Aus der Klinik und Poliklinik für Hals-, Nasen- und Ohrenheilkunde

Klinik der Ludwig-Maximilians-Universität München

Direktor Prof. Dr. Martin Canis

EGF/EGFR-Mediated Epithelial-to-Mesenchymal Transition in the Disease Progression of Head and Neck Squamous Cell Carcinoma

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von

Henrik Schinke

aus
Paderborn

Jahr
2020

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Betreuer(in):	Prof. Dr. Oliver Gires
Zweitgutachter(in):	Prof. Dr. Horst Zitzelsberger
Dekan:	Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung:	30.04.2021

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig und ohne fremde Hilfe angefertigt habe. Textpassagen, die wörtlich oder dem Sinn nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche kenntlich gemacht.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

den 11.05.2021

München, Datum

Henrik Schinke

Name (+ Unterschrift)

Acknowledgements

First, I want to thank my supervisor Olivier Gires for his support, patience, and education. Second, I want to thank Min Pan aka "No1" for introducing me to this project and the countless hours we spent philosophizing. Both, Olivier Gires and Min Pan, were always keen to discuss scientific ideas at any time, which allowed me to progress in my work. Further, I want to thank Kristian Unger, Martin Selmansberger, Julia Hess, and Daniel Samaga for helping me with mathematical/ statistical problems and *R*-coding. Philipp Baumeister greatly supported me with his medical knowledge as a physician and was always open to answer any clinical question from my side. Vera Kohlbauer and Alexandra J. Blancke Soares helped me a lot with sharing experiences regarding experiments. I want to thank Jiefu "Jeff" Zhou and Merve Akyol, my favorite student, for their support with experiments. Gisela Kranz, Darko Libl, and Sabina Schwenk-Zieger have been a great support regarding tumor sample collection, processing, and staining. Finally, I want to thank my girlfriend, family, and close friends; Leonie, Lennart, Renate, Christian, Lucas, Nici, Bubu, Kaan, Dominik, Annika, Mischi, Roger, Annette, and all others for motivating me to become a decent scientist.

List of publications

Thesis related

- Pan, M.* and Schinke, H.* *et al.*, EpCAM ectodomain EpEX is a ligand of EGFR that counteracts EGF-mediated epithelial-mesenchymal transition through modulation of phospho-ERK1/2 in head and neck cancers. *PLoS Biology*, 2018, *; *MP and HS contributed equally to this work.*

Other contributions

- Gires, O., Pan, M., Schinke, H. *et al.*, Expression and function of epithelial cell adhesion molecule EpCAM: where are we after 40 years? *Cancer metastasis reviews*, 2020
- Sharaf, K., Kleinsasser, A., Schwenk-Zieger, S., Gires, O., Schinke, H. *et al.*, Molecular Characterization of Lipoaspirates Used in Regenerative Head and Neck Surgery. *JAMA Facial Plast. Surg.*, 2019

Manuscripts in preparation / under review

- Schinke, H.* and Heider, T.* *et al.*, Digital scoring of EpCAM and Slug expression as prognostic markers in head and neck squamous cell carcinomas. Re-submitted: *Molecular Oncology*, 2020, *; *HS and TH contributed equally to this work.*
- Schinke, H. *et al.*, Partial epithelial-to-mesenchymal transition is prognostic and associates with Slug in head and neck cancer. Pre-print available at bioRxiv: doi:10.1101/2020.10.20.346692, Under review: *Molecular Oncology*, 2020

Zusammenfassung

Bei Plattenepithelkarzinomen des Kopf-Hals-Bereiches (HNSCCs) wurde die Tumorheterogenität als eine Hauptursache für das Fortschreiten der Krankheit beschrieben. Der Epithelial-zu-Mesenchymal-Übergang (EMT) wurde als eine prominente Ursache der molekularen und zelluläre Heterogenität in Tumoren identifiziert. Bei humanem Papillomavirus (HPV) negativen HNSCC Patienten gibt es gegenwärtig keine klinisch akzeptierten molekularbiologischen Prognostikatoren, die in die Therapieentscheidung mit einbezogen werden. Der epidermale Wachstumsfaktorrezeptor (EGFR) wird bei HNSCC-Patienten häufig überexprimiert und ist Zielmolekül für eine adjuvante Behandlung. In früheren Arbeiten haben wir und andere die anhaltende Aktivierung der EGF/EGFR-Achse als Induktor der EMT durch die zytoplasmatischen Kinasen ERK1/2 beschrieben. Wir zeigten, dass die lösliche Ektodomäne von EpCAM, EpEX, in der Lage war, die EMT-Induktion zu blockieren. Ein Ziel der vorliegenden Arbeit war die Genregulation der EGF/EGFR-Achse mit Assoziation zum EMT zu untersuchen und den EMT-blockierenden Einfluss von EpEX auf Transkriptomebene zu definieren. In der vorliegenden Arbeit wurden RNA-Sequenzierungsdaten von EGF-behandelten Krebszelllinien generiert, analysiert, und die anhaltende Aktivierung der EGF/EGFR-Achse als Induktor von EMT bestätigt. Weiterhin wurde eine EGFR-vermittelte EMT Gensignatur von n = 172 hoch- und runter-regulierten EMT-Genen aus den Transkriptomdaten bioinformatisch extrahiert. Mit Hilfe einer forward-feature-selection Methode konnten fünf überlebensassoziierte Gene bei HPV-negativen HNSCC-Patienten der publizierten "The Cancer Genome Atlas" (TCGA, n = 240) Kohorte aus der EGFR-vermittelten EMT Gensignatur identifiziert werden. Die Implementierung der fünf überlebensassoziierten Gene in ein multivariates Cox-Modell ermöglichte es, die Genexpression zu gewichten und so einen prognostischen EGF/EGFR-vermittelten EMT-Risiko-Score mit Assoziationen zu Lymphknoteninfektionen und zum klinischen Stadium des Tumors zu berech-

nen. Der prognostische Wert des EGF/EGFR-vermittelten EMT-Risiko-Scores wurde in zwei unabhängigen und publizierten HPV-negativen HNSCC-Kohorten des "MD Anderson Cancer Center" (MDACC, n = 97) und des "Fred Hutchinson Cancer Research Center" (FHCRC, n = 62) bestätigt. Darüber hinaus konnte gezeigt werden, dass EpEX auf Transkriptomebene kein Repressor der EGF-vermittelten EMT ist und im Vergleich zu EGF ähnliche, aber reduzierte Genregulationsfähigkeiten, zeigt.

In einem zweiten Teilabschnitt der Promotionsarbeit wurde der Einfluss einer EMT Sonderform, der partiellen EMT (pEMT), auf die Tumorerkrankung untersucht. Vorherige Studien zeigten, dass insbesondere bei HNSCC-Patienten des basal-ähnlichen mesenchymalen Subtyps die pEMT mit inter- und intratumoraler Heterogenität assoziiert ist. Unter Verwendung einer publizierten Signatur von pEMT-Genen und der Methode "Single sample scoring molecular phenotype" (SING) wurde die pEMT-Genexpression im HNSCC des basal-ähnlichen und mesenchymalen Subtyps zur Berechnung eines SING Scores zur Quantifizierung der p-EMT herangezogen. Dieser neue pEMT SING-Score war in zwei unabhängigen HPV-negativen HNSCC-Kohorten (TCGA und MDACC) prognostisch, konnte Patienten auf der Grundlage des Gesamtüberlebens stratifizieren, war positiv mit der Lymphknotenmetastasierung korreliert und stark mit der Expression des kanonischen EMT-Transkriptionsfaktors (EMT-TF) SLUG korreliert. Durch exogene Überexpression von SLUG in zwei Krebszelllinien der Kopf- und Halsregion wurde ein zellulärer pEMT-Phänotyp mit erhöhtem invasivem Potential und erhöhter Strahlenresistenz induziert. Der prognostischte Wert der SLUG-Proteinexpression bei HPV-negativen Patienten wurde in einer separaten Kohorte unserer Klinik durch immunohistochemische Quantifizierung von SLUG in kryo-konservierten HNSCC Proben (n = 76) bestätigt. Eine hohe Expression von SLUG war mit einem Tumorrezidiv, einer Lokalisation an den Tumorrändern und einem schlechteren krankheitsfreien Überleben assoziiert.

Die vorgestellten Untersuchungen führen zu der Schlussfolgerung, dass die EGF/EGFR-Achse einen Induktor der EMT bei HPV-negativen HNSCC-Patienten mit Auswirkungen auf den Krankheitsverlauf darstellt. In vorheriger Arbeit zeigten wir, dass die

EGF/EGFR-Achse die kanonischen EMT-TFs SNAIL, SLUG und ZEB1 induziert. Im zweiten Teil dieser Promotionsarbeit wurde gezeigt, dass SLUG als Mediator der pEMT fungiert und dass pEMT bei HPV-negativen HNSCCs Patienten des basal-ähnlichen mesenchymalen Subtyps mit der Schwere der Erkrankung zusammenhängt.

Abstract

In Head and Neck Squamous Cell Carcinomas (HNSCCs), tumor heterogeneity was described as a main driver of disease progression. Epithelial-to-Mesenchymal transition (EMT) was identified as a prominent cause of molecular and cellular heterogeneity in tumors. At present, in Human papillomavirus (HPV) negative HNSCC patients, no clinically accepted molecular prognosticator is available that is implemented into treatment decision making. Epidermal Growth Factor Receptor (EGFR) is frequently overexpressed in HNSCC patients and a target for adjuvant treatment. In previous work, we and others described sustained EGF/EGFR-signaling as an inducer of EMT through the cytoplasmic kinases ERK1/2. We showed that the soluble ectodomain of EpCAM, EpEX, was capable of blocking EMT induction. The aim of the presented thesis was the investigation of the EMT-associated gene regulation by the EGF/EGFR-axis and the identification of the EMT-impairing influence of EpEX at the transcriptome level. In the present thesis, next-generation RNA sequencing data of EGF-treated cancer cell lines was generated, analyzed, and sustained EGF/EGFR-signaling was confirmed as an inducer of EMT. Further, an EGF/EGFR-mediated EMT signature of n = 172 up- and down-regulated EMT genes was extracted from the transcriptome data through bioinformatic analysis. Using a forward feature selection method, five survival-associated genes in HPV-negative HNSCC patients could be identified from the EGF/EGFR-mediated EMT signature in the published "The Cancer Genome Atlas" (TCGA, n = 240) cohort. Implementing the five survival-associated genes into a multivariate Cox model allowed to weigh the gene expression and to compute a prognostic EGF/EGFR-mediated EMT Risk Score with associations to lymph node infection and tumor clinical stage. The prognostic value of the EGF/EGFR-mediated EMT Risk Score was confirmed in two independent and published HPV-negative HNSCC cohorts from the "MD Anderson Cancer Center" (MDACC, n = 97) and the "Fred Hutchinson Cancer Research Center" (FHCRC, n = 62). Further,

at the transcriptome level, EpEX is not a repressor of EGF-mediated EMT and shows similar, but reduced, gene regulation capabilities compared to EGF.

In a second part of this thesis, the influence of a particular form of EMT, so called partial EMT (pEMT), on tumor progression was investigated. For HNSCC patients of the basal-like mesenchymal subtype, pEMT was shown to be associated with inter- and intratumoral heterogeneity by others. Using a published signature of common pEMT genes and the method "Single sample scoring molecular phenotype" (SING), pEMT gene expression in HNSCC of the basal-like and mesenchymal subtype was incorporated to compute a SING score to quantify pEMT. This novel pEMT SING score was prognostic in two independent HPV-negative HNSCC cohorts (TCGA and MDACC), was capable of stratifying patients based on overall survival, was positively correlated with nodal metastasis, and strongly correlated with the expression of canonical EMT transcription factor (EMT-TF) SLUG. By exogenous overexpression of SLUG in two cancer cell lines of the head and neck region, a cellular pEMT phenotype with elevated invasive potential and irradiation resistance was induced. Prognostic value of SLUG protein expression in HPV-negative patients was confirmed in a separate cohort from our clinic by immunohistochemical quantification of SLUG in cryo-conserved HNSCC samples (n = 76). High expression of SLUG was associated with tumor recurrence, localization to tumor edges, and poorer disease-free survival.

The presented investigations lead to the conclusion that EGF/EGFR-signaling represents an inducer of EMT in HPV-negative HNSCC patients that affects disease progression. EGF/EGFR-signaling was shown to induce the canonical EMT-TFs SNAIL, SLUG, and ZEB1. In the second part of the presented work, SLUG was identified as a mediator of pEMT and pEMT was related to disease severity in HPV-negative HNSCC-patients of the basal-like mesenchymal subtype.

Table of Contents

List of Abbreviations

CRISPR	Clustered regularly interspaced short palindromic repeats
Cox model	Cox proportional hazard model
CSC	Cancer stem cell
Ctrl	Control
DE	Differential expression
DEGs	Differentially expressed genes
DFS	Disease-free survival
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EMT	Epithelial-to-Mesenchymal transition
EMT-TF	EMT transcription factors
EpCAM	Epithelial cell adhesion molecule
EpEX	Extracellular domain of EpCAM
EpICD	Intracellular domain of EpCAM
ERK1/2	Extracellular signal–regulated kinase 1/2
Fc	Fragment crystallizable region
FDA	Food and Drug Administration
GLM	Generalized linear model
GO	Gene Ontology (project)
GSEA	Gene set enrichment analysis
HR	(Cox) Hazard ratio
HGFR	Hepatocyte Growth Factor Receptor
HEK293	Human embryonic kidney 293
HNSCC	Head and neck squamous cell carcinoma
HPV	Human papillomavirus
IHC	Immunohistochemistry
JAK	Janus kinase

KM	Kaplan-Meier
KRT	Cytokeratine
LMU	Ludwig-Maximilians-University
LNM	Lymph node metastasis
LM	Linear model
MEK	MAPK–ERK kinase
MET	Mesenchymal-to-Epithelial transition
NES	Normalized enrichment score
OE	Overexpression
OS	Overall survival
PCA	Principal component analysis
pAKT	Phosphorylated AKT
pEMT	Partial EMT
pERK1/2	Phosphorylated extracellular signal–regulated kinase 1/2
PLCγ	Phospholipase C γ
PKC	Protein kinase
PI3	Phosphoinositide-3
qRT-PCR	Quantitative real-time PCR
Raf	Rapidly accelerated fibrosarcoma kinase
Ras	Rat sarcoma gene
RIP	Regulated intramembrane proteolysis
RNASeq	RNA sequencing
SING	Single sample scoring molecular phenotype
STAT	Signal transducer and activator of transcription proteins
TCGA	The Cancer Genome Atlas
TGFα	Transforming growth factor α
TGFβ	Transforming growth factor β
TKI	Tyrosine kinase inhibitor
TNM	Tumor, node, metastasis
ZEB1/2	Zinc finger E-box-binding homeobox 1/2

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Head and Neck Squamous Cell Carcinoma

Head and Neck Squamous Cell Carcinoma (HNSCC) describes a group of cancers in the oral cavities, oropharynx, hypopharynx, and larynx. Risk factors do include heavy smoking, alcohol abuse, and human papillomavirus (HPV) 16 infection. Within HNSCCs, HPV-positive tumors were postulated to rely on a distinct form of disease etiology with P16 up-regulation, P53 degradation, and retinoblastoma protein (Rb) pathway down-regulation through the function of the two viral oncogenes *E6* and *E7*, respectively. In contrast, HPV-negative tumors are often characterized by *TP53*, the P53 encoding gene, mutations and P16 down-regulation (Stransky et al. 2011). Thus, HPV-negative and -positive HNSCCs are generally viewed as two distinct subgroups and HPV infection is a positive prognostic marker for overall and disease-free survival (Marur et al. 2010). Worldwide every year approximately 600,000 patients are diagnosed with HNSCC and due to high rates of intra- and intertumor heterogeneity, standard radio(-chemo)therapy often fails and patients suffer from tumor recurrence or locoregional lymphnode metastasis. Further, the functional and aesthetic relevance of the head and neck region does limit the possibilities of radical surgical approaches. Thus, overall survival rates over five years remain below 50 % (Siegel, Miller, and Jemal 2016; Ferlay et al. 2015).

1.1.1 Molecular Subtypes

Cancers defined as HNSCC cover a spatial region starting, from cranial to caudal, at the nasal cavity and ending with the larynx. Even though all carcinomas defined as HNSCC

originate from squamous epithelium and retain epithelial features to a variable degree, differences in underlying tumor biology were reported. Formerly, HPV infection was the only established molecular characterization option in HNSCC. In 2013, a genomic analysis of tissue microarray sequencing data determined four molecular classes of HNSCCs consistent with signatures previously defined for squamous carcinoma of the lung (Walter et al. 2013). Until today, the definition comprises the atypical, basal-like, classical, and mesenchymal molecular subtypes of HNSCCs. Atypical tumors were strongly associated with HPV infection, which correlated with elevated gene levels of CDKN2A, LIG1, and the transcription factor RPA2. Analysis of gene expression in the basal-like subtype showed high expression of the Tumor Growth Factor TGFα and Epidermal Growth Factor Receptor (EGFR), the transcription factor TP63 and an extracellular matrix associated gene called COL17A1. HNSCC of the classical molecular subtype were associated with the heaviest smoking history and xenobiotic metabolism gene products like AKR1C1/3 or GPX2 were found to be highly expressed. The mesenchymal subtype is characterized by increased expression of mesenchymal markers, such as VIMENTIN (VIM), the transcription factor TWIST1, and the Hepatocyte Growth Factor (HGF). Regarding clinical parameters, the described molecular tumor subtypes do not correlate with age, alcohol consumption, gender, or tumor size (Walter et al. 2013).

Essentially, those findings confirmed previous work based on microarray analysis of n = 60 HNSCC samples that suggested the above-mentioned four subtypes in HNSCCs (Chung et al. 2004).

In 2015, The Cancer Genome Atlas (TCGA) accomplished a comprehensive transcriptome-wide multi-platform characterization of n= 279 tumor samples by RNA sequencing (RNASeq) (Lawrence et al. 2015). Using the sequencing data, the authors were able to segregate the cohort into the four molecular subtypes: atypical (24%), basal-like (31%), classical (18%), and mesenchymal (27%). The computational subtyping of this cohort *via* tumor bulk sequencing classified nearly all HPV-positives patients with oropharyngeal cancers as atypical. The oral cavity sublocalization group consisted, in large proportions, of HPV-negative patients classified as basal-like or mesenchymal subtypes. HPV-negative

patients suffering from a larynx carcinoma were evenly distributed between the atypical and classical molecular subtypes.

It is worth noting that recently published work based on single-cell RNA sequencing is now challenging this standard classification into four molecular subtypes. The authors found that the mesenchymal classification might essentially be caused by large proportions of non-malignant mesenchymal cells, such as fibroblasts or immune cells, within the analyzed tumor bulks. By deconvolution approaches of the bulk TCGA data sets based on multiple linear regression analysis to circumvent influences of non-malignant cells, the basal-like and mesenchymal subtypes fall together into one single cluster re-defined, by the authors, as malignant-basal (Puram et al. 2017). Nonetheless, this re-classification of subtypes within HNSCC needs to be validated by further independent investigations for its final establishment.

1.2 Epidermal Growth Factor Receptor in HNSCC

EGFR is a 170 kDa cell surface receptor, belongs to the prominent ErbB/HER-family, and is known to be involved in the regulation of cell proliferation, migration, differentiation, and survival (Kalyankrishna and Grandis 2006). Constitutive activation or overexpression are frequently found in a variety of cancer types, like lung or HNSCCs, leading to poor prognosis for patients (Lawrence et al. 2015; McKenzie 1991; Santini et al. 1991). Therefore, EGFR is a prominent target for therapeutic approaches using *e.g.* specific antibodies like Cetuximab or selective chemical inhibitors blocking enzymatic activities of this receptor tyrosine kinase (Liang, Zhang, and Zhang 2020). In HNSCCs, function altering EGFR mutations are rare. Mutated EGFR alleles were reported in 4.7 % of HNSCC TCGA tumors but around 15 % of HPV-negative TCGA HNSCC tumors were characterized by allele amplification of EGFR (Lawrence et al. 2015). Further, EGFR may be found endocytosed into the nucleus triggered by ligand or Cetuximab binding, or exposure to irradiation (Zhang et al. 2015). Nonetheless, most attention was paid to the membrane-bound form of EGFR with its activity regulated by the quantity

of available ligands and receptor molecules, as it is overexpressed in between 80-100 % of HNSCCs (Dassonville et al. 1993). The overexpression of EGFR is associated with decreased survival of cancer patients (Byeon, Ku, and Yang 2019; Kalyankrishna and Grandis 2006). Even though a prime candidate, monotherapy with EGFR inhibitors, such as Cetuximab or Erlotinib, a selective Receptor Tyrosine Kinase (RTK) inhibitor, show no superior treatment success compared to standard of care (Vermorken et al. 2007; Cohen 2014).

Upon binding of ligands, EGFR homo- or heterodimerizes with other HER receptors or RTKs like Hepatocyte Growth Factor Receptor (HGFR). Known ligands of the EGFR receptor include: EGF, TGFα, heparin-binding EGF, and EpCAM extracellular domain (EpEX) (Pan et al. 2018; Byeon, Ku, and Yang 2019). Once ligands have bound, the dimerization of EGFR molecules activates signaling cascades such as the pathways of Mitogen Activated Protein Kinase (MAPK), Phosphoinositid-3-kinase (PI3K)/Protein Kinase B (PKB/Akt), Janus Kinase (JAK)/ Signal Transducer and Activator of Transcription Proteins (STAT), and the Phospholipase C γ (PLCγ)/Protein Kinase C (PKC) (Kalyankrishna and Grandis 2006).

1.2.1 EGFR-signaling in cancer

Canonical EGFR pathways are involved in cell differentiation, proliferation, and survival. In different cancer types, the EGF/EGFR-axis was shown to contribute to disease progression by diverse signaling cascades with different outcomes, *e.g.* cellular proliferation or transformation (Normanno et al. 2006). The deregulation or activation of the MAPK/ Extracellular Signal-Regulated Kinase (ERK)- and PI3K/Akt-pathway leads to cellular proliferation and survival (Horn et al. 2015; Affolter et al. 2016). The MAPK/ERK-pathway is constitutively activated in different tumor types. The pathway was shown to be up-regulated in cancer cell lines derived from pancreas, colon, lung, ovary, and kidney tumors (Hoshino et al. 1999). Up-stream of MAPK/ERK, dimerized EGFR activates Raf by binding to the small GTPase Ras. The activated Ras/Raf-complex then phosphorylates MEK (MAPK/ERK-activating kinase), which itself then phosphorylates

MAPK/ERK. Furthermore, the MAPK/ERK-pathway possesses several capacities as its activation can lead to different cellular morphologies and behaviors. An intermediate activation by EGFR ligands, such as EGF, causes transient phosphoERK (pERK) levels leading to an induction of proliferation of epithelial cells. Oppositely, a strong and sustained ERK2, not ERK1, activation induces a reduction in proliferation and activates distinct cellular reprogramming inducing a mesenchymal phenotype demonstrated in HNSCC, human breast, and mouse fibroblast cell lines (Pan et al. 2018; Shin et al. 2010; Murphy, MacKeigan, and Blenis 2004).

For the PI3K/Akt-pathway, activated EGFR directly stimulates PI3K *via* adapter molecules such as the Insulin Receptor Substrate (IRS) proteins or through Ras. This triggers conversion of Phosphatidylinositol (3,4)-Bisphosphate (PIP2) lipids to Phosphatidylinositol (3,4,5)-Trisphosphate (PIP3) by the catalytic domain of PI3K. Then, membrane-bound PIP3 activates Akt *via* phosphorylation leading to activation of protein synthesis and proliferation (Alessi et al. 1997).

Additionally, EGFR activation can trigger PLCγ phosphorylation, which then phosphorylates PKC. Both pathways, PI3K/Akt and PLCγ/PKC ultimately modulate gene translation through mTORC1 (Fan et al. 2009; Koyama et al. 2003).

Further, EGFR recruits and activates JAK1/2 to phosphorylate STAT1 and STAT3. STAT1 and STAT3 then form heterodimers and translocate to the nucleus to activate gene transcription. It has been shown that STAT1/3 activity promotes cell migration *in vitro* (Andl et al. 2004).

As aforementioned, solely targeting EGFR in cancer therapy frequently fails to provide ample benefits to patients and it seems likely that HNSCC cancer cells develop an adaptive response against therapeutic treatments by up-regulation of survival pathways such as the MAPK/ERK-signaling pathway (Rong et al. 2020; Rampias et al. 2014).

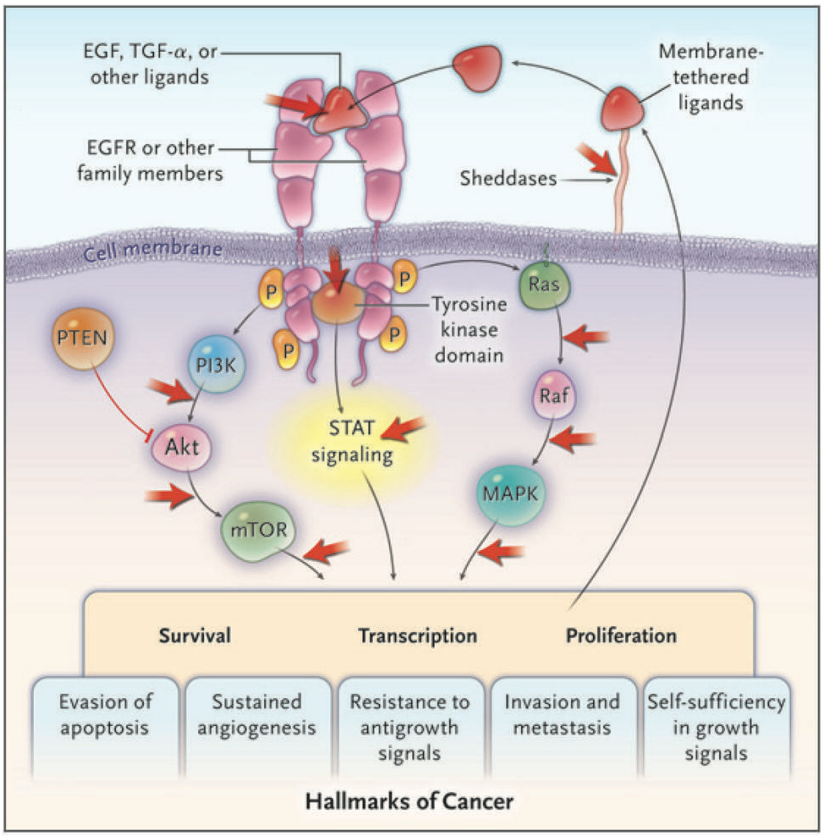

Figure 1.1: Illustration of canonical EGFR-signaling pathways and cellular effects. Effects on Hallmarks of cancer are shown. Only nodal points in the pathway are displayed, indicating targets of therapies in current clinical use or in trials (red arrows). EGF/R: Epidermal growth factor/receptor, TGFα: Transforming growth factor α, PTEN: Phosphatase and Tensin homolog, PI3K: Phosphoinositide-3 kinase, Akt: Proteinkinase B, mTOR: Mammalian target of rapamycin, STAT: Signal transducer and activator of transcription proteins, Raf: Rapidly accelerated fibrosarcoma kinase, Ras: Rat sarcoma gene. MAPK: Mitogen Activated Protein Kinase. Adapted from Gazdar, The New England Journal of Medicine, 2009

1.2.2 Signaling crosstalk with Epithelial Cell Adhesion Molecule

Epithelial Cell Adhesion Molecule (EpCAM) was first described as an antigen expressed in colon carcinomas and identified as a homophilic cell-to-cell adhesion molecule (Herlyn et al. 1979; Litvinov et al. 1994). In the following, a variety of functions, besides cell adhesion, were assigned to EpCAM in cancer and stem cells including promotion of proliferation, epithelial differentiation, and an induction of multipotency in mesenchymal stem cells and, in part controversially to findings in HNSCCs, an induction of Epithelial-to-Mesenchymal transition in endometrial cancer through EGFR-signaling (Osta et al. 2004; Münz et al. 2004; Kuan et al. 2019; Hsu et al. 2016; Gires et al. 2020). High expression of EpCAM was identified as a poor prognosticator, among others, in colorectal, prostate, and breast cancer but associated with increased survival in colonic, esophageal, and HNSCCs (Spizzo et al. 2004; Seeber et al. 2016; Massoner et al. 2014; Went et al. 2006; Kimura et al. 2007; Baumeister et al. 2018).

Cellular signaling and degradation of EpCAM is triggered by a processes termed Regulated Intramembrane Proteolysis (RIP). EpCAM RIP is a sequential cleavage and involves initial shedding of the extracellular domain of EpCAM, called EpEX, by α-secretase (ADAM) and β-secretase (BACE) sheddases. The shedding of EpEX generates a membrane-tethered C-terminal fragment termed EpCTF. The EpCTF fragment is then cleaved by a γ-secretase complex, which generates an extracellular EpCAM-Aβ-like fragment and an intracellular EpICD fragment (Maetzel et al. 2009; Tsaktanis et al. 2015). The EpICD fragement was shown to be involved in nuclear signaling in conjunction with components of the WNT-signaling pathway including Four and a Half LIM domain protein 2 (FHL2), β-catenin, and Lef-1 (Maetzel et al. 2009). The extracellular domain of EpCAM, EpEX, was recently shown to be capable of inducing EGFR-signaling in HNSCC and colon cancers (Pan et al. 2018; Liang et al. 2018). In both cancer types, EpEX could induce phosphorylation of ERK1/2 and Akt. In colon cancer cell lines, shedded EpEX triggered a positive feedback loop of EpCAM proteolysis leading to nuclear localization of EpICD, causing metastasis and poor survival (Liang et al. 2018). In HNSCC, this feedback loops was not reported and high expression of EpCAM in combination with low

levels of EGFR marked a subgroup of patients with exceptionally high survival, independent of the HPV infection status (Pan et al. 2018). Further, *in vitro* studies of HNSCC cell lines led to the conclusion that EpEX is capable of impairing EGF/EGFR-mediated induction of Epithelial-to-Mesenchymal transition (EMT) (Pan et al. 2018).

1.3 Epithelial-to-Mesenchymal transition

EMT is a major cellular differentiation program, which was first described in 1982 by Elizabeth Hay and colleagues in embryonic development (Greenburg and Hay 1982). During this cellular program, epithelial cells loose apical-basal polarity, cell-to-cell contacts, and, simultaneously, gain a spindle-shaped morphology. On a molecular level, expression of epithelial markers, such as E-CADHERIN (ECAD) or Cytokeratins (KRT), is lost and mesenchymal markers, such as VIMENTIN (VIM) or FIBRONECTIN1 (FN1), become expressed (Nieto 2009; Thiery et al. 2009; Kalluri and Weinberg 2009). Under physiological conditions, EMT is a cellular program that is crucial for embryogenesis and wound healing, including tissue regeneration. These two forms are also classified as EMT type I, appearing during embryonic development, and type II that is helping the organism to close wounds after trauma. The type III EMT describes the malignant form in carcinoma progression (Kalluri and Weinberg 2009).
It has been postulated that EMT might be primarily regulated as an epigenetic process and would therefore be rather independent of DNA alterations, such as mutations (Tam and Weinberg 2013; Dongre and Weinberg 2019). Thus, cancer biology assessments related to the EMT program should rather be addressed by transcriptome or proteomic analysis.

So far, six canonical EMT Transcription Factors (TFs) were identified, namely SNAI1 (SNAIL), SNAI2 (SLUG), TWIST1, TWIST2, ZEB1, and ZEB2. These EMT-TFs act as transcriptional repressors to facilitate gene regulation related to cell adhesion, migration, and invasion (Lamouille, Xu, and Derynck 2014; Peinado, Olmeda, and Cano 2007; Bolós et al. 2003). SNAIL and SLUG belong to the Snail family that encompasses the

Figure 1.2: Schematic representation of the major changes during EMT. Epithelial cells (blue) with E-CAD expression start expressing EMT transcription factors SLUG and ZEB1. E-CAD expression is decreased and mesenchymal marker VIM expression is increased. Finally, the EMT program leads to a transition from an epithelial to a mesenchymal cell (orange) without tight cell-cell contacts as shown. EMT: Epithelial-to-Mesenchymal transition, E-CAD: E-CADHERIN, VIM: VIMENTIN.

three members SNAIL, SLUG, and SNAI3. From the Snail family, SNAI3 was the most recent member to be discovered and the knowledge about the structure and functional implementations of human SNAI3 is fairly limited to this date. The Snail family shares a common organization as zinc-finger transcriptional repressors with four to six zinc-fingers (Cys2-His2-type) and a non-conserved N-terminus containing a SNAG domain of seven to nine conserved amino acids. DNA elements containing the E2-box type elements C/A(CAGGTG) are recognized by Snail family members and bound by the zinc-fingers. The repressor capacity is determined by the SNAG domain (Nieto 2002). Through recruitment of the co-repressor CtBP-1, the SNAG domain recruits histone modifying enzymes to silence gene expression (Tripathi et al. 2005; Lin et al. 2010). The central region of Snail proteins is highly divergent, with SNAIL containing a serine/proline-rich region, whereas SLUG contains the name giving SLUG domain, which acts as a negative modulator of SLUG-mediated EMT (Peinado, Olmeda, and Cano 2007; Molina-Ortiz et al. 2012). TWIST1 and TWIST2 encode basic helix-loop-helix (bHLH) transcription factors that are active under non-malignant conditions during embryonic development. TWIST2 lacks a glycine rich domain found in TWIST1, otherwise both TFs share high amino acid similarity of up to 95 % in the bHLH domain and 100 % in the Twist box (Franco et al. 2011). TWIST1/2 form homo- and heterodimers that bind to DNA E-box sequences regulating gene products involved in *e.g.* skull development (Chen and

Behringer 1995). ZEB1 and ZEB2 contain N- and C-terminal zinc-finger clusters and an E-box binding homeobox (Stemmler et al. 2019). All six canonical EMT-TFs have been shown to be involved in down-regulating ECAD and are associated with the up-regulation of mesenchymal genes, such as VIM, FN1, or N-CADHERIN (NCAD).

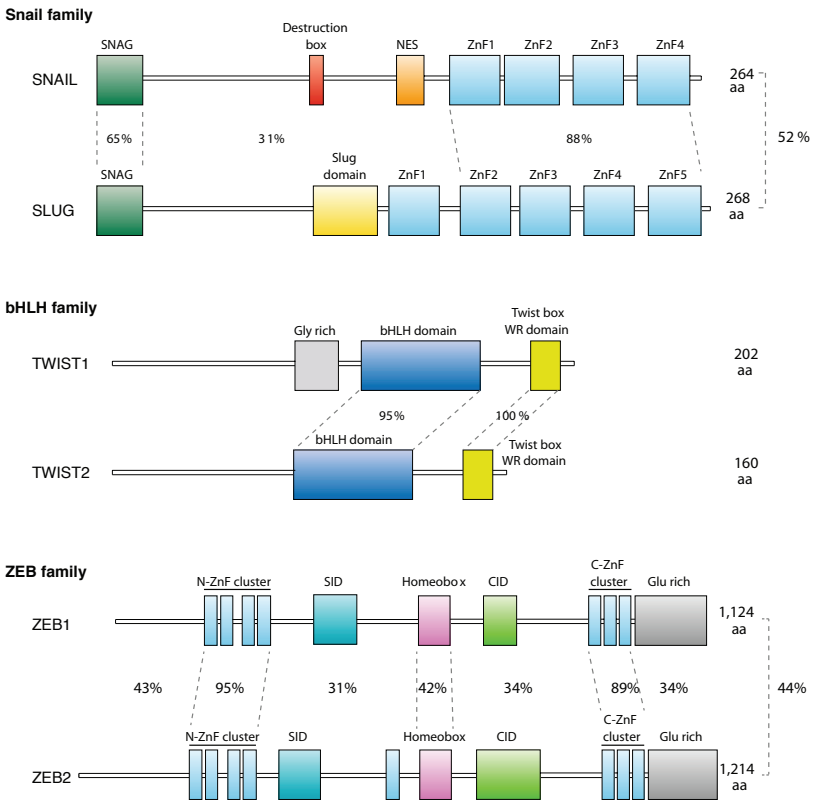

Figure 1.3: Schematic representation of the protein structures of the core EMT-TFs. ZEB1 and ZEB2 are much larger than the other EMT-TFs. Thus, their domain schematics are not drawn to the same scale as those of SNAIL, SLUG, TWIST1, and TWIST2. DNA-binding domains are represented in blue. aa: Amino acids, bHLH: Basic helix–loop–helix, CID: CtBP interaction domain, NES: Nuclear export sequence, SID: Smad interaction domain, SNAG: Snail corepressor binding domain, EMT-TFs: Epithelial-to-Mesenchymal transition transcription factors. Adapted from Stemmler *et al.*, Nature Cell Biology, 2019

1.3.1 The role of EMT in cancer

During cancer progression, cancer cells need to permanently adapt to changing and often hostile conditions. Metastases formation, local invasion, and tumor recurrence require cancer cell motility and resistance to irradiation and chemotherapeutic agents. A potential connection of EMT and tumor progression was established 25 years ago (Hay 1995), and EMT-TFs were identified as important mediators of cellular plasticity (Chaffer et al. 2016, 2013). Type III EMT was shown to be linked to tumorigenesis, invasion, cancer cell stemness, metastasis, and treatment resistance and, thereby, leads to tumor cell plasticity (Lambert, Pattabiraman, and Weinberg 2017; Nieto et al. 2016). EMT-TFs were reported to be involved in double-strand DNA repair, induction of a pro-survival and anti-apoptotic phenotype, and up-regulation of immunosuppressive cytokines (Brabletz et al. 2001; Nieto et al. 2016). In general, various signaling pathways have been described to be capable of inducing the EMT program in cancer cells. Until today, the canonically pathways include TGFβ, RHO-like GTPases, and as aforementioned the EGFR-mediated MAPK and PI3K/Akt pathways (Pan et al. 2018; Lamouille, Xu, and Derynck 2014; Xu, Lamouille, and Derynck 2009; Derynck, Muthusamy, and Saeteurn 2014). Further, stress conditions, such as hypoxia, have been identified as EMT inducers (Zhang et al. 2013). The (EGFR/)MEK/ERK-signaling cascade inducing EMT was found to be triggered by EGF and FGF. Additionally, TGFβ can synergize with EGF to execute the EMT program (Tashiro et al. 2016; Tian et al. 2007; Uttamsingh et al. 2008; Shirakihara et al. 2011).

Many findings about EMT have been provided by cell culture and animal model studies focusing on the expression of particular marker/ effector genes. EMT represents a complex and plastic cellular program, which makes it challenging to define optimal experimental conditions for its studies. This might be, in part, an explanation why the involvement of EMT in cancer progression is of ongoing debate. In 2012 for example, an *in vivo* mouse experiment modeling skin carcinogenesis with squamous cell carcinomas found that EMT promotes local tumor invasion and needs to be reversible for the establishment of macrometastases but seems not to be activated in distant metastases

(Tsai et al. 2012). Further, a lineage-tracing mouse model of spontaneous breast-to-lung metastasis suggested that EMT is not required for metastasis but contributes to recurrence of lung metastasis after chemotherapy (Fischer et al. 2015). A mouse model of pancreatic cancer suggested that SNAIL1- and TWIST1-induced EMT is dispensable for the dissemination of primary tumors but is necessary for chemoresistance (Zheng et al. 2015). However, appending examinations revealed concerns whether those findings are fully sustainable. The mouse model study of breast-to-lung metastasis described FSP1 as a "critical gatekeeping gene" of EMT but all stages of type I EMT were reported in FSP1-knockout mice (Ye et al. 2017). The mouse model of pancreatic cancer was showing a reduction of mesenchymal marker α-SMA in TWIST1-knockout mice but α-SMA expression was shown to be rarely induced upon EMT in this mouse model (Aiello et al. 2017). For HPV-negative HNSCCs, recent work with patient data incorporating a variety of previously published EMT signatures revealed the clinical importance of EMT. This confirmation of clinical relevance in HNSCC tumors was accomplished by the implementation of seven different EMT signatures in two independent cohorts of HPV-negative patients (Heijden et al. 2020).

So far, it should be noted that EMT, as a definition of a biological program, might be regarded as a simplification of several interacting and multi-layered processes, as *e.g.* cancer stemness, Mesenchymal-to-Epithelial transition (MET), and immunosurveillance are often found to be linked to EMT-related processes. Thus, latest investigations in cancer research start to refrain from describing EMT as a simple conversion with distinct steps from an epithelial to a mesenchymal cell but focus more on the underlying processes and transcriptional changes upon the transition.

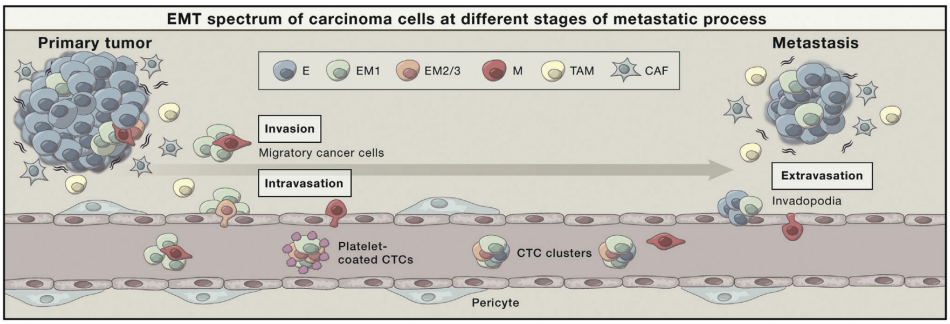

Figure 1.4: Illustration of the metastatic cascade and the involvement of EMT. EMT is a focal event in the primary tumor that may occur when epithelial carcinoma cells (E) interact with CAFs or TAMs. The majority of CTCs exhibit a partial EMT phenotype (EM1 or EM2/3). CTCs can be derived from carcinoma cells that undergo *in situ* EMT in the primary tumor, or they can acquire partial EMT phenotypes in the bloodstream, when exposed to TGFβ derived from associated platelets. Mesenchymal cells undergo MET to colonize their new destination. The plasticity of the carcinoma cells is important to avoid death in the different stages of tumor progression. CAFs: cancer-associated fibroblasts, TAFs: tumor-associated macrophages, CTCs: circulating tumor cells, TGFβ: Tumor growth factor β, EMT: Epithelial-to-Mesenchymal transition, MET: Mesenchymal-to-Epithelial transition. Adapted from Nieto *et al.*, Nature Cell Biology, 2016

1.3.2 The spectrum of EMT in malignancy

During embryonic development, EMT induces a full conversion of epithelial to mesenchymal cells that is required for the formation of the three germ layers of differentiating embryos. During cancer progression, instead of acquiring a complete mesenchymal phenotype, epithelial tumor cells often undergo an incomplete or partial EMT (pEMT) (Kalluri and Weinberg 2009; Yang and Weinberg 2008; Micalizzi, Farabaugh, and Ford 2010). Hereby, cancer cells are co-expressing epithelial and mesenchymal markers. Those cancer cells represent a pEMT phenotype and were reported to cause an elevated risk of metastasis and to show increased resistance to therapeutic drugs in contrast to cells that acquired a fully mesenchymal phenotype (Saitoh 2018). In basal breast cancer for example, hybrid epithelial-mesenchymal cells defined by CD44/CD104 (ITGβ4) expression represent a more tumorigenic phenotype compared to completely mesenchymal cancer cells (Kröger et al. 2019). For HNSCCs of the basal-like and mesenchymal subtype, single cell RNASeq from patient samples allowed to establish a pEMT state of cancer cells that functioned as an independent predictor of nodal metastasis, tumor grade, and adverse pathologic features. Further, for this form of pEMT a gene signature could be identified but, so far, no prognostic values of the pEMT signature was determined (Puram et al. 2017). Postulated mesenchymal markers in HPV-negative HNSCCs were TGFβI, VIM, FN1, PDPN, LAMB3, and LAMC2. Identified epithelial markers were ECAD, EpCAM, SFN, and KRT5, KRT14, KRT15, KRT17, and KRT18 (Puram et al. 2017).
The ongoing debate about the influence of EMT in malignancies is leaning more and more towards a clinical importance of pEMT, as different subpopulations of cells in different EMT states were found to co-exist *in vivo* and showed varying abilities of proliferation, invasion, and metastasis (Pastushenko et al. 2018; Shibue and Weinberg 2017). Recent observations could show that a complete EMT program, induced by *e.g.* TGFβ treatments under cell culture conditions, might rather be non-physiological in cancer cells and is possibly restricted to *in vitro* experiments (Dongre and Weinberg 2019; McFaline-Figueroa et al. 2019). Further, it was suggested that the process of EMT in tumors might reflect a dynamic spectrum of transcriptional changes of cancer cells that cannot

be adequately partitioned into separate EMT stages (McFaline-Figueroa et al. 2019; Dijk et al. 2018; Krishnaswamy et al. 2018). Taken together, the clinical relevance of (p)EMT in cancer as a grounding process for several steps of malignancy development is becoming increasingly apparent but the EMT process is a diverse cellular program which needs further refinement for different cancer subtypes and signaling pathways.

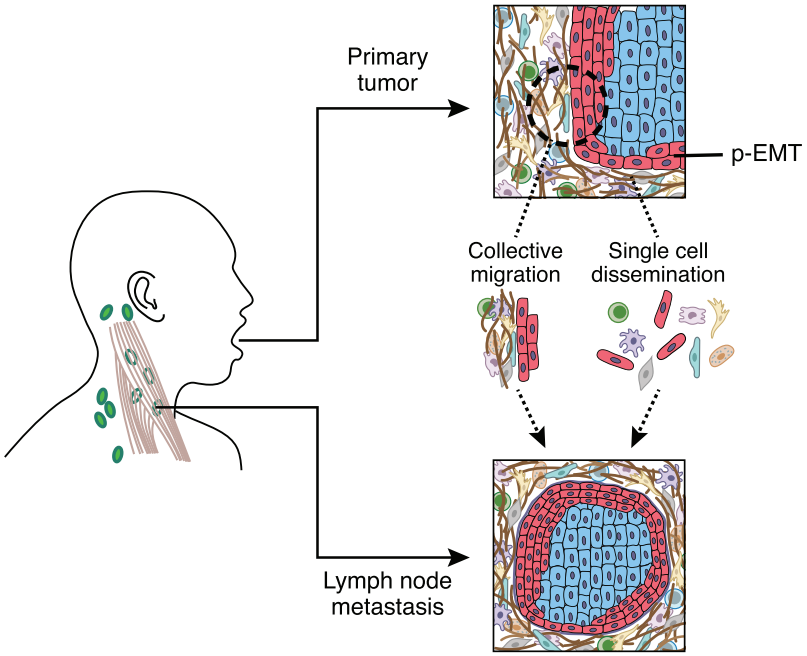

Figure 1.5: Model of the *in vivo* pEMT program associated with invasion and metastasis in basal-like and mesenchymal HNSCC tumors. pEMT cells within a heterogenous tumor are allocated at the tumor edges. Through single-cell dissemination or collective migration tumor cells are capable of invading surrounding connective tissue and forming lymph node metastases. pEMT: Partial EMT, HNSCC: Head and neck squamous cell carcinoma. Adapted from Puram *et al.*, Cell, 2017.

1.4 Aim of this thesis

Even though great improvements were achieved in cancer care, most routine treatments implemented in clinical routine are based on irradiation and non-targeted chemotherapeutic agents, such as cisplatin or 5-fluoruracil, leading to severe side effects in a patient. Therefore, a diagnosis with cancer is still a major threat and a leading cause of death world wide. Recent advantages in technology aim at introducing personalized medicine into health care. Hereby, high-throughput sequencing approaches shall be incorporated to apply a tailor-made treatment approach for each patient.

HNSCCs patients frequently suffer from lymph node metastasis and locoregional tumor recurrence resulting in poor prognosis (Siegel, Miller, and Jemal 2016; Ferlay et al. 2015). The process of EMT enables disease progression but its particular contributions to tumor aggressiveness, recurrences, and metastasis are still not fully understood. Thus, in HPV-negative patients no clinically accepted prognostic biomarker is available and clinical tumor classifications, such as TNM-staging, are serving as a sole base for treatment decision (Heijden et al. 2020). This work shall contribute to the refinement of the measurement and usage of EMT as a parameter in HPV-negative HNSCC prognosis. In HPV-negative HNSCCs, EGFR is a major receptor with prognostic value and the capability of the EGFR-signaling axis to induce EMT has been demonstrated in previous work. In the present work, a transcriptome analysis of EGF/EGFR-mediated EMT in cancer cell lines will be used to shed light on this particular signaling axis of EGFR. Further, the signaling crosstalk of EpCAM, through EpEX, with EGFR and its influence on the EMT process at the transcriptome level shall be defined. The findings from this *in vitro* study on HNSCC cell lines shall then be translated into patient data with the aim to improve stratification tools. Publicly available transcriptome sequencing data from patient cohorts will be investigated to identify survival-associated genes regulated by the EGF/EGFR-mediated EMT signaling axis. In a second part of this work, a published gene signature of the described pEMT in HNSCC patients of the basal-like and mesenchymal subtype will be used to quantify the degree of pEMT. The resulting

pEMT-quantifying surrogate will be implemented to assess the prognostic value of pEMT in HNSCCs. A further aim of the present thesis is the identification of a promising candidate pEMT regulator from the canonical EMT-TFs. The extracted candidate gene will then be further assessed in cell culture experiments to assess a link to a pEMT. Finally, the clinical relevance of the identified pEMT mediator gene will be addressed by the investigation of an in-house patient cohort of HPV-negative HNSCCs.

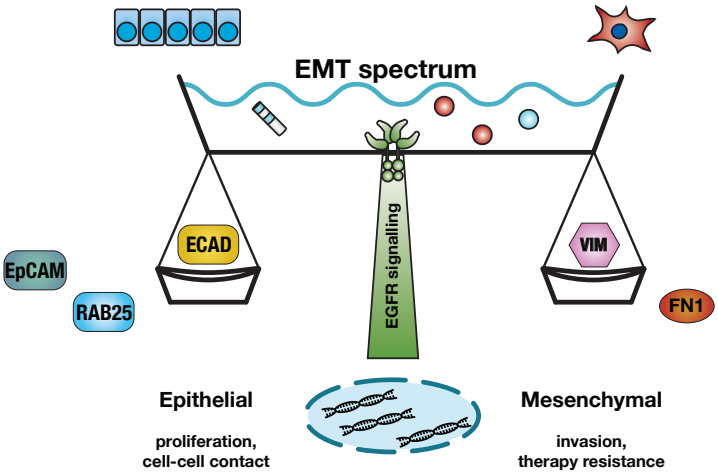

Figure 1.6: Illustration of the continuum of malignant EMT by EGFR-signaling in HNSCC. The EMT spectrum represents a fluid process. By expression of epithelial or mesenchymal building blocks, the cellular phenotype adjusts to a specific position within the spectrum. Further, the amount of a gene expressed contributes to the final cellular phenotype. EMT: Epithelial-to-Mesenchymal transition, EGF/R: Epidermal growth factor/ receptor, HNSCC: Head and neck squamous cell carcinoma, ECAD: E-CADHERIN, EpCAM: Epithelial cell adhesion molecule, VIM: VIMENTIN, FN1: Fibronectin 1, RAB25: Ras-related protein RAB25.

Chapter 2

Materials and Methods

2.1 Human samples and ethics statement

The HNSCC cohort of the Ludwig-Maximilians-University Munich (LMU) included tumor samples from n= 169 patients. For n= 87 patients macroscopically normal mucosa was available. N= 82 patients were identified by p16 staining as HPV-negative and n= 54 as HPV positive. N= 33 patients could not be clearly defined. For n= 76 HPV-negative patients a complete data set with clinical parameters was available. Clinical specimens were tested after written consent for routine procedures based on the approval of the Ethics Committee of the Medical Faculty of the LMU (#087-03;150 #197-11; #426-11) and in accordance with the WMA Declaration of Helsinki and the report of the Department of Health and Human Services Belmont.

2.2 Consumables

Table 2.1: The table shows general consumables.

Item	**Manufacturer**	**Application**
1.5 mL Tube (nuclease-free)	Costar (USA)	Miscellaneous
6-well plates	Falcon Corning (Germany)	Cell culture
8 µm 24-well inserts	Corning (Germany)	Matrigel invasion
24-well plate (flat bottom)	Nunc (Germany)	Matrigel invasion
96-well plate (flat bottom)	Nunc (Germany)	Cell culture
96-well plate (round bottom)	Nunc (Germany)	Cell culture
96-well ultra-low attachment	Nunc (Germany)	Spheroid invasion
Micro tube (1.5 mL/2 mL)	Eppendorf AG (Germany)	Miscellaneous
Quadriperm	Sarstedt (Germany)	IF microscopy
Reagent reservoirs	Costar (USA)	Miscellaneous
Safe seal tips professional	Biozym Scientific (Germany)	Miscellaneous
T25, T75, T125 easy flasks	Nunc (Denmark)	Cell culture

2.3 Chemicals

Table 2.2: The table lists general chemicals that have not been further described in the methods.

Chemical	**Manufacturer**	**Application**
ABC-Kit Vectastain® Elite® PK6100	Vector Laboratories (USA)	IHC
Acrylamide	Merck (Germany)	Western blot
Ammonium persulfate (APS)	BioRad (USA)	Western blot
β-Mercaptoethanol	Merck (Germany)	Western blot
Bovine serum albumin (BSA)	Merck (Germany)	Western blot
Complete protease inhibitor cocktail	Roche (Germany)	Western blot
Crystal Violet	Sigma (USA)	Clonogenic surv.
Dimethyl sulfoxide (DMSO)	Carl Roth GmbH (Germany)	Miscellaneous
DMEM medium	Gibco (Germany)	Cell culture
Ethylenediaminetetraacetic acid (EDTA)	AppliChem GmbH (Germany)	Miscellaneous
Fetal calf serum (FCS)	Merck (Germany)	Cell culture
Geltrex Matrix (Matrigel)	ThermoFisher (Germany)	Matrigel inv.
Chemiluminescent HRP substrate	Millipore (USA)	Western blot
Penicillin/ streptomycin	Merck (Germany)	Cell culture
Polysorbate 20 (Tween 20)	Merck (Germany)	Western blot
PowerUp SYBR Green Master Mix	ThermoFisher (Lithuania)	qPCR
RPMI medium	Gibco (Germany)	Cell culture
Tetramethylethylenediamine (TEMED)	BioRad (Germany)	Western blot
Triton-X 100	Merck (Germany)	Western blot
Tris(hydroxymethyl)aminomethane	AppliChem GmbH (Germany)	Miscellaneous
Trypsin	Merck (Germany)	Cell culture

2.4 General devices

Table 2.3: The table lists general devices that have not been further described in the methods.

Device	**Manufacturer**	**Application**
-80°C freezer	Hera Thermo Corporation (Germany)	Storage
-20°C freezer	Liebherr (Germany)	Storage
-80°C freezer	Hera Thermo Corporation (Germany)	Storage
-20°C freezer	Liebherr (Germany)	Storage
Balance	Sartorius (Germany)	Miscellaneous
EVE Cell counter	NanoEnTek (USA)	Cell culture
Centrifuge 5415 R	Eppendorf (Germany)	Miscellaneous
Centrifuge Rotanta 46 R	Sorvall (Germany)	Miscellaneous
BlueVertical PRiME BV-104	SERVA (Germany)	SDS-PAGE
Fine balance	Sartorius (Germany)	Miscellaneous
Incubator	Binder (Germany)	Cell culture
Laminar flow	Heraeus Instruments (Germany)	Cell culture
Micro tube thermomixer	Eppendorf (Germany)	Miscellaneous
MS1 Minishaker	IKA (Germany)	Vortex
NanoPhotometer	Implen (Germany)	RNA, DNA, Protein conc.
Power supply	Major Science (USA)	Western blot
Power supply	SERVA (Germany)	SDS-PAGE
SB10 omniPAGE mini	Biozym (Germany)	Western blot transfer
Shaker Polymax 2040	Heidolph (Germany)	Miscellaneous
Water bath	GFL (Germany)	Cell culture

2.5 Cell lines and treatments

FaDu and Kyse30 cell lines were obtained from ATCC and DSMZ and were confirmed by STR typing (Helmholtz Center, Munich, Germany). Kyse30 cells were maintained in RPMI 1640 and FaDu cell ins DMEM with 10% FCS, 1% penicillin/ streptomycin, in a 5% CO2 atmosphere at 37 °C. Treatment with EGF (PromoCell PromoKine, Heidelberg, Germany) and EpEX-Fc was conducted in medium after overnight serum starvation. Recombinant EpEX-Fc was produced as described (Tsaktanis et al. 2015). In brief, HEK293 cells were stably transfected with human EpEX-Fc fusion protein in a ps521 vector (kind gift from Pascal Schneider, ISREC, Switzerland) cultured under serum free conditions for EpEX-Fc purification. Then, recombinant EpEX-Fc was purified from supernatant of transfected cells after $3 - 5$ days of culture according to the protocol by Savas and colleagues (Savas et al. 2014). Used as a control treatment for EpEX-Fc treatment, recombinant Fc was purchased from Jackson ImmunoResearch, Baltimore, MD, United States. To treat cells with EpEX-Fc and Fc, 50 nM of protein were applied. To treat cells with EGF-low, 10 ng/mL corresponding to 1.8 nM were applied. To treat cells with EGF-high, 50 ng/mL corresponding to 9 nM were applied. In the co-treatment of EGF with EpEX, 50 ng/mL of EGF and 50 nM of EpEX-Fc were applied. EpEX treatments as mentioned within this dissertation are an abbreviation for EpEX-Fc treatments. These definitions were used throughout the whole extent of this thesis.

2.6 Immunohistochemistry scoring and immunofluorescence

Specific antibodies against SLUG (C19G7, Cell Signaling Technology, NEB, Frankfurt, Germany, #9585, 1:400), pan-Cytokeratine (polyclonal, Invitrogen, Camarillo, USA, #18-0059, 1:200) and E-CADHERIN (24E10, Cell Signaling Technology, NEB, Frankfurt, Germany, #3195, 1:400) were used for immunohistochemistry (IHC) with the avidin-biotin-peroxidase method (Vectastain, Vector laboratories, Burlingame, CA, US) or immunofluorescence (IF) staining in combination with Alexa Fluor-488-conjugated secondary antibody. Confocal microscopy images were recorded with a TCS-SP5 sys-

tem (Leica Microsystems; Wetzlar, Germany). IHC scores were formed using a two-parameters system, which implemented scoring of all specimen in percentages of (tumor) cells and their staining intensities from 0 to 3 (0= negative, 1= mild intensity, 2= moderate intensity, 3= strong intensity, score = sum(% x intensity); resulting max. score 300) as described (Mack and Gires 2008). At least two experienced scorers evaluated IHC specimen independently and blinded for the specimen's identity and clinical outcome.

2.7 Reverse transcription qPCR analysis

In accordance to Schinke *et al.*, total RNA was extracted using RNeasy Mini kit (Qiagen) and reverse transcribed with QuantiTect Reverse Transcription kit (Qiagen) (Schinke et al. 2020). The resulting cDNAs were used for analysis with SYBR-Green master mix and LightCycler480 (Roche) or QuantStudio3 (ThermoFisher). If quantifications were exceeding a cycle threshold (CT) of 35 the respective gene was regarded as not expressed. All quantified values were normalized to an internal GAPDH control. The relative expression value for each target gene compared to the calibrator for that target was calculated as $2^{-\Delta\Delta Ct}$ as described by Livak (Livak and Schmittgen 2001).

2.8 Primers used for qPCR quantification

E-CADHERIN-FW 5'-TGC CCA GAA AAT GAA AAA GG-3'
E-CADHERIN-BW 5'-GTG TAT GTG GCA ATG CGT TC-3'
GAPDH-FW 5'-AGG TCG GAG TCA ACG GAT TT-3'
GAPDH-BW 5'-TAG TTG AGG TCA ATG AAG GG-3'
ITGα5-FW 5'-GGC TTC AAC TTA GAC GCG GAG-3'
ITGα5-BW 5'-TGG CTG GTA TTA GCC TTG GGT-3'
LAMC2-FW 5'-CAA AGG TTC TCT TAG TGC TCG AT-3'
LAMC2-BW 5'-CAC TTG GAG TCT AGC AGT CTC T-3'
MMP10-FW 5'-TCA GTC TCT CTA CGG ACC TCC-3'

MMP10-BW 5'-CAG TGG GAT CTT CGC CAA AAA TA-3'
PDPN-FW 5'-ACC AGT CAC TCC ACG GAG AAA-3'
PDPN-BW 5'-GGT CAC TGT TGA CAA ACC ATC T-3'
TGFβ1-FW 5'-CTT CGC CCC TAG CAA CGA G-3'
TGFβ1-BW 5'-TGA GGG TCA TGC CGT GTT TC-3'
SLUG-FW 5‘-TGA TGA AGA GGA AAG ACT ACAG-3'
SLUG-BW 5'-GCT CAC ATA TTC CTT GTC ACA G-3‘
SNAIL-FW 5‘-GCG AGC TGC AGG ACT CTA AT-3‘
SNAIL-BW 5'-CCT CAT CTG ACA GGG AGG TC-3‘
VIMENTIN-FW 5‘-GAG AAC TTT GCC GTT GAA GC-3‘
VIMENTIN-BW 5'-GCT TCC TGT AGG TGG CAA TC-3‘
ZEB1-FW 5'-TGC ACT GAG TGT GGA AAA GC-3'
ZEB1-BW 5'-TGG TGA TGC TGA AAG AGA CG-3'

2.9 Western blotting

Western blotting was performed as described in Schinke *et al.* (Schinke et al. 2020). Whole cell lysates were extracted with phosphate-buffered saline (PBS) containing 2 % Triton X-100 and protease inhibitors (Roche Complete, Roche Diagnostics, Mannheim, Germany). Protein concentrations were determined by BCA-assay (Thermo Scientific, Schwerte, Germany). Ten to 50 µg of proteins were separated by 10 % SDS-PAGE and visualized with primary SLUG or E-CADHERIN antibodies (C19G7, Cell Signaling Technology, #9585, 1:1000, overnight at 4°C / 24E10, Cell Signaling Technology, #3195, 1:1000, overnight at 4°C) and horseradish peroxidase (HRP)-conjugated secondary antibodies (1:5000, 1 hour at room temperature), and the ECL reagent (Millipore, Darmstadt, Germany) in a Chemidoc XRS imaging system (Bio-Rad, Munich, Germany). An HRP-conjugated specific primary antibody was used to visualize beta-actin (sc-47778 HRP, Santa Cruz).

2.10 Cell proliferation assay

Cell proliferation assays were performed in accordance to Schinke *et al.* (Schinke et al. 2020). Cells were counted using a Leica DMi8 microscope with LAS X software and FIJI. In a 96-well, 2,000 cells were seeded initially for each time point. Cells were left overnight to fully attach to the plate. The next day (time point 0 hours), and 24 or 48 hours later (time points 24 and 48 hours) were measured. For counting, cells were stained with Hoechst 33342 dye (ThermoFisher) for 15 minutes. Then, using the LAS X software, 72 images per well with 100x magnification were taken and merged. In FIJI, images in greyscale (16-bit) were compressed to 8-bit by threshold adjustment (in FIJI: Image>Adjust>Threshold>Apply) to remove noise. Then, the Watershed function was applied to cut any artificially merged pixels (in FIJI: Process>Binary>Watershed). Finally, resulting particles, representing single cells, were counted and summarized (in FIJI: Analyze>Analyze Particles, setting Size:0-Infinity, Circularity: 0-1, tick "Clear results" and "Summarize"). Resulting counted cell numbers were then analyzed using *R* Software.

2.11 Fibroblast spheroid invasion assay

Fibroblast spheroid invasion assays were performed as described in Schinke *et al.* (Schinke et al. 2020). Spheroids of normal human foreskin fibroblasts (PromoCell, C-12352) were grown in Ultra Low Attachment plates (ULA) over 24 hours by seeding $1x10^4$ cells in standard DMEM. Following the formation of fibroblast spheroids, $1x10^4$ FaDu or Kyse cell transfectants were added and co-cultured for additional 48 and 72 hours. Co-cultured spheroids were carefully harvested with a cut 100 µL pipette tip and immediately frozen in tissue-TEK (Sakura Europe) in a cryomold with liquid nitrogen. Then, cryosections of 4 µm thickness were generated and IHC staining was conducted.

2.12 Matrigel invasion assay

Matrigel invasion assays were conducted in accordance to Leslie M. Shaw and Schinke *et al.* (Shaw 2005; Schinke et al. 2020).
Briefly, a total of 1x10^5 of tumor cells was seeded in 1:10 diluted matrigel-coated 24-well membrane chambers (Corning, cell culture inserts, 8 µm pore size, 353097) in serum-free medium. The lower 24-well chambers contained standard medium. After 24 hours of invasion, cells attached on the top were swiped off with a cotton swab and membranes were carefully extracted with a scalpel. Then, cells were fixed with methanol and membranes stained with crystal violet and invaded cells were counted visually.

2.13 Clonogenic survival assay

Clonogenic survival assays were conducted as described in Schinke *et al.* (Schinke et al. 2020). For FaDu 1x10^3 and for Kyse30 5x10^3 cells were plated on a 6-well plate and irradiated 24 hours later. After 14 days for FaDu and 10 days for Kyse30, cells were fixed and stained with crystal violet solution containing methanol. The whole 6-well plate was photographed using the Chemidoc XRS imaging system. Then, to quantify the area of colonies, the ColonyArea Image J Plugin by Guzman was used (Guzmán et al. 2014). Clonogenic survival was calculated by measuring the area of colonies of irradiated relative to respective non-irradiated control plates.

2.14 RNA Sequencing

For the 3'-RNA Sequencing (RNASeq), total RNA was extracted using RNeasy Mini kit (Qiagen), quantified with the Qubit-Fluorometer, and reverse transcribed with QuantiTect Reverse Transcription kit (Qiagen). The quality of RNA was assessed using a Bioanalyzer 2100 System (Agilent Technologies, Inc., USA) with the Agilent RNA 6000 Pico Kit (#5067-1513, Agilent Technologies, Inc., USA). The RNA integrity was evaluated by calculating the percentage of fragments $>$ 200 nucleotides (DV200). Sequencing libraries

were prepared using 50 ng total RNA and the QuantSeq 3'-mRNA-Seq Library Prep Kit FWD for Illumina (SKU: 015.96, Lexogen GmbH, Austria). Library amplification PCR cycles were defined by using the PCR Add-on Kit (SKU: 020.96, Lexogen GmbH, Austria). The individual libraries were amplified with 17 PCR cycles. The Quanti-iT PicoGreen dsDNA Assay Kit (P7589, Invitrogen, USA) and the Bioanalyzer High Sensitivity DNA Analysis Kit (#5067-4626, Agilent Technologies, Inc., USA) were used to assess the quality and quantity of the resulting libraries. Final 3'-RNA-Sequencing with 150 bp paired ends was performed on an Illumina HiSeq4000 platform (Illumina, Inc., USA). Gene expression quantification was carried out by alignment of RNASeq reads using the STAR aligner. Then the *featureCounts* of the *Rsubreads* R package was used for count quantification.

2.15 General data analysis and statistics

Data analysis was performed using *R* (R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, 2017; R version 3.6.1 (2019-07-05)) as described (Schinke et al. 2020). Correlation matrices were calculated and illustrated with CRAN *corrplot* package. Unless otherwise stated, further analysis was performed with built-in packages and functions from the CRAN package *tidyverse.*

2.16 Principal Componant Analysis (PCA)

The PCAs were computed using the *pcaExplorer* package (version 2.14.2). The top 500 expressed genes across all treatments and controls were used to define the Principal Components (PCs). Explained variance of PCs is listed directly on the plots.

2.17 Differentially expressed genes (DEGs)

The *DESeq2* R package (version 1.28.1) was applied to compute DEGs between pairs of samples from different treatments and their respective controls. Before running *DESeq2*, genes with an average expression lower than 100 normalized read counts were excluded. Next, all genes that were significantly differentially expressed at a false discovery rate (FDR) of ≤ 0.05 with an estimated absolute log2 fold change of > 0.5 in the pairwise comparisons were kept for downstream analysis. Gene expression is illustrated by the Volcano plots in the appendix (Appendix Fig. 2 and Fig. 4). The 72 hours treatments with EGF-low, EpEX and EGF with EpEX did not result in any significant DEGs and therefore are not displayed in the Volcano plots. The R package used was *EnhancedVolcano* (version 1.6.0) and the displayed p-value cut-off was 0.25 and the log2 fold change cut-off was 0.5.

2.18 DEGs visualization with Venn diagram and UpSet plot

To illustrate the shared and distinct gene counts from all treatments separated by the two time points 6 and 72 hours, Venn diagrams as constructed by the *venn* with *vennCounts* and *vennDiagram* from the *limma* package (version 3.44.3) were used. To visualize the DEG count intersections in more detail, an UpSet plot was used. The conducted R package was *UpSetR* (version 1.4.0) with the *upset* function. Both, the Venn diagrams and the UpSet plot, do only show DEGs as retrieved from the *DESeq2* package with an adjustest p-value ≤ 0.05.

2.19 Heatmaps with hierachical clustering

The r package *pheatmap* (version 1.0.12) was used for the unsupervised hierarchical clustering of the Top 50 DEGs from the pairwise comparisons defining the DEGs (Appendix Fig. 1 and Fig. 3). *pheatmap* was further applied for unsupervised hierarchical clustering using the Hallmarks EMT signature (MSigDB M5930) in Fig. 4.5 and appendix Fig. 5

visualizing clustering of treated Kyse30 cells. For the heatmap in Fig. 4.16, the function *heatmap.2* from the *gplots* package (version 3.0.4) was used to illustrate scaled gene expression and the distance matrix dendrogram, which was computed with the functions *dist* and *hclust* (both from *stats* package version 4.0.2). The *pheatmap* and *dist* functions computed the respective distance matrix measuring euclidean distance to identify the distances between the rows of a data matrix. The rows or columns of the heatmaps are centered and scaled as indicated in the respective figures using the *R* base function *scale*.

2.20 Gene Set Enrichment Analysis (GSEA)

GSEA was conducted using the *R* package *fgsea* (version 1.14.0) and the GO term Biological Process pathways extracted with *gmtPathways* from c5.bp.v6.2.symbols.gmt. The GSEA was conducted by *fgsea* with ranked genes from the comparisons of treated cells and respective controls, ranked by log2 fold change, without any p-value cut-off, but excluding reads below 100. Permutations were set to 10000, minimum size of genes in pathways was 15, and maximum was 500.

2.21 Survival analysis

To build Cox proportional hazard (PH) models, survival functions, and Kaplan-Meier curves, the R package *survival* (version 3.2-3) with the functions *Surv*, *coxph*, and *survfit* was used. To further visualize the survival analysis, the R package *survminer* (0.4.8) with *ggsurvplot* and *ggforest* was applied.

2.22 EGF/EGFR-mediated EMT Risk Score

All n= 181 genes distinctively shared by EGF-high 72 hours treatments in FaDu and Kyse30 cells were cut to n= 172 because of genes being opposingly deregulated comparing the two cell lines. The resulting n= 172 consistently down- or up-regulated DEGs were

applied to univariate Cox PH model analysis using 5-year overall survival as a clinical endpoint. Genes that were up-regulated in the DE analysis and were showing a Cox PH hazard ratio (HR) > 1 were kept and genes that were down-regulated in the DE analysis and were showing a HR < 1 were kept. No cut off using a p-value was applied. The resulting n= 57 genes were used to build a multivariate Cox PH model. To select genes, a forward feature selection method called Robust Likelihood-Based Survival Modeling was applied (Cho et al. 2009). The method is distributed as the *rbsurv* package (version 2.46.0). The parameters for *rbsurv* defined were; max.n.genes = 20, n.fold = 5, n.iter = 100, n.seq=1. The n.fold was set to n.fold = t in order to secure a minimal number of 10 events per fold. From the n= 57 genes, NCEH1, DDIT4, ITGβ4, FADD, and TIMP1 were selected by *rbsurv*. Then, a multivariate Cox PH model was computed with NCEH1, DDIT4, ITGβ4, FADD, and TIMP1 as features. For each patient, the expression value of the respective gene was multiplied by its Cox model coefficient. Then, the sum of all resulting values was calculated to define the Risk Score. This Risk Score was then used to assess patients survival and association with clinical parameters. The computation of the Risk Score, sometimes in literature referred to as prognostic index, was done in accordance to literature (Hess et al. 2019).

2.23 Pathway activity with PROGENy

Using the *STRINGdb* package (version 2.02.) a string network was extracted applying the function *new* with parameters version= "11", species= 9606, and score_threshold= 400. NCEH1, DDIT4, ITGβ4, FADD, and TIMP1 with *map* were used to extract neighbors with STRING IDs using *get_neighbors*. R package *AnnotationDbi* (version 1.50.3) with *EnsDb.Hsapiens.v86* was applied to map STRING IDs back to gene names. The resulting list was used to acquire gene expression data for the TCGA patients from CGDS as described in the data section. Using the gene expression matrix and the Bioconductor package *progeny* (version 1.10.0), pathway activity scores for n= 14 pathways were computed. The pathways scores were used to compare two risk subgroups of the TCGA

cohort ("Risk-" and "Risk+"). By setting the *progeny* parameter *scale* = *T*, the resulting scaled pathways activity scores were used to investigate linear correlation with the EGF/EGFR-mediated Risk Score directly across all patients.

2.24 TCGA subtype selection for pEMT quantification

In accordance to Schinke *et al.*, top 10,000 protein coding genes from the whole human genome across all TCGA patients were identified (Schinke et al. 2020). The R-package *cgdsr* (Bioconductor) was used to extract expression profiles (cancer study: "hnsc_tcga_pub"/ case list: "hnsc_tcga_pub_all"/ genetic profile: "hnsc_tcga_pub_rna_seq_v2_mrna"). A total of n= 10,000 complete mRNA profiles could be extracted for n= 243 HPV-negative TCGA HNSCC patients. Molecular subtypes for each TCGA patient were subtracted (Lawrence et al. 2015) and patient data was processed in accordance to Puram *et al.* (Puram et al. 2017). Briefly, expression values of extracted top 10,000 protein coding genes across all patients were log2 transformed, centered and a correlation analysis across all patients was computed. Patients with a mean Pearson correlation $> .1$ within their respective subtype and $< .1$ compared to all patients from the other subtypes were kept for further analysis. Then, gene expressions of resulting n= 46 patients from the mesenchymal and n= 38 from the basal-like subtype group were applied to filter out patients with highest influence of non-malignant cells. Therefore, centered gene expression values from indicated marker genes of non-malignant cells (Puram et al. 2017) served to compute an euclidean distance-to-distance matrix using the *dist* function from the R CRAN *stats* package (version 4.0.2). Then, hierarchical clustering was applied using the *hclust* function from *stats* package. The resulting clustering tree was cut at a height of 13 resulting in n= 4 different clusters. The cluster with the largest influence of non-malignant cells across all patients was excluded from downstream analysis. All others were kept for further analysis.

2.25 SING score calculation

SING score calculation was computed as described in Schinke *et al.* (Schinke et al. 2020). Using the R-package *singscore* (version 1.8.0, Bioconductor), a SING score, short for Single Sample Scoring of Molecular Phenotypes (Foroutan et al. 2018), was computed for each TCGA patient with the n= 15 common pEMT genes defined by Puram et al (Puram et al. 2017). For univariate survival analysis, Cox PH ratios (HR) > 1 with logrank p-value ≤ 0.05 was accepted as relevant. To test the validity of results obtained with the pEMT gene set, n= 10,000 random sets of n= 15 genes from the extracted gene pool (top 10,000 protein coding genes, excluding all n= 100 pEMT genes) served to compute Cox PH models with log-rank p-values. For visualization, SING scores of pEMT genes were implemented to dichotomize patients into 25 % lowest ("low", 1st quartile), intermediate 50 % ("medium", 2nd and 3rd quartiles), and 25 % highest groups ("high", 4th quartile). Then, a Cox PH model, median survival times, and log-rank p-values were calculated and included in plots of Kaplan-Meier curves. MD Anderson Cancer Cencer (MDACC) data was retrieved from the GEO object *GSE42743* by using the R Bioconductor packages *GEOquery*, *affy*, *AnnotationDbi*, and *hgu133plus2.db* to extract and map the cDNA microarray expression data, and to receive the according clinical data set. The *affy* function *rma* for robust multi-array average expression measure with default settings was applied to MDACC transcriptome data to receive log2 transformed data. From the n= 10,000 protein coding genes, n= 9831 were recovered in the MDACC data set. Then, as described for the TCGA data, a SING score with n= 15 common pEMT was computed, tested in a Cox PH model, applied to patient dichotomization, and visualized by a Kaplan-Meier curve including a Cox-proportional hazard model, median survival times, and log-rank p-values.

Chapter 3

Data

3.1 Public clinical data

The Cancer Genome Atlas (TCGA) clinical data was processed and kindly provided by Dr. Julia Hess, Head of Radiation Sensitivity Group, HelmholtzZentrum Munich. Transcription data not listed within this section was acquired from the Cancer Genomics Data Server (CGDS) hosted by the Computational Biology Center at Memorial-Sloan-Kettering Cancer Center (MSKCC) using the R-CRAN package *cgdsr* (cancer study: "hnsc_tcga_pub"/ case list: "hnsc_tcga_pub_all"/ genetic profile: "hnsc_tcga_pub_rna_seq_v2_mrna"). Data, clinical and preprocessed transcription data, from the Fred Hutchinson Cancer Research Center (FHCRC) cohort was directly extracted from the NCBI Gene Expression Omnibus (GEO) object GSE41613. Data, clinical and transcription data, from the University of Texas MD Anderson Cancer Center (MDACC) cohort was extracted from the NCBI Gene Expression Omnibus (GEO) object GSE42743. MDACC transcription data was preprocessed as described in the methods. More information can be found on `https://www.ncbi.nlm.nih.gov`.

Table 3.1: Publicly available clinical data of the HNSCC TCGA cohort. Median of described EGF/EGFR-mediated EMT Risk Score defined subgroups Risk- (low) and Risk+ (high). Statistical tests compare Risk- and Risk+. Numeric values are compared by ANOVA. Categorical values are compared by Pearson's Chi-squared test. P-values are listed.

	Risk- (N=120)	**Risk+** (N=120)	**Total** (N=240)	p-value
Time OS in months				0.003
Median	21.5	14.6	16.4	
Mean	26.0	19.3	22.6	
Range	0.0 - 60.0	0.0 - 60.0	0.0 - 60.0	
Status OS				< 0.001
Living	80 (66.7%)	46 (38.3%)	126 (52.5%)	
Deceased	40 (33.3%)	74 (61.7%)	114 (47.5%)	
Primary site				0.097
Hypopharynx	0 (0.0%)	1 (0.8%)	1 (0.4%)	
Larynx	30 (25.0%)	40 (33.3%)	70 (29.2%)	
Oral Cavity	87 (72.5%)	71 (59.2%)	158 (65.8%)	
Oropharynx	3 (2.5%)	8 (6.7%)	11 (4.6%)	
N-status				< 0.001
N0	77 (64.2%)	47 (39.2%)	124 (51.7%)	
N+	43 (35.8%)	73 (60.8%)	116 (48.3%)	
Stage				< 0.001
I	8 (6.7%)	1 (0.8%)	9 (3.8%)	
II	32 (26.7%)	17 (14.2%)	49 (20.4%)	
III	36 (30.0%)	26 (21.7%)	62 (25.8%)	
IV	44 (36.7%)	76 (63.3%)	120 (50.0%)	
RiskScore				< 0.001
Median	-0.645	0.700	-0.098	
Mean	-0.790	0.790	0.000	
Range	-3.257 - -0.101	-0.095 - 2.303	-3.257 - 2.303	

Table 3.2: Publicly available clinical data of the FHCRC and MDACC HNSCC cohorts. Risk- and Risk+ represent subgroups defined according to median of the EGF/EGFR-mediated EMT Risk Score of the TCGA cohort. Statistical tests compare Risk- and Risk+. Numeric values are compared by ANOVA. Categorical values are compared by Pearson's Chi-squared test. P-values are listed.

FHCRC	**Risk-** (N=39)	**Risk+** (N=58)	**Total** (N=97)	p-value
Time OS in monhts				0.004
Median	60.0	35.3	54.4	
Mean	48.0	34.8	40.1	
Range	3.4 - 60.0	0.5 - 60.0	0.5 - 60.0	
Status OS				0.015
Living	26 (66.7%)	24 (41.4%)	50 (51.5%)	
Deceased	13 (33.3%)	34 (58.6%)	47 (48.5%)	
RiskScore				< 0.001
Median	-0.884	0.549	0.149	
Mean	-0.988	0.665	0.000	
Range	-2.833 - -0.116	-0.090 - 2.544	-2.833 - 2.544	
MDACC	**Risk-** (N=27)	**Risk+** (N=35)	**Total** (N=62)	p-value
Time OS in months				0.002
Median	27.8	11.8	20.2	
Mean	29.5	16.1	21.9	
Range	1.6 - 60.0	0.3 - 54.4	0.3 - 60.0	
Status OS				< 0.001
Living	20 (74.1%)	10 (28.6%)	30 (48.4%)	
Deceased	7 (25.9%)	25 (71.4%)	32 (51.6%)	
RiskScore				< 0.001
Median	-0.763	0.764	0.025	
Mean	-0.928	0.716	-0.000	
Range	-2.359 - -0.131	-0.082 - 2.144	-2.359 - 2.144	

3.2 In-house LMU cohort

Table 3.3: Clinical data of the in-house LMU cohort. Cohort was separated according to SLUG IHC quantification. Statistical tests compare SLUG low and SLUG high. Numeric values are compared by ANOVA. Categorical values are compared by Pearson's Chi-squared test. P-values are listed.

	SLUG high (N=59)	**SLUG low** (N=17)	**Total** (N=76)	p-value
Time DFS in months				0.684
Median	14.1	18.7	14.7	
Mean	19.0	20.7	19.4	
Range	1.1 - 60.0	2.8 - 43.7	1.1 - 60.0	
Status DFS				0.008
Living/Recurrence free	31 (52.5%)	15 (88.2%)	46 (60.5%)	
Deceased/Recurrence	28 (47.5%)	2 (11.8%)	30 (39.5%)	
Primary site				0.518
Oral Cavity	21 (35.6%)	4 (23.5%)	25 (32.9%)	
Oropharynx	27 (45.8%)	8 (47.1%)	35 (46.1%)	
Hypopha. & Larynx	11 (18.6%)	5 (29.4%)	16 (21.1%)	
Recurrence				0.097
Free	40 (67.8%)	15 (88.2%)	55 (72.4%)	
Recurrence	19 (32.2%)	2 (11.8%)	21 (27.6%)	
IHC localization				0.100
Homogeneous	34 (57.6%)	12 (70.6%)	46 (60.5%)	
Edge	13 (22.0%)	0 (0.0%)	13 (17.1%)	
NoInfo	12 (20.3%)	5 (29.4%)	17 (22.4%)	

Chapter 4

Results

4.1 EGF/EGFR-mediated EMT induction

In collaborative work of Min Pan and myself, we assessed EGFR and EpCAM protein expression in n = 180 tumor cryosections from HNSCC patients of an in-house LMU cohort and I performed the bioinformatic analysis of the resulting data (Pan et al. 2018). By doing so, we demonstrated that patients with high EGFR protein expression suffer from poorer overall survival. Comparing the EGFR/EpCAM co-expression, we showed that patients with EGFR-low/ EpCAM-high co-expression are characterized by outstandingly high overall and disease-free survival rates compared to patients with EGFR-high/ EpCAM-low co-expression (Fig. 4.1 A-B). The difference in survival rates were also confirmed in HPV-negative patients only (Fig. 4.1 C-D). Further, we reported the capability of EGF to mediate EMT in HNSCC cell lines Kyse30 and FaDu, both expressing high levels of EGFR. A treatment with 10 ng/mL EGF (EGF-low) under serum starvation induced intermediate levels of pERK and pAkt, and induced cell proliferation. A treatment with 50 ng/mL (EGF-high) induced high levels of pERK, expression of EMT-TFs SNAIL, SLUG, and ZEB1, and induced an EMT phenotype at a morphological level after 72 hours (Pan et al. 2018). Further, sustained high pERK levels were measured within a timeframe of 6 hours and were identified as the phospho-protein mediator of EMT in HNSCC cell lines. The extracellular domain of EpCAM, EpEX, was defined as a novel ligand of EGFR, capable of inducing a similar phenotype as EGF-low treatment. Co-treatment with EGF-high and EpEX, abbreviated as EGF with EpEX, led to the inhibition of EMT induction (Pan et al. 2018).

Now, to understand the underlying changes in the transcriptomes we conducted an RNASeq of the four different treatments EGF-low, EGF-high, EpEX, and EGF with EpEX. At 6 and 72 hours bulk RNA was extracted, representing the time points of high pERK levels and phenotypic EMT, respectively. All treatments were conducted in quadruplicates, under serum starvation conditions, and in HNSCC cell lines Kyse30 and FaDu. Cellular morphology of EGF/EGFR-mediated EMT is shown in Fig. 4.2 and the whole procedure is schematically represented in Fig. 4.3.

A Principal Component Analysis (PCA) with the top 500 genes expressed in Kyse30

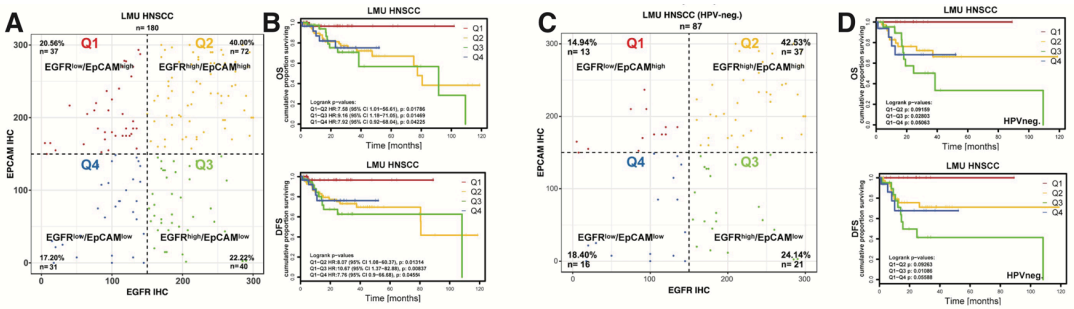

Figure 4.1: EGFR and EpCAM co-expression is prognostic in HNSCCs. EGFR and EpCAM expression of primary HNSCCs was evaluated through IHC stained cryosections of n = 180 patients from the LMU cohort. A) IHC scores of EGFR and EpCAM expression from all n = 180 primary tumors and a n = 87 HPV-negative subcohort (C) were subdivided into quadrants Q1- Q4 according to an IHC cut-off threshold of 150 (IHC score range 0-300). Numbers and percentages of patients within quadrants are noted on the plot. B)-D) Overall and disease-free survival rates of subgroups of patients from (A) and (C) were compared. Kaplan-Meier curves with 95% CI, HRs, and log-rank p-values of the Cox models comparing the subgroups are shown. EGFR: Epidermal growth factor receptor, EpCAM: Epithelial cell adhesion molecule, CI: Confidence intervall, HR: Hazard ratio, IHC: Immunohistochemistry, RPPA: Reversed-phase protein atlas, HNSCC: Head and neck squamous cell carcinoma. Taken from Pan *et al.*, 2018.

cells across all treatments determined a PC1 explaining 46.11 % and a PC2 explaining 13.31 % of variance in the transcriptional data. Plotting PC1 and PC2 shows that the controls of 0 and 6 hours are in close transcriptional proximity. All 6 hours treatments

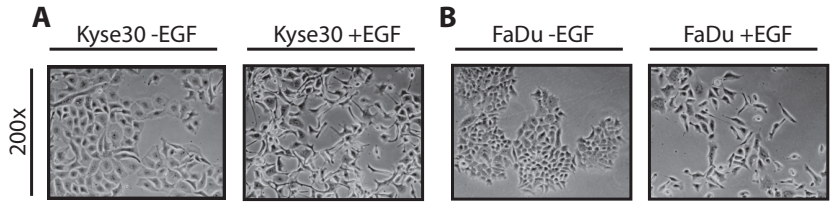

Figure 4.2: Representative micrographs of the morphology of Kyse30 (A) and FaDu (B) cells after EGF-high mediated EMT induction. Cells were treated for 72 hours under serum starvation conditions. 200x magnification. Shown are representative micrographs from n = 4 independent experiments. EGF: Epidermal growth factor, EMT: Epithelial-to-Mesenchymal transition.

do also appear as a cluster, indicating a high degree of transcriptome similarity. After 72 hours, EGF-high treatments are clearly separated from all other 72 hours treatment and control groups (Fig. 4.4). This is in line with the morphological changes related to EMT and depicted after 72 hours. To confirm the findings from the PCA and the induction of an EMT phenotype after 72 hours with EGF-high treatments a Hallmark EMT signature from the MSig data base was implemented (MSigDB: M5930). This signature includes n = 200 genes defining EMT related to wound healing, fibrosis, and metastasis. It was defined using n = 105 previously described gene sets, refined using four data sets of different cancer types and cell lines, and validated in two data sets, one consisting of distal stroma vs. malignant tumors and the other consisting of human lung adenocarcinoma A549 cells. The gene expression of all treatment groups in the Kyse30 cells, excluding the controls, was applied to hierarchical clustering based on euclidean distance. The values in rows, representing the single gene expressions, were centered and scaled for a better visualization. To exclude genes that might not be relevant in the presented HNSCC data set and enhance the clustering precision, the top 25 genes expressed after 72 hours of EGF-high treatment were selected. The hierarchical clustering identified two distinct clusters, one including all 6 hours and EGF-high 72 hours treatments, the other including all other 72 hours treatments. Within the first cluster, EGF-high 72 hours treatments further separated from the 6 hours treatments (Fig. 4.5). A clustering

approach implementing all n = 200 Hallmark EMT genes and n = 200 random genes as a control of clustering can be found in the appendix (Appendix Fig. 5). A GSEA of GO terms "Biological Processes" of Kyse30 cells with all treatment groups was conducted. Terms related to processes such as EMT or cell proliferation were chosen to be displayed to summarize the findings (Fig. 4.6). All 6 hours treatments do show a high degree of GO term regulation, as wound healing, ribosome biogenesis, and keratinization are up-regulated in all four 6 hours groups showing a Normalized Enrichment Score (NES) ≥ 1.5. Terms like chromosome segregation or cellular component assembly involved in morphogenesis are down-regulated in all four groups (NES ≤ 2). Terms enriched in all 6 hours treatments, except for the EpEX treatment, were response to cytokine, immune response, or cell cell signaling (NES = 1.25 to 2). Terms reduced in all 6 hours treatments, except EpEX treatment, were organelle fission and cell division (NES ≥ 1.5). At 6 hours, all treatments were showing similar GO term enrichment profiles, no treatment was clearly distinct from all others, except EpEX treatment was showing a reduced number of regulated GO terms. At 72 hours, most regulation of biological processes was lost in all treatment groups, except the EGF-high treatment, which led to an up-regulation of GO-terms including wound healing, membrane budding, immune response, cell cell signaling, and keratinization with NES values ≥ 1.5. Wound healing, keratinization, and immune response were already enriched at 6 hours in EGF-high treatments. Terms like cell division, negative regulation of locomotion, negative regulation of epithelial to mesenchymal transition, chromosome segregation, or cell cycle phase transition are down-regulated by NES ≥ 2. The EGF-low, EpEX, and EGF with EpEX treatments at 72 hours were showing considerably reduced regulation compared to 6 hours. These GO term regulations indicate a similar transcriptional regulation at 6 hours across all treatments. After 72 hours, essentially only EGF-high is capable of regulating processes suppressing cellular division or negative regulation of EMT.

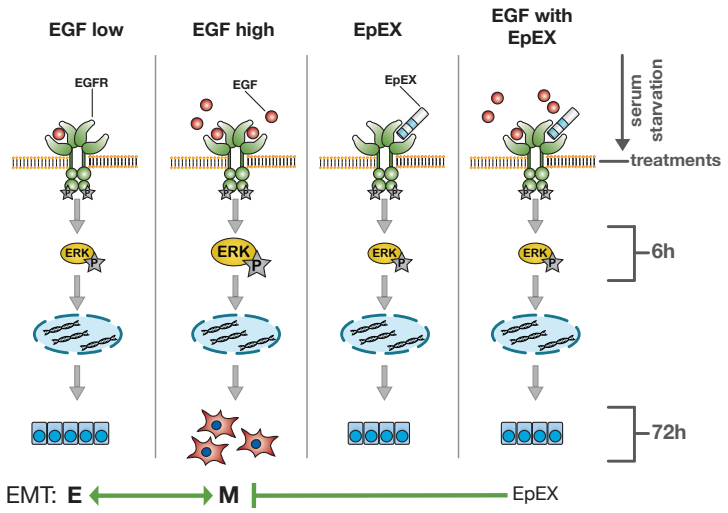

Figure 4.3: Schematic representation of the RNASeq experiment of EGFR-signaling induced by EGF and EpEX. Transient and moderate ERK activity in response to EGFR activation by EGF-low leads to cell proliferation (left panel). Sustained and strong ERK activity induced by EGF-high treatments results in EMT induction and a mesenchymal phenotype is induced (center-left panel). EpEX binds to the extracellular domain of EGFR and activates the ERK downstream signaling pathway similiar to EGF-low treatments (center-right panel). The co-treatment of EpEX with EGF-high (EGF with EpEX) is not inducing EMT and an epithelial phenotype is retained (right panel). Bulk RNA was extracted at 6 and 72 hours from n = 4 independent experiments. Cells were treated under serum starvation conditions in the absence of additional growth factors. RNASeq: RNA sequencing, EGF/R: Epidermal growth factor/ receptor, ERK: Extracellular signal–regulated kinase, EpEX: Extracellular domain of EpCAM.

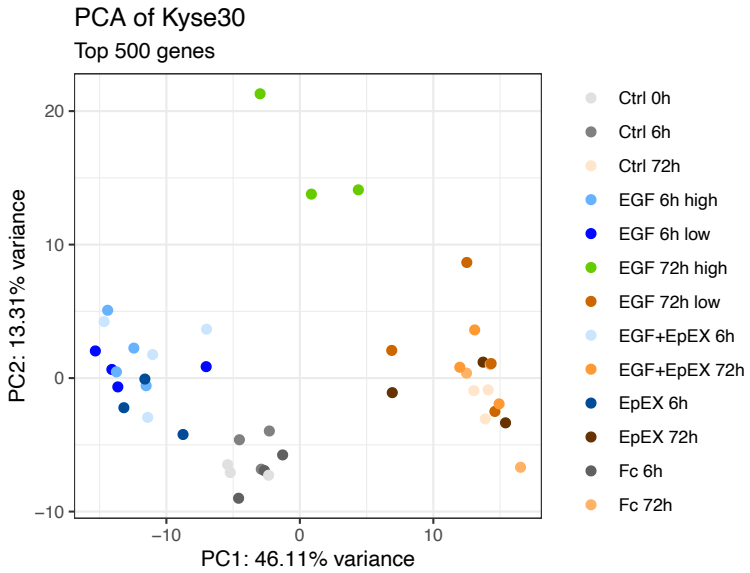

Figure 4.4: PCA of Kyse30 cells. All treatment groups are shown. Top 500 overall expressed genes were implemented in the analysis. Per treatment group n $\geq$ 3 independent experiments were analyzed. PC1 and PC2 with explained variance are shown. Ctr: Contol, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM, Fc: Fragment crystallizable region, PCA: Principal component analysis, PC: Principal component.

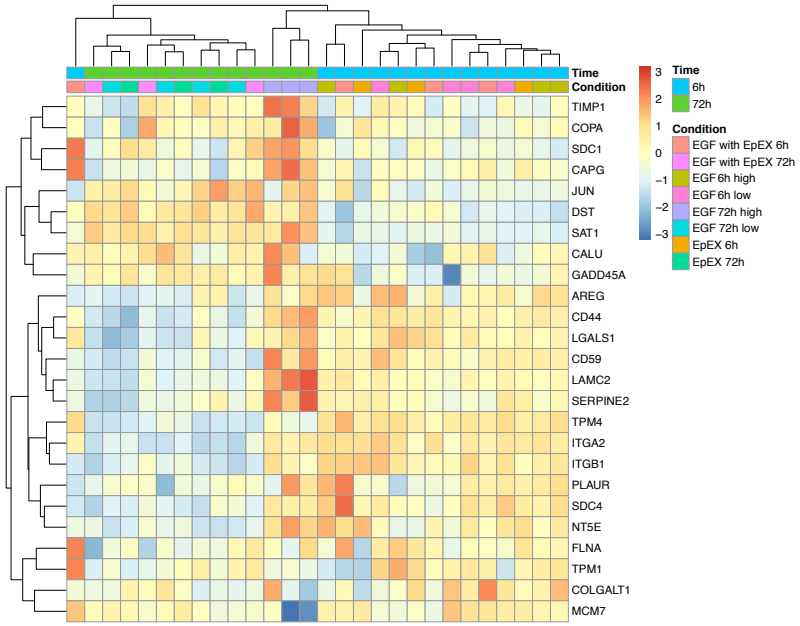

Figure 4.5: Heatmap with hierarchical clustering of treatments excluding controls. Top 25 genes expresssed in Kyse30 EGF-high 72 hours treatments of Hallmarks EMT signature (MSigDB M5930) were selected for clustering. Color represents centered and scaled gene expresson values in rows. First row of colored boxes represents time points: 6 (blue) and 72 hours (green). Second row of boxes defines treatment group with colors (Condition). Genes represented in rows are named. Shown are gene expression values from n $\geq$ 3 independent experiments. EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM, EMT: Epithelial-to-Mesenchymal transition.

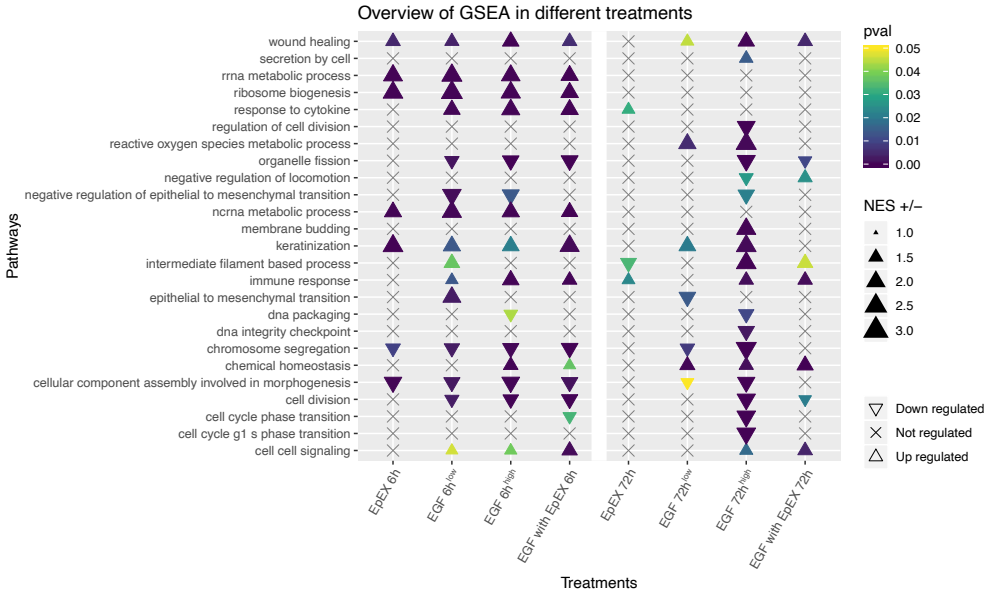

Figure 4.6: Overview of a GSEA with GO terms Biological Processes of Kyse30 cells. All treatment groups compared to the respective controls are shown. Gene expression was ranked according to log2 fold change without p-value cut-off. Pathway names are shown on the left. Color encodes for p-value. Triangle size encodes for NES. Direction of trianlge shows if a pathway is up- or down-regulated. Shown are the results from n $\geq$ 3 independent experiments. NES: Normalized enrichment score, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM, EMT: Epithelial-to-Mesenchymal transition, GSEA: Gene set enrichment analysis, GO: Gene ontology.

4.2 Sustained EGF/EGFR-signaling is necessary to induce EMT

To simplify the interpretation of results, the previously reported analysis of EGF/EGFR-mediated EMT induction was shown for Kyse30 cells only. To avoid making cell type specific assumptions, all of the following steps will be performed for both, Kyse30 and FaDu cells. In order do identify differentially expressed genes, the RNA expression in the treatment groups was compared to the respective controls. All treatments were executed under serum starvation conditions. All EGF treatments were compared to their control groups, meaning prolonged serum starvation without further treatment for 6 or 72 hours. The recombinant EpEX used within the scope of this experiment possessed a C-terminal immunoglobulin G heavy-chain as a tag for purification (Fc-tag). Therefore, the EpEX treatments were compared to serum starved cells that were further treated with Fc for 6 and 72 hours. For further downstream analysis of DEGs, cut-offs applied were > 0.5 in log2 fold change with an adjusted p-value of ≤ 0.05 as defined by the implemented R-package *DESeq2*. For Kyse30 cells at 6 hours, EGF-low treatment resulted in n = 397, EGF-high treatment in n = 612, EpEX in n = 137, and EGF with EpEX in n = 291 DEGs. Within Kyse30 cells, the exclusive DEGs in the EGF-low treatment group were n = 96, n = 264 in EGF-high, n = 53 in EpEX, and n = 64 in EGF with EpEX treatments. Comparing all 6 hours treatments, n = 53 DEGs were overlapping. EGF-low and -high treatments did show n = 284 genes that were differentially expressed in both treatment groups, not exclusive to other treatments. For Kyse30 cells at 72 hours, only the EGF-high treatment group showed n = 1208 DEGs. None of the other treatments sustained transcriptional regulation after 72 hours, except two DEGs in EpEX, of which one was as well found in EGF-high treatment (Fig. 4.7). For FaDu at 6 hours, EGF-low induced regulation of n = 778 genes, EGF-high n = 994 genes, EpEX n = 36 genes, and EGF with EpEX n = 911 genes. The exclusive DEGs in the EGF-low treatment group were n = 118, n = 316 in EGF-high, one in EpEX, and n = 278 in EGF with EpEX treatments. Across all treatments within FaDu cells, n = 33 DEGs were overlapping. In EGF-low and

-high, n = 581 genes were common but not exclusive to the other treatments. In FaDu cells after 72 hours, EGF-high sustained the highest number of DEGs with n = 1536. EGF-low treated cells differentially expressed four genes and EGF with EpEX treated cells n = 103 genes. Treatment with EGF-low induced the regulation of one distinct gene, EGF-high n = 1439, EpEX n = 0, and EGF with EpEX n = 6 distinct genes (Fig. 4.7). These findings of the differential gene expression in both cell lines are in line with the hierarchical clustering of Hallmark EMT genes and the GSEA within Kyse30 cells.

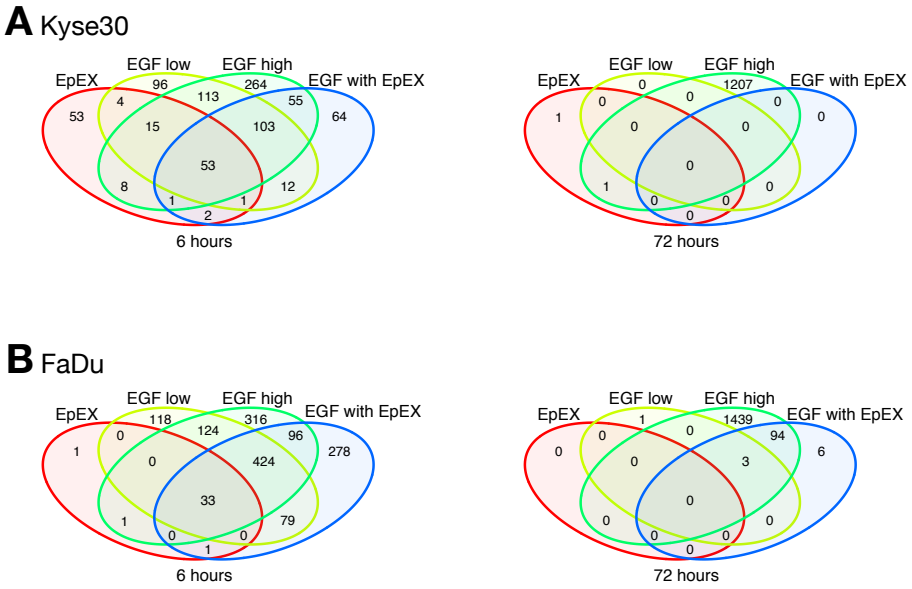

Figure 4.7: Venn diagrams of DEGs in Kyse30 and FaDu cells. Time points 6 and 72 hours are displayed separately. Left panel shows 6 hours, right panel shows 72 hours. A) DEGs in Kyse30 cells. B) DEGs in FaDu cells. Numbers show DEGs in different intersections. Intersects are not exclusive across time points and cell lines. All DEGs plotted meet the criteria of log2 fold change > 0.5 and adjusted p-value ≤ 0.05. Results from n $\geq$ 3 independent experiments are shown. DEGs: Differentially expressed genes, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM.

4.2.1 EpEX does not inhibit EMT through transcriptional regulation

To compare the DEGs in detail and to understand whether the induction of EMT might be interfered by transcriptional counter-regulation in EGF-low or EpEX treatments, the log2 fold change in expression of different treatment groups at 6 hours was plotted. To quantify the similarity of transcriptional regulation in the comparisons, a Pearson correlation analysis was calculated and is indicated within the respective plots. Further, the total number of mutual DEGs is shown. In both cell lines, the comparison of DEGs from EGF-low and -high treatments indicates a very high similarity in transcriptional regulation at 6 hours. The Pearson correlation coefficient r is highly significant and exceeding 0.98, indicating a very strong similarity in log2 fold change of the DEG expression. When comparing EGF-low and -high treatments to EGF with EpEX respectively, the Pearson correlation coefficient r is highly significant and exceeds 0.98 in both comparisons and consistently over both cell lines. Further, the matching gene numbers are not excessively different, as for Kyse30 in EGF-low vs. -high n = 284, EGF-low vs. EGF with EpEX n = 169, and in EGF-high vs. EGF with EpEX n = 212 DEGs were found. In FaDu, the matching DEG numbers were; n = 581 for EGF-low vs. -high, n = 536 for EGF-low vs. EGF with EpEX, and n = 553 for EGF-high vs. EGF with EpEX. This shows that EGF-low and -high treatments are very similar at 6 hours and the co-treatment with EpEX is not inducing a different transcriptional profile. To confirm that EpEX is not counteracting the induction of EMT by a different transcriptional profile, log2 fold change values of EGF-low and -high were compared to treatment with EpEX only. Again, no counter-regulated genes were found and in both cell lines, Kyse30 and FaDu, the Pearson correlation indicates a very high similarity of expression values (for EGF-low and -high vs. EpEX: r= 0.97 in Kyse30/ r= 0.88 in FaDu). Across all comparisons, not a single counter-regulated gene by EpEX only or in the combined treatment EGF with EpEX was found in neither of the two investigated cell lines (Fig. 4.8 and Fig. 4.9). Therefore, EpEX is not directly regulating gene transcription to repress EMT induction by EGF-high treatments.

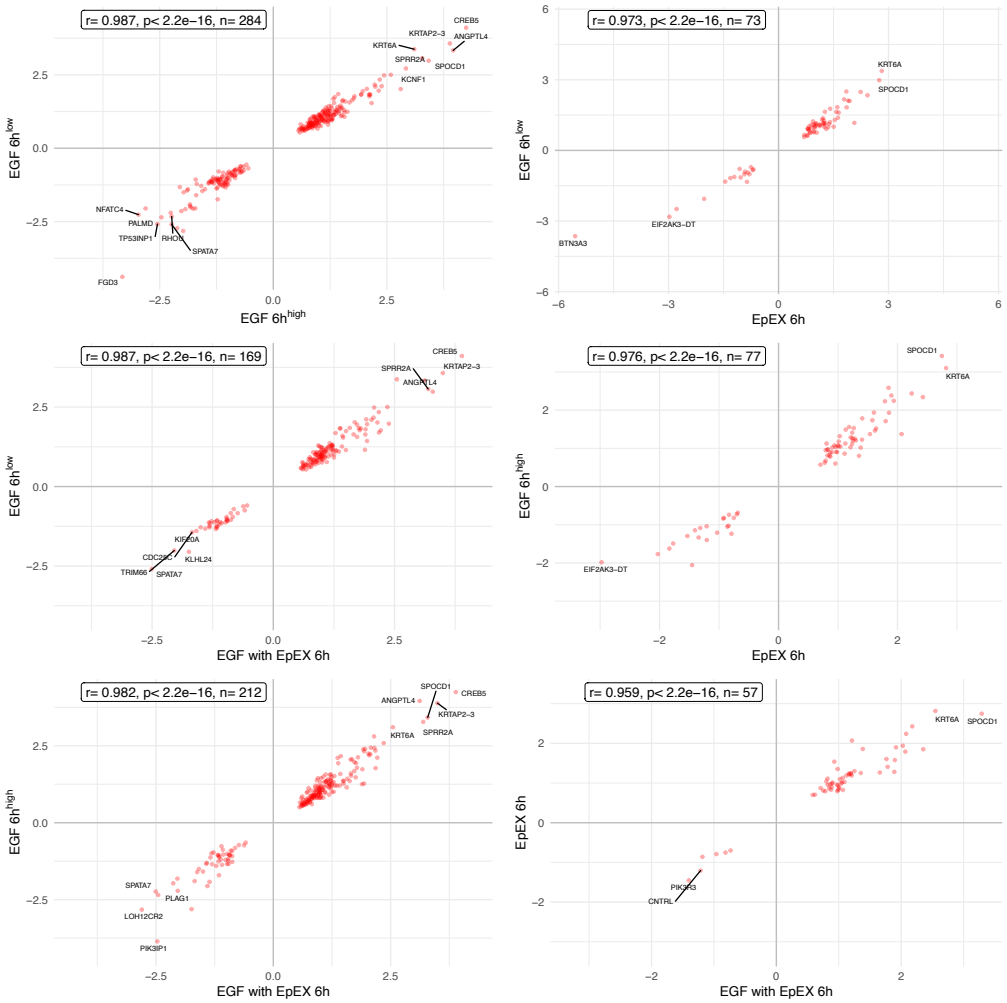

Figure 4.8: Comparison of DEGs in Kyse30 cells after 6 hours. 2.5 % highest or lowest, and most significant genes are labelled. Treatments compared are depicted on the respective axis. Pearson correlation analysis coefficients with p-value and total number of compared genes are shown on the plot in the white box. Axis show log2 fold change of DEGs. All DEGs plotted meet the criteria of log2 fold change > 0.5 and adjusted p-value ≤ 0.05. Results from n ≥ 3 independent experiments are shown. DEGs: Differentially expressed genes, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM.

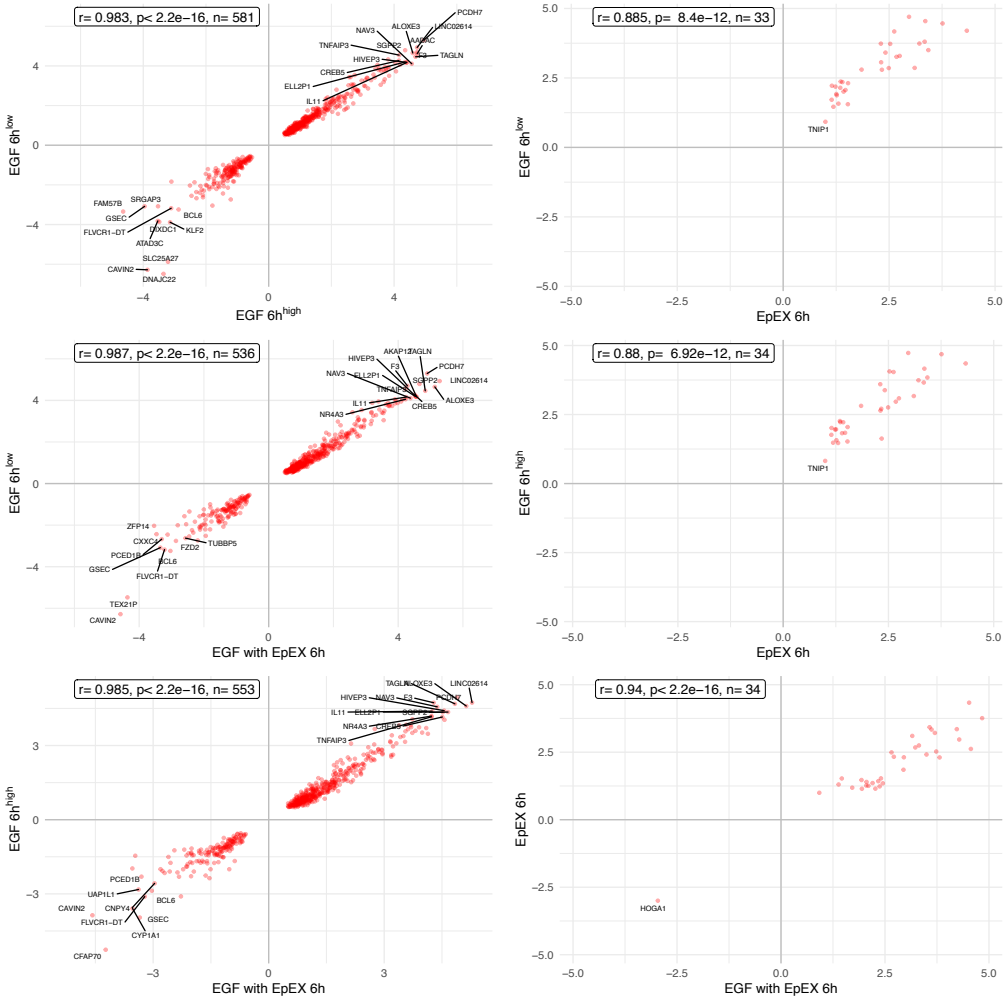

Figure 4.9: Comparison of DEGs in FaDu cells after 6 hours. 2.5 % highest or lowest, and most significant genes are labelled. Treatments compared are depicted on the respective axis. Pearson correlation analysis coefficients with p-value and total number of compared genes are shown on the plot in the white box. Axis show log2 fold change of DEGs. All DEGs plotted meet the criteria of log2 fold change > 0.5 and adjusted p-value ≤ 0.05. Results from n ≥ 3 independent experiments are shown. DEGs: Differentially expressed genes, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM.

4.2.2 Gene regulation by EpEX

Across both cell lines, EpEX was capable of inducing five genes consistently; A-Kinase Anchor Protein 12 (AKAP12), Glioma Pathogenesis-Related Protein 1 (GLIPR1), Inositol Polyphosphate 1-Phosphatase (INPP1), Microtubule-Associated Monooxygenase Calponin-And-LIM-Domain-Containing 2 (MICAL2), and ST3 β-Galactoside α-2,3-Sialyl transferase 1 (ST3GAL1). Those genes were found to be up-regulated to very similar levels in the RNASeq comparing Kyse30 and FaDu cells. The cell growth-related AKAP12 and the metabolic activity-associated GLIPR1 showed the highest difference in DEG expression with a log2 fold change ranging from 2 to 3.5. INPP1 is involved in the inositol phosphate metabolism and was induced by two fold (log 2 fold change = 1). MICAL2, which is involved in cytoskeletal dynamics, and ST3GAL1, which is a sialyltransferase, were induced with a log2 fold change of 1.4 to 1.9. None of those genes was exclusively found in the EpEX treatment group. In fact, the regulation of those five genes was found to be highly similar in all four treatment groups at 6 hours (data not shown). To confirm the findings from the RNASeq data, AKAP12, GLIPR1, INPP1, MICAL2, and ST3GAL1 mRNA transcription was measured by qPCR analysis. The up-regulation of all five genes after 6 hours EpEX treatment relative to Fc treatment (control) was confirmed in both cell lines. Except MICAL2 expression in FaDu cells, all relative mRNA expression values were significantly up-regulated. In the qPCR, AKAP12 was induced by 5 to 17 fold, in Kyse30 and FaDu respectively, which translates into a log2 fold change of 2.3 to 4.1. GLIPR1 was induced by 7.5 fold in Kyse30 and 35 fold in FaDu cells, translating into a log2 fold change of 2.9 for Kyse30 and 5.1 for FaDu cells. In Kyse30, relative mRNA expression of INPP1 was 1.8 (log2 = 0.85) and in FaDu it was 2.5 (log2 = 1.32). MICAL2 was relatively increased by 3.8 fold in Kyse30 (log2 = 1.93) and 4.5 fold in FaDu (log2 = 2.17). For ST3GAL1, qPCR analysis showed an increase of 2.7 in Kyse30 cells (log2 = 1.43) and 7.6 in FaDu cells (log2 = 2.93). Thus, the induction of those genes by EpEX treatment could be validated and the expression values of the confirmation by qPCR fall into a similar range compared to the data measured by RNASeq (Fig. 4.10).

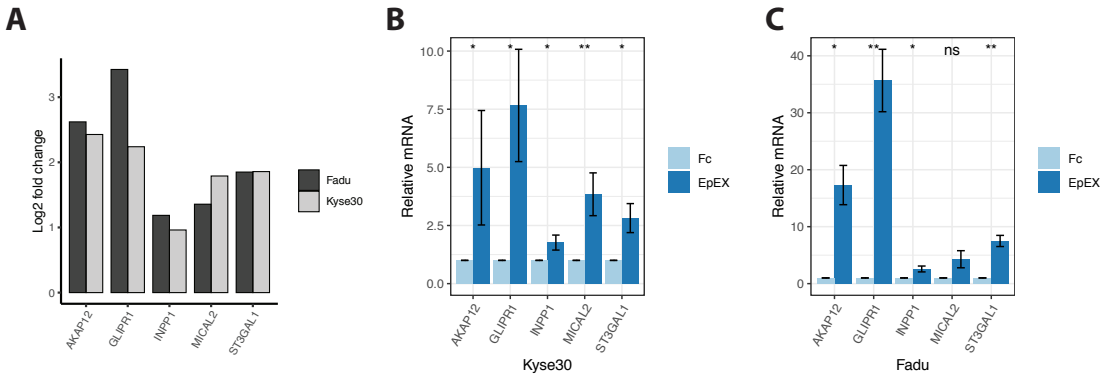

Figure 4.10: Genes regulated by EpEX. A) Log2 fold changes of genes in Kyse30 and FaDu cells defined by RNASeq (log2 fold change > 0.5, adj. p-value ≤ 0.05). B) and C) qPCR relative mRNA expression mean values with SDs of Kyse30 (B) and FaDu cells (C) as defined by the $\Delta\Delta$Ct method. One-way ANOVA with Tukey HSD p-value = not significant; ns, ≤ 0.05; *, ≤ 0.01; **. Results from n ≥ 3 independent experiments are shown. SD: Standard deviation, qPCR: quantitative real-time PCR, Fc: Fragment crystallizable region, EpEX: Extracellular domain of EpCAM, RNASeq: RNA sequencing.

4.3 EGF/EGFR-mediated EMT Risk Score

The presented RNASeq analysis led to the conclusion that at 6 hours there is no substantial difference in gene regulation between EGF-low or EGF-high or any other treatment group. An UpSet plot (Fig. 4.11) allowed to define the distinct gene numbers of single or combined groups of DEG venns. In the UpSet plot, this is illustrated by the dots, representing DEG venns of a single treatment, and connecting lines, representing selected combinations of venns. The distinct numbers of DEGs are indicated and represented by the blue bar charts on top. Since, treatments at 72 hours, except EGF-high, were showing almost no DEGs, they will be excluded from the UpSet plot to facilitate the interpretation of the plotted data. The combination of the DEGs of EpEX treatments from Kyse30 and FaDu cells shows no distinct genes. The same is true when combining EpEX and EGF with EpEX treatments, which confirms the aforementioned findings that EpEX is not a counter-regulator of EGF/EGFR-mediated EMT on transcriptome level. Nonetheless, the dual comparison of EGF with EpEX in Kyse30 and FaDu cells showed four genes being distinct from all other treatments. Those four genes were Proliferation And Apoptosis Adaptor Protein 15 (PEA15), a negative regulator of apoptosis, Plectin (PLEC), encoding an intermediate filament binding protein, and the two zinc fingers ZC3H15 and ZNF792. The distinct numbers of DEGs from the treatment groups that did not result in the induction of EMT, namely EGF-low, EpEX, and EGF with EpEX, were zero, which shows that the gene regulation is most likely not different to EGF-high at 6 hours. In both cell lines, EGF-high was capable of regulating the highest absolute number of genes at both time points. EGF-low was regulating fewer genes in total at 6 hours, and a maximum of four genes at 72 hours. In contrast, EGF-high regulated the transcription of n = 1208 genes in Kyse30 and n = 1536 in FaDu cells at 72 hours (Fig. 4.7). This is depicted as well by the UpSet plot in Fig. 4.11 by the number of DEGs on the left side, represented as bar charts. After 72 hours EGF-high treatment, n = 709 genes and n = 919 genes were different from all other treatments for Kyse30 and FaDu cells, respectively, as shown by the blue bar charts on top. EGF-high treatments at 6

hours regulated n = 12 distinct genes and the combination with EGF-high treatment of 72 hours led to two DEGs. When looking at EGF-high treatments at 72 hours in Kyse30 and FaDu cells, n = 181 DEGs were distinct from all other plotted treatment groups. Since an EGF/EGFR-mediated EMT phenotype was observed at 72 hours in both cell lines (Fig. 4.2, Fig. 4.5, and Fig. 4.6) and was further shown in previous collaborative work from Min Pan and myself (Pan et al. 2018), those n = 181 DEGs were defined as the EGF/EGFR-mediated EMT gene signature and served to extract a prognostic EGF/EGFR-mediated EMT gene signature for HPV-negative HNSCC patients.

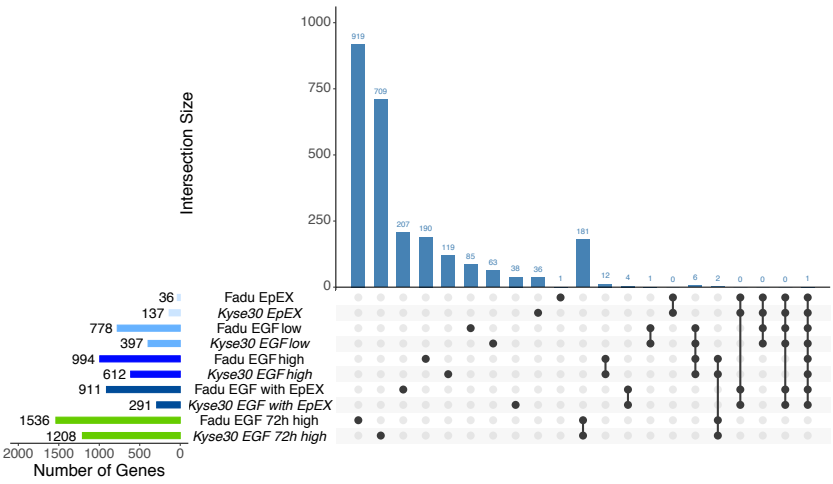

Figure 4.11: Upset plot of DEGs in all 6 hours treatments and 72 hours EGF-high treatments from Kyse30 and FaDu cells. Total number of DEGs is shown by bar charts on the left. Within the table, points represent a selected treatment and lines mark comparisons of different treatments. Distinct numbers of DEGs across all samples from comparisons or single treatments and intersection sizes are shown by bar charts on the top of the graph. All DEGs plotted meet the criteria of log2 fold change > 0.5 and adjusted p-value ≤ 0.05. Results from n $\geq$ 3 independent experiments are shown. DEGs: Differentially expressed genes, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM.

4.3.1 Extracting candidate genes from TCGA patient data

The EGF/EGFR-mediated EMT gene signature was reduced from n = 181 to 172 genes being consistently up- and down-regulatied across FaDu and Kyse30 cells. To assess the impact of those EMT genes on survival, the publicly available TCGA HNSCC cohort consisting of n = 243 HPV-negative patients was used (Lawrence et al. 2015). Complete data including clinical parameters such as tumor primary sites or lymph node metastasis status were available for n = 240 TCGA patients (Tab. 3.1). The overall survival data was available for all n = 240 patients and cut to five years of clinical follow-up. Then, the mRNA expression values of the n = 172 identified EMT genes from the TCGA patients were downloaded by accessing the Memorial Sloan-Kettering Cancer Center (MSKCC) Cancer Genomics Data Server (CGDS) and were further log2 transformed. In order to make a transition from *in vitro* findings to clinical data, all n = 172 genes were assessed by a univariate Cox proportional hazard model (Cox model). Since, EMT is viewed as a process with negative impact on the overall survival of patients, only genes being up-regulated *in vitro* and leading to a Cox hazard ratio (HR) > 1 were kept. Similarly, down-regulated genes from the candidate EMT genes with a Cox HR < 1 were kept for further analysis. This led to a total number of n = 57 genes from the EGF/EGFR-mediated EMT candidate genes with potential relevance in TCGA patients. All n = 57 univariate Cox models are listed alphabetically and separated into up- and down-regulated in Appendix Tab. 1, 2, and 3.

4.4 Computing and evaluating an EGF/EGFR-mediated EMT Risk Score

The identified n = 57 genes were applied to robust likelihood-based survival modeling, a method developed by Cho and colleagues (Cho et al. 2009), for feature selection of a multivariate Cox model. The feature selection resulted in a highly significant multivariate Cox model with a global log-rank p-value of 0.0003 (Fig. 4.12, A). The genes selected as features for the multivariate Cox model were: Neutral Cholesterol Ester Hydrolase 1

(NCEH1), DNA Damage Inducible Transcript 4 (DDIT4), Integrin Subunit β 4 (ITGβ4), Fas Associated Via Death Domain (FADD), and TIMP Metallopeptidase Inhibitor 1 (TIMP1) (Fig. 4.12, top left). TIMP1 is part of the Hallmark EMT signature (MSigDB: M5930). NCEH1 was suggested to promote tumor cell migration (Chiang et al. 2006). DDIT4 is part of the cellular DNA-damage response (Ellisen et al. 2002). FADD mediates apoptotic signals and ITGβ4 mediates cell-matrix adhesion (Chinnaiyan et al. 1996; Dyce et al. 2002). Next, the information of the multivariate Cox model was combined with the expression levels of the selected genes to form a Risk Score. In accordance to Hess *et al.*, the coefficients of the multivariate Cox model were used to weigh RNA expression of the featured genes in patients. Thus, for all n = 240 TCGA patients the expression values of each of the five genes were multiplied by the respective coefficient from the Cox model. The resulting weighted expression values were added up for each patient. Then the sum of values were centered and scaled across all patients to form an EGF/EGFR-mediated EMT Risk Score. To visualize the prognostic value of the EGF/EGFR-mediated EMT Risk Score, TCGA patients were dichotomized into two groups according to the median Risk Score of the whole cohort. Patients with a Risk Score larger then the median were labeled as "Risk+" and patients with a score lower then the median with "Risk-". In a Kaplan-Meier curve with a Cox model, the "Risk+"- and "Risk-"-patients were compared and the "Risk+" group was showing worse prognosis according to overall survival with a 2.41-times elevated relative hazard risk (Fig. 4.12, B). The clinical parameters stage and lymph node metastasis status were both binarized into two groups, I/II vs. III/IV and N0 vs. N+ respectively. For both parameters, the groups related to progression of tumor disease (III/IV and N+) had significantly higher mean values of EGF/EGFR-mediated EMT Risk Score compared to patients in stage I/II or with no lymph node involvement. Further, the patient group denoted as "Risk+" had enriched numbers of stage III/IV and N+ patients (Fig. 4.13, A). To assure that the prognostic value of the Risk Score is not driven by a potential substratification of primary tumor localization groups, a Kaplan-Meier with a Cox model stratified by the risk groups and tumor primary sites was plotted. The subgroup of oropharyngeal cancers consisted of n = 3 patients in "Risk-" and n = 8

in "Risk+". Patients suffering from larynx carcinomas consisted of n = 30 "Risk-"- and n = 40 "Risk+"-patients. Oral cavity carcinomas, the most abundant sublocalization within the investigated cohort, included n = 87 "Risk-"- and n = 71 "Risk+"-patients. The plotted Kaplan-Meier and the significant Cox model (log-rank p-value = 0.00019) show that the tumor primary site is not a surrogate for the Risk Score because all "Risk-"-subgroups show a better survival compared to the "Risk+"-patients in the respective localization group (Fig. 4.13, B). Thus, the computed EGF/EGFR-mediated EMT Risk Score is a prognosticator that is independent of tumor primary site and is positively associated to AJCC staging and lymph node metastasis status in the HPV-negative patients of the TCGA HNSCC cohort.

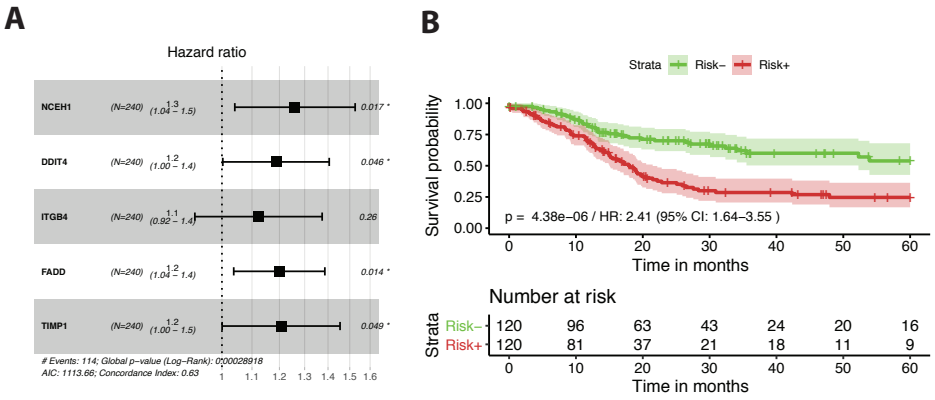

Figure 4.12: EGF/EGFR-mediated EMT Risk Score of TCGA patients. A) Forest plot of a multivariate Cox model after feature selection with RBsurv. Number of events, global log-rank p-value, AIC, and Concordance index are shown. B) Kaplan-Meier curve with 95% CI, HR, HR 95% CI, and log-rank p-value of the Cox model. Patients are stratified into Risk- and Risk+ by the median of the Risk Score. Risk- means below median. Risk+ means above median. Numbers at risk are shown in the table. Overall survival over 5 years of clinical follow-up with time in months were implemented in a Kaplan-Meier curve and Cox models. CI: Confidence intervall, HR: Hazard ratio, AIC: Akaike information criterion, EGF/R: Epidermal growth factor/ receptor, EMT: Epithelial-to-Mesenchymal transition, TCGA: The Cancer Genome Atlas.

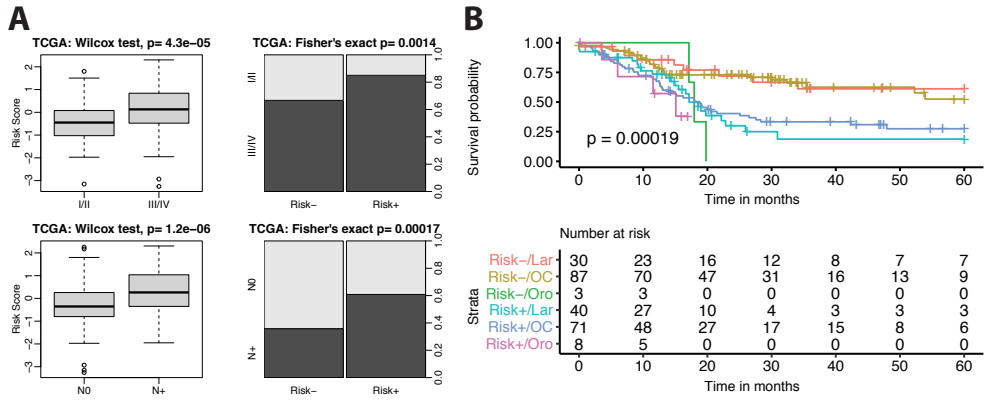

Figure 4.13: EGF/EGFR-mediated EMT Risk Score of TCGA patients. A) Comparisons of the Risk Score with clinical stage and lymph node status. Differences in Risk Score mean expression values between groups were compared by a Wilcox test. P-value is shown. Respective patient numbers in subgroups defined according to median Risk Score values (below: Risk- and above: Risk+) and clinical parameters were compared by Fisher's exact test. P-value is shown. Clinical stages were summarized into I/II and III/IV. N0 means no nodal involvement. N+ means lymph node metastasis of any stage (I-III). B) Kaplan-Meier curve with Cox model log-rank p-value stratified according to Risk subgroups and tumor primary site. Numbers at risk are shown in the table. Overall survival over 5 years of clinical follow-up with time in months were implemented in Kaplan-Meier and Cox model. EGF/R: Epidermal growth factor/ receptor, EMT: Epithelial-to-Mesenchymal transition, TCGA: The Cancer Genome Atlas. Lar: Larynx, OC: Oral cavity, Oro: Oropharynx.

4.4.1 Validation of EGF/EGFR-mediated EMT Risk Score in independent HNSCC cohorts

The validity of the EGF/EGFR-mediated EMT Risk Score was tested in two additional publicly available HNSCC patients cohorts. One validation cohort was from the Fred Hutchinson Cancer Research Center (FHCRC) and the other from the University of Texas MD Anderson Cancer Center (MDACC). The FHCRC and the MDACC data set consisted of HPV-negative oral cavity tumor patients and data of overall survival was available for both cohorts. The overall survival over five years of clinical follow-up was not significantly different across all three evaluated data sets (TCGA, FHCRC, and MDACC) as illustrated by the Kaplan-Meier curve with a Cox log-rank p-value in the Appendix Fig. 7. Further, the transcriptomes of the FHCRC and MDACC patients were quantified with cDNA microarrays and expression data was available for all of the five genes implemented in the multivariate Cox model forming the EGF/EGFR-mediated EMT Risk Score in TCGA patients (Fig. 4.12). The RNA expression values of the FHCRC and MDACC cohorts were weighted by the coefficients of the Cox model previously defined in the TCGA discovery cohort. The resulting values were summed up, centered, and scaled to form an EGF/EGFR-mediated EMT Risk Score. To stratify the FHCRC and MDACC cohorts into "Risk-" and "Risk+", the median Risk Score value from the TCGA data set was applied. To validate the prognostic value of the EGF/EGFR-mediated EMT Risk Score, a Kaplan-Meier curve including a Cox model was plotted. The "Risk+"-patients in the FHCRC cohort were characterized by a 2.28 elevated relative hazard risk and in the MDACC cohort the "Risk+" were marked by a 4.51 elevated risk of death (Fig. 4.14). Therefore, the prognostic value of the EGF/EGFR-mediated EMT Risk Score consisting of weighted expression of the genes NCEH1, DDIT4, ITGβ4, FADD, and TIMP1 could be confirmed in two independent cohorts using a different technique to quantify gene transcription, *i.e.* cDNA microarray measurements versus RNASeq.

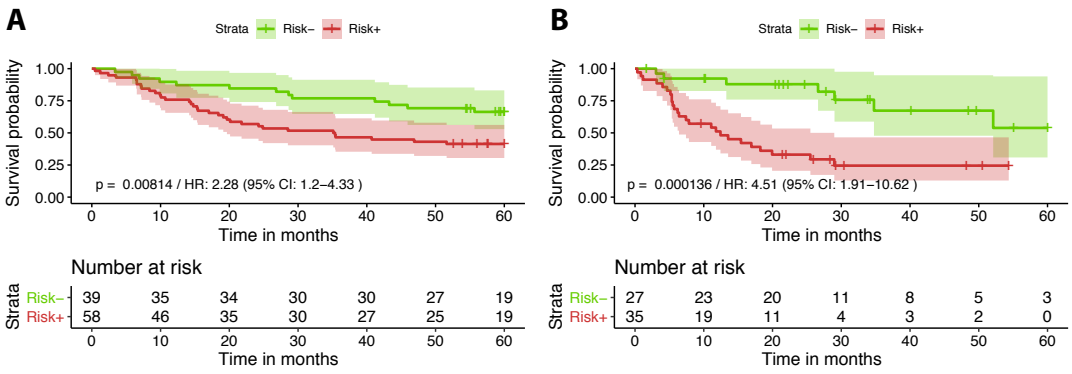

Figure 4.14: Validation of EGF/EGFR-mediated EMT Risk Score in the FHCRC and MDACC HNSCC cohorts. Risk Score was calculated as sum of weighted gene expression according to the multivariate Cox model computed with the TCGA data. Risk subgroups were calculated according to the median Risk Score value from the TCGA HNSCC cohort. Risk- means below and Risk+ means above TCGA median Risk Score. A, B) Kaplan-Meier curves with 95% CI, HRs, HR 95% CIs, and log-rank p-values of the Cox models of the (A) FHCRC and the (B) MDACC cohort. Numbers at risk are shown in the table. Overall survival over 5 years of clinical follow-up with time in months was implemented. CI: Confidence intervall, HR: Hazard ratio, EGF/R: Epidermal growth factor/ receptor, EMT: Epithelial-to-Mesenchymal transition, FHCRC: Fred Hutchinson Cancer Research Center, MDACC: University of Texas MD Anderson Cancer Center, TCGA: The Cancer Genome Atlas.

4.5 Evaluation of EGF/EGFR-mediated EMT pathway activity

So far, five genes were identified that were up-regulated in EGF-treated HNSCC cell lines, associated to an increase of Cox HR of death, and incorporated to form one single prognostic Risk Score in three distinct HPV-negative HNSCC patient cohorts. Those five genes might thus reflect the EMT status in cell lines as well as in patients but do not provide in-depth insights into any interaction network or pathway activity. To identify known interactions, the STRING data base (version 11) was used to extract neighbors of the EGF/EGFR-mediated EMT genes (NCEH1, DDIT4, ITGβ4, FADD, and TIMP1), representing interaction partners at the protein level. To avoid bloating of the network and only subtract confident interactions, a standard cut-off of ≥ 400 was applied to the combined score, which quantifies interaction knowledge defined by the STRING data base. The combined score integrates and ranks protein associations by benchmarking against a common reference (Mering et al. 2005). The identified network was translated back to gene names and consisted of n = 808 unique genes, including the five EMT genes. This list of n = 808 genes was then used to extract the RNA expression profiles of the investigated TCGA patient cohort, as described before. The transcriptional data was log2 transformed. Now, the extracted interaction network of n = 808 genes with their expression profiles was investigated for any pathway regulation related to tumor progression. To identify possible cancer-related pathway activity in those n = 240 TCGA patients, the Pathway RespOnsive GEnes (PROGENy) approach was chosen because it was shown to perform better than other methods in a wider range of conditions by Schubert and colleagues (Schubert et al. 2018). Further, the authors of PROGENy could show that pathway activity is likely to be not reliant on transcription of pathway-related signaling molecules (Schubert et al. 2018), a prominent assumption of comparable methods. Since the extracted transcriptional interaction network related to EGF/EGFR-mediated EMT of a patient cohort should be investigated, consisting of a heterogeneous group of patients at different time points in disease progression, choosing the robust PROGENy approach

over methods, as for example GSEA, seemed adequate. Until now, PROGENy lists 13 pathways; Androgen, EGFR, Estrogen, Hypoxia, Jak-STAT, MAPK, NfκB, p53, PI3K, TGFβ, TNFα, VEGF, and WNT. The pathway activities were quantified based on the extracted EGF/EGFR-mediated transcription network for each patient. To relate the EGF/EGFR-mediated EMT Risk Score with the computed pathway activity quantities, generalized linear models were built and the estimates with p-values are illustrated in Fig. 4.15. The Risk Score showed a positive association with the EGFR (estimate = 0.218, p-val = 0.00067), Estrogen (estimate = 0.3, p-val $<$ 0.0001), Hypoxia (estimate = 0.302, p-val $<$ 0.0001), MAPK (estimate = 0.244, p-val = 0.00013), NfκB (estimate = 0.307, p-val $<$ 0.0001), TGFβ (estimate = 0.264, p-val $<$ 0.0001), and TNFα (estimate = 0.375, p-val $<$ 0.0001) pathway activities. All other pathways activities were not significantly associated with the Risk Score. In Fig. 4.15 the EGF/EGFR-mediated EMT Risk Score was compared to computed pathway activities. Using the Risk Score median, the TCGA cohort was dichotomized into "Risk-"-and "Risk+"-patients as previously. Next, it was investigated whether the positive associations of the Risk Score with higher activity in pathways was reflected in the dichotomized subgroups. Compared to "Risk-", "Risk+"-patients are marked by higher activities of the Estrogen (estimate = 0.386, p-value = 0.001), Hypoxia (estimate = 0.713, p-value $<$ 0.001), MAPK (estimate = 0.311, p-value = 0.023), NfκB (estimate = 0.412, p-value = 0.002), TGFβ (estimate = 0.360, p-value = 0.003), and TNFα (estimate = 0.503, p-value $<$ 0.001) pathways. The EGFR pathway activity was enriched (estimate = 0.288, p-value = 0.053) as was the Jak-STAT pathway activity (estimate = -0.199, p-values = 0.097) but the changes in activity did not meet the criteria to be accepted as significant (Tab. 4.1).

Table 4.1: PROGENy pathway activities in TCGA patients. Risk- and Risk+ subgroups are compared. Results for Risk+ patients (above median Risk Score) are shown. Estimated coefficients of linear models with SE, t-statistic, resulting p-value, and analyzed pathway are shown. p-value: 0.1- 0.05, ., $\leq$ 0.05, *, $\leq$ 0.01, **, $\leq$ 0.001, ***, $\leq$ 0.0001, ****. SE: Standard error, TCGA: The Cancer Genome Atlas.

estimate	std.error	statistic	p.value	pathway	signif
0.041	0.130	0.317	0.751	Androgen	
0.288	0.148	1.948	0.053	EGFR	.
0.386	0.119	3.258	0.001	Estrogen	**
0.713	0.125	5.692	0.000	Hypoxia	***
-0.199	0.120	-1.664	0.097	JAK-STAT	.
0.311	0.136	2.296	0.023	MAPK	*
0.412	0.131	3.149	0.002	NFkB	**
-0.111	0.127	-0.873	0.384	p53	
-0.048	0.128	-0.380	0.704	PI3K	
0.360	0.121	2.979	0.003	TGFb	**
0.503	0.126	3.982	0.000	TNFa	***
0.018	0.121	0.144	0.885	VEGF	
-0.212	0.160	-1.330	0.185	WNT	

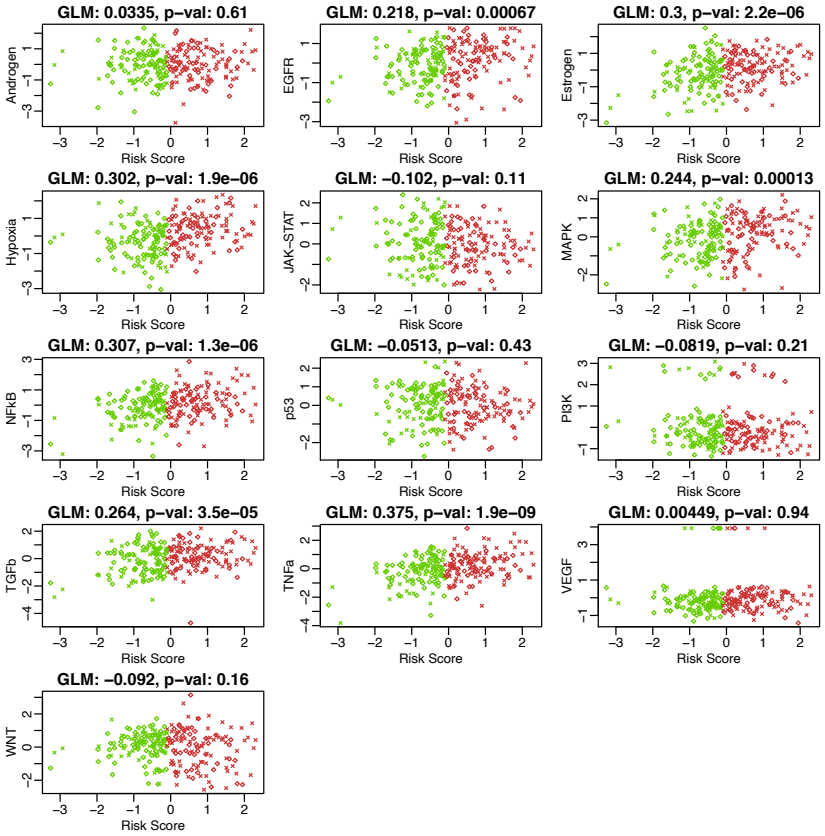

Figure 4.15: PROGENy pathway activity scores were scaled across all TCGA patients and compared to the EGF/EGFR-mediated EMT Risk Score. GLM coefficients with p-values are shown. Pathway activity scores are depicted on the y-axis, Risk Scores on the x-axis. Crosses represent patients with an event during 5 years of clinical follow-up. Squares represent patients without events. Green means patients below Risk Score median, red means above median. EGF/R: Epidermal growth factor/ receptor, EMT: Epithelial-to-Mesenchymal transition, TCGA: The Cancer Genome Atlas, GLM: Generalized Linear Model.

4.6 Identification of a pEMT mediator in HNSCC patients

Recently, a gene signature was defined describing partial EMT in oral cavity HNSCC patients based upon single cell RNASeq. (Puram et al. 2017). The authors could show that the pEMT associates with disease progression in the basal-like and mesenchymal subtype. The pEMT program was linked to clinical parameters, such as nodal metastasis or tumor grade. Cells expressing high levels of the genes of the pEMT signature were often found at the primary tumor edge. Further, it was shown that a subgroup of patients of the mesenchymal subtype were marked by an excessive influence of non-malignant cells distorting the results of bulk sequencing of tumors. In the present work, this pEMT signature was quantified in HNSCC patients and the prognostic value with associations to disease progression was assessed.

4.6.1 Evaluating expression profiles of TCGA patients

All HPV-negative TCGA patients available (n = 240) were implemented in the analysis. The previously identified molecular subgroups of the patients (Lawrence et al. 2015), atypical, basal-like, classical, and mesenchymal, were further refined by a Pearson correlation analysis implementing the top 10,000 protein coding genes across all patients in accordance to Puram *et al.* (Puram et al. 2017). Briefly, patients with a Pearson correlation of > 0.1 within their respective molecular subtype group and < 0.1 correlation towards all other subtypes were kept. This selection refined the number of patients to n = 125 with good transcriptome correlation within their respective subgroup (atypical = 19, basal-like = 38, classical = 22, mesenchymal = 46, Fig. 4.16). It has been described, that tumor bulk sequencing approaches are prone to the influence of non-cancerous cells and this was shown for the TCGA cohort (Puram et al. 2017). In order to minimize the influence of non-malignant cells within the analyzed tumor bulks, marker genes for T-cells, fibroblasts, macrophages, dendritic cells, endothelial cells, and B/plasma-cells were used to conduct hierarchical clustering to identify patients with increased amounts of non-cancerous cells. The cluster analysis resulted in two main clusters, one representing

bulks with very low influence of non-malignant marker genes tested and the other with influences to different degrees. The cluster with influence from non-malignant cells in the tumor bulk could be substratified into two distinct clusters, where one was marked by relatively high gene expression from non-malignant cell markers (Fig. 4.16). This group consistent solely of patients of the mesenchymal molecular subtype, which is in accordance with previous findings (Puram et al. 2017). The patients from this cluster were excluded from the following analysis and, thus, a sub-cohort of n = 55 HPV-negative TCGA patients of the basal-like and mesenchymal subtype with relatively low influence of non-malignant cells was identified (Fig. 4.16).

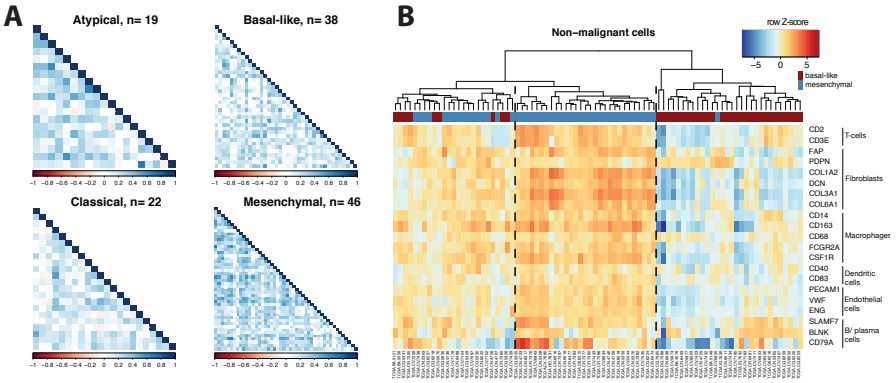

Figure 4.16: TCGA patient selection for pEMT quantification by SING scoring. A) Pearson correlation matrices of selected HPV-negative TCGA patients' transcriptomes. Top 10,000 protein coding genes expressed were analyzed. Numbers of patients per molecular subtype are noted. Color encodes for Pearson coefficients. Insignificant p-values are blanked. Significance niveau ≤ 0.01. Patients with correlation > 0.1 within molecular subtype and correlation < 0.1 to others were kept and analyzed in (B). B) Heatmap with hierachical clustering of basal-like and mesenchymal patients based on non-malignant cell marker genes expression. Heatmap colors define gene expression Z-scores across rows. Red marks basal-like and blue marks mesenchymal patients. Marker genes in rows are noted on the right side. TCGA sample ID is noted below the heatmap. Black dotted lines mark subcluster with high non-malignant cell marker genes expressions. pEMT: Partial Epithelial-to-Mesenchymal transition, HPV: Human papillomavirus, TCGA: The Cancer Genome Atlas, SING: Single sample scoring molecular phenotype. Adapted from Schinke *et al.*, 2020.

4.7 Quantification of pEMT by SING scoring

Different gene set scoring methods are available to quantify the concordance of gene signatures and transcriptomes. Techniques to define stable scores for individual samples are relativy scarce and often susceptible to instability. In 2018, a research group developed the Single sample scoring of molecular phenotypes (SING score) method, which is capable of providing stable quantification scores on an individual level and outperformed comparable approaches defining phenotypic landscapes (Foroutan et al. 2018). The SING score method was used here to quantify pEMT in the extracted n = 55 TCGA patients by applying the common pEMT signature of n = 15 genes as defined by Puram *et al.* and using the background of the top 10,000 protein coding genes expressed across all patients (common pEMT genes: SERPINE1, TGFβI, MMP10, LAMC2, P4HA2, PDPN, ITGα5, LAMA3, CDH13, TNC, MMP2, EMP3, INHBA, LAMB3, VIM). The pEMT SING score was used to stratify TCGA patients into "low", "medium", and "high" subgroups, representing the lowest 25 %, intermediate 50 %, and highest 25 %. Using the pEMT SING score as a feature in a Cox model of overall survival showed prognostic value of the method (Cox model coefficient = 6.67, log-rank p-value = 0.04, data not shown), which was further reflected by a Kaplan-Meier curve with a Cox model of the "low", "medium", and "high" pEMT SING score groups as stratificators. Patients described as "high" had a significantly worse overall survival prognosis compared to "medium" and "low" patients (pairwise log-rank p-values: "high" vs. "medium" = 0.04, "high" vs. "low" = 0.005, data not shown) and the overall model log-rank p-value was 0.0033 (Fig. 4.17 A). To estimate the false-positive rate (α-error), 15 randomly picked gene sets excluding pEMT genes were tested. The pEMT SING score of the random gene set was computed and the performance was tested in a Cox model. This procedure was repeated 10,000 times and the proportion of relevant results from a random pEMT SING score according to a log-rank p-value ≤ 0.05 with a HR > 1 was 4.18 %, representing an α-error in the acceptable range below 5 % (Fig. 4.17 B). To test if the pEMT SING score is correlated with nodal metastasis, the pEMT SING score was plotted against the lymph node status

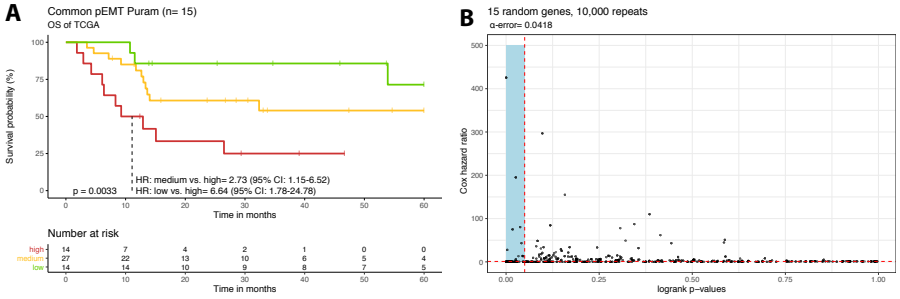

Figure 4.17: Quantification of pEMT in selected TCGA patients by SING scoring. A) Kaplan-Meier curves with median survival times and Cox model log-rank p-value of TCGA patients dichotomized according to pEMT SING score values. HRs with 95% CIs comparing subgroups are noted. Numbers at risk are shown in the table. B) α-error rate of pEMT SING score was assessed by computing SING scores of 10,000 randomly picked gene sets of 15 genes. Random SING scores were applied to a Cox model. Cox models featuring random SING scores with a HR > 1 and p-value ≤ 0.05 (red dotted lines) were accepted as relevant (blue box). Overall survival over 5 years of clinical follow-up with time in months was assessed. CI: Confidence intervall, HR: Hazard ratio, pEMT: Partial Epithelial-to-Mesenchymal transition, HPV: Human papillomavirus, SING: Single sample scoring molecular phenotype, TCGA: The Cancer Genome Atlas. Adapted from Schinke *et al.*, 2020.

of the analyzed patients (N0 = 18 patients, N1 = 15 patients, N2 = 10 patients, N3 = 1). A correlation analysis showed that the pEMT SING score was positively associated with nodal metastasis (Kendall's $\tau = 0.30$, p-value = 0.005) and significant differences between the mean levels of pEMT SING score of the different N-statuses were observed (Kruskal-Wallis = 0.026, Fig. 4.18). To validate the prognostic value of the SING score method, the MDACC cohort was investigated as a second independent cohort. Again, the same background of formerly defined 10,000 protein coding genes expressed in the MDACC patients was applied to quantify the pEMT. In contrast to the TCGA cohort, the resulting pEMT SING scores for the majority of patients from the MDACC validation cohort was scored below 0. This means that the respective overall ranks of the pEMT genes within the MDACC cohort are lower compared to the TCGA cohort. A

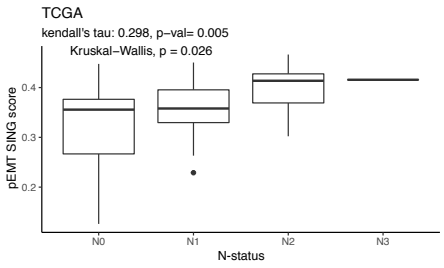

Figure 4.18: pEMT SING score and lymph node status. Boxplot showing median as line, 1. to 3. quartile as box, and 1.5x interquartile range whiskers of pEMT SING scores of different lymph node statuses (N-status) in TCGA patients. Kruskal-Wallis test shows difference in means across N-statuses. Kendall's τ shows positive association of N-status and pEMT SING score. N0 = 18 patients, N1 = 15 patients, N2 = 10 patients, N3 = 1 patient. pEMT: Partial Epithelial-to-Mesenchymal transition, SING: Single sample scoring molecular phenotype, TCGA: The Cancer Genome Atlas. Adapted from Schinke *et al.*, 2020.

quantile-quantile plot of the pEMT SING score for both cohorts, TCGA and MDACC, shows that the pEMT SING score is well described by a normal distribution and high absolute scores are associated to low gene dispersion, as measured by the median absolute deviation (Fig. 4.19). As for the TCGA patients, the pEMT SING score in the MDACC cohort was tested in a Cox model before stratifying the patients into the three subgroups. The Cox model was significant with a log-rank p-value of 0.0002 (Cox model coefficient = 7.24, data not shown). The prognostic value of the pEMT SING score was confirmed by a Kaplan-Meier curve with a Cox model showing an overall log-rank p-value of 0.00048, comparing the "low" (lowest 25 %), "medium" (intermediate 50 %), and "high" (highest 25 %) patients in accordance to the TCGA patients (Fig. 4.20). In a pairwise log-rank comparison, the p-value for "high" vs. "medium" was 0.025, for "high" vs. "low" it was 0.001, and for "medium" vs. "low" it was 0.025 (data not shown).

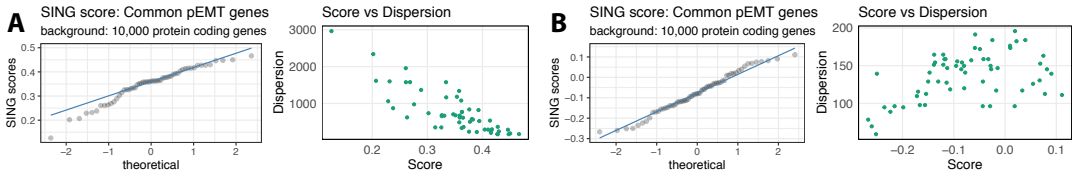

Figure 4.19: Metrics of quantification of pEMT by SING scoring in the (A) TCGA and (B) MDACC cohorts. Left side shows a quantile-quantile plot with bisector (blue line) of computed pEMT SING scores. Right panel shows distribution of dispersion and scores. Dispersion is defined by median absolute deviance. pEMT: Partial Epithelial-to-Mesenchymal transition, SING: Single sample scoring molecular phenotype, TCGA: The Cancer Genome Atlas, MDACC: University of Texas MD Anderson Cancer Center. Adapted from Schinke *et al.*, 2020.

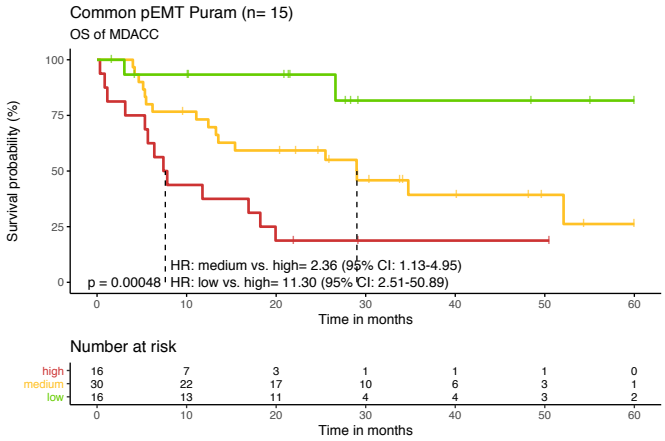

Figure 4.20: Confirmation of prognostic value of pEMT SING score in the MDACC cohort. Kaplan-Meier curves with median survival times and Cox model log-rank p-value of MDACC patients dichotomized according to pEMT SING score values. HRs with 95% CIs comparing subgroups are noted. Overall survival over 5 years of clinical follow-up with time in months was assessed. Numbers at risk are shown in the table. CI: Confidence intervalls, HR: Hazard ratio, pEMT: Partial Epithelial-to-Mesenchymal transition, SING: Single sample scoring molecular phenotype, MDACC: University of Texas MD Anderson Cancer Center. Adapted from Schinke *et al.*, 2020.

4.8 Identification of potential regulators of pEMT

In order to address whether the pEMT SING score does actually quantify a partial differentiation status of tumor cells that is characterized by a retention of epithelial markers and gained expression of mesenchymal markers, epithelial and mesenchymal markers were plotted along the quantified pEMT spectrum. To do so, the patients from the TCGA and MDACC cohorts were ranked according to the pEMT SING score from low to high and marker gene expression was plotted as a line graph with smoothing using Locally Estimated Scatterplot Smoothing (LOESS). Chosen as epithelial markers were: ECAD, as a standard marker, CLDN7, a variable epithelial marker, EpCAM, used to define epithelial phenotype in cell lines, and KRT14, a non-variable pEMT marker, all four were described in literature (Puram et al. 2017). RAB25 was described as a top five epithelial marker defined in cancer patients (Tan et al. 2014). As seen in the plots, RAB25 represented the only epithelial marker that clearly decreased with increasing pEMT rank in both cohorts, TCGA and MDACC. CLDN7 and ECAD were showing a mild but not consistent decrease associated with the pEMT rank. KRT14 expression levels were showing no substantial decrease in both cohorts. EpCAM was mildly increasing in the TCGA patients with increasing pEMT. Conclusively, no clear and consistent decrease for epithelial markers according to the pEMT rank, except RAB25, was depicted in the TCGA and MDACC patients (Fig. 4.21 A). To investigate the mesenchymal status, VIM and FN1, two standard mesenchymal markers in HNSCC, and the canonical EMT transcriptions factors reported in the expression data, SLUG, ZEB1, and ZEB2, were plotted. VIM was not consistently increasing with pEMT across the two cohorts. FN1 was showing a clear increase along the ranked pEMT spectrum. All three EMT transcription factors were increasing with progressing pEMT rank, and SLUG was showing overall higher expression values compared to ZEB1 and ZEB2 (Fig. 4.21 B). Thus, it can be concluded that an increase in pEMT SING score was associated with enhanced FN1 and EMT-TFs expression values along with a retention of epithelial markers except RAB25. The three canonical EMT-TFs SLUG, ZEB1, and ZEB2 were

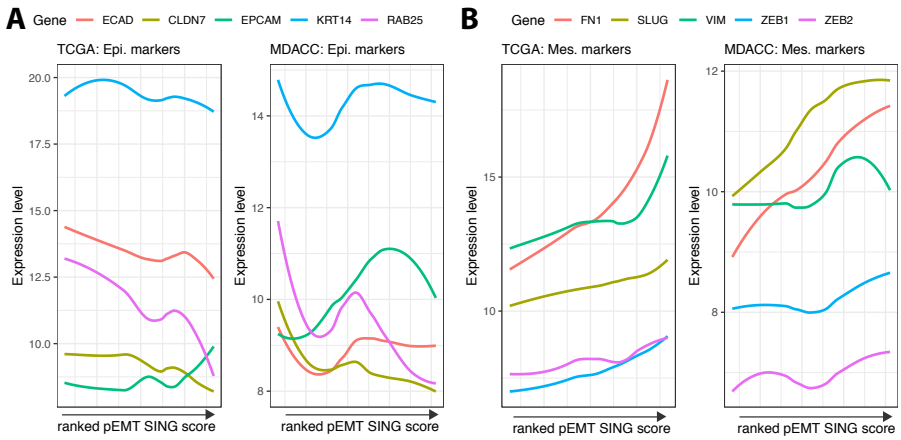

Figure 4.21: pEMT SING score quantifies pEMT. TCGA and MDACC patients were ranked by pEMT SING scores. Log2 gene expression levels of selected epithelial (A) and mesenchymal markers (B) are plotted across increasing ranks of the patients. Line smoothing by LOESS was applied. Color of lines encodes for respective marker gene as shown on the top of the plots. LOESS: Locally weighted scatterplot smoothing, pEMT: Partial Epithelial-to-Mesenchymal transition, SING: Single sample scoring molecular phenotype, TCGA: The Cancer Genome Atlas, MDACC: University of Texas MD Anderson Cancer Center. Adapted from Schinke *et al.*, 2020.

found in the top 10,000 protein coding genes of the TCGA and MDACC HNSCC patients. To identify potential regulators of pEMT that are consistently expressed in patients of the TCGA and the MDACC cohorts, SLUG, ZEB1, and ZEB2 were analyzed by a Pearson correlation against all 15 common pEMT genes. Further, the six genes; TGFβI, LAMC2, PDPN, ITGα5, VIM, and MMP10, described as the top pEMT genes, and their correlation to the three EMT-TFs are plotted in further detail with noted non-significant p-values (significance niveau: ≤ 0.01). In both cohorts, SLUG was correlating best with the pEMT genes compared to ZEB1 and ZEB2. Out of the 15 common pEMT genes, only MMP2 and VIM were showing better correlation with ZEB1 and ZEB2, in both patient cohorts. In the TCGA cohort, SLUG was positively correlated to all 15 common pEMT genes. ZEB1 and ZEB2 were not correlated with SERPINE1, MMP10, LAMC2, P4HA2, LAMA3, and LAMB3. SLUG, ZEB1, and ZEB2 were showing positive

correlations with TGFβI, PDPN, ITGα5, CDH13, TNC, MMP2, EMP3, INHBA, and VIM. Except MMP2 and VIM, SLUG was showing the strongest correlation overall. Investigating the top pEMT markers, SLUG was outperforming ZEB1 and ZEB2 for all genes except VIM, as aforementioned (Fig. 4.22). In the MDACC, ITGα5, CDH14, and EMP3 were not significantly correlated with any of the EMT transcription factors. Except those genes, SLUG was showing a strong positive correlation to all remaining common pEMT genes. In contrast, ZEB1 and ZEB2 were not significantly correlated to any of the pEMT genes, except a negative correlation with LAMA3. None of the top pEMT markers, except VIM, showed a significant positive association with ZEB1 or ZEB2. Thus, it is concluded that out of the three EMT-TFs found in the expression data of the TCGA and MDACC patients, SLUG is showing the strongest correlations to the pEMT genes on a single gene level (Fig. 4.22). To investigate whether the correlation

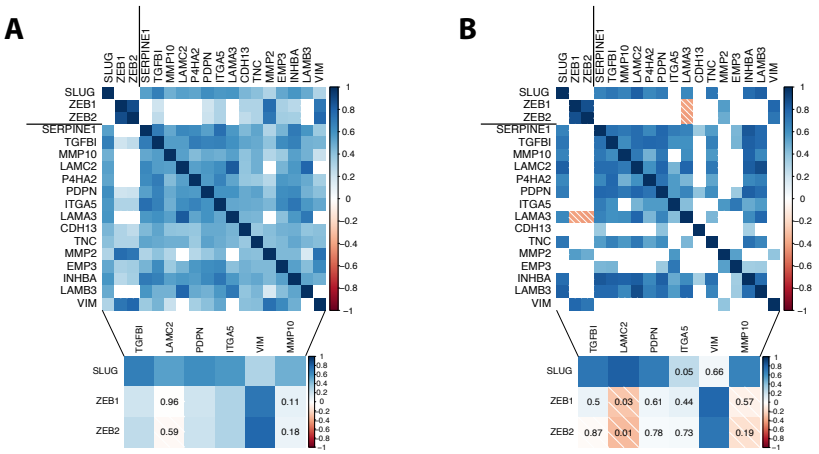

Figure 4.22: Correlation matrices of pEMT and EMT-TF genes in the (A) TCGA and (B) MDACC cohorts. Color encodes for Pearson correlation coefficient. Top panel shows correlation of EMT-TFs and common pEMT genes with insignificant p-values blanked. Lower panel shows correlation of EMT-TFs and top 6 pEMT genes with insignificant p-values stated on the plot. Significance niveau was ≤ 0.01. Black lines in the top panel separate EMT-TFs from pEMT genes. pEMT: Partial Epithelial-to-Mesenchymal transition, SING: Single sample scoring molecular phenotype, TCGA: The Cancer Genome Atlas, MDACC: University of Texas MD Anderson Cancer Center, EMT-TFs: EMT transcription factors. Adapted from Schinke *et al.*, 2020.

from a gene to gene comparison is also reflected in the pEMT SING score, expression levels of SLUG, ZEB1, and ZEB2 were plotted against pEMT SING score levels. As a quantification of association, a Spearman's rank correlation analysis was conducted. In the TCGA cohort, Spearman's ρ in correlation with the pEMT SING score for SLUG was 0.52 (p-value < 0.0001), for ZEB1 it was 0.39 (p-value $= 0.0034$), and for ZEB2 it was 0.27 (p-value $= 0.045$). In the MDACC cohort, Spearman's ρ was 0.51 for SLUG (p-value < 0.00001), 0.28 for ZEB1 (p-value $= 0.028$), and not significant for ZEB2 (ρ $= 0.21$, p-value $= 0.11$) (Fig. 4.23). Thus, in both HNSCC patient cohorts, SLUG is showing the best correlation with the pEMT genes on a single gene level and outperforms ZEB1 and ZEB2 when associating the pEMT SING score with their respective expression levels.

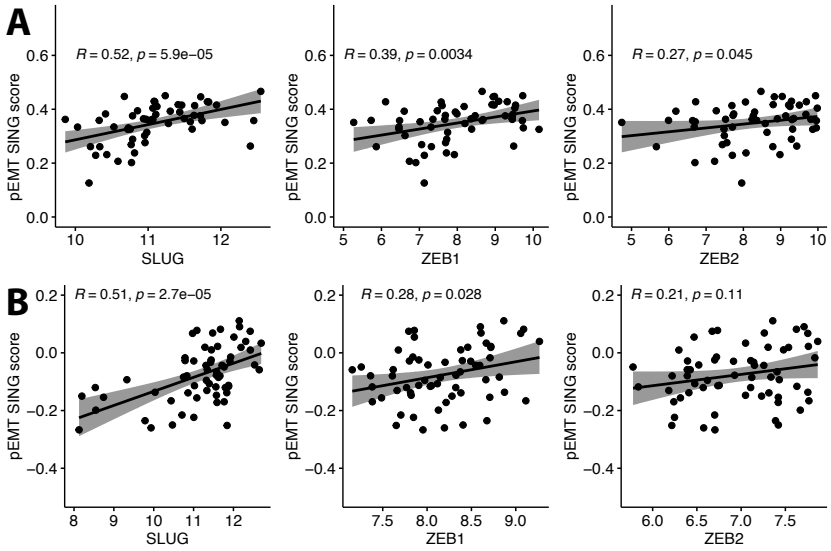

Figure 4.23: Spearman's rank correlation of pEMT SING scores with single EMT-TFs for the TCGA (A) and MDACC (B) cohorts. Spearman's ρ and corresponding p-values are depicted. Linear regression line with 95% confidence intervall is plotted. pEMT: Partial Epithelial-to-Mesenchymal transition, SING: Single sample scoring molecular phenotype, TCGA: The Cancer Genome Atlas, MDACC: University of Texas MD Anderson Cancer Center, EMT-TFs: EMT transcription factors. Adapted from Schinke *et al.*, 2020.

4.9 SLUG induces a pEMT phenotype *in vitro*

Since SLUG represents the best candidate for an induction of pEMT relevant in HNSCC patients, it was overexpressed in two cell lines of the head an neck region, FaDu and Kyse30, to explore effects on cellular phenotypes and behavior functions. Exogenous expression of SLUG using a 141-pCAG-3SIP plasmid with a CMV promoter and selecting for puromycin resistance led to a clear induction of SLUG protein expression compared to empty vector control cell lines (Fig. 4.24 A). Using confocal imaging and immunostaining, the exogenously expressed SLUG protein was found located in the nucleus comparable to physiological conditions. Compared to the control (Ctrl) cell lines, the SLUG-over-expressing (SLUG-OE) cells showed much higher SLUG protein levels within the nucleus (Fig. 4.24 B). The cell morphology of FaDu SLUG-OE cells appeared to reflect an epithelial phenotype similar to the Ctrl cell line but with potentially less dense cell-to-cell contacts, as visualized by an increase in phase contrast (Fig. 4.24 C). In Kyse30, the SLUG-OE cells entered a mesenchymal state compared to Ctrl cells as judged by the cellular morphology. Cell-to-cell contacts were substantially reduced and the cells became spindle-shaped (Fig. 4.24 C). Visualizing ECAD on the cell membrane *via* immunofluorescence staining showed a qualitative reduction of membranous ECAD protein in FaDu and, more distinctly, in Kyse30 SLUG-OE cells in relation to the respective Ctrl cell lines (Fig. 4.24 D). To quantify the ECAD reduction in SLUG-OE cells in comparison to Ctrl cells, a western blot from whole cell lysates was performed and quantified from three independent experiments. FaDu SLUG-OE cells were showing a moderate ECAD protein reduction of 16.33 % and Kyse30 SLUG-OE cells a reduction of 83.44 %. Both reductions were statistically significant, indicating a switch towards a more mesenchymal cellular state (Fig. 4.25 A and B). The analysis via qPCR linked an induction of mesenchymal marker VIMENTIN and EMT transcription factor ZEB1 to SLUG-OE compared to Ctrl and wild type cells, confirming the induction of an EMT in the SLUG-OE cells for both cell types, FaDu and Kyse30 (Fig. 4.25 C). Cells that have undergone a full EMT are characterized by a reduction in their proliferation (Diepen-

bruck and Christofori 2016). Proliferation rates of the SLUG-OE cells were determined by automated cell counting using a Leica DMi8 microscope and FIJI software by quantifying Hoechst-stained nuclei. A fold change proliferation (fcp) rate relative to day 0 was calculated. FaDu SLUG-OE cells were showing no reduction of proliferation after 24 and 48 hours compared to the corresponding Ctrl cells (Fig. 4.25 D). Kyse30 SLUG-OE cells were showing a significant but minor reduction in proliferation rate after 24 hours (Ctrl: 2.00 fcp, SLUG-OE: 1.84 fcp) and after 48 hours (Ctrl: 3.39 fcp, SLUG-OE: 3.11 fcp, Fig. 4.25 D). A clear induction of a full mesenchymal phenotype was not linked to SLUG-OE but both cancer cell lines were showing characteristics linked to EMT, thus an induction of a pEMT phenotype is concluded (Fig. 4.25).

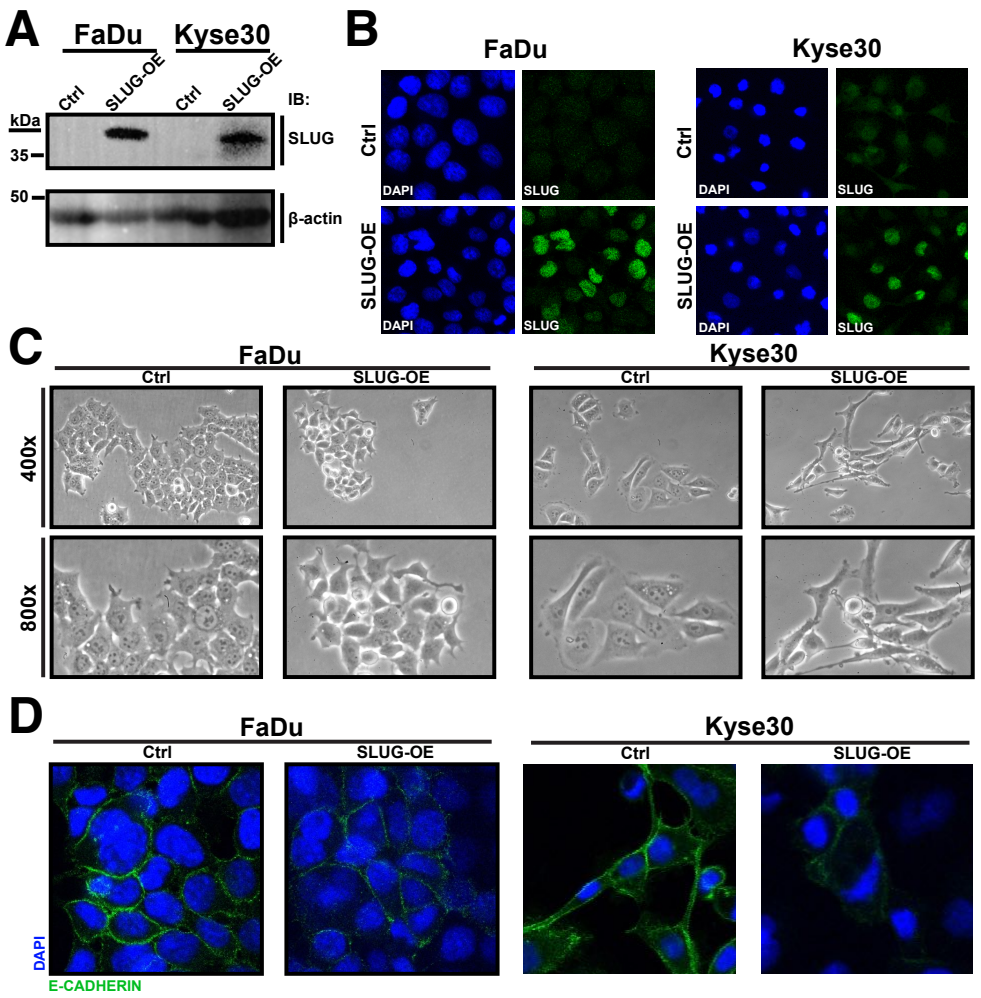

Figure 4.24: SLUG overexpression in FaDu and Kyse30 cells. A) Western blot analysis of Ctrl and SLUG-OE cells β-actin was used as a loading control. B) Representative images of immunofluoresence of SLUG protein in Ctrl and SLUG-OE cells. SLUG protein is shown in green, and DAPI, marking nuclei, in blue. C) 400x - 800x magnified micrographs of cellular morphology of FaDu and Kyse30 SLUG-OE and Ctrl cells. D) Immunofluorescence of mebranous E-CADHERIN in Ctrl and SLUG-OE cells from FaDu and Kyse30. All images shown are representatives from n $\geq$ 3 independent experiments. Ctrl: Control, SLUG-OE: SLUG overexpression. Adapted from Schinke *et al.*, 2020.

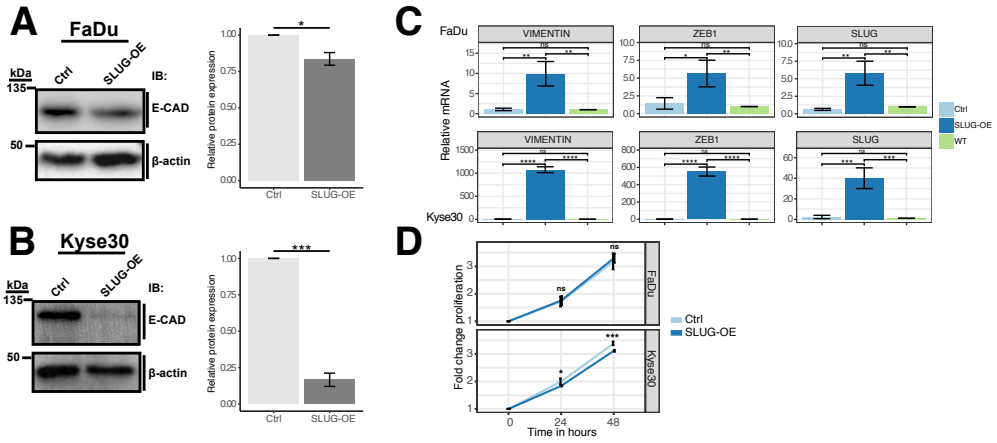

Figure 4.25: Induction of pEMT by SLUG-OE. A) and B) Western blot with quantification of E-CAD protein comparing Ctrl and SLUG-OE in FaDu and Kyse30 cells. β-actin was used as a housekeeping gene. Means compared by Wilcox test: p-value ≤ 0.05,* ; ≤ 0.01, **; ≤ 0.001, ***. C) qPCR data of VIMENTIN, ZEB1, and SLUG mRNA expression in Ctrl, SLUG-OE, and WT cells. Values were normalized to Ctrl cells and were analyzed using the $\Delta\Delta$Ct method. Normalized relative mRNA levels are plotted. Upper panel shows FaDu cells. Lower panel shows Kyse30 cells. D) Proliferation analysis by automated cell counting from 0 to 48 hours. Values were normalized to 0 hours. Differences in mean values were compared by one-way ANOVA with Tukey HSD. p-values: ns, not significant; ≤ 0.05, *; ≤ 0.01, **; ≤ 0.001, ***. Upper panel shows FaDu cells. Lower panel shows Kyse30 cells. Shown are results from n $\geq$ 3 independent experiments. Ctrl: Control, SLUG-OE: SLUG overexpression, IB: Immunoblot, E-CAD: E-CADHERIN, WT: Wild type. Adapted from Schinke *et al.*, 2020.

4.9.1 Cells in pEMT show increased invasion and irradiation resistance

An increased invasion of cancer cells into the surrounding connective tissue by overcoming the barriers of the extracellular matrix is a result of EMT (Santamaria et al. 2017; Lu et al. 2003; Hay 1995). The SLUG-OE cells were tested for their invasive potential into human skin fibroblast spheroids. To do so, pre-formed spheroids of skin fibroblasts were co-cultured for 48 and 72 hours with 10,000 Fadu Ctrl, FaDu SLUG-OE, Kyse30 Ctrl or Kyse30 SLUG-OE cells. Co-cultured spheroids were harvested and cryosections were stained with pan-cytokeratin-specific antibodies to visualize carcinoma cells. Cytokeratins are not expressed by human skin firboblasts, which therefore remain unstained. (Fig. 4.26 A). For the Ctrl cell lines of FaDu and Kyse30, a clear invasion into the underlying fibroblast spheroid was neither observed after 48 hours, nor after 72 hours. In contrast, FaDu SLUG-OE cells showed a moderate increase in invasion after 72 hours of co-culture, with singles cells and small aggregates of carcinoma cells infiltrating fibroblast spheroids (Fig. 4.26 B). Effects of SLUG overexpression were more distinct in Kyse30 cells, which displayed a strong invasive phenotype. After 48 hours, Kyse30 SLUG-OE cells had penetrated into the first layers of the fibroblast sphere and the edges between the two distinct cell populations were less sharp compared to the Ctrl cells. After 72 hours this difference was more severe, many Kyse30 SLUG-OE cells were found in the inner area of the fibroblast spheroids (Fig. 4.26 C). In order to quantify the observed effects on spheroid invasion, a matrigel invasion assay was performed. 100,000 FaDu Ctrl, FaDu SLUG-OE, Kyse30 Ctrl, or Kyse30 SLUG-OE cells were seeded in matrigel-coated 24-well membrane chambers and were allowed to invade for 24 hours. The quantification of the invaded cells compared to Ctrl cells showed a 2.31-fold higher invasion compared to FaDU SLUG-OE cells (2.29 ± 0.45 cell/mm^2 and 0.99 ± 0.28 cells/mm^2, respectively). Kyse30 cells revealed to be generally more invasive compared to FaDu cells, with Kyse30 Ctrl cells showing a 6.19-fold higher invasive potential compared to FaDu Ctrl cells. However, matrigel invasion in Kyse30 was significantly increased by 5.10-fold upon expression of SLUG compared to Ctrl cells (31.17 ± 6.23 cells/mm^2 and 6.13 ± 4.45

cells/mm^2, respectively). Hence, the matrigel invasion assay confirmed the enhanced invasion of SLUG-OE cell lines observed in 3D co-cultures with fibroblast spheroids (Fig. 4.26 D). Besides an increase in inasive potential, the EMT program in cancer was linked to irradiation resistance (Lambert, Pattabiraman, and Weinberg 2017; Nieto et al. 2016). In order to test the effects of SLUG overexpression on irradiation, a clonogenic assay was performed. One thousand FaDu cells were plated per well on a 6-well plate and irradiated the next day. After two weeks, clonogenic survival was calculated by measuring the area of colonies of irradiated relative to non-irradiated control plates. To evaluate the area of colonies grown relative to the corresponding untreated control, the ColonyArea Image J Plugin by Guzmán *et al.* was used (Guzmán et al. 2014). With an irradiation dose of 2 Gy, FaDu SLUG-OE cells displayed a colony formation capacity of 42.53 % relative to FaDu SLUG-OE untreated cells (0 Gy). In contrast, non-irradiated FaDu Ctrl cells had a colony formation capacity of 21.96 % compared to untreated FaDu Ctrl cells. For 4 Gy, FaDu SLUG-OE and Ctrl cells recovered 4.91 % and 2.20 %, respectively. For 6 Gy, 0.33 % and 0.18 % was recovered respectively (Fig. 4.26 E). Five thousand Kyse30 cells were plated per well on a 6-well plate and irradiated the next day. After 10 days, clonogenic survival was calculated. Kyse30 Ctrl cells showed less colony formation with 4 and 6 Gy compared to Kyse30 SLUG-OE cells, reporting a significant difference for 4 Gy. After irradiation with 4 Gy, Kyse30 SLUG-OE cells recovered 12.70 % of clonogenic area relative to untreated SLUG-OE-cells and irradiated Ctrls recovered 4.85 % compared to untreated Ctrl cells (Fig. 4.26 F). Conclusively, the overexpression of SLUG was increasing the invasive potential of cancer cells and, at least for FaDu, elevated the resistance to irradiation consistently.

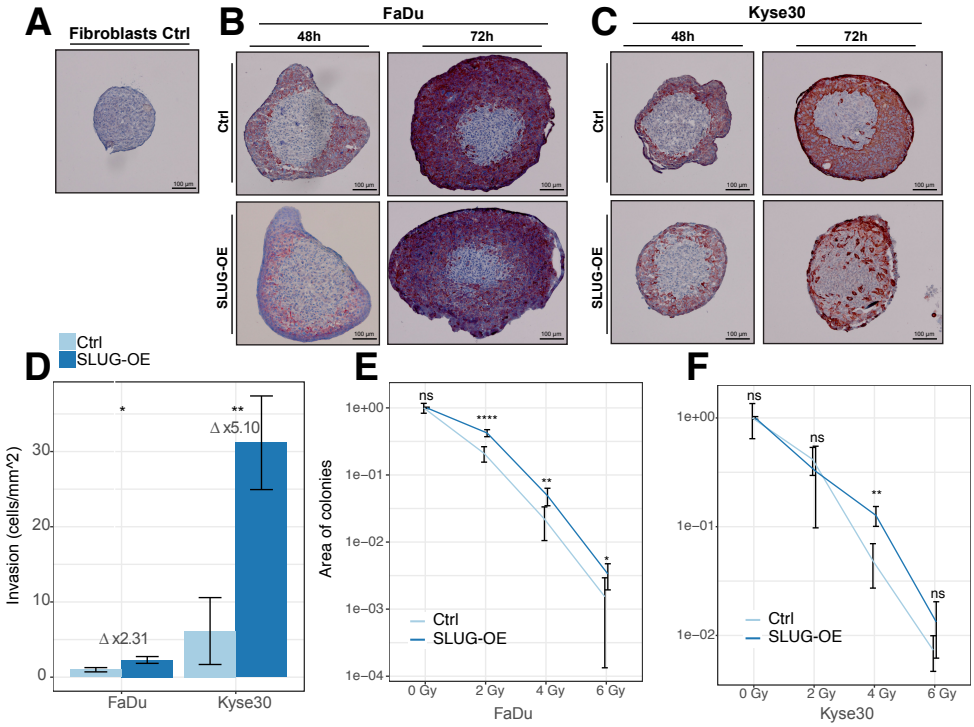

Figure 4.26: SLUG overexpression induces invasion and resistance to irradiation. B)-C) Representative IHC cryosections of fibroblast spheroid invasion by FaDu and Kyse30 cell transfectants (Ctrl and SLUG-OE). Pan-cytokeratine stained cancer cells only, as shown by unstained fibroblasts in Fibroblasts Ctrl panel (A). Invasion was assessed after 48 and 72 hours of 3D co-culture. D) Matrigel invasion assay quantifying the invasion rate. The mean of relative invasion as defined by invaded cells per square mm is plotted with SD values. Student's t-test p-value is shown: p-value ≤ 0.05, *; ≤ 0.01, *. E) Colony formation assay with 0, 2, 4, and 6 Gy (x-axis). Area of colonies were normalized to respective untreated (0 Gy) cells and plotted on a log10 scale on the y-axis. Mean area with SDs of colonies are shown. Student's t-test was used to compare mean values of Ctrl and SLUG-OE cells. p-values are indicated as ns, not significant; ≤ 0.05, *; ≤ 0.01, **; ≤ 0.001,**; ≤ 0.0001, ***; ≤ 0.00001, ****. Results from n ≥ 3 independent experiments are shown. Ctrl: Control, SLUG-OE: SLUG overexpression, Gy: Gray, SD: Standard deviation, IHC: Immunohistochemistry. Adapted from Schinke *et al.*, 2020.

4.10 SLUG in disease progression of HNSCC patients

To explore whether protein levels of SLUG are prognostic, a quantification of immunohistochemistry (IHC) staining of SLUG primary tumor cryosections within a cohort of HNSCC patients from the University Hospital, LMU Munich, Germany (LMU; n = 76 HPV-negative patients) was conducted. Protein levels of SLUG were quantified as described (Mack and Gires 2008). Furthermore, the sublocalizations of SLUG within tumor areas was assessed and categorized as homogeneous or as peripheral at the edge of the tumor area. Examples of the two SLUG localizations are shown in Fig. 4.27. In the

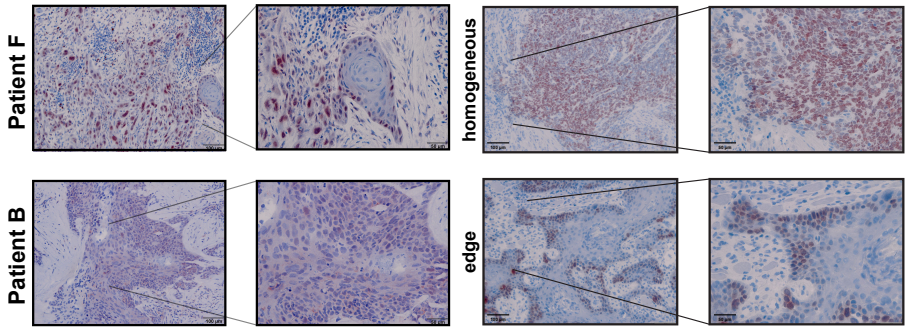

Figure 4.27: SLUG protein staining and sublocalization in tumor sections. Left panel shows representative images from IHC cryosections of SLUG high (patient F) and low (patient B) stainings. Right panel shows representative images of homogenous staining or localization to tumor edges within section. IHC: Immunohistochemistry. Adapted from Schinke *et al.*, 2020.

TCGA data, a positive link between severity of lymph node metastasis and the pEMT SING Score was shown in the present work (Fig. 4.18). And SLUG expression was strongly correlating with pEMT (Fig. 4.22 and 4.23). In the independent in-house LMU HNSCC cohort, patients suffering from recurrence showed a reduced overall survival over five years. Interestingly, patients with tumor recurrence had significantly higher SLUG protein levels (Fig. 4.28 A and B). Further, patients with predominant SLUG staining on the edges of tumor sections showed elevated SLUG levels compared to patients with a homogeneous distribution (Fig. 4.28 C). Additionally, the subgroup of patients with tumor recurrence and SLUG located to the tumor edges significantly expressed two

times more SLUG protein than patients of the homogeneous and recurrence-free subgroup, comparing the mean expression levels (Fig. 4.28 D). This is inline with previous findings about pEMT and SLUG expression in HNSCCs (Puram et al. 2017; Parikh et al. 2019). To further support the link between SLUG and pEMT in patients and disease progression, especially with respect to the formation of lymph node metastasis, n = 16 triplets from patients consisting of normal mucosa, primary tumor, and lymph node metastases were stained to quantify SLUG expression. When comparing the levels of SLUG protein, normal mucosa was showing very low expression levels compared to tumor and lymph node metastases (LNM) tissue (mean/ median levels; mucosa = 6.73/ 3.75, tumor = 69.92/ 66.25, LNM = 97.97/ 102.50). Interestingly, affected lymph nodes from the same patient were showing elevated SLUG expression levels compared to primary tumors (Fig. 4.28 A). Additionally, Disease-Free Survival (DFS) according to SLUG IHC scores was assessed in the whole LMU HNSCC cohort. Patients expressing the lowest levels of SLUG protein (1. quartile, lowest 25%) showed significantly better DFS over five years of clinical follow-up (Fig. 4.28 B). Thus, SLUG is linked to pEMT on a transcriptional level, induces a pEMT phenotype *in vitro*, and is associated with recurrence and degree of lymph node infection at the protein level in patients.

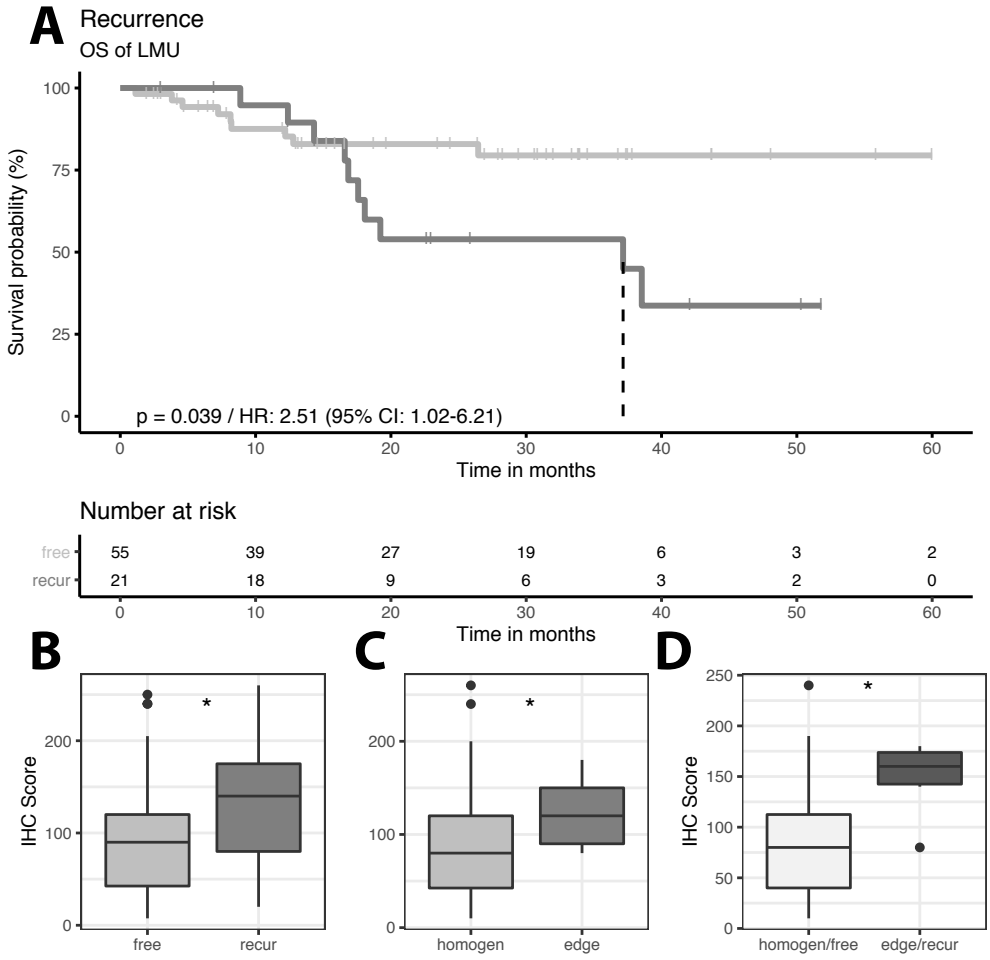

Figure 4.28: SLUG protein is linked to tumor recurrence in HNSCC LMU patients. A) Kaplan-Meier curve with Cox model log-rank p-value, HR, and HR 95% CI of overall survival based on recurrence status over 5 years of clinical follow-up. Numbers at risk are shown in table. B)-D) Box plots with median as line, 1.-3. quartiles as squares, and 1.5x interquartile range as whiskers. Mean values of SLUG IHC scores (y-axis) were compared based on (B) recurrence status, (C) staining localization and D) recurrence free/ homogenous and recurrence/ edge staining localization. Wilcox test p-value is shown. p-values $\leq$ 0.05, *. CI: Confidence intervall, HR: Hazard ratio, OS: overall survival, LMU: Ludwig-Maximilians-University, recur: Recurrence, homogen: Homogeneous, IHC: Immunohistochemistry. Adapted from Schinke *et al.*, 2020.

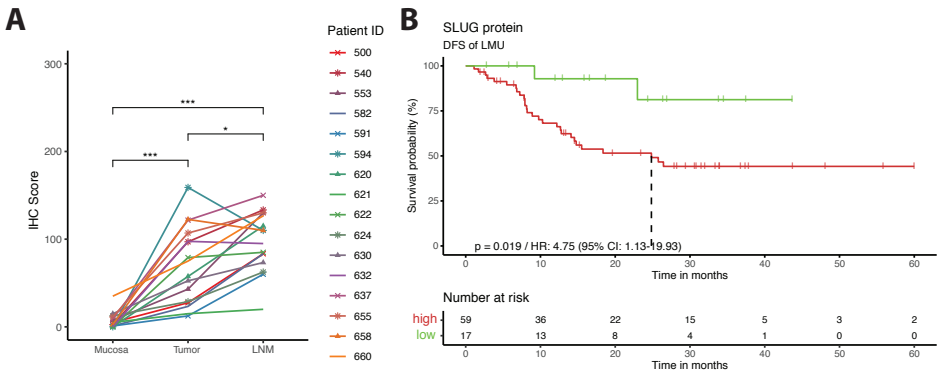

Figure 4.29: SLUG protein is linked to disease progression in patients from the LMU HNSCC cohort. A) SLUG IHC score differences between normal mucosa, primary tumor, and LNM within n = 16 patients from the LMU cohort. Colored lines and symbols encode for n = 16 individual patients. Differences in means were assessed by one-way ANOVA with Tukey HSD. p-values ≤ 0.05, *; ≤ 0.01, **; ≤ 0.001, ***; ≤ 0.0001, ****. B) Kaplan-Meier curve with Cox model log-rank p-value, HR, and HR 95% confidence intervall (CI) of SLUG high (2.-4. quartiles) vs. low (1. quartile) with all patients from the LMU cohort. DFS was assessed over five years of clinical clinical follow. DFS: Disease-Free Surival, HNSCC: Head and neck squamous cell carcinoma, LMU: Ludwig-Maximilians-University, LNM: Lymph node metastases, CI: Confidence intervall, HR: Hazard ratio, IHC: Immunohistochemistry. Adapted from Schinke *et al.*, 2020.

Chapter 5

Discussion

5.1 Analysis of the EGF/EGFR-mediated transcriptome

In HNSCCs, primary tumor heterogeneity is a driver of disease progression and the EGFR-signaling axis is frequently deregulated (Lawrence et al. 2015; Puram et al. 2017; Mroz et al. 2015; Affolter et al. 2016; Horn et al. 2015; Hoshino et al. 1999). Previous work identified pERK as the major mediator of sustained EGFR activity and described it as the main driver of EMT, which itself enables cancer progression and spread (Pan et al. 2018; Lambert, Pattabiraman, and Weinberg 2017; Nieto et al. 2016). Patients expressing high levels of EGFR but low levels of EpCAM were characterized by the poorest survival rates. Simultaneously, patients with low levels of EGFR but high levels of EpCAM were the best survivors. This in combination with *in vitro* findings led to the hypothesis that EpEX, the extracellular domain of EpCAM, is capable of inhibiting EMT induction by reducing pERK levels (Pan et al. 2018). It was unknown whether EpEX acts as a competitive inhibitor of EGF for the binding to EGFR molecules or if it triggers a transcriptional counter-regulation of EMT. Further, the differences of transcriptional regulation between EGF-low and EGF-high treatments were unknown. These questions were addressed within the present doctoral thesis through the RNA sequencing analysis of the four treatment groups EGF-low, EGF-high, EpEX, and EGF(-high) with EpEX at 6 and 72 hours (Experimental overview illustrated by Fig. 4.3). After 6 hours, the EGF-low, EGF-high, EpEX, and EGF with EpEX treatments were showing very similar transcriptomes but different from their respective controls (Figs. 4.4, 4.6, 4.8, 4.9). Until 72 hours, all treatment groups, except EGF-high, had lost their regulatory

capability. Accordingly, the transcriptomes of those treatments were similar to control groups and no DEGs were detected (Figs. 4.4 and 4.7). The EGF-high treatment group was showing the highest number of DEGs and a persistent transcriptional regulation (Figs. 4.7 and 4.11). The reported gene regulation could further be linked to the EMT program. Genes from the Hallmark EMT signature (MSigDB: M5930) were most prominently up-regulated in the EGF-high 72 hours treatment group and the enriched GO terms "biological processes" were further confirming an EMT phenotype (Figs. 4.5 and 4.6). This is in line with the idea that sustained and strong pERK-signaling, as induced by EGF-high, is necessary for EMT induction and transient pERK levels fail to do so (Pan et al. 2018).

The RNASeq-based transcriptome analysis allowed to quantify similarities of gene regulation by comparing the reported DEG expression values with a Pearson correlation analysis. Besides supporting the idea of high transcriptional similarity of all treatment groups, this allowed to conclude that EpEX is not capable of inducing gene regulation countering EMT. After 6 hours, no single counter-regulated gene was identified, neither in the treatment with EpEX alone, nor in the co-treatment with EGF. In fact, the scatter plots with high Pearson correlations reported nearly identical gene regulation. Overall, EpEX is regulating far less genes by itself, which does not exclude the possibility that it would counter-regulate EGF-specific genes to block EMT induction. However, the co-treatment of EpEX and EGF is not showing any counter-regulation and the total numbers of DEGs are in the same range as EGF-low and -high (Figs. 4.8 and 4.9). If any gene transcription would be reduced or counter-regulated actively by EpEX, the Pearson correlation analysis and the scatter plot would be demonstrating it. Theoretically, EpEX could completely block the transcription of certain EMT genes, which then would not be detected in a DE analysis. However, this seems unlikely since EpEX and EGF with EpEX treatments were not showing any outstandingly different gene regulation patterns at 6 hours (Figs. 4.5, 4.4, 4.8, 4.9, 4.11). Not exclusive to the other treatment groups, five genes induced by EpEX were identified throughout both cell lines tested (Fig. 4.10). All of those five genes were regulated in both cell lines and to comparable levels across all

6 hours treatments (data not shown). The RNASeq analysis presented in this doctoral thesis led to the conclusion that EpEX is not actively repressing or counter-regulating gene expression at the transcriptional level induced by EGF treatments. In this work, no evidence of EpEX directly or indirectly supporting EMT was reported, which is in contrast to findings in colon cancer (Liang et al. 2018) but in line with good survival of EGFR-low/EpCAM-high HNSCC patients (Fig. 4.1). This suggests that EpEX is competing with EGF for binding to the extracellular domain of EGFR and is capable of triggering the same signaling axis but to weaker levels as EGF-high. This is in line with recent work from Chen and colleagues using domain mutated EpCAM and EpEX to investigate binding to the extracellular domain of EGFR and signaling cascade activation (Chen et al. 2020). The group could demonstrate that mutant EpEX, lacking the EGF-like domain I, is no longer able to trigger Akt- and MAPK-signaling in colon cancer cells.

The program of EMT in cancer seems to have multifaceted triggers, *e.g.* ligand binding of EGF and TGFβ or hypoxia, with a variety of cellular outcomes (Ye et al. 2017; Aiello et al. 2017; Fischer et al. 2015; Zheng et al. 2015; Lamouille, Xu, and Derynck 2014; Derynck, Muthusamy, and Saeteurn 2014; Zhang et al. 2013; Xu, Lamouille, and Derynck 2009). Since, EGFR represents an important and well described receptor in HNSCCs (Pan et al. 2018; Santini et al. 1991), the RNASeq presented in this work focused on EGFR-signaling induced by EGF as a trigger for EMT. All 6 hours treatments, including treatments that were known to lack the capability of inducing EMT, were showing similar transcriptional profiles. Therefore, further analysis focused on exclusive gene regulation by cells that have undergone EMT and showed a mesenchymal phenotype, *i.e* the EGF-high 72 hours treatment group. By comparing all 6 hours treatments with EGF-high 72 hours treatments, n = 172 DEG candidates consistently regulated by EGF-high treatment after 72 hours in both HNSCC cell lines were identified (Fig. 4.11). Recent research raised the question how well *in vitro* experiments with ligands inducing a full EMT are modeling the non-binary EMT program in cancer progression (Dongre and Weinberg 2019; McFaline-Figueroa et al. 2019). Thus, to transfer *in vitro* results from the present

analysis to patient data and ensure consistency with the hypothesis that EMT is leading to adverse clinical effects, all n = 172 gene candidates were investigated in univariate Cox models using overall survival as an outcome variable and publicly available transcriptome data from HPV-negative HNSCC TCGA patients. If a gene was up-regulated in EMT cells *in vitro*, a HR > 1 in TCGA patients was accepted, and, vice versa, a HR < 1 in TCGA patients if the gene was down-regulated in EGF treated cells. By applying this selection, potential non-physiological *in vitro*-effects in the transcriptome data of treated cancer cell lines, not reflected in the heterogeneous transcriptomes of patients, could be limited. The resulting n = 57 gene candidates were used to form a multivariate survival model assessing EGF/EGFR-mediated EMT effects on patient survival. A forward feature selection method called robust likelihood-based survival modeling (Cho et al. 2009) was used to select survival-associated genes based on the partial likelihood of the Cox model. From the n = 57 genes, a group of five survival-associated genes was identified and implemented into the multivariate Cox model. The Cox model performed well in explaining the overall survival of the n = 240 TCGA patients (global log-rank p-value = 0.000289, Fig. 4.12). Further, the selected genes, namely NCEH1, DDIT4, ITGβ4, FADD, and TIMP1, have been associated with cellular processes related to cancer progression (Chiang et al. 2006; Ellisen et al. 2002; Dyce et al. 2002; Chinnaiyan et al. 1996). ITGβ4, TIMP1, and FADD have been described as negative prognosticators in HNSCCs (González-Moles et al. 2020; Carpén et al. 2019; Kurokawa et al. 2008). High expression of DDIT4 was associated with poorer survival in colorectal cancer, melanoma, breast cancer, glioblastoma, and ovarian cancer (Pinto et al. 2017). The multivariate Cox model coefficients were used to weigh the respective gene expression in cancer patients and to form a cumulative risk factor implementing all five genes. This EGF/EGFR-mediated EMT Risk Score was performing well in Cox models and allowed to stratify HPV-negative HNSCC patients of three independent cohorts (discovery data: TCGA in Fig. 4.12 and validation data: FHCRC and MDACC in Fig. 4.14). The association of the Risk Score with lymph node metastasis and the clinical stage shown in Fig. 4.13 underlines its functional value and corroborates the notion that EMT plays a role in the

metastatic cascade and tumor progression (Stemmler et al. 2019; Lambert, Pattabiraman, and Weinberg 2017; Nieto et al. 2016; Chaffer et al. 2016; Tsai et al. 2012). The transcriptome of the TCGA cohort was quantified by RNASeq and the transcriptomes of the MDACC and FHCRC cohorts by cDNA microarrays. This supports the robustness of the described EGF/EGFR-mediated EMT Risk Score but made it necessary to center and scale the Risk Score because gene expression levels were differing across the data sets. Therefore, so far the Risk Score is not eligible for an implementation into clinical routine for limit value definition to enable risk assessment and to predict risk of EGF/EGFR-mediated EMT of treatment-naive single patients. Sufficiently large and independent patient cohorts with comparable quantification methods would be needed to address the definition of a predictive EGF/EGFR-mediated EMT Risk Score limit value. Describing the whole complexity of the EGFR-signaling cascade by only five genes is not sufficient. In order to further judge the functional value of the EGF/EGFR-mediated EMT Risk Score in accordance to a potential regulatory network involved, the STRING data base was used to built a network around the selected five genes (Szklarczyk et al. 2019). By using the STRING data base, this network is per default restricted to described protein-protein interactions. This network of 808 genes with protein-protein interactions has to be viewed with caution because it assumes a direct proportional relationship between RNA levels and protein expression. Thus, it was essentially only used to extract gene expression data of TCGA patients.

The application of the resulting gene expression matrix to the PROGENy analysis allowed to define cancer-related pathway activity with potential relation to EGF/EGFR-mediated EMT in patients by comparing it to the Risk Score. This comparison was done by building (generalized) linear models with the continuous Risk Score but also by comparing the stratified TCGA patient groups "Risk-" and "Risk+". The Risk Score was positively associated with the EGFR and MAPK pathway but not the PI3K or JAK-STAT pathway (Fig. 4.15 and Tab. 4.1). This supports the functional connection to EGF/EGFR-mediated EMT and corroborates the MAPK-pERK axis (Pan et al. 2018; Blaj et al. 2017; Ichikawa et al. 2015; Shin et al. 2010), not pAkt-signaling (Wang et

al. 2018; Gao et al. 2015), as the mediator of EMT in HNSCCs. The positive correlations of the Risk Score to pathways related to EMT and disease progression in HNSCCs and breast cancers; Estrogen, Hypoxia, NFκB, TFGβ, and TNFα (Dong et al. 2013; Imani et al. 2016; Thiery2009; Diepenbruck2016; Liu et al. 2018), might further hint towards a pleiotropic outcomes of EGF/EGFR-mediated EMT. In HNSCC, common mediators/ modulators of EMT postulated include HIF1-α (hypoxia), NFκB, TFGβ, EGF, and EGFR activation (Chen et al. 2013). In contrast to other cancer types where EGFR gene amplification or mutation is implicated, overexpression of EGFR, without gene amplification, appears to drive pathogenesis in HNSCCs (Lawrence et al. 2015; Rogers et al. 2005; Dassonville et al. 1993). HNSCC studies reported that EGF is secreted by cancer-associated macrophages and EGF protein was located near the epithelium of oral mucosa in the stroma (Gao et al. 2018; Rogers et al. 2005), where increased EGF levels associated with the degree of epithelial malignancy. Further, increased expression of TGFβ in tumor tissue samples was reported in 91 % of cases in an esophageal squamous cell carcinoma study (Talukdar et al. 2020), formally not an HNSCC but in close proximity, and TGFβ can synergize with EGF to activate EGFR-signaling (Uttamsingh et al. 2008). Taken together, those findings suggests that (EGF/)EGFR-mediated EMT is apparent in HNSCC patients and treating cancer cell lines with EGF to assess effects of EMT is a reasonable model and a transition to patient data is eligible.

5.2 A pEMT mediator in basal-like mesenchymal HNSCCs

In the previous part of this work, the EGFR-signaling axis was investigated with a focus on EMT induction and its effects. The analytic approach was strictly focused on EGF/EGFR-mediated EMT and its effects on patient survival. The second part of the present work is focusing on a specific form of EMT, so called partial EMT. Puram *et al.* used a single cell RNASeq experiment and defined pEMT and an associated gene signature in HNSCC. Further, they postulated that the mesenchymal subgroup classification actually reflects sequenced tumor bulks with high rates of non-malignant cells,

such as fibroblast, distorting the sequencing results towards a mesenchymal phenotype (Puram, Parikh, and Tirosh 2018; Puram et al. 2017). This view could be confirmed in the present thesis with hierarchical clustering of TCGA patients using non-malignant cell marker genes. Hereby, a subgroup of patients with high non-malignant cell marker gene expression was identified (Fig. 4.16). This subgroup expressing the highest relative levels of non-malignant cell markers solely consisted of sequenced tumors previously defined as the mesenchymal subtype (Lawrence et al. 2015). This supports the idea by Puram *et al.* that the molecular subtypes of HNSCC might need refinement, but also shows that bulk sequencing data can be resolved with established clustering methods. Since, the defined pEMT phenotype was functionally described for patients of the basal-like and mesenchymal subtype and to avoid distortion of the downstream analysis, analyzed basal-like and mesenchymal TCGA patients with the highest influence on gene expression by non-malignant cells were excluded (Fig. 4.16). By excluding patients and not deconvoluting the gene expression matrix, the cohort was reduced to n = 55 individuals but the order of gene expression levels was retained, which was highly important for the following quantification of the pEMT gene signature. Using the SING scoring technique described by Foroutan *et al.* as a robust single sample scoring method for molecular phenotypes, a score quantifying the relative expression of the common 15 pEMT genes in internal relation to the top 10,000 protein coding genes for each single patient was defined. This pEMT SING score was performing well in Cox models and allowed to significantly stratify the TCGA cohort, serving as a discovery data set (log-rank p-value = 0.0033), and the MDACC cohort, serving as a validation data set (log-rank p-value = 0.00048), into "low"-, "middle"-, and "high"-risk subgroups referring to risk of death (Figs. 4.17 and 4.17). For both cohorts, the "high"-risk group was showing the poorest survival rates with median overall survival of approximately 10 months. By testing n = 10,000 randomly picked gene sets of n = 15 genes from the top 10,000 protein coding genes, excluding the pEMT signature, an α-error rate below 5 % (α-error = 0.0418) was reported, showing that the described method has reliable information content (Fig. 4.17). Disease progression, such as lymph node metastasis, was described to be linked to pEMT (Parikh

et al. 2019; Puram et al. 2017). Clinical data of the lymph node status was available for the TCGA patients and a positive correlation with the pEMT SING score was found supporting the prognostic value of pEMT and the link to disease progression in patients (Kendall's $\tau = 0.30$, p-value = 0.005, Fig. 4.18).

Epithelial and mesenchymal marker gene expression from TCGA and MDACC patients was plotted along the pEMT SING score rank to demonstrate that pEMT rather than EMT or mesenchymal gene expression was quantified. Consistently across both cohorts, epithelial and mesenchymal genes were co-expressed in patients with a high SING score, represented by a high pEMT rank, showing that the tumors of these individuals are reflecting a pEMT state (Fig. 4.21). RAB25 and ECAD were listed as part of the top 5 epithelial marker genes in a generic EMT signature derived from bladder, gastric, ovarian, breast, lung, colorectral, and bladder cancer (Tan et al. 2014). According to the pEMT SING score, RAB25, out of five epithelial markers including ECAD, was the only marker gene substantially decreasing expression with increasing pEMT in HNSCC (Fig. 4.21).

Recently, a meta-analysis of the canonical EMT-TFs; TWIST1, TWIST2, SNAIL, SLUG, ZEB1, and ZEB2, in HNSCC cohorts reported OS prognostic values for TWIST1 (HR = 1.61), SNAIL (HR = 2.17), SLUG (HR = 1.90), and ZEB1 (HR = 2.70) (Wan et al. 2020). From the three EMT-TFs found within the top 10,000 protein coding genes in patients (SLUG, ZEB1, ZEB2), SLUG was showing the highest gene expression levels consistently across both cohorts and SLUG was positively correlating with ten out of the 15 common pEMT genes in the TCGA and MDACC cohort (Figs. 4.21 and 4.22). On the other hand, ZEB1 and ZEB2 were consistently correlating with only two genes from the signature, MMP2 and VIM (Fig. 4.22). This shows that on a single gene level, SLUG is representing the best candidate to mediate the pEMT in HNSCCs. This assumption was further reflected by the strong, highly significant and consistent correlation of SLUG with the pEMT SING score (TCGA: Spearman's $\rho = 0.52$, p-value = 0.000059 and MDACC: $\rho = 0.51$, p-value = 0.000027) and outperforming ZEB1 and ZEB2 (Fig. 4.23). Therefore, SLUG was chosen as the most promising candidate to investigate for associations with

pEMT effects *in vitro* and in cancer patients. In HNSCC, SNAIL and TWIST1 have been described as mediators of EMT and tumor progression (Li et al. 2019; Lee et al. 2012). Since, SNAIL and TWIST1 expression was not reported in the defined patient data, their impact on pEMT could not be assessed upon this doctoral thesis but cannot be formally excluded. ZEB1 expression correlated to mesenchymal marker VIM expression in HNSCC patients (Fig. 4.22) and SLUG was shown to be capable of inducing ZEB1 expression in HT-29 and MDCK cells (Guaita et al. 2002) and cell lines of the head and neck region (Fig. 4.25). Further, SLUG was reported to regulate expression and activity of ZEB1 in melanoma (Wels et al. 2011). Thus, higher levels of ZEB1 might mark more excessive SLUG-induced pEMT in cancer cells that could partly explain the higher OS HR compared to SLUG reported by Wan *et al.* (Wan et al. 2020). By overexpressing SLUG in FaDu cells, a cellular phenotype related to a minor ECAD decrease, VIM and ZEB1 increase, and no clear decrease in proliferation was reported (Figs. 4.24 and 4.25). This shows that SLUG expression is capable of inducing a pEMT phenotype in cancer cells. In Kyse30 cells, the SLUG-OE was inducing a phenotype, which represented a rather mesenchymal state, as ECAD levels were decreased extensively and the cellular morphology became spindle-shaped (Figs. 4.24 and 4.25). Nonetheless, in both cell lines of the head and neck region the overexpression of SLUG was elevating the invasive potential of the cells. SLUG-OE cells were showing a higher capability of invading a fibroblast spheroid and these findings were confirmed by a quantification *via* a matrigel invasion assay in both cell lines, FaDu and Kyse30 (Fig. 4.26). EMT was reported to protect cancer cells from irradiation (Stemmler et al. 2019). In both cancer cell lines, exogenous SLUG expression was sufficient to significantly increase their irradiation resistance to 2, 4, and 6 Gray in FaDu, and to 4 Gray in Kyse30 (Fig. 4.26 E-F). Kyse30 cells might naturally represent a more mesenchymal state with elevated invasive potential of Ctrl cells compared to respective FaDu cells (Fig. 4.26). This would explain the higher overall susceptibility of FaDu cells to irradiation compared to Kyse30 cells, as shown by the colony formation assay (Fig. 4.26 E-F). Taken together, SLUG overexpression was inducing invasion and irradiation resistance *in vitro*, characteristics

related to tumor recurrence and metastasis *in vivo*. In the present work, a link between pEMT and SLUG was established by transcriptome analysis of patient data and, now, *in vitro* experiments show an association with adverse clinical effects, which add to the understanding of reappearance and spread of primary tumors. To reveal the effects of SLUG at the protein level in patients, IHC cryosections of n = 76 HPV-negative HNSCC primary tumors were stained and quantified. The SLUG staining patterns of 77.6 % of the evaluated cryosections could be unambiguously categorized as either homogeneously distributed or as localized towards the edge of tumor sections (Tab. 3.3 and Fig. 4.27). SLUG expression and pEMT was reported to be found on leading edges of tumors in HNSCC (Parikh et al. 2019; Puram et al. 2017). Interestingly, LMU patients with a tumor recurrence or SLUG staining primarily localized to the leading tumor edges were showing higher SLUG levels in their primary tumors (Fig. 4.28 B-C). When comparing SLUG levels of patients of the subgroups homogeneous/ recurrence free *vs.* edge/ recurrence, patients suffering from recurrence with SLUG on tumor edges were showing two times higher mean levels of SLUG (Fig. 4.28 D).

The EMT-TF SLUG was shown to be induced by EGF/EGFR-mediated EMT and represents a direct target of EGFR-signaling (Pan et al. 2018; Chen et al. 2009). In a collaborative work with Min Pan, we demonstrated that EGFR-high/ EpCAM-low compared to EGFR-low/ EpCAM-high HNSCC patients show significantly increased levels of pERK and SLUG, and pERK-SLUG-high patients suffer from poorer OS and DFS (Appendix Fig. 8). SLUG expression and pEMT in cancer cells located to the tumor edges could be triggered by EGF/EGFR-signaling through paracrine secretion of EGF by *e.g.* cancer-associated macrophages (Gao et al. 2018; Rogers et al. 2005). This mechanism could lead to tumor cells on the primary tumor edges possessing an aggressive phenotype causing recurrence and spread to lymph nodes. In line with this hypothesis and corroborated by the *in vitro* findings of elevated invasion and irradiation resistance upon SLUG overexpression, SLUG levels within the same patient were increased in lymph nodes compared to primary tumors and lower to not expressed in normal mucosa (Fig. 4.29 A). Further, SLUG was showing prognostic value in a Cox model with disease-free

survival defining the clinical endpoint (log-rank p-value = 0.019, Fig. 4.29 B). These findings identifying SLUG as a promising candidate for pEMT induction in cancer are in agreement with a meta-analysis report defining SLUG as most impactful on the risk of metastatic breast cancer compared to TWIST1, SNAIL1, and ZEB1 (Imani et al. 2016). The fact that SLUG is not expressed in healthy tissue, at least not the mucosa of the head and neck region, might enable it as a therapeutic target.

The pEMT program quantified and connected to SLUG expression within this work was defined in a subgroup of HPV-negative HNSCC tumors of the basal-like or mesenchymal subtype, entitled by Puram *et al.* as malignant-basal, and predominantly consisted of oral cavity tumors (Fig. 4.16). SLUG protein expression was investigated in HPV-negative LMU patients of unknown molecular subtype. The LMU cohort consisted of 32.9 % oral cavity patients (Tab. 3.3) and therefore most likely did not comprise malignant-basal patients only. It would be interesting to know whether the SLUG-driven pEMT program represents a general process in HNSCC tumor progression.

Taken together, this work evaluated the effects of EMT on HNSCC patients and *via* a transcriptome analysis allowed to identify five genes, which could mediate the clinically adverse effects of enhanced EGFR-signaling. In the following, HNSCC tumor bulks with low influence of non-malignant cells were extracted, pEMT was succesfully quantified, the prognostic value of pEMT was demonstrated, and SLUG was identified as a mediator of pEMT. Generally, EMT is a cellular program with different triggers and it would be interesting to see whether the EGF/EGFR-mediated EMT represents the main signaling pathway in HNSCC. Additional measurements of EMT-inducing ligands, *e.g.* the concentration of soluble EGF and TGFβ, within the tumor microenvironment *in vivo* are necessary. Further, using tumor cell spheroids with and without EGF-high treatment followed by *e.g.* staining of cryosections for SLUG, epithelial-, and mesencyhmal markers and single cell RNASeq analysis might allow to assess the structural organization of *in vivo*-like tumors/ spheroids. Prior to those experiments, it seems conducive to use transcriptome data to define the position of the selected tumor cell lines within the EMT spectrum, as the SLUG-OE experiments were already pointing towards different

points in the EMT spectrum of FaDu and Kyse30 cells. Future research further needs to address how well cell culture conditions, especially 2D-cultures, model EMT relevant in tumor progression, as this has been questioned for TGF$\beta-$mediated EMT (Dongre and Weinberg 2019; McFaline-Figueroa et al. 2019).

The process of EMT *in vivo* most likely represents a spatial and temporal spectrum causing the reported high rates of heterogeneity of HNSCC tumors (McFaline-Figueroa et al. 2019; Dijk et al. 2018; Puram et al. 2017; Mroz et al. 2015) and complicating the transition from *in vitro* findings to cancer disease in patients. This work aimed to illuminate a certain axis of the spectrum and defined a certain cellular state of this fluent spectrum with succeeding prerequisites for cancer disease progression. Future cancer research needs to implement rising high-throughput technology and big data experiments to further entangle the EMT spectrum, consolidating solid grounds for innovative treatments and paving the way to a system of truly personalized cancer care.

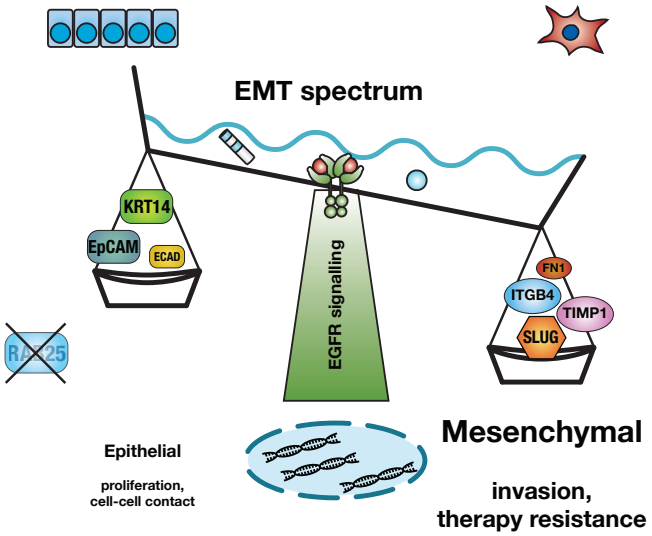

Figure 5.1: Illustration of the EMT spectrum influenced by EGFR-signaling. The EMT status of a particular cell is influenced by its mircoenvironment. EGFR activation by excessive EGF binding leads to induction of EMT. The resulting position in the EMT spectrum is influenced by regulation extend of different epithelial and mesenchymal genes. EGF/EGFR-mediated EMT leads to induction of ITGB4, TIMP1, NCEH1, DDIT4, and FADD. A subtype of (p)EMT in malignant-basal HNSCC is driven by expression of EMT-TF SLUG and mesenchymal marker FN1, and decrease of ECAD and RAB25. Other epithelial markers such as KRT14 or EpCAM are not decreased. (p)EMT: Partial Epithelial-to-Mesenchymal transition, EGF/R: Epidermal growth factor/ receptor, HNSCC: Head and neck squamous cell carcinoma, pEMT: Partial EMT, HNSCC: Head and neck squamous cell carcinoma, ECAD: E-CADHERIN, EpCAM: Epithelial cell adhesion molecule, KRT14: Cytokeratin 14, FN1: Fibronectin 1, RAB25: Ras-related protein RAB25, ITGB4: Integrin β 4, TIMP1: TIMP metallopeptidase inhibitor 1.

References

Affolter, Annette, Marie-France Muller, Katharina Sommer, Albrecht Stenzinger, Karim Zaoui, Katja Lorenz, Thomas Wolf, et al. 2016. "Targeting irradiation-induced mitogen-activated protein kinase activation in vitro and in an ex vivo model for human head and neck cancer." *Head & Neck* 38 Suppl 1 (April): E2049–61. https://doi.org/10.1002/hed.24376.

Aiello, Nicole M., Thomas Brabletz, Yibin Kang, M. Angela Nieto, Robert A. Weinberg, and Ben Z. Stanger. 2017. "Upholding a role for EMT in pancreatic cancer metastasis." *Nature* 547 (7661): E7–E8. https://doi.org/10.1038/nature22963.

Alessi, D R, S R James, C P Downes, A B Holmes, P R Gaffney, C B Reese, and P Cohen. 1997. "Characterization of a 3-phosphoinositide-dependent protein kinase which phosphorylates and activates protein kinase Balpha." *Current Biology : CB* 7 (4): 261–69. https://doi.org/10.1016/s0960-9822(06)00122-9.

Andl, Claudia D, Takaaki Mizushima, Kenji Oyama, Mark Bowser, Hiroshi Nakagawa, and Anil K Rustgi. 2004. "EGFR-induced cell migration is mediated predominantly by the JAK-STAT pathway in primary esophageal keratinocytes." *American Journal of Physiology. Gastrointestinal and Liver Physiology* 287 (6): G1227–37. https://doi.org/10.1152/ajpgi.00253.2004.

Baumeister, Philipp, Alessandra Hollmann, Julia Kitz, Artemis Afthonidou, Florian Simon, Julius Shakhtour, Brigitte Mack, et al. 2018. "High Expression of EpCAM and Sox2 is a Positive Prognosticator of Clinical Outcome for Head and Neck Carcinoma." *Scientific Reports* 8 (1): 14582. https://doi.org/10.1038/s41598-018-32178-8.

Blaj, Cristina, Eva Marina Schmidt, Sebastian Lamprecht, Heiko Hermeking, Andreas Jung, Thomas Kirchner, and David Horst. 2017. "Oncogenic effects of high MAPK

activity in colorectal cancer mark progenitor cells and persist irrespective of RAS mutations." *Cancer Research* 77 (7): 1763–74. https://doi.org/10.1158/0008-5472.CAN-16-2821.

Bolós, Victoria, Hector Peinado, Mirna A Pérez-Moreno, Mario F Fraga, Manel Esteller, and Amparo Cano. 2003. "The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors." *Journal of Cell Science* 116 (Pt 3): 499–511. https://doi.org/10.1242/jcs.00224.

Brabletz, T, A Jung, S Reu, M Porzner, F Hlubek, L A Kunz-Schughart, R Knuechel, and T Kirchner. 2001. "Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment." *Proceedings of the National Academy of Sciences of the United States of America* 98 (18): 10356–61. https://doi.org/10.1073/pnas.171610498.

Byeon, Hyung Kwon, Minhee Ku, and Jaemoon Yang. 2019. "Beyond EGFR inhibition: multilateral combat strategies to stop the progression of head and neck cancer." *Experimental and Molecular Medicine* 51 (1). https://doi.org/10.1038/s12276-018-0202-2.

Carpén, Timo, Timo Sorsa, Lauri Jouhi, Taina Tervahartiala, Caj Haglund, Stina Syrjänen, Jussi Tarkkanen, et al. 2019. "High levels of tissue inhibitor of metalloproteinase-1 (TIMP-1) in the serum are associated with poor prognosis in HPV-negative squamous cell oropharyngeal cancer." *Cancer Immunology, Immunotherapy : CII* 68 (8): 1263–72. https://doi.org/10.1007/s00262-019-02362-4.

Chaffer, Christine L, Nemanja D Marjanovic, Tony Lee, George Bell, Celina G Kleer, Ferenc Reinhardt, Ana C D'Alessio, Richard A Young, and Robert A Weinberg. 2013. "Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity." *Cell* 154 (1): 61–74. https://doi.org/10.1016/j.cell.2013.06.005.

Chaffer, Christine L, Beatriz P San Juan, Elgene Lim, and Robert A Weinberg. 2016. "EMT, cell plasticity and metastasis." *Cancer Metastasis Reviews* 35 (4): 645–54. https://doi.org/10.1007/s10555-016-9648-7.

Chen, Chao, Miriam Zimmermann, Inge Tinhofer, Andreas M. Kaufmann, and Andreas E. Albers. 2013. "Epithelial-to-mesenchymal transition and cancer stem(-like) cells in head and neck squamous cell carcinoma." *Cancer Letters* 338 (1): 47–56. https://doi.org/10.1016/j.canlet.2012.06.013.

Chen, Haoming, Genfeng Zhu, Yong Li, Ravi N Padia, Zheng Dong, Zhixing K Pan, Kebin Liu, and Shuang Huang. 2009. "Extracellular signal-regulated kinase signaling pathway regulates breast cancer cell migration by maintaining slug expression." *Cancer Research* 69 (24): 9228–35. https://doi.org/10.1158/0008-5472.CAN-09-1950.

Chen, Hao-Nien, Kang-Hao Liang, Jun-Kai Lai, Chun-Hsin Lan, Mei-Ying Liao, Shao-Hsi Hung, Yi-Ting Chuang, Kai-Chi Chen, William Wei-Fu Tsuei, and Han-Chung Wu. 2020. "EpCAM signaling promotes tumor progression and protein stability of PD-L1 through the EGFR pathway." *Cancer Research*, September. https://doi.org/10.1158/0008-5472.CAN-20-1264.

Chen, Z F, and R R Behringer. 1995. "Twist Is Required in Head Mesenchyme for Cranial Neural Tube Morphogenesis." *Genes & Development* 9 (6): 686–99. https://doi.org/10.1101/gad.9.6.686.

Chiang, Kyle P, Sherry Niessen, Alan Saghatelian, and Benjamin F Cravatt. 2006. "An enzyme that regulates ether lipid signaling pathways in cancer annotated by multidimensional profiling." *Chemistry & Biology* 13 (10): 1041–50. https://doi.org/10.1016/j.chembiol.2006.08.008.

Chinnaiyan, A M, C G Tepper, M F Seldin, K O'Rourke, F C Kischkel, S Hellbardt, P H Krammer, M E Peter, and V M Dixit. 1996. "FADD/MORT1 is a common mediator of CD95 (Fas/APO-1) and tumor necrosis factor receptor-induced apoptosis." *The*

Journal of Biological Chemistry 271 (9): 4961–5. https://doi.org/10.1074/jbc.271.9.4961.

Cho, Hyung Jun, Ami Yu, Sukwoo Kim, Jaewoo Kang, and Seung Mo Hong. 2009. "Robust likelihood-based survival modeling with microarray data." *Journal of Statistical Software* 29 (1). https://doi.org/10.18637/jss.v029.i01.

Chung, Christine H, Joel S Parker, Gamze Karaca, Junyuan Wu, William K Funkhouser, Dominic Moore, Dale Butterfoss, et al. 2004. "Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression." *Cancer Cell* 5 (5): 489–500. https://doi.org/10.1016/s1535-6108(04)00112-6.

Cohen, Roger B. 2014. "Current challenges and clinical investigations of epidermal growth factor receptor (EGFR)- and ErbB family-targeted agents in the treatment of head and neck squamous cell carcinoma (HNSCC)." *Cancer Treatment Reviews* 40 (4): 567–77. https://doi.org/10.1016/j.ctrv.2013.10.002.

Dassonville, O, J L Formento, M Francoual, A Ramaioli, J Santini, M Schneider, F Demard, and G Milano. 1993. "Expression of epidermal growth factor receptor and survival in upper aerodigestive tract cancer." *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 11 (10): 1873–8. https://doi.org/10.1200/JCO.1993.11.10.1873.

Derynck, Rik, Baby Periyanayaki Muthusamy, and Koy Y Saeteurn. 2014. "Signaling pathway cooperation in TGF-β-induced epithelial-mesenchymal transition." *Current Opinion in Cell Biology* 31 (December): 56–66. https://doi.org/10.1016/j.ceb.2014.09.001.

Diepenbruck, Maren, and Gerhard Christofori. 2016. "Epithelial-mesenchymal transition (EMT) and metastasis: Yes, no, maybe?" *Current Opinion in Cell Biology* 43: 7–13. https://doi.org/10.1016/j.ceb.2016.06.002.

Dijk, David van, Roshan Sharma, Juozas Nainys, Kristina Yim, Pooja Kathail, Ambrose

J. Carr, Cassandra Burdziak, et al. 2018. "Recovering Gene Interactions from Single-Cell Data Using Data Diffusion." *Cell* 174 (3): 716–729.e27. https://doi.org/10.1016/j.cell.2018.05.061.

Dong, Jing, Shi-Wen Jiang, Yanru Niu, Ling Chen, Shuyan Liu, Tianzhong Ma, Xiancai Chen, Liyan Xu, Zhongjing Su, and Haibin Chen. 2013. "Expression of estrogen receptor α and β in esophageal squamous cell carcinoma." *Oncology Reports* 30 (6): 2771–6. https://doi.org/10.3892/or.2013.2770.

Dongre, Anushka, and Robert A. Weinberg. 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.

Dyce, Orville H, Amy F Ziober, Randal S Weber, Kaoru Miyazaki, Samir S Khariwala, Michael Feldman, and Barry L Ziober. 2002. "Integrins in head and neck squamous cell carcinoma invasion." *The Laryngoscope* 112 (11): 2025–32. https://doi.org/10.1097/00005537-200211000-00021.

Ellisen, Leif W, Kate D Ramsayer, Cory M Johannessen, Annie Yang, Hideyuki Beppu, Karolina Minda, Jonathan D Oliner, Frank McKeon, and Daniel A Haber. 2002. "REDD1, a developmentally regulated transcriptional target of p63 and p53, links p63 to regulation of reactive oxygen species." *Molecular Cell* 10 (5): 995–1005. https://doi.org/10.1016/s1097-2765(02)00706-2.

Fan, Qi-Wen, Christine Cheng, Zachary A Knight, Daphne Haas-Kogan, David Stokoe, C David James, Frank McCormick, Kevan M Shokat, and William A Weiss. 2009. "EGFR signals to mTOR through PKC and independently of Akt in glioma." *Science Signaling* 2 (55): ra4. https://doi.org/10.1126/scisignal.2000014.

Ferlay, Jacques, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. 2015. "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012." *International Journal of Cancer* 136 (5): E359–86. https:

//doi.org/10.1002/ijc.29210.

Fischer, Kari R, Anna Durrans, Sharrell Lee, Jianting Sheng, Fuhai Li, Stephen T C Wong, Hyejin Choi, et al. 2015. "Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance." *Nature* 527 (7579): 472–76. https://doi.org/10.1038/nature15748.

Foroutan, Momeneh, Dharmesh D. Bhuva, Ruqian Lyu, Kristy Horan, Joseph Cursons, and Melissa J. Davis. 2018. "Single sample scoring of molecular phenotypes." *BMC Bioinformatics*. https://doi.org/10.1186/s12859-018-2435-4.

Franco, Hector L., José Casasnovas, José R. Rodríguez-Medina, and Carmen L. Cadilla. 2011. "Redundant or separate entities? - Roles of Twist1 and Twist2 as molecular switches during gene transcription." *Nucleic Acids Research* 39 (4): 1177–86. https://doi.org/10.1093/nar/gkq890.

Gao, Jiujiao, Qiu Yan, Jiao Wang, Shuai Liu, and Xuesong Yang. 2015. "Epithelial-to-Mesenchymal transition induced by TGF-β1 Is mediated by AP1-Dependent EpCAM expression in MCF-7 cells." *Journal of Cellular Physiology* 230 (4): 775–82. https://doi.org/10.1002/jcp.24802.

Gao, Lu, Wei Zhang, Wen-Qun Zhong, Zhuo-Jue Liu, Hui-Min Li, Zi-Li Yu, and Yi-Fang Zhao. 2018. "Tumor associated macrophages induce epithelial to mesenchymal transition via the EGFR/ERK1/2 pathway in head and neck squamous cell carcinoma." *Oncology Reports* 40 (5): 2558–72. https://doi.org/10.3892/or.2018.6657.

Gires, Olivier, Min Pan, Henrik Schinke, Martin Canis, and Patrick A Baeuerle. 2020. "Expression and function of epithelial cell adhesion molecule EpCAM: where are we after 40 years?" *Cancer Metastasis Reviews*, June. https://doi.org/10.1007/s10555-020-09898-3.

González-Moles, Miguel Ángel, Ángela Ayén, Isábel González-Ruiz, Teresa de Porras-Carrique, Lucía González-Ruiz, Isábel Ruiz-Ávila, and Pablo Ramos-García. 2020.

"Prognostic and Clinicopathological Significance of FADD Upregulation in Head and Neck Squamous Cell Carcinoma: A Systematic Review and Meta-Analysis." *Cancers* 12 (9). https://doi.org/10.3390/cancers12092393.

Greenburg, G, and E D Hay. 1982. "Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells." *The Journal of Cell Biology* 95 (1): 333–39. https://doi.org/10.1083/jcb.95.1.333.

Guaita, Sandra, Isabel Puig, Clara Francí, Marta Garrido, David Domínguez, Eduard Batlle, Elena Sancho, Shoukat Dedhar, Antonio García De Herreros, and Josep Baulida. 2002. "Snail induction of epithelial to mesenchymal transition in tumor cells is accompanied by MUC1 repression and ZEB1 expression." *Journal of Biological Chemistry* 277 (42): 39209–16. https://doi.org/10.1074/jbc.M206400200.

Guzmán, Camilo, Manish Bagga, Amanpreet Kaur, Jukka Westermarck, and Daniel Abankwa. 2014. "ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays." *PloS One* 9 (3): e92444. https://doi.org/10.1371/journal.pone.0092444.

Hay, E D. 1995. "An overview of epithelio-mesenchymal transformation." *Acta Anatomica* 154 (1): 8–20. https://doi.org/10.1159/000147748.

Heijden, Martijn van der, Paul B. M. Essers, Caroline V. M. Verhagen, Stefan M. Willems, Joyce Sanders, Reinout H. de Roest, David M. Vossen, et al. 2020. "Epithelial-to-mesenchymal transition is a prognostic marker for patient outcome in advanced stage HNSCC patients treated with chemoradiotherapy." *Radiotherapy and Oncology* 147: 186–94. https://doi.org/10.1016/j.radonc.2020.05.013.

Herlyn, M, Z Steplewski, D Herlyn, and H Koprowski. 1979. "Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies." *Proceedings of the National Academy of Sciences of the United States of America* 76 (3): 1438–42. https://doi.org/10.1073/pnas.76.3.1438.

Hess, Julia, Kristian Unger, Cornelius Maihoefer, Lars Schuttrumpf, Ludmila Wintergerst, Theresa Heider, Peter Weber, et al. 2019. "A five-microRNA signature predicts survival and disease control of patients with head and neck cancer negative for HPV infection." *Clinical Cancer Research* 25 (5): 1505–16. https://doi.org/10.1158/1078-0432.CCR-18-0776.

Horn, Dominik, Jochen Hess, Kolja Freier, Jürgen Hoffmann, and Christian Freudlsperger. 2015. "Targeting EGFR-PI3K-AKT-mTOR signaling enhances radiosensitivity in head and neck squamous cell carcinoma." *Expert Opinion on Therapeutic Targets* 19 (6): 795–805. https://doi.org/10.1517/14728222.2015.1012157.

Hoshino, R, Y Chatani, T Yamori, T Tsuruo, H Oka, O Yoshida, Y Shimada, et al. 1999. "Constitutive activation of the 41-/43-kDa mitogen-activated protein kinase signaling pathway in human tumors." *Oncogene* 18 (3): 813–22. https://doi.org/10.1038/sj.onc.1202367.

Hsu, Ya-Ting, Pawel Osmulski, Yao Wang, Yi-Wen Huang, Lu Liu, Jianhua Ruan, Victor X Jin, Nameer B Kirma, Maria E Gaczynska, and Tim Hui-Ming Huang. 2016. "EpCAM-Regulated Transcription Exerts Influences on Nanomechanical Properties of Endometrial Cancer Cells That Promote Epithelial-to-Mesenchymal Transition." *Cancer Research* 76 (21): 6171–82. https://doi.org/10.1158/0008-5472.CAN-16-0752.

Ichikawa, Kenji, Yuji Kubota, Takanori Nakamura, Jane S. Weng, Taichiro Tomida, Haruo Saito, and Mutsuhiro Takekawa. 2015. "MCRIP1, an ERK Substrate, Mediates ERK-Induced Gene Silencing during Epithelial-Mesenchymal Transition by Regulating the Co-Repressor CtBP." *Molecular Cell* 58 (1): 35–46. https://doi.org/10.1016/j.molcel.2015.01.023.

Imani, Saber, Hossein Hosseinifard, Jingliang Cheng, Chunli Wei, and Junjiang Fu. 2016. "Prognostic Value of EMT-inducing Transcription Factors (EMT-TFs) in Metastatic Breast Cancer: A Systematic Review and Meta-analysis." *Scientific Reports* 6 (Jan-

uary): 1–10. https://doi.org/10.1038/srep28587.

Kalluri, Raghu, and Robert A Weinberg. 2009. "The basics of epithelial-mesenchymal transition." *The Journal of Clinical Investigation* 119 (6): 1420–8. https://doi.org/10.1172/JCI39104.

Kalyankrishna, Shailaja, and Jennifer R Grandis. 2006. "Epidermal Growth Factor Receptor Biology in Head and Neck Cancer." *Journal of Clinical Oncology* 24 (17): 2666–72. https://doi.org/10.1200/JCO.2005.04.8306.

Kimura, Hitoshi, Hiroyuki Kato, Ahmad Faried, Makoto Sohda, Masanobu Nakajima, Yasuyuki Fukai, Tatsuya Miyazaki, Norihiro Masuda, Minoru Fukuchi, and Hiroyuki Kuwano. 2007. "Prognostic significance of EpCAM expression in human esophageal cancer." *International Journal of Oncology* 30 (1): 171–79.

Koyama, Noriko, Masanori Kashimata, Hideaki Sakashita, Hiroshi Sakagami, and Edward W Gresik. 2003. "EGF-stimulated signaling by means of PI3K, PLCgamma1, and PKC isozymes regulates branching morphogenesis of the fetal mouse submandibular gland." *Developmental Dynamics : An Official Publication of the American Association of Anatomists* 227 (2): 216–26. https://doi.org/10.1002/dvdy.10309.

Krishnaswamy, Smita, Nevena Zivanovic, Roshan Sharma, Dana Pe'Er, and Bernd Bodenmiller. 2018. *Learning time-varying information flow from single-cell epithelial to mesenchymal transition data.* Vol. 13. 10. https://doi.org/10.1371/journal.pone.0203389.

Kröger, Cornelia, Alexander Afeyan, Jasmin Mraz, Elinor Ng Eaton, Ferenc Reinhardt, Yevgenia L. Khodor, Prathapan Thiru, et al. 2019. "Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells." *Proceedings of the National Academy of Sciences of the United States of America* 116 (15): 7353–62. https://doi.org/10.1073/pnas.1812876116.

Kuan, I. I., Chi Chiu Lee, Chien Hsu Chen, Jean Lu, Yuan Sung Kuo, and Han Chung

Wu. 2019. "The extracellular domain of epithelial cell adhesion molecule (EpCAM) enhances multipotency of mesenchymal stem cells through EGFR-LIN28-LET7 signaling." *Journal of Biological Chemistry* 294 (19): 7769–86. https://doi.org/10.1074/jbc.RA119.007386.

Kurokawa, Akira, Masaki Nagata, Nobutaka Kitamura, Arhab A Noman, Makoto Ohnishi, Tokio Ohyama, Takanori Kobayashi, Susumu Shingaki, and Ritsuo Takagi. 2008. "Diagnostic value of integrin alpha3, beta4, and beta5 gene expression levels for the clinical outcome of tongue squamous cell carcinoma." *Cancer* 112 (6): 1272–81. https://doi.org/10.1002/cncr.23295.

Lambert, Arthur W, Diwakar R Pattabiraman, and Robert A Weinberg. 2017. "Emerging Biological Principles of Metastasis." *Cell* 168 (4): 670–91. https://doi.org/10.1016/j.cell.2016.11.037.

Lamouille, Samy, Jian Xu, and Rik Derynck. 2014. "Molecular mechanisms of epithelial-mesenchymal transition." *Nature Reviews. Molecular Cell Biology* 15 (3): 178–96. https://doi.org/10.1038/nrm3758.

Lawrence, Michael S., Carrie Sougnez, Lee Lichtenstein, Kristian Cibulskis, Eric Lander, Stacey B. Gabriel, Gad Getz, et al. 2015. "Comprehensive genomic characterization of head and neck squamous cell carcinomas." *Nature* 517 (7536): 576–82. https://doi.org/10.1038/nature14129.

Lee, Keun Woo, Jeong Hoon Kim, Songying Han, Chang Ohk Sung, In Gu Do, Young Hyeh Ko, Sung Hee Um, and Seok Hyung Kim. 2012. "Twist1 is an independent prognostic factor of esophageal squamous cell carcinoma and associated with its epithelial-mesenchymal transition." *Annals of Surgical Oncology* 19 (1): 326–35. https://doi.org/10.1245/s10434-011-1867-0.

Li, Ching Fei, Jia Yang Chen, Yang Hui Ho, Wen Hao Hsu, Liang Chun Wu, Hsin Yi Lan, Dennis Shin Shian Hsu, Shyh Kuan Tai, Ying Chih Chang, and Muh Hwa Yang. 2019. "Snail-induced claudin-11 prompts collective migration for tumour progression." *Na-*

ture Cell Biology 21 (2): 251–62. https://doi.org/10.1038/s41556-018-0268-z.

Liang, Kang-Hao Hao, Hsien-Cheng Cheng Tso, Shao-Hsi Hsi Hung, I-I I. Kuan, Jun-Kai Kai Lai, Feng-Yi Yi Ke, Yi-Ting Ting Chuang, et al. 2018. "Extracellular domain of EpCAM enhances tumor progression through EGFR signaling in colon cancer cells." *Cancer Letters* 433 (March): 165–75. https://doi.org/10.1016/j.canlet.2018.06.040.

Liang, Yuan, Tiehua Zhang, and Jie Zhang. 2020. "Natural tyrosine kinase inhibitors acting on the epidermal growth factor receptor: Their relevance for cancer therapy." *Pharmacological Research* 161 (August): 105164. https://doi.org/10.1016/j.phrs.2020.105164.

Lin, Yiwei, Yadi Wu, Junlin Li, Chenfang Dong, Xiaofeng Ye, Young-In Chi, B Mark Evers, and Binhua P Zhou. 2010. "The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1." *The EMBO Journal* 29 (11): 1803–16. https://doi.org/10.1038/emboj.2010.63.

Litvinov, S V, H A Bakker, M M Gourevitch, M P Velders, and S O Warnaar. 1994. "Evidence for a role of the epithelial glycoprotein 40 (Ep-CAM) in epithelial cell-cell adhesion." *Cell Adhesion and Communication* 2 (5): 417–28. https://doi.org/10.3109/15419069409004452.

Liu, Shuli, Lei Shi, Yang Wang, Dongxia Ye, Houyu Ju, Hailong Ma, Wenyi Yang, et al. 2018. "Stabilization of Slug by NF-κB is Essential for TNF-α -Induced Migration and Epithelial-Mesenchymal Transition in Head and Neck Squamous Cell Carcinoma Cells." *Cellular Physiology and Biochemistry* 47 (2): 567–78. https://doi.org/10.1159/000489990.

Livak, K J, and T D Schmittgen. 2001. "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." *Methods (San Diego, Calif.)* 25 (4): 402–8. https://doi.org/10.1006/meth.2001.1262.

Lu, Zhimin, Sourav Ghosh, Zhiyong Wang, and Tony Hunter. 2003. "Downregulation of caveolin-1 function by EGF leads to the loss of E-cadherin, increased transcriptional activity of β-catenin, and enhanced tumor cell invasion." *Cancer Cell* 4 (6): 499–515. https://doi.org/10.1016/S1535-6108(03)00304-0.

Mack, Brigitte, and Olivier Gires. 2008. "CD44s and CD44v6 expression in head and neck epithelia." *PLoS ONE* 3 (10): e3360. https://doi.org/10.1371/journal.pone.0003360.

Maetzel, Dorothea, Sabine Denzel, Brigitte Mack, Martin Canis, Philip Went, Michael Benk, Cuong Kieu, et al. 2009. "Nuclear signalling by tumour-associated antigen EpCAM." *Nature Cell Biology* 11 (2): 162–71. https://doi.org/10.1038/ncb1824.

Marur, Shanthi, Gypsyamber D'Souza, William H Westra, and Arlene A Forastiere. 2010. "HPV-associated head and neck cancer: a virus-related cancer epidemic." *The Lancet. Oncology* 11 (8): 781–89. https://doi.org/10.1016/S1470-2045(10)70017-6.

Massoner, P, T Thomm, B Mack, G Untergasser, A Martowicz, K Bobowski, H Klocker, O Gires, and M Puhr. 2014. "EpCAM is overexpressed in local and metastatic prostate cancer, suppressed by chemotherapy and modulated by MET-associated miRNA-200c/205." *British Journal of Cancer* 111 (5): 955–64. https://doi.org/10.1038/bjc.2014.366.

McFaline-Figueroa, José L, Andrew J Hill, Xiaojie Qiu, Dana Jackson, Jay Shendure, and Cole Trapnell. 2019. "A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition." *Nature Genetics* 51 (9): 1389–98. https://doi.org/10.1038/s41588-019-0489-5.

McKenzie, S J. 1991. "Diagnostic utility of oncogenes and their products in human cancer." *Biochimica et Biophysica Acta* 1072 (2-3): 193–214. https://doi.org/10.1016/0304-419x(91)90014-c.

Mering, Christian von, Lars J. Jensen, Berend Snel, Sean D. Hooper, Markus Krupp, Mathilde Foglierini, Nelly Jouffre, Martijn A. Huynen, and Peer Bork. 2005. "STRING: Known and predicted protein-protein associations, integrated and transferred across organisms." *Nucleic Acids Research* 33 (DATABASE ISS.): 433–37. https://doi.org/10.1093/nar/gki005.

Micalizzi, Douglas S, Susan M Farabaugh, and Heide L Ford. 2010. "Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression." *Journal of Mammary Gland Biology and Neoplasia* 15 (2): 117–34. https://doi.org/10.1007/s10911-010-9178-9.

Molina-Ortiz, Patricia, Ana Villarejo, Matthew MacPherson, Vanesa Santos, Amalia Montes, Serhiy Souchelnytskyi, Francisco Portillo, and Amparo Cano. 2012. "Characterization of the SNAG and SLUG domains of Snail2 in the repression of E-cadherin and EMT induction: modulation by serine 4 phosphorylation." *PloS One* 7 (5): e36132. https://doi.org/10.1371/journal.pone.0036132.

Mroz, Edmund A, Aaron D Tward, Rebecca J Hammon, Yin Ren, and James W Rocco. 2015. "Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas." *PLoS Medicine* 12 (2): e1001786. https://doi.org/10.1371/journal.pmed.1001786.

Murphy, Leon O, Jeffrey P MacKeigan, and John Blenis. 2004. "A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration." *Molecular and Cellular Biology* 24 (1): 144–53. https://doi.org/10.1128/mcb.24.1.144-153.2004.

Münz, Markus, Cuong Kieu, Brigitte Mack, Bärbel Schmitt, Reinhard Zeidler, and Olivier Gires. 2004. "The carcinoma-associated antigen EpCAM upregulates c-myc and induces cell proliferation." *Oncogene* 23 (34): 5748–58. https://doi.org/10.1038/sj.onc.1207610.

Nieto, M Angela. 2009. "Epithelial-Mesenchymal Transitions in development and disease:

old views and new perspectives." *The International Journal of Developmental Biology* 53 (8-10): 1541–7. https://doi.org/10.1387/ijdb.072410mn.

Nieto, M. Angela. 2002. "The snail superfamily of zinc-finger transcription factors." *Nature Reviews Molecular Cell Biology* 3 (3): 155–66. https://doi.org/10.1038/nrm757.

Nieto, M. Angela Angela, Ruby Yun Yun-Ju Y. J. J Huang, Rebecca A A. Jackson, and Jean Paul Paul Thiery. 2016. "Emt: 2016." *Cell* 166 (1): 21–45. https://doi.org/10.1016/j.cell.2016.06.028.

Normanno, Nicola, Antonella De Luca, Caterina Bianco, Luigi Strizzi, Mario Mancino, Monica R Maiello, Adele Carotenuto, Gianfranco De Feo, Francesco Caponigro, and David S Salomon. 2006. "Epidermal growth factor receptor (EGFR) signaling in cancer." *Gene* 366 (1): 2–16. https://doi.org/https://doi.org/10.1016/j.gene.2005.10.018.

Osta, Walid A, Yian Chen, Kaidi Mikhitarian, Michael Mitas, Mohamed Salem, Yusuf A Hannun, David J Cole, and William E Gillanders. 2004. "EpCAM is overexpressed in breast cancer and is a potential target for breast cancer gene therapy." *Cancer Research* 64 (16): 5818–24. https://doi.org/10.1158/0008-5472.CAN-04-0754.

Pan, Min, Henrik Schinke, Elke Luxenburger, Gisela Kranz, Julius Shakhtour, Darko Libl, Yuanchi Huang, et al. 2018. "EpCAM ectodomain EpEX is a ligand of EGFR that counteracts EGF-mediated epithelial-mesenchymal transition through modulation of phospho-ERK1/2 in head and neck cancer2s." *PLoS Biology* 16 (9): 1–36. https://doi.org/10.1371/journal.pbio.2006624.

Parikh, Anuraag S., Sidharth V. Puram, William C. Faquin, Jeremy D. Richmon, Kevin S. Emerick, Daniel G. Deschler, Mark A. Varvares, Itay Tirosh, Bradley E. Bernstein, and Derrick T. Lin. 2019. "Immunohistochemical quantification of partial-EMT in oral cavity squamous cell carcinoma primary tumors is associated with nodal metastasis." *Oral Oncology* 99 (September): 104458. https://doi.org/10.1016/j.

oraloncology.2019.104458.

Pastushenko, Ievgenia, Audrey Brisebarre, Alejandro Sifrim, Marco Fioramonti, Tatiana Revenco, Soufiane Boumahdi, Alexandra Van Keymeulen, et al. 2018. "Identification of the tumour transition states occurring during EMT." *Nature* 556 (7702). https://doi.org/10.1038/s41586-018-0040-3.

Peinado, Héctor, David Olmeda, and Amparo Cano. 2007. "Snail, ZEB and bHLH factors in tumour progression: An alliance against the epithelial phenotype?" *Nature Reviews Cancer* 7 (6): 415–28. https://doi.org/10.1038/nrc2131.

Pinto, Joseph A, Christian Rolfo, Luis E Raez, Alexandra Prado, Jhajaira M Araujo, Leny Bravo, Williams Fajardo, et al. 2017. "In silico evaluation of DNA Damage Inducible Transcript 4 gene (DDIT4) as prognostic biomarker in several malignancies." *Scientific Reports* 7 (1): 1526. https://doi.org/10.1038/s41598-017-01207-3.

Puram, Sidharth V., Anuraag S. Parikh, and Itay Tirosh. 2018. "Single cell RNA-seq highlights a role for a partial EMT in head and neck cancer." *Molecular and Cellular Oncology* 5 (3): 1–3. https://doi.org/10.1080/23723556.2018.1448244.

Puram, Sidharth V., Itay Tirosh, Anuraag S. Parikh, Anoop P. Patel, Keren Yizhak, Shawn Gillespie, Christopher Rodman, et al. 2017. "Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer." *Cell* 171 (7): 1611–1624.e24. https://doi.org/10.1016/j.cell.2017.10.044.

Rampias, T, A Giagini, S Siolos, H Matsuzaki, C Sasaki, A Scorilas, and A Psyrri. 2014. "RAS/PI3K Crosstalk and Cetuximab Resistance in Head and Neck Squamous Cell Carcinoma." *Clinical Cancer Research* 20 (11): 2933 LP–2946. https://doi.org/10.1158/1078-0432.CCR-13-2721.

Rogers, Susanne J., Kevin J. Harrington, Peter Rhys-Evans, Pornchai O-Charoenrat, and Suzanne A. Eccles. 2005. "Biological significance of c-erbB family oncogenes in head and neck cancer." *Cancer and Metastasis Reviews* 24 (1): 47–69. https:

//doi.org/10.1007/s10555-005-5047-1.

Rong, Chao, Marie F Muller, Fang Xiang, Alexandra Jensen, Wilko Weichert, Gerald Major, Peter K Plinkert, Jochen Hess, and Annette Affolter. 2020. "Adaptive ERK signalling activation in response to therapy and in silico prognostic evaluation of EGFR-MAPK in HNSCC." *British Journal of Cancer* 123 (2): 288–97. https://doi.org/10.1038/s41416-020-0892-9.

Saitoh, Masao. 2018. "JB special review-cellular plasticity in epithelial homeostasis and diseases: Involvement of partial EMT in cancer progression." *Journal of Biochemistry* 164 (4): 257–64. https://doi.org/10.1093/jb/mvy047.

Santamaria, Patricia G., Gema Moreno-Bueno, Francisco Portillo, and Amparo Cano. 2017. "EMT: Present and future in clinical oncology." *Molecular Oncology* 11 (7): 718–38. https://doi.org/10.1002/1878-0261.12091.

Santini, J, J L Formento, M Francoual, G Milano, M Schneider, O Dassonville, and F Demard. 1991. "Characterization, quantification, and potential clinical value of the epidermal growth factor receptor in head and neck squamous cell carcinomas." *Head & Neck* 13 (2): 132–39. https://doi.org/10.1002/hed.2880130209.

Savas, Jeffrey N, Joris De Wit, Davide Comoletti, Roland Zemla, Anirvan Ghosh, and John R 3rd Yates. 2014. "Ecto-Fc MS identifies ligand-receptor interactions through extracellular domain Fc fusion protein baits and shotgun proteomic analysis." *Nature Protocols* 9 (9): 2061–74. https://doi.org/10.1038/nprot.2014.140.

Schinke, Henrik, Min Pan, Merve Akyol, Jiefu Zhou, Gisela Kranz, Darko Libl, Florian Simon, et al. 2020. "Partial epithelial-to-mesenchymal transition is prognostic and associates with Slug in head and neck cancer." *bioRxiv*. https://doi.org/10.1101/2020.10.20.346692.

Schubert, Michael, Bertram Klinger, Martina Klünemann, Anja Sieber, Florian Uhlitz, Sascha Sauer, Mathew J. Garnett, Nils Blüthgen, and Julio Saez-Rodriguez. 2018.

"Perturbation-response genes reveal signaling footprints in cancer gene expression." *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-017-02391-6.

Seeber, Andreas, Gerold Untergasser, Gilbert Spizzo, Luigi Terracciano, Alessandro Lugli, Armin Kasal, Florian Kocher, et al. 2016. "Predominant expression of truncated EpCAM is associated with a more aggressive phenotype and predicts poor overall survival in colorectal cancer." *International Journal of Cancer* 139 (3): 657–63. https://doi.org/10.1002/ijc.30099.

Shaw, Leslie M. 2005. "Tumor cell invasion assays." *Methods in Molecular Biology (Clifton, N.J.)* 294: 97–105. https://doi.org/10.1385/1-59259-860-9:097.

Shibue, Tsukasa, and Robert A Weinberg. 2017. "EMT, CSCs, and drug resistance: the mechanistic link and clinical implications." *Nature Reviews. Clinical Oncology* 14 (10): 611–29. https://doi.org/10.1038/nrclinonc.2017.44.

Shin, Sejeong, Christopher A Dimitri, Sang-Oh Yoon, William Dowdle, and John Blenis. 2010. "ERK2 but not ERK1 induces epithelial-to-mesenchymal transformation via DEF motif-dependent signaling events." *Molecular Cell* 38 (1): 114–27. https://doi.org/10.1016/j.molcel.2010.02.020.

Shirakihara, Takuya, Kana Horiguchi, Keiji Miyazawa, Shogo Ehata, Tatsuhiro Shibata, Ikuo Morita, Kohei Miyazono, and Masao Saitoh. 2011. "TGF-β regulates isoform switching of FGF receptors and epithelial-mesenchymal transition." *The EMBO Journal* 30 (4): 783–95. https://doi.org/10.1038/emboj.2010.351.

Siegel, Rebecca L, Kimberly D Miller, and Ahmedin Jemal. 2016. "Cancer statistics, 2016." *CA: A Cancer Journal for Clinicians* 66 (1): 7–30. https://doi.org/10.3322/caac.21332.

Spizzo, Gilbert, Philip Went, Stephan Dirnhofer, Peter Obrist, Ronald Simon, Hanspeter Spichtin, Robert Maurer, et al. 2004. "High Ep-CAM expression is associated with poor prognosis in node-positive breast cancer." *Breast Cancer Research and Treatment*

86 (3): 207–13. https://doi.org/10.1023/B:BREA.0000036787.59816.01.

Stemmler, Marc P, Rebecca L Eccles, Simone Brabletz, and Thomas Brabletz. 2019. "Non-redundant functions of EMT transcription factors." *Nature Cell Biology* 21 (1): 102–12. https://doi.org/10.1038/s41556-018-0196-y.

Stransky, Nicolas, Ann Marie Egloff, Aaron D. Tward, Aleksandar D. Kostic, Kristian Cibulskis, Andrey Sivachenko, Gregory V. Kryukov, et al. 2011. "The mutational landscape of head and neck squamous cell carcinoma." *Science* 333 (6046): 1157–60. https://doi.org/10.1126/science.1208130.

Szklarczyk, Damian, Annika L Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, et al. 2019. "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets." *Nucleic Acids Research* 47 (D1): D607–D613. https://doi.org/10.1093/nar/gky1131.

Talukdar, Jayasree, Kangkana Kataki, Eyashin Ali, Bikash Narayan Choudhury, Munindra Narayan Baruah, Mallika Bhattacharyya, Sahana Bhattacharjee, and Subhash Medhi. 2020. "Altered expression of TGF-β1 and TGF-βR2 in tissue samples compared to blood is associated with food habits and survival in esophageal squamous cell carcinoma." *Current Problems in Cancer*, July, 100617. https://doi.org/10.1016/j.currproblcancer.2020.100617.

Tam, Wai Leong, and Robert A Weinberg. 2013. "The epigenetics of epithelial-mesenchymal plasticity in cancer." *Nature Medicine* 19 (11): 1438–49. https://doi.org/10.1038/nm.3336.

Tan, Tuan Zea, Qing Hao Miow, Yoshio Miki, Tetsuo Noda, Seiichi Mori, Ruby Yun-Ju Huang, and Jean Paul Thiery. 2014. "Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients." *EMBO Molecular Medicine* 6 (10): 1279–93. https://doi.org/10.15252/emmm.201404208.

Tashiro, Etsu, Shizuka Henmi, Hiroyuki Odake, Seitaro Ino, and Masaya Imoto. 2016. "Involvement of the MEK/ERK pathway in EGF-induced E-cadherin down-regulation." *Biochemical and Biophysical Research Communications* 477 (4): 801–6. https://doi.org/10.1016/j.bbrc.2016.06.138.

Thiery, Jean Paul, Hervé Acloque, Ruby Y J Huang, and M Angela Nieto. 2009. "Epithelial-mesenchymal transitions in development and disease." *Cell* 139 (5): 871–90. https://doi.org/10.1016/j.cell.2009.11.007.

Tian, Ya-Chung, Yung-Chang Chen, Chiz-Tzung Chang, Cheng-Chieh Hung, Mai-Szu Wu, Aled Phillips, and Chih-Wei Yang. 2007. "Epidermal growth factor and transforming growth factor-beta1 enhance HK-2 cell migration through a synergistic increase of matrix metalloproteinase and sustained activation of ERK signaling pathway." *Experimental Cell Research* 313 (11): 2367–77. https://doi.org/10.1016/j.yexcr.2007.03.022.

Tripathi, Manish K, Smita Misra, Sheetal V Khedkar, Nalo Hamilton, Charletha Irvin-Wilson, Chakradhari Sharan, Linda Sealy, and Gautam Chaudhuri. 2005. "Regulation of BRCA2 gene expression by the SLUG repressor protein in human breast cells." *The Journal of Biological Chemistry* 280 (17): 17163–71. https://doi.org/10.1074/jbc.M501375200.

Tsai, Jeff H., Joana Liu Donaher, Danielle A. Murphy, Sandra Chau, and Jing Yang. 2012. "Spatiotemporal Regulation of Epithelial-Mesenchymal Transition Is Essential for Squamous Cell Carcinoma Metastasis." *Cancer Cell* 22 (6): 725–36. https://doi.org/10.1016/j.ccr.2012.09.022.

Tsaktanis, Thanos, Heidi Kremling, Miha Pavšič, Ricarda Von Stackelberg, Brigitte Mack, Akio Fukumori, Harald Steiner, et al. 2015. "Cleavage and cell adhesion properties of human epithelial cell adhesion molecule (HEPCAM)." *Journal of Biological Chemistry* 290 (40): 24574–91. https://doi.org/10.1074/jbc.M115.662700.

Uttamsingh, S., X. Bao, K. T. Nguyen, M. Bhanot, J. Gong, J. L. K. Chan, F. Liu,

T. T. Chu, and L. H. Wang. 2008. "Synergistic effect between EGF and TGF-β1 in inducing oncogenic properties of intestinal epithelial cells." *Oncogene* 27 (18): 2626–34. https://doi.org/10.1038/sj.onc.1210915.

Vermorken, Jan B, José Trigo, Ricardo Hitt, Piotr Koralewski, Eduardo Diaz-Rubio, Frédéric Rolland, Rainald Knecht, Nadia Amellal, Armin Schueler, and José Baselga. 2007. "Open-Label, Uncontrolled, Multicenter Phase II Study to Evaluate the Efficacy and Toxicity of Cetuximab As a Single Agent in Patients With Recurrent and/or Metastatic Squamous Cell Carcinoma of the Head and Neck Who Failed to Respond to Platinum-Based The." *Journal of Clinical Oncology* 25 (16): 2171–7. https://doi.org/10.1200/JCO.2006.06.7447.

Walter, Vonn, Xiaoying Yin, Matthew D Wilkerson, Christopher R Cabanski, Ni Zhao, Ying Du, Mei Kim Ang, et al. 2013. "Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes." *PloS One* 8 (2): e56823. https://doi.org/10.1371/journal.pone.0056823.

Wan, Yuehan, Haichao Liu, Ming Zhang, Zhengxian Huang, Han Zhou, Yue Zhu, Yifan Tao, et al. 2020. "Prognostic value of epithelial-mesenchymal transition-inducing transcription factors in head and neck squamous cell carcinoma: A meta-analysis." *Head and Neck* 42 (5): 1067–76. https://doi.org/10.1002/hed.26104.

Wang, Meng He, Rui Sun, Xiao Min Zhou, Mei Yin Zhang, Jia Bin Lu, Yang Yang, Li Si Zeng, et al. 2018. "Epithelial cell adhesion molecule overexpression regulates epithelial-mesenchymal transition, stemness and metastasis of nasopharyngeal carcinoma cells via the PTEN/AKT/mTOR pathway." *Cell Death and Disease* 9 (1). https://doi.org/10.1038/s41419-017-0013-8.

Wels, Christian, Shripad Joshi, Petra Koefinger, Helmut Bergler, and Helmut Schaider. 2011. "Transcriptional activation of ZEB1 by slug leads to cooperative regulation of the epithelial-mesenchymal transition-like phenotype in melanoma." *Journal of Investigative Dermatology* 131 (9): 1877–85. https://doi.org/10.1038/jid.2011.

142.

Went, P, M Vasei, L Bubendorf, L Terracciano, L Tornillo, U Riede, J Kononen, R Simon, G Sauter, and P A Baeuerle. 2006. "Frequent high-level expression of the immunotherapeutic target Ep-CAM in colon, stomach, prostate and lung cancers." *British Journal of Cancer* 94 (1): 128–35. https://doi.org/10.1038/sj.bjc.6602924.

Xu, Jian, Samy Lamouille, and Rik Derynck. 2009. "TGF-beta-induced epithelial to mesenchymal transition." *Cell Research* 19 (2): 156–72. https://doi.org/10.1038/cr.2009.5.

Yang, Jing, and Robert A Weinberg. 2008. "Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis." *Developmental Cell* 14 (6): 818–29. https://doi.org/10.1016/j.devcel.2008.05.009.

Ye, Xin, Thomas Brabletz, Yibin Kang, Gregory D. Longmore, M. Angela Nieto, Ben Z. Stanger, Jing Yang, and Robert A. Weinberg. 2017. "Upholding a role for EMT in breast cancer metastasis." *Nature* 547 (7661): E1–E6. https://doi.org/10.1038/nature22816.

Zhang, Jiali, Qian Cheng, Yi Zhou, Yu Wang, and Xinming Chen. 2013. "Slug is a key mediator of hypoxia induced cadherin switch in HNSCC: Correlations with poor prognosis." *Oral Oncology* 49 (11): 1043–50. https://doi.org/10.1016/j.oraloncology.2013.08.003.

Zhang, Jun, Nabil F Saba, Georgia Zhuo Chen, and Dong M Shin. 2015. "Targeting HER (ERBB) signaling in head and neck cancer: An essential update." *Molecular Aspects of Medicine* 45 (November): 74–86. https://doi.org/10.1016/j.mam.2015.07.001.

Zheng, Xiaofeng, Julienne L Carstens, Jiha Kim, Matthew Scheible, Judith Kaye, Hikaru Sugimoto, Chia-Chin Wu, Valerie S LeBleu, and Raghu Kalluri. 2015. "Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance

in pancreatic cancer." *Nature* 527 (7579): 525–30. https://doi.org/10.1038/nature16064.

Tables

Table 1: Candidate DOWN genes from EGF/EGFR-mediated EMT genes found in cell lines by univariate Cox models. Direction of DE is indicated. DOWN means log2 fold change in cell lines $<$ -0.5. For DOWN genes only genes with HR $<$ 1 in TCGA patients were kept. Cox model estimated coefficients with SE, t-statistic, and p-value are shown. p-value: .1 - 0.05, .; $\leq$ 0.05, *; $\leq$ 0.01, **; $\leq$ 0.001, ***. EGF/R: Epidermal growth factor receptor, EMT: Epithelial-to-Mesenchymal transition, DE: Differential expression, HR: Hazard ratio, SE: Standard error, TCGA: The Cancer Genome Atlas.

term	Hazard ratio (HR)	std.error	statistic	p.value	signif	DE
ADGRL2	0.883	0.075	-1.660	0.097	.	Down
BLMH	0.970	0.126	-0.241	0.810		Down
DCP2	0.997	0.195	-0.017	0.986		Down
GPSM2	0.971	0.114	-0.261	0.794		Down
H19	0.976	0.043	-0.564	0.572		Down
ICK	0.986	0.102	-0.136	0.892		Down
KCNK2	0.986	0.020	-0.693	0.488		Down
MAP3K3	0.850	0.179	-0.909	0.363		Down
MBNL3	0.832	0.078	-2.351	0.019	*	Down
MSRB3	0.987	0.079	-0.168	0.866		Down
MTF2	0.878	0.249	-0.523	0.601		Down
NFIB	0.920	0.099	-0.838	0.402		Down
PRDM11	0.949	0.104	-0.507	0.612		Down
SLC35B4	0.933	0.181	-0.383	0.701		Down
TRIM29	0.920	0.065	-1.285	0.199		Down
TRPS1	0.856	0.080	-1.955	0.051	.	Down

Table 2: Candidate UP genes from EGF/EGFR-mediated EMT genes found in cell lines by univariate Cox models. Direction of DE is indicated. UP means log2 fold change in cell lines > 0.5. For UP genes only genes with HR > 1 in TCGA patients were kept. Cox model estimated coefficients with SE, t-statistic, and p-value are shown. p-value: .1 - 0.05, .; ≤ 0.05, *; ≤ 0.01, **; ≤ 0.001, ***. EGF/R: Epidermal growth factor receptor, EMT: Epithelial-to-Mesenchymal transition, DE: Differential expression, HR: Hazard ratio, SE: Standard error, TCGA: The Cancer Genome Atlas.

term	Hazard ratio (HR)	std.error	statistic	p.value	signif	DE
ADIRF	1.075	0.072	1.002	0.316		Up
ASPH	1.290	0.119	2.144	0.032	*	Up
BAIAP2L1	1.171	0.146	1.086	0.277		Up
CORO2B	1.040	0.055	0.721	0.471		Up
DDIT4	1.218	0.084	2.353	0.019	*	Up
EHBP1	1.368	0.142	2.215	0.027	*	Up
FADD	1.227	0.071	2.877	0.004	**	Up
FGFBP1	1.006	0.057	0.112	0.911		Up
FKBP1A	1.213	0.192	1.003	0.316		Up
FSCN1	1.106	0.131	0.768	0.442		Up
FXYD5	1.061	0.100	0.588	0.557		Up
HMGA1	1.164	0.120	1.262	0.207		Up
IDS	1.079	0.167	0.454	0.650		Up
ITGB4	1.128	0.109	1.102	0.270		Up
ITPR3	1.090	0.139	0.622	0.534		Up
LAMA3	1.038	0.067	0.551	0.581		Up
LAMB3	1.160	0.092	1.614	0.107		Up
LDHA	1.111	0.134	0.788	0.431		Up
MANCR	1.008	0.046	0.178	0.859		Up
MAPK6	1.009	0.139	0.065	0.948		Up

Table 3: Candidate UP genes from EGF/EGFR-mediated EMT genes found in cell lines by univariate Cox models. Direction of DE is indicated. UP means log2 fold change in cell lines > 0.5. For UP genes only genes with HR > 1 in TCGA patients were kept. Cox model estimated coefficients with SE, t-statistic, and p-value are shown. p-value: .1 - 0.05, .; ≤ 0.05, *; ≤ 0.01, **; ≤ 0.001, ***. EGF/R: Epidermal growth factor receptor, EMT: Epithelial-to-Mesenchymal transition, DE: Differential expression, HR: Hazard ratio, SE: Standard error, TCGA: The Cancer Genome Atlas.

term	Hazard ratio (HR)	std.error	statistic	p.value	signif	DE
MPV17	1.114	0.208	0.521	0.603		Up
MYH16	1.010	0.030	0.346	0.729		Up
NCEH1	1.249	0.098	2.270	0.023	*	Up
NT5E	1.059	0.061	0.945	0.345		Up
PLEKHB2	1.179	0.241	0.683	0.495		Up
POMP	1.564	0.173	2.581	0.010	**	Up
RAB11A	1.036	0.164	0.214	0.830		Up
RAB3B	1.030	0.032	0.933	0.351		Up
RAC2	1.170	0.098	1.603	0.109		Up
RALB	1.149	0.184	0.755	0.450		Up
S100A10	1.135	0.127	0.999	0.318		Up
SERPINE2	1.043	0.061	0.691	0.489		Up
SH3KBP1	1.045	0.092	0.479	0.632		Up
SHANK2	1.064	0.048	1.300	0.194		Up
SHC1	1.013	0.194	0.067	0.947		Up
SQSTM1	1.264	0.137	1.702	0.089	.	Up
STRAP	1.236	0.205	1.034	0.301		Up
TIMP1	1.242	0.091	2.395	0.017	*	Up
TMSB10	1.181	0.136	1.227	0.220		Up
TNNT1	1.047	0.048	0.958	0.338		Up
UCA1	1.009	0.030	0.308	0.758		Up

Figures

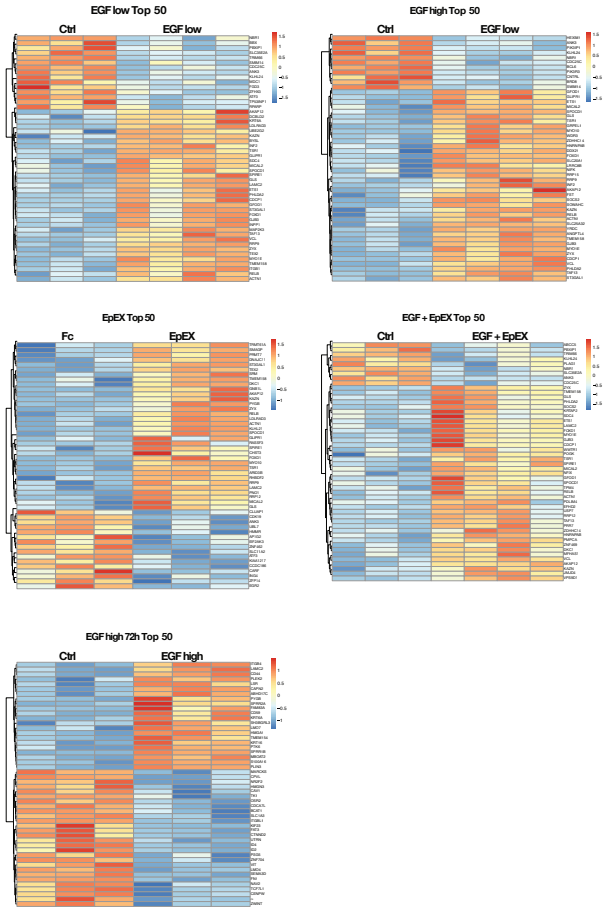

Figure 1: Heatmap with hierarchical clustering of Top50 expressed genes in treatment groups vs. control of Kyse30. Color encodes for row Z-score. Treatment and control groups are noted on the top. Gene names are noted on the right side. Ctrl: Control, Fc: Fragment crystallizable region, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM.

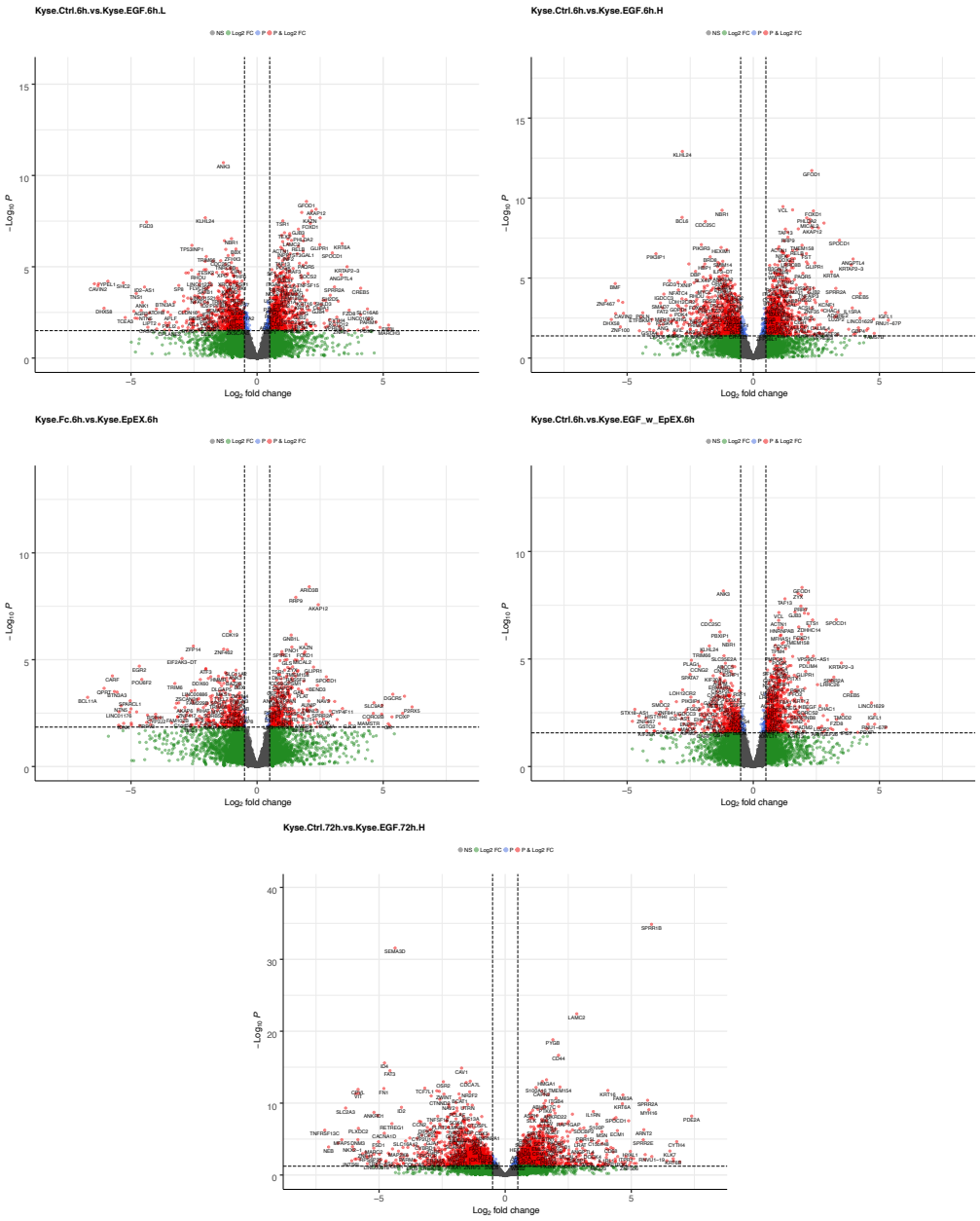

Figure 2: Volcano plot of differentially expressed genes in Kyse30. Lines show thresholds of log2 fold change (log2 FC > 0.5) and p-value ($p \leq 0.05$). Ctrl: Control, Fc: Fragment crystallizable region, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM.

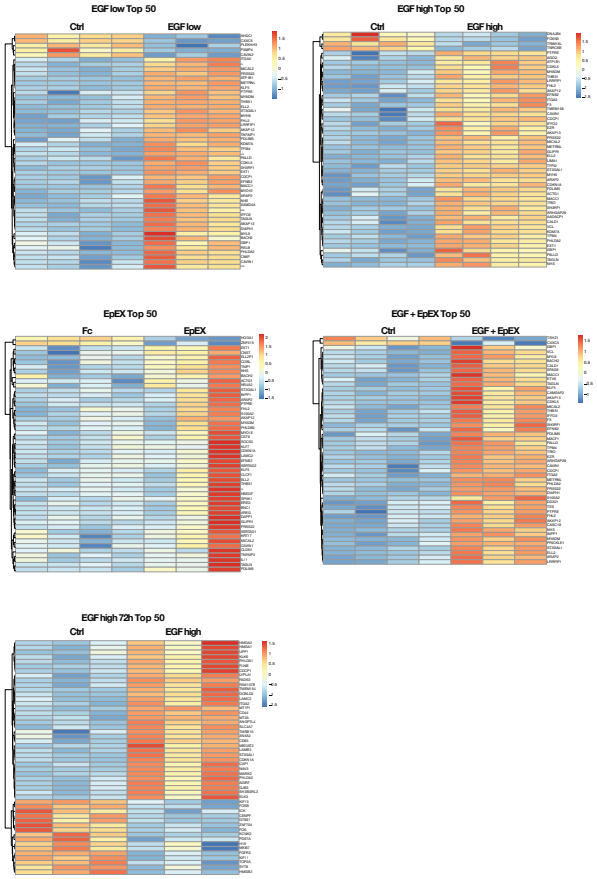

Figure 3: Heatmap with hierarchical clustering of Top50 expressed genes in treatment groups vs. control of FaDu. Color encodes for row Z-score. Treatment and control groups are noted on the top. Gene names are noted on the right side. Ctrl: Control, Fc: Fragment crystallizable region, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM.

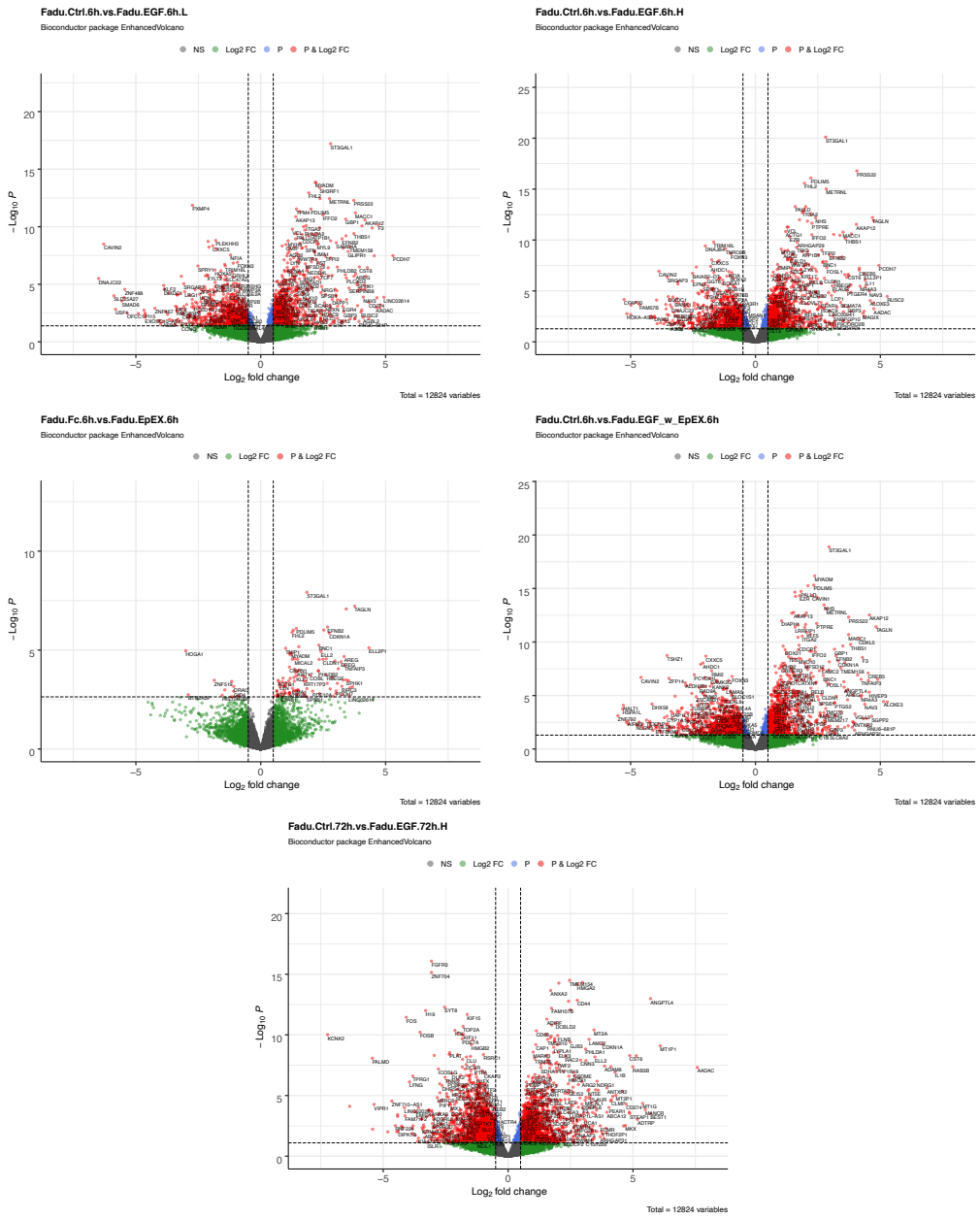

Figure 4: Volcano plot of differentially expressed genes in FaDu. Lines show thresholds of log2 fold change (log2 FC > 0.5) and p-value ($p \leq 0.05$). Ctrl: Control, Fc: Fragment crystallizable region, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM.

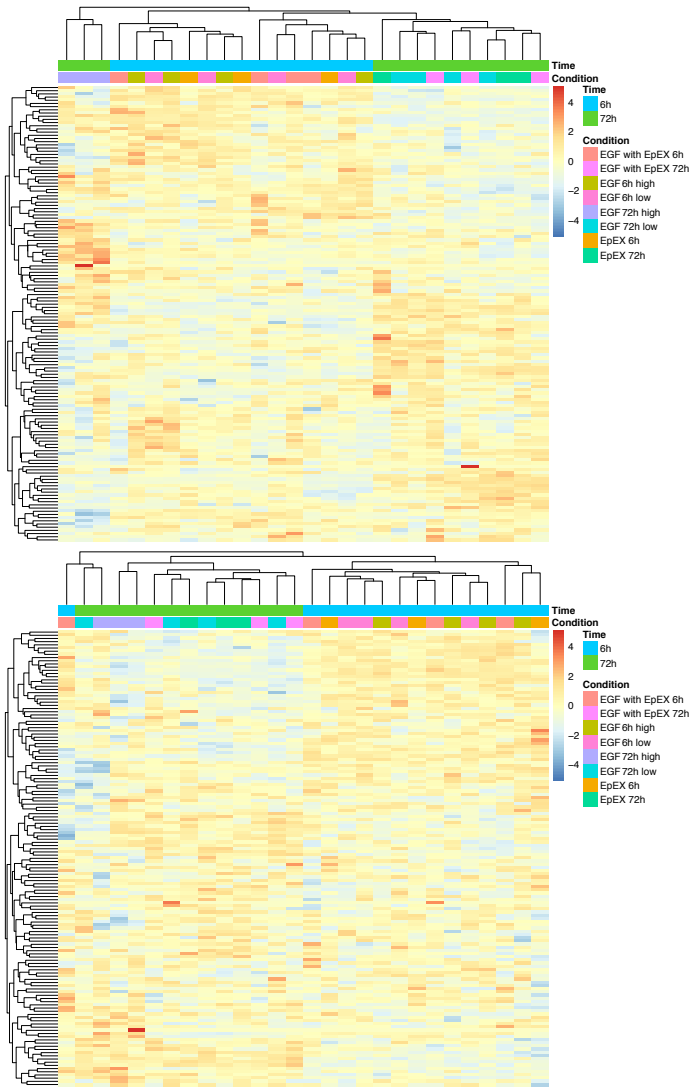

Figure 5: A heatmap with hierarchical clustering of treatments, excluding controls, including all n= 200 Hallmark EMT genes (left) and random n= 200 genes (right) as a control in Kyse30 cells. Applying all n= 200 Hallmark EMT genes to clustering separates EGF-high 72 hours treatments from all others. Controling cluster formation with n= 200 random genes shows no separation of EGF-high 72 hours from others. Color represents centered and scaled gene expresson values in rows. EMT: Epithelial-to-Mesenchymal transition, EGF: Epidermal growth factor, EpEX: Extracellular domain of EpCAM.

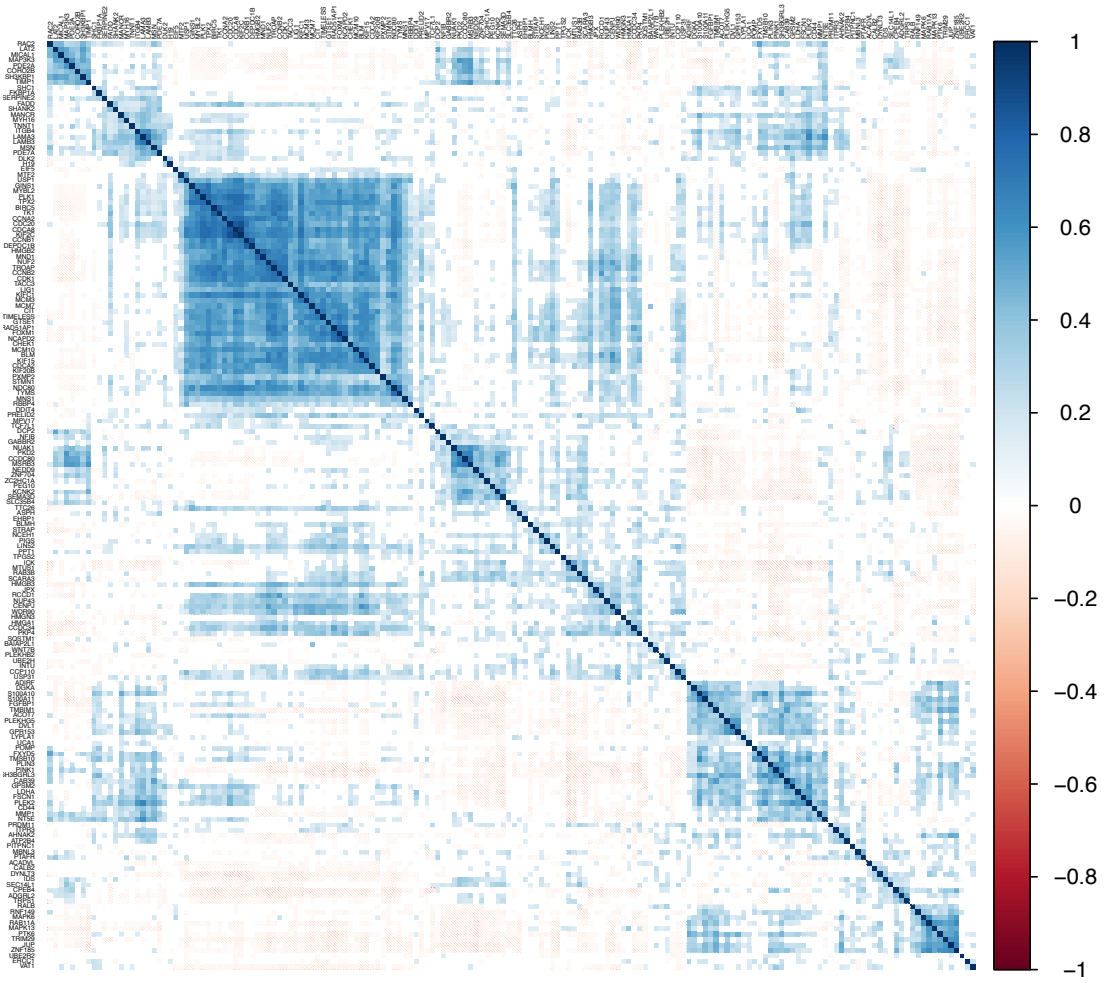

Figure 6: Correlation matrix with hierachical clustering of a Pearson correlation anaylsis in TCGA patients of all n = 172 EGF/EGFR-mediated EMT genes. Color encodes for Pearson correlation coefficient. Insignificant values are blanked. Signifiance niveau ≤ 0.01. Gene names are noted on the side. EMT: Epithelial-to-Mesenchymal transition, TCGA: The Cancer Genome Atlas, EGF/R: Epidermal growth factor/receptor.

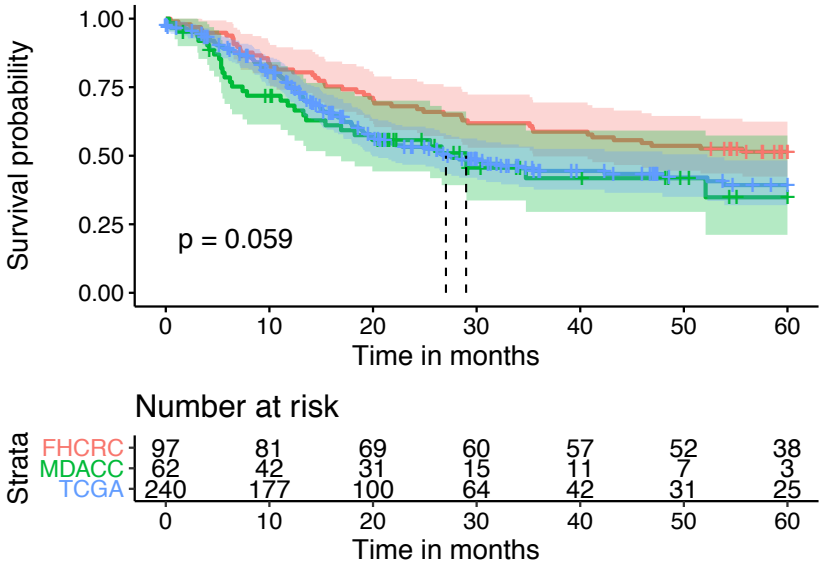

Figure 7: Kaplan-Meier curve with 95% CIs and Cox model log-rank p-value and median overall surival times of TCGA, MDACC, and FHCRC cohorts. Clinical follow-up over 5 years is plotted. Log-rank p-value shows that overall survival of the cohorts are not significantly different between the cohorts. Time in months is shown. Numbers at risk are shown in table. CI: Confidence intervall, TCGA: The Cancer Genome Atlas, MDACC: University of Texas MD Anderson Cancer Center, FHCRC: Fred Hutchinson Cancer Research Center.

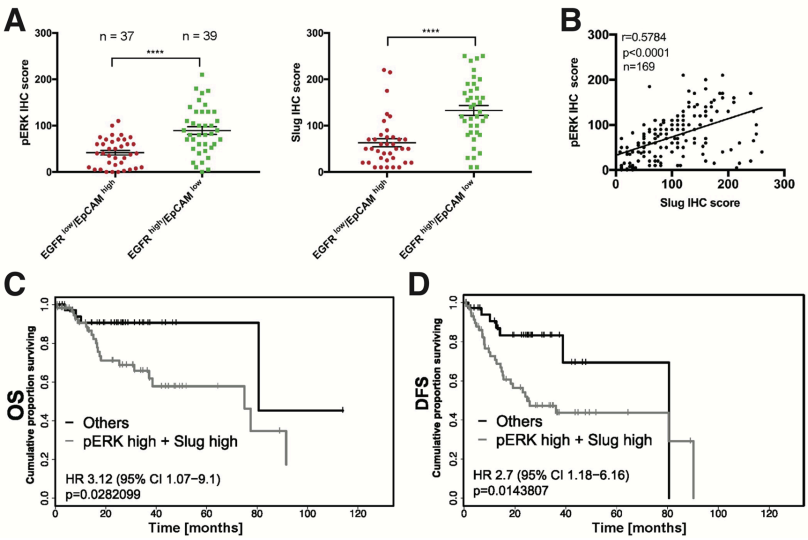

Figure 8: pERK and SLUG expressions are associated with EGFR/EpCAM co-expression subgroups and are prognostic in HNSCC. LMU HNSCC tumor cryosections were stained for EpCAM, EGFR, pERK, and SLUG by IHC. IHC scores were evaluated with a range from 0-300. LMU patients were subdivided into EGFR/EpCAM subgroups according to a cut-off threshold of 150. B) pERK and SLUG IHC scores were compared from EGFR-low/ EpCAM-high and EGFR-high/ EpCAM-low co-expressing patients. Shown are mean (lines) and Student t-test. p-value: ****, < 0.0001. C) pERK and SLUG IHC scores of patients were compared by Spearman correlation. r- and p-values are shown. D) OS (n= 98) and DFS (n= 97) rates of pERK- and SLUG-high patients were assessed by a Kaplan-Meier curve with a Cox model. HR, 95% CI, and log-rank p-values are shown. Patients with EGFR and EpCAM exression < 125 or > 175 were included. Patients with pERK and/or SLUG > 175 (pERK + Slug high) were compared against all others. Time in months was assessed. pERK: Phospho extracellular signal–regulated kinase, EGFR: Epidermal growth factor receptor, EpCAM: Epithelial cell adhesion molecule, LMU: Ludwig-Maximilians-University, IHC: Immunohistochemistry, CI: Confidence interval, HR: Hazard ratio, OS: Overall survival, DFS: Disease-free survival, HNSCC: Head and neck squamous cell carcinoma. Taken from Pan *et al.*, 2018