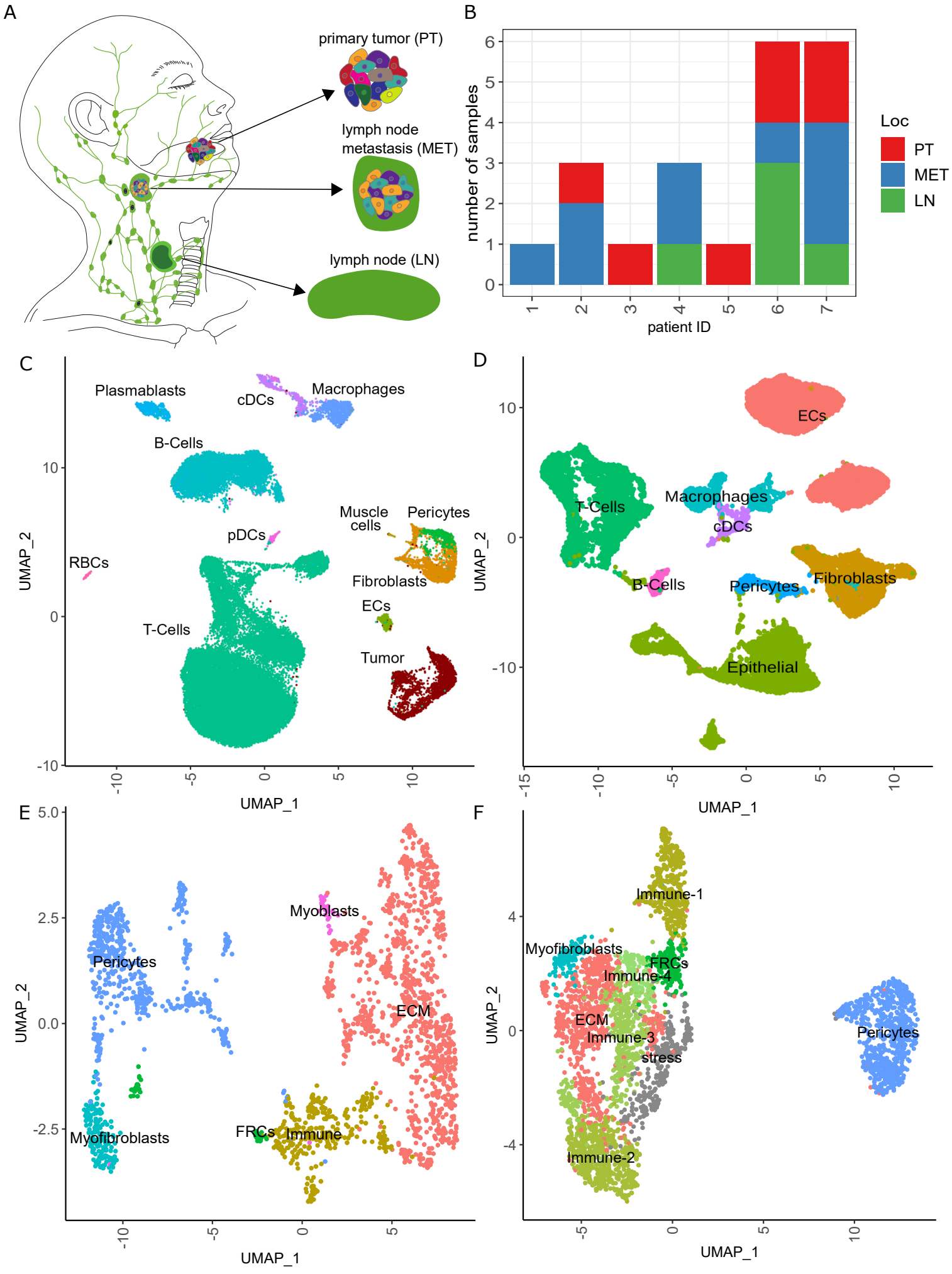
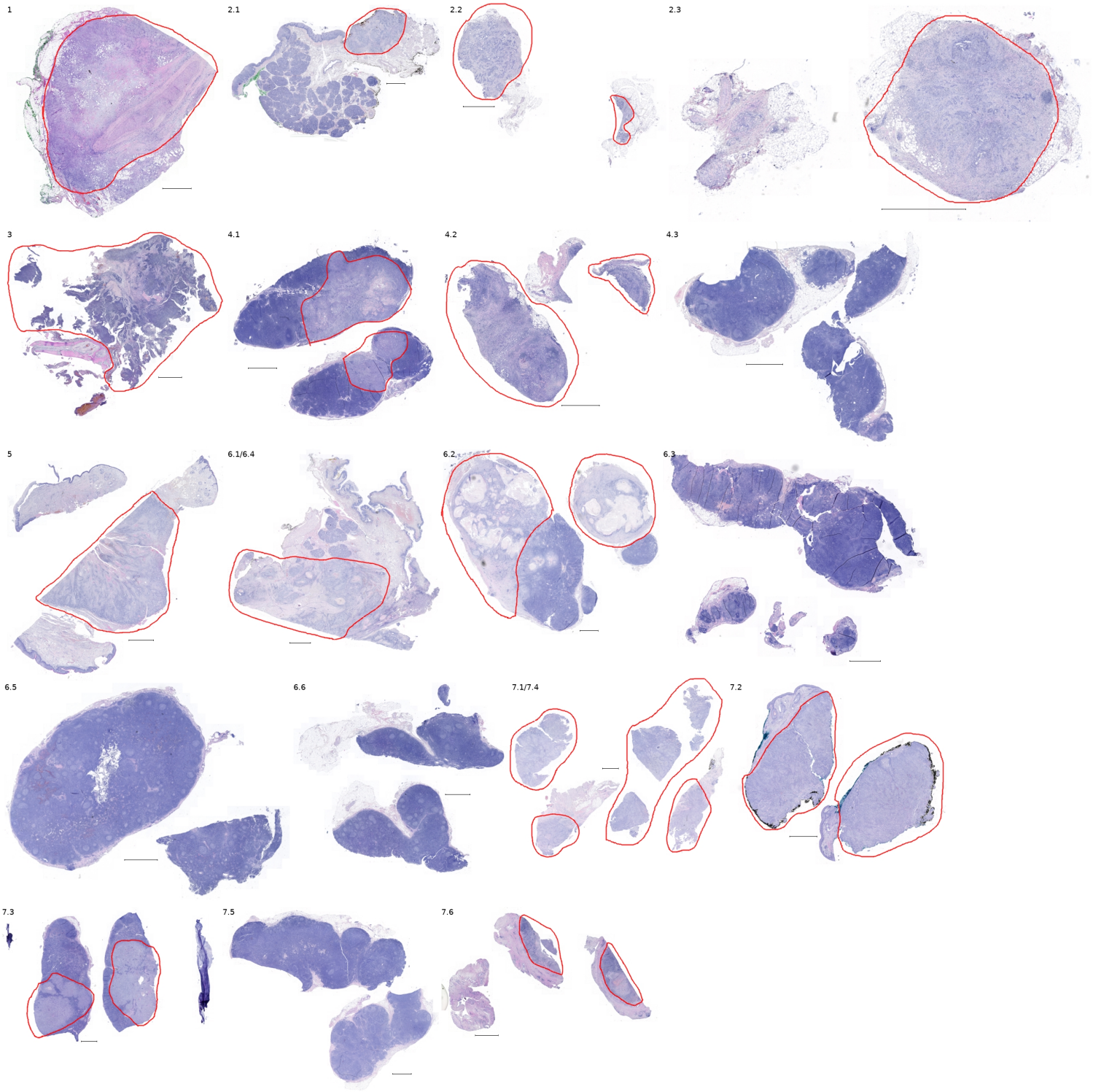
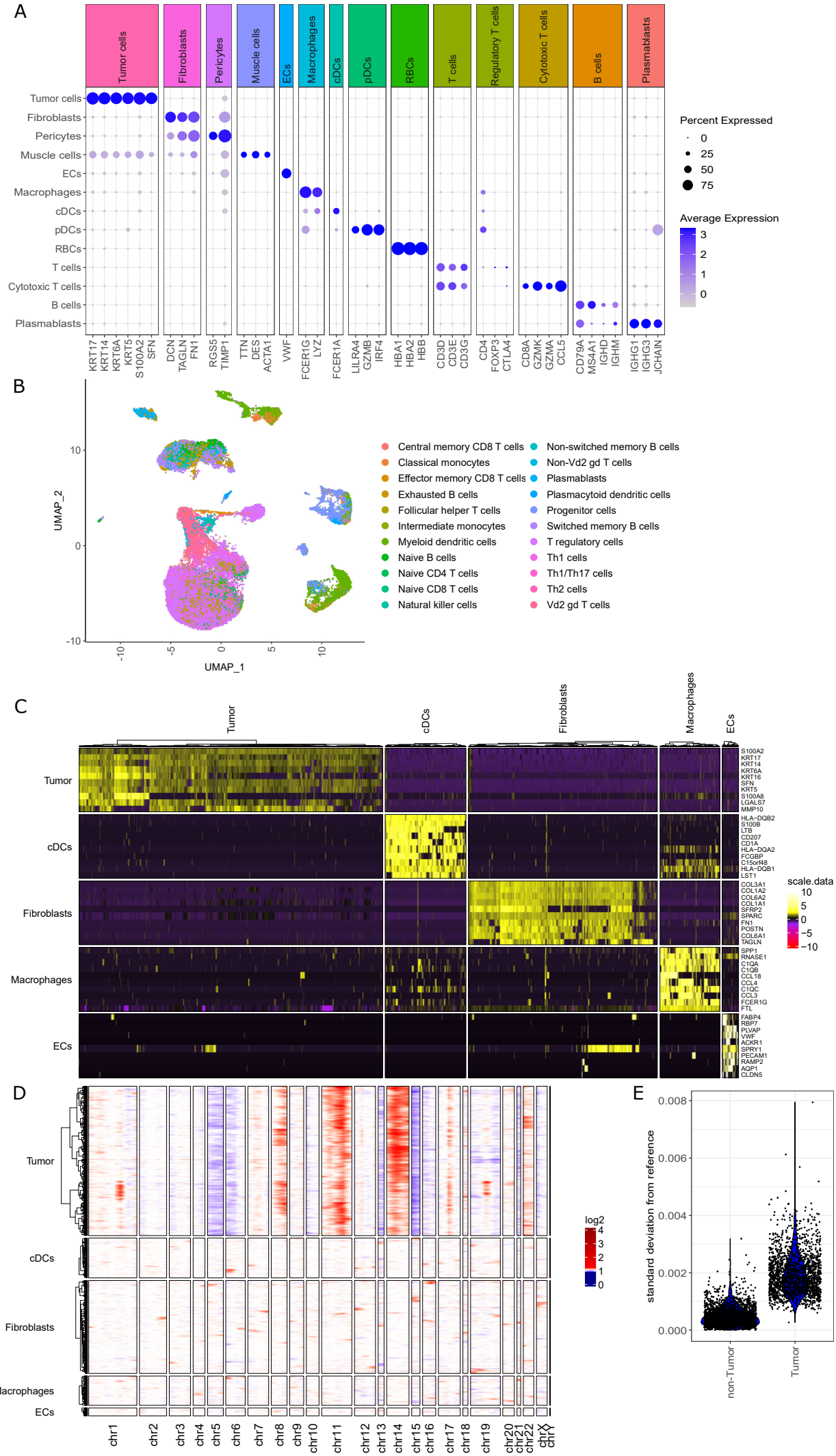

A
primary tumor (PT)
lymph node metastasis (MET)
lymph node (LN)
B
number of samples
patient ID
Loc
PT
MET
LN
C
Plasmablasts
cDCs
Macrophages
B-Cells
Muscle cells
Pericytes
pDCs
RBCs
Fibroblasts
ECs
T-Cells
Tumor
UMAP_1
UMAP_2
D
ECs
T-Cells
Macrophages
cDCs
B-Cells
Pericytes
Fibroblasts
Epithelial
UMAP_1
UMAP_2
E
Myoblasts
Pericytes
ECM
Myofibroblasts
FRCs
Immune
UMAP_1
UMAP_2
F
Immune-1
Myofibroblasts
FRCs
Immune-4
ECM
Immune-3
stress
Pericytes
Immune-2
UMAP_1
UMAP_2

Figure 5 | The OSCC microenvironment is composed of heterogenous fibroblasts from which immunomodulatory cells are present across the metastatic cascade. (A) Overview of types of analyzed OSCC samples and their localization within the head and neck area. **(B)** Number of samples across patients colored by their respective tissue origin; for patient #4, the primary tumor could not be analyzed due to incorrect specimen processing; for patients #6 and #7 two regions of the primary tumor were analyzed, denoted as sample #6.1 and #6.4 and #7.1 and #7.4, respectively. **(C)** UMAP of 41,284 cells based on OSCC scRNAseq data from our cohort and colored by cell type. **(D)** UMAP of 21,037 cells based on CD45-negative and HPV-negative primary HNSCC from Kürten *et al.* and colored by cell type. **(E)** UMAP of 1,595 fibroblasts and 551 pericytes from C colored by the respective phenotypes derived from shared-nearest neighbor clusters. **(F)** UMAP of 2,920 fibroblasts and 683 pericytes from D colored by the respective phenotypes derived from shared-nearest neighbor clusters.

Supplementary Figure 1

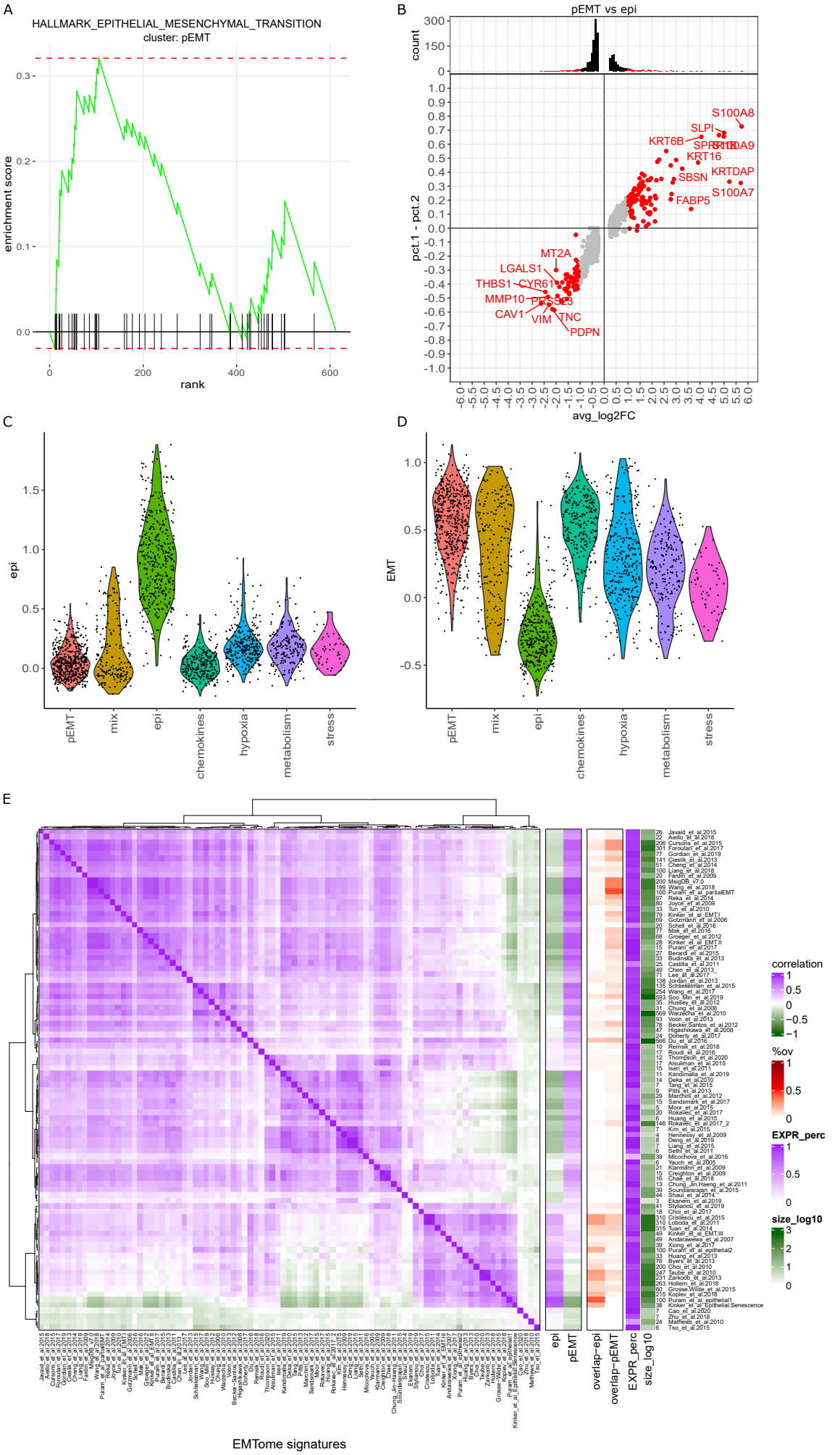


Supplementary Figure 1 | Histology of OSCC primary and metastatic tumors. Whole-slide image H&E staining for FFPE sections of OSCC samples. Scale bars depict 2 mm. High-resolution pictures are available through DOI: 10.6084/m9.figshare.20905837.v1.


Tumor cells
Fibroblasts
Pericytes
Muscle cells
ECs
Macrophages
cDCs
pDCs
RBCs
T cells
Regulatory T cells
Cytotoxic T cells
B cells
Plasmablasts
KRT17 KRT14 KRT6A KRT5 S100A2 SFN
DCN TAGLN FN1
RGS5 TIMP1
TTN DES ACTA1
VWF
FCER1G LYZ
FCER1A
LILRA4 GZMB IRF4
HBA1 HBA2 HBB
CD3D CD3E CD3G
CD4 FOXP3 CTLA4
CD8A GZMK GZMA CCL5
CD79A MS4A1
IGHD IGHM
IGHG1 IGHG3 JCHAIN
Percent Expressed
0
25
50
75
Average Expression
3
2
1
0
B
UMAP_2
UMAP_1
Central memory CD8 T cells
Classical monocytes
Effector memory CD8 T cells
Exhausted B cells
Follicular helper T cells
Intermediate monocytes
Myeloid dendritic cells
Naive B cells
Naive CD4 T cells
Naive CD8 T cells
Natural killer cells
Non-switched memory B cells
Non-Vd2 gd T cells
Plasmablasts
Plasmacytoid dendritic cells
Progenitor cells
Switched memory B cells
T regulatory cells
Th1 cells
Th1/Th17 cells
Th2 cells
Vd2 gd T cells
C
Tumor
cDCs
Fibroblasts
Macrophages
ECs
scale.data
D
log2
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22 chrX chrY
E
standard deviation from reference
non-Tumor
Tumor

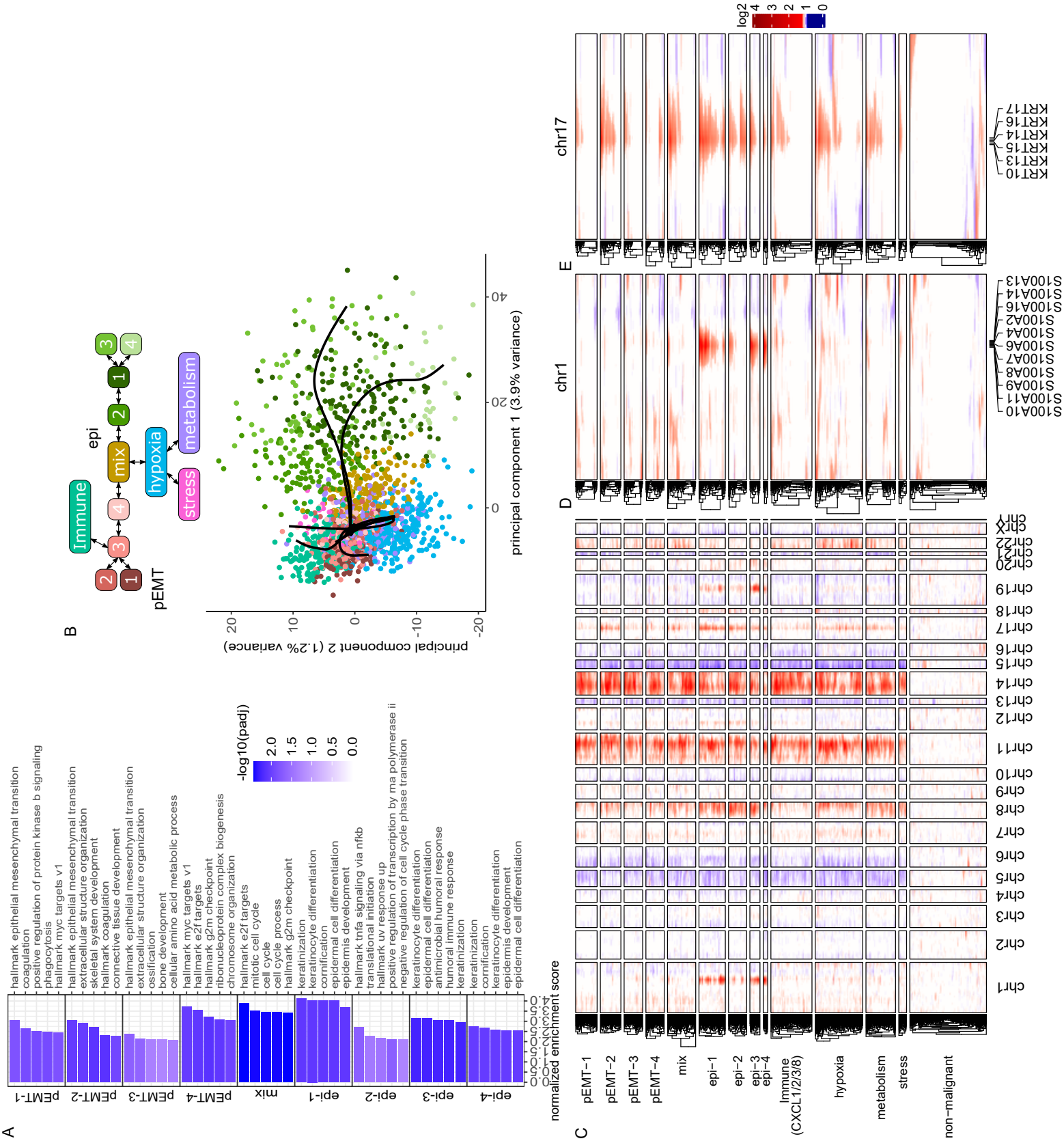
Supplementary Figure 2 | Cell type identification by marker genes, automated reference-based annotation, differential expression and inferred CNVs. (A) Expression of marker genes (x-axis) for each cell type (y-axis) in our cohort. Dots are colored by the average log-normalized gene expression and the dot size represents the percentage of cells with detected expression of the respective gene within the cell type. **(B)** UMAP of 41,284 cells from our cohort cells colored by SingleR annotations using the Monaco bulk RNA dataset on shared-nearest neighbor clusters with resolution 100. **(C)** Heatmap for scaled, log-normalized gene expression of all cell types from patient #1 and their top 10 DEGs (rows) against all other cells. DEGs are sorted from highest to lowest log2 foldchange. **(D)** Inferred CNVs across cells (rows) of different cell types without mitochondrial genes from patient #1. 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. **(E)** Standard deviation of the log2 inferCNV values to the mean of non-malignant cells compared between non-malignant and malignant cells of patient #1.

Supplementary Figure 3



Supplementary Figure 3 | PEMT and epithelial differentiating gene expression signatures are comparable to previously published EMT signatures. (A) EMT hallmark gene set enrichment plot for log2 fold changes of pEMT cells against all other cells of lymph node metastasis from patient 1. Shown is the stepwise calculated enrichment score, black lines indicate genes present in the respective gene set. **(B)** Average log2 fold change of gene expression (x-axis) and differences in cellular fractions expressing the respective gene (y-axis) between pEMT and epithelial differentiated cell clusters. Labelled in red are genes with log2 foldchange below or above 1 that are included in the epithelial differentiation or pEMT signature, respectively, with top 10 genes named. The histogram on top shows the number of genes across the log2 fold change with in total 100 bins. **(C, D)** Average expression scores (y-axis) of the pEMT (C) and epithelial differentiation (D) signatures across tumor phenotypes from patient #1 depicted in figure 2A (x-axis) color-coded by these clusters. **(E)** Heatmap of correlation coefficients of GSVA scores between 91 EMP-related signatures of malignant cells, derived from the EMTome database and selected publications (9, 10, 14). On the right side, the correlation coefficients between GSVA scores of EMT signatures from the EMTome database and of epithelial differentiation and pEMT signatures from patient #1 (right) are shown and next to it, annotated as ”EXPR_perc”, is the fraction of genes with non-zero expression and the size of the respective EMT signature in log10 scale with the respective number next to it. Rows and columns are hierarchically clustered using a spearman correlation distance (1-cor(x,y)) and ward.D2 method.

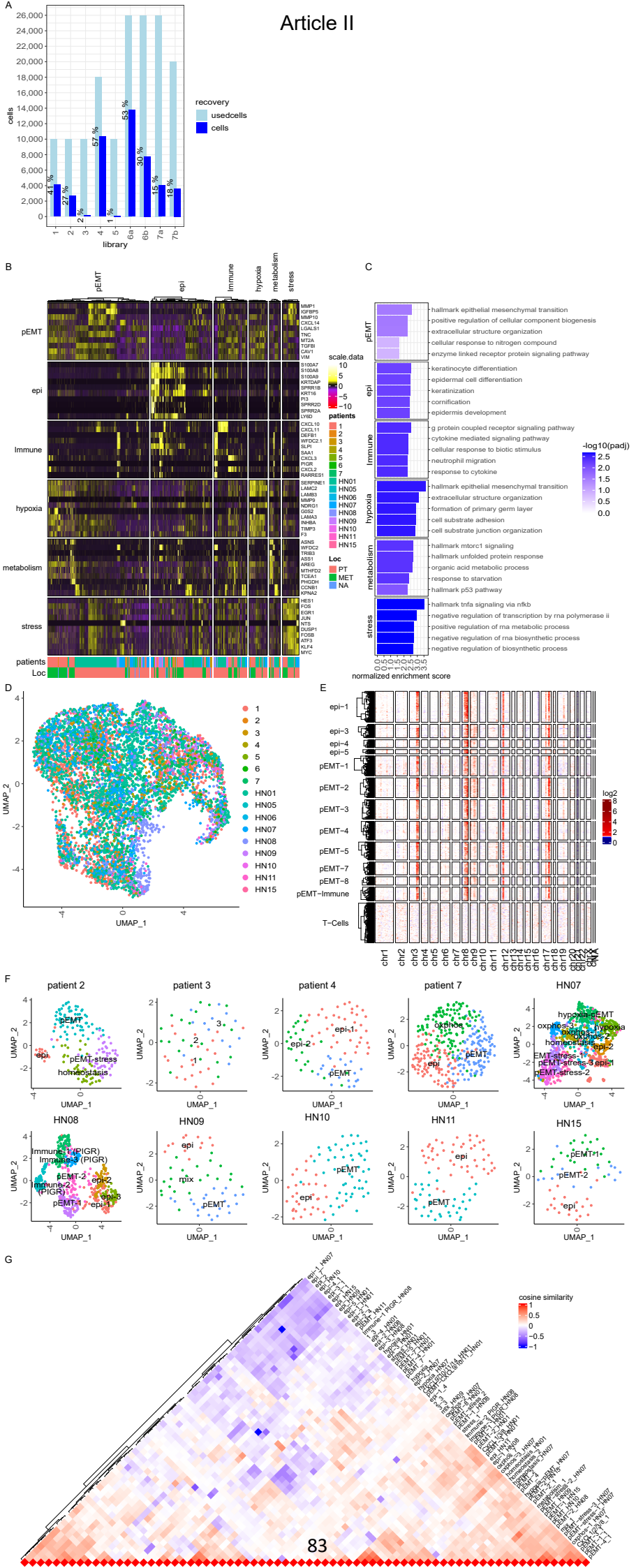
Supplementary Figure 4



Supplementary Figure 4 | Extended analysis of the tumor phenotype characterization for the lymph node metastasis of patient #1. (A) 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). **(B)** First two PCs of OSCC cells with all six principal curves that are derived from trajectory inference. Graph on top visualizes the relationship between malignant phenotypes, described by the principal curves forming a branching trajectory. Cells responding to environmental conditions form their own branch, indicating that the strong reactive response determines their predominant phenotype. **(C)** Inferred CNVs across tumor cells of patient #1 (rows) for all chromosomes (columns). 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. **(D, E)** Inferred CNVs across tumor cells (rows) of chromosome 1 (D) and chromosome 17 (E) showing genes (columns) ordered by genome position. The signal on chromosome 1 is located on a genomic position on which S100 genes are accumulating and the signal on chromosome 17 on a location with accumulation of cytokeratins; most of these genes are highly expressed in the more epithelial differentiated cells.

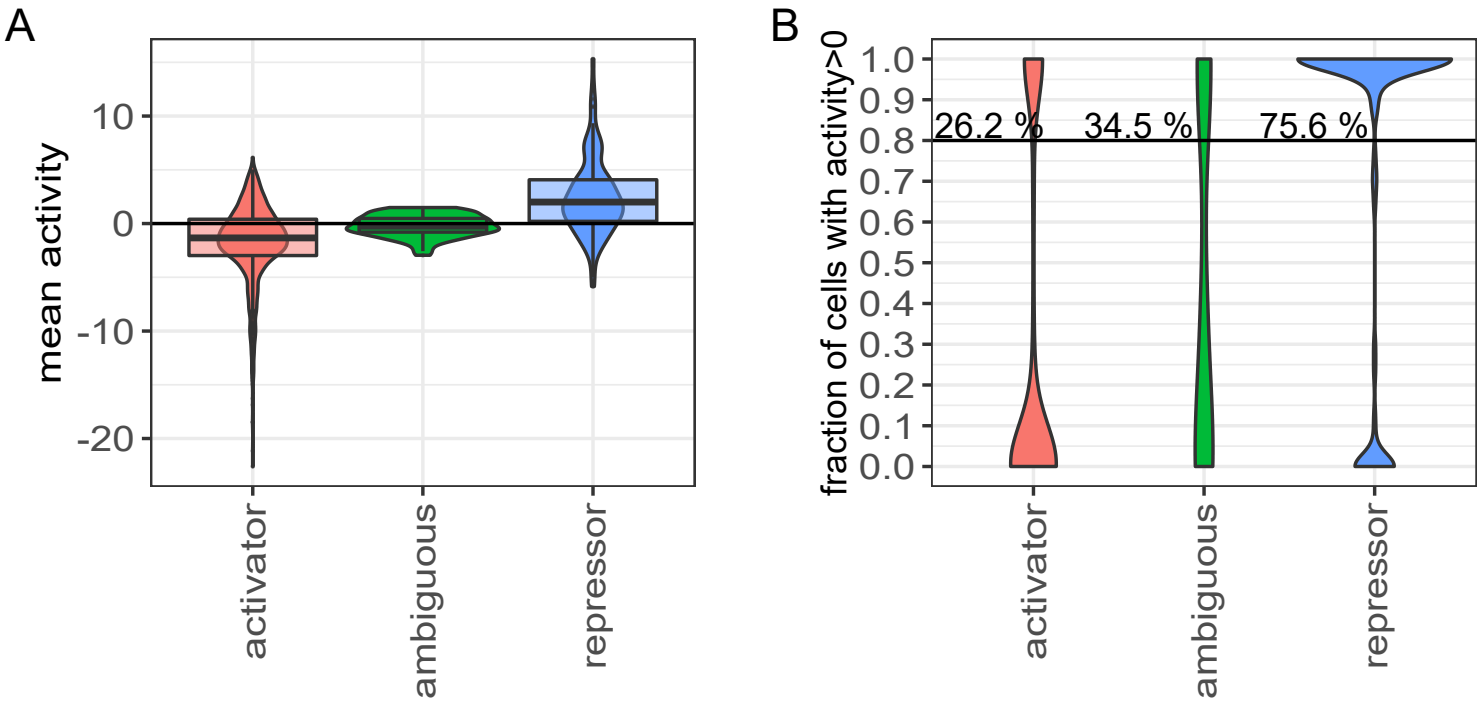
Supplementary Figure 5

Article II



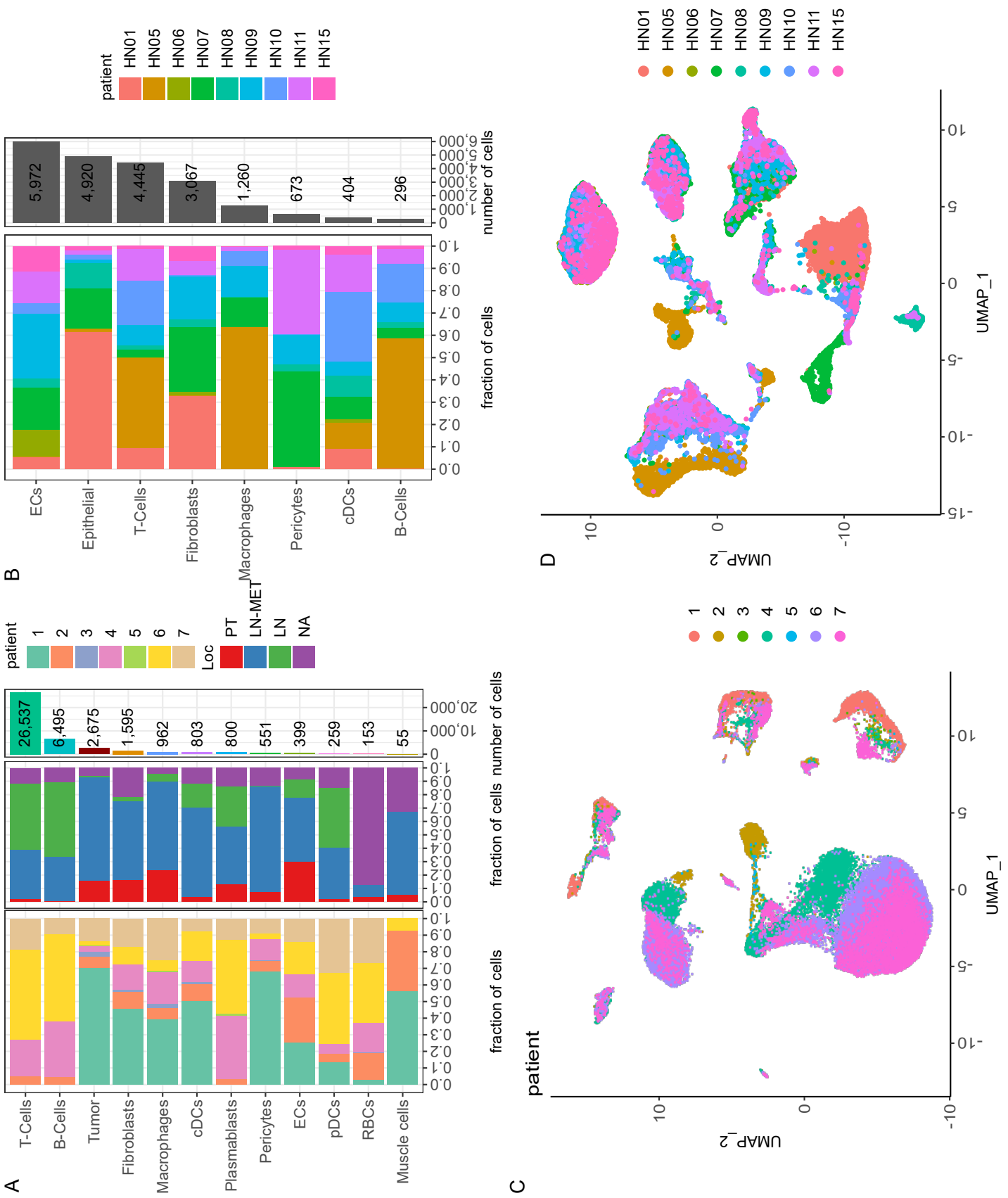
Supplementary Figure 5 | Malignant phenotypes characterized across all analyzed patients. (A) Number of cells (y-axis) for each library (x-axis) showing the cells that are used for 10x Genomics scRNAseq (light blue) and all recovered, i.e., detected, cells after sequencing (blue). Based on manufacturers information a recovery rate around 50% is expected. **(B)** Heatmap for scaled, log-normalized gene expression of tumor cells (columns) split by respective phenotype depicted in Figure 3C and the top 10 DEGs (rows) of the respective phenotype against all other tumor cells. DEGs are sorted from highest to lowest log2 foldchange and row sections are ordered the same as column section. On bottom, the respective patient and localization is annotated for each cell. **(C)** 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). **(D)** UMAP of OSCC cells as depicted in figure 3C with PCs corrected for patient-specific effects using harmony. Cells are annotated according to their patient id. **(E)** Inferred CNVs across EMP-related OSCC cells from patient HN01 (rows) for all chromosomes (columns). Cells split by their EMP phenotype do not show any differences in their inferred CNVs pattern. 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. **(F)** UMAPs of malignant cells from all respective patients. Cells are annotated SNN clusters and renamed according to the predominant phenotype. **(G)** Same plot as depicted in Figure 3F with the names of all patient-specific clusters as shown in E followed be the patient id.

Supplementary Figure 6



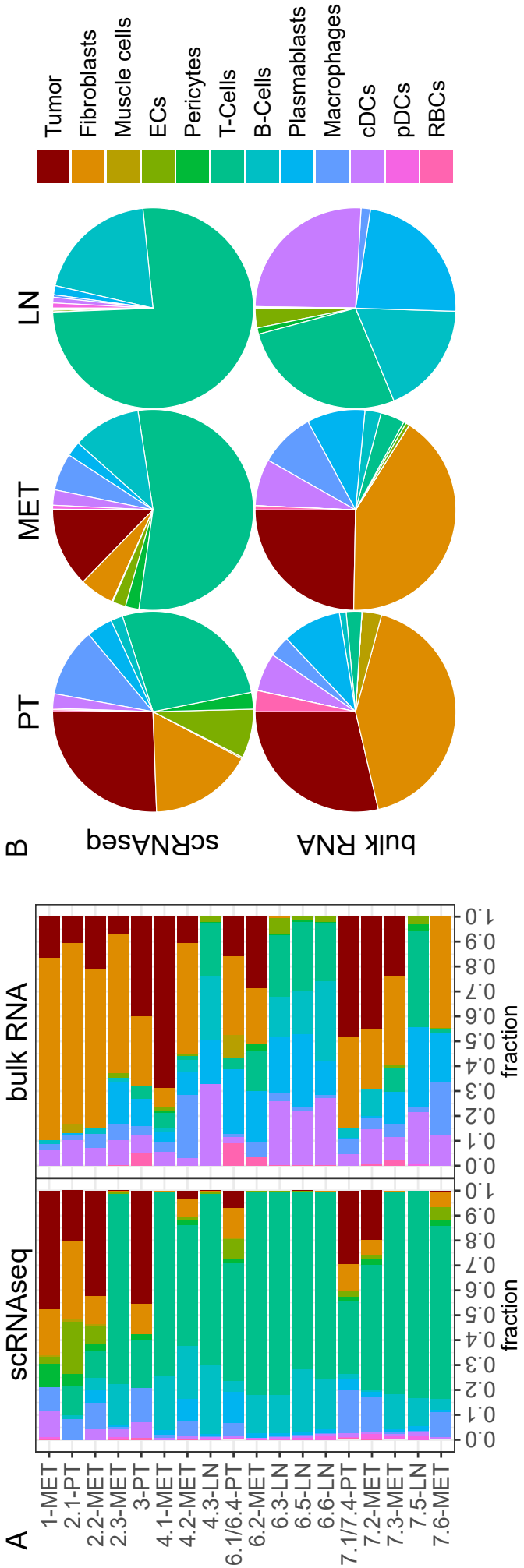
Supplementary Figure 6 | Inferred transcription factor activity might be biased by activator or repressor function. (A) Distribution of the mean activity of all cells from patient #1 for all transcription factors split by repressor, ambiguous and activators. Repressors and activators are defined based on more than 90% of the target genes being either repressed or upregulated, transcription factors with less than 90% for both are in the ambiguous class. **(B)** Distribution of the fraction of cells within a respective cell cluster with a transcription factor activity of greater than 0. The clusters include all cell types and malignant cell clusters from patient #1 split by activators, ambiguous and repressors. Clusters with high fraction of cells with activity greater than 0 indicate an active transcription factor, which is more prominent across repressors than for activators.

Supplementary Figure 7



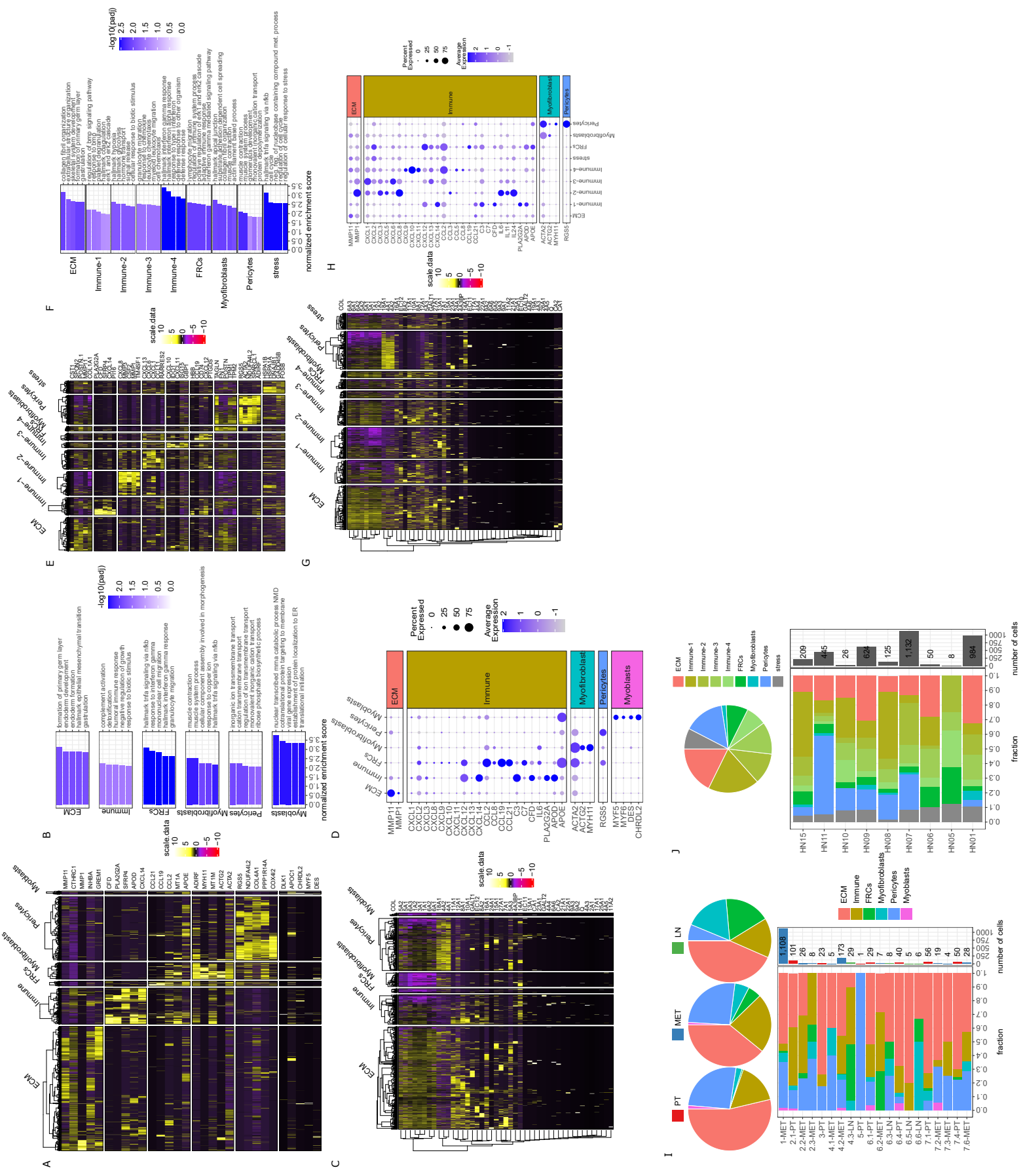
Supplementary Figure 7 | Cell type abundances across patients and tissues. (A) Relative fractions (x-axis) of cell types (y-axis) across different patients (left) or tissue types (middle) with the absolute number of cells per cell type (right), colored by cell types from Figure 5C. “NA” denotes cells that could not be demultiplexed from hashed samples and hence could not be assigned to a tissue type. **(B)** Relative fractions (x-axis) of cell types (y-axis) across different patients (left) with absolute numbers per cell type (right). **(C)** UMAP of 41,284 cells based on OSCC scRNAseq data from our cohort and colored by patients. **(D)** UMAP of 21,037 cells based on CD45-negative and HPV-negative primary HNSCC from Kürten *et al.* and colored by patients.

Supplementary Figure 8



Supplementary Figure 8 | Bulk transcriptomes reveal the cellular composition of OSCC across tissue types. (A) Fractions of cell types (x-axis) across all samples (y-axis) including primary tumors (PT), metastatic lymph nodes (MET) and tumor-free lymph nodes (LN) for cells detected by scRNAseq (left panel) or deconvoluted from bulk transcriptome analysis (right panel). **(B)** Pie charts showing the average fraction of cell types across samples from each tissue type, derived from scRNAseq data (top) and bulk transcriptome deconvolution (bottom) and colored by cell type. cDCs: conventional dendritic cells; pDCs: plasmacytoid dendritic cells; RBCs: red blood cells; ECs: endothelial cells.

Supplementary Figure 9



Supplementary Figure 9 | Characterization of OSCC-derived fibroblasts. (A) Heatmap of scaled, log-normalized expression of the top 5 differentially expressed genes (DEGs) (rows) for fibroblasts and pericytes (columns) split by their respective phenotype. DEGs are sorted from highest towards lowest log2 foldchange and row sections are ordered like column sections. **(B)** Normalized enrichment scores (NES) of top 5 enriched gene sets for each fibroblast phenotype. Gene sets are sorted from highest to lowest NES and the bar chart is colored by negative decadic logarithm of Benjamini-Hochberg adjusted p-values (padj). **(C)** Scaled, log-normalized expression of collagens (COL) (rows) across fibroblasts and pericytes split by respective phenotypes (columns). Rows are clustered by their similarity using the Euclidean distance and ward.D2 method. **(D)** Selected genes (y-axis) expressed across phenotypes (x-axis). Dots are colored by averaged log-normalized gene expression and dot size represents the percentage of cells expressed in this phenotype, i.e., cells with more than 1 unique molecular identifier (UMI) detected in the respective gene. **(E-F)** Analog to A-D for fibroblasts and pericytes from Kürten *et al.* dataset. **(I)** Composition of phenotypes across tissue types in pie charts (top) and across samples as bar chart (bottom). Pie charts show the average fraction of phenotypes across fibroblasts and pericytes for each tissue type. The bar chart shows the fraction of phenotypes (x-axis) across samples (y-axis) on the left with the absolute abundance of cells on the right side, colored by tissue type. **(J)** Similar plot as in I for the Kürten *et al.* dataset. As all samples represent primary tumors, they were summarized in one pie chart.

Supplementary Methods

Cell identification

For every sequenced library, we separately identified cells from the barcodes by evaluation of four quality criteria inspired by Luecken *et al.* [1] : (1) number of unique molecular identifiers (UMIs, nCount), (2) number of genes (nFeature), (3) percentage of mitochondrial gene expression (percent.mito) and (4) number of expressed housekeeper genes (n.exp.hkgenes), derived from Tirosh et al. without the mouse gene PRPS1L3 [2]. Accordingly, we chose manual filtering threshold for (1) and (2) s by first evaluating the histogram and combined view of both (first plot). Then, we apply the respective UMI or gene threshold and evaluate the criteria (3) and (4) by combined view (second plot). Lastly, we reevaluate each quality criteria and their combination as well as downstream analysis and adjust thresholds.

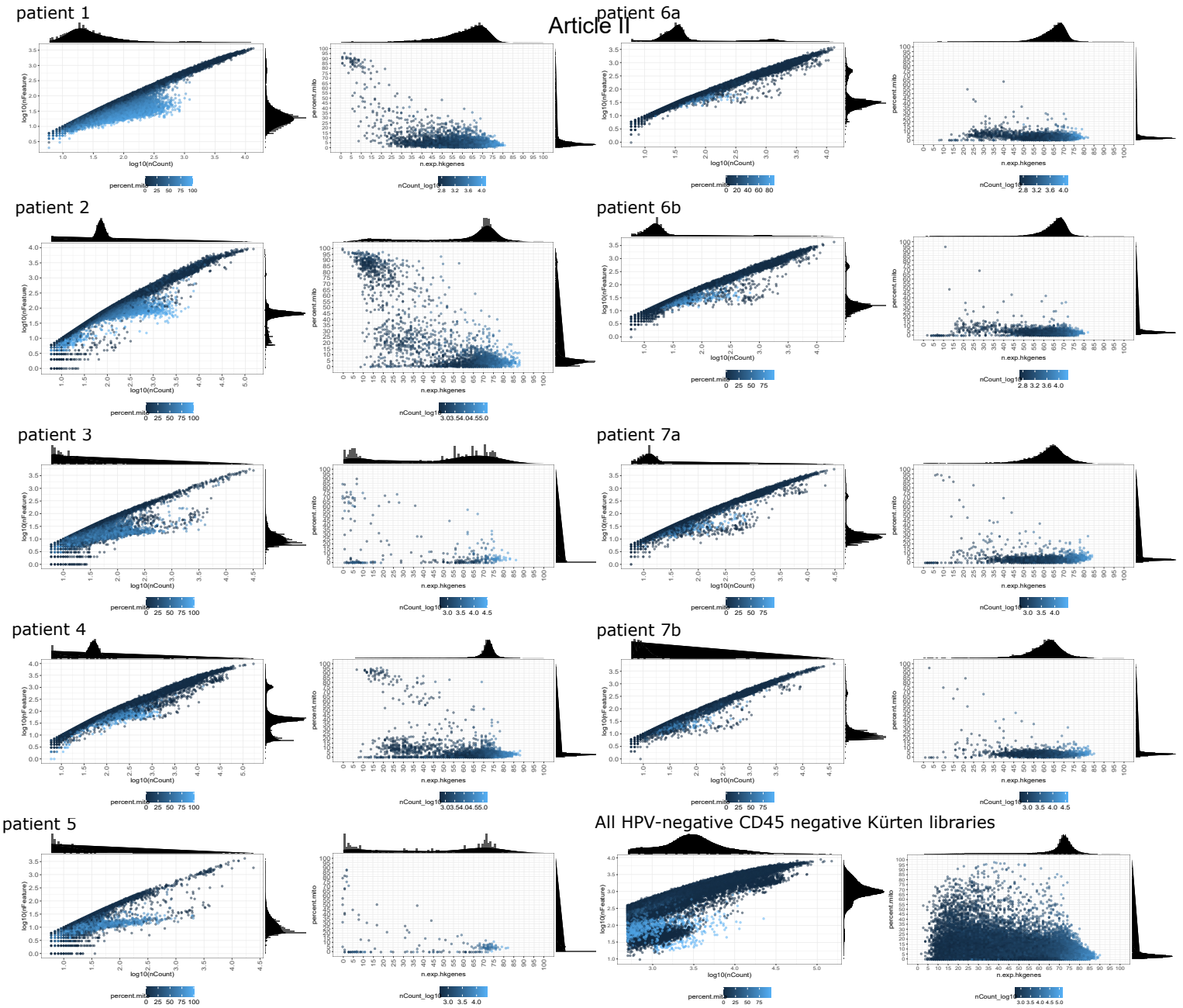


Figure 1 | Cell identification based on quality criteria evaluation. Quality criteria for all analyzed scRNAseq libraries. All barcodes with > 5 UMIs are shown as number of UMIs (nCount) against the number of genes (nFeature) colored by the percentage of mitochondrial gene expression with respective histograms (left plot). On the right side are all barcodes after applying the number of UMI or genes threshold showing the number of expressed housekeeper genes (maximum of 97 genes) against the percentage of mitochondrial gene expression colored by number of UMIs with respective histograms. The Kürten dataset was filtered collectively and not for every individual patient since it does not contain multiplexed hashed samples.

Hashing

For demultiplexing we first normalized the hashtag oligo (HTOs) expression matrix of all identified cells by centered log ratio transformation. Then, we chose manual thresholds of HTO expression for each HTO antibody and assigned samples, doublets, and cells without clear assignment. These thresholds were chosen based on combination of several criteria: We performed HTODemux from the Seurat R package once with default 99 % positive quantile for the fitted negative binomial distribution and once with the quantile greater than 50% that results the maximum numbers of singlets. The HTO expression was also visualized in scatter plots, histograms and UMAPs based on principal components to visualize the similarity of cells.

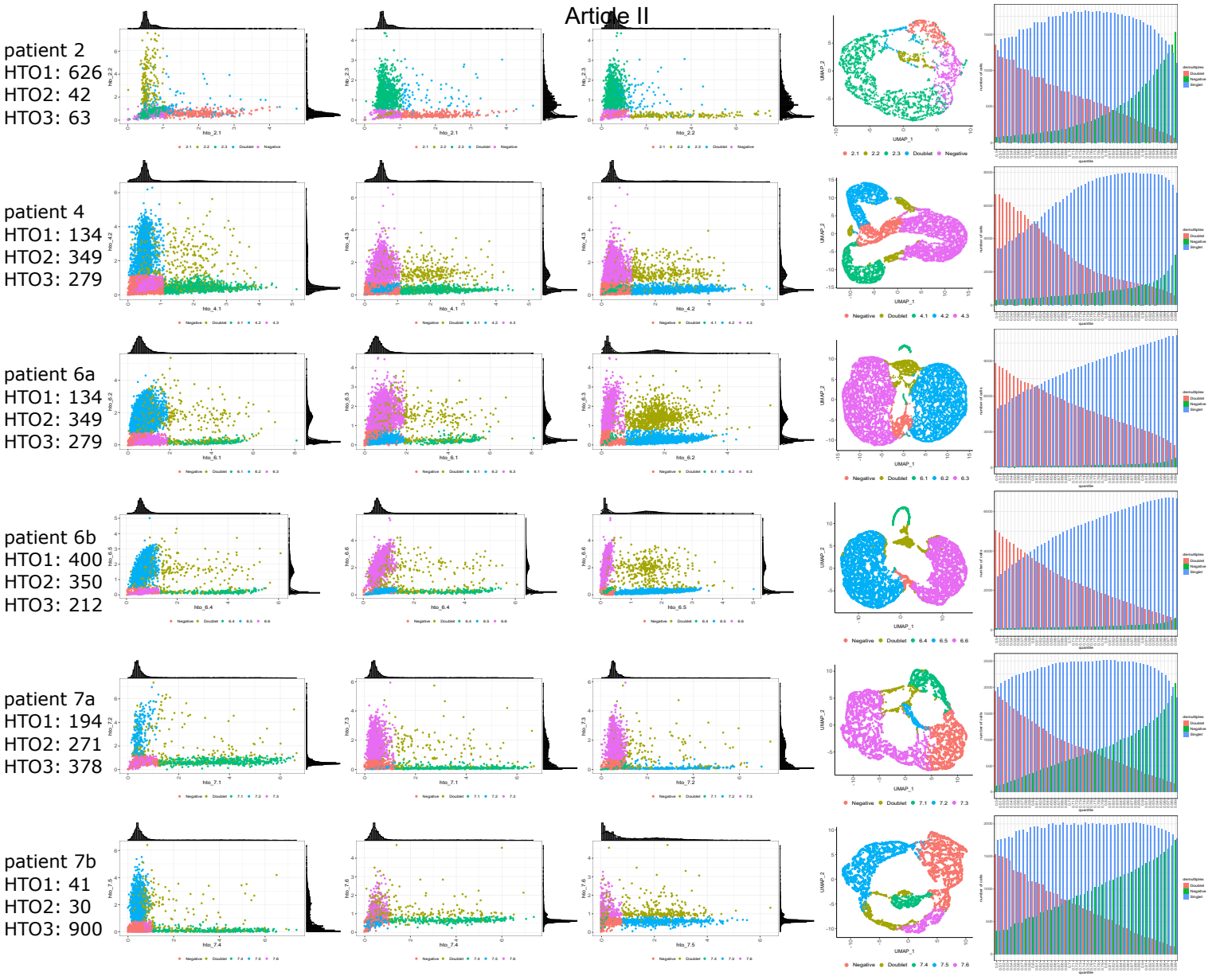


Figure 2 | Demultiplexing of the HTO expression matrix. The first three plots from the left show the centered log-ratio normalized expression of the three HTOs in all three different combinations. The color code shows the identified doublets, singlets and samples based on the manually determined HTO expression thresholds for demultiplexing. The fourth plot shows the UMAPs based on the principal components of the HTO expression matrix, clustering cells based on their HTO expression similarity. Color code correspond to the first three plots. The last plot on the right shows the number of cells (y-axis) identified as singlets, doublets, and negatives with different positive quantile parameter of HTODemux function from 0.5 to 0.99 (x-axis). All these information and plots were considered for choosing the manual HTO expression thresholds for demultiplexing the samples.

Note about trajectory and velocyto analysis:

The trajectory derived with Slingshot yields multiple curves that reflects a branching trajectory connecting the given clusters and expressed them as multiple linear curves. Also, a single sample represents just 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 ongoing evolutional processes within the sample. However, 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. Hence, RNA velocity can give insights into the near future developmental processes within the sample.

References

1 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.

2 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.

patient	sex	age at diagnosis	smoking in packyears	alcohol abuse	HPV status	localisation	type	pT	pN	pM	samples	localization per sample
1	m	82	60	yes	n.d.	bottom lip / mouth corner	PEC	Tx	N3b(4/29, ECE+)	M0	1	MET
2	m	66	50	yes	0%	anterior mouth base	PEC	T3	N0 (0/21)	M0	2.1;2.2;2.3	PT;MET;MET
3	m	67	80	no	0%	soft palate left	PEC	T3	N0(0/7)	M0	3	PT
4	w	70	0	no	0%	edge of tongue left	PEC	T2	N3b(3/48, ECE+)	M0	4.1;4.2;4.3	MET;MET;LN
5	w	84	0	no	0%	cheek left	PEC	T4a	N0(0/33)	M0	5	PT
6	m	37	40	yes	0%	tongue base left	PEC	T3	N1 (1/50, ECE-)	M0	6.1;6.2;6.3;6.4;6.5;6.6	PT;MET;LN;PT;LN;LN
7	m	64	80	yes	n.d.	mouth base left	PEC	T4a	N3b(4/40, ECE+)	M0	7.1;7.2;7.3;7.4;7.5;7.6	PT;MET;MET;PT;LN;MET

Total Seq per sample Hashing Antibody	percentage tumor by histology per sample	ZEB1 and SPRR1B co expression per sample	library per sample	cells used per library	recovered cells per library	10X technology
-	40	+	1	10.000	4.121	5' v2
C0251;C0252;C0253	30;40;30	o;o;+	2;2;2	10.000	2.703	5' v2
-	80	+	3	10.000	171	5' v2
A0253;A0251;A0252	60;50;0	+;o;n.d.	4	18.000	10.345	3' v2
-	60	o	5	10.000	106	5' v2
A0251;A0252;A0253;A0251;A0252;A0253	50;60;0;50;0;0	o;+;n.d.;as 6.4;n.d.;n.d.	a;a;a;b;b;b	a:26000;b:26000	a:13774;b:7783	3' v2
A0251;A0252;A0253;A0251;A0252;A0253	70;80;70;70;0;20	+;+;+;as 7.1;n.d.;+	a;a;a;b;b;b	a:26000;b:20000	a:4029;b:3643	3' v2

4 Discussion

4.1 Genomic diversity in Merkel cell carcinoma

This work verified the previously shown substantial differences in the mutational landscape between MCPyV- and UV-associated MCCs (Becker et al., 2017; Goh et al., 2015; Harms et al., 2015; Horny et al., 2021; Knepper et al., 2019; Starrett et al., 2017; Wong et al., 2015). Our study shows that these genomic differences are reflected in the respective MCC cell lines and are based on the underlying carcinogenesis: UV-light induced MCC cell lines have high numbers of SNVs, small insertions and deletions (InDels) and CNVs, from which SNVs show a characteristic UV-light induced DNA damage signature and include pathogenic variants in tumor-suppressor genes *TP53* and *RB1* (Horny et al., 2021). In contrast, MCC cell lines with an integrated MCPyV genome have low numbers of genomic mutations but characteristic CNV patterns.

Both MCC types are distinguished by these genomic differences, as the presence of integrated MCPyV genome and UV-light induced DNA damage signatures are mutually exclusive (González-Vela et al., 2017; Knepper et al., 2019; Starrett et al., 2020). However, additional genomic heterogeneity within the two MCC types exists: While *TP53* and *RB1* abrogation is common in UV-associated MCCs, genes such as *PTEN*, *KMT2D*, *PIK3CA*, *KIT*, *NOTCH* family genes, *MYC* and MCPyV genes have pathogenic variants within smaller fractions of MCC samples (Goh et al., 2015; González-Vela et al., 2017; Harms et al., 2015; Knepper et al., 2019; Starrett et al., 2020; Veija et al., 2016; Wong et al., 2015). Furthermore, a recent study proposes three MCC subtypes based on their copy number patterns associated with amplification and loss of cancer-related genes (Hill et al., 2021). Also, MCC samples show a wide range of TMB levels, including samples with intermediate TMB levels for both MCC types (Goh et al., 2015; Harms et al., 2015; Knepper et al., 2019; Starrett et al., 2020; Wong et al., 2015). In general, the TMB is a continuous variable that is often binarized into high and low levels using variable thresholds. This simplifies distinction of MCC types but causes a loss of information about their genomic heterogeneity, especially in cases with intermediate TMB levels close to the used threshold.

In this work, we found a focal amplification and increased expression of *c-MYC* in the UV-associated cell line UM-MCC34 that likely represents a genomic subtype of MCC (Horny et al., 2021). Amplification of *MYC* family genes was previously reported in MCC, especially for *L-MYC* (Knepper et al., 2019; Paulson et al., 2009). The initial report by Paulson *et al.* reported 9/23 tumors (39 %) with amplified *L-MYC*, while more recent studies report lesser frequencies predominantly in UV-associated MCC: 12/117 (10 %) tumors classified as “TMB-high” by

Knepper *et al.* have either amplified *c-MYC* or *L-MYC* and Carter *et al.* reported 1/28 (3.6 %) tumors classified as MCPyV-negative as affected (Carter et al., 2018; Knepper et al., 2019; Paulson et al., 2009). *MYC* aberrations are common in small cell lung cancer which also undergoes neuroendocrine transformation, has abrogation of *TP53* and *RB1* function and is therefore similar to MCC (Becker et al., 2017; Gazdar et al., 2017). Amplification of *c-MYC* and *L-MYC* are shown to be important determinants of molecular subtypes during small cell lung cancer development: *c-MYC* is associated with a less neuroendocrine *NEUROD1*+ subtype and *L-MYC* with a more neuroendocrine subtype driven by *ASCL1*+ (Ireland et al., 2020; Mollaoglu et al., 2017; Patel et al., 2021). Furthermore, *C-MYC*-related transcriptional programs have higher expression of genes involved in NOTCH-signaling and EMT, hence are likely to affect metastatic spread (Patel et al., 2021). *C-MYC* also activates NOTCH to drive the conversion from *ASCL1*+ to *NEUROD1*+ subtypes, indicating NOTCH mutations could be relevant also for MCC (Ireland et al., 2020). Translating these findings to MCC indicates that *MYC* status and mutations in NOTCH signaling may characterize genomic subtypes of MCC with possible implications for tumor progression and metastatic spread.

A multitude of intrinsic and extrinsic factors controls the evolutionary processes that results in the genomic heterogeneity of cancer, hence, several other, rarer genomic subtypes might exist in MCC that are not yet detected (Stratton et al., 2009). One of those genomic subtypes could be related to upregulated APOBEC expression which induces regional accumulation of genomic mutations in so-called hotspots and are commonly found in virally induced cancer (Roberts et al., 2013). Within a targeted panel sequencing study, Starrett *et al.* found one sample with an APOBEC-related mutational signature indicating the possible existence of such a rare subtype (Starrett et al., 2020). However, if only few APOBEC-hotspots are present, for example due to a late onset of APOBEC activity, these may not be detected in MCC samples due of the low TMB in MCPyV-positive samples and Starrett *et al.* state this could be counteracted by additional whole-genome sequencing studies (Starrett et al., 2017). Moreover, improvements in bioinformatic methods to detect mutational signatures specifically aim to derive signatures within mutational hotspots and thus might increase sensitivity for detecting APOBEC activity in MCC (Manders et al., 2022; Maura et al., 2019). Another example for an MCC subtype is the co-existence of UV-associated MCCs with *in-situ* SCC, both having large genomic similarity and thus potentially originating from the same cell type (Carter et al., 2018; DeCoste et al., 2022; Kervarrec et al., 2022). Also contributing to the genomic heterogeneity, a recent MCC study found 9/56 patients with low levels of defective DNA mismatch repair from which one is evidenced of having a high microsatellite instability (Gambichler et al., 2021).

4.2 Heterogeneity of EMP states in metastatic OSCC

EMP involves a variety of dynamic processes in which multiple intermediate or partial states could exist (Pal et al., 2021; Yang et al., 2020). The simple distinction between one epithelial, one mesenchymal and one partial EMT state likely underrepresents the complexity of the EMP continuum (Pal et al., 2021). By using single cell transcriptomes, this work gave insights into EMP-related heterogeneity in OSCC cells and revealed another intermediate EMP state with more epithelial characteristics compared to the partial EMT state (Horny et al., 2023). This phenotype is characterized by the activity of the EMT-related transcription factor ZEB1 which is co-expressed on protein level with the keratinocyte-envelope protein cornifin B. Interestingly, previous OSCC studies have not reported the association of EMT with ZEB1 as these focused on mRNA expression and neglected transcription factor activity or protein abundance (Kürten et al., 2021; Puram et al., 2017). Due to the dropout of lowly expressed transcription factors in scRNAseq, deriving the transcription factor activity may be more accurate compared to mRNA expression levels (Kharchenko et al., 2014). However, the transcription factor activity might also be overestimated for transcriptional repressors, likely due to high dropout rates in scRNAseq and lesser accuracy with lower gene coverage (Holland et al., 2020; Horny et al., 2023).

The in this work intensively analyzed metastasis developed one year after removal of the primary tumor and has a completely disrupted lymph node structure, without any T and B cells detected in scRNAseq (Horny et al., 2023). Within this period, the metastatic lesion either maintained the original phenotypic heterogeneity from diverse founder cells or newly developed it. Thus, the existence of heterogenous EMP phenotypes within a lymph node metastasis can be explained by different hypotheses about the metastatic spread.

One hypothesis is that the founder cell clones had a partial EMT phenotype with high plasticity, enabling differentiation into multiple phenotypes. Indeed, we found partial EMT signatures present in several cell populations which show transcriptional patterns associated with angiogenesis, ECM- and immunomodulation and response to hypoxia, low glucose, and stress (Horny et al., 2023). In addition, partial EMT cells may promote metastatic spread: a previous study in the breast cancer mouse mammary tumor virus-polyoma middle tumor-antigen (MMTV-PyMT) mouse model concluded that metastatic spread is more likely if epithelial tumor cells developed into a partial EMT state rather than a full mesenchymal phenotype (Lüönd et al., 2021). In fact, a full EMT may disable metastatic outgrowth as observed in prostate and bladder cancer models (Celià-Terrassa et al., 2012). Similarly, the identified ZEB1^{+}/SPRR1B^{+} cells may have originated from partial EMT cells and developed into a partial MET rather than a full epithelial phenotype. Supporting this hypothesis, we observed (I) simultaneous expression of epithelial and mesenchymal genes such as *MMP1* and *KRT16* as well as *ZEB1*

and *SPRR1B*, (II) development of ZEB1^{+}/SPRR1B^{+} cells on RNA velocity trajectories directed towards epithelial differentiated cell populations within a metastatic lesion, and (III) lower inferred CNV burden in partial EMT cells compared to ZEB1^{+}/SPRR1B^{+} cells, suggesting an earlier existence of partial EMT cells (Horny et al., 2023). The active transcriptional repression of epithelial genes by ZEB1 might hamper progression towards a full MET, but may not prevent a partial MET (Sánchez-Tilló et al., 2010). A partial MET could also be achieved through other mechanisms similar to a previously found protein internalization mechanism for EMT (Aiello et al., 2018).

In contrast, ZEB1^{+}/SPRR1B^{+} cells might have developed from epithelial tumor cells and reflect another intermediate EMP state with more epithelial traits than previously described partial EMT states. The existence of several intermediate EMP states with different stability was previously proposed (Jolly et al., 2017; Pal et al., 2021). These could have developed upon EMT induction either in the primary tumor and migrated to the metastatic site or by reinduced EMT of migrated epithelial cells in the metastatic lesion. Indeed, we detected ZEB1^{+}/SPRR1B^{+} cells within primary tumor samples, which supports this hypothesis (Horny et al., 2023). Furthermore, previous studies indicate partial EMT cells preferably travel in collective, heterogenous clusters that can also contain more epithelial cells expressing E-cadherin (Aiello et al., 2018). These clusters might also include fibroblasts, as this work found similar fibroblast populations in primary and metastatic lesions. Thus, multiple founder cells with varying degrees of EMP from the primary tumor might cause cellular heterogeneity in metastatic lesions.

Both scenarios may also have occurred simultaneously: first, a multitude of intermediate EMP states may had formed in the primary tumor and disseminated to the metastatic site with multiple founder cells. These might then inherit the phenotypic plasticity enabling further development into other phenotypes, potentially including the ZEB1^{+}/SPPR1B^{+} or other epithelial differentiating or partial EMP states, possibly as a response to the new tissue environment. The exact developmental trajectory and origin of intermediate EMP states might also be heterogenous between OSCC patients, which could explain the large transcriptional heterogeneity between malignant cells of different patients (Horny et al., 2023). Unfortunately, this work could not investigate the developmental trajectories from primary to metastatic lesions due to low scRNAseq recovery of malignant cells that may have shed light on the origin of these phenotypes (Horny et al., 2023). ScRNAseq also does not exhaustively captures all cells and only represent a single snapshot in tumor evolution, hence possibly leaving more unstable EMP states left to uncover (Pal et al., 2021). The same may be true for CAFs that have large transcriptional similarity with partial EMT cells which could additionally contribute

to underestimate the abundance of malignant partial EMT cells in bulk transcriptome cell type deconvolution approaches (Tyler & Tirosh, 2021).

The transcriptional heterogeneity in OSCC is also present in the TME for which we found diverse phenotypes with potential effects on tumor progression and immune escape (Horny et al., 2023). Interestingly, we found similar CAF phenotypes in primary and metastatic lesions which indicate a similar origin (Horny et al., 2023). CAFs could possibly be created by polarization of tissue-resident cells, single or collective dissemination from a malignant seeding site or originate from bone marrow tissue (Kalluri, 2016; Lambert et al., 2017).

4.3 Conclusions and outlook

This work gives insights into the genomic and transcriptional heterogeneity within malignant and microenvironmental cells. It shows that MCC cell lines are suitable genomic models of UV-light and MCPyV-induced MCC. This is important, as for example previous studies frequently used variant MCC cell lines that do not display neuroendocrine growth patterns and are epigenetically more similar to SCC than classical MCC cell lines (Gravemeyer et al., 2021). Additionally, the cell line UM-MCC34 represents a genomic MCC subtype with an amplified and overexpressed *c-MYC* gene that can serve as a model to investigate the effect of *MYC* aberrations in MCC.

In the future, other genomic MCC subtypes might be revealed by capturing non-exonic regions, as most studies focused on targeted or exome sequencing experiments (Goh et al., 2015; Harms et al., 2015; Knepper et al., 2019; Starrett et al., 2020; Wong et al., 2015). Many genomic subtypes might be rare, and their detection could be facilitated in larger sample cohorts which are hard to acquire due to the low incidence of MCC. Uncovering these genomic subtypes would be the first step: Subsequent studies need to elucidate the functional consequences of these subtypes and their impact on tumor progression and clinical outcome. These insights in the genomic heterogeneity are of great interest for personalized medicine approaches that will also apply to rare cancers such as MCC (Cohen et al., 2016; Knepper et al., 2021). Especially, predicting the response to immune checkpoint therapy is a major challenge. Current studies already evaluated potential predictive markers such as clinical characteristics and some genomic properties such as TMB and specific mutations (Long et al., 2022; Spassova et al., 2022; Yarchoan et al., 2017).

Additionally, this work investigated the EMP-related heterogeneity within primary and metastatic OSCC lesions and discovered an intermediate EMP state associated with ZEB1 and cornifin B expression (Horny et al., 2023). These results indicate a role of ZEB1 beyond EMT induction possibly in mediating partial MET. Finally, we found similar CAF phenotypes in

primary and metastatic lesions indicating their important function for tumor progression and immune escape mechanisms.

The discovery of different EMP phenotypes might contribute to the derivation of biomarkers for predicting the metastatic risk of OSCC patients. Previous studies showed that mesenchymal and partial EMT gene expression is associated with increased metastatic risk, decreased progression-free intervals and worse prognosis (Cook & Vanderhyden, 2022; Jung et al., 2020; Parikh et al., 2019; Puram et al., 2017; Schinke et al., 2022). The heterogenous TME can also influence the prognosis, for example a 4-genes based signature derived from CAFs correlated with poor survival in HNSCC (Yang et al., 2022). A current study derived specific CAF markers from scRNAseq data that can be utilized to study the metastatic risk more in detail (Kazakova et al., 2022). This doctoral work also indicates that partial EMT is linked with tumor-supportive properties, as we found gene expression patterns in partial EMT cells related to Immune evasion, angiogenesis and response to hypoxia, hypoglycemia, and stress (Horny et al., 2023). Earlier studies showed that the involvement of EMT in immunosuppression is linked to metastases (Dongre et al., 2017; Kudo-Saito et al., 2009). Immune evasion and cell motility is further influenced by fibroblast subpopulations which could also have an impact on prognosis (Horny et al., 2023). Factors inducing EMP changes in malignant cells might pose future therapeutic targets to prevent development of metastases (Huang et al., 2022; Jehanno et al., 2022). However, treatment-induced EMT is also known to promote tumor progression and worsen the clinical outcome (Redfern et al., 2018).

Hence, future research should comprehensively investigate the complex network determining tumor heterogeneity by integrating multiple levels of genomic, epigenetic, and spatial information as well as sampling from primary and metastatic lesions of the same patient. For example, integrating single-cell transcriptomic and genomic data can give hints about the ancestry of different phenotypes (Campbell et al., 2019). Recent advances in high-resolution spatial transcriptomics can capture additional information and will give insights into local interactions and signaling between malignant and their surrounding cells (Palla et al., 2022). For analyzing the spatial information with the combined information from multiple data layers, enhanced bioinformatic methods will be needed (Longo et al., 2021; Velten et al., 2022). The gained knowledge from these approaches could contribute to understand carcinogenesis, tumor progression and locoregional invasion and ultimately guide the derivation of predictive biomarkers supporting clinical therapy decision making.

5 References

Aiello, N. M., Maddipati, R., Norgard, R. J., Balli, D., Li, J., Yuan, S., Yamazoe, T., Black, T., Sahmoud, A., Furth, E. E., Bar-Sagi, D., & Stanger, B. Z. (2018). EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Developmental Cell*, *45*(6), 681-695.e684. https://doi.org/10.1016/j.devcel.2018.05.027

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., Alexandrov, L. B., et al. Consortium, P. (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), 94-101. https://doi.org/10.1038/s41586-020-1943-3

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., et al. PedBrain, I. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415-421. https://doi.org/10.1038/nature12477

Anderson, N. M., & Simon, M. C. (2020). The tumor microenvironment. *Current Biology*, *30*(16), R921-r925. https://doi.org/10.1016/j.cub.2020.06.081

Bartoschek, M., Oskolkov, N., Bocci, M., Lövrot, J., Larsson, C., Sommarin, M., Madsen, C. D., Lindgren, D., Pekar, G., Karlsson, G., Ringnér, M., Bergh, J., Björklund, Å., & Pietras, K. (2018). Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nature Communications*, *9*(1), 5150. https://doi.org/10.1038/s41467-018-07582-3

Batlle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J., & García de Herreros, A. (2000). The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nature Cell Biology*, *2*(2), 84-89. https://doi.org/10.1038/35000034

Becker, J. C., Stang, A., DeCaprio, J. A., Cerroni, L., Lebbé, C., Veness, M., & Nghiem, P. (2017). Merkel cell carcinoma. *Nature Reviews Disease Primers*, *3*(1), 17077. https://doi.org/10.1038/nrdp.2017.77

Becker, J. C., Ugurel, S., Leiter-Stoppke, U., Meier, F., Gutzmer, R., Haferkamp, S., Zimmer, L., Livingstone, E., Eigentler, T., Hauschild, A., Kiecker, F., Hassel, J. C., Mohr, P., Fluck, M., Thomas, I., Garzarolli, M., Grimmelmann, I., Drexler, K., Eckhardt, S., & Schadendorf, D. (2022). 787O Adjuvant immunotherapy with nivolumab (NIVO) versus observation in completely resected Merkel cell carcinoma (MCC): Disease-free survival (DFS) results from ADMEC-O, a randomized, open-label phase II trial. *Annals of Oncology*, *33*, S903. https://doi.org/10.1016/j.annonc.2022.07.913

Berrios, C., Padi, M., Keibler, M. A., Park, D. E., Molla, V., Cheng, J., Lee, S. M., Stephanopoulos, G., Quackenbush, J., & DeCaprio, J. A. (2016). Merkel Cell Polyomavirus Small T Antigen Promotes Pro-Glycolytic Metabolic Perturbations Required for Transformation. *PLOS Pathogens*, *12*(11), e1006020. https://doi.org/10.1371/journal.ppat.1006020

Bichakjian, C. K., Olencki, T., Aasi, S. Z., Alam, M., Andersen, J. S., Blitzblau, R., Bowen, G. M., Contreras, C. M., Daniels, G. A., Decker, R., Farma, J. M., Fisher, K., Gastman, B., Ghosh, K., Grekin, R. C., Grossman, K., Ho, A. L., Lewis, K. D., Loss, M., et al. Engh, A. M. (2018). Merkel Cell Carcinoma, Version 1.2018, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network*, *16*(6), 742-774. https://doi.org/10.6004/jnccn.2018.0055

Biffi, G., & Tuveson, D. A. (2021). Diversity and Biology of Cancer-Associated Fibroblasts. *Physiological Reviews*, *101*(1), 147-176. https://doi.org/10.1152/physrev.00048.2019

Brabletz, S., Schuhwerk, H., Brabletz, T., & Stemmler, M. P. (2021). Dynamic EMT: a multi-tool for tumor progression. *The EMBO Journal*, *40*(18), e108647. https://doi.org/10.15252/embj.2021108647

Campbell, K. R., Steif, A., Laks, E., Zahn, H., Lai, D., McPherson, A., Farahani, H., Kabeer, F., O'Flanagan, C., Biele, J., Brimhall, J., Wang, B., Walters, P., Consortium, I., Bouchard-Côté, A., Aparicio, S., & Shah, S. P. (2019). clonealign: statistical integration of independent single-cell RNA and DNA sequencing data from human cancers. *Genome Biology*, *20*(1), 54. https://doi.org/10.1186/s13059-019-1645-z

Cano, A., Pérez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F., & Nieto, M. A. (2000). The transcription factor Snail controls epithelial–mesenchymal transitions by repressing E-cadherin expression. *Nature Cell Biology*, *2*(2), 76-83. https://doi.org/10.1038/35000025

Carter, M. D., Gaston, D., Huang, W.-Y., Greer, W. L., Pasternak, S., Ly, T. Y., & Walsh, N. M. (2018). Genetic profiles of different subsets of Merkel cell carcinoma show links between combined and pure MCPyV-negative tumors. *Human Pathology*, *71*, 117-125. https://doi.org/10.1016/j.humpath.2017.10.014

Celià-Terrassa, T., Meca-Cortés, Ó., Mateo, F., Martínez de Paz, A., Rubio, N., Arnal-Estapé, A., Ell, B. J., Bermudo, R., Díaz, A., Guerra-Rebollo, M., Lozano, J. J., Estarás, C., Ulloa, C., ρlvarez-Simón, D., Milà, J., Vilella, R., Paciucci, R., Martínez-Balbás, M., García de Herreros, A., et al. Thomson, T. M. (2012). Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *The Journal of Clinical Investigation*, *122*(5), 1849-1868. https://doi.org/10.1172/JCI59218

Chan, I. S., Bhatia, S., Kaufman, H. L., & Lipson, E. J. (2018). Immunotherapy for Merkel cell carcinoma: a turning point in patient care. *Journal for ImmunoTherapy of Cancer*, *6*(1), 23. https://doi.org/10.1186/s40425-018-0335-9

Chaturvedi, A. K., Engels, E. A., Pfeiffer, R. M., Hernandez, B. Y., Xiao, W., Kim, E., Jiang, B., Goodman, M. T., Sibug-Saber, M., Cozen, W., Liu, L., Lynch, C. F., Wentzensen, N., Jordan, R. C., Altekruse, S., Anderson, W. F., Rosenberg, P. S., & Gillison, M. L. (2011). Human Papillomavirus and Rising Oropharyngeal Cancer Incidence in the United States. *Journal of Clinical Oncology*, *29*(32), 4294-4301. https://doi.org/10.1200/jco.2011.36.4596

Chen, Y.-P., Yin, J.-H., Li, W.-F., Li, H.-J., Chen, D.-P., Zhang, C.-J., Lv, J.-W., Wang, Y.-Q., Li, X.-M., Li, J.-Y., Zhang, P.-P., Li, Y.-Q., He, Q.-M., Yang, X.-J., Lei, Y., Tang, L.-L., Zhou, G.-Q., Mao, Y.-P., Wei, C., et al. Ma, J. (2020). Single-cell transcriptomics reveals regulators underlying immune cell diversity and immune subtypes associated with prognosis in nasopharyngeal carcinoma. *Cell Research*, *30*(11), 1024-1042. https://doi.org/10.1038/s41422-020-0374-x

Cheng, J., Park, D. E., Berrios, C., White, E. A., Arora, R., Yoon, R., Branigan, T., Xiao, T., Westerling, T., Federation, A., Zeid, R., Strober, B., Swanson, S. K., Florens, L., Bradner, J. E., Brown, M., Howley, P. M., Padi, M., Washburn, M. P., & DeCaprio, J. A. (2017). Merkel cell polyomavirus recruits MYCL to the EP400 complex to promote oncogenesis. *PLOS Pathogens*, *13*(10), e1006668. https://doi.org/10.1371/journal.ppat.1006668

Cheng, J., Rozenblatt-Rosen, O., Paulson, K. G., Nghiem, P., & DeCaprio, J. A. (2013). Merkel cell polyomavirus large T antigen has growth-promoting and inhibitory activities. *Journal of Virology*, *87*(11), 6118-6126. https://doi.org/10.1128/jvi.00385-13

Cheung, K. J., Padmanaban, V., Silvestri, V., Schipper, K., Cohen, J. D., Fairchild, A. N., Gorin, M. A., Verdone, J. E., Pienta, K. J., Bader, J. S., & Ewald, A. J. (2016). Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proceedings of the National Academy of Sciences*, *113*(7), E854-863. https://doi.org/10.1073/pnas.1508541113

Christensen, E., Naidas, A., Chen, D., Husic, M., & Shooshtari, P. (2022). TMExplorer: A tumour microenvironment single-cell RNAseq database and search tool. *PLOS ONE*, *17*(9), e0272302. https://doi.org/10.1371/journal.pone.0272302

Cohen, P. R., Tomson, B. N., Elkin, S. K., Marchlik, E., Carter, J. L., & Kurzrock, R. (2016). Genomic portfolio of Merkel cell carcinoma as determined by comprehensive genomic profiling: implications for targeted therapeutics. *Oncotarget*, *7*(17). https://doi.org/10.18632/oncotarget.8032

Cook, D. P., & Vanderhyden, B. C. (2020). Context specificity of the EMT transcriptional response. *Nature Communications*, *11*(1), 2142. https://doi.org/10.1038/s41467-020-16066-2

Cook, D. P., & Vanderhyden, B. C. (2022). Transcriptional census of epithelial-mesenchymal plasticity in cancer. *Science Advances*, *8*(1), eabi7640. https://doi.org/10.1126/sciadv.abi7640

Costa, A., Kieffer, Y., Scholer-Dahirel, A., Pelon, F., Bourachot, B., Cardon, M., Sirven, P., Magagna, I., Fuhrmann, L., Bernard, C., Bonneau, C., Kondratova, M., Kuperstein, I., Zinovyev, A., Givel, A.-M., Parrini, M.-C., Soumelis, V., Vincent-Salomon, A., & Mechta-Grigoriou, F. (2018). Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *Cancer Cell*, *33*(3), 463-479.e410. https://doi.org/10.1016/j.ccell.2018.01.011

Cramer, J. D., Burtness, B., Le, Q. T., & Ferris, R. L. (2019). The changing therapeutic landscape of head and neck cancer. *Nature Reviews Clinical Oncology*, *16*(11), 669-683. https://doi.org/10.1038/s41571-019-0227-z

D'Cruz, A. K., Vaish, R., Kapre, N., Dandekar, M., Gupta, S., Hawaldar, R., Agarwal, J. P., Pantvaidya, G., Chaukar, D., Deshmukh, A., Kane, S., Arya, S., Ghosh-Laskar, S., Chaturvedi, P., Pai, P., Nair, S., Nair, D., & Badwe, R. (2015). Elective versus Therapeutic Neck Dissection in Node-Negative Oral Cancer. *New England Journal of Medicine*, *373*(6), 521-529. https://doi.org/10.1056/NEJMoa1506007

Dagogo-Jack, I., & Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology*, *15*(2), 81-94. https://doi.org/10.1038/nrclinonc.2017.166

DeCoste, R. C., Walsh, N. M., Gaston, D., Ly, T. Y., Pasternak, S., Cutler, S., Nightingale, M., & Carter, M. D. (2022). RB1-deficient squamous cell carcinoma: the proposed source of combined Merkel cell carcinoma. *Modern Pathology*. https://doi.org/10.1038/s41379-022-01151-2

Dhasarathy, A., Phadke, D., Mav, D., Shah, R. R., & Wade, P. A. (2011). The Transcription Factors Snail and Slug Activate the Transforming Growth Factor-Beta Signaling Pathway in Breast Cancer. *PLOS ONE*, *6*(10), e26514. https://doi.org/10.1371/journal.pone.0026514

Dittberner, A., Friedl, B., Wittig, A., Buentzel, J., Kaftan, H., Boeger, D., Mueller, A. H., Schultze-Mosgau, S., Schlattmann, P., Ernst, T., & Guntinas-Lichius, O. (2020). Gender Disparities in Epidemiology, Treatment, and Outcome for Head and Neck Cancer in Germany: A Population-Based Long-Term Analysis from 1996 to 2016 of the Thuringian Cancer Registry. *Cancers*, *12*(11), 3418. https://doi.org/10.3390/cancers12113418

Dongre, A., Rashidian, M., Reinhardt, F., Bagnato, A., Keckesova, Z., Ploegh, H. L., & Weinberg, R. A. (2017). Epithelial-to-Mesenchymal Transition Contributes to Immunosuppression in Breast Carcinomas. *Cancer Research*, *77*(15), 3982-3989. https://doi.org/10.1158/0008-5472.Can-16-3292

Dongre, A., & Weinberg, R. A. (2019). New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nature Reviews Molecular Cell Biology*, *20*(2), 69-84. https://doi.org/10.1038/s41580-018-0080-4

Eisemann, N., Waldmann, A., Geller, A. C., Weinstock, M. A., Volkmer, B., Greinert, R., Breitbart, E. W., & Katalinic, A. (2014). Non-Melanoma Skin Cancer Incidence and Impact of Skin Cancer Screening on Incidence. *Journal of Investigative Dermatology*, *134*(1), 43-50. https://doi.org/10.1038/jid.2013.304

Elyada, E., Bolisetty, M., Laise, P., Flynn, W. F., Courtois, E. T., Burkhart, R. A., Teinor, J. A., Belleau, P., Biffi, G., Lucito, M. S., Sivajothi, S., Armstrong, T. D., Engle, D. D., Yu, K. H., Hao, Y., Wolfgang, C. L., Park, Y., Preall, J., Jaffee, E. M., et al. Tuveson, D. A. (2019). Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discovery*, *9*(8), 1102-1123. https://doi.org/10.1158/2159-8290.Cd-19-0094

Feng, H., Shuda, M., Chang, Y., & Moore, P. S. (2008). Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. *Science*, *319*(5866), 1096-1100. https://doi.org/10.1126/science.1152586

Fondain, M., Dereure, O., Uhry, Z., Guizard, A. V., Woronoff, A. S., Colonna, M., Molinie, F., Bara, S., Velten, M., Marrer, E., Grosclaude, P., Lapôtre-Ledoux, B., Tretarre, B., & Guillot, B. (2018). Merkel cell carcinoma in France: a registries-based, comprehensive epidemiological survey. *Journal of the European Academy of Dermatology and Venereology*, *32*(8), 1292-1296. https://doi.org/10.1111/jdv.14798

Galbo, P. M., Jr, Zang, X., & Zheng, D. (2021). Molecular Features of Cancer-associated Fibroblast Subtypes and their Implication on Cancer Pathogenesis, Prognosis, and Immunotherapy Resistance. *Clinical Cancer Research*, *27*(9), 2636-2647. https://doi.org/10.1158/1078-0432.Ccr-20-4226

Gambichler, T., Abu Rached, N., Tannapfel, A., Becker, J. C., Vogt, M., Skrygan, M., Wieland, U., Silling, S., Susok, L., Stücker, M., Meyer, T., Stockfleth, E., Junker, K., Käfferlein, H. U., Brüning, T., & Lang, K. (2021). Expression of Mismatch Repair Proteins in Merkel Cell Carcinoma. *Cancers*, *13*(11), 2524. https://doi.org/10.3390/cancers13112524

Gauci, M.-L., Aristei, C., Becker, J. C., Blom, A., Bataille, V., Dreno, B., Del Marmol, V., Forsea, A. M., Fargnoli, M. C., Grob, J.-J., Gomes, F., Hauschild, A., Hoeller, C., Harwood, C., Kelleners-Smeets, N., Kaufmann, R., Lallas, A., Malvehy, J., Moreno-Ramirez, D., et al. Lebbé, C. (2022). Diagnosis and treatment of Merkel cell carcinoma: European consensus-based interdisciplinary guideline – Update 2022. *European Journal of Cancer*, *171*, 203-231. https://doi.org/10.1016/j.ejca.2022.03.043

Gazdar, A. F., Bunn, P. A., & Minna, J. D. (2017). Small-cell lung cancer: what we know, what we need to know and the path forward. *Nature Reviews Cancer*, *17*(12), 725-737. https://doi.org/10.1038/nrc.2017.87

George, J., Lim, J. S., Jang, S. J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., Müller, C., Dahmen, I., Jahchan, N. S., Park, K.-S., Yang, D., Karnezis, A. N., Vaka, D., Torres, A., Wang, M. S., et al. Thomas, R. K. (2015). Comprehensive genomic profiles of small cell lung cancer. *Nature*, *524*(7563), 47-53. https://doi.org/10.1038/nature14664

Goh, G., Walradt, T., Markarov, V., Blom, A., Riaz, N., Doumani, R., Stafstrom, K., Moshiri, A., Yelistratova, L., Levinsohn, J., Chan, T. A., Nghiem, P., Lifton, R. P., & Choi, J. (2015). Mutational landscape of MCPyV-positive and MCPyV-negative Merkel cell carcinomas with implications for immunotherapy. *Oncotarget*, *7*(3). https://doi.org/10.18632/oncotarget.6494

González-Vela, M. d. C., Curiel-Olmo, S., Derdak, S., Beltran, S., Santibañez, M., Martínez, N., Castillo-Trujillo, A., Gut, M., Sánchez-Pacheco, R., Almaraz, C., Cereceda, L., Llombart, B., Agraz-Doblas, A., Revert-Arce, J., López Guerrero, J. A., Mollejo, M., Marrón, P. I., Ortiz-Romero, P., Fernandez-Cuesta, L., et al. Vaqué, J. P. (2017). Shared Oncogenic Pathways Implicated in Both Virus-Positive and UV-Induced Merkel Cell Carcinomas. *Journal of Investigative Dermatology*, *137*(1), 197-206. https://doi.org/10.1016/j.jid.2016.08.015

Gravemeyer, J., Lange, A., Ritter, C., Spassova, I., Song, L., Picard, D., Remke, M., Horny, K., Sriram, A., Gambichler, T., Schadendorf, D., Hoffmann, D., & Becker, J. C. (2021). Classical and Variant Merkel Cell Carcinoma Cell Lines Display Different Degrees of Neuroendocrine Differentiation and Epithelial-Mesenchymal Transition. *Journal of Investigative Dermatology*, *141*(7), 1675-1686.e1674. https://doi.org/10.1016/j.jid.2021.01.012

Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., Vadas, M. A., Khew-Goodall, Y., & Goodall, G. J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature Cell Biology*, *10*(5), 593-601. https://doi.org/10.1038/ncb1722

Harms, K. L., Healy, M. A., Nghiem, P., Sober, A. J., Johnson, T. M., Bichakjian, C. K., & Wong, S. L. (2016). Analysis of Prognostic Factors from 9387 Merkel Cell Carcinoma Cases Forms the Basis for the New 8th Edition AJCC Staging System. *Annals of Surgical Oncology*, *23*(11), 3564-3571. https://doi.org/10.1245/s10434-016-5266-4

Harms, P. W., Vats, P., Verhaegen, M. E., Robinson, D. R., Wu, Y.-M., Dhanasekaran, S. M., Palanisamy, N., Siddiqui, J., Cao, X., Su, F., Wang, R., Xiao, H., Kunju, L. P., Mehra, R., Tomlins, S. A., Fullen, D. R., Bichakjian, C. K., Johnson, T. M., Dlugosz, A. A., & Chinnaiyan, A. M. (2015). The Distinctive Mutational Spectra of Polyomavirus-Negative Merkel Cell Carcinoma. *Cancer Research*, *75*(18), 3720-3727. https://doi.org/10.1158/0008-5472.Can-15-0702

Hayward, N. K., Wilmott, J. S., Waddell, N., Johansson, P. A., Field, M. A., Nones, K., Patch, A.-M., Kakavand, H., Alexandrov, L. B., Burke, H., Jakrot, V., Kazakoff, S., Holmes, O., Leonard, C., Sabarinathan, R., Mularoni, L., Wood, S., Xu, Q., Waddell, N., et al. Mann, G. J. (2017). Whole-genome landscapes of major melanoma subtypes. *Nature*, *545*(7653), 175-180. https://doi.org/10.1038/nature22071

Hesbacher, S., Pfitzer, L., Wiedorfer, K., Angermeyer, S., Borst, A., Haferkamp, S., Scholz, C. J., Wobser, M., Schrama, D., & Houben, R. (2016). RB1 is the crucial target of the Merkel cell polyomavirus Large T antigen in Merkel cell carcinoma cells. *Oncotarget*, *7*(22), 32956-32968. https://doi.org/10.18632/oncotarget.8793

Hill, N. T., Kim, D., Busam, K. J., Chu, E. Y., Green, C., & Brownell, I. (2021). Distinct Signatures of Genomic Copy Number Variants Define Subgroups of Merkel Cell Carcinoma Tumors. *Cancers*, *13*(5), 1134. https://doi.org/10.3390/cancers13051134

Holland, C. H., Tanevski, J., Perales-Patón, J., Gleixner, J., Kumar, M. P., Mereu, E., Joughin, B. A., Stegle, O., Lauffenburger, D. A., Heyn, H., Szalai, B., & Saez-Rodriguez, J. (2020). Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biology*, *21*(1), 36. https://doi.org/10.1186/s13059-020-1949-z

Horny, K., Gerhardt, P., Hebel-Cherouny, A., Wülbeck, C., Utikal, J., & Becker, J. C. (2021). Mutational Landscape of Virus- and UV-Associated Merkel Cell Carcinoma Cell Lines Is Comparable to Tumor Tissue. *Cancers*, *13*(4), 649. https://doi.org/10.3390/cancers13040649

Horny, K., Sproll, C., Peiffer, L., Furtmann, F., Gerhardt, P., Gravemeyer, J., Stoecklein, N. H., Spassova, I., & Becker, J. C. (2023). Mesenchymal-epithelial transition in lymph node metastases of oral squamous cell carcinoma is accompanied by ZEB1 expression. *Journal of Translational Medicine*, *21*(1), 267. https://doi.org/10.1186/s12967-023-04102-w

Huang, Y., Hong, W., & Wei, X. (2022). The molecular mechanisms and therapeutic strategies of EMT in tumor progression and metastasis. *Journal of Hematology & Oncology*, *15*(1), 129. https://doi.org/10.1186/s13045-022-01347-8

Ireland, A. S., Micinski, A. M., Kastner, D. W., Guo, B., Wait, S. J., Spainhower, K. B., Conley, C. C., Chen, O. S., Guthrie, M. R., Soltero, D., Qiao, Y., Huang, X., Tarapcsák, S., Devarakonda, S., Chalishazar, M. D., Gertz, J., Moser, J. C., Marth, G., Puri, S., et al. Oliver, T. G. (2020). MYC Drives Temporal Evolution of Small Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate. *Cancer Cell*, *38*(1), 60-78.e12. https://doi.org/10.1016/j.ccell.2020.05.001

Jacobs, D., Huang, H., Olino, K., Weiss, S., Kluger, H., Judson, B. L., & Zhang, Y. (2021). Assessment of Age, Period, and Birth Cohort Effects and Trends in Merkel Cell Carcinoma Incidence in the United States. *JAMA Dermatology*, *157*(1), 59-65. https://doi.org/10.1001/jamadermatol.2020.4102

Jehanno, C., Vulin, M., Richina, V., Richina, F., & Bentires-Alj, M. (2022). Phenotypic plasticity during metastatic colonization. *Trends in Cell Biology*, *32*(10), 854-867. https://doi.org/10.1016/j.tcb.2022.03.007

Ji, A. L., Rubin, A. J., Thrane, K., Jiang, S., Reynolds, D. L., Meyers, R. M., Guo, M. G., George, B. M., Mollbrink, A., Bergenstråhle, J., Larsson, L., Bai, Y., Zhu, B., Bhaduri, A., Meyers, J. M., Rovira-Clavé, X., Hollmig, S. T., Aasi, S. Z., Nolan, G. P., et al. Khavari, P. A. (2020). Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell*, *182*(2), 497-514.e422. https://doi.org/10.1016/j.cell.2020.05.039

Johnson, D. E., Burtness, B., Leemans, C. R., Lui, V. W. Y., Bauman, J. E., & Grandis, J. R. (2020). Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers*, *6*(1), 92. https://doi.org/10.1038/s41572-020-00224-3

Jolly, M. K., Ware, K. E., Gilja, S., Somarelli, J. A., & Levine, H. (2017). EMT and MET: necessary or permissive for metastasis? *Molecular Oncology*, *11*(7), 755-769. https://doi.org/10.1002/1878-0261.12083

Jones, D., Pereira, E. R., & Padera, T. P. (2018). Growth and Immune Evasion of Lymph Node Metastasis. *Frontiers in Oncology*, *8*. https://doi.org/10.3389/fonc.2018.00036

Joshi, R. S., Kanugula, S. S., Sudhir, S., Pereira, M. P., Jain, S., & Aghi, M. K. (2021). The Role of Cancer-Associated Fibroblasts in Tumor Progression. *Cancers*, *13*(6), 1399. https://doi.org/10.3390/cancers13061399

Jung, A. R., Jung, C. H., Noh, J. K., Lee, Y. C., & Eun, Y. G. (2020). Epithelial-mesenchymal transition gene signature is associated with prognosis and tumor microenvironment in head and neck squamous cell carcinoma. *Sci Rep*, *10*(1), 3652. https://doi.org/10.1038/s41598-020-60707-x

Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nature Reviews Cancer*, *16*(9), 582-598. https://doi.org/10.1038/nrc.2016.73

Kazakova, A. N., Anufrieva, K. S., Ivanova, O. M., Shnaider, P. V., Malyants, I. K., Aleshikova, O. I., Slonov, A. V., Ashrafyan, L. A., Babaeva, N. A., Eremeev, A. V., Boichenko, V. S., Lukina, M. M., Lagarkova, M. A., Govorun, V. M., Shender, V. O., & Arapidi, G. P. (2022). Deeper insights into transcriptional features of cancer-associated fibroblasts: An integrated meta-analysis of single-cell and bulk RNA-sequencing data. *Frontiers in Cell and Developmental Biology*, *10*, 825014. https://doi.org/10.3389/fcell.2022.825014

Kervarrec, T., Appenzeller, S., Samimi, M., Sarma, B., Sarosi, E.-M., Berthon, P., Le Corre, Y., Hainaut-Wierzbicka, E., Blom, A., Benethon, N., Bens, G., Nardin, C., Aubin, F., Dinulescu, M., Jullie, M.-L., Pekár-Lukacs, Á., Calonje, E., Thanguturi, S., Tallet, A., et al. Schrama, D. (2022). Merkel Cell Polyomavirus–Negative Merkel Cell Carcinoma Originating from In Situ Squamous Cell Carcinoma: A Keratinocytic Tumor with Neuroendocrine Differentiation. *Journal of Investigative Dermatology*, *142*(3, Part A), 516-527. https://doi.org/10.1016/j.jid.2021.07.175

Kharchenko, P. V., Silberstein, L., & Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nature Methods*, *11*(7), 740-742. https://doi.org/10.1038/nmeth.2967

Kinker, G. S., Greenwald, A. C., Tal, R., Orlova, Z., Cuoco, M. S., McFarland, J. M., Warren, A., Rodman, C., Roth, J. A., Bender, S. A., Kumar, B., Rocco, J. W., Fernandes, P. A. C. M., Mader, C. C., Keren-Shaul, H., Plotnikov, A., Barr, H., Tsherniak, A., Rozenblatt-Rosen, O., et al. Tirosh, I. (2020). Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nature Genetics*, *52*(11), 1208-1218. https://doi.org/10.1038/s41588-020-00726-6

Knepper, T. C., Montesion, M., Russell, J. S., Sokol, E. S., Frampton, G. M., Miller, V. A., Albacker, L. A., McLeod, H. L., Eroglu, Z., Khushalani, N. I., Sondak, V. K., Messina, J. L., Schell, M. J., DeCaprio, J. A., Tsai, K. Y., & Brohl, A. S. (2019). The Genomic Landscape of Merkel Cell Carcinoma and Clinicogenomic Biomarkers of Response to Immune Checkpoint Inhibitor Therapy. *Clinical Cancer Research*, *25*(19), 5961-5971. https://doi.org/10.1158/1078-0432.Ccr-18-4159

Knepper, T. C., Panchaud, R. A., Muradova, E., Cohen, L., DeCaprio, J. A., Khushalani, N. I., Tsai, K. Y., & Brohl, A. S. (2021). An analysis of the use of targeted therapies in patients with advanced Merkel cell carcinoma and an evaluation of genomic correlates of response. *Cancer Medicine*, *10*(17), 5889-5896. https://doi.org/10.1002/cam4.4138

Knight, L. M., Stakaityte, G., Jennifer, J. W., Abdul-Sada, H., Griffiths, D. A., Howell, G. J., Wheat, R., Blair, G. E., Steven, N. M., Macdonald, A., Blackbourn, D. J., & Whitehouse, A. (2015). Merkel Cell Polyomavirus Small T Antigen Mediates Microtubule Destabilization To Promote Cell Motility and Migration. *Journal of Virology*, *89*(1), 35-47. https://doi.org/10.1128/JVI.02317-14

Kopp, H.-G., Placke, T., & Salih, H. R. (2009). Platelet-Derived Transforming Growth Factor-β Down-Regulates NKG2D Thereby Inhibiting Natural Killer Cell Antitumor Reactivity. *Cancer Research*, *69*(19), 7775-7783. https://doi.org/10.1158/0008-5472.Can-09-2123

Kudo-Saito, C., Shirako, H., Takeuchi, T., & Kawakami, Y. (2009). Cancer Metastasis Is Accelerated through Immunosuppression during Snail-Induced EMT of Cancer Cells. *Cancer Cell*, *15*(3), 195-206. https://doi.org/10.1016/j.ccr.2009.01.023

Kürten, C. H. L., Kulkarni, A., Cillo, A. R., Santos, P. M., Roble, A. K., Onkar, S., Reeder, C., Lang, S., Chen, X., Duvvuri, U., Kim, S., Liu, A., Tabib, T., Lafyatis, R., Feng, J., Gao, S.-J., Bruno, T. C., Vignali, D. A. A., Lu, X., et al. Ferris, R. L. (2021). Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nature Communications*, *12*(1), 7338. https://doi.org/10.1038/s41467-021-27619-4

Kwun, Hyun J., Shuda, M., Feng, H., Camacho, Carlos J., Moore, Patrick S., & Chang, Y. (2013). Merkel Cell Polyomavirus Small T Antigen Controls Viral Replication and Oncoprotein Expression by Targeting the Cellular Ubiquitin Ligase SCF Fbw7. *Cell Host & Microbe*, *14*(2), 125-135. https://doi.org/10.1016/j.chom.2013.06.008

Lambert, A. W., Pattabiraman, D. R., & Weinberg, R. A. (2017). Emerging Biological Principles of Metastasis. *Cell*, *168*(4), 670-691. https://doi.org/10.1016/j.cell.2016.11.037

Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N., & Werb, Z. (2018). Tumour heterogeneity and metastasis at single-cell resolution. *Nature Cell Biology*, *20*(12), 1349-1360. https://doi.org/10.1038/s41556-018-0236-7

Li, J., Wang, X., Diaz, J., Tsang, S. H., Buck, C. B., & You, J. (2013). Merkel Cell Polyomavirus Large T Antigen Disrupts Host Genomic Integrity and Inhibits Cellular Proliferation. *Journal of Virology*, *87*(16), 9173-9188. https://doi.org/10.1128/JVI.01216-13

Long, J., Wang, D., Wang, A., Chen, P., Lin, Y., Bian, J., Yang, X., Zheng, M., Zhang, H., Zheng, Y., Sang, X., & Zhao, H. (2022). A mutation-based gene set predicts survival benefit after immunotherapy across multiple cancers and reveals the immune response landscape. *Genome Medicine*, *14*(1), 20. https://doi.org/10.1186/s13073-022-01024-y

Longo, S. K., Guo, M. G., Ji, A. L., & Khavari, P. A. (2021). Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews Genetics*, *22*(10), 627-644. https://doi.org/10.1038/s41576-021-00370-8

Lüönd, F., Sugiyama, N., Bill, R., Bornes, L., Hager, C., Tang, F., Santacroce, N., Beisel, C., Ivanek, R., Bürglin, T., Tiede, S., van Rheenen, J., & Christofori, G. (2021). Distinct contributions of partial and full EMT to breast cancer malignancy. *Developmental Cell*, *56*(23), 3203-3221.e3211. https://doi.org/10.1016/j.devcel.2021.11.006

Majmundar, A. J., Wong, W. J., & Simon, M. C. (2010). Hypoxia-Inducible Factors and the Response to Hypoxic Stress. *Molecular Cell*, *40*(2), 294-309. https://doi.org/10.1016/j.molcel.2010.09.022

Manders, F., Brandsma, A. M., de Kanter, J., Verheul, M., Oka, R., van Roosmalen, M. J., van der Roest, B., van Hoeck, A., Cuppen, E., & van Boxtel, R. (2022). MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genomics*, *23*(1), 134. https://doi.org/10.1186/s12864-022-08357-3

Maura, F., Degasperi, A., Nadeu, F., Leongamornlert, D., Davies, H., Moore, L., Royo, R., Ziccheddu, B., Puente, X. S., Avet-Loiseau, H., Campbell, P. J., Nik-Zainal, S., Campo, E., Munshi, N., & Bolli, N. (2019). A practical guide for mutational signature analysis in hematological malignancies. *Nature Communications*, *10*(1), 2969. https://doi.org/10.1038/s41467-019-11037-8

McFaline-Figueroa, J. L., Hill, A. J., Qiu, X., Jackson, D., Shendure, J., & Trapnell, C. (2019). A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nature Genetics*, *51*(9), 1389-1398. https://doi.org/10.1038/s41588-019-0489-5

Meng, J., Chen, S., Han, J.-X., Qian, B., Wang, X.-R., Zhong, W.-L., Qin, Y., Zhang, H., Gao, W.-F., Lei, Y.-Y., Yang, W., Yang, L., Zhang, C., Liu, H.-J., Liu, Y.-R., Zhou, H.-G., Sun, T., & Yang, C. (2018). Twist1 Regulates Vimentin through Cul2 Circular RNA to

Promote EMT in Hepatocellular Carcinoma. *Cancer Research*, *78*(15), 4150-4162. https://doi.org/10.1158/0008-5472.Can-17-3009

Mollaoglu, G., Guthrie, M. R., Böhm, S., Brägelmann, J., Can, I., Ballieu, P. M., Marx, A., George, J., Heinen, C., Chalishazar, M. D., Cheng, H., Ireland, A. S., Denning, K. E., Mukhopadhyay, A., Vahrenkamp, J. M., Berrett, K. C., Mosbruger, T. L., Wang, J., Kohan, J. L., et al. Oliver, T. G. (2017). MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. *Cancer Cell*, *31*(2), 270-285. https://doi.org/10.1016/j.ccell.2016.12.005

Narayanan, D. L., Saladi, R. N., & Fox, J. L. (2010). Ultraviolet radiation and skin cancer. *International Journal of Dermatology*, *49*(9), 978-986. https://doi.org/10.1111/j.1365-4632.2010.04474.x

Nieto, M. A., Huang, Ruby Y.-J., Jackson, Rebecca A., & Thiery, Jean P. (2016). EMT: 2016. *Cell*, *166*(1), 21-45. https://doi.org/10.1016/j.cell.2016.06.028

Nik-Zainal, S., Alexandrov, Ludmil B., Wedge, David C., Van Loo, P., Greenman, Christopher D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, Lucy A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, King W., Mudie, Laura J., et al. Stratton, Michael R. (2012). Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell*, *149*(5), 979-993. https://doi.org/10.1016/j.cell.2012.04.024

Nik-Zainal, S., Kucab, J. E., Morganella, S., Glodzik, D., Alexandrov, L. B., Arlt, V. M., Weninger, A., Hollstein, M., Stratton, M. R., & Phillips, D. H. (2015). The genome as a record of environmental exposure. *Mutagenesis*, *30*(6), 763-770. https://doi.org/10.1093/mutage/gev073

Noguti, J., De Moura, C. F. G., De Jesus, G. P. P., Da Silva, V. H. P., Hossaka, T. A., Oshima, C. T. F., & Ribeiro, D. A. (2012). Metastasis from Oral Cancer: An Overview. *Cancer Genomics - Proteomics*, *9*(5), 329. http://cgp.iiarjournals.org/content/9/5/329.abstract

Pal, A., Barrett, T. F., Paolini, R., Parikh, A., & Puram, S. V. (2021). Partial EMT in head and neck cancer biology: a spectrum instead of a switch. *Oncogene*, *40*(32), 5049-5065. https://doi.org/10.1038/s41388-021-01868-5

Palla, G., Fischer, D. S., Regev, A., & Theis, F. J. (2022). Spatial components of molecular tissue biology. *Nature Biotechnology*, *40*(3), 308-318. https://doi.org/10.1038/s41587-021-01182-1

Parikh, A. S., Puram, S. V., Faquin, W. C., Richmon, J. D., Emerick, K. S., Deschler, D. G., Varvares, M. A., Tirosh, I., Bernstein, B. E., & Lin, D. T. (2019). Immunohistochemical quantification of partial-EMT in oral cavity squamous cell carcinoma primary tumors is associated with nodal metastasis. *Oral Oncology*, *99*, 104458. https://doi.org/10.1016/j.oraloncology.2019.104458

Park, D. E., Cheng, J., Berrios, C., Montero, J., Cortés-Cros, M., Ferretti, S., Arora, R., Tillgren, M. L., Gokhale, P. C., & DeCaprio, J. A. (2019). Dual inhibition of MDM2 and MDM4 in virus-positive Merkel cell carcinoma enhances the p53 response. *Proceedings of the National Academy of Sciences*, *116*(3), 1027-1032. https://doi.org/10.1073/pnas.1818798116

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., De Clercq, S., Minguijón, E., Balsat, C., Sokolow, Y., Dubois, C., De Cock, F., Scozzaro, S., Sopena, F., Lanas, A., et al. Blanpain, C. (2018). Identification of the tumour transition states occurring during EMT. *Nature*, *556*(7702), 463-468. https://doi.org/10.1038/s41586-018-0040-3

Patel, A. S., Yoo, S., Kong, R., Sato, T., Sinha, A., Karam, S., Bao, L., Fridrikh, M., Emoto, K., Nudelman, G., Powell, C. A., Beasley, M. B., Zhu, J., & Watanabe, H. (2021). Prototypical oncogene family Myc defines unappreciated distinct lineage states of small cell lung cancer. *Science Advances*, *7*(5), eabc2578. https://doi.org/10.1126/sciadv.abc2578

Paulson, K. G., Lemos, B. D., Feng, B., Jaimes, N., Peñas, P. F., Bi, X., Maher, E., Cohen, L., Helen Leonard, J., Granter, S. R., Chin, L., & Nghiem, P. (2009). Array-CGH Reveals Recurrent Genomic Changes in Merkel Cell Carcinoma Including Amplification of L-

Myc. *Journal of Investigative Dermatology*, *129*(6), 1547-1555. https://doi.org/10.1038/jid.2008.365

Paulson, K. G., Park, S. Y., Vandeven, N. A., Lachance, K., Thomas, H., Chapuis, A. G., Harms, K. L., Thompson, J. A., Bhatia, S., Stang, A., & Nghiem, P. (2018). Merkel cell carcinoma: Current US incidence and projected increases based on changing demographics. *Journal of the American Academy of Dermatology*, *78*(3), 457-463.e452. https://doi.org/10.1016/j.jaad.2017.10.028

Pelon, F., Bourachot, B., Kieffer, Y., Magagna, I., Mermet-Meillon, F., Bonnet, I., Costa, A., Givel, A.-M., Attieh, Y., Barbazan, J., Bonneau, C., Fuhrmann, L., Descroix, S., Vignjevic, D., Silberzan, P., Parrini, M. C., Vincent-Salomon, A., & Mechta-Grigoriou, F. (2020). Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. *Nature Communications*, *11*(1), 404. https://doi.org/10.1038/s41467-019-14134-w

Pfeifer, G. P., You, Y.-H., & Besaratinia, A. (2005). Mutations induced by ultraviolet light. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *571*(1), 19-31. https://doi.org/10.1016/j.mrfmmm.2004.06.057

Pucci, F., Garris, C., Lai, C. P., Newton, A., Pfirschke, C., Engblom, C., Alvarez, D., Sprachman, M., Evavold, C., Magnuson, A., von Andrian, U. H., Glatz, K., Breakefield, X. O., Mempel, T. R., Weissleder, R., & Pittet, M. J. (2016). SCS macrophages suppress melanoma by restricting tumor-derived vesicle-B cell interactions. *Science*, *352*(6282), 242-246. https://doi.org/10.1126/science.aaf1328

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C. L., Mroz, E. A., Emerick, K. S., Deschler, D. G., Varvares, M. A., Mylvaganam, R., Rozenblatt-Rosen, O., Rocco, J. W., Faquin, W. C., Lin, D. T., Regev, A., & Bernstein, B. E. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell*, *171*(7), 1611-1624.e1624. https://doi.org/10.1016/j.cell.2017.10.044

Qian, J., Olbrecht, S., Boeckx, B., Vos, H., Laoui, D., Etlioglu, E., Wauters, E., Pomella, V., Verbandt, S., Busschaert, P., Bassez, A., Franken, A., Bempt, M. V., Xiong, J., Weynand, B., van Herck, Y., Antoranz, A., Bosisio, F. M., Thienpont, B., et al. Lambrechts, D. (2020). A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Research*, *30*(9), 745-762. https://doi.org/10.1038/s41422-020-0355-0

Redfern, A. D., Spalding, L. J., & Thompson, E. W. (2018). The Kraken Wakes: induced EMT as a driver of tumour aggression and poor outcome. *Clinical & Experimental Metastasis*, *35*(4), 285-308. https://doi.org/10.1007/s10585-018-9906-x

Reichgelt, B. A., & Visser, O. (2011). Epidemiology and survival of Merkel cell carcinoma in the Netherlands. A population-based study of 808 cases in 1993–2007. *European Journal of Cancer*, *47*(4), 579-585. https://doi.org/10.1016/j.ejca.2010.11.002

Reymond, N., d'Água, B. B., & Ridley, A. J. (2013). Crossing the endothelial barrier during metastasis. *Nature Reviews Cancer*, *13*(12), 858-870. https://doi.org/10.1038/nrc3628

Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G. V., Carter, S. L., Saksena, G., Harris, S., Shah, R. R., Resnick, M. A., Getz, G., & Gordenin, D. A. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature Genetics*, *45*(9), 970-976. https://doi.org/10.1038/ng.2702

Rubió-Casadevall, J., Hernandez-Pujol, A. M., Ferreira-Santos, M. C., Morey-Esteve, G., Vilardell, L., Osca-Gelis, G., Vilar-Coromina, N., & Marcos-Gragera, R. (2016). Trends in incidence and survival analysis in non-melanoma skin cancer from 1994 to 2012 in Girona, Spain: A population-based study. *Cancer Epidemiology*, *45*, 6-10. https://doi.org/10.1016/j.canep.2016.09.001

Ruscetti, M., Quach, B., Dadashian, E. L., Mulholland, D. J., & Wu, H. (2015). Tracking and Functional Characterization of Epithelial–Mesenchymal Transition and Mesenchymal Tumor Cells during Prostate Cancer Metastasis. *Cancer Research*, *75*(13), 2749-2759. https://doi.org/10.1158/0008-5472.Can-14-3476

Saini, N., Roberts, S. A., Klimczak, L. J., Chan, K., Grimm, S. A., Dai, S., Fargo, D. C., Boyer, J. C., Kaufmann, W. K., Taylor, J. A., Lee, E., Cortes-Ciriano, I., Park, P. J., Schurman, S. H., Malc, E. P., Mieczkowski, P. A., & Gordenin, D. A. (2016). The Impact of Environmental and Endogenous Damage on Somatic Mutation Load in Human Skin Fibroblasts. *PLOS Genetics*, *12*(10), e1006385. https://doi.org/10.1371/journal.pgen.1006385

Samuel, R. J., Matthews, A. G., & Holme, S. A. (2015). Merkel cell carcinoma in Scotland 2000–10. *British Journal of Dermatology*, *173*(4), 1073-1075. https://doi.org/10.1111/bjd.13869

Sánchez-Tilló, E., Lázaro, A., Torrent, R., Cuatrecasas, M., Vaquero, E. C., Castells, A., Engel, P., & Postigo, A. (2010). ZEB1 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1. *Oncogene*, *29*(24), 3490-3500. https://doi.org/10.1038/onc.2010.102

Saxena, K., Jolly, M. K., & Balamurugan, K. (2020). Hypoxia, partial EMT and collective migration: Emerging culprits in metastasis. *Translational Oncology*, *13*(11), 100845. https://doi.org/10.1016/j.tranon.2020.100845

Schadendorf, D., Lebbé, C., Zur Hausen, A., Avril, M. F., Hariharan, S., Bharmal, M., & Becker, J. C. (2017). Merkel cell carcinoma: Epidemiology, prognosis, therapy and unmet medical needs. *European Journal of Cancer*, *71*, 53-69. https://doi.org/10.1016/j.ejca.2016.10.022

Schadendorf, D., Nghiem, P., Bhatia, S., Hauschild, A., Saiag, P., Mahnke, L., Hariharan, S., & Kaufman, H. L. (2017). Immune evasion mechanisms and immune checkpoint inhibition in advanced merkel cell carcinoma. *OncoImmunology*, *6*(10), e1338237. https://doi.org/10.1080/2162402X.2017.1338237

Schinke, H., Pan, M., Akyol, M., Zhou, J., Shi, E., Kranz, G., Libl, D., Quadt, T., Simon, F., Canis, M., Baumeister, P., & Gires, O. (2022). SLUG-related partial epithelial-to-mesenchymal transition is a transcriptomic prognosticator of head and neck cancer survival. *Molecular Oncology*, *16*(2), 347-367. https://doi.org/10.1002/1878-0261.13075

Schrama, D., Hesbacher, S., Angermeyer, S., Schlosser, A., Haferkamp, S., Aue, A., Adam, C., Weber, A., Schmidt, M., & Houben, R. (2016). Serine 220 phosphorylation of the Merkel cell polyomavirus large T antigen crucially supports growth of Merkel cell carcinoma cells. *International Journal of Cancer*, *138*(5), 1153-1162. https://doi.org/10.1002/ijc.29862

Shuda, M., Feng, H., Kwun, H. J., Rosen, S. T., Gjoerup, O., Moore, P. S., & Chang, Y. (2008). T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. *Proceedings of the National Academy of Sciences*, *105*(42), 16272-16277. https://doi.org/10.1073/pnas.0806526105

Shuda, M., Kwun, H. J., Feng, H., Chang, Y., & Moore, P. S. (2011). Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator. *The Journal of Clinical Investigation*, *121*(9), 3623-3634. https://doi.org/10.1172/JCI46323

Siebels, S., Czech-Sioli, M., Spohn, M., Schmidt, C., Theiss, J., Indenbirken, D., Günther, T., Grundhoff, A., & Fischer, N. (2020). Merkel Cell Polyomavirus DNA Replication Induces Senescence in Human Dermal Fibroblasts in a Kap1/Trim28-Dependent Manner. *mBio*, *11*(2). https://doi.org/10.1128/mBio.00142-20

Solé-Boldo, L., Raddatz, G., Schütz, S., Mallm, J.-P., Rippe, K., Lonsdorf, A. S., Rodríguez-Paredes, M., & Lyko, F. (2020). Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Communications Biology*, *3*(1), 188. https://doi.org/10.1038/s42003-020-0922-4

Song, Y., Azari, F. S., Tang, R., Shannon, A. B., Miura, J. T., Fraker, D. L., & Karakousis, G. C. (2021). Patterns of Metastasis in Merkel Cell Carcinoma. *Annals of Surgical Oncology*, *28*(1), 519-529. https://doi.org/10.1245/s10434-020-08587-3

Spassova, I., Ugurel, S., Kubat, L., Zimmer, L., Terheyden, P., Mohr, A., Björn Andtback, H., Villabona, L., Leiter, U., Eigentler, T., Loquai, C., Hassel, J. C., Gambichler, T., Haferkamp, S., Mohr, P., Pfoehler, C., Heinzerling, L., Gutzmer, R., Utikal, J. S., et al.

Becker, J. C. (2022). Clinical and molecular characteristics associated with response to therapeutic PD-1/PD-L1 inhibition in advanced Merkel cell carcinoma. *Journal for ImmunoTherapy of Cancer*, *10*(1), e003198. https://doi.org/10.1136/jitc-2021-003198

Stakaitytė, G., Nwogu, N., Dobson, S. J., Knight, L. M., Wasson, C. W., Salguero, F. J., Blackbourn, D. J., Blair, G. E., Mankouri, J., Macdonald, A., & Whitehouse, A. (2018). Merkel Cell Polyomavirus Small T Antigen Drives Cell Motility via Rho-GTPase-Induced Filopodium Formation. *Journal of Virology*, *92*(2). https://doi.org/10.1128/jvi.00940-17

Stang, A., Becker, J. C., Nghiem, P., & Ferlay, J. (2018). The association between geographic location and incidence of Merkel cell carcinoma in comparison to melanoma: An international assessment. *European Journal of Cancer*, *94*, 47-60. https://doi.org/10.1016/j.ejca.2018.02.003

Starrett, G. J., Marcelus, C., Cantalupo, P. G., Katz, J. P., Cheng, J., Akagi, K., Thakuria, M., Rabinowits, G., Wang, L. C., Symer, D. E., Pipas, J. M., Harris, R. S., & DeCaprio, J. A. (2017). Merkel Cell Polyomavirus Exhibits Dominant Control of the Tumor Genome and Transcriptome in Virus-Associated Merkel Cell Carcinoma. *mBio*, *8*(1), e02079-02016. https://doi.org/10.1128/mBio.02079-16

Starrett, G. J., Thakuria, M., Chen, T., Marcelus, C., Cheng, J., Nomburg, J., Thorner, A. R., Slevin, M. K., Powers, W., Burns, R. T., Perry, C., Piris, A., Kuo, F. C., Rabinowits, G., Giobbie-Hurder, A., MacConaill, L. E., & DeCaprio, J. A. (2020). Clinical and molecular characterization of virus-positive and virus-negative Merkel cell carcinoma. *Genome Medicine*, *12*(1), 30. https://doi.org/10.1186/s13073-020-00727-4

Stemmler, M. P., Eccles, R. L., Brabletz, S., & Brabletz, T. (2019). Non-redundant functions of EMT transcription factors. *Nature Cell Biology*, *21*(1), 102-112. https://doi.org/10.1038/s41556-018-0196-y

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, *458*(7239), 719-724. https://doi.org/10.1038/nature07943

Tam, S. Y., Wu, V. W. C., & Law, H. K. W. (2020). Hypoxia-Induced Epithelial-Mesenchymal Transition in Cancers: HIF-1α and Beyond. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/fonc.2020.00486

Tauriello, D. V. F., Palomo-Ponce, S., Stork, D., Berenguer-Llergo, A., Badia-Ramentol, J., Iglesias, M., Sevillano, M., Ibiza, S., Cañellas, A., Hernando-Momblona, X., Byrom, D., Matarin, J. A., Calon, A., Rivas, E. I., Nebreda, A. R., Riera, A., Attolini, C. S.-O., & Batlle, E. (2018). TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature*, *554*(7693), 538-543. https://doi.org/10.1038/nature25492

Topalian, S. L., Bhatia, S., Amin, A., Kudchadkar, R. R., Sharfman, W. H., Lebbé, C., Delord, J.-P., Dunn, L. A., Shinohara, M. M., Kulikauskas, R., Chung, C. H., Martens, U. M., Ferris, R. L., Stein, J. E., Engle, E. L., Devriese, L. A., Lao, C. D., Gu, J., Li, B., et al. Nghiem, P. (2020). Neoadjuvant Nivolumab for Patients With Resectable Merkel Cell Carcinoma in the CheckMate 358 Trial. *Journal of Clinical Oncology*, *38*(22), 2476-2487. https://doi.org/10.1200/jco.20.00201

Tran, K. B., Lang, J. J., Compton, K., Xu, R., Acheson, A. R., Henrikson, H. J., Kocarnik, J. M., Penberthy, L., Aali, A., Abbas, Q., Abbasi, B., Abbasi-Kangevari, M., Abbasi-Kangevari, Z., Abbastabar, H., Abdelmasseh, M., Abd-Elsalam, S., Abdelwahab, A. A., Abdoli, G., Abdulkadir, H. A., et al. Murray, C. J. L. (2022). The global burden of cancer attributable to risk factors, 2010-19: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *400*(10352), 563-591. https://doi.org/10.1016/S0140-6736(22)01438-6

Tyler, M., & Tirosh, I. (2021). Decoupling epithelial-mesenchymal transitions from stromal profiles by integrative expression analysis. *Nature Communications*, *12*(1), 2592. https://doi.org/10.1038/s41467-021-22800-1

Veija, T., Sarhadi, V. K., Koljonen, V., Bohling, T., & Knuutila, S. (2016). Hotspot mutations in polyomavirus positive and negative Merkel cell carcinomas. *Cancer Genetics*, *209*(1), 30-35. https://doi.org/10.1016/j.cancergen.2015.11.006

Velten, B., Braunger, J. M., Argelaguet, R., Arnol, D., Wirbel, J., Bredikhin, D., Zeller, G., & Stegle, O. (2022). Identifying temporal and spatial patterns of variation from multimodal

data using MEFISTO. *Nature Methods*, *19*(2), 179-186. https://doi.org/10.1038/s41592-021-01343-9

Verhaegen, M. E., Mangelberger, D., Harms, P. W., Vozheiko, T. D., Weick, J. W., Wilbert, D. M., Saunders, T. L., Ermilov, A. N., Bichakjian, C. K., Johnson, T. M., Imperiale, M. J., & Dlugosz, A. A. (2015). Merkel Cell Polyomavirus Small T Antigen Is Oncogenic in Transgenic Mice. *Journal of Investigative Dermatology*, *135*(5), 1415-1424. https://doi.org/10.1038/jid.2014.446

von der Grün, J., Winkelmann, R., Meissner, M., Wieland, U., Silling, S., Martin, D., Fokas, E., Rödel, C., Rödel, F., & Balermpas, P. (2019). Merkel Cell Polyoma Viral Load and Intratumoral CD8+ Lymphocyte Infiltration Predict Overall Survival in Patients With Merkel Cell Carcinoma. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.00020

Wherry, E. J., & Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. *Nature Reviews Immunology*, *15*(8), 486-499. https://doi.org/10.1038/nri3862

Wong, S. Q., Waldeck, K., Vergara, I. A., Schröder, J., Madore, J., Wilmott, J. S., Colebatch, A. J., De Paoli-Iseppi, R., Li, J., Lupat, R., Semple, T., Arnau, G. M., Fellowes, A., Leonard, J. H., Hruby, G., Mann, G. J., Thompson, J. F., Cullinane, C., Johnston, M., et al. Tothill, R. W. (2015). UV-Associated Mutations Underlie the Etiology of MCV-Negative Merkel Cell Carcinomas. *Cancer Research*, *75*(24), 5228-5234. https://doi.org/10.1158/0008-5472.Can-15-1877

Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G., Dedhar, S., Derynck, R., Ford, H. L., Fuxe, J., García de Herreros, A., Goodall, G. J., Hadjantonakis, A.-K., Huang, R. Y. J., Kalcheim, C., et al. On behalf of the, E. M. T. I. A. (2020). Guidelines and definitions for research on epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology*, *21*(6), 341-352. https://doi.org/10.1038/s41580-020-0237-9

Yang, Y., Ma, B., Han, L., Xu, W., Du, X., Wei, W., Liao, T., Ji, Q., Qu, N., & Wang, Y. (2022). Integrated single-cell and bulk RNA sequencing analyses reveal a prognostic signature of cancer-associated fibroblasts in head and neck squamous cell carcinoma. *Frontiers in Genetics*, *13*, 1028469. https://doi.org/10.3389/fgene.2022.1028469

Yao, D., Dai, C., & Peng, S. (2011). Mechanism of the Mesenchymal–Epithelial Transition and Its Relationship with Metastatic Tumor Formation. *Molecular Cancer Research*, *9*(12), 1608-1620. https://doi.org/10.1158/1541-7786.Mcr-10-0568

Yarchoan, M., Hopkins, A., & Jaffee, E. M. (2017). Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *New England Journal of Medicine*, *377*(25), 2500-2501. https://doi.org/10.1056/NEJMc1713444

Yu, M., Bardia, A., Wittner, B. S., Stott, S. L., Smas, M. E., Ting, D. T., Isakoff, S. J., Ciciliano, J. C., Wells, M. N., Shah, A. M., Concannon, K. F., Donaldson, M. C., Sequist, L. V., Brachtel, E., Sgroi, D., Baselga, J., Ramaswamy, S., Toner, M., Haber, D. A., & Maheswaran, S. (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science*, *339*(6119), 580-584. https://doi.org/10.1126/science.1228522

Yu, Y., Xiao, C. H., Tan, L. D., Wang, Q. S., Li, X. Q., & Feng, Y. M. (2014). Cancer-associated fibroblasts induce epithelial–mesenchymal transition of breast cancer cells through paracrine TGF-β signalling. *British Journal of Cancer*, *110*(3), 724-732. https://doi.org/10.1038/bjc.2013.768

Zámborszky, J., Szikriszt, B., Gervai, J. Z., Pipek, O., Póti, Á., Krzystanek, M., Ribli, D., Szalai-Gindl, J. M., Csabai, I., Szallasi, Z., Swanton, C., Richardson, A. L., & Szüts, D. (2017). Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. *Oncogene*, *36*(6), 746-755. https://doi.org/10.1038/onc.2016.243

zur Hausen, H. (1991). Viruses in Human Cancers. *Science*, *254*(5035), 1167-1173. https://doi.org/10.1126/science.1659743

6 Acknowledgements

Obtaining a doctoral degree is a long road that you walk together with many people. I am grateful for my colleagues from the “Translational skin cancer research” group, who brought joy and fun into my student life. Thanks for the great atmosphere, the support in my research projects, and for sharing this experience. I want to thank Jürgen C. Becker, who guided me through this time. He taught me to question my results, to shape my research into graspable visualizations, and to look at things from different perspectives. Special thanks go to Ivelina Spassova-Gießler, Lukas Peiffer, Frauke Furtmann, Angela Hebel-Cherouny, Corinna Wülbeck, and Patricia Gerhardt who performed most of the wet-lab experiments for my research projects. I cannot list all the small contributions of all the group members here, but I am truly grateful for the friendly help, constructive cooperation, and great discussions. I would like to sincerely thank Christoph Sproll from the University Hospital of the Heinrich-Heine University Düsseldorf for the productive collaboration, clinical expertise, and scientific enthusiasm.

During my time in Essen, I learned many things about bioinformatics that would not have been possible without all the people supporting me in becoming a bioinformatician. I want to thank Jan Gravemeyer for your advice when R was making problems again, for all the interesting scientific discussions about MCC biology and also other topics as politics, nature, and sloth-eating predator birds. I like to thank Daniel Hoffmann for encouraging me to be thorough and to think outside my research area, and for opening my eyes to all the problems with statistical significance. I also like to thank my first student Noorhan Kamal for your enthusiasm and hard work during your internship. I am also glad to always have Sabine Kaul supporting me when it comes to solving administrative issues and removing obstacles that get in my way.

Sadly, not everyone could live long enough to see this dissertation. I thank Ashwin Sriram for your joyfulness and passion for science that will be sorely missed. Ich danke meinen kürzlich verstorbenen Großmüttern Magdalene Westermann und Maria Horny: Danke für den steten Frohsinn, danke für eure Zuversicht, danke, dass ihr immer stolz auf meine Leistungen wart und danke, dass ihr stets für unsere Familie da wart.

Without my family and friends, I would be a different and much unhappier person. I thank all my friends who have accompanied me on this journey, celebrated all my achievements and cheered me up when I was feeling miserable. Despite the long distance between us, I want to especially thank Severina for indeed sharing the pain and glory of obtaining a doctoral degree. A great thank you goes to my girlfriend Annica for patiently listening to all my scientific stories and complaints, and for supporting me with your love. Schließlich bin ich am meisten meinen Eltern und Geschwistern dankbar. Danke, dass ihr immer da seid, wenn ich euch brauche.

7 Eidesstattliche Erklärung

Erklärung:

Hiermit erkläre ich, gem. § 6 Abs. (2) g) der Promotionsordnung der Fakultät für Biologie zur Erlangung der Dr. rer. nat., dass ich das Arbeitsgebiet, dem das Thema „Genomic and Phenotypic Heterogeneity of Malignant Tumors and their Microenvironment“ zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Kai Horny befürworte und die Betreuung auch im Falle eines Weggangs, wenn nicht wichtige Gründe dem entgegenstehen, weiterführen werde.

Essen, den _____ _____ _____

Name des Betreuers an der Universität Duisburg-Essen

Unterschrift

Erklärung:

Hiermit erkläre ich, gem. § 7 Abs. (2) d) + f) der Promotionsordnung der Fakultät für Biologie zur Erlangung des Dr. rer. nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient, bei der Abfassung der Dissertation nur die angegeben Hilfsmittel benutzt und alle wörtlich oder inhaltlich übernommenen Stellen als solche gekennzeichnet habe.

Essen, den _____ _____

Unterschrift des Doktoranden

Erklärung:

Hiermit erkläre ich, gem. § 7 Abs. (2) e) + g) der Promotionsordnung der Fakultät für Biologie zur Erlangung des Dr. rer. nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät/Fachbereich abgelehnt worden ist.

Essen, den _____ _____

Unterschrift des Doktoranden