MASTERARBEIT

Titel der Masterarbeit

"The role of Neuropilin 2 in hepatocellular carcinoma: from meta-analysis to target characterization"

verfasst von

Christian Gröger, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 066 834
Studienrichtung lt. Studienblatt:	Masterstudium Molekulare Biologie
Betreuerin / Betreuer:	A. o. Univ. Prof. Dr. Wolfgang Mikulits

Table of contents

1 Abstract

1.1 Abstract

The epithelial mesenchymal transition (EMT) is a major process involved in tumor progression and metastasis. Several studies have investigated the mechanisms of EMT in different cell models utilizing different modes of inductions and cell lines of various origins. Nevertheless, no attempt has been made to combine these studies and investigate the effectors of EMT in a comprehensive meta-analysis. Therefore, we gathered 24 publicly available and published EMT datasets, elaborated their relationships between each other and extracted an EMT-core gene list defining the majority of the shared EMT regulated genes within these studies. Consequently, we correlated this meta-analysis with a hepatocellular carcinoma (HCC) cell model termed 3p and 3sp. These cells were isolated from a single HCC patient and the 3sp cells were found to derive from 3p cells through EMT. One gene of particular interest was Neuropilin 2 (NRP2). This gene showed a moderate number of upregulations in the meta-analysis as well as a high upregulation in 3sp cells compared to their epithelial 3p counterpart. NRP2 has been implicated in modulating and enhancing various signaling pathways such as vascular endothelial growth factor (VEGF) and transforming growth factor beta (TGF-β) signaling. We addressed the question whether NRP2 exhibits a functional role during EMT or in mesenchymal cells that have undergone EMT. We were able to show that NRP2 knockdown reduced cell migration, invasion, proliferation and adhesion in mesenchymal 3sp cells. Furthermore, we found significant correlations of NRP2 expression with grading in HCC patient samples both on mRNA and on protein level. Besides this functional impact, we found reduced phosphorylation of focal adhesion kinase (FAK), Smad2 and c-Jun N-terminal kinase/stress-activated protein kinase (JNK/SAPK) upon NRP2 knockdown. Several cell lines with NRP2 overexpression constructs as well as shRNA against NRP2 have been generated to elucidate the effect of NRP2 overexpression or knockdown *in vitro* and *in vivo*. In conclusion, NRP2 might play an important role in EMT and cancer progression as indicated by the meta-analysis, by *in vitro* characterization and clinical data.

1.2 Zusammenfassung

Die epithelial zu mesenchymale Transition (EMT) ist ein wichtiger Prozess in der Tumorprogression und Metastasierung. Mehrere Studien untersuchten die Mechanismen der EMT in verschiedenen Modellsystemen mit unterschiedlichen Induktionen. Es wurde jedoch noch kein Versuch unternommen die Studien zu kombinieren und die Effektoren der EMT in einer zusammenfassenden Metaanalyse zu untersuchen. Dafür sammelten wir 24 öffentlich verfügbare und publizierte EMT Datensätze, um ihre Relationen zueinander zu untersuchen und eine Liste von EMT-Genen zu extrahieren, die in der Mehrheit der Studien vorkamen. Diese Metaanalyse wurde mit dem hepatozellulären Karzinom (HCC) Modell 3p und 3sp verglichen. Diese Zellen stammen von einem HCC Patienten, dessen Hepatomzellen eine EMT gezeigt haben. Ein Gen von speziellem Interesse war Neuropilin 2 (NRP2). Dieses Gen zeigte, im Vergleich zu ihren epithelialen 3p Pendant, eine erhöhte Regulation in der Metaanalyse sowie in EMT-transformierten 3sp Zellen. NRP2 wurde mit Modulation und Verstärkung verschiedener Signalwege, wie den Signalwegen des vaskulären endothelialen Wachstumsfaktor (VEGF) und dem transformierenden Wachstumsfaktor (TGF-β) in Verbindung gesetzt. Es ergab sich daraus die Frage, ob NRP2 einen funktionellen Einfluss auf den EMT Prozess oder in mesenchymalen Zellen, die bereits die EMT durchlaufen hatten, besitzt. Wir konnten zeigen, dass der „Knockdown" von NRP2 zur Verringerung der Zellmigration, Zellinvasion, Proliferation und Adhäsion in 3sp Zellen führt. Es war weiters möglich, signifikante Korrelationen zwischen der NRP2 Expression und der HCC Differenzierung in Patienten sowohl auf mRNA als auch auf Proteinebene zu finden. Neben diesen funktionellen Implikationen von NRP2 untersuchten wir ebenfalls die Signalwege von NRP2 in HCC Zelllinien und fanden eine reduzierte Phosphorylierung von FAK, Smad2 und JNK/SAPK nach „Knockdown" von NRP2. Zudem wurden mehrere Zelllinien mit einer konstitutiven Überexpression beziehungsweise mit einem stabilen „Knockdown" von NRP2 generiert, um die positive und negative Modulation der NRP2 Expression *in vitro* und *in vivo* zu untersuchen. Zusammenfassend lässt sich anhand der Meta-analyse, der *in vitro* Charakterisierung und in den klinischen Daten sagen, dass NRP2 eine wichtige Rolle in der EMT und HCC Entwicklung innehaben könnte.

2 Introduction

2.1 Epithelial and mesenchymal cells

A comparison of epithelial versus mesenchymal cells as well as their specific characteristics is given in detail in 3.2. To complete this section, a few things are left to be mentioned in the following. Epithelial cells are located with their basal side on top of the basal lamina (or basement membrane) which is an extracellular matrix (ECM) mainly consisting of laminins and collagen IV. This meshwork provides mechanical support and forms a barrier between connective tissue and epithelial cells. Furthermore, it also regulates cell polarity, differentiation and migration. Epithelial cells are connected via tight, gap, adhesion and desmosomal junctions. Tight junctions prevent the diffusion of molecules through intercellular gaps and support the apical/basolateral polarity. Gap junctions, however, enable cells to exchange small molecules [1]. Cadherins are responsible for direct cell-cell adhesions. They are Ca^{2+}-dependent hemophilic transmembrane proteins which form an adhesion belt through association via β-catenin and other anchor proteins with actin filaments [1]. Finally, desmosomal junctions provide epithelia with the necessary mechanical strength. They consist of cadherins linked to intermediate filaments through adaptor proteins such as desmoplakin [2]. Epithelial cells generally express E-cadherin (epithelial cadherin, cadherin-1, CDH1). In contrast, fibroblasts and nerve cells commonly express N-cadherin (neuronal cadherin, cadherin-2, CDH2). The different cadherins are necessary for contact guidance and tissue assembly in the course of embryonic development [1,3].

2.2 The epithelial to mesenchymal transition (EMT)

In general, EMT describes the conversion of highly organized epithelial cells into mesenchymal cells possessing enhanced migratory abilities. An overview on EMT, its classification into three subtypes and the reverse process, mesenchymal to epithelial transition (MET) is outlined in 3.2. Characteristic epithelial markers are E-cadherin, cytokeratins, adhesion molecules such as zona occludens 1 (ZO-1) and desmoplakin as well as collagen IV. Typical mesenchymal markers are N-cadherin, OB-cadherin, fibroblast specific protein-1 (FSP-1), several ECM proteins (vimentin, fibronectin) as well as a number of transcription factors (Snail, Slug, SIP1/ZEB1, Twist, Goosecoid, LEF-1, FOXC2) [4]. An overview on the process of EMT as well as the respective marker proteins is displayed in Figure 1.

7
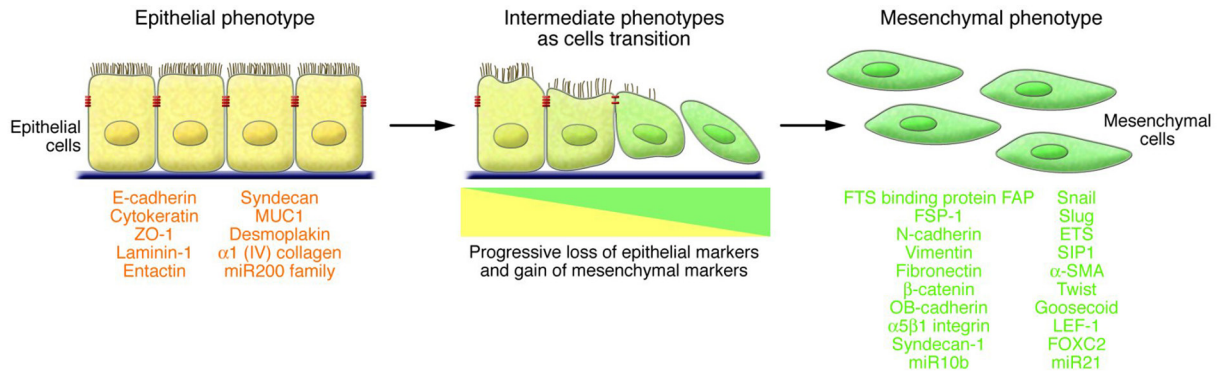
Figure 1: The epithelial to mesenchymal transition (EMT), adapted from [4]. ZO-1, zona occludens 1; MUC1, mucin 1, cell surface associated; miR, microRNA; FTS, fused toes homolog; FAP, fibroblast activation protein, alpha; ETS, E-twenty six; SIP1, survival of motor neuron protein interacting protein 1; SMA, smooth muscle actin; LEF-1, Lymphoid enhancer-binding factor-1; FOXC2, forkhead box C2.

2.3 Signaling in EMT

EMT is regulated by a multitude of signaling pathways during physiological as well as pathological events [5]. An overview on the most important pathways is given in Figure 2. EMT is regulated by a plethora of extracellular signals including ECM components (e. g. hyaluronic acid) and soluble growth factors (e. g. TGF-β and hepatocyte growth factor, HGF) [5]. For example, HGF confers signaling through the receptor tyrosine kinase (RTK) c-Met and the mitogen-activated protein kinase (MAPK) cascade in order to reduce E-cadherin expression [6]. An additional pathway, which is essential for EMT, can be activated by growth factors such as fibroblast growth factor (FGF) which signals through phosphatidylinositol-3'-kinase (PI3K) [7]. Furthermore, Notch signaling has been shown to be involved during EMT [8]. Numerous other pathways, microRNAs (miRNAs, miRs) and hypoxia have also been reported in EMT [4]. TGF-β signaling exerts an important role in EMT, as recently reviewed [9]. Pathways induced by TGF-β interact with pathways such as Wnt, Ras, Hedgehog (Hh) and Notch which results in EMT. Several transcription factors (SNAI1, ZEB1/2, TWIST1/2, β-catenin) form complexes with Smads in order to regulate the expression of both epithelial and mesenchymal genes. TGF-β signaling is depicted in detail in 2.7.

Interestingly, the majority of EMT inducers and pathways lead to a downregulation of E-cadherin [10]. This downregulation is fundamental for EMT and is achieved, amongst many others, via transcriptional repression and promoter hypermethylation [11]. However, E-cadherin expression or activity can also be lost due to mutations [12]. Important transcriptional repressors are the zinc finger proteins Snail and Slug, homodomain/zinc finger proteins ZEB1 and ZEB2 and basic helix-loop-helix (bHLH) proteins E12/E47 and Twist [4,13]. The loss of E-cadherin results in the dissociation of epithelial junctions [14]. Consequently, β-catenin is able to translocate to the nucleus in order to induce Wnt signaling together with TCF4/LEF.

8
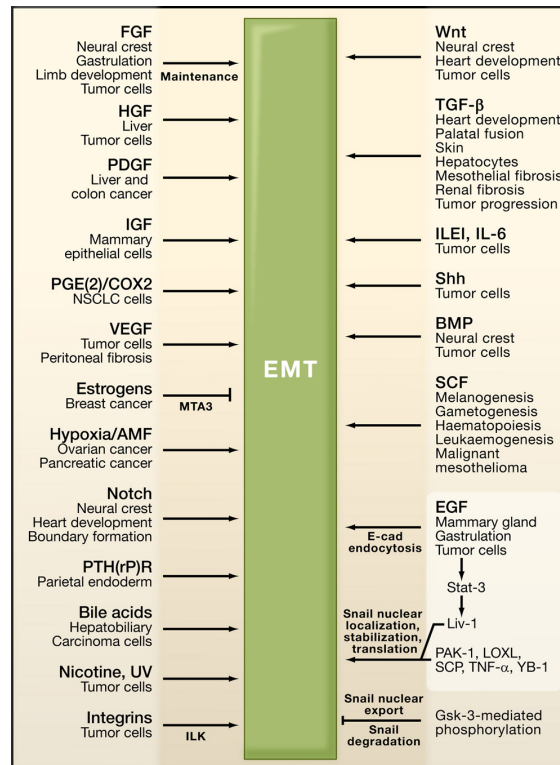
Figure 2: Signaling pathways involved in EMT, adapted from [5]. Biological processes, pathways and their corresponding EMT inducing signals are shown. AMF, autocrine motility factor; E-cad, E-cadherin; EGF, epidermal growth factor; FGF, fibroblast growth factor; BMP, bone morphogenetic protein; IGF, insulin-like growth factor; ILEI, interleukin-related protein; ILK, integrin-linked kinase; IL-6, interleukin-6; LOXL, lysyl oxidase-like proteins; MTA3, metastasis-associated protein 3; PDGF, platelet-derived growth factor; TGFβ, transforming growth factor β; TNF-α, tumor necrosis factor α; PAK-1, p21-activated kinase 2; PTH(rP)R, parathyroid hormone related peptide receptor; SCF, stem cell factor; SCP, small C-terminal domain phosphatase; UV, ultraviolet light; VEGF, vascular endothelial growth factor, YB-1, Y-box binding protein.

2.4 EMT in cancer and hepatocellular carcinoma (HCC)

The ability to invade tissue and produce metastasis is one of the hallmarks of cancer (Figure 3) and it is responsible for most human cancer deaths [15]. Cancer cells can escape from primary lesions though collective cell migration or individually in the form of amoeboid or mesenchymal cells [16]. The latter are a result of epithelial cells which have undergone EMT [5]. As mentioned in 3.2, cycles of EMT and MET are thought to drive invasion and metastasis in carcinoma cells (Figure 4) [16]. The probably most important pathway in EMT is TGF-β signaling which is discussed in the following sections. Interestingly, EMT is not only thought to be responsible for cancer cell dissemination but also contributes to resistance to oncogene-induced premature senescence [17]. The transcription factors which confer this ability are TWIST1 and TWIST2. These transcriptional repressors, which are often reported in EMT, inhibit p21 and p16 which thereby protects cells from senescence [17]. In addition, EMT correlates with resistance to chemotherapeutic agents. It may also be involved in immune-suppressive and -resistance processes [5,18].

9
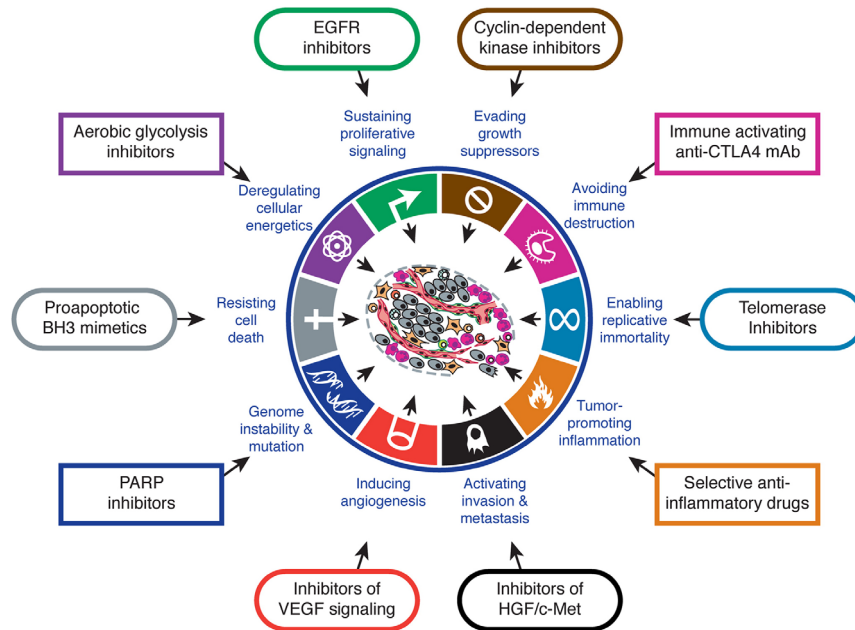
Figure 3: The hallmarks of cancer, adapted from [15]. EMT plays a prominent role in activating invasion and metastasis but is also involved in other hallmarks.
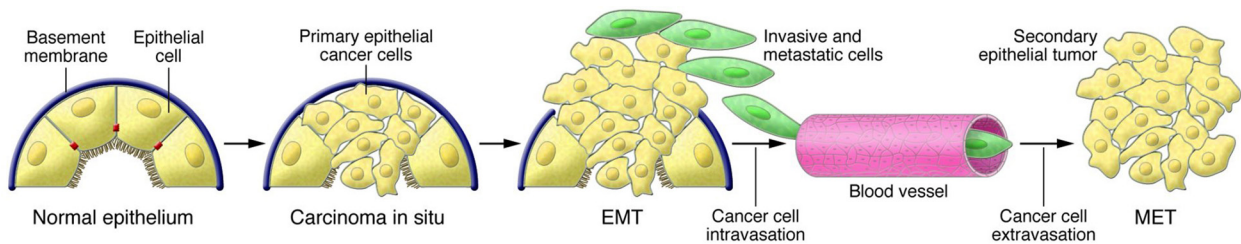
Figure 4: EMT and MET contribute to cancer metastasis, adapted from [4]. Cancer formation from a normal epithelium to an invasive carcinoma progresses through several stages. The latter stage involves epithelial cells which undergo EMT and detach from the basement membrane. The next stage of malignant tumor growth includes EMT, an angiogenic switch and alterations of the ECM composition. The metastatic phase occurs when EMT enables cells to enter and survive in the blood stream as well as to leave the vessels in order to establish micro- and/or macro-metastases through MET.

In 2007, cancer caused about 13% of all human deaths worldwide and the numbers are increasing due to changes in life styles and higher longevity [19]. HCC is a primary liver cancer and with 600,000 deaths annually it is denoted the sixth most common cancer and the third leading cause of cancer related deaths among solid tumors worldwide [20]. HCC commonly occurs in the background of chronic liver disease and cirrhosis which are caused by hepatitis B (HBV) or hepatitis C (HCV) infection, alcoholic liver disease, nonalcoholic steatohepatitis, aflatoxin, obesity, diabetes mellitus or predispositions such as hemochromatosis [21]. Together, HCV, HBV and aflatoxin are responsible for 80% of HCC cases [22]. As HCC is often diagnosed at advanced stages, more than 80% of patients are diagnosed with inoperable HCC and poor prognosis [23]. Curative treatment options are sparse and include liver resection or liver transplantation [24,25]. However, due to metastasis and further development, the recurrence

rate is over 60% after resection [23,26]. E-cadherin and β-catenin levels have been reported to be frequently decreased in HCC patients and correlated with tumor progression, intrahepatic metastasis and poor patient survival [27]. Another study reported that low E-cadherin levels and nuclear translocation of β-catenin correlated with an increase of Snail, Slug and laminin-5, which is reminiscent to EMT [28]. As expected, there are frequently altered genes and regulatory pathways. Nevertheless, individual HCCs display highly heterogeneous changes and disruption patterns [22]. Furthermore, the accelerated proliferation, typical for hepatocytes, as well as the selective growth of monoclonal cell populations are increased in HCC [22]. The process which ultimately leads to the formation of HCC includes the accumulation of genetic changes, increasing genetic stability and alterations of the liver tissue and environment [22]. In the course of HCC progression, fibrosis and inflammation change the microenvironment leading to the death of many hepatocytes and subsequent infiltration of inflammatory cells which deposit connective tissue [22]. This is especially important in HCC because fibrotic and cirrhotic stages most frequently precede HCC [29]. Further development stages involve foci formation of altered hepatocytes which progress into dysplastic nodules resulting in full blown HCC [22].

Every type of cancer as well as HCC in particular displays alterations in numerous signaling pathways which can either be positively or negatively changed by various mechanisms. These pathways are complex, sometimes redundant and necessary for several cellular processes (Figure 5).
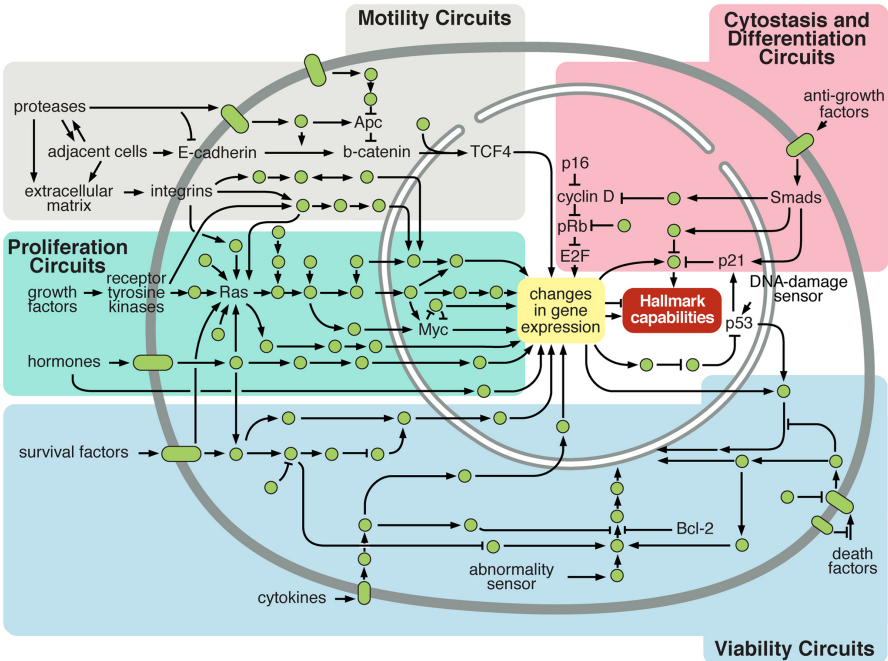
Figure 5: Signaling networks regulating the cancer cell, adapted from [15]. Signaling pathways operating within the normal cell are reprogrammed to enhance the hallmark capabilities for cancer cells. Several pathways are shown here in colored fields, which specifically conduct these capabilities. However, this picture is rather simplified, not showing the considerable number of connections between these pathways as well as the various signals from the tumor microenvironment.

Moreover, there are multiple connections between the different pathways. Thus, the alteration of a single gene or the aberrant activity of a single protein can have diverse consequences [15,22]. Additionally, due to the nature of the hepatocytes, HCC is highly resistant to currently available chemotherapies [23,26]. Since there is no systemic therapy usable for the treatment of advanced HCC, the research focuses on the identification of novel targeted, so called "personalized", therapies [23,26]. Recent findings indicate that NRP2 carries multiple roles in regulating several signaling pathways involved in cancer progression [30,31]. However, no publications are available about the role of NRP2 in HCC so far.

2.5 Neuropilins (NRPs): structure, ligands and isoforms

Neuropilin 2 (NRP2) and its structurally related homologue Neuropilin 1 (NRP1) are 130-140 kD single-pass transmembrane proteins that play multifunctional roles in development, immunity and cancer [31]. They form homodimers as well as NRP1/NRP2 heterodimers [32]. The first to be described was NRP1 (also named CD304 or BDCA-1). It was originally named as the antigen of a monoclonal antibody against neuronal cell-surface proteins presumably involved in neuronal recognition between the optic-nerve fibers and visual centers of *Xenopus laevis* [33,34]. The name NRP resulted from the binding of the antibody to the superficial layer of the neuropile in the tadpole optic tectum [35]. Subsequent studies on NRP1 in transgenic and deficient *Danio rerio* (zebrafish) and mice established the essential role of NRP1 in embryonic neuronal and cardiovascular development [36-39]. NRP1 and NRP2 are receptors for class 3 semaphorins, secreted polypeptides with roles in axon guidance, various members of the vascular endothelial growth factor (VEGF) family, transforming growth factor beta 1 (TGF-β1) and latency associated protein (LAP), hepatocyte growth factor (HGF) and various integrins. However, functional responses may only occur when co-expressed with other receptors such as plexins, VEGF receptors (VEGFRs), TGF-β receptor I and II (TβRI and TβRII) and HGF receptor (c-Met) in the case of semaphorins as well as VEGFs, TGF-β1, LAP and HGF, respectively [30,31]. A comprehensive overview of ligands and co-receptors for NRPs is given in Table 1.

NRP2 is a transmembrane glycoprotein of up to 926 amino acids encoded by 17 exons. It shares an overall amino acid homology of 44% and a similar domain structure with NRP1 [40]. The extracellular regions of Neuropilins consist of two CUB (complement binding factors C1s/C1r, Uegf, BMP1) (a1/a2) domains, two Factor V/VIII homology (b1/b2) domains and a MAM (meprin, A5 antigen, receptor tyrosine phosphatase μ) (c) domain. A single transmembrane domain and a small cytoplasmic domain (43 amino acids for NRP2 and 44 for NRP1) completes the proteins (Figure 6) [30].

Table 1: Neuropilin ligands, adapted from [31].

Ligand	NRP1	NRP2	Ligand	NRP1	NRP2
VEGF-A$_{121}$	+		TGF-β1 and LAP	+	+
VEGF-A$_{145}$		+	TβRI and TβRII	+	+
VEGF-A$_{165}$	+	+	HGF and c-Met	+	+
VEGF-B$_{167}$	+		PDGF and PDGFR	+	
VEGF-C	+	+	FGF-1, 2, 4, 7*	+	
VEGF-D	+	+	FGF receptor-1*	+	
VEGF-E	+		Integrins (α5β1; αvβ3; other)	+	+
PlGF-2	+	+	Fibronectin	+	
VEGFR	+ (R1/R2)	+ (R1/R2/R3)	Galectin-1	+	
Heparin	+		L1-CAM	+	
SEMA3A	+		Glut-1	+	
SEMA3B,C,D,F	+	+	Peptides (CendR: $^{R/}_{K}$XX$^{R/}_{K}$; others)	+	+
SEMA3G		+	Neurotrophin R	+	
Plexin-A1 to A4; D1	+	+			

+, indicates binding to NRPs; R, receptor; PlGF-2 placenta growth factor 2; *, binding occurs but with no effect in HUVEC treated with FGF-2 [41].
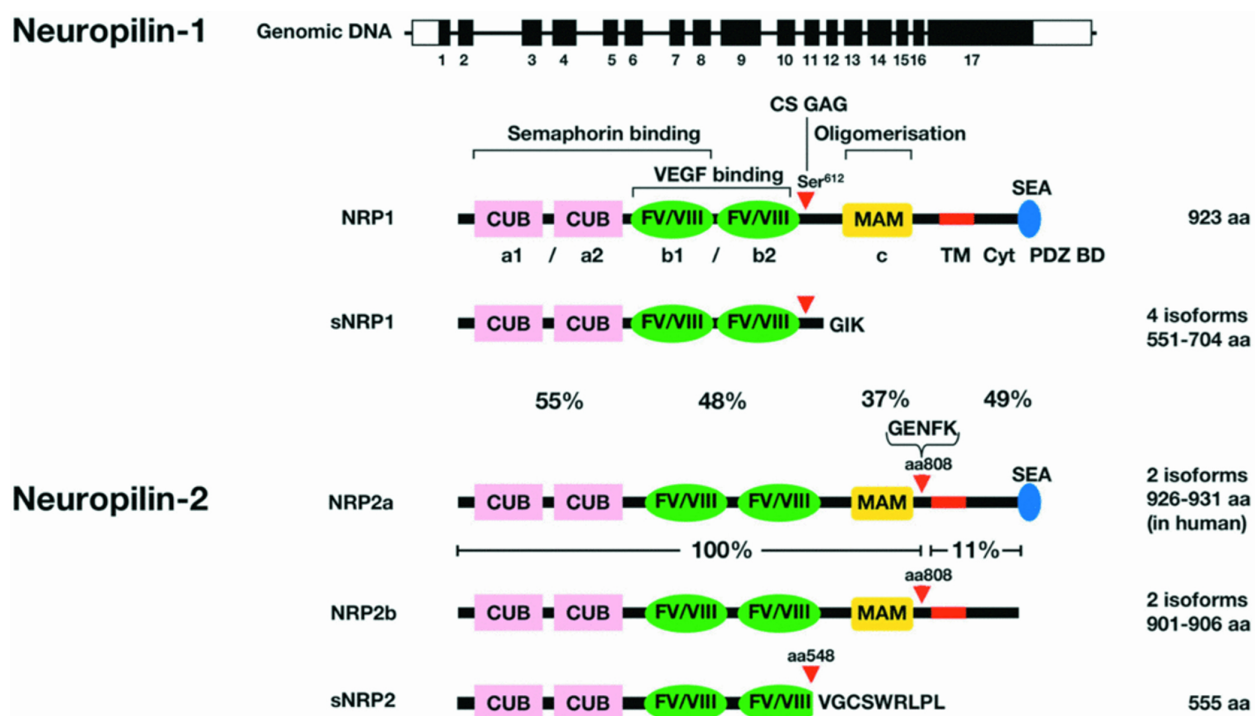
Figure 6: Gene organization and protein structure of Neuropilins [30]. In NRP1, the red arrowheads indicate the positions of Ser612, the major site of glycosaminoglycan (GAG) modification. In NRP2, the red arrowheads indicate an insertion at amino acid 808 of the five residues GEDFL as well as the addition of nine amino acids VGCSWRLPL after the truncated b2 domain at residue 548 in soluble NRP2 (sNRP2). The amino acid homology percentages between the domains of NRP1 and NRP2a isoforms are indicated below the structure of sNRP1. The percentage amino acid homologies between NRP2a and 2b isoforms are indicated below the structure of NRP2a.

13

The three C-terminal amino acids, SEA, present in the NRP2a isoform and NRP1 form a consensus PDZ (post synaptic density protein 95, Drosophila disc large tumor suppressor, and zonula occludens 1 protein) domain binding motif which mediates the association with a PDZ domain containing protein called neuropilin interacting protein-1 (NIP1), synectin or RGS-GAIP-interacting protein (GIPC) [42].

The approximately 110 amino acids of the CUB domains (a1/a2) share homology with complement binding factors C1s/C1r, sea urchin fibropellin (Uegf) and bone morphogenetic protein 1 (BMP1). Such domains are found in functionally diverse proteins generally regulated during development. Four conserved cysteine residues, contained in most CUB domains, probably form two disulfide bridges predicted to form a *β*-barrel similar to the structure of immunoglobulins. These domains mediate associations of proteins in the complement system. In the case of NRPs, the CUB domains are necessary for the binding of semaphorins [30].

The two tandem regions b1/b2, each consists of approximately 150 amino acids, share homology with the C-terminal domains of Factor V and VIII. In these blood coagulation factors, the C-terminal domains are part of a functional domain that facilitates binding of membrane phosphatidylserine on the surface of endothelial cells and platelets. In NRP1, the b1 domain is necessary for VEGF-A$_{165}$ binding. Optimal binding, however, requires also the b2 domain. Several studies indicate the importance of a terminal arginine residue for efficient binding to b1/b2 domains [43,44]. Strikingly, nearly all VEGFs or peptides binding to NRP1 contain a C-terminal arginine, concordant with the C-end rule [45]. Moreover, The b1/b2 domains are necessary for the binding of semaphorins [30]. Investigation of the crystal structure of the NRP1 b1 domain showed a strong similarity but also differences to the structure of the C2 domain of Factor V/VIII [46].

The MAM domain of 170 amino acids is a region often found in the extracellular domains of functionally diverse proteins. These domains are believed to confer homo- or oligomerization and the regulation of protein stability. Hence, this domain may play a role in the oligomerization of NRP1 as well as other domains including the transmembrane domain (TM) [47-49]. A synthetic peptide derived from the NRP1 TM region disrupted its functions as a co-receptor [50].

Neuropilins can harbor post-translational modifications. NRP1 can be modified by glycosaminoglycan (GAG) resulting in a significant higher molecular mass but there is no evidence for a GAG-modified form of NRP2. However, both NRP1 and NRP2 seem to be modified by asparagine (N)-linked glycosylation but little is known about the effects on ligand binding [51,52].

NRP2 can bind to at least five classes of ligands with distinct biological functions. These ligands have been mentioned above. The interaction of NRPs with the corresponding receptors and the hypothetical impact on cell signaling is depicted in Figure 7.
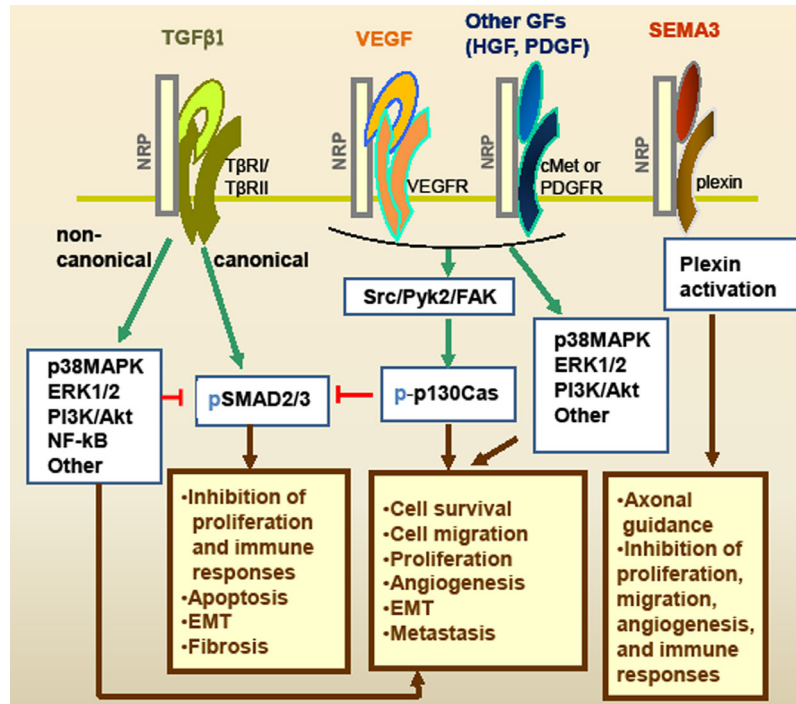
Figure 7: The co-receptor functions of Neuropilins [31].

NRPs are present in all examined vertebrates including mammals, chicken and the zebrafish. Furthermore, they are highly conserved between species. However, no NRPs have been found so far in non-vertebrate organisms, although homologues of semaphorins and the VEGFs have been identified in invertebrates such as *Drosophila* and *Caenorhabditis elegans*. The human genes for NRPs are located on Chromosome 10p12 (NRP1) and 2q34 (NRP2) [53,54]. The molecular masses for both full-length proteins encoded by these genes are between 130 and 140 kD [53]. The NRP2 gene spans >112 kb and contains 17 exons [53]. Considering the strong similarities of NRP1 and NRP2 in terms of exon-intron organization, their exon sizes and positions of splice sites suggests that these genes might have been generated by a gene duplication event [30].

NRP2 can be membrane-bound or soluble. The membrane-bound NRP2 can exist in two major isoforms, NRP2a and NRP2b (Figure 8). NRP2a shares 44% overall amino acid homology with NRP1. The extra-cytoplasmic domains of NRP2b are identical with NRP2a but the transmembrane and cytoplasmic regions display only 11% homology with NRP2a (Figure 6). Four NRP2a isoforms are generated in the mouse by alternative splicing resulting in the insertion of 0, 5, 17 and 22 amino acids after residue 809 [40]. Nevertheless, only two isoforms of NRP2a have been cloned in humans so far, $NRP2a_{(17)}$ and $NRP2a_{(22)}$, which are homologues of the respective mouse isoforms. The ligand-binding properties of these isoforms seem to be identical. However, the insertion of amino acids might modify their ability for dimerization with co-receptors or homodimerization.

15

NRP2b shows little homology with NRP2a after amino acid 808 [53]. The different cytoplasmic region lacks the C-terminal PDZ domain sequence, SEA, which is necessary for interaction with synectin. $NRP2b_{(0)}$ and $NRP2b_{(5)}$ also result from alternative splicing similar to NRP2a. Their significant differences in terms of the cytoplasmic regions suggest that NRP2a and NRP2b might have different functions. This is supported by the differential expression of the isoforms in tissues. Both NRP2a and NRP2b are highly expressed in the brain. NRP2a is mostly expressed in the liver, lung, small intestine, kidney and heart, whereas NRP2b is preferentially expressed in heart and skeletal muscle [53]. As suggested by this distribution, the isoforms are often coexpressed.
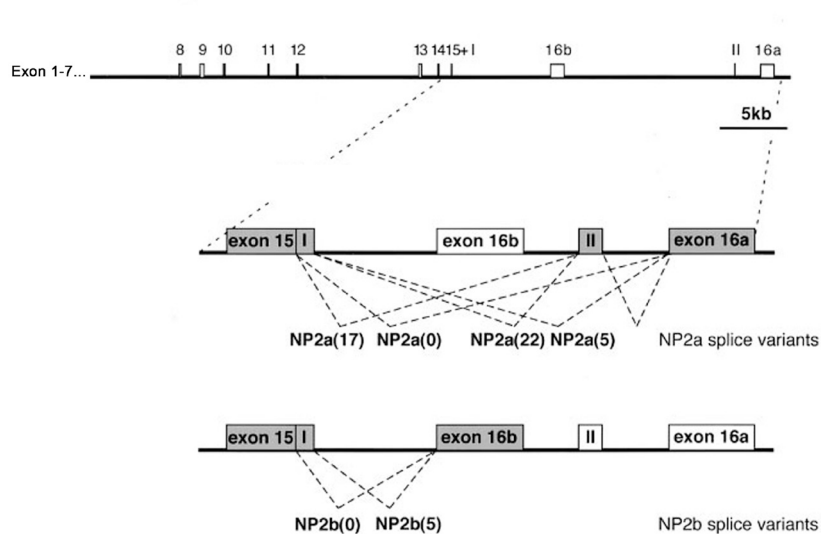
Figure 8: Possible NRP2a and NRP2b splice variants, adapted from [55]. The gene spans 18 exons. This counting system includes exon I and II and treats exons 16a and 16b according to their isoform indication. The exons 16a, 16b, I, and II at the 3' end can be alternatively spliced resulting in the different splice variants. No other gene has been predicted within the NRP2 locus. The two major isoforms NRP2a and NRP2b are generated by alternative splicing of either exon 16a or 16b. Notably, exon 15 contains an internal splice acceptor site which generates an isoform lacking 5 amino acids of the $NRP2a/b_{(5)}$ variant called $NRP2a/b_{(0)}$. Intron 16 contains exon II which has 17 additional amino acids which can be added to $NRP2a_{(0)}$ or $NRP2a_{(5)}$, thereby generating $NRP2a_{(17)}$ and $NRP2a_{(22)}$, respectively. These isoforms are not possible for NRP2b variants due to the position of exon II at the 3' end after exon 16b.

The soluble isoform $_{S9}NRP2$ (555 amino acids, 62.5 kD) results also from alternative splicing. It consists of the two a1/a2 domain, the b1 domain and a truncated b2 domain followed by nine amino acids encoded by intron 9 [53].

2.6   NRP2 functions in development

The disruption of NRP2 in mice (NRP2-null) is not lethal. These mice survive to adulthood but display severe reduction of small lymphatic vessels and capillaries as well as abnormalities in the guidance and fasciculation of cranial and spinal nerves [56-58]. This is in contrast to NRP1-null mice which die between E12 and 13.5 with a variety of cardiovascular and neuronal

16

defects [37,38,59]. Mice double deficient for NRP1 and NRP2 display even earlier embryonic mortality at E8 with a more severe phenotype which is similar to the VEGF-A$_{165}$ and VEGFR2 knockouts [60]. In the zebrafish embryo, NRP2a is expressed in the telencephalon and anterior pituitary while NRP2b is present in the telencephalon, thalamus, hypothalamus and epiphysis [61,62].

Further insights have been gained by the generation of tissue-specific knockouts and mutant mice. The expression patterns of NRPs and their semaphorin ligands can specifically orchestrate the homing of specific subsets of axons during the guidance of motor axons from the spinal cord in the course of vertebrate limb bud development. Both lateral and medial axons of the lateral motor column (LMC) express NRP1 as they leave the spinal cord but NRP2 is restricted to the medial LMC axons. In the whole limb bud, chemorepulsive Sema3A is expressed, whereas Sema3F is present in the dorsal limb bud. Deficiency of either NRP1 or Sema3A results in premature and disorganized invasion of the limb bud by motor axons. Both lateral LMC and medial LMC axons exhibited misprojection. Targeted disruption of either Sema3F or NRP2 causes selective misprojection of medial LMC axons to the dorsal limb whereas lateral LMC axons which normally express only NRP1 track regularly to the dorsal limb [63,64].

The expression of NRPs might also have an important role during vascular development, especially in fate determination of arteries and veins. In 1-day-old chicken embryos, NRP1 and NRP2 are co-expressed in the blood islands but in the 13-somite embryos, expression becomes more restricted. NRP1 is expressed in the arterial and NRP2 is expressed in the venous regions of the primary vascular plexus before the blood flow starts [65]. By the 26-somite stage, which displays a working vasculature, NRP1 and NRP2 are differently expressed in arteries and veins [65]. Similar expression patterns can also be found in the zebrafish embryo. NRP1 is mostly expressed in the dorsal aorta, whereas NRP2 is preferentially present in the posterior cardinal vein [66]. However, it remains unclear whether differential expression of NRPs is necessary for the embryonic segregation of arterial and venous cells in mammalian embryos. Nevertheless, the relatively mild phenotype in NRP2-deficient mice mentioned above, suggests that the expression of NRP2 might be less important than NRP1 expression.

In neuronal cells, the most important biological function of NRPs is mediation of chemorepulsion. NRP2 is necessary for the Sema3F-induced repulsion of superior cervical ganglia [40,67]. This chemorepulsion by Sema3A and 3F has also been reported in porcine aortic endothelial cells as well as human endothelial cells which express NRP1 and NRP2, respectively [41,68-71]. However, the ability of Sema3A to inhibit endothelial-cell migration still requires further investigation [72]. The roles of NRP1 on migration, adhesion to the extracellular matrix and chemotaxis in endothelial and vascular smooth muscle cells have been thoroughly

investigated until today [51,72-76]. However, there are few studies about the role of NRP2 in these processes so far.

The involvement of NRP1 in the immune system has been investigated extensively [30,31]. Special attention should be drawn to a small subpopulation of NRP1 positive plasmacytoid dendritic cells (NRP1+ pDCs) which are important for fighting viral infections [77]. NRP1 displays a possible function in pDCs regarding interferon-α (IFN-α) production [78]. In addition, NRP1 can serve as receptor for Human T-lymphotropic virus Type I (HTLV-1) and maybe also other viruses [79]. The role of NRP2 in DCs is not as established as for NRP1 but there is evidence that NRP2 promotes CCL21-driven chemotaxis involving polysialic acid bound to both NRP2a and NRP2b in mature DCs [80,81].

2.7 NRP2 signaling mechanisms

NRPs are believed to function as co-receptors that can bind extracellular ligands and complex with other transmembrane molecules in order to form a holoreceptor but without another signaling receptor they are not able to transduce a biological signal.

Plexins, L1CAM and Hedgehog signaling

NRP2 is able to form complexes with several plexins (A1, A2, A3 and B1) independently of the presence of semaphorins [82-84]. Crystal structures, mutagenic analysis and deletion studies have proposed a 2:2:2 complex of Sema3A, plexin A1 and NRP1 in which NRP1 might act only for binding while plexins transduce signaling [48,85]. NRP1 has been shown to bind to cell-adhesion molecule L1 (L1CAM) and mediate chemorepulsive Sema3A signals [86]. As mentioned previously, NRP2 is not necessary for Sema3A signaling. Concomitantly, NRP2 was unable to complex with LCAM1 [86]. NRPs have also been reported to regulate Hedgehog (Hh) signaling as well as survival and self-renewal of cancer stem cells [87-90].

VEGFs and VEGF receptors

The role of NRP1 in VEGFR2 signaling in endothelial and tumor cells has been elaborated extensively but with differing results. NRP1 seems to act as a receptor for VEGF-A and as a co-receptor for VEGFR2/KDR [91-93]. However, whether or not NRP1 increases the affinity of VEGFR2/KDR for VEGF-A or if the NRP1-VEGFR2 complex formation requires VEGF binding is still uncertain. Nevertheless, the formation of the NRP1-VEGFR2 complex increases VEGF binding [72,91-95]. Heparin also enhances the affinity of VEGF for NRP1 and it seems to assist the complex formation of VEGF, NRP1 and VEGFR2 [30,96,97]. NRP1 appears to enhance VEGF-induced VEGFR2 signaling through several pathways but it is not essential. The specific

inhibition of NRP1-VEGFR2 complex formation or VEGF-A binding to NRP1 only leads to attenuation rather than inhibition of VEGFR2 signal transduction [72,98]. A connection of NRP1 and VEGFR1 has not been investigated so far. NRP2 has been reported to interact with VEGF-C in a heparin-independent manner and with VEGF-D in a heparin-dependent manner [99]. Following stimulation by VEGF-C and -D, NRP2 showed co-localization and -internalization with VEGFR3 [99]. Complex formation of NRP2 with VEGFR2 or VEGFR3 lowered the activation threshold of VEGFR2; thus, VEGF-A and VEGF-C induced cell survival and migration was increased [100]. VEGFR3 and VEGF-C are both involved in lymphatic vascular development which might explain the phenotype of NRP2-knockout mice and a possible role of NRP2 in lymphangiogenesis [56,101]. Furthermore, there is evidence for a NRP2-VEGFR1 complex but the function and relevance of this complex needs further investigation [102].

TGF-β signaling

As mentioned previously, NRPs are able to bind active TGF-β1 (the most important of the three isoforms), LAP-TGF-β1 (latent form) and its receptors, type I (TβRI, or ALK5) and type II (TβRII). The third component is the type III (TβRIII or betaglycan) receptor [103]. TβRIII binds TGF-β prior to recruiting TβRII which phosphorylates TβRI, thereby forming a heterotetrameric serine/threonine kinase complex. Consequently, in the canonical pathway, TβRI phosphorylates the receptor-associates Smads (R-Smads) Smad2 and Smad3 which then form a heteromeric complex with Smad 4 followed by translocation to the nucleus (Figure 9). Non-canonical pathways include signaling of the TGF-β receptors through JNK, p38 MAPK, ERK1/2, PI3K/Akt, and Rho-like GTPases (Figure 9, C and D). Furthermore, these pathways exhibit an extensive cross-talk between each other [104]. Interaction of TRAF6 with the TGF-β receptor complex leads to the activation of TGF-β-activated kinase 1 (TAK1) which then activates the JNK, p38 MAPK and nuclear factor kappa B (NFκB) pathways (Figure 9, D). The cross-talk of the non-canonical pathways can also include antagonizing canonical signaling by inhibition of Smad3. Tumor progression might be influenced by the balance of the canonical and non-canonical pathways. Hence, a transition from canonical to non-canonical signaling might increase metastasis [105].

The large latent complex (LLC) is the usual form of TGF-β secretion. It consists of LAP-TGF-β (also called small latent complex) which is covalently bound to one molecule of latent TGF-β binding protein (LTBP) [103,106]. LAP and TGF-β are not covalently linked. The activation of TGF-β, which releases mature TGF-β totally or partially, is an important process in regulating TGF-β signaling. Several cell surface proteins can capture latent TGF-β such as integrins, NRP1 and glycoprotein A repetitions predominant (GARP) [31]. Activation of latent TGF-β can occur in different ways. For instance, $α_V β6$ integrins bind the RGD-motif of LAP and

thereby to other ECM components via LTBP [31]. Subsequent traction forces seem to induce conformational changes in LAP resulting in the release of mature TGF-β. Another example is the activation by matrix metalloproteinase (MMP) enzymes as conveyed by $\alpha_V\beta8$ integrins [31]. Nevertheless, there is evidence for a third mechanism involving the activation of latent TGF-β by NRPs after it has bound to an integrin. Cancer cells often co-express RGD-binding integrins and NRPs. Hence, this type of activation might play an important role especially in cancer cells [31]. However, the process how NRPs might activate TGF-β remains unknown. There is a RKFK motif in the b2 domain of NRP1 and NRP2 and a RKPK peptide which are, at least in soluble peptide form, able to activate LAP-TGF-β1 [107,108]. These basic peptides possibly compete with TGF-β1 for LAP binding, thereby or additionally inducing conformational changes in LAP leading to the release of mature TGF-β1.
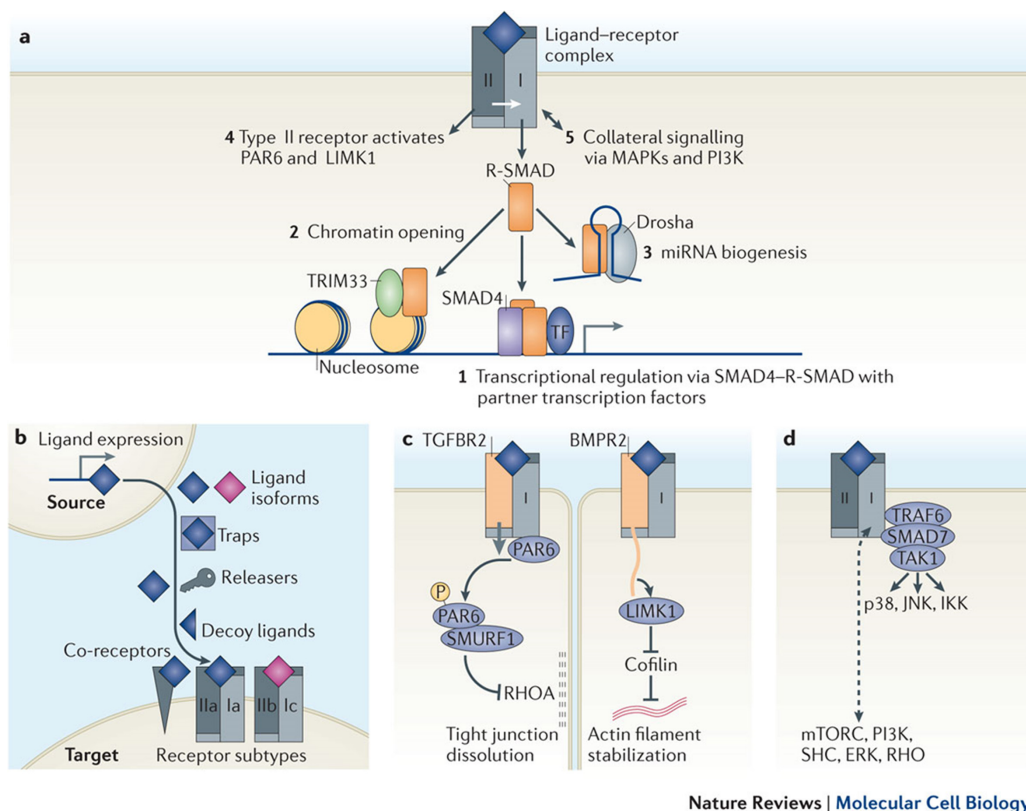
Nature Reviews | Molecular Cell Biology

Figure 9: Canonical and non-canonical TGF-β signaling, adapted from [109]. A: Signaling pathway 1 is named the canonical Smad pathway while pathway 2 is also mediated through Smads. Signaling pathways 3-5 represent the non-canonical TGF-β signaling pathways. **(1)** Receptor-Smad proteins (R-Smads) are phosphorylated and form transcriptional complexes with other factor to regulate hundreds of genes, for details see text above. **(2)** R-Smads are also able to associate with TRIM33 (tripartite motif containing 33) after activation. This protein disables repressive histone marks, thereby opening the chromatin for the access of Smad complexes. **(3)** R-Smads are able to contribute to miRNA processing by Drosha during the generation of several Smad-binding miRNA precursors. **(4)** TGF-β and BMP type II receptors signal through activation of partitioning defective 6 (PAR6) and LIM kinase 1 (LIMK1), respectively. **(5)** Collateral signaling of TGF-β and BMP receptors occurs through several mitogen-activated protein kinases (MAPKs) and PI3K pathways. **B:** The seven classes of variables influencing the access of TGF-β ligands to TGF-β receptors are depicted here. **C:** TGF-β receptor type II (TβRII) is able to signal directly through PAR6 to recruit the E3 ubiquitin ligase SMURF1 which targets RHOA for degradation. This disbands tight junctions in epithelial cells. BMP receptor type II, on the other hand, binds and activates LIMK1 which inhibits Cofilin. This stabilizes actin filaments. **D:** Non-canonical TGF-β signaling via TRAF6 and TAK1, for details see text above.

TGF-β is able to confer tumor suppressive as well as tumor promoting signaling. This ambiguity is called the TGF-β paradox. For instance, TβRII and Smad4 are frequently found inactivated in human cancers [110]. TGF-β may inhibit angiogenesis in several cancers such as diffuse-type gastric carcinoma [111]. TGF-β can regulate cyclin-dependent kinase inhibitors (CDKIs) and MYC expression to block cell proliferation [112,113]. TGF-β antagonists such as c-Ski and SnoN can bind Smad3 and 4 which inhibits the tumor suppressive TGF-β effects [114]. Furthermore, the activation of death-associated protein kinase (DAPK) through TGF-β induces apoptosis [115]. In some HCCs, TGF-β has been reported to activate autophagy genes which mediate autophagy and growth arrest [116]. In contrast, TGF-β may exhibit also tumor promoting effects. In several mesenchymal and cancer cells, it can stimulate proliferation through the platelet-derived growth factor (PDGF) pathway [117,118]. Additionally, TGF-β exhibits anti-apoptotic and pro-angiogenic effects on cancer cells [119,120]. There was also an involvement of TGF-β reported in the regulation of cancer stem cells (CSCs) [121].

Other signaling mechanisms

Besides their role as co-receptors, there is evidence for an independent signaling function of NRPs. NRP1 and NRP2 both contain a C-terminal consensus PDZ-binding region within their small cytoplasmic domain. This region is associated with the PDZ protein neuropilin-interacting protein-1 (NIP1, synectin or GIPC1) as shown by a yeast two-hybrid screen and in human cells [42,122]. Disrupted vessel formation was observed in zebrafish expressing NRP1 lacking the SEA motif. Knockdown of either NRP1 or synectin in zebrafish generated comparable phenotypes [122]. This supports a functional role of the NRP1-synectin association in angiogenesis. Moreover, synectin knockdown in human endothelial cells abrogated NRP1-mediated endothelial migration [122]. However, Sema3A-mediated functions do not require the PDZ domain [48]. Synectin has been shown to associate with up to 20 proteins [30]. It binds to syndecan-4, integrin α5 and α6, M-SemaF and rho guanine nucleotide exchange factor (GEF) and it binds and modulates G-protein-coupled receptors [123-126]. The binding of synectin to NRP1 points to the possibility that NRP1 is involved in signaling networks independently from other receptors and suggests that there might be other so far unidentified proteins interacting with NRP1. There is also evidence for a role of the cytoplasmic region of NRP1 which is not dependent on synectin. NRP1, in response to either VEGF, HGF or PDGF, has been reported to contribute to phosphorylation of the integrin adapter molecule p130Cas (also called breast cancer anti-estrogen resistance, BCAR1) as well as downstream signaling [127,128]. Phosphorylation of p130Cas is generally performed by Src and FAK. Notably, TGF-β is able to activate Src and FAK [129,130]. Nevertheless, the Src/FAK pathway seems to be bypassed by NRP1 [127,128]. Phosphorylated p130Cas contributes to the formation of dedicator of

cytokinesis 1 (DOCK1) at the cell membrane, thereby activating several pathways which promote cell migration, invasion, proliferation and survival [129]. Furthermore, there is an important connection to TGF-β signaling in cancer. On one hand, canonical TGF-β signaling can be inhibited by p130Cas via interaction with Smad3. On the other hand, p130Cas is able to enhance non-canonical TGF-β signaling by increasing p38 MAPK signaling [131,132].

Additionally, there is evidence that protein-protein interactions of the cytoplasmic domain of NRP1 can mediate intracellular signaling [31]. In conclusion, the mechanisms how NRPs modulate signaling of their co-receptors are still unknown. A possible explanation may be that NRPs capture soluble ligands and enhance their availability for other signaling receptors. A second hypothesis could be that NRPs might be involved in endocytosis and/or receptor complex routing. This can improve signaling in selected cases such as clathrin-mediated endocytosis which thereby enhances TGF-β receptor signaling [133,134].

2.8   NRP2 in cancer

The scope of human tumor cell lines and human neoplasms expressing NRPs is quite diverse [31,73,135,136]. In cell lines, NRPs confer VEGF and semaphorin effects on proliferation, survival and migration [137-139]. NRP2 expression has been found in lung cancer, neuroblastoma, pancreatic cancer, osteosarcoma, bladder cancer and prostate cancer [140-145]. NRP1 was suggested to be more expressed in carcinomas and NRP2 was reported to be more prevalently present in non-carcinoma neoplasms. However, there is no clear discrimination of the types of neoplasms which express NRP1 or NRP2 as they are often co-expressed [69,136,146,147]. The expression of NRPs, however, has been reported to vary from one tumor to another [148,149]. Expression of both NRP1 and NRP2 correlated significantly with tumor progression and poor prognosis in patients suffering of non-small-cell lung carcinoma as well as poor survival in breast cancer patients [140,150].

In contrast, NRP2 possibly exhibits antagonistic roles in tumor cell regulation. In murine melanoma xenografts, Sema3F induces a non-metastatic phenotype with little vascularization [69]. Interestingly, Sema3F and Sema3B, which is also a NRP2 ligand, are both localized on the chromosome at 3p21.3, which is generally deleted in human lung cancers [151,152]. Therefore, Semaphorin ligands for NRP2 show potential as tumor suppressors which have to be further investigated. In a human colon cancer xenotransplantation study, NRP2 has been reported to increase tumor growth when forcibly expressed, or decreased tumor growth, prevented tumor formation and increased apoptosis when reduced via knockdown [153,154].

The biological mechanisms of NRPs in cancer are difficult to elucidate due to the high numbers of proteins with which NRPs can interact. Hence, possible contributions of NRPs can

influence cancer cell adhesion, migration, invasion, proliferation and metastasis as well as the interaction of stromal cells (fibroblasts, endothelial cells, immune cells) with cancer cells which express NRPs. An antibody which binds to NRP2 and thereby blocked VEGF-C binding was able to inhibit tumoral lymphangiogenesis and inhibited metastasis in some tumor models [31]. In addition, NRP2 possibly enhances the expression of metastatic or anti-apoptotic genes and enhances chemoresistance through a mechanism which involves β-catenin [155].

TGF-β is produced in many tumors and exhibits an important but also complex role, known as the TGF-β paradox. In early neoplastic stages, TGF-β confers a suppressive role whereas in advanced lesions it exerts a tumor promoting effect including promotion of metastasis and inhibition anti-tumor immunity [103]. The TGF-β paradox, however, is still poorly understood. The fact that NRPs are able to bind TGF-β1 and LAP-TGF-β1 suggests a role of NRPs in the modulation of TGF-β signaling. Notably, VEGF-A$_{165}$ competed with free LAP, LAP-TGF-β1 and TGF-β1 for binding to NRP1. Thus, indicating a possible binding of these proteins to the same or contiguous binding site [31]. Moreover, NRP1 and NRP2 have been shown to enhance canonical Smad2/3 signaling in response to TGF-β1 through interactions with TβRI and TβRII [156,157].

EMT is regarded as a critical event in tumor progression and metastasis. Three signaling pathways which are enhanced by NRPs have been linked to EMT, namely TGF-β, Hh and HGF/c-Met [30,31]. Interactions of PDGFs with NRPs might also be of relevance for EMT. NRP1 has been shown to interact with PDGF-A and B [158]. PDGF-C can induce VEGF-independent angiogenesis and maybe also EMT [159]. PDGF-D is an important inducer of EMT [160]. However, interactions of NRPs with PDGF-C and PDGF-D have not been reported so far [31]. Strikingly, NRP2 has been reported to promote and induce EMT by acting as a co-receptor for TGF-β1 and enhancing TGF-β1-Smad signaling [153]. In this colon cancer xenograft model, NRP2 also increased tumor formation.

The targeting of NRPs for cancer therapy is of great interest and numerous studies have been performed [136,161]. Approaches included NRP blockade with antibodies or peptides, knockdown with siRNA or shRNA and administration of soluble NRP1. There is a small molecule ligand (EG00229) which inhibited VEGF-A binding to NRP1, reduced A549 lung carcinoma viability and enhanced the potency of cytotoxic agents (paclitaxel and 5-fluoracil) [162]. A combined anti-VEGF and anti-NRP1 therapy using monoclonal antibodies displayed synergistic effects in murine cancer models [72]. As Sema3s possibly confer tumor suppressive effect, it seems of particular interest to block VEGF binding. An antibody binding to the VEGF binding site of NRP2 (anti-NRP2[B]) has been demonstrated to inhibit lymphatic tumor development and metastasis [163]. Consequently, a combined therapy blocking both NRP1 and NRP2 seems promising. As mentioned above, NRP1 and NRP2 have overlapping but not exact

the same roles in angiogenesis and cancer. A phase I clinical trial by Genentech, Inc. of the human anti-NRP1 antibody MNRP1685A lead to transient platelet depletion [164]. This antibody was then combined with an anti-VEGF antibody (bevacizumab) which resulted, with or without paclitaxel, in a high incidence of proteinuria [165]. Therefore, anti-NRP1 therapy itself may not be toxic but produces noteworthy side effects in combination with other anti-cancer drugs.

It has been reported previously that NRP1 binds small peptides which are then rapidly internalized. Hence, it could serve as a target which permits internalization of several drugs due to its frequent expression in cancer cells. Especially for large drugs which cannot penetrate the cell membrane, this would be quite useful. The screening of a phage peptide library revealed that numerous cell-penetrating peptides which bind to NRP1 possess a consensus sequence ($^R/_K$XX$^R/_K$), mostly with a C-terminal arginine (R), occasionally also with a lysine (K) [45]. Binding of these peptides seems to occur in the electronegative pocket within the b1 domain of NRP1, consistently with the binding of the C-terminal domain of VEGF, as described above. This binding pattern was named as the C-end rule (CendR) by these authors. There are tumor-homing cyclic peptides named iRGD (CRGD$^K/_R$GP$^D/_E$C) attaching to RGD-binding integrins which are important for cancer therapy [166,167]. These peptides were cleaved by a furin-like protease on the membranes of tumor cells which exposed a CendR motif (RGD$^K/_R$) able to bind to NRP1. Notably, subsequent to the binding of the peptide to NRP1, both are internalized but the internalization mechanism has not been investigated so far. In addition to the elucidation of the internalization, the specificity and safety of CendR peptides needs to be established. CendR motifs appear to play important roles in several biological situations. For example, LAP-TGF-β1 becomes quickly internalized into tumor cells after binding to NRP1 [107]. Additionally, several bacterial and animal toxins (e. g. imperatoxin A from scorpions) as well as some viruses (e. g. HTLV-1) possess a CendR motif [45,168,169]. Hence, their cell-penetrating ability may be of great interest for further research.

2.9   Introduction to bioinformatics and microarrays

Bioinformatics is an interdisciplinary field which deals with methods for efficient storing, handling and analyzing of biological data with first pioneers in 1980 [170]. The Introduction of the Sanger chain termination method as the first industrial-grade DNA sequencing method was the beginning of the generation of more biological data [171]. Moreover, new technologies were developed leading to the next generation sequencing (NGS) methods which produce massive amounts of data per experiment. These methods, together with microarray experiments, contributed to the rising problem of data handling, computing and reproduction [172]. A variety

of computational methods and approaches are nowadays available to tackle the issues of data management and interpretation.

The first publication using a DNA microarray was in 1995 [173] where the first spotting of oligonucleotide probes on a glass surface in high density transformed the way of studying gene expression and drove bioinformatics onwards. The placement of immobilized probes derived from a reference genome to identify target sequences instead of sequences from the tested samples, as in Northern blotting, was a remarkable breakthrough. Further miniaturization allowed the pre-fabrication of such arrays that could be used on nearly every sample anywhere. Many laboratories developed their own arrays and techniques and the cost of commercial available microarrays decreased rapidly. Protocols stabilized of the past years and researchers can choose microarrays for a great variety of applications [174]. Microarrays are a well-established technology with the original intention to investigate gene expression. However, the technology developed further and is now also used in clinical diagnostics [175-177], environmental analysis [178,179] and food analysis [180].

DNA microarrays consist of oligonucleotides or DNA fragments called probes which are attached to a surface. These probes are complementary to their target sequence of a specific marker gene. For oligonucleotide probes, short (15-30mer) and long (50-100mer) fragments can be applied. Short oligonucleotides display higher precision at the detection of short polymorphisms and the discrimination of a single mismatch is feasible with optimized reaction conditions. Nevertheless, they show lower capacity for target binding due to reversible hybridization kinetics. Long oligonucleotides, on the other hand, display a higher target binding capacity but a reduced specificity for the identification of single polymorphisms [181].

There are numerable ways to detect a hybridized target on the microarray. Fluorescent labeling is the most commonly used method. Fluorescently labeled targets are in general prepared through the incorporation of nucleotides which are covalently linked to a fluorophore, a fluorescent dye, using enzymes. Some examples are Klenow-based random labeling (gDNA or PCR amplicons), the usage of labeled primers during polymerase chain reaction (PCR) or labeling during *in vitro* transcription [182].

Data from high-throughput methods such as microarrays or NGS required computational resources and knowledge which low-throughput methods did not. Hence, data analysis and bioinformatics became an important part of studies performed with microarrays. The first step was the standardization of vocabularies and analysis methods, for instance Gene Ontology [183]. The scientific community contributed for a great part to the advance of bioinformatics with, for example, the formation of the Microarray Gene Expression Data (MGED) Society and the development of data analysis tools such as Bioconductor [183,184]. Initial microarray experiments showed a great discrepancy if repeated or reproduced by different laboratories.

Thus, guidelines such as Minimum Information About a Microarray Experiment (MIAME) and public databases for data sharing such as the Gene Expression Omnibus (GEO) or ArrayExpress (AE) where created [185-187].

Genome-wide gene expression data are distributed by a number of different resources nowadays. These include primary repositories, added-value databases, topical databases and integrative database [174]. Primary archives, such as GEO and AE, mostly contain gene expression and other genomic data from publications in peer-reviewed journals. Most journals require data submissions to meet the MIAME criteria and to be uploaded to either AE, GEO or the DNA Data Bank of Japan (DDBJ). Furthermore, the AE database imports great amounts of data available in GEO [188]. Added-value databases use information from primary data to tackle questions such as "where is a particular gene expressed?" or "what is the difference in gene expression between the normal or the disease state?" [174]. The largest of these databases is the Gene Expression Atlas at the European Bioinformatics Institute (EBI) which covers more than 3,000 studies with 100,000 different assays from AE [189]. Topical databases, on the other hand, use primary expression data only for a certain research domain, such as Oncomine and GCOD for cancer data sets [174]. Integrated repositories combine public expression data with other methods such as additional annotation, pathways, phenotypes, drug data and literature-based text mining [174]. This can be application-driven, such as IntOGen for cancer or the Connectivity Map for drug response and disease or species-driven, such as WormBase or the Mouse Genome Informatics (MGI) [174].

The idea to analyze several different gene expression experiments in a single meta-analysis and compare the results has been performed years ago and suffered from severe problems of reproducibility. For example, an attempt to replicate the results from 18 microarray studies published in the years 2005-2006, which were deposited at GEO and AE, was only successful for two [190]. Six were only "partially" reproduced and for ten it was not possible. This discrepancy was due to the lack of data. Hence, MIAME compliance of gene expression studies is crucial for further research. There are two possibilities to reuse and analyze public expression data: reuse of raw data and reuse of processed or summary data [174]. The combination of many dataset improves the ability to detect weak signals. Thus, most meta-analyses gain their power through numbers. However, the reuse of raw data is not trivial. There are many differences in the available platforms and one must account for such biases when using different platforms. Nevertheless, the usage of data from the same platform is much more feasible and can be performed using an identical pipeline of methods for all datasets [174]. In contrast, the use of summary-level data such as p-values or effect sizes (e. g. fold change) is a much popular way because of its inter-platform flexibility. Therefore, this method was used in the presented study.

2.10 Aim of the study

EMT is a crucial event in cancer progression and a major process involved in metastasis. However, the pathways controlling EMT are still poorly understood. Many model systems for EMT have been investigated so far but there was no approach to elucidate the similarities of these EMT models. Hence, a meta-analysis which should elaborate the common EMT regulators of these studies was performed [191]. The aim of this study was to analyze the relationships of different EMT models with each other and identify those genes which were deregulated in the majority of them. The following section deals with the *"Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression"* by Christian J Gröger, Markus Grubinger, Thomas Waldhör, Klemens Vierlinger and Wolfgang Mikulits which was published in PLOS ONE in 2012 [191].

NRP2 has been implicated in the regulation of various signaling pathways of which many are important for cancer progression [30,31]. Moreover, it has been shown to promote EMT in a colon cancer cell line through cooperation with the TGF-β-Smad pathway [153]. Consequently, the investigation of the role of NRP2 in the progression of HCC and EMT during HCC was promising. The meta-analysis of EMT models revealed that NRP2 was upregulated in some of those studies. Notably, NRP2 showed a strong upregulation in the EMT-transformed 3sp cell line compared to its epithelial counterpart, the 3p cell line. The aim of this thesis was to investigate a novel putative regulator of EMT and derive a promising candidate gene, namely NRP2, and investigate its involvement HCC progression as described in section 4.

3   Manuscript "Meta-analysis of EMT signatures"

Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression

Christian J Gröger[1], Markus Grubinger[1], Thomas Waldhör[2], Klemens Vierlinger[3] and Wolfgang Mikulits[1]*

[1]Department of Medicine I, Division: Institute of Cancer Research, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria; [2]Department of Epidemiology, Centre of Public Health, Medical University of Vienna, Vienna, Austria; [3]Austrian Institute of Technology, Vienna, Austria.

*Corresponding author: Wolfgang Mikulits, Department of Medicine I, Division: Institute of Cancer Research, Medical University of Vienna, Vienna, Austria. Tel: +43 1 40160 57527; Fax: +43 1 40160 957519; E-mail: wolfgang.mikulits@meduniwien.ac.at

3.1   Abstract

The epithelial to mesenchymal transition (EMT) represents a crucial event during cancer progression and dissemination. EMT is the conversion of carcinoma cells from an epithelial to a mesenchymal phenotype that associates with a higher cell motility as well as enhanced chemoresistance and cancer stemness. Notably, EMT has been increasingly recognized as an early event of metastasis. Numerous gene expression studies (GES) have been conducted to obtain transcriptome signatures and marker genes to understand the regulatory mechanisms underlying EMT. Yet, no meta-analysis considering the multitude of GES of EMT has been performed to comprehensively elaborate the core genes in this process. Here we report the meta-analysis of 18 independent and published GES of EMT which focused on different cell types and treatment modalities. Computational analysis revealed clustering of GES according to the type of treatment rather than to cell type. GES of EMT induced via transforming growth factor-β and tumor necrosis factor-α treatment yielded uniformly defined clusters while GES of models with alternative EMT induction clustered in a more complex fashion. In addition, we identified those up- and downregulated genes which were shared between the multitude of GES. This core gene list includes well known EMT markers as well as novel genes so far not described in this process. Furthermore, several genes of the EMT-core gene list significantly

correlated with impaired pathological complete response in breast cancer patients. In conclusion, this meta-analysis provides a comprehensive survey of available EMT expression signatures and shows fundamental insights into the mechanisms that are governing carcinoma progression.

3.2 Introduction

The epithelial to mesenchymal transition (EMT) has been originally described as an essential process of metazoan embryogenesis [1]. In the past decade, EMT has been realized as a critical event in carcinoma progression as epithelial tumor cells acquire a mesenchymal phenotype that allows them to detach from the primary tumor and to invade into the local tissue [2]. In general, polarized epithelial cells are organized by cell-cell junctions and cell-anchoring complexes to form apical and basolateral surfaces. In contrast, mesenchymal cells form irregularly shaped structures in the absence of tight adhesions to the neighboring cells and reduced cell contact to the substratum. Mesenchymal cells have an elongated shape compared to epithelia and display an anterior-posterior polarity that enables enhanced migration through reduced adhesion forces. While epithelial cells invade collectively in clusters, mesenchymal cells show individual cell movement that allows them to disseminate from bulk cells [3]. In addition, a partial EMT displaying different levels of E-cadherin expression has been observed that might still lead to collective cell invasion [4].

EMT has been classified into three subtypes [5]. Type 1 EMT is required for embryogenesis to provide gastrulation and formation of neural crest cells that differentiate into various cell types without systemic spreading. Type 2 EMT is involved in tissue regeneration and fibrosis of different organs such as the kidney, liver, lung and intestine leading to the accumulation of connective tissue. Type 3 EMT associates with a gain in malignancy of carcinoma cells. Neoplastic epithelial cells induced to undergo EMT are frequently localized at the invasive front of the primary tumor and initiate the cascade of tumor cell dissemination by local cell invasion which is followed by the entry into the vasculature. Notably, EMT represents a transient and reversible process that can lead to a mesenchymal to epithelial transition (MET) upon metastatic colonization [5,6]. Cycles of EMT and MET are assumed to be involved in metastasis formation at distal sites [3]. Yet, the molecular basis for the changes in epithelial plasticity by EMT and MET is still an open issue and its role in cancer patients is a matter of debate. Signaling molecules and inducers of type 3 EMT confer the resistance of cancer cells to apoptosis and oncogene-induced senescence as well as chemoresistance [6]. Recent findings indicate that EMT provides mesenchymal cells with stem cell features that enable carcinoma cells to generate metastasis at secondary sites [3]. These cancer stem cells, also termed

cancer initiating cells, share phenotypic and functional characteristics with migratory embryonic cells displaying a mesenchymal phenotype [6].

Profiling of the transcriptome using microarrays has been widely used to elucidate the expression patterns during EMT under different conditions which revealed novel biomarkers and molecular mechanisms from single studies. A meta-analysis usually describes the combination of a large number of studies from different samples and tissues or the comparison of own data with published data [7,8]. Recent progress in the establishment of gene expression datasets enables to identify new markers and relevant mechanisms which were underestimated in single studies but emerged from a meta-analysis. By now, a plethora of gene expression studies (GES) covering a wide variety of cell types undergoing EMT together with various modes of induction are available. Yet to our knowledge, no meta-analysis dealing with these EMT studies has been performed so far.

Changes in a biological system require a concerted alteration of gene expression sets. Bioinformatic enrichment analysis tools investigate gene expression sets for such changes. These tools examine the overrepresentation of gene sets in comparison to the whole genome, map an input list of genes to biological categories in online databases and statistically assess the overrepresentation of genes for each biological category or annotation such as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and gene ontology (GO) terms [9]. The use of several single enrichment tools for the same input list and the consideration of only consistently enriched categories have been reported to be a very promising strategy [10,11].

We gathered data from 18 published and independent GES of EMT and extracted gene lists of significantly up- and downregulated genes for cluster analysis. This approach revealed gene clusters according to treatment modalities rather than to cell type. We subsequently extracted an EMT-core list consisting of 130 genes with official gene symbols and names which was further investigated by enrichment analysis with several single enrichment tools. Notably, selected genes from the EMT-core list significantly correlated with impaired pathological complete response (pCR) in breast cancer patients. This analysis proposes that the EMT-core gene list is relevant for the recognition of the molecular mechanisms of EMT. In addition, the cluster analysis shows novel insights into the relationships of EMT processes across different cell types and induction modes.

3.3 Results

3.3.1 Data collection of gene expression studies (GES)

To assess the similarities between published GES and define a core gene list of human EMT, we analyzed 18 independent GES of EMT. These 18 independent and published GES consisted of 24 datasets in total (Table 1). Several authors reported EMT kinetics of different cell types or dose-dependent effects of EMT inducers within single studies. Nevertheless, only the particular testing point showing the strongest effect or EMT phenotype, as reported by the authors, has been selected. Takahashi *et al.* published two related GES, of which one consisted of two datasets, resulting in three datasets of one independent study [12]. Taube *et al.* reported 5 datasets published within one GES with similar expression patterns and different modes of EMT induction [13]. Processed data (normalized and generally logarithmized data) were downloaded from the Gene expression Omnibus (GEO) and ArrayExpress (AE) databases and annotated with BioConductor and NetAffx. Numerous GES, available on GEO and AE, were excluded as they either did not provide processed data or did not contain replicates or have not been published. Due to the variety of microarray formats as well as different normalization and filtering methods used in the literature, we used processed instead of raw data in order to maintain the quality criteria applied by the authors during the data preprocessing. Two-tailed Student's *t*-test was used to compute p-values. Significantly up- and downregulated genes were selected to meet a fold change greater than 2 or lower than 0.5 and a p-value below 0.05.

3.3.2 GES cluster analysis

We generated a matrix containing gene symbols across the analyzed GES (n=14,113) that are all uniquely reported. Significantly up- and downregulated genes of each GES were transferred into the matrix according to their type of regulation. Upregulated genes were labeled with 1, downregulated genes with -1 and not differentially regulated genes with 0 (Table S1). This data distribution consisted of 88.22% not differentially regulated genes and 11.78% up- or downregulated genes and is significantly different to a binomial distribution with those parameters (p<0.0001). In order to determine a cutoff for the number of GES sharing a particular gene used for cluster analysis, the binomial distribution function provided by R as well as the preliminary hierarchical clustering results of each cutoff option were analyzed (data not shown). From this we decided to investigate the clustering of genes shared between at least 10 datasets (n=365; p<0.0001; Figure 1). In addition, this analysis showed clusters of GES according to the mode of EMT stimulus rather than to cell type (Figure 2A). Interestingly, a more

stringent clustering of genes shared between at least 14 of the analyzed GES datasets provided similar clusters, despite the fact that this list contains only 41 genes (Figure 2B and Figure S1).

3.3.3 Generation of the EMT-core gene list

Based on the cluster analysis of the GES, we aimed to define a meaningful EMT-core gene list which describes the majority of the involved genes across the analyzed GES. The cluster analysis of the genes shared between at least 10 datasets contained 365 genes (Table S2). However, it does not show whether a gene is up- or downregulated across different GES. Therefore, the list was filtered to keep only genes which were either up- or downregulated in at least 10 of the GES datasets. The resulting list contained 130 genes of which 67 are up- and 63 are downregulated (Table 2 and Table S3). This selection of genes could be further classified into five categories ((i) cell adhesion and migration, (ii) development, cell differentiation and proliferation, (iii) angiogenesis and wound healing, (iv) metabolism, (v) others or unclassified) according to single enrichment analysis as described below. Several genes were also present in more than one of those categories (Table S3). In conclusion, this resulting EMT-core gene list contains 130 genes which were derived from a multitude of cell types and EMT initiation methods.

3.3.4 Consistently enriched KEGG pathway and GO term analysis of the EMT-core gene list

To further analyze the EMT-core list consisting of 130 genes, a rigorous single enrichment analysis combined with stringent selection criteria was performed. First, an enriched KEGG pathway or GO term had to contain at least 5 genes from the input list and a p-value below 0.05 to be considered significant. An enumeration of significantly enriched terms and pathways is shown in Table 3. Second, a significantly enriched KEGG pathway or GO term had to be observed in at least 4 out of 5 used bioinformatic tools. Third, a consistently enriched KEGG pathway or GO term had to be identified in both the EMT-core gene list and the 365 gene list. Using these criteria, we obtained 6 KEGG pathways, 20 GO biological processes and 15 GO molecular functions consistently enriched in both lists (Table 4). The KEGG pathways consisted of the MAPK signaling pathway, axon guidance, focal adhesion, ECM-receptor interaction, regulation of actin cytoskeleton and pathways in cancer. The GO biological processes could be grouped into processes involved in tissue development, wound healing, cell migration or cell proliferation. The GO molecular functions consisted of ECM and cytoskeleton constituents, peptidase inhibitors and the binding of collagen, growth factors, heparin and integrin. As expected, the list with 365 genes comprised all significantly enriched pathways and GO terms from the 130 genes EMT-core list except for 2 GO biological processes (ECM organization and

lung development). Several more KEGG pathways, GO biological processes and molecular functions could be identified in the list with 365 genes (Table 3 and 4). All these pathways, biological processes and molecular functions are well known to be involved in EMT [5,14-16], and thus confirm the integrity of our EMT-core gene list. In addition, both the EMT-core list and the list with 365 genes display comparable enrichment ratios of KEGG pathways and GO biological processes (Figure 3) as well as GO molecular functions (Figure S2). Therefore, the list containing 365 genes may be considered as an amelioration of the EMT-core list by containing additional genes that might have an ambiguous role in EMT. In summary, our EMT-core list of 130 genes and its amelioration containing 365 genes show strong enrichment of EMT-relevant processes.

3.3.5 Clinical relevance of the EMT-core gene list

The EMT-core gene list contains several genes with yet unidentified roles in cancer progression and/or EMT. We aimed to investigate the clinical relevance of this selection of genes. Therefore, we correlated their expression with overall survival of patients suffering from squamous cell lung carcinomas (SCC) [17] and pathological complete response (pCR) of breast cancer patients [18]. From the downregulated genes of the EMT-core gene list, low FXYD3 expression showed a trend to poor overall survival of SCC patients (p=0.17) and low expression of LAD1 (p=0.00074), SLC7A5 (p=0.0093) and SLPI (p=0.043) significantly correlated with worse pCR of breast cancer patients. From the upregulated genes of the EMT-core gene list, high PTX3 expression tends to poor overall survival of SCC patients (p=0.16) and high expression of NID2 (p=0.0091), SPOCK1 (p=0.038) and SULF1 (p=0.00029) significantly correlated with impaired pCR of breast cancer patients. These correlations demonstrate that the comparison of different data sets is a powerful tool to identify novel relevant target genes that do not emerge from single studies.

3.4 Discussion

Over the past decade a considerable number of GES that deal with EMT have been accumulating in the literature. These cover a variety of cell types which display EMT and include different modes of EMT induction. So far, these resources have only been partially used to compare single findings with those in the literature [8,19,20]. To our knowledge, no attempt has been made to investigate the majority of the independent GES of EMT for their relations to each other. Although we are aware that gene expression data of EMT are not complete, we analyzed the currently available GES to generate an EMT-core list of genes altered most frequently during the EMT process, as depicted in the flow chart (Figure S3).

Cluster analysis of genes shared between at least 10 GES datasets revealed clusters of GES with the same or a similar treatment type. The GES in which EMT was induced by TNF-α either alone or in combination with TGF-β, by TGF-β alone or by different transcription factors consistently grouped together. These clusters persisted when genes shared between at least 14 datasets were used for cluster analysis. A clear clustering of different types of EMT induction, however, would have only been possible if an adequate number of GES on each of these EMT initiation methods existed. Since several treatment modalities are only represented once in the literature, such GES cluster to their most related treatment type.

One cluster predominantly emerged from GES of TGF-β-induced EMT which consisted of 13 datasets. Interestingly, the cluster includes the exogenous expression of Six1 (Micalizzi *et al*; GSE23655; [20]) which has been shown to enhance tumor-promoting TGF-β signaling, and Runx2 (Baniwal *et al*; GSE24261; [21]) that acts downstream of TGF-β signaling [22-25]. Hence, this supports the clustering of these studies together with others using TGF-β as EMT initiator. The study by van Zijl *et al.* (GSE26391; [26]) described the analysis of epithelial and mesenchymal hepatocellular carcinoma cells derived from the same tumor patient. The clustering of this study along with other studies with TGF-β-induced EMT suggests an involvement of TGF-β signaling during the establishment of the mesenchymal cell line.

The cluster of GES with TNF-α as EMT inducer contained the study by Takahashi *et al.* which analyzed the ARPE19 cell line treated with either TNF-α alone (GSE15205_TNFa), TNF-α together with TGF-β (GSE12548) or TGF-β alone (GSE15205_TGFb) in order to induce EMT [12]. The two datasets with TNF-α treatment formed a consistent cluster. However, the third dataset which was obtained from the exclusive treatment with TGF-β clustered to other GES describing EMT initiation by TGF-β. Hence, these data suggest a stronger impact of the EMT stimulus on the clustering rather than the cell type.

One cluster consisted mainly of the datasets from Taube *et al.* (GSE24202; [13]) who reported the induction of EMT in HMLE cells using overexpression of Twist, Snail, Goosecoid and TGF-β as well as the knockdown of E-cadherin. Consistent with the data reported by Taube *et al*, the datasets from Snail- and Twist-induced EMT were the most similar within this cluster. This finding is concordant with the fact that Twist is a direct target of Snail [27]. The high number of datasets in this study might lead to an overrepresentation within the cluster analysis. Furthermore, the use of the same cell line as well as transcription factors with similar targets such as Twist and Snail might lead to a high level of similarity within the datasets of this particular study.

The cluster comprising of Ke *et al.* (E-TABM-949; [28]) who utilized high cell density culturing of EPT2 cells and Ohashi *et al.* (GSE27424; [29]) who described a NOTCH3 knock-down in EPC2 cells displays a low level of relation to other clusters due to the unique types of

EMT induction. It appears likely that on the one hand these GES form a cluster due to the lack of relationship to the other clusters. On the other hand, it might also suggest a relation of their types of EMT initiation as well.

We found a variety of well-known markers of EMT upregulated in our EMT-core gene list such as CDH2, CDH11, COL1A1, COL3A1, FBLN5, FN1, HAS2, LOX, MMP2, PLAT, SERPINE1, VIM, WNT5A and ZEB1 [15,30,31]. Furthermore, we detected downregulated genes reported to be reduced in EMT such as ANK3, CDH1, CXADR, PRSS8 and SYK [15,32-34], several downregulated epithelial cell markers such as EPCAM, JUP, KRT15, KRT17, OCLN, PKP2 and PPL [5,15] and a number of downregulated tumor suppressors such as KLK10, MTUS1, OAS1 and SERPINB1 [35-38]. Together, these genes provide a solid verification of our EMT-core gene list. Besides those genes confirming the integrity of our gene list, however, genes with unknown functions as well as an unknown or unclear relation to cancer and/or EMT emerged which are novel candidates for further investigation. Upregulated genes include MAP1B, NID2, PTX3, SPOCK1, SULF1, TAGLN and TMEM158 while downregulated genes comprised ABLIM1, LAD1, FAM169A, FXYD3, SLC7A5, SLPI, TMEM30B and TPD52L1. Two meta-analyses of EMT in breast cancer considering different cell lines or types of EMT induction have been reported. These have identified EMT-core gene lists with 200 and 251 genes [13,39], however, overlapping with approximately 10% only. Our EMT-core list containing 130 genes shows a poor overlap of 7% with the list of Choi *et al.* [39] but an overlap of 55% with Taube *et al.* [13]. Both the lists by Choi *et al.* and Taube *et al.* contain unmapped identifiers (IDs) such as array IDs, expressed sequence tags and locus IDs. We used consistently enriched pathway analysis to further investigate these gene lists. Notably, our EMT-core list displayed more enriched KEGG pathways and GO terms than the gene lists of Choi *et al.* and Taube *et al.* (Table 3 and 4). Upon reducing the stringency of analysis to two genes within an enriched category, the enrichment for the list of Choi *et al.* did not improve whereas nearly all KEGG pathways and GO terms enriched in our EMT-core list could be observed in the list of Taube *et al.* (data not shown, Table 4).

The EMT-core list contains several genes with unknown functions and relations to cancer and/or EMT. We were able to show that FXYD3 and PTX3 expression is associated with poor overall patient survival in SCC patients and LAD1, SLC7A5, SLPI, NID2, SPOCK1 and SULF1 correlated significantly with impaired pCR in breast cancer patients. FXYD3 has been shown to be involved in tumor cell proliferation and to be downregulated by TGF-β signaling [40,41]. PTX3 has been reported to be a lung cancer biomarker [42]. NID2 has been shown to be elevated during phorbol 12-myristate 13-acetate-induced invasion of several human tumor cell lines and as a potential tumor biomarker [43,44]. SPOCK1 has been reported to be involved in neuronal attachment and matrix metalloproteinase activation [45,46]. SULF1 has been shown to

be a potential biomarker for gastric cancer which can be induced by TGF-β1 [47,48]. LAD1 is an adaptor protein involved in ERK5 and JNK pathways [49]. SLPI has been reported to act anti-tumorigenic for certain tumors as well as to promote migration and invasion in others [50-52]. Hence, these genes seem to be promising candidates for further investigation. Taken together, we propose that the EMT-core list of 130 genes is highly relevant for EMT and the cluster analysis represents a useful overview on the relationships of currently available GES of EMT.

3.5 Materials and Methods

3.5.1 Data collection and annotation

Processed microarray data were downloaded from the websites of GEO (available: http://www.ncbi.nlm.nih.gov/geo/) and AE (available: http://www.ebi.ac.uk/arrayexpress/) by using "EMT" as keyword for published GES until February 2012. The downloaded GES were annotated to retrieve official gene symbols, EntrezID and gene names using BioConductor 2.9 (available: http://www.bioconductor.org/; accessed: 2012 Jan 02) [53] and the online tool NetAffx (available: http://www.affymetrix.com/analysis/index.affx; accessed: 2012 June 25). BioConductor was used within the R environment [54]. Annotated data was imported to MS-Excel 2010 and log2 transformed. Subsequently, fold changes and p-values using two-sided Student's *t*-test were calculated. Significantly up- and downregulated genes were selected and separated from each other when showing a fold change greater than 2 or below 0.5 and a p-value below 0.05. Upregulated genes were ordered from highest to lowest fold change. Vice versa, downregulated genes were arranged from lowest to highest fold change. Duplicates were removed afterwards. Gene symbols have been used for further analysis and will be referred to as genes.

3.5.2 Cluster analysis

The up- and downregulated genes from each study were summarized, ordered and duplicates were removed to obtain a list of all uniquely reported genes across all studies. Upregulated genes were labeled with 1 and downregulated genes were labeled with -1. Genes that were not significantly deregulated within a GES and genes which were found to be both up- and downregulated within a study were labeled with 0. The distribution of the observed number of up- and downregulated genes was tested against a binomial distribution with parameter p=11.78% by means of a chi-squared test. We calculated the possibilities of drawing each cutoff option for cluster analysis (>1, >2, >3, and so forth) by chance with the binomial distribution function provided by R (probability = 11.78%). The possibilities to draw each cutoff option by

chance were compared to preliminary cluster analyses of each cutoff option in order to determine a suitable cutoff. The clustering was performed in BioConductor 2.9 embedded in R 2.14.1 (64 bit) with the packages gdata [55], gplots [56] and heatmap.plus [57] using hierarchical heatmap clustering with Manhattan distance function.

3.5.3 Consistently enrichment of KEGG pathways and GO terms

The gene lists were analyzed using five different bioinformatic enrichment tools. A comprehensive overview of the used tools and their characteristics is shown in Table S4. The tools FatiGO and GeneCodis were used on the Babelomics 4 platform [58], which provided access to both programs at once. The selection criteria for significantly enriched pathways were a p-value or FDR below 0.05 and a minimum of 5 genes of the input list within an enriched category. Furthermore, consistently enriched GO terms and KEGG pathways were identified in at least 4 of 5 programs in both the EMT-core gene list and the 365 gene list. Enrichment ratios (number of observed genes divided by the number of expected genes for a GO or KEGG category) have been obtained by WebGestalt, or alternatively, have been calculated as described by Zhang *et al.* with the data from FatiGO [59].

3.5.4 Correlation of the EMT-core list with clinical data

Microarray and clinical data for patients with squamous cell lung carcinomas (n=130) reported by Raponi *et al.* [17] with the accession GDS2373 were downloaded from GEO. Microarray and clinical data for breast cancer patients (n=133) reported by Hess *et al.* [18] were downloaded from the MD Anderson Cancer Center website (available: http://bioinformatics.mdanderson.org/pubdata.html; accessed 2012 Sep 07). Patients were divided into high and low expressing groups for selected genes within the EMT-core list. The p-values were computed using two-sided Student's *t*-test. Survival analysis for the data by Raponi *et al.* was performed with the chi-squared test of equality using the survival package in R [60]. P-values below 0.05 were considered significant.

3.6　References

1. Hay ED (1995) An overview of epithelio-mesenchymal transformation. Acta Anat (Basel) 154: 8-20.
2. Hay ED (2005) The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it. Dev Dyn 233: 706-720.
3. van Zijl F, Krupitza G, Mikulits W (2011) Initial steps of metastasis: cell invasion and endothelial transmigration. Mutat Res 728: 23-34.
4. Friedl P, Alexander S (2011) Cancer invasion and the microenvironment: plasticity and reciprocity. Cell 147: 992-1009.
5. Kalluri R, Weinberg RA (2009) The basics of epithelial-mesenchymal transition. J Clin Invest 119: 1420-1428.
6. Thiery JP, Acloque H, Huang RY, Nieto MA (2009) Epithelial-mesenchymal transitions in development and disease. Cell 139: 871-890.
7. Loboda A, Nebozhyn MV, Watters JW, Buser CA, Shaw PM, *et al*. (2011) EMT is the dominant program in human colon cancer. BMC Med Genomics 4: 9.
8. Tobin NP, Sims AH, Lundgren KL, Lehn S, Landberg G (2011) Cyclin D1, Id1 and EMT in breast cancer. BMC Cancer 11: 417.
9. Huang da W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1-13.
10. Rhee SY, Wood V, Dolinski K, Draghici S (2008) Use and misuse of the gene ontology annotations. Nat Rev Genet 9: 509-515.
11. Lascorz J, Chen B, Hemminki K, Forsti A (2011) Consensus pathways implicated in prognosis of colorectal cancer identified through systematic enrichment analysis of gene expression profiling studies. PLoS One 6: e18867.
12. Takahashi E, Nagano O, Ishimoto T, Yae T, Suzuki Y, *et al*. (2010) Tumor necrosis factor-alpha regulates transforming growth factor-beta-dependent epithelial-mesenchymal transition by promoting hyaluronan-CD44-moesin interaction. J Biol Chem 285: 4060-4073.
13. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, *et al*. (2010) Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci U S A 107: 15449-15454.
14. Yilmaz M, Christofori G (2010) Mechanisms of motility in metastasizing cells. Mol Cancer Res 8: 629-642.
15. Zeisberg M, Neilson EG (2009) Biomarkers for epithelial-mesenchymal transitions. J Clin Invest 119: 1429-1437.
16. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646-674.
17. Raponi M, Zhang Y, Yu J, Chen G, Lee G, *et al*. (2006) Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer Res 66: 7466-7472.
18. Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, *et al*. (2006) Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. J Clin Oncol 24: 4236-4244.
19. Hwang WL, Yang MH, Tsai ML, Lan HY, Su SH, *et al*. (2011) SNAIL regulates interleukin-8 expression, stem cell-like activity, and tumorigenicity of human colorectal carcinoma cells. Gastroenterology 141: 279-291, 291 e271-275.
20. Micalizzi DS, Christensen KL, Jedlicka P, Coletta RD, Baron AE, *et al*. (2009) The Six1 homeoprotein induces human mammary carcinoma cells to undergo epithelial-mesenchymal transition and metastasis in mice through increasing TGF-beta signaling. J Clin Invest 119: 2678-2690.
21. Baniwal SK, Khalid O, Gabet Y, Shah RR, Purcell DJ, *et al*. (2010) Runx2 transcriptome of prostate cancer cells: insights into invasiveness and bone metastasis. Mol Cancer 9: 258.
22. Micalizzi DS, Wang CA, Farabaugh SM, Schiemann WP, Ford HL (2010) Homeoprotein Six1 increases TGF-beta type I receptor and converts TGF-beta signaling from suppressive to supportive for tumor growth. Cancer Res 70: 10371-10380.
23. Farabaugh SM, Micalizzi DS, Jedlicka P, Zhao R, Ford HL (2012) Eya2 is required to mediate the pro-metastatic functions of Six1 via the induction of TGF-beta signaling, epithelial-mesenchymal transition, and cancer stem cell properties. Oncogene 31: 552-562.
24. Lee KS, Hong SH, Bae SC (2002) Both the Smad and p38 MAPK pathways play a crucial role in Runx2 expression following induction by transforming growth factor-beta and bone morphogenetic protein. Oncogene 21: 7156-7163.
25. Chimge NO, Baniwal SK, Little GH, Chen YB, Kahn M, *et al*. (2011) Regulation of breast cancer metastasis by Runx2 and estrogen signaling: the role of SNAI2. Breast Cancer Res 13: R127.

26. van Zijl F, Mall S, Machat G, Pirker C, Zeillinger R, *et al*. (2011) A human model of epithelial to mesenchymal transition to monitor drug efficacy in hepatocellular carcinoma progression. Mol Cancer Ther 10: 850-860.

27. Ip YT, Park RE, Kosman D, Yazdanbakhsh K, Levine M (1992) dorsal-twist interactions establish snail expression in the presumptive mesoderm of the Drosophila embryo. Genes Dev 6: 1518-1530.

28. Ke XS, Li WC, Hovland R, Qu Y, Liu RH, *et al*. (2011) Reprogramming of cell junction modules during stepwise epithelial to mesenchymal transition and accumulation of malignant features in vitro in a prostate cell model. Exp Cell Res 317: 234-247.

29. Ohashi S, Natsuizaka M, Naganuma S, Kagawa S, Kimura S, *et al*. (2011) A NOTCH3-mediated squamous cell differentiation program limits expansion of EMT-competent cells that express the ZEB transcription factors. Cancer Res 71: 6836-6847.

30. Zoltan-Jones A, Huang L, Ghatak S, Toole BP (2003) Elevated hyaluronan production induces mesenchymal and transformed properties in epithelial cells. J Biol Chem 278: 45801-45810.

31. Ren D, Minami Y, Nishita M (2011) Critical role of Wnt5a-Ror2 signaling in motility and invasiveness of carcinoma cells following Snail-mediated epithelial-mesenchymal transition. Genes Cells 16: 304-315.

32. Kumar S, Park SH, Cieply B, Schupp J, Killiam E, *et al*. (2011) A pathway for the control of anoikis sensitivity by E-cadherin and epithelial-to-mesenchymal transition. Mol Cell Biol 31: 4036-4051.

33. Chen LM, Verity NJ, Chai KX (2009) Loss of prostasin (PRSS8) in human bladder transitional cell carcinoma cell lines is associated with epithelial-mesenchymal transition (EMT). BMC Cancer 9: 377.

34. Sung YM, Xu X, Sun J, Mueller D, Sentissi K, *et al*. (2009) Tumor suppressor function of Syk in human MCF10A in vitro and normal mouse mammary epithelium in vivo. PLoS One 4: e7445.

35. Talieri M, Alexopoulou DK, Scorilas A, Kypraios D, Arnogiannaki N, *et al*. (2011) Expression analysis and clinical evaluation of kallikrein-related peptidase 10 (KLK10) in colorectal cancer. Tumour Biol 32: 737-744.

36. Seibold S, Rudroff C, Weber M, Galle J, Wanner C, *et al*. (2003) Identification of a new tumor suppressor gene located at chromosome 8p21.3-22. FASEB J 17: 1180-1182.

37. Mandal S, Abebe F, Chaudhary J (2011) 2'-5' oligoadenylate synthetase 1 polymorphism is associated with prostate cancer. Cancer 117: 5509-5518.

38. Chou RH, Wen HC, Liang WG, Lin SC, Yuan HW, *et al*. (2012) Suppression of the invasion and migration of cancer cells by SERPINB family genes and their derived peptides. Oncol Rep 27: 238-245.

39. Choi YL, Bocanegra M, Kwon MJ, Shin YK, Nam SJ, *et al*. (2010) LYN is a mediator of epithelial-mesenchymal transition and a target of dasatinib in breast cancer. Cancer Res 70: 2296-2306.

40. Yamamoto H, Okumura K, Toshima S, Mukaisho K, Sugihara H, *et al*. (2009) FXYD3 protein involved in tumor cell proliferation is overproduced in human breast cancer tissues. Biol Pharm Bull 32: 1148-1154.

41. Yamamoto H, Mukaisho K, Sugihara H, Hattori T, Asano S (2011) Down-regulation of FXYD3 is induced by transforming growth factor-beta signaling via ZEB1/deltaEF1 in human mammary epithelial cells. Biol Pharm Bull 34: 324-329.

42. Diamandis EP, Goodglick L, Planque C, Thornquist MD (2011) Pentraxin-3 is a novel biomarker of lung carcinoma. Clin Cancer Res 17: 2395-2399.

43. Rouleau C, Roy A, St Martin T, Dufault MR, Boutin P, *et al*. (2006) Protein tyrosine phosphatase PRL-3 in malignant cells and endothelial cells: expression and function. Mol Cancer Ther 5: 219-229.

44. Kuk C, Gunawardana CG, Soosaipillai A, Kobayashi H, Li L, *et al*. (2010) Nidogen-2: a new serum biomarker for ovarian cancer. Clin Biochem 43: 355-361.

45. Marr HS, Edgell CJ (2003) Testican-1 inhibits attachment of Neuro-2a cells. Matrix Biol 22: 259-266.

46. Nakada M, Yamada A, Takino T, Miyamori H, Takahashi T, *et al*. (2001) Suppression of membrane-type 1 matrix metalloproteinase (MMP)-mediated MMP-2 activation and tumor invasion by testican 3 and its splicing variant gene product, N-Tes. Cancer Res 61: 8896-8902.

47. Junnila S, Kokkola A, Mizuguchi T, Hirata K, Karjalainen-Lindsberg ML, *et al*. (2010) Gene expression analysis identifies over-expression of CXCL1, SPARC, SPP1, and SULF1 in gastric cancer. Genes Chromosomes Cancer 49: 28-39.

48. Yue X, Li X, Nguyen HT, Chin DR, Sullivan DE, *et al*. (2008) Transforming growth factor-beta1 induces heparan sulfate 6-O-endosulfatase 1 expression in vitro and in vivo. J Biol Chem 283: 20397-20407.

49. Sun W, Wei X, Kesavan K, Garrington TP, Fan R, *et al*. (2003) MEK kinase 2 and the adaptor protein Lad regulate extracellular signal-regulated kinase 5 activation by epidermal growth factor via Src. Mol Cell Biol 23: 2298-2308.

50. Wen J, Nikitakis NG, Chaisuparat R, Greenwell-Wild T, Gliozzi M, *et al*. (2011) Secretory leukocyte protease inhibitor (SLPI) expression and tumor invasion in oral squamous cell carcinoma. Am J Pathol 178: 2866-2878.

51. Amiano NO, Costa MJ, Reiteri RM, Payes C, Guerrieri D, *et al.* (2012) Antitumor effect of SLPI on mammary but not colon tumor growth. J Cell Physiol.

52. Choi BD, Jeong SJ, Wang G, Park JJ, Lim DS, *et al.* (2011) Secretory leukocyte protease inhibitor is associated with MMP-2 and MMP-9 to promote migration and invasion in SNU638 gastric cancer cells. Int J Mol Med 28: 527-534.

53. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.

54. R Development Core Team (2011) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

55. Warnes GR, Bolker B, Gorjanc G, Grothendieck G, Korosec A, *et al.* (2011) gdata: Various R programming tools for data manipulation. R package version 2.8.2. CRAN website. Available: http://CRAN.R-project.org/package=gdata. Accessed 2012 June 25.

56. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, *et al.* gplots: Various R programming tools for plotting data. R package version 2.10.1.; 2011. CRAN website. Available: http://CRAN.R-project.org/package=gplots. Accessed 2012 June 25.

57. Day A (2007) heatmap.plus: Heatmap with more sensible behavior. R package version 1.3. CRAN website. Available: http://CRAN.R-project.org/package=heatmap.plus. Accessed 2012 June 25.

58. Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, *et al.* (2010) Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. Nucleic Acids Res 38: W210-213.

59. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33: W741-748.

60. Thernau T (2012) A Package for Survival Analysis in S. CRAN website. Available: http://CRAN.R-project.org/package=survival. Accessed 2012 June 25.

61. Andarawewa KL, Erickson AC, Chou WS, Costes SV, Gascard P, *et al.* (2007) Ionizing radiation predisposes nonmalignant human mammary epithelial cells to undergo transforming growth factor beta induced epithelial to mesenchymal transition. Cancer Res 67: 8662-8670.

62. Tay PN, Tan P, Lan Y, Leung CH, Laban M, *et al.* (2010) Palladin, an actin-associated protein, is required for adherens junction formation and intercellular adhesion in HCT116 colorectal cancer cells. Int J Oncol 37: 909-926.

63. Drake JM, Strohbehn G, Bair TB, Moreland JG, Henry MD (2009) ZEB1 enhances transendothelial migration and represses the epithelial phenotype of prostate cancer cells. Mol Biol Cell 20: 2207-2217.

64. Sartor MA, Mahavisno V, Keshamouni VG, Cavalcoli J, Wright Z, *et al.* (2010) ConceptGen: a gene set enrichment and gene set relation mapping tool. Bioinformatics 26: 456-463.

65. Papageorgis P, Lambert AW, Ozturk S, Gao F, Pan H, *et al.* (2010) Smad signaling is required to maintain epigenetic silencing during breast cancer progression. Cancer Res 70: 968-978.

66. Hills CE, Willars GB, Brunskill NJ (2010) Proinsulin C-peptide antagonizes the profibrotic effects of TGF-beta1 via up-regulation of retinoic acid and HGF-related signaling pathways. Mol Endocrinol 24: 822-831.

67. Leshem O, Madar S, Kogan-Sakin I, Kamer I, Goldstein I, *et al.* (2011) TMPRSS2/ERG promotes epithelial to mesenchymal transition through the ZEB1/ZEB2 axis in a prostate cancer model. PLoS One 6: e21650.

68. Maupin KA, Sinha A, Eugster E, Miller J, Ross J, *et al.* (2010) Glycogene expression alterations associated with pancreatic cancer epithelial-mesenchymal transition in complementary model systems. PLoS One 5: e13002.

69. Hesling C, Fattet L, Teyre G, Jury D, Gonzalo P, *et al.* (2011) Antagonistic regulation of EMT by TIF1gamma and Smad4 in mammary epithelial cells. EMBO Rep 12: 665-672.

70. Wang L, Mezencev R, Bowen NJ, Matyunina LV, McDonald JF (2011) Isolation and characterization of stem-like cells from a human ovarian cancer cell line. Mol Cell Biochem 363: 257-268.

3.7 Figures
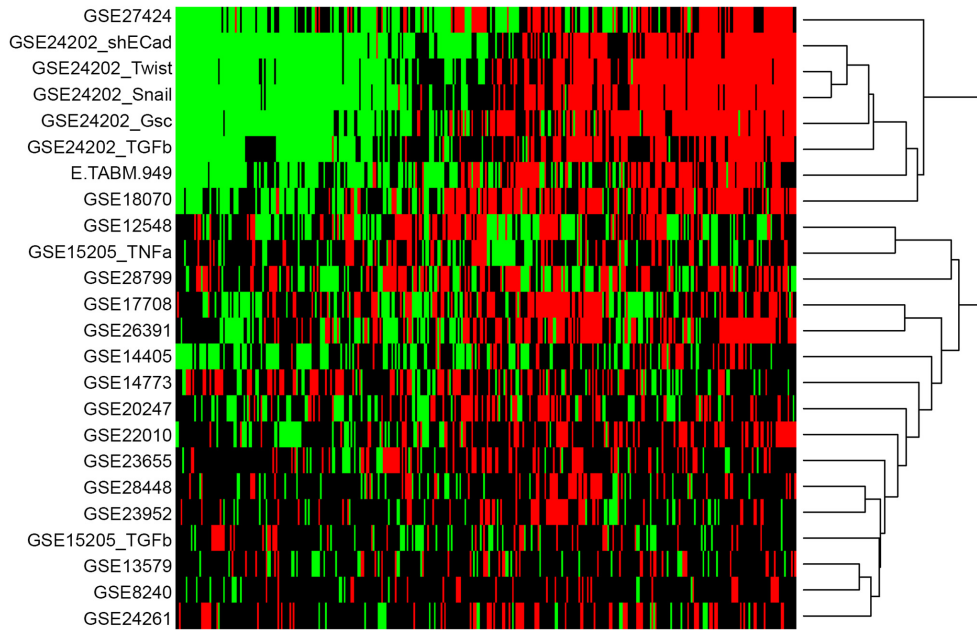
Figure 1: Cluster analysis of genes shared between at least 10 GES datasets shows distinguishable and significant clusters. Genes shared between at least 10 out of 24 datasets were used for Manhattan hierarchical clustering. The type of regulation within a particular study was visualized via heatmap. Columns: genes shared between at least 10 datasets (n=365); rows: analyzed GES (24 datasets in total); green: downregulated genes; red: upregulated genes; black: genes not regulated. GSE: Gene expression omnibus (GEO) series record; E.TABM: ArrayExpress (AE) series record; TGF, transforming growth factor; TNF, tumor necrosis factor.
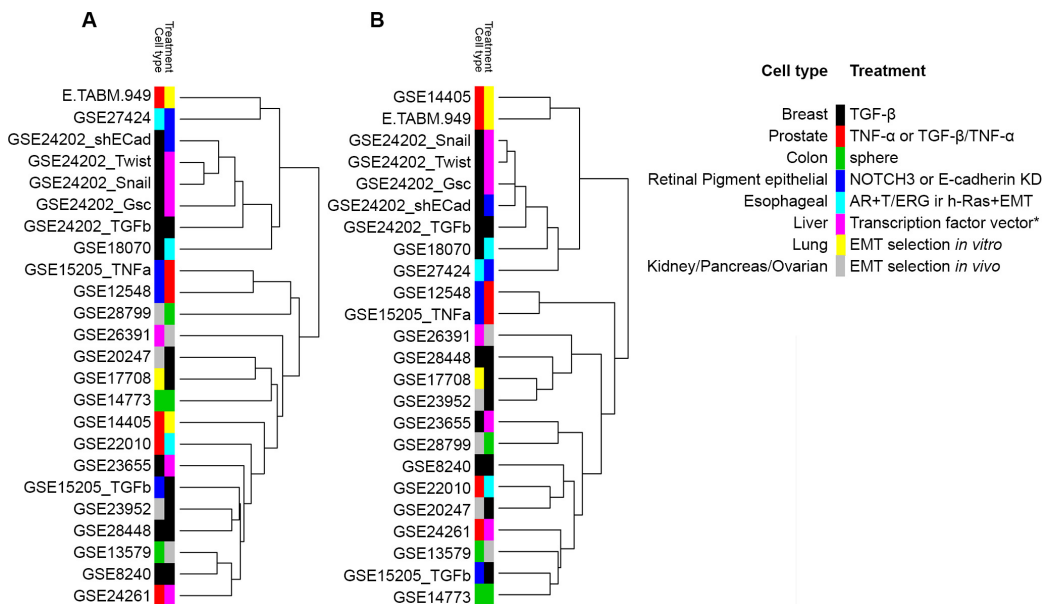
Figure 2: Gene expression studies cluster according to the mode of EMT initiation rather than to cell type. The cell type and treatment modality of EMT was annotated and revealed clustering according to the mode of EMT induction. The clustering persisted when genes shared between at least 14 GES datasets were used for the analysis. **(A)** Hierarchical clustering of 365 genes shared between at least 10 datasets. **(B)** Hierarchical clustering of 41 genes shared between at least 14 datasets. The legend indicates cell type and treatment modality (right panel). *, Transcription factor vectors: Runx2, Six1, Snail, Twist and Goosecoid. GSE: Gene expression omnibus (GEO) series record; E.TABM: ArrayExpress (AE) series record; TGF, transforming growth factor; TNF, tumor necrosis factor.
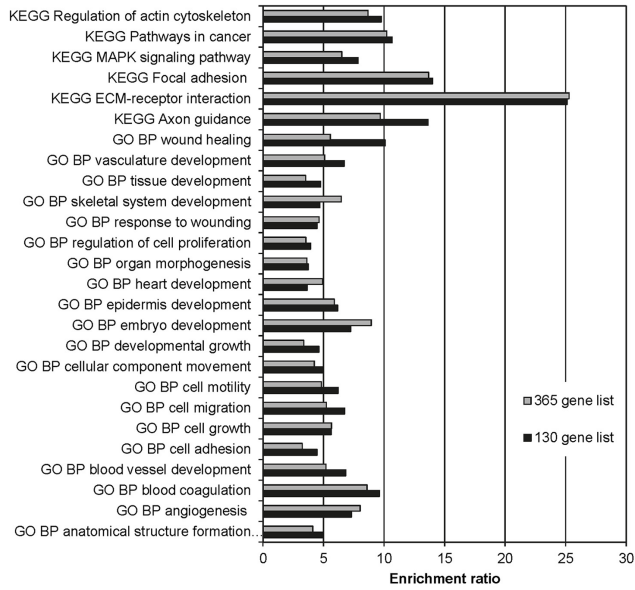
Figure 3: The 130 genes EMT-core list and the 365 genes list exhibit comparable enrichment ratios of GO biological processes and KEGG pathways. The enrichment ratio is the number of observed genes divided by the number of expected genes for a given term or pathway. Enrichment ratios were obtained from WebGestalt or calculated with data from FatiGO. GO, gene ontology; BP, biological process; KEGG, Kyoto encyclopedia of genes and genomes.

3.8 Supporting information
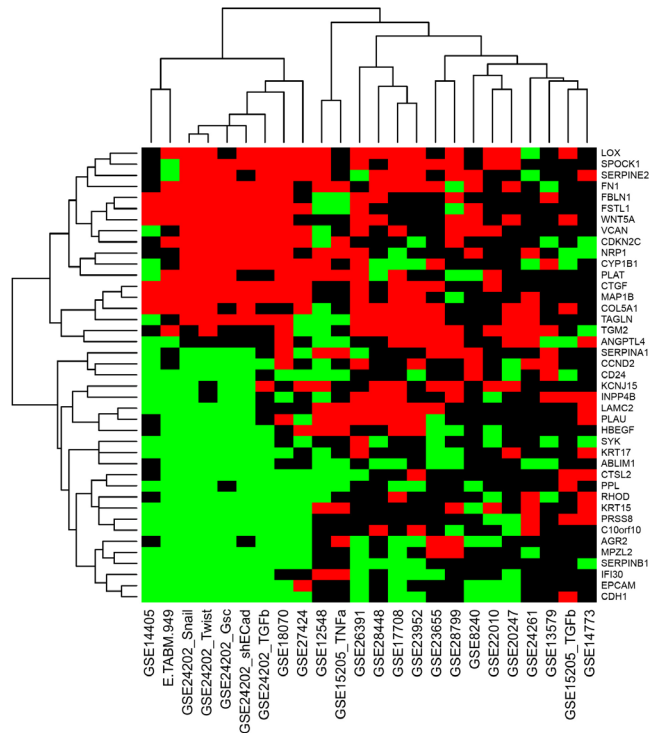
Figure S1: Cluster analysis of genes shared between at least 14 GES datasets shows persistent and distinct clusters. Genes shared between at least 14 out of 24 datasets were used for Manhattan hierarchical clustering. The type of regulation within a particular study was visualized via heatmap. Columns: genes shared between at least 14 GES datasets (n=41); rows: analyzed GES (24 datasets in total); green: downregulated genes; red: upregulated genes; black: genes not regulated. GSE: Gene expression omnibus (GEO) series record; E.TABM: ArrayExpress (AE) series record; TGF, transforming growth factor; TNF, tumor necrosis factor.
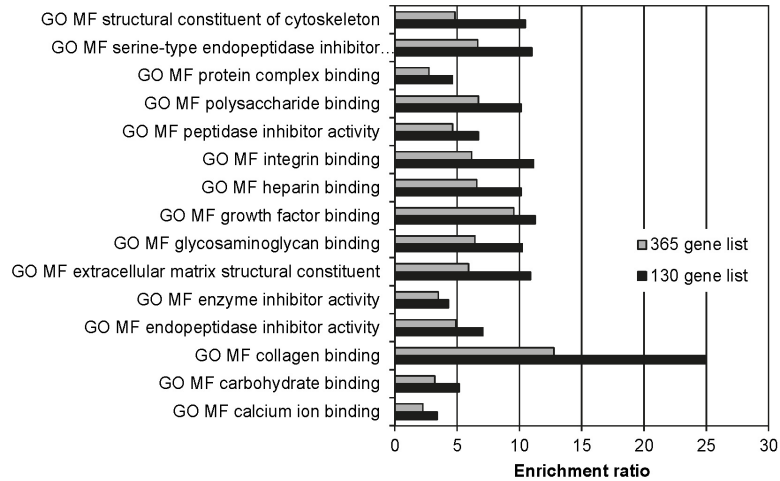
Figure S2: The 130 genes EMT-core list and the 365 genes list exhibit comparable enrichment ratios of GO molecular functions. The enrichment ratio is the number of observed genes divided by the number of expected genes for a given term or pathway. Enrichment ratios were obtained from WebGestalt or calculated with data from FatiGO. GO, gene ontology; BP, biological process; MF, molecular function; KEGG, Kyoto encyclopedia of genes and genomes.
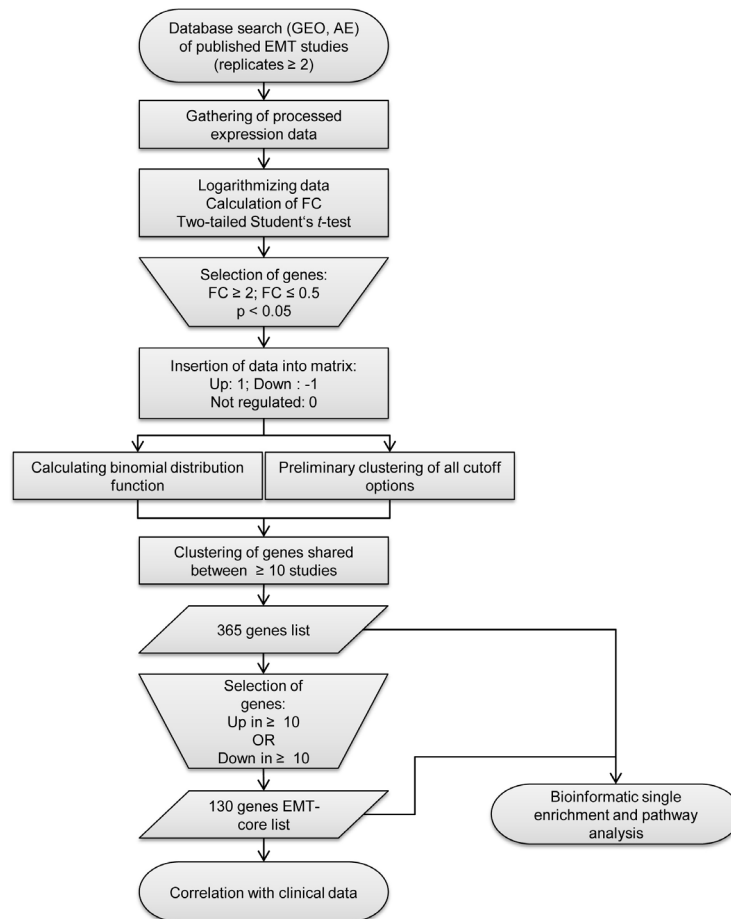
Figure S3: Flow chart depicting the generation of the EMT-core gene list. Overview of the meta-analysis, the generation of the 365 genes list, the EMT-core list as well as the single enrichment analysis and the correlation with clinical data. GEO, Gene expression Omnibus; AE, Array Express; EMT, epithelial to mesenchymal transition; FC, Fold change; Up, upregulated; Down, downregulated.

Table S1: Matrix containing significantly up- and downregulated genes across the analyzed GES datasets.

Available at PLOS ONE (16.05.2013):

http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0051136.s004

Table S2: List of 365 genes significantly regulated in at least 10 GES datasets.

Available at PLOS ONE (16.05.2013):

http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0051136.s005

Table S3: EMT-core gene list of 130 up- or downregulated genes shared between at least 10 GES datasets.

Symbol	EntrezID	Name	Other categories
Upregulated genes			
a) Cell adhesion and migration			
ADAM12	8038	ATP-binding cassette, sub-family A (ABC1), member 1	b
CDH11	1009	cadherin 11, type 2, OB-cadherin (osteoblast)	b
CDH2	1000	cadherin 2, type 1, N-cadherin (neuronal)	b
COL1A1	1277	collagen, type I, alpha 1	b
COL3A1	1281	collagen, type III, alpha 1	b
COL5A1	1289	collagen, type V, alpha 1	b
COL6A1	1291	collagen, type VI, alpha 1	b
COL6A3	1293	collagen, type VI, alpha 3	b
CTGF	1490	connective tissue growth factor	b, c, d, e
CYP1B1	1545	cytochrome P450, family 1, subfamily B, polypeptide 1	b, d
DLC1	10395	deleted in liver cancer 1	b
FBLN1	2192	fibulin 1	b
FBLN5	10516	fibulin5	b
FGF2	2247	fibroblast growth factor 2 (basic)	b, c, d, e
FGFR1	2260	fibroblast growth factor receptor 1	b, c, e
FN1	2335	fibronectin 1	b, c, e
HAS2	3037	hyaluronan synthase 2	b
LAMC2	3918	laminin, gamma 2	
LUM	4060	lumican	b
MMP2	4313	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	b, c, e
MYL9	10398	myosin, light chain 9, regulatory	b
NID2	22795	nidogen 2 (osteonidogen)	
NR2F1	7025	nuclear receptor subfamily 2, group F, member 1	b, d
NRP1	8829	neuropilin 1	b, c
PLAT	5327	plasminogen activator, tissue	b, c
PPAP2B	8613	phosphatidic acid phosphatase type 2B	b, c
PRKCA	5578	protein kinase C, alpha	b, c, d, e
RECK	7434	reversion-inducing-cysteine-rich protein with kazal motifs	b, c
SERPINE1	5054	serpin peptidase inhibitor, clade E, member 1	b, c, e

SERPINE2	5270	serpin peptidase inhibitor, clade E, member 2	b, c
SPOCK1	6695	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1	b
TGM2	7052	transglutaminase	b, c, e
TNFAIP6	7130	tumor necrosis factor, alpha-induced protein 6	b, c
TPM1	7168	tropomyosin 1 (alpha)	b, c
VCAN	1462	versican	b, c
WNT5A	7474	wingless-type MMTV integration site family, member 5A	b, c, e

b) Development / cell differentiation and proliferation

CDKN2C	1031	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	d, e
EMP3	2014	epithelial membrane protein 3	d, e
FBN1	2200	fibrillin 1	
IGFBP3	3486	insulin-like growth factor binding protein 3	d, e
IL1R1	3554	interleukin 1 receptor, type I	
LTBP1	4052	latent transforming growth factor beta binding protein 1	
MME	4311	membrane metallo-endopeptidase	d
PMP22	5376	peripheral myelin protein 22	
PTGER2	5732	prostaglandin E receptor 2 (subtype EP2), 53kDa	
PTX3	5806	pentraxin 3, long	c, d
SRGN	5552	serglycin	c, e
SULF	23213	sulfatase 1	c
SYNE1	23345	spectrin repeat containing, nuclear envelope 1	e
TAGLN	6876	transgelin	
TUBA1A	7846	tubulin, alpha 1a	
VIM	7431	vimentin	e
ZEB1	6935	zinc finger E-box binding homeobox 1	

c) Angiogenesis and wound healing

DCN	1634	decorin	a, b, e
LOX	4015	lysyl oxidase	a, b, d
TFPI	7035	tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)	

d) Metabolism

ABCA1	19	ATP-binding cassette, sub-family A (ABC1), member 1	b
GALNT10	55568	GalNAc-T10	
SLC22A4	6583	solute carrier family 22 (organic cation/ergothioneine transporter), member 4	

Others or unclassified

C5orf13	9315	chromosome 5 open reading frame 13	
CDK14	5218	cyclin-dependent kinase 14	
EML1	2009	echinoderm microtubule associated protein like 1	
FSTL1	11167	follistatin-like 1	
LTBP2	4053	latent transforming growth factor beta binding protein 2	
MAP1B	4131	microtubule-associated protein 1B	
RGS4	5999	regulator of G-protein signaling 4	

| SYT11 | 23208 | synaptotagmin XI | |
| TMEM158 | 25907 | transmembrane protein 158 (gene/pseudogene) | |

Downregulated genes

a) Cell adhesion and migration

CD24	100133941	CD24 molecule	b, e
CDH1	999	cadherin 1, type 1, E-cadherin (epithelial)	b, e
CXADR	1525	coxsackie virus and adenovirus receptor	b, c
CXCL16	58191	chemokine (C-X-C motif) ligand 16	b
DSG3	1830	desmoglein 3	e
ELF3	1999	E74-like factor 3 (ets domain transcription factor, epithelial-specific)	b, c, e
EPCAM	4072	epithelial cell adhesion molecule	b
EPHA1	2041	EPH receptor A1	b
JUP	3728	junction plakoglobin	b
MPZL2	10205	myelin protein zero-like 2	b
OVOL2	58495	ovo-like 2 (Drosophila)	b
PLXNB1	5364	plexin B1	b
S100P	6286	S100 calcium binding protein P	
SLC7A5	8140	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5	b, c
SYK	6850	spleen tyrosine kinase	b, c, d

b) Development / cell differentiation and proliferation

ABLIM1	3983	actin binding LIM protein 1	a
ADRB2	154	adrenergic, beta-2-, receptor, surface	c, e
ALDH1A3	220	aldehyde dehydrogenase 1 family, member A3	e
ANK3	288	ankyrin 3, node of Ranvier (ankyrin G)	
BIK	638	BCL2-interacting killer (apoptosis-inducing)	e
CA2	760	carbonic anhydrase II	
CTSL2	1515	cathepsin L2	d, e
FGFR2	2263	fibroblast growth factor receptor 2	
FGFR3	2261	fibroblast growth factor receptor 3	
FST	10468	follistatin	
GJB3	2707	gap junction protein, beta 3, 31kDa	
IFI30	10437	interferon, gamma-inducible protein 30	
IL18	3606	interleukin 18 (interferon-gamma-inducing factor)	a, c, e
KLK7	5650	kallikrein-related peptidase 7	
KRT15	3866	keratin 15	
KRT17	3872	keratin 17	
LSR	51599	lipolysis stimulated lipoprotein receptor	d
MAP7	9053	microtubule-associated protein 7	d
MBP	4155	myelin basic protein	
OCLN	4950	occludin	e
PKP2	5318	plakophilin 2	a
PPL	5493	periplakin	

PRSS8	5652	protease, serine, 8	d
RAPGEF5	9771	Rap guanine nucleotide exchange factor (GEF) 5	
SPINT1	6692	serine peptidase inhibitor, Kunitz type 1	

d) Metabolism

GPX3	2878	glutathione peroxidase 3 (plasma)	b
SLC27A2	11001	solute carrier family 27 (fatty acid transporter), member 2	
SMPDL3B	27293	sphingomyelin phosphodiesterase, acid-like 3B	
SORL1	6653	sortilin-related receptor, L(DLR class) A repeats containing	a
ST6GALNAC2	10610	ST6 -N-acetylgalactosaminide alpha-2,6-sialyltransferase 2	

Others or unclassified

AGR2	10551	anterior gradient homolog 2 (Xenopus laevis)	
C10orf10	11067	chromosome 10 open reading frame 10	
CDS1	1040	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 1	
FAM169A	26049	family with sequence similarity 169, member A	
FXYD3	5349	FXYD domain containing ion transport regulator 3	
KLK10	5655	kallikrein-related peptidase 10	
LAD1	3898	ladinin 1	
MTUS1	57509	microtubule associated tumor suppressor 1	
PLS1	5357	plastin 1	
PRRG4	79056	proline rich Gla (G-carboxyglutamic acid) 4 (transmembrane)	
RHOD	29984	ras homolog gene family, member D	
SERPINB1	1992	serpin peptidase inhibitor, clade B (ovalbumin), member 1	
SLPI	6590	secretory leukocyte peptidase inhibitor	
TMEM30B	161291	transmembrane protein 30B	
TPD52L1	7164	tumor protein D52-like 1	e
TSPAN1	10103	tetraspanin 1	
ZHX2	22882	zinc fingers and homeoboxes 2	
ZNF165	7718	zinc finger protein 165	

Table S4: Enrichment tools used in this study and their properties.

Tool	First ref.	Key statistical method	Multiple testing correction method(s)
ConsensusPathDB*	1	Hypergeometric	FDR
FatiGO	2	Fisher's exact	3 methods (B-H, B-Y, permutation)
GeneCodis	3	Hypergeometric	FDR[#], permutation
ToppFun	4	Hypergeometric	Bonferroni, FDR[#]
WebGestalt	5	Hypergeometric	B-H[#], B-Y, Bonferroni, holm, hommel

*, uses gene ontology levels 2-4 only; [#,] indicates the multiple testing method which was used.

References to Table S4: available at PLOS ONE (16.05.2013):

http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0051136.s007

3.9 Tables

Table 1: Gene expression studies of EMT used for meta-analysis.

First author	Acc.	Ref.	Cell type	Cell origin	Treatment modality	Platform	Samples*
Ke	E-TABM-949	[28]	EP156T / EPT2	Prostate	high cell density[a]	Agil WHG 4x44K G4112F	2
Andarawewa	GSE8240	[61]	MCF10A	Breast	TGF-β + irradiation[a]	Affy HTU133A	3
Takahashi[c]	GSE12548/ GSE15205	[12]	ARPE19	Retinal pigment	TGF-β + TNF-α[a]/ TGF-β or TNF-α[a]	Affy U133Plus2	3
Tay	GSE13759	[62]	HCT116/E1	Colon	serial transplantation[b]	Affy U133A	3
Drake	GSE14405	[63]	PC-3 / TEM4-18	Prostate	transendothelial migration[a]	Affy U133Plus2	2
Hwang	GSE14773	[19]	CRC	Colon	spheroid formation[a]	Affy U133Plus2	2
Sartor	GSE17708	[64]	A549	Lung	TGF-β[a]	Affy U133Plus2	3
Papageorgis	GSE18070	[65]	MCF10CA1h	Breast	H-Ras + carcinoma[b]	Affy U133Plus2	3
Hills	GSE20247	[66]	HK2	Kidney	TGF-β + Cpep[a]	Illum HWG-6 v3.0	3
Leshem	GSE22010	[67]	PrEC-hTERT	Prostate	AR + T/ERG[a]	Affy HG 1.0 ST	4
Micalizzi	GSE23655	[20]	MCF7	Breast	Six1 vector[a]	Affy HTU133A	6
Maupin	GSE23952	[68]	Panc-1	Pancreas	TGF-β[a]	Affy U133Plus2	3
Taube[d]	GSE24202	[13]	HMLE	Breast	TGF-β1; Snail1, Twist, Gsc vectors; siRNA against E-Cadherin [a]	Affy HTU133A	3
Baniwal	GSE24261	[21]	PCa C4-2B/Rx2dox	Prostate	Runx2 vector[a]	Illum HR-8 v3.0	4
van Zijl	GSE26391	[26]	3p/3sp	Liver	tumor cell recovery[b]	Affy HG 1.0 ST	2
Ohashi	GSE27424	[29]	EPC2-hTERT	Esophagus	Notch3 knock-down (shRNA)[a]	Affy U133Plus2	3
Hesling	GSE28448	[69]	HMEC-TR	Breast	TGF-β + siRNA against TIFγ[a]	Affy U133Plus2	2
Wang	GSE28799	[70]	OVCAR-3	Ovary	spheroid formation[a]	Affy U133Plus2	3

*, lowest number of samples per class (control or test subject).
a, *in vitro*; b, *in vivo;* c, consists of two studies with three datasets in total; d, consists of five datasets.
Abbreviations: Affy, Affymetrix; Agil, Agilent; AR, androgen receptor; Illum, Illumina; sh, small hairpin; si, small interfering; T/ERG, TMPRSS2/ERG; TGF, transforming growth factor; TNF, tumor necrosis factor.

Table 2: EMT-core list of 130 genes shared between at least 10 GES datasets.

	Upregulated	Downregulated
Cell adhesion and migration	ADAM12, CDH11, CDH2, COL1A1, COL3A1, COL5A1, COL6A1, COL6A3, CTGF, CYP1B1, DLC1, FBLN1, FBLN5, FGF2, FGFR1, FN1, HAS2, LUM, MMP2, MYL9, NID2, NR2F1, NRP1, PLAT, PPAP2B, PRKCA, RECK, SERPINE1, SERPINE2, SPOCK1, TGM2, TNFAIP6, TPM1, VCAN, WNT5A	CD24, CDH1, CXADR, CXCL16, DSG3, ELF3, EPCAM, EPHA, JUP, MPZL2, OVOL2, PLXNB1, S100P, SLC7A5, SYK
Development, cell differentiation and proliferation	CDKN2C, EMP3, FBN1, IGFBP3, IL1R1, LTBP1, MME, PMP22, PTGER2, PTX3, SRGN, SULF1, SYNE1, TAGLN, TUBA1A, VIM, ZEB1	ABLIM1, ADRB2, ALDH1A3, ANK3, BIK CA2, CTSL2, FGFR2, FGFR3, FST, GJB3, IFI30, IL18, KLK7, KRT15, KRT17, LSR, MAP7, MBP, OCLN, PKP2, PPL, PRSS8, RAPGEF5, SPINT1
Angiogenesis and wound healing	DCN, LOX, TFPI	*no gene with a major classification**
Metabolism	ABCA1, GALNT10, SLC22A4	GPX3, SLC27A2, SMPDL3B, SORL1, ST6GALNAC2
Others or unclassified	C5orf13, CDK14, EML1, FSTL1, LTBP2, MAP1B, RGS4, SYT11, TMEM158	AGR2, C10orf10, CDS1, FAM169A, FXYD3, KLK10, LAD1, MTUS1, PLS1, PRRG4, RHOD, SERPINB1, SLPI, TMEM30B, TPD52L1, TSPAN1, ZHX2, ZNF165

Categories have been chosen according to the GO classifications of the enrichment tools. Genes may be present in more than one category.
*see Table S3 for more information.

Table 3: Number of enriched terms and pathways in all lists detected by the enrichment tools.

Tool	130 gene list			365 gene list			GSE13195 core list			GSE24202 core list		
	BP	MF	KEGG	BP	MF	KEGG	BP	MF	KEGG	BP	MF	KEGG
ConsensusPathDB	305	31	9	558	61	31	62	10	6	247	34	8
FatiGO	178	28	9	452	72	36	0	0	2	172	28	10
GeneCodis	34	16	8	155	45	46	59	17	4	240	48	7
ToppFun	241	21	1	610	45	5	0	0	1	127	14	0
WebGestalt	40	28	6	40	40	37	5	4	4	40	30	8

The numbers of enriched terms and pathways found by the particular enrichment tools are displayed. BP, GO biological process; MF, GO molecular function; KEGG, KEGG pathway. GSE13195 core list of Choi *et al*., GSE24202 core list of Taube *et al*. [13,39].

Table 4: Consistently enriched GO terms and KEGG pathways and their occurrence in the analyzed gene lists.

Term ID	Category	Term size*	130 gene list		365 gene list		GSE13915 core list		GSE24202 core list	
			Tools	Genes	Tools	Genes	Tools	Genes	Tools	Genes
GO biological process										
GO:0048646	anatomical structure formation involved in morphogenesis	390	4	24	4	62	0	-	4	22
GO:0001525	angiogenesis	189	4	16	4	38	0	-	4	14
GO:0007596	blood coagulation	182	4	13	4	29	0	-	3	13
GO:0001568	blood vessel development	288	4	25	5	54	0	-	5	20
GO:0007155	cell adhesion	953	5	36	5	76	2	19	5	41
GO:0016049	cell growth	226	4	13	4	34	0	-	4	14
GO:0016477	cell migration	405	5	32	5	67	1	13	5	35
GO:0048870	cell motility	484	4	33	4	69	1	13	5	35
GO:0006928	cellular component movement	666	4	36	5	73	1	16	5	41
GO:0009790	embryo development	619	4	18	4	46	0	-	3	20
GO:0008544	epidermis development	218	5	16	4	32	2	6	5	26
GO:0007507	heart development	230	5	15	4	28	1	6	3	10
GO:0009887	organ morphogenesis	800	5	21	5	54	0	-	5	34
GO:0042127	regulation of cell proliferation	823	4	28	5	81	1	14	5	37
GO:0050793	regulation of developmental process	1005	4	34	4	88	0	-	4	32
GO:0009611	response to wounding	776	5	31	5	85	0	-	4	34
GO:0001501	skeletal system development	394	4	14	4	35	0	-	5	20
GO:0009888	tissue development	808	4	38	4	93	1	12	5	52
GO:0001944	vasculature development	294	4	25	4	56	0	-	5	20
GO:0042060	wound healing	270	4	20	5	50	0	-	3	19
GO molecular function										
GO:0005509	calcium ion binding	1033	4	22	4	55	0	-	4	34
GO:0030246	carbohydrate binding	380	4	15	4	29	1	7	4	14
GO:0005518	collagen binding	40	4	5	5	12	0	-	0	-
GO:0004866	endopeptidase inhibitor activity	179	4	9	4	19	0	-	4	9
GO:0004857	enzyme inhibitor activity	327	4	10	4	26	2	8	4	13
GO:0005201	ECM constituent	105	5	7	5	12	0	-	4	7
GO:0005539	glycosaminoglycan binding	146	4	13	5	24	0	-	5	10
GO:0019838	growth factor binding	127	4	13	4	26	0	-	5	14
GO:0008201	heparin binding	108	4	9	4	17	0	-	3	7
GO:0005178	integrin binding	57	4	6	5	9	0	-	4	7
GO:0030414	peptidase inhibitor activity	192	5	9	5	20	0	-	4	9
GO:0030247	polysaccharide binding	165	4	14	4	27	0	-	5	13
GO:0032403	protein complex binding	199	4	11	4	20	0	-	2	8
GO:0004867	serine-type endopeptidase inhibitor activity	118	4	9	4	14	0	-	3	7
GO:0005200	structural constituent of cytoskeleton	92	5	8	5	10	0	-	5	13
KEGG pathway										
map04360	axon guidance	126	4	6	4	11	0	-	4	6
map04512	ECM-receptor interaction	92	5	7	5	18	0	-	1	5
map04510	focal adhesion	207	4	9	5	23	0	-	3	8
map04010	MAPK signaling pathway	289	3	7	4	15	0	-	0	-
map05200	pathways in cancer	329	4	11	5	28	1	5	2	8
map04810	regulation of actin cytoskeleton	209	4	7	4	16	0	-	2	6

*According to FatiGO category size in genome.

The maximum number of genes from the category present in the input list is displayed. ID, identity; GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes. GSE13195 core list of Choi *et al.*, GSE24202 core list of Taube *et al.* [13,39].

Abstract to the following section

The data gained from the meta-analysis described previously was compared with microarray expression data of the 3p/3sp EMT cell model. One particular gene was slightly below the threshold used during the meta-analysis but was one of the 100 genes displaying a high fold change in 3sp compared to 3p cells. This gene was NRP2, a transmembrane co-receptor which has been described to interact with many pathways involved in cancer and EMT. The aim of the following study was to investigate the role of NRP2 in HCC and EMT *in vitro* and *in vivo*.

4   Target characterization of NRP2

4.1   Materials and Methods

4.1.1 RNA isolation and reverse transcription to cDNA

Cells were grown in 6 cm or 6-well plates (CytoOne, Starlab, Germany) and harvested at a maximum confluency of 80% with RLT buffer (QIAGEN, Germany). RNA isolation was performed using the RNeasy Mini Kit (QIAGEN, Germany). The sample was homogenized by passing 10 times through a 20-gauge needle. 1 volume 70% ethanol was added, the mix was transferred into the provided spin columns and centrifuged for 1 min at 13,000 rpm in a table top centrifuge. Flow-through was discarded, 700 µL buffer RW1 were added to the column and centrifuged again. Flow-through was discarded, 500 µL buffer RPE was added and centrifuged again. This step was repeated once. To remove residual traces of ethanol, the column was transferred into a new waste tube and centrifuged for 2 min at 13,000 rpm. The column was then transferred into a new reaction tube and RNA was eluted by addition of 30 µL RNase-free ddH$_2$O, incubating for 1 min at room temperature and centrifuging for 1 min at 13,000 rpm. RNA was subsequently used for reverse transcription using QuantiTect Reverse Transcription Kit (QIAGEN, Germany). 500 ng RNA were treated with 1 µL gDNA wipeout buffer in a total volume of 7 µL (adjusted with ddH$_2$O) for 2 min at 42°C in order to remove genomic DNA and RNA secondary structures. The solution was transferred on ice, 0.5 µL primer mix, 2 µL RT buffer and 0.5 µL reverse transcriptase were added and incubated at 42°C from 30 min. The reverse transcriptase was then heat-inactivated at 95°C for 3 min. Remaining RNA was stored at -80°C. cDNA was diluted to required concentrations and stored at -20°C for further use in PCR and qPCR.

4.1.2 Polymerase chain reaction and sequencing

Taq PCR was used for primer trials and NRP2 isoform detection. In a tube of PuReTaq Ready-To-Go PCR beads (GE Healthcare, UK), 10 µL of a 1:100 diluted cDNA sample and 15 µL primer pair mixture (4 µM of each primer) were mixed. A proofreading polymerase, Phusion High-Fidelity DNA Polymerase (New England Biolabs, Germany), was used to generate inserts for cloning. PCR was performed in a Primus 96 advanced cycler (Peqlab). Inserts were sequenced after they have been successfully cloned into pWPI as described below by VBC Biotech (Vienna, Austria) using pWPI *Mss*I sequencing and NRP2 sequencing primers. PCR programs are depicted in Table 2 and Table 3. Primers used are described in Table 4.

Table 2: PCR program for NRP2 isoform detection

Step	Function	Temp	Time	
1	Denaturation	95°C	2 min	
2	Denaturation	95°C	45 sec	
3	Annealing	58°C	20 sec	Repeat 3 times
4	Extension	72°C	45 sec	
5	Denaturation	95°C	45 sec	
6	Annealing	56°C	20 sec	Repeat 32 times
7	Extension	72°C	45 sec	
8	Elongation	72°C	10 min	
9	Store	4°C	∞	

Table 3: PCR program and reaction mixture for NRP2 insert generation

Step	Function	Temp	Time	
1	Denaturation	98°C	2 min	
2	Denaturation	98°C	15 sec	
3	Annealing	54°C	30 sec	Repeat 4 times
4	Extension	72°C	45 sec	
5	Denaturation	98°C	5 sec	
6	Annealing	58°C	30 sec	Repeat 21 times
7	Extension	72°C	45 sec	
8	Denaturation	98°C	5 sec	Repeat 15 times
9	Annealing/Ext	72°C	45 sec	
10	Elongation	72°C	10 min	
11	Store	4°C	∞	

Reagent	Volume [µL]
5x Buffer GC	10
DMSO	1.5
1:10 cDNA	2
primer mix 10 µM	2.5
dNTPs 10 mM	1.5
Phusion	1
ddH$_2$O	31.5
total volume	50

Table 4: Primers for isoform detection, insert generation and sequencing

	forward	reverse	Amplicon [bp]
NRP2 iso1-3	AGCAGCTCCAGCCCAAACGA	CTTTGTCGGTCGAGGGGGCG	Iso 1: 553 Iso 2: 538 Iso 3: 487
NRP2 iso4-5	AGCAGCTCCAGCCCAAACGA	CGGAGACCAGCACCAGCACG	Iso 4: 547 Iso 5: 532
NRP2 iso6	GGTCACAGATGCTCCCTGCTC	TTACAGAGGCCTCCAAGAACAGC	385
NRP2_Iso1-3_CO-MssI_h_F1/ R1	GAGAGTTTAAACATGGATATG TTCCCTCTCACCTGG	GAGAGTTTAAACTCATGCCTCGG AACAACACTTTTG	2796
pWPI sequencing	GTGCAGGGGAAAGAATAGTAGAC	AAGCGGCTTCGGCCAGTAAC	-
NRP2 sequencing	GTCTGGCCGGATTGCTAATG	GGAAGCCAGCCAGGAGAG	-

Iso 1, NRP2a$_{(22)}$; Iso 2, NRP2a$_{(17)}$; Iso3, NRP2a$_{(0)}$; Iso 4, NRP2b$_{(5)}$, Iso 5, NRP2b$_{(0)}$; Iso 6, $_{S9}$NRP2

4.1.3 Agarose gel electrophoresis and gel extraction

Agarose gel electrophoresis was performed with 0.8-2% agarose gels in TAE buffer. Agarose was dissolved in 1xTAE Buffer (4.84 g/L Tris, 1.141 mL/L glacial acetic acid, 1 mM EDTA pH 8) by heating. Ethidium bromide or SYBR Safe (Invitrogen, USA) was added to the gel in a ratio of 1:10,000 after the solution cooled. Gels were poured into the casting chambers and left for polymerization with applied combs. The polymerized gel was transferred into the buffer chambers filled with TAE buffer. Aliquots of PCR products were diluted with 5x loading buffer (60 mM Tris/HCl pH 7.5, 40% sucrose (w/v), 0.2% bromophenol blue) and 10-15 µL of the sample aliquots as well as 5 µL of 100bp or 1kbp Plus DNA Gene Ruler (Fermentas, USA) were loaded onto the gel. Gels were run with 10V/cm. Stained DNA was visualized using the GelDoc system (Bio-Rad, USA) under UV light.

4.1.4 Quantitative PCR

Primers for quantitative PCR (qPCR) were designed using PerlPrimer or PrimerBlast (NCBI) and were analyzed for any secondary structures (dimers, loops, hairpins, runs) in PerlPrimer, GeneRunner and PrimerExpress 3.0. Primers were designed to have a melting temperature of 60°C ± 2°C and to amplify an amplicon of 200 bp at maximum. In a Fast Reaction Tube or 96-well plate (Applied Biosystems, USA), 3 µL cDNA (1:100 diluted in ddH$_2$O), 2 µL primer mix (4 µM stock of forward and reverse primer in ddH$_2$O) and 5 µL Fast SYBR green master mix (Applied Biosystems, USA) were mixed. Triplicates for RhoA and NRP2 were analyzed by the 7500 Fast Real Time PCR System (Applied Biosystems, USA). Conditions used for qPCR are shown in Table 5. Primers used for qPCR are described in Table 6.

Table 5: Conditions for qPCR

Step	Function	Temp [°C]	Time [sec]
1	Enzyme activation	95	20
2	Denaturation	95	3
3	Anneal/Extend	60	30
4	Melting curve analysis	55-95°C	60 sec / 1°C

Table 6: Primers used for qPCR

	forward	reverse	Amplicon [bp]
NRP2 Iso1-5	CTGTGGGTCATCCGTGAGGAC	ATGGGTTCCATGCAGTTCTCCAG	185
hRhoA	CCATCATCCTGGTTGGGAAT	CCATGTACCCAAAAGCGC	141
mRhoA	TGGTTGGGAACAAGAAGGAC	TGGTCTTTGCTGAACACTCC	150

Iso1-5, see Table 4

4.1.5 Preparation of chemically competent Escherichia coli

A 20 mL culture of *E. coli* DH5α in LB (10 g/L NaCl, 10 g/L Tryptone, 5 g/L yeast extract) was prepared from a glycerol stock and incubated overnight at 37°C and 180 rpm. From this overnight culture, 4 mL were used to inoculate 2 Erlenmeyer flasks with 200 mL LB. A sample was taken for the time point t_0 and the flasks were incubated at 37°C and 180 rpm. Samples were taken on a regularly basis and OD_{600} was measured using a photometer until an OD_{600} of 0.4. All following steps were performed on ice. The culture was poured into pre-chilled 50 mL reaction tubes, incubated 10 min on ice and centrifuged for 7 min at 3000 rpm and 4°C in a centrifuge (Heraeus, Thermo Scientific, Germany). Supernatant was decanted and cells were resuspended in 10 mL ice-cold $CaCl_2$ solution (100 mM $CaCl_2$, 10 mM PIPES, autoclaved). Cells were centrifuged for 5 min at 2500 rpm and 4°C and supernatant was again decanted. Cells were resuspended in 10 mL ice-cold $CaCl_2$ solution, pooled and incubated 30 min on ice. Pooled cells were centrifuged for 5 min at 2500 rpm and 4°C and each pellet was resuspended in 4 mL ice-cold glycerol/$CaCl_2$ solution (100 mM $CaCl_2$, 10 mM PIPES, 15% (w/v) glycerol, sterile filtrated). Resuspended cells were aliquoted in pre-chilled reaction tubes, transferred on dry ice and stored at -80°C.

4.1.6 Cloning, ligation and transformation

Vector preparation

The vector pWPI, obtained from Addgene, was transformed into *E. coli* DH5α. 50 ng pWPI were prepared in a reaction tube on ice and 100 µL chemically competent DH5α were added. The tube was gently snipped and incubated on ice for 30 min. The mixture was heat shocked at 42°C for 45 seconds and transferred on ice for 2 min. 1 mL room temperature SOC medium (20 g/L Tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$,

20 mM D-glucose) was added to the tube and bacteria were allowed to incubate in a horizontal shaker at 37°C for 1 h. Afterwards, cells were centrifuged, resuspended in 50 µL volume and spread on a pre-warmed petri dish with LB-agar (LB with 2% Agar-agar) containing 75 mg/L ampicillin (LB-Amp) with a sterile Drigalski spatula. The dish was incubated at 37°C overnight. A clone from this plate was used to inoculate a 5 mL liquid LB-Amp overnight culture at 37°C with shaking which was subsequently used to inoculate 150 mL LB-Amp overnight culture at 37°C with shaking. This culture was then harvested and pWPI was isolated from the cells using the QIAGEN Plasmid Midi Kit (QIAGEN, Germany) according to manufacturer's protocol. Furthermore, isolated pWPI was cut and dephosphorylated using *Mss*I (Fermentas, Germany) and FastAP (Fermentas, Germany) in a single reaction and then purified using the QIAquick Gel Extraction kit (QIAGEN, Germany) according to manufacturer's protocol.

<u>Cloning of NRP2 insert</u>

Inserts for cloning were obtained using Phusion PCR from 3sp cDNA and gel extraction from gels stained with SYBR Safe as described above. Purified inserts were cut with the restriction enzyme *Mss*I (Fermentas) which was loaded on an agarose gel stained with SYBR Safe and extracted via gel purification using the QIAquick Gel extraction kit. The ligation reaction of 20 µL total volume was prepared on ice and contained 100 ng pWPI, cut insert in a ratio of 1:5-10, 2 µL 10x T4 buffer, 2 µL PEG 4000 50% (w/v), 5U T4 ligase and ad 20 µL ddH$_2$O. The reaction was incubated overnight at 4°C and then 4h at 16°C. Of this ligation reaction, 3 µL were used to transfect 100 µL chemically competent *E. coli* DH5α as described above. Transfected cells were spread on LB-Amp plates, incubated overnight at 37°C and clones were picked to inoculate 5 mL overnight cultures as described above. 1.5 mL of this overnight culture were harvested in a reaction tube and plasmid DNA was purified using the QIAGEN Plasmid Mini Kit (QIAGEN, Germany) according to manufacturer protocols.

Purified plasmid DNA was further analyzed to ensure the correct size and orientation of the insert. The purified plasmid samples were digested in a reaction tube containing 3 µL plasmid, 0.5 U *Mss*I, 1µL 10X FastDigest Green Buffer (Fermentas, Germany) and 5.5 µL ddH$_2$O for 15 min at 37°C. The enzyme was heat-inactivated for 10 min at 65°C and samples were loaded directly onto an agarose gel. Samples containing insert were digested in a reaction tube containing 3 µL plasmid, 0.5 U *Eco91*I, 0.5 U *Cla*I, 1µL 10X FastDigest Green Buffer (Fermentas, Germany) and 5 µL ddH$_2$O for 15 min at 37°C. Enzymes were heat-inactivated for 15 min at 65°C and samples were directly loaded onto an agarose gel. Plasmids containing NRP2 inserts of correct size and orientation were sent for sequencing as described above.

4.1.7 Protein sample preparation and Bradford protein analysis

The following procedure was performed on ice. Cells were washed twice with ice-cold PBS. Ripa buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1mM ß-glycerophosphate pH 7.2, 0.5% sodium deoxycholate, 5 mM $MgCl_2$, 1% Nonidet NP-40) containing phosphatase (1 mM NaF), protease (10 µg/mL mM aprotinin, 10 µg/mL leupeptin) and membrane-bound enzyme inhibitors (1 mM NaV_3O_4) was added and cells were harvested with a rubber scraper. The cell suspension was pipetted into reaction tubes containing the protease inhibitor PMSF (100 mM PMSF in 97% ethanol) with a final concentration of 0.5 mM. The tubes were snap frozen in liquid hydrogen and thawed for 20 min on ice. The thawed mixture was centrifuged in a table centrifuge at 16,000 x g for 20 min at 4°C. The supernatant containing the proteins was removed and aliquots were diluted 1:10 in ddH_2O for the measurement of the protein concentration. A dilution series of BSA in ddH_2O (0.125, 0.25, 0.375, 0.5, 0.75, 1.0, 1.5 mg/mL protein) was used for the calibration curve. 5 µL sample or BSA dilution were pipetted in triplicates in a 96-well plate and 195 µL of 1:5 diluted Bradford reagent (Bio-Rad Protein assay, Bio-Rad, USA) was added. The plate was then analyzed using a colorimetric microplate reader at 620 nm and protein concentrations were calculated in MS Excel 2010. Protein samples were then diluted to the desired concentration with Ripa buffer and 5x sodium dodecyl sulfate (SDS)-loading buffer (250 mM Tris pH 6.8, 10 % SDS, 30 % glycerol, 5 % β-mercaptoethanol, 100 mM dithiothreitol, some crumbs bromophenol blue), incubated at 95°C for 3 min, spinned down, vortexed and loaded on gels.

4.1.8 SDS-polyacrylamide gel electrophoresis

Proteins were separated according to their size using SDS-polyacrylamide gel electrophoresis (PAGE). Separation gels containing polyacrylamide (PAA) (Table 7) with a given percentage were casted between the spaced glass plates of the gel casting equipment (Bio Rad , USA) and overlayed with isopropanol, which was discarded after polymerization.

Table 7: Separation gel (sufficient for 2 gels)

	10% PAA	12% PAA
ddH_2O	6.2 mL	4.42 mL
2 M Tris, pH 8.8	4.5 mL	2.25 mL
30% Acrylamide, 1% N,N'-Methylenebisacrylamide	4.45 mL	5.2 mL

+90 µL ammonium persulfate (APS, 10%) + 15 µL tetramethylethylenediamine (TEMED, Sigma)

The stacking gel was prepared (Table 8) and transferred on top of the separation gel after removal of the isopropanol. Gel combs were applied and gel was polymerized.

Table 8: Stacking gel (sufficient for 2 gels)

ddH$_2$O	3 mL
1 M Tris/HCl, pH 6.8	0.5 mL
30% Acrylamide,	0.5 mL
1% N,N'-Methylenebisacrylamide	
+40 µL APS (10%) + 8 µL TEMED	

Protein samples of equal concentration between 10 and 30 µg were loaded and the gel was operated at 5V/cm in an electrophoresis equipment (Bio-Rad, USA) filled up with electrophoresis buffer (2.5 mM Tris, 19.2 mM glycine, 0.1% SDS)

4.1.9 Western blot

Proteins were transferred from the SDS-PAGE gel onto a nitrocellulose membrane (Protran, Whatmen, UK) through wet Western blotting for further analysis. The blotting sandwich was assembled, while avoiding air bubbles, in blotting buffer (2.5 mM Tris, 19.2 mM glycine, 15% (v/v) mL methanol, 0.02% SDS) and consisted of a sponge on which 2 filterpapers (Whatmen, UK) were placed. The nitrocellulose membrane was placed on the filter papers followed by the SDS-PAGE gel, two filter papers and a sponge on top. The blotting sandwich was closed and transferred into the blotting cartridge containing ice-cold blotting buffer. The blotting was performed on ice for 1 h at 100V. Afterwards, the membrane was washed with ddH$_2$O and stained with Ponceau S (0.2% Ponceau in 3% trichloroacetic acid). The membranes were cut with a razor blade to separate desired protein size ranges. The membranes were blocked with 4% BSA in TBS (20 mM Tris, 150 mM NaCl, pH 7.5) containing 0.1% Tween-20 (TBST) for 1h at room temperature under agitation. The blocked membranes were directly incubated with 3 mL of the first antibody diluted in blocking solution over night at 4°C under agitation. Primary antibodies were stored and the membranes were washed three times for 15 min with TBST and then incubated with the second antibody of the appropriate species in TBST for 1 h at room temperature under agitation. After incubation, the membranes were washed three times for 15 min with TBST and incubated in luminol/coumarin working solution (200 mL 0.1 M Tris pH 8.8, 500 µL of 13.07 mg/L p-coumaric acid in DMSO, 1 mL of 44.3 g/L luminol in DMSO) plus 3 µL/mL 3% H$_2$O$_2$ for 1 min, then transferred into an exposure cassette and developed. For reprobing with primary antibodies, membranes were stripped in 3 mL stripping solution (20 mM Tris/HCl pH 6.8, 2% SDS) containing 7 µL/mL ß-mercaptoethanol for 30 minutes at 55°C. Membranes were thoroughly washed with ddH$_2$O and TBST before proceeding with blocking and subsequent incubation with primary antibodies as described above.

4.1.10 Immunofluorescence

Cells were grown in 24-well plates on cover slips or on microscope slides using a 12-well flexiPERM (Greiner Bio-One, Austria), both coated with collagen. Cells were washed two times with room temperatured PBS and then incubated with Histofix 4% formaldehyde solution (Roth Lactan, Austria) for 8 min at 37°C. The fixed cells were then washed three times with PBS for 15 min at room temperature and stored at 4°C in PBS. The fixed cells were permeabilized using 0.25% Triton-X (Roth Lactan, Austria) in PBS for 5 min followed by washing twice with PBS for 5 min. Blocking was performed with 4% serum of the animal in which the desired secondary antibody was produced (Vector Laboratories, USA) in PBS with 1% BSA for 45 min at room temperature. The first antibody was used 1:100 in 1% BSA/PBS with 1% serum for 1.5 h at room temperature followed by washing three times with PBS for 15 min. All following steps were performed in the dark. The secondary antibody was used 1:250 in 1% BSA/PBS for 1 h at room temperature followed by washing three times for 15 min with PBS. The slide or coverslip was then embedded in Roti-Mount FluorCare (Roth Lactan, Austria) and signals were detected using the confocal laser microscope LSM-700 (Zeiss, Germany).

4.1.11 Immunohistochemistry

Animal tumor tissue was fixed overnight in Histofix 4% formaldehyde solution (Roth Lactan, Austria) after extraction, washed in PBS and stored in 70% ethanol. The tissue was then embedded in paraffin. The paraffin block was cut using a microtome creating 4 µm thick slices which were attached onto microscope slides. The slides were deparaffinized twice for 20 min in xylene, transferred into isopropanol twice for 10 min and rehydrated in a descending alcohol series (97%, 80%, 70%, 60% ethanol and ddH$_2$O; each for 2 min). The rehydrated tissue was boiled in 10 mM citric acid, pH 6, for 20 min and cooled at room temperature for 20 min in order to retrieve antigens. The slides were put into PBS for 5 min and then treated with 3% H$_2$O$_2$ in ddH$_2$O for 10 min in order to saturate endogenous peroxidases. Slides were washed in PBS for 5 min and transferred onto a Shandon Coverplate (Thermo Scientific, Germany) and placed in a histo-cassette rack. The tissue was permeabilized with 150 µL 0.25% Triton-X in PBS for 5 min and then washed with PBS three times for 5 min. Blocking was performed with 150 µL 4% serum of the animal in which the desired secondary antibody was produced (Vector Laboratories, USA) in PBS for 1 h at room temperature. Of the first antibody, 100 µL 1:100 in 1% serum-containing PBS overnight at 4°C were used. The slides were washed afterwards three times with PBS for 15 min and treated with 100 µL of the secondary antibody 1:200 in PBS for 1 h at room temperature. The slides were then washed three times with PBS for 15 min. The ABC complex (Vector Laboratories, USA) was prepared by adding one drop of

solutions A and B to 2.5 mL PBS followed by an incubation of 30 min at room temperature. 250 µL of ABC complex were added to the slides and incubated for 45 min at room temperature. The slides were washed with PBS three times for 10 min. 3,3'-Diaminobenzidine (DAB) substrate (Vector Laboratories, USA) was prepared by adding one drop DAB, two drops H_2O_2 and one drop buffer solution to 2.5 mL ddH_2O. The slides were removed from the histo-cassette rack and treated with the DAB substrate between 1 and 10 min until sufficient staining was achieved. The staining reaction was stopped by transferring the slides into PBS. Slides were washed afterwards with ddH_2O for 5 min, and stained with Mayer's hemalaun solution (Roth Lactan, Austria). Superfluous staining was removed by washing the slides for 10 min under running tap water and 5 min in ddH_2O. The slides were dehydrated in an ascending alcohol series and put into xylene as described above. Slides were then embedded in Entellan (Merck) and pictures were taken using a Nikon TMS (Nikon, Japan) microscope.

4.1.12 Antibodies

Primary and secondary antibodies used for Western blot, immunofluorescence and immunohistochemistry are shown in Table 9 and Table 10.

Table 9: 1st Antibodies

Antibody	Company	Dilution	Reactivity	Source	Size [kD]
Western blot					
NRP2	R&D	1:4000	h, m	Goat	120
β-Actin	Sigma-Aldrich	1:2500	wide range	Rabbit	42
p-FAK	Invitrogen	1:1000	h, m, rat, c, fr	Rabbit	125
FAK	Santa Cruz	1:1000	h, m, rat, c	Rabbit	125
p-Smad2	Upstate	1:500	h, m, rat	Rabbit	60
Smad 2/3	BD	1:1000	h, m, rat, dog	Mouse	58
p-JNK/SAPK	Cell Signaling	1:1000	h, m, rat	Rabbit	54/46
JNK/SAPK	Cell Signaling	1:1000	h, m, rat	Rabbit	54/46
α-Tubulin	Sigma-Aldrich	1:1000	h, m, rat, b, c	Mouse	50
Immunohistochemistry					
NRP2	R&D	1:100	h, m	Goat	120
E-cadherin	BD	1:100	h, m, rat, dog	Mouse	120
GFP	Santa Cruz	1:100	wide range	Rabbit	27
Immunofluorescence					
NRP2	R&D	1:100	h, m	Goat	120
Smad2/3	BD	1:100	h, m, rat, dog	Mouse	58
Hoechst 33258	Sigma-Aldrich	1:10,000	DNA	-	
Phalloidin-rhodamine	Invitrogen	1:1,000	Actin fibers	*Amanita phalloides*	

h, human; m, mouse; c, chicken; fr, frog; b, bovine.

Table 10: 2nd Antibodies

Conjugate	Company	Dilution	Reactivity	Source
Western blot				
Peroxidase	Vector Laboratories	1:10,000	Rabbit IgG	Goat
Peroxidase	Vector Laboratories	1:10,000	Mouse IgG	Horse
Peroxidase	Vector Laboratories	1:5,000	Goat IgG	Horse
Immunohistochemistry				
Biotin	Vector Laboratories	1:200	Goat IgG	Rabbit
Biotin	Vector Laboratories	1:200	Rabbit IgG	Goat
Biotin	Vector Laboratories	1:200	Mouse IgG	MoM kit
Immunofluorescence				
AF 488	Invitrogen	1:1000	Mouse IgG	Goat
Cy3	Jackson ImmunoResearch	1:200	Goat IgG	Donkey

4.1.13 Cell lines, media and reagents

3p and 3sp

The 3p and 3sp cells were isolated from the HCC of a single patient [192]. The 3p cell line shows an epithelial phenotype as well as expression of epithelial markers such as E-cadherin and keratin 8. Furthermore, they display intact adherens junctions as well as low migratory and invasive capabilities. 3sp cells display a mesenchymal phenotype with relocalization of actin from the cell membrane to stress fibers and the presence of mesenchymal markers such as the transcription factors LEF1, SNAI1, SNAI2, and ZEB1 as well as an upregulation of vimentin. This cell line shows high migratory and invasive potential. The common genomic identity of 3p and 3sp cells was verified by short tandem repeat analysis. Moreover, array-comparative genomic hybridization (aCGH) analysis showed an identical cellular origin of these cells. Thus, suggesting that 3sp cells arose from 3p cells through EMT in the patient.

SNU398

The HCC cell line SNU398 was isolated and described between 1995 and 1999 [193,194]. This cell line exhibits mutations in p53 and β-catenin as well as HBV integration [195]. They display a late TGF-β signature as well as a poor response of canonical TGF-β signaling because of decreased expression of ELF and TβRII [196-198]. SNU398 show an invasive phenotype with a high metastatic potential *in vivo* [199].

MIM-Ras

MIM-Ras cells were established from the MIM-1-4 cell line which are immortalized hepatocytes isolated from p19$^{ARF-/-}$ mice through stable retroviral transmission with oncogenic v-Ha-Ras and green fluorescent protein (GFP) [200,201]. This cell line displays an epithelial

morphology, contact-inhibited growth and shows a malignant phenotype. MIM-Ras cells with puromycin resistance instead of GFP for selection (designated MIM-pRas) were used to generate NRP2 overexpressing cells.

MIM-RT

MIM-RT cells were generated through long-term treatment of MIM-Ras cells with TGF-β [201]. In contrast to MIM-Ras cells, this cell line displays a mesenchymal phenotype without epithelial polarization. These cells grow in polylayers without contact inhibition and show increased malignancy compared to MIM-Ras cells.

4.1.14 Cultivation of cells

Murine hepatocyte cell lines were grown on collagen-coated culture dishes (rat-tail collagen, BD Transduction Lab, USA) in RPMI 1640 plus 10% fetal calf serum (FCS) and antibiotics (50 U Penicillin and 50 µg/mL Streptomycin). For MIM-RT, 1 ng/mL TGF-β1 (R&D Systems, USA) was added to the culture medium. The human HCC cell lines 3p, 3sp and SNU398 were grown in RPMI 1640 plus 10% FCS and antibiotics (50 U Penicillin and 50 µg/mL Streptomycin). All cells were kept in a humidity controlled incubator at 37˚C in 5% CO_2. For stimulation, cells were serum starved for 24h and re-stimulated with RPMI 1640 plus 10% FCS for 30 min prior to sample preparation for Western blots.

Splitting of cells

Splitting of the cells was performed by removal of the culture medium, washing of the cells with PBS and addition of trypsin/EDTA (0.1% trypsin/ 0.01% EDTA). Cells were resuspended in RPMI 1640 plus 10% FCS and antibiotics and split in desired ratio.

Freezing and thawing of cells

Medium was removed, cells were washed with PBS and trypsinized using trypsin/EDTA as described above. Cells were resuspended in RPMI 1640 plus 10% FCS and centrifuged at 200 x g for 6 min. The medium was removed and the cell pellet was resuspended in FCS containing 5% DMSO (Merck, Germany). The cell suspension was transferred into round bottom freezing vials (1 mL suspension/vial) and incubated on ice for 15 min. Cells were stored at -80°C and, for long term storage, transferred into liquid N_2 after at least one night at -80°C.

For thawing, cells were quickly thawed at 37°C and resuspended in 3 mL RPMI 1640 plus 10% FCS. The cell suspension was centrifuged at 200 x g for 6 min and medium was removed. Cells were resuspended in 3 mL RPMI 1640 plus 10% FCS and antibiotics, transferred into a 6 cm culture dish.

4.1.15 Transfection of cells with siRNA

Cells were transfected with siRNA (ON-TARGETplus SMARTpool, Dharmacon, Thermo Scientific, UK) against human NRP2 at a concentration of 120 nM as described in the following. Cells were seeded in 6-well plates the day before transfection in an appropriate concentration as determined using a cell analyzer (Schärfe System, Germany). On the next day, transfection was performed. In one reaction tube 150 µL Opti-MEM (Invitrogen) were added to 3 µL siRNA and in a second tube 40 µL Opti-MEM were added to 4.5 µL Oligofectamine (Invitrogen, UK). Both tubes were mixed by gentle pipetting. After ten minutes incubation at room temperature, the content of the first tube was carefully transferred into the second tube. The mixture was mixed by flicking the tube. The tube was incubated for 15 min at room temperature. Afterwards, 300 µL Opti-MEM was added and the total volume (500 µL) of the tube was carefully transferred onto cells which have been seeded one day earlier at 150,000 cells per well in a 6-well culture plate. The cells were incubated at 37°C for 5 h and then 1 mL RPMI with 15% FCS and 1.5x P/S were added. After incubation for further 24 h at 37°C, the medium was changed and cells were used on the next day for experiments (48 h post infection).

4.1.16 Dense curve

Cells were seeded at a concentration of $1.5*10^4$ in 24-well plates in quadruplicates. At each chosen timepoint, cells were washed with PBS, trypsinized and counted using a cell analyzer (Schärfe System, Germany).

4.1.17 Cell migration and invasion assay

Cell invasion assays were performed using 8 µm cell culture inserts for 24-well plates (Corning, US). Inserts were coated on ice by adding 50 µL Matrigel solution (1 mg/µL, BD Biosciences, USA) to the plate followed by removal of 20 µL. The plates were incubated for 30 min at 37°C. Cells were seeded at a concentration of $4*10^5$/mL, determined using a cell analyzer (Schärfe System, Germany), into the insert in RPMI containing 1% FCS. The lower chamber was filled with 600 µL RPMI with 10% FCS. Cells were incubated for 24 h at 37°C. Afterwards, culture media was removed and residual cells were removed from the upper part of the membrane with cotton swabs. The lower part of the membrane was washed twice with PBS

for 5 min and then fixed with room temperature Histofix 4% formaldehyde solution (Roth Lactan, Graz, Austria) for 25 min. Fixed cells were washed once with PBS for 5 min and stained with Hoechst 33342 (Roche, Austria, diluted 1:10,000 in PBS) for 15 min at room temperature. Stained cells were washed once in ddH$_2$O for 5 min and embedded onto microscope slides using Mowiol. Pictures of random fields were taken using a fluorescence microscope (Nikon TMS, Nikon, Japan) and counted in ImageJ. Cell migration assays were performed as described above but with 8 µm ThinCerts (Greiner Bio-One, Austria) without coating using a cell concentration of 2*10^5/mL for 24 h.

4.1.18 Cell adhesion assay

Experiments to assess cell adhesion were performed in flat-bottom 96-well plates (CytoOne, Starlab, Germany). ECM proteins were human collagen I (C7999), laminin (L2020), vitronectin (SRP3186) (Sigma-Aldrich, Austria) and fibronectin (#354008, BD Bioscience, USA), which were reconstituted to 1 mg/mL according to manufacturer's protocols. Plates were coated on ice with 50 µL of selected ECM proteins diluted to 10 µg/mL in PBS and incubated at 4°C overnight. On the next day, coating solutions were removed and 150 µL 2% BSA in PBS were added for 30 min at RT to block remaining protein binding sites. Uncoated wells were also blocked with BSA to provide a control. Immediately prior to adding cells, BSA was removed and wells were washed three times with PBS. Cells were trypsinized and harvested in medium with 10% FCS. Cells were washed three times with medium without FCS. Washed cells were resuspended in blank medium and cell number was determined with a cell analyzer (CASY; Schärfe Systems, Germany). Cells were diluted to desired concentrations (3sp cells: 1.5*10^5/mL) and 100 µL were applied per well. After incubation for 1 h at 37°C, medium was carefully decanted and non-adherent cells were removed by washing five times with 200 µL PBS. Cells were fixed using 50 µL of 96% ethanol for 10 min and stained with 50 µL of 0.1% crystal violet (in 8% ethanol) for 30 min. Excessive crystal violet was removed by washing 10 times with 200 µL ddH$_2$O. Crystal violet bound to nuclei of cells was released by treatment with 1% SDS for 1 h. Optical density was measured at 570 nm using a plate reader (Synergy HT, BioTek, Germany).

4.1.19 Mouse xenograft model

SCID mice were purchased from Charles River Laboratories (USA). For tumor xenografts, MIM-Ras (5*10^6), MIM-RT (5*10^6) and SNU398 (5*10^6) each resuspended in 100 µL PBS were injected subcutaneously into the lower sides of the abdomen of female SCID mice. Mice were monitored every 2-3 days and tumors were removed before they reached a size of 1 cm³ on

each flank. Tumor volume (mm³) was calculated by (length (mm) x width (mm) x width (mm))/2, where length was the longest axis.

4.1.20 Correlation of NRP2 with clinical data

Frozen tissue of paired lesion and adjacent tissue samples from human HCCs (n=38) were a kind gift from the group of Marco Chilosi (University of Verona, Verona, Italy). RNA was extracted, transcribed into cDNA and analyzed via qPCR as described above in our lab. Data were analyzed using Microsoft Excel 2010 and Graphpad Prism 5. Microarray and clinical data for patients with squamous cell lung carcinomas (n=130) [17] were downloaded from GEO and for breast cancer patients (n=133) [18] from the MD Anderson Cancer Center website (available: http://bioinformatics.mdanderson.org/pubdata.html; accessed 2012 Sep 07) and further analyzed as described previously [191]. Statistical analysis was performed as described in 4.1.22 below.

4.1.21 Tissue microarray

A tissue microarray containing 216 cases was obtained from US Biomax (LV2161, US Biomax, USA). Clinical data included age, gender, grade, stage and TNM stage. HCC (n=176), mixed HCC with cholangiocellular carcinoma (n=3), cholangiocellular carcinoma (n=29), normal liver tissue (n=8) and skin melanoma (n=1) tissues were present on the array. The array was stained for NRP2 as described in 4.1.11. The array was subsequently evaluated by two independent researchers (Philipp Wittmann and Christian Gröger) without prior knowledge about the corresponding clinical data of the tissue. Data were summarized and differing results were discussed until a consensus was achieved. NRP2 intensity score (NIS) was calculated by addition of the staining intensity (0, none; 1, low; 2, medium; 3, high) with the score of the percentage of positively membrane-stained cells (0-24%, 1; 25-49%, 2; 50-74%, 3; 75-100%, 4). Hence, the NIS could range between 0 and 7. A NIS of 1, however, was not possible because both the staining intensity and the percentage of positive stains could either be 0 or 1 which resulted in a NIS of 0 or 2, respectively. Statistical analysis of the clinical data was performed as described in 4.1.22 below.

4.1.22 Statistics

Data were shown as means ±SD unless indicated otherwise. The statistical significance of differences was evaluated using two-sided Student's *t*-test unless indicated otherwise. For the tissue microarray analysis, 2 out of 3 statistical tests for normality (Kolmogorov-Smirnov test, D'Agostino and Pearson omnibus normality test, Shapiro-Wilk normality test) indicated a non-

Gaussian distribution of the NIS. Thus, Kruskal-Wallis test and one-sided Mann Whitney U-test were performed to detect differences between the grades. Significant differences between experimental groups were *$P < 0.05$, **$P < 0.01$, or ***$P < 0.005$.

4.2 Results

4.2.1 Meta-analysis of gene expression studies reveals NRP2 as potential EMT effector

The analysis of 24 gene expression datasets revealed that NRP2 is upregulated in 6 of 24 but not downregulated in any (data not shown). Furthermore, it was of great interest that NRP2 showed increased expression in mesenchymal 3sp cells compared to epithelial 3p cells. These data are available in the GSE26391 dataset [202]. NRP2 displayed a 25-fold upregulation in 3sp cells. The 3p/3sp EMT model showed some relationship to other EMT models where EMT was induced via TGF-β. NRP2 has been reported to promote EMT by cooperation with TGF-β signaling in a colon cancer cell model. Thus, we decided to investigate the role of NRP2 in the EMT of the 3p/3sp cell line as well as in primary HCC.
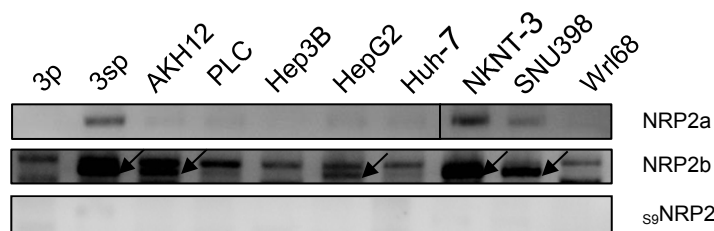
4.2.2 NRP2 isoforms are present in selected HCC cell lines

We further analyzed the expression of NRP2 in available HCC cell lines. Quantitative PCR revealed that NRP2 transcripts are present in 3sp, NKNT-3 and SNU398 cells. The level compared to 3sp was approximately 40% in both NKNT-3 and SNU398 (Figure 10).
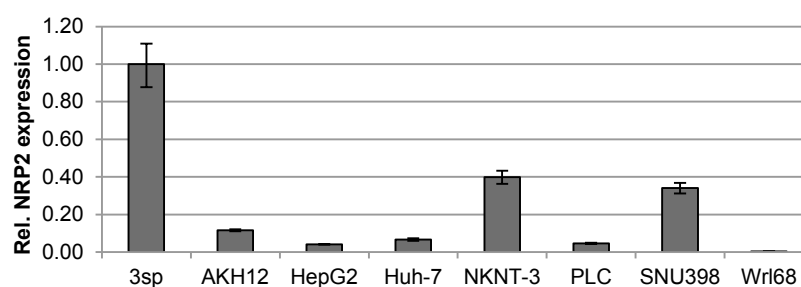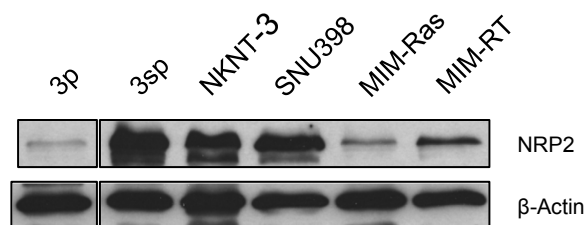

Figure 10: NRP2 is expressed in 3sp, NKNT-3 and SNU398 cell lines. Quantitative PCR was performed with primers which identify all isoforms. Relative expression to RhoA is shown. Analysis was performed in duplicates.

The qPCR did not distinguish between the different NRP2 isoforms NRP2a$_{(22)}$, NRP2a$_{(17)}$, NRP2a$_{(5)}$, NRP2a$_{(0)}$, NRP2b$_{(5)}$, NRP2b$_{(0)}$ and $_{s9}$NRP2. Therefore, the isoform presence was determined in the mentioned HCC cell lines using isoform specific PCR primers for NRP2a, NRP2b and $_{s9}$NRP2 isoforms (Figure 11).

Figure 11: Expression of NRP2 isoforms in selected HCC cell lines. Upper panel: The specific band indicates the presence of $NRP2a_{(17)}$ and/or $NRP2a_{(22)}$. The presence of $NRP2a_{(5)}$ or $NRP2a_{(0)}$ would have resulted in a band 50 bp smaller than the band of the other NRP2a isoform which was not detectable. **Middle panel:** The bands specific for the isoforms $NRP2b_{(5)}$ and $NRP2b_{(0)}$ are indicated with arrows beneath an unspecific band. **Lower panel:** No specific band indicating the presence of $_{s9}NRP2$ was observed.

It was possible to detect the isoforms $NRP2a_{(22)}$ and/or $NRP2a_{(17)}$ in 3sp, NKNT-3 and SNU398. Small amounts were also observed in AKH12 and HepG2 cells. Isoforms $NRP2a_{(5)}$ and $NRP2a_{(0)}$ would have shown a band of a smaller size and were therefore not present in any analyzed cell line. Interestingly, also the isoforms $NRP2b_{(5)}$ and/or $NRP2b_{(0)}$ were detectable in 3sp, AKH12, HepG2, NKNT-3 and SNU398. However, there was no $_{s9}NRP2$ in any analyzed cell line. This data is consistent with the known isoform distribution of NRP2 as $NRP2a_{(5)}$, $NRP2a_{(0)}$ and $_{s9}NRP2$ have not been found yet in human cells [30]. Nevertheless, NRP2a isoforms have been reported to be mainly expressed in organs such as the liver whereas the NRP2b isoforms are generally present in the heart and skeletal muscle; thus, the discrimination of NRP2 isoforms expression may not be that simple. Western blot analysis showed that NRP2 is expressed in 3sp, SNU398 and the mouse cell line MIM-RT (Figure 12) and in very low amounts in MIM-Ras cells (Figure 12). NRP2 was barely detectable in 3p cells (Figure 12).

Figure 12: NRP2 expression in selected HCC cell lines. NRP2 is expressed in the human HCC cell lines 3sp, NKNT-3 and SNU398 as well as in the murine hepatoma cell lines MIM-Ras and MIM-RT. β-Actin was used as loading control.

4.2.3 Influences of EMT and TGF-β on NRP2 expression

NRP2 is known to cooperate with TGF-β signaling to promote EMT [153]. Furthermore, we detected an increase of NRP2 in MIM-RT cells which have arose from EMT of MIM-Ras cells under TGF-β treatment (Figure 13, A). Similar results were obtained with the A549 lung adenocarcinoma cell line which undergoes EMT after TGF-β treatment (Figure 13, B). However, qPCR time course analysis of MIM-Ras cells treated with TGF-β did not reveal any effect of TGF-β treatment on NRP2 expression (Figure 13, C).

66
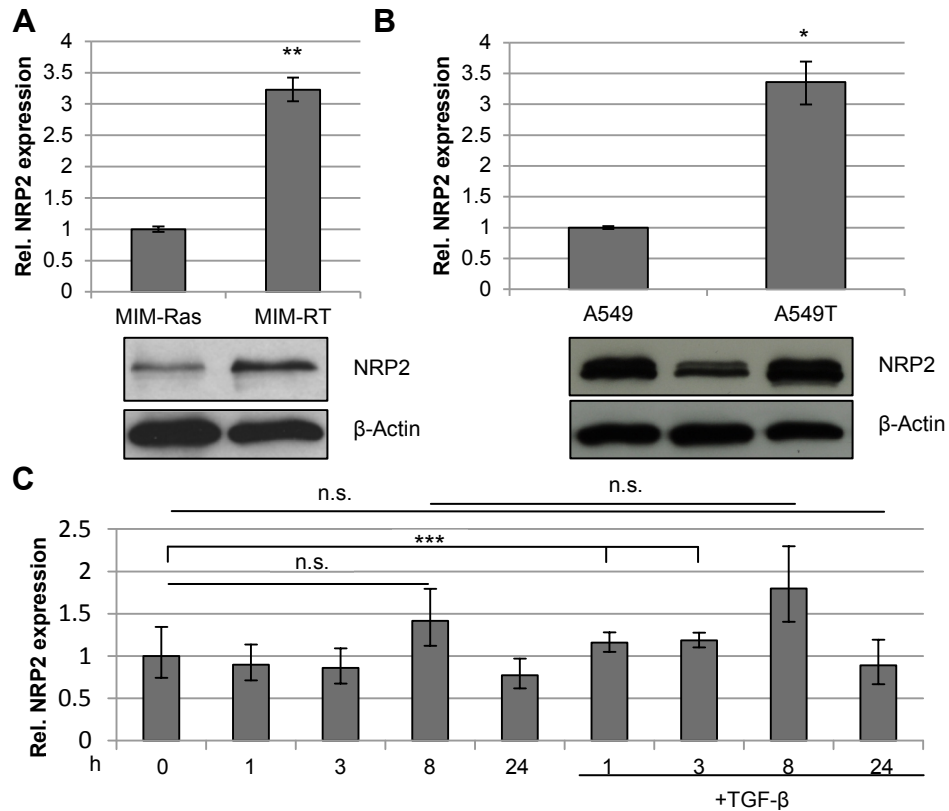
Figure 13: NRP2 expression increases during TGF-β-induced EMT but is no direct target of TGF-β treatment.
(A) NRP2 expression increased when MIM-Ras cells have undergone EMT after TGF-β treatment as measured by qPCR (upper panel) and Western blot (lower panel). **(B)** NRP2 expression increased when A549 cells have undergone EMT after TGF-β treatment as analyzed by qPCR (upper panel) and Western blot (lower panel). **(C)** Quantitative PCR of NRP2 expression in MIM-Ras cells with or without TGF-β over the indicated time course revealed no differences of NRP2 expression. Relative expression compared to RhoA is given in qPCR analysis. Analysis was performed in duplicates. n.s., not significant; *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$. β-Actin was used as loading control.

4.2.4 NRP2 mediates cell migration, invasion, proliferation and adhesion in vitro

To investigate NRP2 functions *in vitro*, 3sp cell were treated with siRNA against NRP2 (Figure 14) and analyzed regarding cell migration, invasion and proliferation. Notably, transient knockdown of NRP2 reduced cell migration and invasion of 3sp cells (Figure 15).
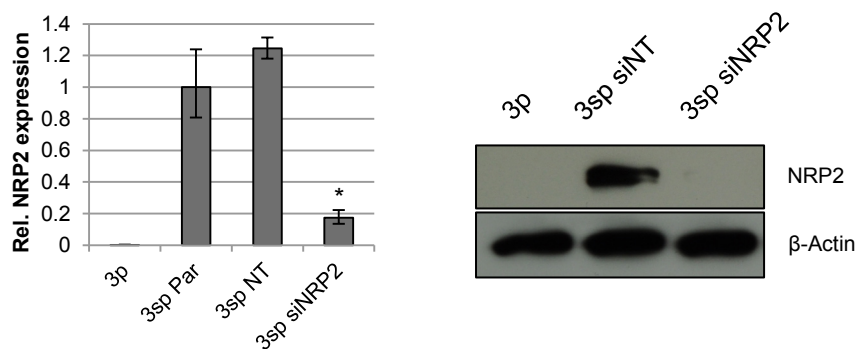
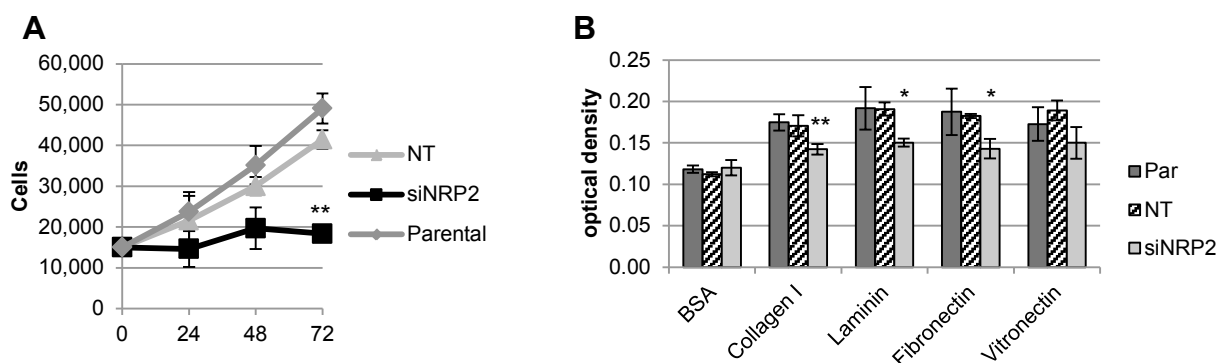**Figure 14: NRP2 knockdown with siRNA in 3sp cells.** Knockdown with siNRP2 achieved less than 20% of the original protein level as shown by qPCR (left panel) and Western blot (right panel). Relative expression compared to RhoA is shown. Analysis was performed in duplicates. Par, 3sp parental cells; NT, 3sp treated with non-target siRNA; siNRP2, 3sp treated with siRNA against NRP2. *, $P<0.05$. β-Actin was used as loading control.
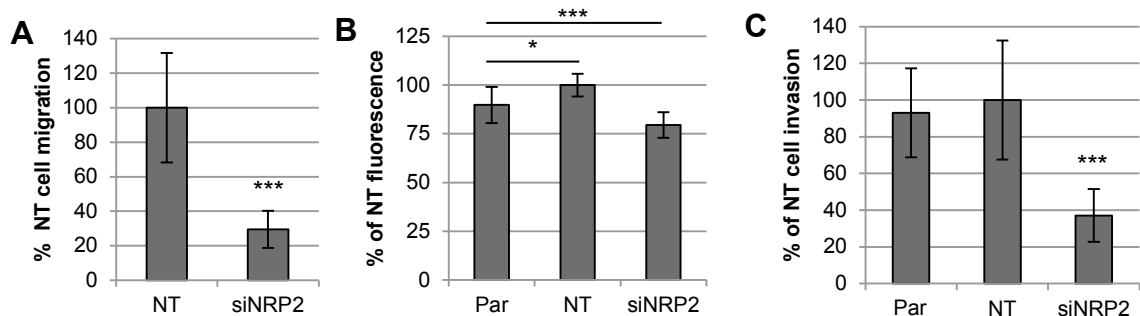
Figure 15: Knockdown of NRP2 inhibits cell migration and invasion in 3sp cells. (A) Cell migration assessed by transwell assay. **(B)** Cell migration assessed by platypus assay. **(C)** Cell invasion assessed by transwell assay. Par, 3sp parental cells; NT, 3sp treated with non-target siRNA; siNRP2, 3sp treated with siRNA against NRP2. *, $P<0.05$; ***, $P<0.001$. Experiments were performed twice with similar results.

Cell proliferation was also reduced after knockdown of NRP2 (Figure 16, A). NRP2 was reported to be involved in integrin and FAK signaling. Hence, we were interested in investigating the effect of NRP2 knockdown on cell adhesion of 3sp cells *in vitro*. Notably, we observed a reduction of adhesion on collagen I, fibronectin and laminin but not vitronectin (Figure 16, B).


Figure 16: NRP2 knockdown impairs cell proliferation and adhesion in 3sp cells. (A) Density curve of 3sp cells treated with non-target siRNA (NT), siNRP2 or without siRNA reveals an inhibition of cell proliferation when treated with siNRP2. Analysis was performed in triplicates. **(B)** Cell adhesion assay performed with 3sp cells. Cell adhesion was measured after 1 h of plating on different human ECM proteins. Analysis was performed in quadruplicates. Par, 3sp parental cells; NT, 3sp treated with non-target siRNA; siNRP2, 3sp treated with siRNA against NRP2. *, $P<0.05$; **, $P<0.01$. Experiments were repeated two times with similar results.

4.2.5 NRP2 detection in tumor xenografts

Next we addressed the effects of NRP2 expression in xenotransplants. Therefore, preliminary data on the abilities to detect NRP2 *in vivo* was gathered. Tumor slices of MIM-Ras and MIM-RT driven tumors were stained successfully to detect NRP2 (Figure 17, A and B). Using E-cadherin and GFP as markers for epithelial MIM-Ras tumor cells, it was possible to detect NRP2 only in tumor cells and not in host cells (Figure 17, A). To ensure the specificity of the antibody on human tissue, SNU398 tumors were generated in SCID mice.

It was possible the detect NRP2 in human tissue with a high specificity as displayed by discrete membrane staining of tumor cells (Figure 17, C).
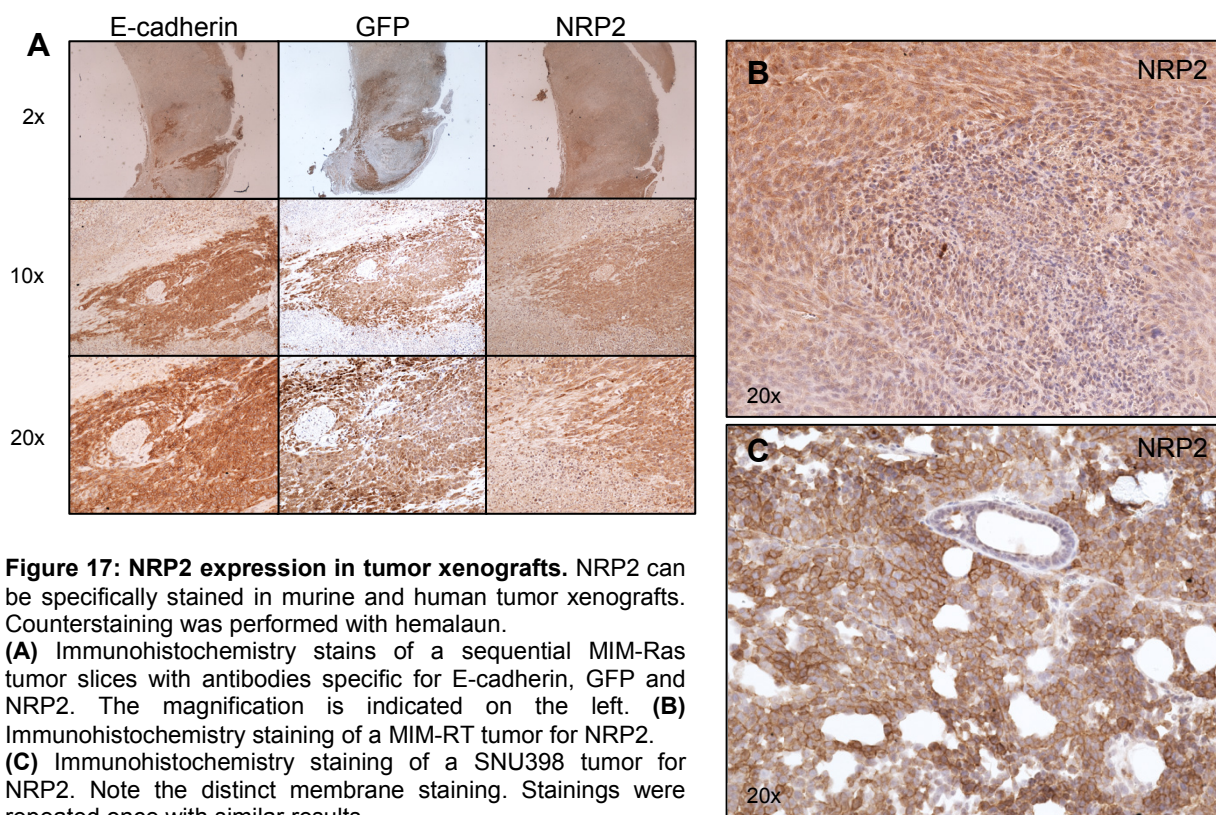

Figure 17: NRP2 expression in tumor xenografts. NRP2 can be specifically stained in murine and human tumor xenografts. Counterstaining was performed with hemalaun.
(A) Immunohistochemistry stains of a sequential MIM-Ras tumor slices with antibodies specific for E-cadherin, GFP and NRP2. The magnification is indicated on the left. **(B)** Immunohistochemistry staining of a MIM-RT tumor for NRP2. **(C)** Immunohistochemistry staining of a SNU398 tumor for NRP2. Note the distinct membrane staining. Stainings were repeated once with similar results.

4.2.6 NRP2 expression in HCC patients

Human HCC patient samples were kindly delivered as frozen tissue samples from the group of Marco Chilosi (University of Verona, Italy). RNA was extracted and transcribed to cDNA. These samples consisted of a pair of tumor tissue and adjacent liver tissue from a single patient (n=38). Quantitative PCR on NRP2 of the tumor samples revealed a significant correlation of NRP2 expression with higher tumor grade (Figure 18, A). Interestingly, the data of the adjacent tissue displayed an inverse situation in which high NRP2 expression significantly correlated with lower tumor grading (Figure 18, B). Analysis of the tumor and adjacent tissue pairs showed a significant difference between the lower and higher stage sample pairs (Figure 19).

To evaluate NRP2 as a tumor marker using publicly available data, the datasets used in the meta-analysis previously described were analyzed [191]. In the breast cancer dataset (133 patients) by Hess *et al.* [203], NRP2 showed a significant correlation (p=0.015) with higher BMN grade but not with pathological complete response (pCR) (p=0.96). There was no correlation detectable in the lung cancer dataset [204].

A

B
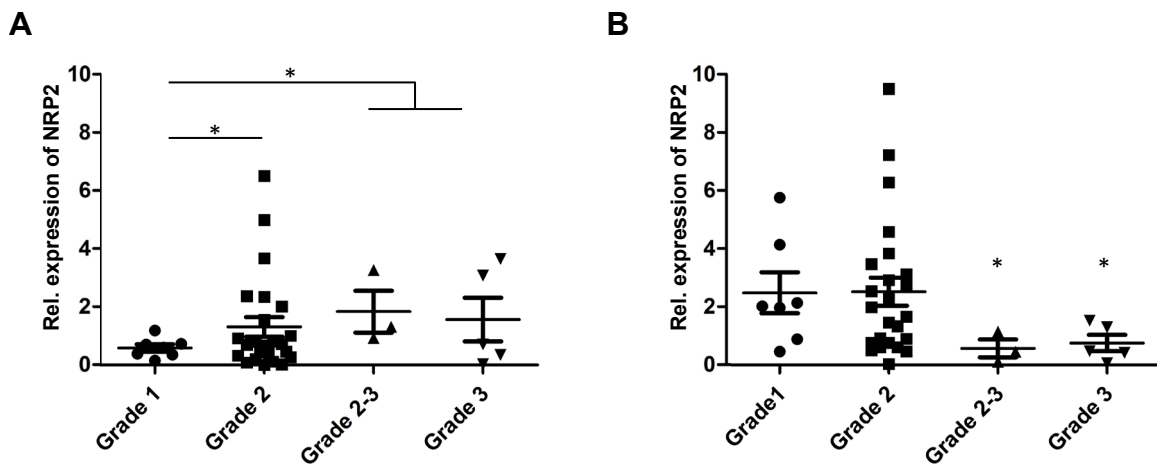
Figure 18: NRP2 expression correlates with higher grade in human HCC samples. (A) NRP2 expression in lesions of the respective grades (n=38). **(B)** NRP2 expression in adjacent tissue of the respective grades (n=38). Relative expression of NRP2 compared to RPL41 is shown. Horizontal lines indicate the respective mean values. Whiskers indicate the standard error of mean (SEM). One-sided *t*-test was used to compute p-values. Analysis was performed in triplicates. *, *P*<0.05.
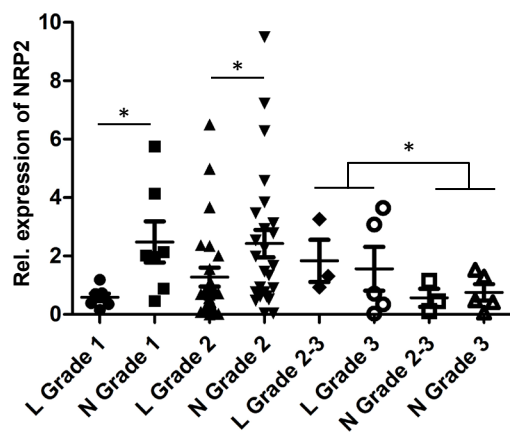
Figure 19: NRP2 expression in paired lesions and adjacent tissue. Lesions of grade 1 or 2 show less NRP2 expression than the corresponding adjacent tissue of grade 1 or 2. Vice versa, lesions of grade 2-3 and 3 display more NRP2 expression than the adjacent tissue of those grades. Relative expression of NRP2 compared to RPL41 is shown. Horizontal lines indicate the respective mean values. Whiskers indicate the standard error of mean (SEM). L, Lesion; N, Adjacent tissue. Paired *t*-test was used to compute p-values. Analysis was performed in triplicates. *, *P*<0.05.

Besides the correlation of NRP2 mRNA levels with cancer grading, we aimed to investigate the situation in terms of NRP2 protein levels. We used a liver cancer tissue microarray (LV2161, US Biomax, USA) containing 216 cases of HCC, cholangiocellular carcinoma (CCC) and normal liver tissue with clinical data. NRP2 protein levels were assessed by immunohistochemistry and scored via combining a score for the staining intensity and for the positively membrane-stained cells resulting in the NRP2 intensity score (NIS, Table 11).

Table 11: Clinicopathologic features and NRP2 presence in liver cancer tissue.

Clinicopathologic features	NRP2 protein present	NIS 0-4	NIS 5-7
Total liver cancer specimens	188/203 (92.6%)	146/203 (71.9%)	57/203 (28.1%)
HCC specimens	155/156 (99.4%)	108/156 (69.2%)	48/156 (30.8%)
CCC specimens	23/27 (85.2%)	19/27 (70.4%)	8/27 (29.6%)
Clinical stages			
Stage I (T1N0M0)	8/8 (100%)	4/8 (50%)	4/8 (50%)
Stage II (T2N0M0)	67/77 (87%)	58/77 (75.3%)	19/77 (24.7%)
Stage III (T3-4N0-1M0)	108/119 (90.8%)	79/119 (66.4%)	40/119 (33.6%)
Histopathological grades			
Grade 1	18/25 (72%)	20/25 (80%)	5/25 (20%)
Grade 2	122/129 (94.6%)	89/129 (69%)	40/129 (31%)
Grade 3	40/40 (100%)	32/40 (80%)	8/40 (20%)

NIS, NRP2 intensity score; HCC, hepatocellular carcinoma; CCC, cholangiocellular carcinoma.

As depicted in Table 11, the clinical data concerning stages and TNM stages showed an insufficient distribution with a low amount of early stage samples (Stage I, n=8) and an emphasis on late stages (Stage II and III). Therefore, we mainly focused on the grades (Grade 1-3) which also showed an emphasis on higher grades but at least a sufficient number of samples with a low grade (Grade 1, n=25). The total number of specimens is lower than the amount spotted in the microarray due to the absence of 5 cores. There were no CCC specimens of grade 3 represented on the tissue microarray. Statistical analysis of the different grades revealed that grade 1 tissues of total liver specimens displayed a significantly lower NIS than grade 2 or grade 3 tumor tissues (Table 12). This was also applicable when HCC or CCC specimens were analyzed separately. However, it was not possible to detect significant differences between grades 2 and 3.

Table 12: Statistical analysis of NRP2 intensity staining with grades of different tissue types.

	Statistical analysis (P-values)	
Total liver specimens	**Mann Whitney U-test**	**Kruskal-Wallis test**
Grade 1, 2 and 3	-	0.0290, *
Grade 1 and 2	0.0048, **	-
Grade 1 and 3	0.0130, *	-
Grade 2 and 3	0.4199	-
HCC specimens		
Grade 1, 2 and 3	-	0.1680
Grade 1 and 2	0.0388, *	-
Grade 1 and 3	0.0471, *	-
Grade 2 and 3	0.4963	-
CCC specimens		
Grade 1 and 2	0.0350, *	-

HCC, hepatocellular carcinoma; CCC, cholangiocellular carcinoma. *, $P<0.05$; **, $P<0.01$.

Examples for tissue specimens stained for NRP2 are given in Figure 20.
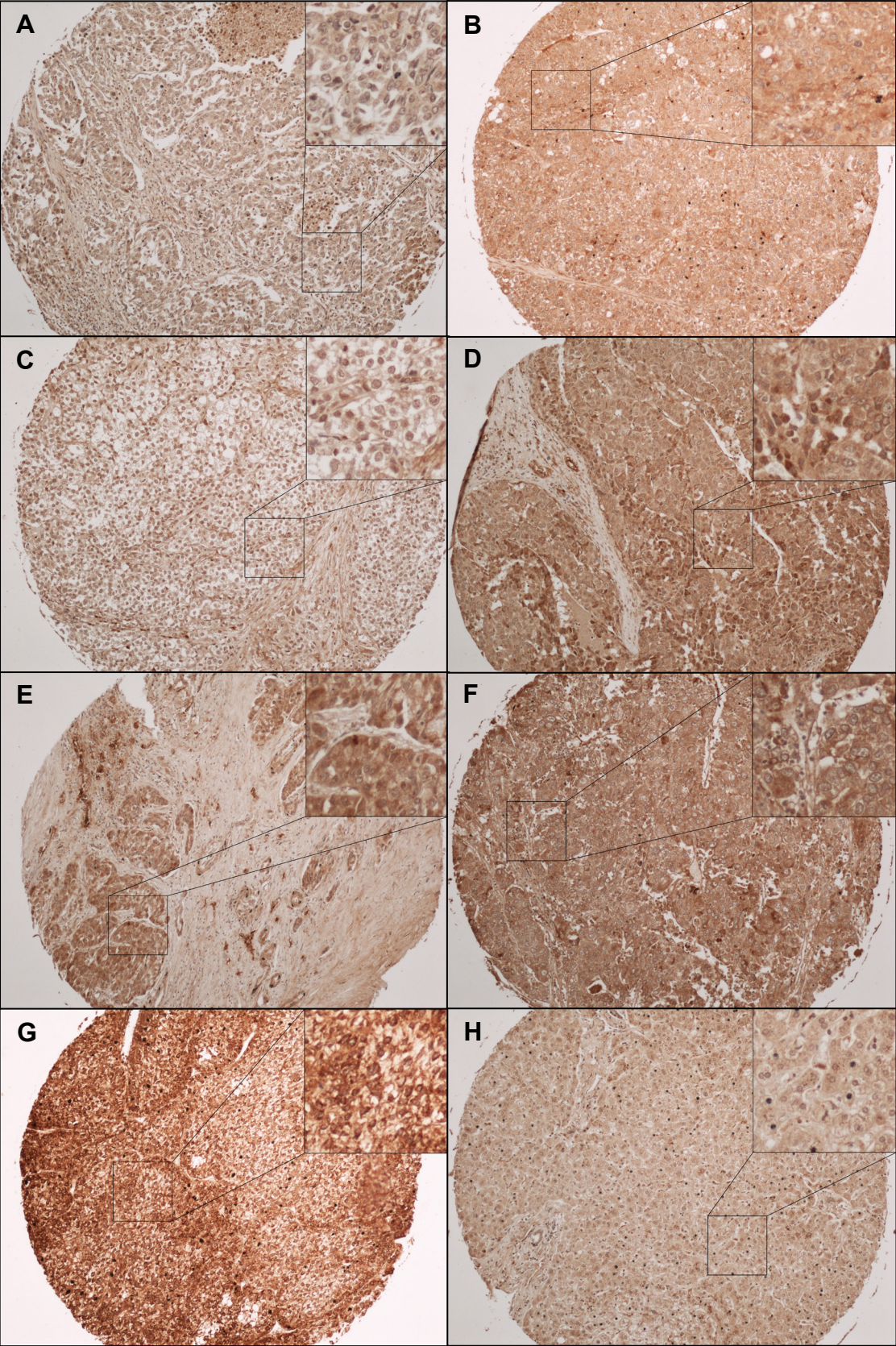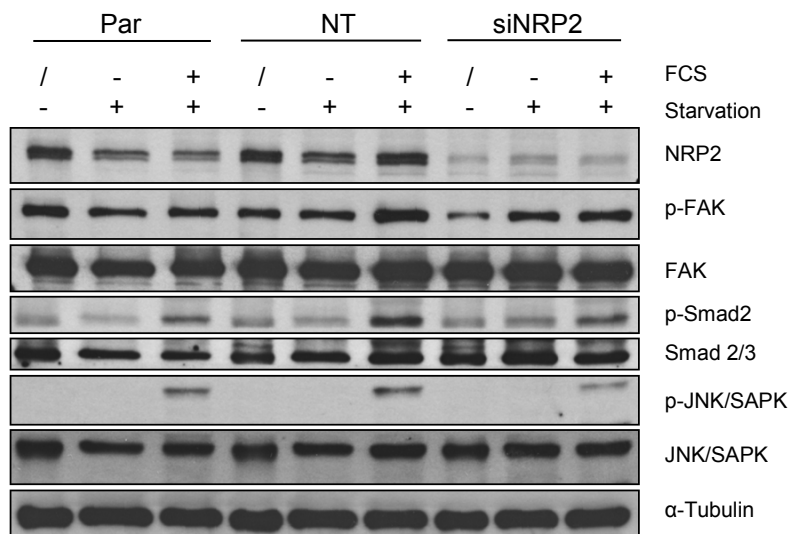

Figure 20: Liver cancer tissue cores stained for NRP2. (A) HCC sample of grade 2 with NIS=0. **(B)** Grade 2 HCC sample with NIS=2. **(C)** HCC sample of grade 1 with NIS=3. **(D)** HCC sample of grade 2 with NIS=4. **(E)** Grade 2 HCC with NIS=5. **(F)** HCC of grade 3 with NIS=6. **(G)** HCC sample of grade 2 with NIS=7. **(H)** Healthy liver tissue sample with NIS=0.

4.2.7 Investigation of NRP2 signaling

NRP2 has been shown to interact with several receptors and modulate various signaling pathways. In order to identify the signaling pathways important for the 3sp cells and possibly also for HCC in general, NRP2 expression was silenced by siRNA in 3sp cells and signaling responses analyzed. Upon serum starvation for 24h and restimulation with FCS for 30 min, a decrease of p-FAK, p-Smad2 and p-JNK/SAPK in 3sp siNRP2 cells compared to 3sp NT cells was observed (Figure 21).

Par			NT			siNRP2			
/	-	+	/	-	+	/	-	+	FCS
-	+	+	-	+	+	-	+	+	Starvation

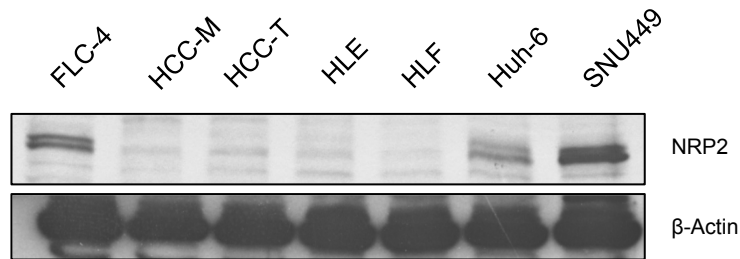| | | | | | | | | | NRP2 |
									p-FAK
									FAK
									p-Smad2
									Smad 2/3
									p-JNK/SAPK
									JNK/SAPK
									α-Tubulin

Figure 21: Knockdown of NRP2 leads to reduced phosphorylation of FAK, Smad2 and JNK/SAPK. NRP2 expression was reduced in 3sp cells using siRNA. Cells were serum starved for 24h and restimulated with Serum containing 10% FCS for 30 min. Compared to NT cells, siNRP2 treated cells showed no increase in FAK and Smad2 phosphorylation. The phosphorylation of JNK/SAPK in siNRP2 cells was lower as of Par or NT cells. Par, 3sp parental cells; NT, 3sp treated with non-target siRNA; siNRP2, 3sp treated with siRNA against NRP2. α-Tubulin was used as loading control. Experiments were repeated two times.

4.2.8 Generation of stable NRP2 overexpression/knockdown cells

The usage of siRNA enabled the initial investigation of NRP2 *in vitro*. However, in order to elucidate NRP2 functions *in vivo*, cells must be analyzed overexpressing or downregulating NRP2. To generate cells which overexpress NRP2, the cDNA of NRP2 from 3sp cells was amplified and cloned into the vector pWPI. Interestingly, it was possible to clone both isoforms NRP2a$_{(22)}$ and NRP2a$_{(17)}$ as determined by sequencing. The vector was packaged into lentiviral particles which were used to infect MIM-pRas, 3p, 3sp and PLC cells. Infected cells were sorted until they showed >95% GFP expression and are available for further research. The generation of cells downregulating NRP2 displayed some difficulties. First attempts to knockdown NRP2 expression with commercial lentiviral shRNA vectors (Genecopoeia, USA) were not successful (data not shown). Thus, other commercial shRNA constructs (Sigma-Aldrich, Austria) based on the widely used pLKO.1 vector will be used to produce MIM-Ras, MIM-RT, 3sp and possibly

SNU449 cells with reduced NRP2 expression. In addition, these shRNA constructs have been reported to successfully abrogate NRP2 expression in prostate cancer cell lines [145]. Since SNU398 cells showed disadvantages when used in *in vitro* experiments, we considered to screen other cell lines for NRP2 abundance in order to find a suitable cell model for *in vitro* and *in vivo* studies on NRP2. It was possible to screen 7 additional HCC cell lines which we kindly received from the lab of Steven Dooley (University of Heidelberg, Mannheim, Germany) for NRP2 expression (Figure 22).

Figure 22: Abundance of NRP2 in additional cell lines. NRP2 expression is present in FLC-4, Huh-6 and SNU449 cells. β-Actin was used as loading control.

NRP2 was expressed in FLC-4, Huh-6 and SNU449 cells. FLC-4 displayed less NRP2 than SNU449 cells. Huh-6 showed the lowest expression and as they are actually a hepatoblastoma and no HCC cell line they showed the least potential as an additional model for NRP2 research. Hence, FLC-4, which exhibit a more epithelial phenotype (data not shown), and SNU449, which display a more mesenchymal phenotype (data not shown), may be used as further cell models to elucidate NRP2 functions *in vitro* and *in vivo*.

4.3 Discussion

NRP2 has been reported to be involved in the regulation and modulation of many signaling pathways including VEGF, TGF-β, Hh, HGF, integrin and plexin/semaphorin signaling [31]. NRP2 expression has been found in many cancers and it was found to promote EMT in a TGF-β-dependent way in a colon cancer cell model [153]. Furthermore, we found NRP2 upregulated in several cell models of EMT [191] including the HCC EMT model 3p/3sp. As there are no studies on the role of NRP2 in HCC progression or its involvement in EMT of HCC we aimed to investigate NRP2. We were able to show that downregulation of NRP2 *in vitro* resulted in a decrease of cell migration, invasion, proliferation and adhesion. Many signaling pathways regulated by NRP2 mediate these effects such as VEGF, integrin and TGF-β signaling. As NRP2 carries out mainly enhancing effects on these pathways, the strong upregulation of NRP2 in 3sp cells suggests that in this cell system NRP2 is needed to enhance several signaling pathways to the required level. One mechanism reported to upregulate NRP2 is loss of phosphatase and tensin homolog (PTEN) and subsequent enhanced c-Jun expression [145]. The NRP2 promoter contains c-Jun binding sites which positively enhances NRP2 expression upon promoter binding in prostate cancer cells [145]. This effect is also modulated by JNK1 expression and activity. Nevertheless, this upregulation of NRP2 expression is independent from another reported mechanism where the nuclear receptor subfamily 2, group F, member 2 (NR2F2) modulates NRP2 expression [205]. As PTEN loss is a common mutation in many cancers, such a pathway, or a similar one, may be able to increase NRP2 expression.

Interestingly, NRP2 is also implicated in EMT. The cooperation of NRP2 with TβRI promoted EMT in a colon cancer cell line [153]. We were able to find NRP2 upregulated in several EMT cell models, as mentioned previously, as well as in the mesenchymal 3sp cell line. In some of these cell lines, EMT was induced by TGF-β. For example, the lung adenocarcinoma cell line A549 undergoes EMT (A549T) after TGF-β treatment for 72 h [206]. Hence, it seemed of great interest if active TGF-β signaling could increase the expression of NRP2. However, it was not possible to find a connection of TGF-β-signaling which could increase NRP2 expression.

Investigation of microarray data of 3p/3sp, MIM-Ras/MIM-RT and A549/A549T cells as well as preliminary Western blot analysis (data not shown) revealed a 10-fold increase of c-Jun in A549T cells but not in 3sp or MIM-RT cells. As A549T cells were the only human transient EMT model analyzed, this upregulation might only occur in human reversible EMT models and not in models where EMT was permanent, such as 3sp cells, or stable over a longer period of time, such as the MIM-RT cells. However, there were no changes in PTEN, JNK1 and NR2F2 expression in A549T cells. The upregulation of NRP2 and c-Jun in this cell model may not occur over the PTEN-c-Jun axis or the transcription factor NR2F2 which suggests another responsible

pathway. In addition, there were no significant differences in PTEN, c-Jun, JNK1 and NR2F2 protein levels in 3sp and MIM-RT cells. It remains unknown whether NRP2 was upregulated in those cell lines by c-Jun or NR2F2 during EMT with other pathways taking over after the completion of EMT or other still unknown pathways were responsible for NRP2 upregulation. In conclusion, it seems that NRP2 is upregulated during the (TGF-β-induced) EMT of some cells lines but the reason why NRP2 was upregulated in those EMT cell lines remains unknown.

Also unknown remains if the different isoforms of NRP2 have different roles in cancer progression and EMT. The isoforms NRP2a and NRP2b seem to be preferentially expressed either in the intestine and several organs or in the heart and skeletal muscle, respectively [53]. However, this discrimination is not black and white, since several isoforms are frequently present [31]. Apart from this distribution not much is known about the different functions of NRP2a and NRP2b. NRP2b, however, lacks the c-terminal domain of NRP2a which confers signaling via synectin [30]. The short cytoplasmic tail of NRP2b might not be able to transduce any signaling via direct protein-protein interaction and possibly remains as a sole co-receptor modulating other signaling pathways. In addition to the two major isoforms, both NRP2a and NRP2b exist with different insertions of amino acids [30]. Depending on the splicing, 5 amino acids are introduced in NRP2a or NRP2b or not. Not much is known on the effect of these 5 amino acids if they are present or not. It has been suggested that they might alter their ability to associate with VEGFR2 or to homodimerize [30]. However, it is problematic to address questions specific to a single isoform as often several are expressed. We were able to detect both NRP2a and NRP2b in several HCC cell lines using isoform specific PCR primers. NRP2a is, in contrast to NRP2b, able to confer signaling via synectin. Therefore, we aimed to clone NRP2a from 3sp cells. Interestingly, it was possible to clone both $NRP2a_{(22)}$ and $NRP2a_{(17)}$. The application of the overexpression vector of a single NRP2a isoform might enable further research on the role of these isoforms in a NRP2-lacking background.

Expression of NRP2 has been found in many cancers but the situation in HCC has not been addressed so far [30,31]. Therefore, we aimed to elucidate if NRP2 might be important in HCC. Using publicly available microarray data, we determined a significant correlation of NRP2 mRNA levels with BMN grading in breast cancer patients. Furthermore, we were able to find a significant correlation of high NRP2 mRNA levels with higher grading in tumor samples from HCC patients. Interestingly, there was an inverse correlation in adjacent tumor tissue samples as high NRP2 expression significantly correlated with lower grading. This might be explained by the increased tumor surveillance and the resulting presence of several immune cells which express NRP2. In addition, healthy hepatocytes might express NRP2 since NRP2 expression has been reported in the liver. Higher tumor stages would possibly not display such high levels of immune cells as well as less healthy hepatocytes in the adjacent tissue. However, the

verification of this hypothesis would need a closer immunohistological investigation of the sampled tumor tissue. Since this analysis of the tissue was not possible, we addressed this question using a liver cancer tissue microarray. Notably, it was possible to find significant correlation of NRP2 protein expression with higher grading. This was applicable for all analyzed liver cancer types as well as for HCC and CCC separately. Together, NRP2 mRNA and protein presence seems to associate with higher grading in HCC as well as in other cancers.

The impact of NRP2 on the signaling of various pathways has been mentioned above. As there is not much known about the signaling of NRP2 in HCC, we elucidated some of its functions *in vitro*. We were able to identify a reduction of FAK phosphorylation upon NRP2 knockdown in 3sp cells. FAK is involved in several pathways including VEGF and integrin signaling (Figure 7). It is also involved in the formation of focal adhesions and assembly of signaling complexes [207]. Whether the knockdown of NRP2 exhibits a direct effect on FAK phosphorylation or through the downregulation of an upstream signal remains unclear. However, the lower level of FAK phosphorylation in NRP2 knockdown cells might contribute to the reduced cell migration, invasion, proliferation and adhesion. In addition, reduced phosphorylation of Smad2 and JNK/SAPK was observed upon NRP2 knockdown in 3sp cells. Whereas Smad2 phosphorylation represents the canonical TGF-β signaling, JNK/SAPK phosphorylation indicates involvement of non-canonical TGF-β signaling. Other parts of non-canonical TGF-β signaling such as NFκB, ERK1/2 and p38, however, remained unchanged during reduced NRP2 expression (data not shown). Thus, NRP2 seems to affect both canonical and non-canonical TGF-β signaling in 3sp cells.

It has been reported that NRP2 is not able to bind to fibronectin [30,31]. Notably, we were able to measure a decrease of cell adhesion when cells treated with siRNA against NRP2 were placed upon ECM proteins such as collagen I, laminin and fibronectin. This suggests that NRP2 might not be directly involved in cell adhesion but maybe alters other proteins necessary to facilitate adhesion to those proteins. The decrease of FAK phosphorylation in cells with reduced NRP2 expression could be a possible explanation. However, more investigation should be performed to clarify the role of NRP2 in cell adhesion.

In conclusion, we were able to demonstrate that several HCC cells express NRP2 isoforms and that NRP2 knockdown reduced cell migration, invasion, proliferation and adhesion *in vitro*. NRP2 expression correlated with higher grading in HCC and breast cancer patient samples. Moreover, NRP2 is upregulated during EMT of several HCC cells and other cell lines. The involvement of NRP2 in these processes strengthens the need for a successful therapeutic targeting in cancer patients. However, due to the wide effects of NRP2 on various pathways, additional research has to elucidate the complex functions of NRP2 in order to find specific druggable targets.

4.4 Future perspectives

This thesis sheds light on the cellular functions of NRP2 in human HCC cell lines *in vitro* as well as the situation in HCC patient samples. However, there is still work to be done on the *in vivo* situation. Cell lines transfected with shRNA against NRP2 as well as overexpression constructs of the isoforms $NRP2a_{(22)}$ and $NRP2a_{(17)}$ have been generated during this study. The use of these cells might help to verify the data gathered from siRNA treated cells and it allows investigating overexpression of NRP2 in cells without endogenous NRP2. It remains questionable whether NRP2 overexpression would induce EMT in such cells as the knockdown did not show any signs of MET in mesenchymal cells. However, due to the known cooperation of NRP2 with TGF-β signaling, NRP2 overexpression might be able to alter or even enhance the EMT of MIM-Ras cells to MIM-RT cells under TGF-β treatment. The usage of two different overexpressed NRP2 isoforms also enables to identify possible differences in their functions regarding EMT or cancer progression. In addition, tumor formation and metastasis of NRP2 knockdown or overexpressing cells can be investigated in mouse xenografts. In order to gain insight on the binding partners of NRP2 in the analyzed cell lines, one could immunoprecipitate NRP2 and investigate the proteins via mass spectrometry. The application of a phospho-protein array will also point to possible binding partners and affected signaling pathways. The influence of c-Jun and NR2F2 on NRP2 expression remains unclear in the analyzed cell lines. Application of siRNA against c-Jun or JNK1 would enable to detect any NRP2 regulation by these proteins. Furthermore, time course analysis with qPCR of MIM-Ras and A549 cell under TGF-β treatment would allow more detailed insights into the change of NRP2, c-Jun, PTEN, JNK1 and NR2F2 expression levels during EMT. The developmental and normal cellular functions for NRP2 have been established for the nervous system and blood vessel formation. As the NRP2 knockout mouse displayed abnormalities in these systems and no differences in other organs, the role of NRP2 in the liver remains unknown. In order to address questions regarding liver regenerative and protective functions of NRP2, the available NRP2-floxed mouse (Jackson Laboratory, USA) would have to be crossed with an AFP-Cre mouse, which would generate a liver-specific knockout of NRP2. Those mice could be challenged by, for instance, CCl_4 induced fibrosis, partial hepatectomy or endogenous tumor formation models. Moreover, the liver-specific NRP2 knockout mouse could be crossed with the $p19^{ARF-/-}$ mouse to produce immortalized hepatocytes lacking NRP2. Together with the already gathered information on NRP2 in HCC, these suggestions would clarify the role of NRP2 in EMT and in HCC progression.

5 References

1. Alberts B, *et al*. (2008) Molecular Biology of the Cell. Garland Science Taylor&Francis Group. pp. 1131-1169.
2. Stokes DL (2007) Desmosomes from a structural perspective. Curr Opin Cell Biol 19: 565-571.
3. Berx G, van Roy F (2009) Involvement of members of the cadherin superfamily in cancer. Cold Spring Harb Perspect Biol 1: a003129.
4. Kalluri R, Weinberg RA (2009) The basics of epithelial-mesenchymal transition. J Clin Invest 119: 1420-1428.
5. Thiery JP, Acloque H, Huang RY, Nieto MA (2009) Epithelial-mesenchymal transitions in development and disease. Cell 139: 871-890.
6. Birchmeier C, Birchmeier W, Gherardi E, Vande Woude GF (2003) Met, metastasis, motility and more. Nat Rev Mol Cell Biol 4: 915-925.
7. Larue L, Bellacosa A (2005) Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways. Oncogene 24: 7443-7454.
8. Wang Z, Li Y, Kong D, Sarkar FH (2010) The role of Notch signaling pathway in epithelial-mesenchymal transition (EMT) during development and tumor aggressiveness. Curr Drug Targets 11: 745-751.
9. Fuxe J, Vincent T, Garcia de Herreros A (2010) Transcriptional crosstalk between TGF-beta and stem cell pathways in tumor cell invasion: role of EMT promoting Smad complexes. Cell Cycle 9: 2363-2374.
10. Thiery JP (2002) Epithelial-mesenchymal transitions in tumour progression. Nat Rev Cancer 2: 442-454.
11. Peinado H, Portillo F, Cano A (2004) Transcriptional regulation of cadherins during development and carcinogenesis. Int J Dev Biol 48: 365-375.
12. Berx G, Becker KF, Hofler H, van Roy F (1998) Mutations of the human E-cadherin (CDH1) gene. Hum Mutat 12: 226-237.
13. Zeisberg M, Neilson EG (2009) Biomarkers for epithelial-mesenchymal transitions. J Clin Invest 119: 1429-1437.
14. Thiery JP, Sleeman JP (2006) Complex networks orchestrate epithelial-mesenchymal transitions. Nat Rev Mol Cell Biol 7: 131-142.
15. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646-674.
16. van Zijl F, Krupitza G, Mikulits W (2011) Initial steps of metastasis: cell invasion and endothelial transmigration. Mutat Res 728: 23-34.
17. Ansieau S, Bastid J, Doreau A, Morel AP, Bouchet BP, *et al*. (2008) Induction of EMT by twist proteins as a collateral effect of tumor-promoting inactivation of premature senescence. Cancer Cell 14: 79-89.
18. Yang AD, Fan F, Camp ER, van Buren G, Liu W, *et al*. (2006) Chronic oxaliplatin resistance induces epithelial-to-mesenchymal transition in colorectal cancer cell lines. Clin Cancer Res 12: 4147-4153.
19. Jemal A, Bray F, Center MM, Ferlay J, Ward E, *et al*. (2011) Global cancer statistics. CA Cancer J Clin 61: 69-90.
20. Parkin DM (2001) Global cancer statistics in the year 2000. Lancet Oncol 2: 533-543.
21. El-Serag HB, Rudolph KL (2007) Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology 132: 2557-2576.
22. Thorgeirsson SS, Grisham JW (2002) Molecular pathogenesis of human hepatocellular carcinoma. Nat Genet 31: 339-346.
23. Huynh H (2010) Molecularly targeted therapy in hepatocellular carcinoma. Biochem Pharmacol 80: 550-560.
24. Llovet JM, Burroughs A, Bruix J (2003) Hepatocellular carcinoma. Lancet 362: 1907-1917.
25. Sun VC, Sarna L (2008) Symptom management in hepatocellular carcinoma. Clin J Oncol Nurs 12: 759-766.
26. Zhang YY, Xia HH (2008) Novel therapeutic approaches for hepatocellulcar carcinoma: fact and fiction. World J Gastroenterol 14: 1641-1642.
27. Zhai B, Yan HX, Liu SQ, Chen L, Wu MC, *et al*. (2008) Reduced expression of E-cadherin/catenin complex in hepatocellular carcinomas. World J Gastroenterol 14: 5665-5673.
28. Giannelli G, Bergamini C, Fransvea E, Sgarra C, Antonaci S (2005) Laminin-5 with transforming growth factor-beta1 induces epithelial to mesenchymal transition in hepatocellular carcinoma. Gastroenterology 129: 1375-1383.
29. Yu H, Pardoll D, Jove R (2009) STATs in cancer inflammation and immunity: a leading role for STAT3. Nat Rev Cancer 9: 798-809.
30. Pellet-Many C, Frankel P, Jia H, Zachary I (2008) Neuropilins: structure, function and role in disease. Biochem J 411: 211-226.
31. Prud'homme GJ, Glinka Y (2012) Neuropilins are multifunctional coreceptors involved in tumor initiation, growth, metastasis and immunity. Oncotarget 3: 921-939.
32. Herzog B, Pellet-Many C, Britton G, Hartzoulakis B, Zachary IC (2011) VEGF binding to NRP1 is essential for VEGF stimulation of endothelial cell migration, complex formation between NRP1 and VEGFR2, and signaling via FAK Tyr407 phosphorylation. Mol Biol Cell 22: 2766-2776.
33. Takagi S, Tsuji T, Amagai T, Takamatsu T, Fujisawa H (1987) Specific cell surface labels in the visual centers of Xenopus laevis tadpole identified using monoclonal antibodies. Dev Biol 122: 90-100.
34. Takagi S, Hirata T, Agata K, Mochii M, Eguchi G, *et al*. (1991) The A5 antigen, a candidate for the neuronal recognition molecule, has homologies to complement components and coagulation factors. Neuron 7: 295-307.
35. Fujisawa H, Takagi S, Hirata T (1995) Growth-associated expression of a membrane protein, neuropilin, in Xenopus optic nerve fibers. Dev Neurosci 17: 343-349.

36. Gu C, Rodriguez ER, Reimert DV, Shu T, Fritzsch B, *et al.* (2003) Neuropilin-1 conveys semaphorin and VEGF signaling during neural and cardiovascular development. Dev Cell 5: 45-57.
37. Kitsukawa T, Shimizu M, Sanbo M, Hirata T, Taniguchi M, *et al.* (1997) Neuropilin-semaphorin III/D-mediated chemorepulsive signals play a crucial role in peripheral nerve projection in mice. Neuron 19: 995-1005.
38. Kawasaki T, Kitsukawa T, Bekku Y, Matsuda Y, Sanbo M, *et al.* (1999) A requirement for neuropilin-1 in embryonic vessel formation. Development 126: 4895-4902.
39. Lee P, Goishi K, Davidson AJ, Mannix R, Zon L, *et al.* (2002) Neuropilin-1 is required for vascular development and is a mediator of VEGF-dependent angiogenesis in zebrafish. Proc Natl Acad Sci U S A 99: 10470-10475.
40. Chen H, Chedotal A, He Z, Goodman CS, Tessier-Lavigne M (1997) Neuropilin-2, a novel member of the neuropilin family, is a high affinity receptor for the semaphorins Sema E and Sema IV but not Sema III. Neuron 19: 547-559.
41. Guttmann-Raviv N, Shraga-Heled N, Varshavsky A, Guimaraes-Sternberg C, Kessler O, *et al.* (2007) Semaphorin-3A and semaphorin-3F work together to repel endothelial cells and to inhibit their survival by induction of apoptosis. J Biol Chem 282: 26294-26305.
42. Cai H, Reed RR (1999) Cloning and characterization of neuropilin-1-interacting protein: a PSD-95/Dlg/ZO-1 domain-containing protein that interacts with the cytoplasmic domain of neuropilin-1. J Neurosci 19: 6519-6527.
43. Vander Kooi CW, Jusino MA, Perman B, Neau DB, Bellamy HD, *et al.* (2007) Structural basis for ligand and heparin binding to neuropilin B domains. Proc Natl Acad Sci U S A 104: 6152-6157.
44. Parker MW, Xu P, Li X, Vander Kooi CW (2012) Structural basis for selective vascular endothelial growth factor-A (VEGF-A) binding to neuropilin-1. J Biol Chem 287: 11082-11089.
45. Teesalu T, Sugahara KN, Kotamraju VR, Ruoslahti E (2009) C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration. Proc Natl Acad Sci U S A 106: 16157-16162.
46. Lee CC, Kreusch A, McMullan D, Ng K, Spraggon G (2003) Crystal structure of the human neuropilin-1 b1 domain. Structure 11: 99-108.
47. Chen H, He Z, Bagri A, Tessier-Lavigne M (1998) Semaphorin-neuropilin interactions underlying sympathetic axon responses to class III semaphorins. Neuron 21: 1283-1290.
48. Nakamura F, Tanaka M, Takahashi T, Kalb RG, Strittmatter SM (1998) Neuropilin-1 extracellular domains mediate semaphorin D/III-induced growth cone collapse. Neuron 21: 1093-1100.
49. Roth L, Nasarre C, Dirrig-Grosch S, Aunis D, Cremel G, *et al.* (2008) Transmembrane domain interactions control biological functions of neuropilin-1. Mol Biol Cell 19: 646-654.
50. Nasarre C, Roth M, Jacob L, Roth L, Koncina E, *et al.* (2010) Peptide-based interference of the transmembrane domain of neuropilin-1 inhibits glioma growth in vivo. Oncogene 29: 2381-2392.
51. Shintani Y, Takashima S, Asano Y, Kato H, Liao Y, *et al.* (2006) Glycosaminoglycan modification of neuropilin-1 modulates VEGFR2 signaling. EMBO J 25: 3045-3055.
52. Fukahi K, Fukasawa M, Neufeld G, Itakura J, Korc M (2004) Aberrant expression of neuropilin-1 and -2 in human pancreatic cancer cells. Clin Cancer Res 10: 581-590.
53. Rossignol M, Gagnon ML, Klagsbrun M (2000) Genomic organization of human neuropilin-1 and neuropilin-2 genes: identification and distribution of splice variants and soluble isoforms. Genomics 70: 211-222.
54. Rossignol M, Beggs AH, Pierce EA, Klagsbrun M (1999) Human neuropilin-1 and neuropilin-2 map to 10p12 and 2q34, respectively. Genomics 57: 459-460.
55. Walz A, Rodriguez I, Mombaerts P (2002) Aberrant sensory innervation of the olfactory bulb in neuropilin-2 mutant mice. J Neurosci 22: 4025-4035.
56. Yuan L, Moyon D, Pardanaud L, Breant C, Karkkainen MJ, *et al.* (2002) Abnormal lymphatic vessel development in neuropilin 2 mutant mice. Development 129: 4797-4806.
57. Chen H, Bagri A, Zupicich JA, Zou Y, Stoeckli E, *et al.* (2000) Neuropilin-2 regulates the development of selective cranial and sensory nerves and hippocampal mossy fiber projections. Neuron 25: 43-56.
58. Giger RJ, Cloutier JF, Sahay A, Prinjha RK, Levengood DV, *et al.* (2000) Neuropilin-2 is required in vivo for selective axon guidance responses to secreted semaphorins. Neuron 25: 29-41.
59. Polleux F, Morrow T, Ghosh A (2000) Semaphorin 3A is a chemoattractant for cortical apical dendrites. Nature 404: 567-573.
60. Takashima S, Kitakaze M, Asakura M, Asanuma H, Sanada S, *et al.* (2002) Targeting of both mouse neuropilin-1 and neuropilin-2 genes severely impairs developmental yolk sac and embryonic angiogenesis. Proc Natl Acad Sci U S A 99: 3657-3662.
61. Yu HH, Houart C, Moens CB (2004) Cloning and embryonic expression of zebrafish neuropilin genes. Gene Expr Patterns 4: 371-378.
62. Bovenkamp DE, Goishi K, Bahary N, Davidson AJ, Zhou Y, *et al.* (2004) Expression and mapping of duplicate neuropilin-1 and neuropilin-2 genes in developing zebrafish. Gene Expr Patterns 4: 361-370.
63. Sahay A, Kim CH, Sepkuty JP, Cho E, Huganir RL, *et al.* (2005) Secreted semaphorins modulate synaptic transmission in the adult hippocampus. J Neurosci 25: 3613-3620.
64. Tran TS, Kolodkin AL, Bharadwaj R (2007) Semaphorin regulation of cellular morphology. Annu Rev Cell Dev Biol 23: 263-292.
65. Herzog Y, Guttmann-Raviv N, Neufeld G (2005) Segregation of arterial and venous markers in subpopulations of blood islands before vessel formation. Dev Dyn 232: 1047-1055.

80

66. Martyn U, Schulte-Merker S (2004) Zebrafish neuropilins are differentially expressed and interact with vascular endothelial growth factor during embryonic vascular development. Dev Dyn 231: 33-42.

67. Giger RJ, Urquhart ER, Gillespie SK, Levengood DV, Ginty DD, *et al*. (1998) Neuropilin-2 is a receptor for semaphorin IV: insight into the structural basis of receptor function and specificity. Neuron 21: 1079-1092.

68. Miao HQ, Soker S, Feiner L, Alonso JL, Raper JA, *et al*. (1999) Neuropilin-1 mediates collapsin-1/semaphorin III inhibition of endothelial cell motility: functional competition of collapsin-1 and vascular endothelial growth factor-165. J Cell Biol 146: 233-242.

69. Bielenberg DR, Hida Y, Shimizu A, Kaipainen A, Kreuter M, *et al*. (2004) Semaphorin 3F, a chemorepulsant for endothelial cells, induces a poorly vascularized, encapsulated, nonmetastatic tumor phenotype. J Clin Invest 114: 1260-1271.

70. Serini G, Valdembri D, Zanivan S, Morterra G, Burkhardt C, *et al*. (2003) Class 3 semaphorins control vascular morphogenesis by inhibiting integrin function. Nature 424: 391-397.

71. Kessler O, Shraga-Heled N, Lange T, Gutmann-Raviv N, Sabo E, *et al*. (2004) Semaphorin-3F is an inhibitor of tumor angiogenesis. Cancer Res 64: 1008-1015.

72. Pan Q, Chanthery Y, Liang WC, Stawicki S, Mak J, *et al*. (2007) Blocking neuropilin-1 function has an additive effect with anti-VEGF to inhibit tumor growth. Cancer Cell 11: 53-67.

73. Soker S, Takashima S, Miao HQ, Neufeld G, Klagsbrun M (1998) Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. Cell 92: 735-745.

74. Murga M, Fernandez-Capetillo O, Tosato G (2005) Neuropilin-1 regulates attachment in human endothelial cells independently of vascular endothelial growth factor receptor-2. Blood 105: 1992-1999.

75. Gerhardt H, Ruhrberg C, Abramsson A, Fujisawa H, Shima D, *et al*. (2004) Neuropilin-1 is required for endothelial tip cell guidance in the developing central nervous system. Dev Dyn 231: 503-509.

76. Shimizu M, Murakami Y, Suto F, Fujisawa H (2000) Determination of cell adhesion sites of neuropilin-1. J Cell Biol 148: 1283-1293.

77. Dzionek A, Inagaki Y, Okawa K, Nagafune J, Rock J, *et al*. (2002) Plasmacytoid dendritic cells: from specific surface markers to specific cellular functions. Hum Immunol 63: 1133-1148.

78. Grage-Griebenow E, Loseke S, Kauth M, Gehlhar K, Zawatzky R, *et al*. (2007) Anti-BDCA-4 (neuropilin-1) antibody can suppress virus-induced IFN-alpha production of plasmacytoid dendritic cells. Immunol Cell Biol 85: 383-390.

79. Ghez D, Lepelletier Y, Lambert S, Fourneau JM, Blot V, *et al*. (2006) Neuropilin-1 is involved in human T-cell lymphotropic virus type 1 entry. J Virol 80: 6844-6854.

80. Rey-Gallardo A, Escribano C, Delgado-Martin C, Rodriguez-Fernandez JL, Gerardy-Schahn R, *et al*. (2010) Polysialylated neuropilin-2 enhances human dendritic cell migration through the basic C-terminal region of CCL21. Glycobiology 20: 1139-1146.

81. Rey-Gallardo A, Delgado-Martin C, Gerardy-Schahn R, Rodriguez-Fernandez JL, Vega MA (2011) Polysialic acid is required for neuropilin-2a/b-mediated control of CCL21-driven chemotaxis of mature dendritic cells and for their migration in vivo. Glycobiology 21: 655-662.

82. Rohm B, Ottemeyer A, Lohrum M, Puschel AW (2000) Plexin/neuropilin complexes mediate repulsion by the axonal guidance signal semaphorin 3A. Mech Dev 93: 95-104.

83. Tamagnone L, Artigiani S, Chen H, He Z, Ming GI, *et al*. (1999) Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. Cell 99: 71-80.

84. Cheng HJ, Bagri A, Yaron A, Stein E, Pleasure SJ, *et al*. (2001) Plexin-A3 mediates semaphorin signaling and regulates the development of hippocampal axonal projections. Neuron 32: 249-263.

85. Antipenko A, Himanen JP, van Leyen K, Nardi-Dei V, Lesniak J, *et al*. (2003) Structure of the semaphorin-3A receptor binding module. Neuron 39: 589-598.

86. Castellani V, De Angelis E, Kenwrick S, Rougon G (2002) Cis and trans interactions of L1 with neuropilin-1 control axonal responses to semaphorin 3A. EMBO J 21: 6348-6357.

87. Cao Y, Wang L, Nandy D, Zhang Y, Basu A, *et al*. (2008) Neuropilin-1 upholds dedifferentiation and propagation phenotypes of renal cell carcinoma cells by activating Akt and sonic hedgehog axes. Cancer Res 68: 8667-8672.

88. Hillman RT, Feng BY, Ni J, Woo WM, Milenkovic L, *et al*. (2011) Neuropilins are positive regulators of Hedgehog signal transduction. Genes Dev 25: 2333-2346.

89. Beck B, Driessens G, Goossens S, Youssef KK, Kuchnio A, *et al*. (2011) A vascular niche and a VEGF-Nrp1 loop regulate the initiation and stemness of skin tumours. Nature 478: 399-403.

90. Hamerlik P, Lathia JD, Rasmussen R, Wu Q, Bartkova J, *et al*. (2012) Autocrine VEGF-VEGFR2-Neuropilin-1 signaling promotes glioma stem-like cell viability and tumor growth. J Exp Med 209: 507-520.

91. Whitaker GB, Limberg BJ, Rosenbaum JS (2001) Vascular endothelial growth factor receptor-2 and neuropilin-1 form a receptor complex that is responsible for the differential signaling potency of VEGF(165) and VEGF(121). J Biol Chem 276: 25520-25531.

92. Soker S, Miao HQ, Nomi M, Takashima S, Klagsbrun M (2002) VEGF165 mediates formation of complexes containing VEGFR-2 and neuropilin-1 that enhance VEGF165-receptor binding. J Cell Biochem 85: 357-368.

93. Mac Gabhann F, Popel AS (2005) Differential binding of VEGF isoforms to VEGF receptor 2 in the presence of neuropilin-1: a computational model. Am J Physiol Heart Circ Physiol 288: H2851-2860.

94. Oh H, Takagi H, Otani A, Koyama S, Kemmochi S, *et al*. (2002) Selective induction of neuropilin-1 by vascular endothelial growth factor (VEGF): a mechanism contributing to VEGF-induced angiogenesis. Proc Natl Acad Sci U S A 99: 383-388.

95. Wang L, Zeng H, Wang P, Soker S, Mukhopadhyay D (2003) Neuropilin-1-mediated vascular permeability factor/vascular endothelial growth factor-dependent endothelial cell migration. J Biol Chem 278: 48848-48860.

96. Uniewicz KA, Fernig DG (2008) Neuropilins: a versatile partner of extracellular molecules that regulate development and disease. Front Biosci 13: 4339-4360.

97. Fuh G, Garcia KC, de Vos AM (2000) The interaction of neuropilin-1 with vascular endothelial growth factor and its receptor flt-1. J Biol Chem 275: 26690-26695.

98. Jia H, Bagherzadeh A, Hartzoulakis B, Jarvis A, Lohr M, *et al*. (2006) Characterization of a bicyclic peptide neuropilin-1 (NP-1) antagonist (EG3287) reveals importance of vascular endothelial growth factor exon 8 for NP-1 binding and role of NP-1 in KDR signaling. J Biol Chem 281: 13493-13502.

99. Karpanen T, Heckman CA, Keskitalo S, Jeltsch M, Ollila H, *et al*. (2006) Functional interaction of VEGF-C and VEGF-D with neuropilin receptors. FASEB J 20: 1462-1472.

100. Favier B, Alam A, Barron P, Bonnin J, Laboudie P, *et al*. (2006) Neuropilin-2 interacts with VEGFR-2 and VEGFR-3 and promotes human endothelial cell survival and migration. Blood 108: 1243-1250.

101. Adams RH, Alitalo K (2007) Molecular regulation of angiogenesis and lymphangiogenesis. Nat Rev Mol Cell Biol 8: 464-478.

102. Gluzman-Poltorak Z, Cohen T, Shibuya M, Neufeld G (2001) Vascular endothelial growth factor receptor-1 and neuropilin-2 form complexes. J Biol Chem 276: 18688-18694.

103. Prud'homme GJ (2007) Pathobiology of transforming growth factor beta in cancer, fibrosis and immunologic disease, and therapeutic considerations. Lab Invest 87: 1077-1091.

104. Zhang YE (2009) Non-Smad pathways in TGF-beta signaling. Cell Res 19: 128-139.

105. Parvani JG, Taylor MA, Schiemann WP (2011) Noncanonical TGF-beta signaling during mammary tumorigenesis. J Mammary Gland Biol Neoplasia 16: 127-146.

106. Wang R, Zhu J, Dong X, Shi M, Lu C, *et al*. (2012) GARP regulates the bioavailability and activation of TGFbeta. Mol Biol Cell 23: 1129-1139.

107. Glinka Y, Prud'homme GJ (2008) Neuropilin-1 is a receptor for transforming growth factor beta-1, activates its latent form, and promotes regulatory T cell activity. J Leukoc Biol 84: 302-310.

108. Young GD, Murphy-Ullrich JE (2004) Molecular interactions that confer latency to transforming growth factor-beta. J Biol Chem 279: 38032-38039.

109. Massague J (2012) TGFbeta signalling in context. Nat Rev Mol Cell Biol 13: 616-630.

110. Levy L, Hill CS (2006) Alterations in components of the TGF-beta superfamily signaling pathways in human cancer. Cytokine Growth Factor Rev 17: 41-58.

111. Komuro A, Yashiro M, Iwata C, Morishita Y, Johansson E, *et al*. (2009) Diffuse-type gastric carcinoma: progression, angiogenesis, and transforming growth factor beta signaling. J Natl Cancer Inst 101: 592-604.

112. Datto MB, Li Y, Panus JF, Howe DJ, Xiong Y, *et al*. (1995) Transforming growth factor beta induces the cyclin-dependent kinase inhibitor p21 through a p53-independent mechanism. Proc Natl Acad Sci U S A 92: 5545-5549.

113. Yagi K, Furuhashi M, Aoki H, Goto D, Kuwano H, *et al*. (2002) c-myc is a downstream target of the Smad pathway. J Biol Chem 277: 854-861.

114. Deheuninck J, Luo K (2009) Ski and SnoN, potent negative regulators of TGF-beta signaling. Cell Res 19: 47-57.

115. Jang CW, Chen CH, Chen CC, Chen JY, Su YH, *et al*. (2002) TGF-beta induces apoptosis through Smad-mediated expression of DAP-kinase. Nat Cell Biol 4: 51-58.

116. Kiyono K, Suzuki HI, Matsuyama H, Morishita Y, Komuro A, *et al*. (2009) Autophagy is activated by TGF-beta and potentiates TGF-beta-mediated growth inhibition in human hepatocellular carcinoma cells. Cancer Res 69: 8844-8852.

117. Battegay EJ, Raines EW, Seifert RA, Bowen-Pope DF, Ross R (1990) TGF-beta induces bimodal proliferation of connective tissue cells via complex control of an autocrine PDGF loop. Cell 63: 515-524.

118. Sintich SM, Lamm ML, Sensibar JA, Lee C (1999) Transforming growth factor-beta1-induced proliferation of the prostate cancer cell line, TSU-Pr1: the role of platelet-derived growth factor. Endocrinology 140: 3411-3415.

119. Ehata S, Hanyu A, Hayashi M, Aburatani H, Kato Y, *et al*. (2007) Transforming growth factor-beta promotes survival of mammary carcinoma cells through induction of antiapoptotic transcription factor DEC1. Cancer Res 67: 9694-9703.

120. Sanchez-Elsner T, Botella LM, Velasco B, Corbi A, Attisano L, *et al*. (2001) Synergistic cooperation between hypoxia and transforming growth factor-beta pathways on human vascular endothelial growth factor gene expression. J Biol Chem 276: 38527-38535.

121. Iwasaki H, Suda T (2009) Cancer stem cells and their niche. Cancer Sci 100: 1166-1172.

122. Wang L, Mukhopadhyay D, Xu X (2006) C terminus of RGS-GAIP-interacting protein conveys neuropilin-1-mediated signaling during angiogenesis. FASEB J 20: 1513-1515.

123. Gao Y, Li M, Chen W, Simons M (2000) Synectin, syndecan-4 cytoplasmic domain binding PDZ protein, inhibits cell migration. J Cell Physiol 184: 373-379.

124. El Mourabit H, Poinat P, Koster J, Sondermann H, Wixler V, *et al*. (2002) The PDZ domain of TIP-2/GIPC interacts with the C-terminus of the integrin alpha5 and alpha6 subunits. Matrix Biol 21: 207-214.

125. Wang LH, Kalb RG, Strittmatter SM (1999) A PDZ protein regulates the distribution of the transmembrane semaphorin, M-SemF. J Biol Chem 274: 14137-14146.

126. Liu M, Horowitz A (2006) A PDZ-binding motif as a critical determinant of Rho guanine exchange factor function and cell phenotype. Mol Biol Cell 17: 1880-1887.

127. Pellet-Many C, Frankel P, Evans IM, Herzog B, Junemann-Ramirez M, *et al*. (2011) Neuropilin-1 mediates PDGF stimulation of vascular smooth muscle cell migration and signalling via p130Cas. Biochem J 435: 609-618.

128. Evans IM, Yamaji M, Britton G, Pellet-Many C, Lockie C, *et al*. (2011) Neuropilin-1 signaling through p130Cas tyrosine phosphorylation is essential for growth factor-dependent migration of glioma and endothelial cells. Mol Cell Biol 31: 1174-1185.

129. Cabodi S, del Pilar Camacho-Leal M, Di Stefano P, Defilippi P (2010) Integrin signalling adaptors: not only figurants in the cancer story. Nat Rev Cancer 10: 858-870.

130. Tian M, Neil JR, Schiemann WP (2011) Transforming growth factor-beta and the hallmarks of cancer. Cell Signal 23: 951-962.

131. Kim W, Seok Kang Y, Soo Kim J, Shin NY, Hanks SK, *et al*. (2008) The integrin-coupled signaling adaptor p130Cas suppresses Smad3 function in transforming growth factor-beta signaling. Mol Biol Cell 19: 2135-2146.

132. Wendt MK, Smith JA, Schiemann WP (2009) p130Cas is required for mammary tumor growth and transforming growth factor-beta-mediated metastasis through regulation of Smad2/3 activity. J Biol Chem 284: 34145-34156.

133. Di Guglielmo GM, Le Roy C, Goodfellow AF, Wrana JL (2003) Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. Nat Cell Biol 5: 410-421.

134. Le Roy C, Wrana JL (2005) Clathrin- and non-clathrin-mediated endocytic regulation of cell signalling. Nat Rev Mol Cell Biol 6: 112-126.

135. Klagsbrun M, Takashima S, Mamluk R (2002) The role of neuropilin in vascular and tumor biology. Adv Exp Med Biol 515: 33-48.

136. Ellis LM (2006) The role of neuropilins in cancer. Mol Cancer Ther 5: 1099-1107.

137. Bachelder RE, Crago A, Chung J, Wendt MA, Shaw LM, *et al*. (2001) Vascular endothelial growth factor is an autocrine survival factor for neuropilin-expressing breast carcinoma cells. Cancer Res 61: 5736-5740.

138. Chabbert-de Ponnat I, Buffard V, Leroy K, Bagot M, Bensussan A, *et al*. (2006) Antiproliferative effect of semaphorin 3F on human melanoma cell lines. J Invest Dermatol 126: 2343-2345.

139. Miao HQ, Lee P, Lin H, Soker S, Klagsbrun M (2000) Neuropilin-1 expression by tumor cells promotes tumor angiogenesis and progression. FASEB J 14: 2532-2539.

140. Kawakami T, Tokunaga T, Hatanaka H, Kijima H, Yamazaki H, *et al*. (2002) Neuropilin 1 and neuropilin 2 co-expression is significantly correlated with increased vascularity and poor prognosis in nonsmall cell lung carcinoma. Cancer 95: 2196-2201.

141. Fakhari M, Pullirsch D, Abraham D, Paya K, Hofbauer R, *et al*. (2002) Selective upregulation of vascular endothelial growth factor receptors neuropilin-1 and -2 in human neuroblastoma. Cancer 94: 258-263.

142. Cohen T, Herzog Y, Brodzky A, Greenson JK, Eldar S, *et al*. (2002) Neuropilin-2 is a novel marker expressed in pancreatic islet cells and endocrine pancreatic tumours. J Pathol 198: 77-82.

143. Handa A, Tokunaga T, Tsuchida T, Lee YH, Kijima H, *et al*. (2000) Neuropilin-2 expression affects the increased vascularization and is a prognostic factor in osteosarcoma. Int J Oncol 17: 291-295.

144. Sanchez-Carbayo M, Socci ND, Lozano JJ, Li W, Charytonowicz E, *et al*. (2003) Gene discovery in bladder cancer progression using cDNA microarrays. Am J Pathol 163: 505-516.

145. Goel HL, Chang C, Pursell B, Leav I, Lyle S, *et al*. (2012) VEGF/neuropilin-2 regulation of Bmi-1 and consequent repression of IGF-IR define a novel mechanism of aggressive prostate cancer. Cancer Discov 2: 906-921.

146. Vales A, Kondo R, Aichberger KJ, Mayerhofer M, Kainz B, *et al*. (2007) Myeloid leukemias express a broad spectrum of VEGF receptors including neuropilin-1 (NRP-1) and NRP-2. Leuk Lymphoma 48: 1997-2007.

147. Marcus K, Johnson M, Adam RM, O'Reilly MS, Donovan M, *et al*. (2005) Tumor cell-associated neuropilin-1 and vascular endothelial growth factor expression as determinants of tumor growth in neuroblastoma. Neuropathology 25: 178-187.

148. Jubb AM, Strickland LA, Liu SD, Mak J, Schmidt M, *et al*. (2012) Neuropilin-1 expression in cancer and development. J Pathol 226: 50-60.

149. Jubb AM, Sa SM, Ratti N, Strickland LA, Schmidt M, *et al*. (2012) Neuropilin-2 expression in cancer. Histopathology 61: 340-349.

150. Ghosh S, Sullivan CA, Zerkowski MP, Molinaro AM, Rimm DL, *et al*. (2008) High levels of vascular endothelial growth factor and its receptors (VEGFR-1, VEGFR-2, neuropilin-1) are associated with worse outcome in breast cancer. Hum Pathol 39: 1835-1843.

151. Sekido Y, Bader S, Latif F, Chen JY, Duh FM, *et al*. (1996) Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung cancer deletion region and demonstrate distinct expression patterns. Proc Natl Acad Sci U S A 93: 4120-4125.

152. Tomizawa Y, Sekido Y, Kondo M, Gao B, Yokota J, *et al*. (2001) Inhibition of lung cancer cell growth and induction of apoptosis after reexpression of 3p21.3 candidate tumor suppressor gene SEMA3B. Proc Natl Acad Sci U S A 98: 13954-13959.

153. Grandclement C, Pallandre JR, Valmary Degano S, Viel E, Bouard A, *et al*. (2011) Neuropilin-2 expression promotes TGF-beta1-mediated epithelial to mesenchymal transition in colorectal cancer cells. PLoS One 6: e20444.

154. Gray MJ, Van Buren G, Dallas NA, Xia L, Wang X, *et al*. (2008) Therapeutic targeting of neuropilin-2 on colorectal carcinoma cells implanted in the murine liver. J Natl Cancer Inst 100: 109-120.

155. Samuel S, Gaur P, Fan F, Xia L, Gray MJ, *et al*. (2011) Neuropilin-2 mediated beta-catenin signaling and survival in human gastro-intestinal cancer cell lines. PLoS One 6: e23208.

156. Cao Y, Szabolcs A, Dutta SK, Yaqoob U, Jagavelu K, *et al*. (2010) Neuropilin-1 mediates divergent R-Smad signaling and the myofibroblast phenotype. J Biol Chem 285: 31840-31848.

157. Glinka Y, Stoilova S, Mohammed N, Prud'homme GJ (2011) Neuropilin-1 exerts co-receptor function for TGF-beta-1 on the membrane of cancer cells and enhances responses to both latent and active TGF-beta. Carcinogenesis 32: 613-621.

158. Ball SG, Bayley C, Shuttleworth CA, Kielty CM (2010) Neuropilin-1 regulates platelet-derived growth factor receptor signalling in mesenchymal stem cells. Biochem J 427: 29-40.

159. Li X, Kumar A, Zhang F, Lee C, Li Y, *et al*. (2010) VEGF-independent angiogenic pathways induced by PDGF-C. Oncotarget 1: 309-314.

160. Devarajan E, Song YH, Krishnappa S, Alt E (2012) Epithelial-mesenchymal transition in breast cancer lines is mediated through PDGF-D released by tissue-resident stem cells. Int J Cancer 131: 1023-1031.

161. Rizzolio S, Tamagnone L (2011) Multifaceted role of neuropilins in cancer. Curr Med Chem 18: 3563-3575.

162. Jarvis A, Allerston CK, Jia H, Herzog B, Garza-Garcia A, *et al*. (2010) Small molecule inhibitors of the neuropilin-1 vascular endothelial growth factor A (VEGF-A) interaction. J Med Chem 53: 2215-2226.

163. Caunt M, Mak J, Liang WC, Stawicki S, Pan Q, *et al*. (2008) Blocking neuropilin-2 function inhibits tumor cell metastasis. Cancer Cell 13: 331-342.

164. Darbonne WC DX, Dhawan P, Hartley, Tarrant DJ,, Taylor CH, Shih LM, Brachmann RK, Phung, Q WC, LoRusso P, Patnaik A, Xiang H,, V R (2011) Mechanism for platelet reduction in antineuropilin-1 (MNRP1685A)–treated phase I patients. J Clin Oncol 29: 2011 (suppl; abstr e13598)

165. Weekes CD LP, Ramakrishnan V, Shih, LM DW, Hegde P, Xin PY, Yu R, Xiang, H BR, Patnaik A (2011) A phase Ib study for MNRP1685A (anti-NRP1) administered intravenously with bevacizumab with or without paclitaxel to patients with advanced solid tumors. J Clin Oncol 2011; 29: (suppl; abstr 3050).

166. Sugahara KN, Teesalu T, Karmali PP, Kotamraju VR, Agemy L, *et al*. (2009) Tissue-penetrating delivery of compounds and nanoparticles into tumors. Cancer Cell 16: 510-520.

167. Sugahara KN, Teesalu T, Karmali PP, Kotamraju VR, Agemy L, *et al*. (2010) Coadministration of a tumor-penetrating peptide enhances the efficacy of cancer drugs. Science 328: 1031-1035.

168. Lambert S, Bouttier M, Vassy R, Seigneuret M, Petrow-Sadowski C, *et al*. (2009) HTLV-1 uses HSPG and neuropilin-1 for entry by molecular mimicry of VEGF165. Blood 113: 5176-5185.

169. Gurrola GB, Capes EM, Zamudio FZ, Possani LD, Valdivia HH (2010) Imperatoxin A, a Cell-Penetrating Peptide from Scorpion Venom, as a Probe of Ca-Release Channels/Ryanodine Receptors. Pharmaceuticals (Basel) 3: 1093-1107.

170. Johnson G, Wu TT (2000) Kabat database and its applications: 30 years after the first variability plot. Nucleic Acids Res 28: 214-218.

171. Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94: 441-448.

172. Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet 11: 31-46.

173. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470.

174. Rung J, Brazma A (2013) Reuse of public genome-wide gene expression data. Nat Rev Genet 14: 89-99.

175. Yoo SM, Lee SY, Chang KH, Yoo SY, Yoo NC, *et al*. (2009) High-throughput identification of clinically important bacterial pathogens using DNA microarray. Mol Cell Probes 23: 171-177.

176. Yoo SM, Choi JY, Yun JK, Choi JK, Shin SY, *et al*. (2010) DNA microarray-based identification of bacterial and fungal pathogens in bloodstream infections. Mol Cell Probes 24: 44-52.

177. Jarvinen AK, Laakso S, Piiparinen P, Aittakorpi A, Lindfors M, *et al*. (2009) Rapid identification of bacterial pathogens using a PCR- and microarray-based assay. BMC Microbiol 9: 161.

178. Lee DY, Shannon K, Beaudette LA (2006) Detection of bacterial pathogens in municipal wastewater using an oligonucleotide microarray and real-time quantitative PCR. J Microbiol Methods 65: 453-467.

179. Bodrossy L, Stralis-Pavese N, Murrell JC, Radajewski S, Weilharter A, *et al*. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. Environ Microbiol 5: 566-582.

180. Suo B, He Y, Paoli G, Gehring A, Tu SI, *et al*. (2010) Development of an oligonucleotide-based microarray to detect multiple foodborne pathogens. Mol Cell Probes 24: 77-86.

181. Schrenzel J KT, Bodrossy L and Francois P (2009) Introduction to Microarray-Based Detection Methods. Detection of Highly Dangerous Pathogens: Microarray Methods for the Detection of BSL3 and BSL4 Agents.: WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

182. Kostic T, Stessl B, Wagner M, Sessitsch A, Bodrossy L (2010) Microbial diagnostic microarray for food- and water-borne pathogens. Microb Biotechnol 3: 444-454.

183. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25-29.

184. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, *et al*. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.

185. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, *et al*. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29: 365-371.

186. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207-210.

187. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, *et al*. (2003) ArrayExpress--a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 31: 68-71.

188. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, *et al*. (2013) ArrayExpress update--trends in database growth and links to data analysis tools. Nucleic Acids Res 41: D987-990.

189. Kapushesky M, Adamusiak T, Burdett T, Culhane A, Farne A, *et al*. (2012) Gene Expression Atlas update--a value-added database of microarray and sequencing-based functional genomics experiments. Nucleic Acids Res 40: D1077-1081.

190. Ioannidis JPA, Allison DB, Ball CA, Coulibaly I, Cui X, *et al*. (2009) Repeatability of published microarray gene expression analyses. Nat Genet 41: 149-155.

191. Gröger CJ, Grubinger M, Waldhör T, Vierlinger K, Mikulits W (2012) Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression. PLoS One 7: e51136.

192. Sagmeister S, Eisenbauer M, Pirker C, Mohr T, Holzmann K, *et al*. (2008) New cellular tools reveal complex epithelial-mesenchymal interactions in hepatocarcinogenesis. Br J Cancer 99: 151-159.

193. Park JG, Lee JH, Kang MS, Park KJ, Jeon YM, *et al*. (1995) Characterization of cell lines established from human hepatocellular carcinoma. Int J Cancer 62: 276-282.

194. Lee JH, Ku JL, Park YJ, Lee KU, Kim WH, *et al*. (1999) Establishment and characterization of four human hepatocellular carcinoma cell lines containing hepatitis B virus DNA. World J Gastroenterol 5: 289-295.

195. Ku JL, Park JG (2005) Biology of SNU cell lines. Cancer Res Treat 37: 1-19.

196. Caja L, Sancho P, Bertran E, Fabregat I (2011) Dissecting the effect of targeting the epidermal growth factor receptor on TGF-beta-induced-apoptosis in human hepatocellular carcinoma cells. J Hepatol 55: 351-358.

197. Coulouarn C, Factor VM, Thorgeirsson SS (2008) Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. Hepatology 47: 2059-2067.

198. Kitisin K, Ganesan N, Tang Y, Jogunoori W, Volpe EA, *et al*. (2007) Disruption of transforming growth factor-beta signaling through beta-spectrin ELF leads to hepatocellular cancer through cyclin D1 activation. Oncogene 26: 7103-7110.

199. Chun E, Lee KY (2004) Bcl-2 and Bcl-xL are important for the induction of paclitaxel resistance in human hepatocellular carcinoma cells. Biochem Biophys Res Commun 315: 771-779.

200. Mikula M, Fuchs E, Huber H, Beug H, Schulte-Hermann R, *et al*. (2004) Immortalized p19ARF null hepatocytes restore liver injury and generate hepatic progenitors after transplantation. Hepatology 39: 628-634.

201. Fischer AN, Herrera B, Mikula M, Proell V, Fuchs E, *et al*. (2005) Integration of Ras subeffector signaling in TGF-beta mediated late stage hepatocarcinogenesis. Carcinogenesis 26: 931-942.

202. van Zijl F, Mall S, Machat G, Pirker C, Zeillinger R, *et al*. (2011) A human model of epithelial to mesenchymal transition to monitor drug efficacy in hepatocellular carcinoma progression. Mol Cancer Ther 10: 850-860.

203. Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, *et al*. (2006) Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. J Clin Oncol 24: 4236-4244.

204. Raponi M, Zhang Y, Yu J, Chen G, Lee G, *et al*. (2006) Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer Res 66: 7466-7472.

205. Tang K, Rubenstein JL, Tsai SY, Tsai MJ (2012) COUP-TFII controls amygdala patterning by regulating neuropilin expression. Development 139: 1630-1639.

206. Sartor MA, Mahavisno V, Keshamouni VG, Cavalcoli J, Wright Z, *et al*. (2010) ConceptGen: a gene set enrichment and gene set relation mapping tool. Bioinformatics 26: 456-463.

207. Hall JE, Fu W, Schaller MD (2011) Focal adhesion kinase: exploring Fak structure to gain insight into function. Int Rev Cell Mol Biol 288: 185-225.

6 Abbreviations

μL	microliter	MAPK	mitogen-activated protein kinase
aa	amino acids	MET	mesenchymal to epithelial transition
AE	ArrayExpress	MGED	Microarray Gene Expression Data
AF	AlexaFluor	MIAME	Minimum Information About a Microarray Experiment
ALK	activin A receptor type II-like kinase	min	minutes
AMF	autocrine motility factor	miRNA	microRNA
Amp	ampicillin	mL	milliliter
ANK3	ankyrin-G	mm	millimeter
APS	ammonium persulfate	MMP	matrix metalloproteinase
Arg	arginine	MMP2	matrix metalloproteinase-2
ARPE19	retinal pigment epithelium cells 19	MS	Microsoft
Asp	asparagine	MTA3	metastasis-associated protein 3
BCAR1	breast cancer anti-estrogen resistance 1	MTUS1	mitochondrial tumor suppressor 1
bHLH	basic helix-loop-helix	MUC1	mucin 1
BMP	bone morphogenetic protein	MYC	myelocytomatosis
bp	base pairs	N_2	nitrogen
BSA	bovine serum albumin	Na	sodium
$CaCl_2$	calcium chloride	NaCl	sodium chloride
CDH1	E-cadherin	NaV_3O_4	sodium orthovanadate
CDH11	OB-cadherin	NCBI	National Center for Biotechnology Information
CDH2	N-cadherin	NFκB	nuclear factor κ B
CDKIs	cyclin-dependent kinase inhibitors	NGS	next generation sequencing
CendR	C-end rule	NID2	nidogen 2
cm	centimeter	NIP1	neuropilin interacting protein-1
COL1A1	collagen, type I, alpha 1	nm	nanometer
COL3A1	collagen, type 3, alpha 1	NRP1	neuropilin 1
CPPs	cell-penetrating peptides	NRP2	neuropilin 2
CSCs	cancer stem cells	OAS1	2'-5'-oligoadenylate synthetase 1
CUB	complement binding factors C1s/C1r, Uegf, BMP1	OCLN	occludin
CXADR	coxsackievirus and adenovirus receptor	OD_{600}	optical density (600 nm)
DAB	3,3'-Diaminobenzidine	P/S	penicillin/streptomycin
DAPK	death-associated protein kinase	PAA	polyacrylamide
DCs	dendritic cells	PAGE	polyacrylamide electrophoresis
DDBJ	DNA Data Bank of Japan	PAK-1	p21-activated kinase 2
ddH_2O	double distilled water or equal	PAR6	partitioning defective 6
DMEM	Dulbecco's Modified Eagle Medium	PBS	phosphate buffered saline
DMSO	dimethyl sulfoxide	PCR	polymerase chain reaction
DNA	deoxyribonucleic acid	pCR	pathological complete response
dNTPs	deoxynucleoside triphosphates	pDCs	plasmacytoid dendritic cells
DOCK1	dedicator of cytokinesis 1	PDGF	platelet-derived growth factor
E. coli	*Escherichia coli*	PDZ	post synaptic density protein 95, Drosophila disc large tumor suppressor, and zonula occludens-1 protein
EBI	European Bioinformatics Institute	PEG	polyethylene glycol
ECM	extracellular matrix	Phe	phenylalanine
EDTA	ethylenediaminetetraacetic acid	PI3K	phosphatidylinositol-3'-kinase
EGF	epidermal growth factor	PIPES	piperazine-N,N′-bis(2-ethanesulfonic acid)
ELF	ETS-related transcription factor 1	PKP2	plakophilin-2
EMT	epithelial to mesenchymal transition	PLAT	tissue plasminogen activator
EPCAM	epithelial cell adhesion molecule	PMSF	phenylmethanesulfonylfluoride
ERK	extracellular signal-regulated kinases	PPL	periplakin
E-TABM	AE series record	PRSS8	prostasin
FAK	focal adhesion kinase	PTH(rP)R	parathyroid hormone related peptide receptor
FBLN5	fibulin 5	PTX3	pentraxin 3, long
FCS	fetal calf serum	qPCR	quantitative PCR
FGF	fibroblast growth factor	RGD	Arg-Gly-Asp
FN1	fibronectin 1	RGS	regulator of G-protein signaling
FOXC2	forkhead box C2	RKFK	Arg-Lys-Phe-Lys

Abbreviations continued:

FSP-1	fibroblast specific protein-1	RNA	ribonucleic acid
FXYD3	FXYD domain containing ion transport regulator 3	rpm	revolutions per minute
GAG	glycosaminoglycan	RPMI	Roswell Park Memorial Institute medium
GAIP	Gα-interacting protein	R-SMADS	receptor Smads
GARP	glutamic acid-rich protein	RTK	receptor tyrosine kinase
GCOD	GeneChip Oncology Database	SAPK	stress-activated protein kinase
gDNA	genomic DNA	SCC	squamous cell lung carcinomas
GEF	guanine exchange factor	SCF	stem cell factor
GEO	Gene Expression Omnibus	SCID	severe combined immunodeficiency
GES	gene expression studies	SCP	small C-terminal domain phosphatase
GFP	green fluorescent protein	SD	standard deviation
GIPC	GAIP interacting protein, C terminus	SDS	sodium dodecyl sulfate
Gly	glycine	SEA	sea urchin sperm protein, enterokinase, agrin
GO	gene ontology	SEM	standard error of mean
GSE	GEO series record	Sema	semaphorin
GTP	guanosine triphosphate	SERPINB1	leukocyte elastase inhibitor
h	hours	SERPINE1	serine protease inhibitor (serpin) protein 1
H_2O_2	hydrogen peroxide	shRNA	short hairpin RNA
HAS2	hyaluronan synthase 2	SIP1/ZEB1	survival of motor neuron protein interacting protein 1 / Zinc finger E-box-binding homeobox 1
HBV	hepatitis B	siRNA	small interfering RNA
HCC	hepatocellular carcinoma	SLC7A5	solute carrier family 7 member 5
HCV	hepatitis C	SLPI	secretory leukocyte peptidase inhibitor
HGF	hepatocyte growth factor	SMURF1	Smad specific E3 ubiquitin protein ligase 1
Hh	hedgehog	SNAI1	Snail
HTLV-1	Human T-lymphotropic virus 1	SNAI2	Slug
IDs	identifiers	SNU	Seoul National University
IF	Immunofluorescence	SOC	super optimal broth with glucose
IGF	insulin-like growth factor	SPOCK1	sparc/osteonectin, cwcv and kazal-like domains proteoglycan
IH	Immunohistochemistry	Src	sarcoma
IL-6	interleukin 6	SULF1	sulfatase 1
ILEI	interleukin-related protein	SYK	spleen tyrosine kinase
ILK	integrin-linked kinase	TAE	Tris, acetic acid, EDTA
JNK	c-Jun N-terminal kinase	TAGLN	transgelin
JUP	junction plakoglobin	TAK1	TGF-β-activated kinase 1
kb	kilo base pairs	TBS	Tris buffered saline
kD	kilo Dalton	TBST	TBS containing 0.1% Tween-20
KDR	kinase insert domain receptor	TEMED	tetramethylethylenediamine
KEGG	Kyoto Encyclopedia of Genes and Genomes	TGF-β	transforming growth factor-β
KLK10	kallikrein-10	TM	transmembrane
KRT15	keratin 15	TMEM30B	transmembrane protein 30B
KRT17	keratin 17	TNF-α	tumor necrosis factor α
L	liter	TPD52L1	tumor protein D52-like 1
L1CAM	cell-adhesion molecule L1	TRAF6	TNF receptor associated factor protein 6
LAD1	leukocyte adhesion deficiency-1	TRIM33	tripartite motif containing 33
LAP	latency associated protein	TβRI	TGF-β receptor I
LB	lysogeny broth	TβRII	TGF-β receptor II
LEF-1	lymphoid enhancer-binding factor 1	TβRIII	TGF-β receptor III
LIM	Lin11, Isl-1 & Mec-3	Uegf	sea urchin fibropellin
LIMK1	LIM kinase 1	UV	ultraviolet light
LLC	large latent complex	V	volt
LMC	lateral motor column	VEGF	vascular endothelial growth factor
LOX	lysyl oxidase	VEGFR	VEGF receptor
LOXL	lysyl oxidase-like proteins	VIM	vimentin
LTBP	latent TGF-β binding protein	WB	Western blot
Lys	lysine	WNT	wingless-type
MAM	meprin/A5-protein/PTPmu	YB-1	Y-box binding protein
MAP1B	microtubule-associated protein 1B	ZO-1	zona occludens-1

7   Acknowledgements

8 Curriculum Vitae

Personal data

First name	Christian
Last name	Gröger
Birthday	05.03.1988
Email address	christian.groeger@yahoo.de

Education

2002-2007	HBLVA für chemische Industrie (Higher Technical School for Biochemistry) Rosensteingasse, Vienna, Austria
2008-2010	Bachelor of Science, *Molecular Biotechnology*, FH Campus Wien (University of Applied Sciences), Vienna, Austria
2010-2013	Master of Science, *Molecular Biology*, specialization on Molecular Medicine, University of Vienna, Austria

Work experience

2004-2007	several lab internships in analytical chemistry at the OMV and ZFT
2010	Bachelor thesis "Optimization of the AIT pathogen array for a high-throughput analysis" in the lab of Angela Sessitsch for 9 months, AIT, Seibersdorf, Austria
2011	several lab internships in the labs of Tim Skern, Christian Seiser, Wolfgang Mikulits and Thomas Rattei at the MFPL, the Medical University of Vienna and the University of Vienna
2011	lab internship at Boehringer-Ingelheim, Vienna, Austria for 3 months
2012-2013	Master thesis in the lab of Wolfgang Mikulits, Institute of Cancer Research, Vienna, Austria.

Scientific achievements

2010-2012	Lecture and poster presentations at scientific conferences of the ÖGHMP, ÖGMBT, CCC-DTSU and RESOLVE
2012	Publication: Gröger CJ, Grubinger M, Waldhör T, Vierlinger K and Mikulits W. *Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression.* PLoS One. 2012; 7(12):e51136.

Scholarship

2010	For outstanding academic achievements by the FH Campus Wien (University of Applied Sciences), Vienna, Austria

Languages

German	first language
English	C1 level, Cambridge Certificate in Advanced English