Stony Brook University

TGF-β/Smad signaling through DOCK4 facilitates lung adenocarcinoma metastasis

A Dissertation Presented

by

Jia-Ray Yu

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Genetics

Stony Brook University

May 2015

Stony Brook University

The Graduate School

Jia-Ray Yu

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Linda Van Aelst
Professor, Cold Spring Harbor Laboratory

David A. Tuveson
Professor, Cold Spring Harbor Laboraty

Gerald H. Thomsen
Professor, Department of Biochemistry and Cell Biology
Stony Brook University

Mikala Egeblad
Associate Professor, Cold Spring Harbor Laboratory

Mark R. Philips
Professor, Department of Medicine, Cell Biology, and Pharmacology
New York University School of Medicine

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

TGF-β/Smad signaling through DOCK4 facilitates lung adenocarcinoma metastasis

by

Jia-Ray Yu

Doctor of Philosophy

in

Genetics

Stony Brook University

2015

The mechanisms by which transforming growth factor β (TGF-β) promotes lung adenocarcinoma (ADC) metastasis are largely unknown. Here, we report that in lung ADC cells TGF-β potently induces expression of dedicator of cytokinesis 4 (DOCK4), but not other DOCK-family members, via the Smad pathway, and that DOCK4 induction mediates TGF-β's pro-metastatic effects by enhancing tumor cell extravasation. TGF-β-induced DOCK4 stimulates lung ADC cell protrusion, motility, and invasion, without affecting epithelial-to-mesenchymal transition. These processes, which are fundamental to tumor cell extravasation, are driven by DOCK4-mediated Rac1 activation, unveiling a novel link between TGF-β and Rac1. Thus, our findings uncover the atypical Rac1 activator DOCK4 as a key component of the TGF-β/Smad pathway that promotes lung ADC cell extravasation and metastasis.

Dedication

I would first like to thank my thesis advisor, Dr. Linda Van Aelst, for her continuous support and guidance throughout my graduate training. I certainly believe what I learned from her will become my life-long asset and be beneficial to my future career development. Also, I would like to thank my thesis committee members, Drs. David Tuveson, Gerald Thomsen, Mikala Egeblad, and Mark Philips, for their time and efforts to advise and support me with my thesis and future career plan. To the past and current members in Van Aelst lab, I appreciate your help and kindness, and I will always remember the moments we enjoyed together. I am also very grateful to my beloved family and friends. I would not have been able to make my ways through the darkest moments in my life without your company and support. Lastly, I would like to give my special thanks to Dr. Zhan Yao, for his kindness, advice, and discussion, as a friend, and also as a colleague.

Table of Contents

Chapter 3

Evidence that DOCK4 mediates TGF-β driven lung ADC metastasis

Chaper 4

Concluding remarks and future perspectives

Chapter 5

Material and methods

List of Figures/Tables/Illustrations

Figures:

Tables:

Vita, Publications and/or Fields of Study

Yu JR, Tai Y, Jin Y, Hammell MC, Wilkinson JE, Roe JS, Vakoc CR, and Van Aelst L. TGF-β/Smad signaling through DOCK4 facilitates lung adenocarcinoma metastasis. *Genes & Development*. 2015 Feb 1; 29(3):250-261.

Chapter 1

Background information

1.1 Lung Adenocarcinoma (ADC)

Lung cancer is the leading cause of cancer death and second most common cancer type in both male and female in the United States (American Cancer Society 2014). Although more than 95% of lung tumors are of epithelial origin, diverse subtypes of lung cancer have been characterized based on histopathological features, sites of occurrence, and molecular signatures. Lung cancer is roughly classified into two major subtypes, small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), accounting for about 15% and 85% of total lung cancer cases, respectively. SCLC is featured by small tumor cell size with scant cytoplasm, often found at central airways, and highly correlated with patient smoking history (Travis et al. 2013). NSCLC consists of three major subtypes, adenocarcinoma (ADC), squamous cell carcinoma (SCC), and large cell carcinoma, with the higher frequency of incidence of ADC and SCC. Lung SCC arises from squamous epithelium constituting a protective superficial layer of the bronchi. Clinically, Lung SCC is often found at large bronchi and its disease onset is associated with patient smoking history. In contrast, lung ADC is originated from the epithelial cells with secretive properties including Superfactant C (SP-C) expressing type II alveolar cells and Clara cell secretory protein (CCSP) expressing cells. Therefore, in contrast to SCC, lung ADC is often found at distal airways or lung parenchyma and in non-smokers. At the genetic and molecular levels, lung SCC and ADC also present distinct mutational landscapes and gene expression profiles. For instance, lung SCC is featured with high occurrence of activating mutations in PI3KCA and NOTCH1, and amplification of TP63 and SOX2. However, the most frequent driver oncogenes found in lung ADC are KRAS and EGFR. The tumor suppressor gene TP53 is

frequently inactivated in both lung ADC and SCC (Hoadley et al. 2014). By comprehensively analyzing the genomic DNA and RNA sequencing data of 12 tumor types, a recent study revealed that the mutational status and molecular signature of human lung SCC resembles SCCs arisen from other tissues such as head and neck squamous cell carcinoma (HNSCC), whereas lung ADC presents a unique gene expression pattern distinct from other tumor types, with a high rate of activating mutations of KRAS and EGFR (Hoadley et al. 2014). Currently, the therapeutic strategy for lung ADC is highly dependent on the EGFR mutation status, given that the ATP analogue inhibitor targeting the tyrosine kinase pocket of EGFR has been developed and approved for clinical application (Lynch et al. 2004). Although KRAS had been considered undruggable for decades, a recent study reported the preliminary development of a new allosteric inhibitor targeting $KRAS^{G12C}$ activation mutation, shedding a light for potential targeted therapy for lung ADC with mutant KRAS (Ostrem et al. 2013).

1.2 Cancer metastasis and epithelial-to-mesenchymal transition (EMT)

Metastasis is the final stage of cancer progression and is responsible for more than 90% of cancer-related death (Padua and Massague, 2009). Cancer metastasis is a multi-step process that requires coordination of several dynamic molecular and cellular features of tumor cells. During the development of solid tumor metastasis, tumor cells have to leave the primary sites, invade through nascent tissues, intravasate into blood stream, survive in the circulation, extravasate into the distant organ, and survive/colonize in the new microenvironment (Massague, 2008; Padua and Massague, 2009). At the initial step of metastasis, disseminated tumor cells must lose the cell-cell contact and adhesion from the primary tumor mass, gain the ability to degrade the extracellular matrix (ECM) and invade and move across tissue to enter circulation. These processes are often associated with at least partial or full epithelial-to-mesenchymal

transition (EMT) phenotypes (Valastyan and Weinberg, 2011). During EMT, tumor cells from epithelial origin manifest downregulation or dislocalilzation of E-cadherin, an essential component of adherens junctions that maintain epithelium integrity, and increase of vimentin and stress fibers formation to enhance actin-based cell motility. Also, tumor cells that underwent EMT also gain expression of proteases, such as collagenase MMP9, to facilitate ECM degradation. EMT is also implicated in acquisition of "cancer stemness" traits to enhance the survivability and tumor-initiating capability for distant metastasis formation (Valastyan and Weinberg, 2011). Although the association between EMT and primary tumor dissemination has been extensively studied, emerging evidence show that the gain of cellular features during EMT also facilitate the later step of cancer metastasis, such as extravasation, with tumor cells undergoing changes in shape and migratory behaviors during their intravascular transit to site of metastasis (Strell and Entschladen 2008; Stoletov et al. 2010; Reymond et al. 2013). Although EMT can be driven by both intrinsic and extrinsic factors, highlighted in tumors of epithelial origin, the paracrine cytokine signaling such as TNF-α or TGF-β derived from tumor stroma plays critical roles to promote tumor cell EMT (Valastyan and Weinberg, 2011). The tumor stroma consists of several cell types that produce high dose of these cytokines, including cancer-associated fibroblasts (CAFs), mesenchymal stem cells (MSCs), and tumor-associated macrophages. During tumor progression, the tumor stromal cells of hematopoietic origin are recruited to primary tumor sites and become important sources of these cytokines, as a natural response to intratumoral hypoxic or inflammatory conditions (Padua et al, 2008).

1.3 The transforming growth factor β (TGF-β) pathway in cancer

The TGF-β signaling pathway is conserved in metazoans, from worms and flies to mice and human. In mammals, TGF-β regulates multiple cellular responses including apoptosis, cell

cycle progression, differentiation, and EMT during development, tissue homeostasis, and cancer progression. The signaling basis of the TGF-β pathway consists of the canonical signal transducer Mothers against decapentaplegic homolog (SMAD) and non-canonical, non-SMAD pathways. TGF-β is produced and secreted as a latent form in complex with carrier proteins termed latent TGF-β binding proteins (LTBPs), requiring further proteolysis for its activation. Once activated, TGF-β binds to TGF-β receptor II (TGFBR2) and then recruits receptor I (TGFBR1) to form a heterdimeric complex. TGFBR2 then activates TGFBR1 by phosphorylation of its GS domain. In the canonical TGF-β signaling pathway, activated TGFBR1 phosphorylates the receptor-coupled SMAD proteins (R-SMADs), SMAD2 and SMAD3. Phosphorylated and activated SMAD2/3 then form a heterotrimetic complex with the co-SMAD, SMAD4, shuttle to the nucleus, and bind to a minimal four nucleotide AGAC termed SMAD binding element (SBE) to regulate gene expression (Massague et al. 2005). In human cancers, TGF-β is a well-known pleiotropic cytokine that plays a highly context-dependent role at different tumor stages. In normal epithelium or during early oncogenesis, activation of the TGF-β pathway elicits a cell-autonomous, anti-proliferative response through direct induction of several cell cycle inhibitors including $p15^{INK4B}$, $p19^{ARF}$, and $p21^{WAF1/CIP1}$ (Massague 2008). Thus, during tumor progression and evolution, advanced tumor cells often lose the tumor suppressive arm of TGF-β by shaping the TGF-β dependent response for gene expression. In some cases, advanced tumor cells become able to alleviate the TGF-β signaling by down-regulating TGFBR2 expression (e.g. lung and breast cancers) or even abrogate it by losing the expression of functional TGFBR2 or SMAD4 (e.g. colon and pancreatic cancers) (Elliot et al, 2005). Nonetheless, in later stages of tumor progression, advanced tumors often recruit abundant immune cells and accumulate higher dose of TGF-β as a natural response to intratumoral

hypoxic and inflammatory conditions. Tumor cells retaining the core TGF-β signaling machinery can thus take advantage of the paracrine TGF-β signaling to gain their invasiveness to disseminate from the primary sites through the induction of EMT and other prometastatic genes, including MMP9 and Angiopoietin-like 4 (ANGPTL4). While disseminated tumor cells enter the blood stream, the myeloid cells in circulation become a bioavailable source of TGF-β to maintain their metastatic phenotype. At least, platelets, highlighted by recent reports, can activate the TGF-β signaling in circulating tumor cells, through direct contact and attachment, in the blood stream (Labelle et al, 2011; Yu et al, 2013). As metastatic tumor cells exit from the blood stream and enter a distant organ, they no longer maintain high TGF-β signaling activity, due to lack of the presence of paracrine TGF-β, to re-epithelialize and restore their proliferative activities, facilitating metastatic colonization and growth.

1.4 The Rho family GTPases in canceer

The Rho family GTPases is a family of small signaling G-proteins with molecular weights of ~21kDa and is belong to the Ras superfamily. The Rho family GTPases has 22 mammalian members, with Rac1, Cdc42, and RhoA being the most extensively studied members. The Rho proteins are renowned for their roles in controlling actin and cytoskeletal dynamics, gene expression, cell proliferation, and signal transduction downstream of diverse sources such as G-protein coupled receptors (GPCRs), receptor tyrosine kinases (RTKs), and integrin signaling pathways. Numerous downstream effectors of the active Rho proteins have been identified to date, including p21-activated kinase (PAK) for Rac1/Cdc42 and Rho-associated protein kinase (ROCK) for RhoA (Sahai and Marshall 2002). The activation and inactivation of Rho proteins is controlled by two classes of proteins termed guanine nucleotide exchange factors (GEFs) and guanosine triphosphatase activating proteins (GAPs). While in the

inactive state, the Rho proteins are GDP-bound; during their activation, GEFs catalyze the GDP to GTP switch, whereas GAPs promote GTP hydrolysis and inactivate the Rho proteins. In human cancers, the roles of Rho proteins have been demonstrated involving in different steps of tumor progression. Early studies have shown that the activity of RhoA, Rac1, or Cdc42 is required for Ras mediated cellular transformation in fibroblasts, and later studies also confirmed that as least Rac1 is essential for Ras driven tumorigenesis by regulating the proliferation and survival of tumor cells in mice (Sahai and Marshall, 2002; Kissil et al. 2007). During tumor cell dissemination and metastasis, the activities of Rho proteins are coordinated to disrupt adherens junctions and facilitate cell movement by regulating actin-based cytoskeleton in different manners. For example, during tumor cell EMT, RhoA signaling through ROCK and MLC stimulates actin stress fiber formation, Rac/WAVE promotes lamellipodia formation, and Cdc42/WASP facilitates filopodia formation (Zhang 2009). However, in response to differential microenvironments, tumor cells can display different modes of movement upon activation of the Rho proteins. Recent studies have shown that melanoma cells manifest mesenchymal-like movement driven by Rac1 and amoeboid-like movement by Cdc42 while they are plated on thick collagen covered by matrigel, named as 2.5D cell culture. (Sanz-Moreno et al. 2008; Gadea et al. 2008). Besides, in 3D matrigel culture, tumor cells can form invadopodia dependent on the activity of Cdc42 to facilitate ECM degradation (Yamaguchi et al. 2005). Unlike the early discoveries of oncogenic mutations in Ras proteins, the mutations in Rho proteins in human cancers had not been found until recently. Through genomic sequencing of human tumor samples, a few groups have identified the mutations in RhoA and Rac1 in several types of human cancers, including $Rac1^{P29S}$ in melanoma and $RhoA^{G17V}$ in diffuse-type gastric carcinoma and T cell lymphoma (Hodis et al. 2012; Kakiuchi et al. 2014; Sakata-Yanagimoto et al. 2014).

Importantly, biochemical evidence showed that these mutations appear to be gain-of-function, implicating the tumor promoting roles of RhoA and Rac1 in these cancer types.

1.5 The dedicators of cytokinesis (DOCK) family in cancer

The DOCK family is conserved in metazoans, from worms and flies to mice and human, with one ortholog in *Caenorhabditis elegans* (CED-5), one in *Drosophila melanogaster* (Myoblast city, Mbc), and eleven in *Mus musculus* and *Homo sapiens* (DOCK1 to DOCK11). In mammals, the DOCK proteins are GEFs for Rho family GTPases, Rac and Cdc42, with DOCK1 to DOCK5 specific for Rac, DOCK9 to DOCK11 for Cdc42, and DOCK6 to DOCK8 for Rac and potentially Cdc42 (Cote and Vouri 2002; Laurin and Cote 2014). Distinct from Dbl homology (DH) domain Rho-GEFs, the DOCK proteins utilize the DOCK homology region 2 (DHR2) domain to activate Rac or Cdc42 and thus are considered atypical GEFs. In addition, conserved in all 11 DOCK proteins, the DOCK homology region 1 (DHR1) domain has been shown to interact with phosphatidylinositol 3,4,5 trisphosphate (PIP3) or phosphatidic acid (PA) for their membrane recruitment (Laurin and Cote 2014). Despite that all 11 DOCK proteins harbor the DHR1 and DHR2 domains, the SH3 domain and proline-rich (PxxP) motif are only found in DOCK1 to DOCK5 (Illustration 1). In these DOCK family members, the N-terminal SH3 domain binds to the C-terminal PxxP motif to form an autoinhibitory loop and restrict their GEF activities. The full activation of these DOCK members requires the protein-protein interaction between their SH3 domain and the PxxP motif of the Engulfment and Cell Motility (ELMO) family proteins, to alleviate the autoinhibition. The temporal and spatial activation of Rac/Cdc42 by DOCK proteins has been reported to regulate cell migration and survival in different types of cancer cells. For instance, DOCK1/Rac signaling is required for PDGF dependent cell survival in glioma cells and HER2 dependent proliferation/tumorigenesis in

breast cancer mouse model (Laurin and Cote 2014). In melanoma cells, DOCK3 controls Rac-dependent elongated cell shape and mesenchymal-like movement, whereas DOCK10-driven Cdc42 activation promotes rounded cell shape and amoeboid-like movement in 2.5D culture conditions, with cells plated on a thick layer of collagen, covered with low concentration of matrigel. Furthermore, DOCK1 and DOCK3 have been shown to form complex with p130Cas family proteins p130Cas and NEDD9, respectively, and mediate integrin-dependent Rac activation in breast cancer and melanoma cells (Laurin and Cote 2014). Lastly, DOCK4 has been suggested to function as a tumor suppressor whose loss facilitates spontaneous sarcoma formation in the TP53 and NF2 depleted mouse model (Yajnik et al. 2004).

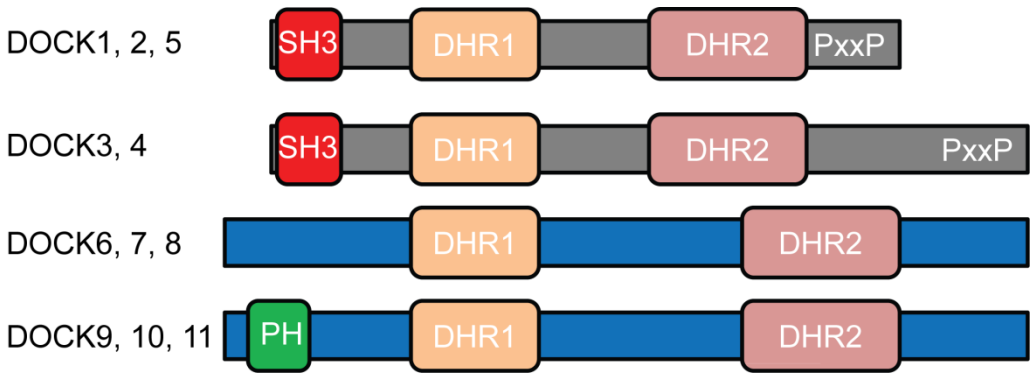

Illustration 1: DOCK family proteins

Chapter 2

Discovery of DOCK4 as a TGF-β/Smad target gene in human lung ADC

2.1 Introduction

The cytokine TGF-β plays an important, albeit complex, role in epithelial tumorigenesis (Derynck et al. 2001; Elliott and Blobe 2005; Padua and Massague 2009; Jakowlew 2010). During early stages of tumorigenesis, TGF-β typically functions as a tumor suppressor. At later stages, however, as tumors grow and progress, TGF-β produced by both tumor and stromal cells within the tumor microenvironment, as a natural response to hypoxic and inflammatory conditions, can act as a potent promoter of multiple steps of the metastatic process. These include not only local motility/invasion and entry of cancer cells into the blood stream (intravasation), but also their exit from the blood vessels (extravasation) and survival at the distant organ sites (Ma et al. 2008; Massague 2008; Padua et al. 2008; Giampieri et al. 2009; Padua and Massague 2009; Labelle et al. 2011; Valastyan and Weinberg 2011; Calon et al. 2012; Yuan et al. 2014). The relevance of TGF-β signaling for disease progression has been particularly recognized in tumors where cancer cells retain the core TGF-β signaling components, as is often the case in breast and lung cancers (Kang et al. 2003; Elliott and Blobe 2005; Massague 2008; Padua and Massague 2009). Indeed, in lung ADC, the most common subtype of lung cancer with a high mortality rate, increased TGF-β1 expression correlates with tumor progression and poor patient survival, and various experimental model systems support a pro-metastatic role for TGF-β in these tumors (Lund et al. 1991; Hoffman et al. 2000; Hasegawa et al. 2001; Gibbons et al. 2009; Nemunaitis et al. 2009; Toonkel et al. 2010; Provencio et al. 2011; Vazquez et al. 2013). However, a major remaining challenge is the identification of TGF-β target

genes that drive the different steps of metastasis, especially since TGF-β modulates gene expression in a highly cell- and context-specific manner (Padua and Massague 2009; Mullen et al. 2011; Massague 2012). While some progress has been made in the context of breast cancer metastasis (Michl et al. 2005; Padua et al. 2008; Gregory et al. 2011; Sethi et al. 2011; Shibue et al. 2013), the genes and mechanisms that mediate the pro-metastatic effects of TGF-β in lung ADC, particularly those mediating the extravasation step, remain largely unknown.

To explore molecular mechanisms that could mediate the pro-metastatic effects of TGF-β in lung ADC, we took a candidate gene approach and started off by scrutinizing members of the DOCK180-related protein superfamily. The DOCK180 family, of which in total 11 mammalian members have been identified (termed DOCK1/DOCK180 to DOCK11) emerged as a novel class of Rac/Cdc42 GTPase guanine nucleotide exchange factors (GEFs) (Cote and Vuori 2002; Meller et al. 2005). This class of proteins has been implicated in diverse cell-type specific processes (Laurin and Cote 2014), with some of its members (*i.e.* DOCK1, DOCK3 and DOCK10) playing distinct roles in the progression and/or metastasis of diverse tumor types, including melanoma, breast cancer, and glioblastoma, by engaging in different protein-protein interactions (Gadea et al. 2008; Sanz-Moreno et al. 2008; Feng et al. 2012; Laurin et al. 2013). As to whether any of the DOCK proteins play(s) a role in the progression and/or metastasis of lung ADC has hitherto not been investigated. Our findings presented here demonstrate that the DOCK4 family member plays a critical role in mediating TGF-β's pro-metastatic effects in lung ADC. We find that in lung ADC cells expression of DOCK4, but not other DOCK180-family members, is rapidly and robustly induced by TGF-β in a Smad dependent manner, and that DOCK4 induction is essential for TGF-β-driven lung ADC metastasis. Blockade of TGF-β-induced DOCK4 attenuates the ability of lung ADC cells to extravasate into distant organ sites,

resulting in a marked reduction in metastatic burden in mice. At the cellular level, TGF-β-induced DOCK4 elicits lung ADC cell protrusive activity, motility and invasion, and intriguingly does so via Rac1 activation. So far, Rac1 has been linked to TGF-β via non-canonical, non-Smad pathway(s) (Zhang 2009). Thus, our findings identify the atypical Rac1 activator DOCK4 as a novel, key component of the TGF-β/Smad pathway that promotes lung ADC cell extravasation and metastasis.

2.2 Specific research goals

Given that DOCK4 appears to be an interesting candidate gene that is potentially involved in the biology of metastasis and the mechanistic link between TGF-β and Rac pathways have been long missing, the specific aims of this study are set to elucidate the molecular and phenotypic features of DOCK4 in TGF-β driven lung ADC metastasis, including 1) understanding the molecular mechanism as how DOCK4 is regulated by TGF-β signaling. 2) modeling TGF-β driven metastasis for lung ADC in mice and assess the function of DOCK4 in this context. 3) stepwise dissecting the role of DOCK4 involved in the metastatic cascade. 4) investigating the function of DOCK4 at cellular level and how it possibly regulates cellular behaviors. 5) exploring the potential clinical correlation of DOCK4 in human lung ADC.

2.3 TGF-β induces DOCK4 expression in lung ADC cells

Upon examining the expression profiles of all 11 DOCK180-family members in the human lung ADC cell line A549, subjected or not subjected to TGF-β treatment, by real-time quantitative PCR, we found that all members, with the exception of DOCK2, were expressed in A549 cells. Strikingly, however, only DOCK4 mRNA levels, but not those of any other DOCK180-family member, were robustly upregulated by TGF-β treatment (Fig. 1A). Besides,

we noted DOCK4 to be the most strongly upregulated gene, among all Rho-family GEFs, when analyzing a publically available dataset comprising gene expression profiles of TGF-β treated A549 cells (NCBI, Gene Expression Omnibus GSE17708) (Fig. 2A). Although in this dataset DOCK2 and DOCK9 levels appeared to mildly increase, we did not find the similar fashion in the validation using qPCR, indicating these trends might be due to the noise of the microarray experiments. Importantly, TGF-β-induced upregulation of DOCK4 was also seen at the protein level (Fig. 1B). The increase in DOCK4 protein levels was not only observed in the KRAS-mutant A549 cell line, but also in several other lung ADC cell lines carrying either a KRAS mutation (H441) or EGFR mutations (HCC4006, H1975, and PC9), as well as in a KRAS and EGFR wild-type lung ADC cell line (H1793) (Fig. 1B). All above cell lines displayed increased Smad3 phosphorylation levels in response to TGF-β (Fig. 1B). Interestingly, no increase in DOCK4 expression by TGF-β was observed in any of the TGF-β-responsive breast cancer and melanoma cell lines we examined (Fig. 2B, and data not shown), suggesting that the effect of TGF-β on DOCK4 expression is tumor-type dependent. Noteworthy, a recent study also implicated the WNT/TCF pathway in lung ADC metastasis (Nguyen et al. 2009). However, we did not detect any change in DOCK4 protein levels upon treatment of A549 cells with WNT3A (Fig. 2C), implying that DOCK4 is not likely a target gene of the WNT/TCF pathway in lung ADC metastasis.

A key pathway in the regulation of TGF-β-induced gene expression is the canonical Smad pathway, albeit non-canonical non-Smad pathways have been implicated as well (Padua and Massague 2009; Zhang 2009). To explore whether the Smad pathway is responsible for TGF-β-induced DOCK4 upregulation, we used two previously described Smad4 shRNAs (shSmad4#1 and shSmad4#2), which we confirmed to be effective in reducing Smad4 protein

levels (Fig. 1C and Fig. 2D). Smad4 is an essential component of the Smad pathway; on TGF-β stimulation it forms complexes with receptor-phosphorylated Smad2/Smad3 proteins, which translocate into the nucleus and regulate gene transcription (Padua and Massague 2009). Stable A549 and HCC4006 cell lines expressing shSmad4#1, shSmad4#2, or control shRNA (shCtrl) were generated, and subjected to TGF-β treatment. As expected, in shCtrl-expressing cells, TGF-β triggered a robust increase in DOCK4 expression. This increase, however, was largely prevented in shSmad4#1- and shSmad4#2-expressing cells (Fig. 1D,E and Fig. 2D,E). Moreover, when we used the Smad3 inhibitor SIS3, or the TGF-β type I receptor inhibitor SB432542, to interfere with the TGF-β/Smad pathway, we similarly found that SIS3 and SB432542 blocked the TGF-β-induced increase in DOCK4 expression (Fig. 1F-H and Fig. 2F). Noteworthy, DOCK4 mRNA upregulation was also seen upon acute TGF-β treatment (3 and 7 h), and was entirely abrogated when cells were treated with the transcription inhibitor, Actinomycin D (Fig. 2G). These data indicate that TGF-β transcriptionally upregulates DOCK4 in lung ADC cells via the canonical Smad pathway.

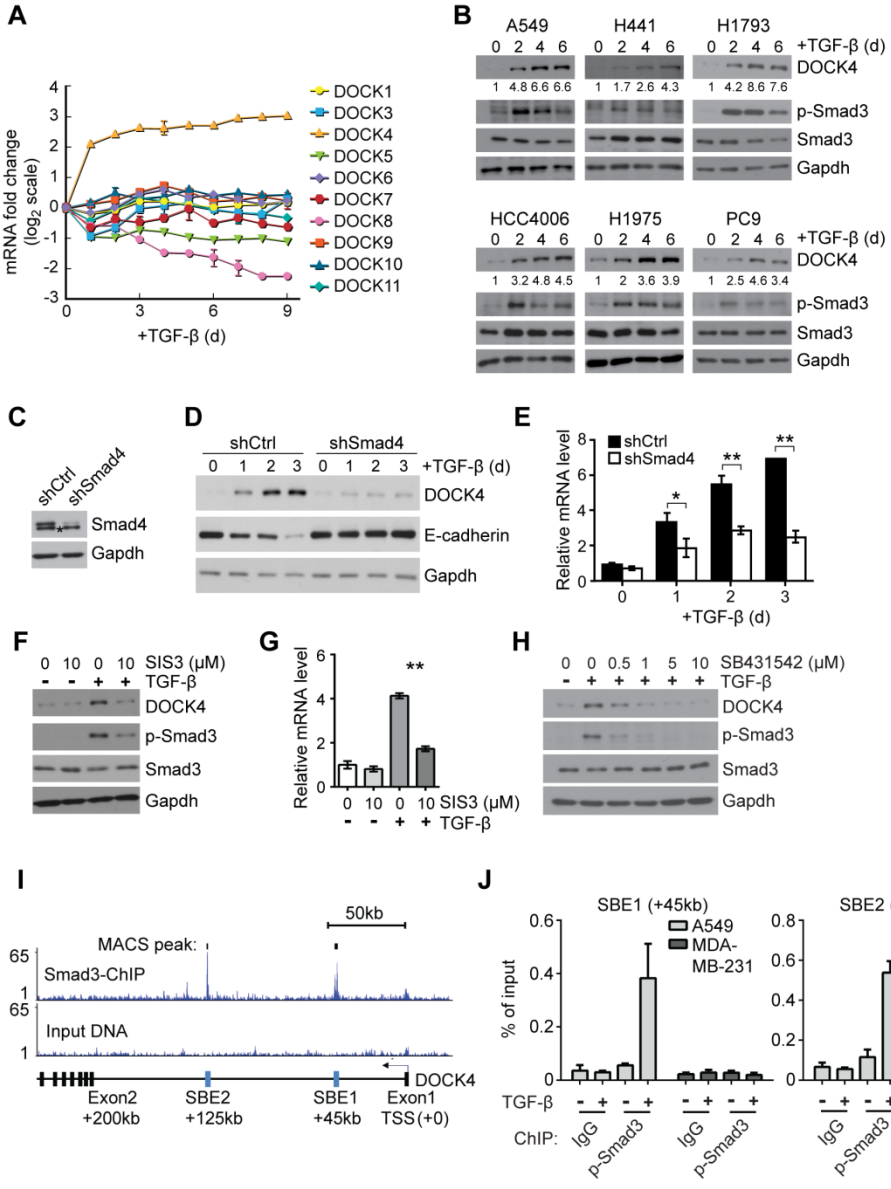

Figure 1. TGF-β induces DOCK4 expression in human lung ADC cells via the Smad pathway. (*A*) qPCR analysis of DOCK180-family members mRNAs in A549 cells treated with 2 ng/ml TGF-β. (*B*) Western blot analysis of DOCK4, p-Smad3, and Smad3 in human lung ADC cell lines treated with TGF-β. DOCK4 levels were normalized to Gapdh and then to a value of 1.0 for day 0. (*C*) Western blot analysis of Smad4 in A549 cells stably expressing shCtrl or shSmad4#1. Asterisk depicts a non-specific band. (*D*) Western blot analysis of DOCK4 and E-cadherin and (*E*) qPCR analysis of DOCK4 mRNA in shCtrl- and shSmad4#1-expressing A549 cells treated with TGF-β. (*F*) Western blot analysis of DOCK4, p-Smad3, and Smad3 and (*G*) qPCR analysis of DOCK4 mRNA in A549 cells treated with 10 μM p-Smad3 inhibitor SIS3 and/or 2 ng/ml TGF-β for 24 h. (*H*) Western blot analysis of DOCK4, p-Smad3, and Smad3 in A549 cells treated with 0-10 μM TGF-β type I receptor inhibitor SB431542 and 2 ng/ml TGF-β for 24 h. qPCR data in *A*, *E*, and *G* were normalized to Gapdh, and presented as mean ± SD (n = 3). *P <

0.05, **$P < 0.01$, by an unpaired two-tailed Student's *t*-test. (*I*) Top, ChIP-seq occupancy profile of Smad3 along with input DNA on human DOCK4 locus obtained in A549 cells shown in reads per half a million. MACS peaks depict two validated peaks with a false discovery rate (FDR) less than 5%. Data were obtained from GEO database (GSE51509). Bottom, validated transcript models for DOCK4 from the hg18 genome assembly. Black bars, exons. Blue bars, Smad-binding elements (SBEs) within the corresponding MACS peaks. TSS, transcriptional start site. (*J*) ChIP assay for p-Smad3 binding to two SBEs in the first intron of DOCK4. A549 and MDA-MB-231 cells left untreated or treated with 2 ng/ml TGF-β for 5 h were harvested and processed for ChIP with isogenic IgG or anti-p-Smad3 antibody. The enrichment of the precipitated DNA by p-Smad3 antibody versus the IgG was analyzed by qPCR using primers flanking SBE1 and SBE2. Data are shown as fold of DNA enrichment and presented as mean ± SD (n =3).

2.4 DOCK4 is a direct target gene of TGF-β/Smad pathway

To evaluate whether DOCK4 is a direct target of the TGF-β/Smad pathway, we first tested whether new protein synthesis is required for TGF-β-induced transcriptional activation of DOCK4. We found that treatment of A549 cells with the protein synthesis inhibitor cycloheximide (CHX) did not prevent TGF-β-induced upregulation of DOCK4 mRNA (Fig. 2H), indicating that increased DOCK4 expression is not a secondary effect of TGF-β/Smad signaling activation. We next searched for potential Smad-binding elements (SBEs) in the DOCK4 promoter region using TRANSFAC and FIMO from the MEME suite (Grant et al. 2011). However, we did not detect any canonical SBE within 20 kb of the DOCK4 transcriptional start site (data not shown). Since prior studies had shown that a large proportion of Smad binding sites are found outside of promoter proximal regions at putative enhancer elements (Kennedy et al. 2011; Morikawa et al. 2011; Schlenner et al. 2012; Gaunt et al. 2013), we considered whether Smad proteins occupy distal SBEs at the DOCK4 locus. To this end, we analyzed available Smad3 ChIP-seq data (GSE51509) obtained from TGF-β-stimulated A549 cells (Isogaya et al. 2014) and found two significant Smad3 peaks in the first intron of DOCK4 at +45 kb and +125 kb (Fig. 1I). Notably, each of these two sites contains a single SBE. To test whether p-Smad3 directly binds to the two putative SBEs, we designed primers flanking the two putative SBEs and performed anti-p-Smad3 chromatin immunoprecipitation followed by qRT-PCR (ChIP-qPCR). We found that p-Smad3 binds to both SBEs in a TGF-β-dependent manner (Fig. 1J). Importantly, binding of p-Smad3 to the two SBEs was detected in lung ADC A549 cells, but not in breast cancer MDA-MD-231 cells (Fig. 1J), in which no TGF-β-induced upregulation of DOCK4 was observed (Fig. 2B). Taken together, these data strongly suggest that DOCK4 is a direct TGF-β/Smad target gene in lung ADC cells.

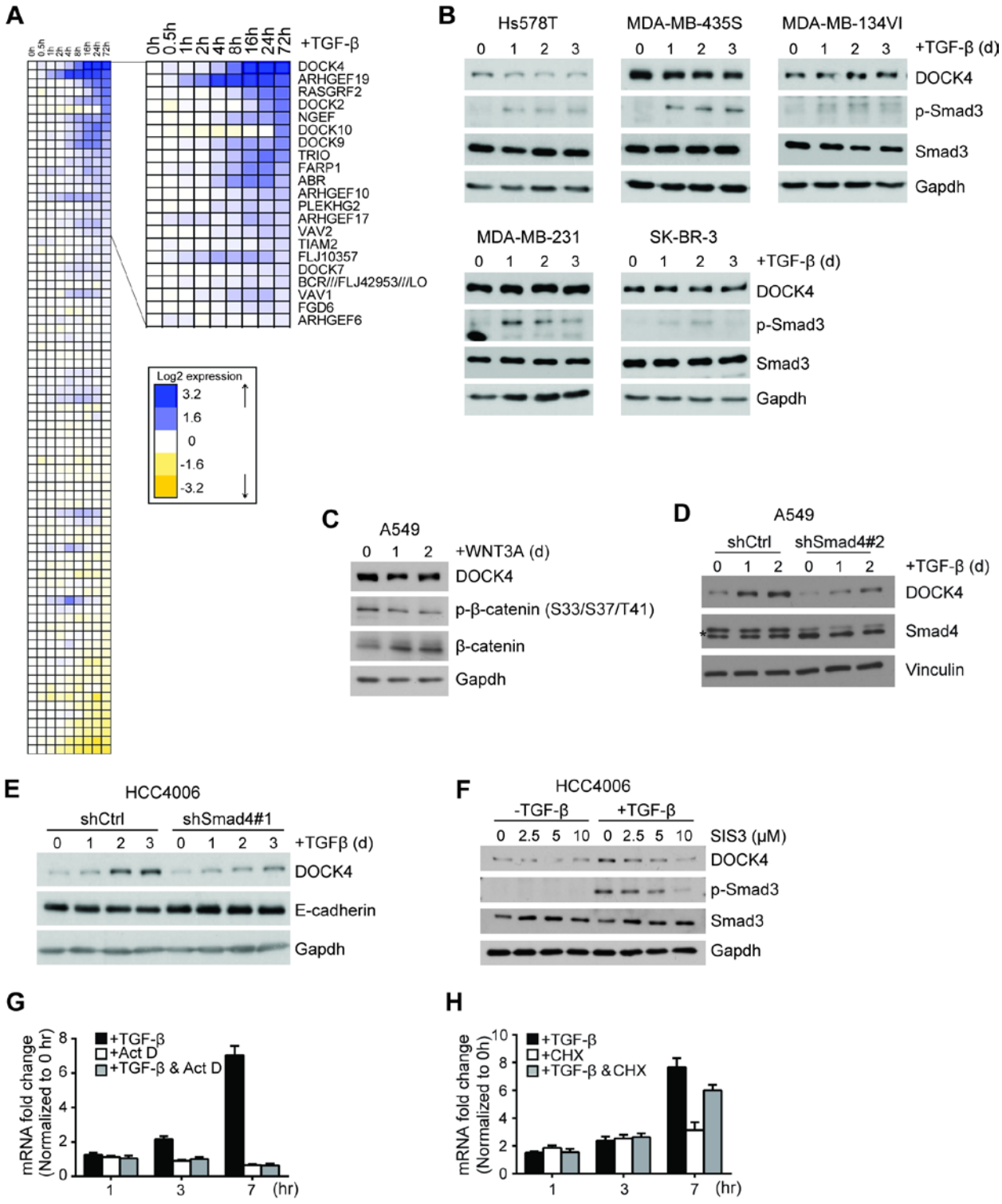

Figure 2. TGF-β/Smad signaling induces DOCK4 expression in lung ADC cells, but not breast cancer cells. *(A)* Heat map showing differential expression of all 83 Rho family GEFs in TGF-β treated A549 cells over a 72-h time window. Gene expression data are presented on a log2 scale. Blue and yellow indicate upregulation and downregulation, respectively, by TGF-β. The top 21

TGF-β upregulated Rho family-GEFs are listed on the right. Original data were retrieved from Gene Expression Omnibus (GEO) with accession number GSE17708. (*B*) Western blot analysis of DOCK4, p-Smad3, and Smad3 in a panel of breast cancer cell lines treated with 2 ng/ml TGF-β over a 3-d time window. Gapdh was used as a loading control. (*C*) Western blot analysis of DOCK4, p-β-catenin, and β-catenin in A549 cells treated with 100 ng/ml WNT3A over a 2-d time window. (*D*) Western blot analysis of DOCK4 and Smad4 in shCtrl- and shSmad4#2-expressing A549 cells treated with 2 ng/ml TGF-β over a 2-d time window. Asterisk depicts a non-specific band. (*E*) Western blot analysis of DOCK4 and E-cadherin in shCtrl- and shSmad4#1-expressing HCC4006 cells treated with 2 ng/ml TGF-β over a 3-d time window. (*F*) Western blot analysis of DOCK4, p-Smad3, and Smad3 in HCC4006 cells treated with 0 to 10 μM SIS3 and/or 2 ng/ml TGF-β for 24 h. (*G*) qPCR analysis of DOCK4 mRNA in A549 cells treated with 2 ng/ml TGF-β, 2 μg/ml Actinomycin D (Act D), or both. (*H*) qPCR analysis of DOCK4 mRNA in A549 cells treated with 2 ng/ml TGF-β, 50 μg/ml Cycloheximide (CHX), or both. qPCR data in G and H were normalized to Gapdh, and presented as mean ± SD (n = 3).

2.5 High DOCK4 expression is correlated with activation of TGF-β signaling and poor recurrence-free survival in human lung ADC patients

To extend our findings beyond cells in culture and determine a possible relevance to human lung ADC disease, we examined whether the levels of DOCK4 and p-Smad3 (used as readout for activity of TGF-β signaling) were correlated in human lung ADC. To this end, we performed immunohistochemistry (IHC) on human lung ADC tissue microarrays (TMAs), using anti-DOCK4 and anti-p-Smad3 specific antibodies (Fig. 4A) (Siebert et al. 2011). We observed that DOCK4 expression was significantly higher in tumor tissues compared to adjacent normal tissues (Fig. 4B,C). Moreover, and, importantly, we observed a strong and significant positive correlation between DOCK4 and p-Smad3 levels in the tumor tissues (Fig. 3A,B), indicating that DOCK4 levels positively correlate with activated TGF-β signaling in human lung ADC.

We further investigated whether DOCK4 expression was correlated with the clinical outcome of human lung ADC. To this end, we analyzed a publicly available microarray dataset containing gene expression profiles of 182 human lung adenocarcinomas and clinical follow-up information. Both Cox proportional univariate and multivariate analyses revealed an inverse correlation between DOCK4 expression and patient recurrence-free survival (Fig. 3C,D, and Table 1). Moreover, contingency analyses revealed that high DOCK4 expression is strongly correlated with high frequency of recurrence events, and weakly but significantly with advanced tumor stage (≥ stage III) (Table 2). Notably, no such correlations were observed for DOCK3 expression; DOCK3 is the most closely related to DOCK4 family member (Fig. 5A,B and Table 1). Also, upon analyzing publicly available breast cancer datasets, no inverse correlation between DOCK4 expression and patient recurrence free-survival was observed (Fig. 5C,D), consistent with our findings that TGF-β does not induce DOCK4 expression in breast cancer cell lines.

Together, these results suggest that DOCK4 is a potential prognostic factor that predicts disease relapse in human lung ADC patients.

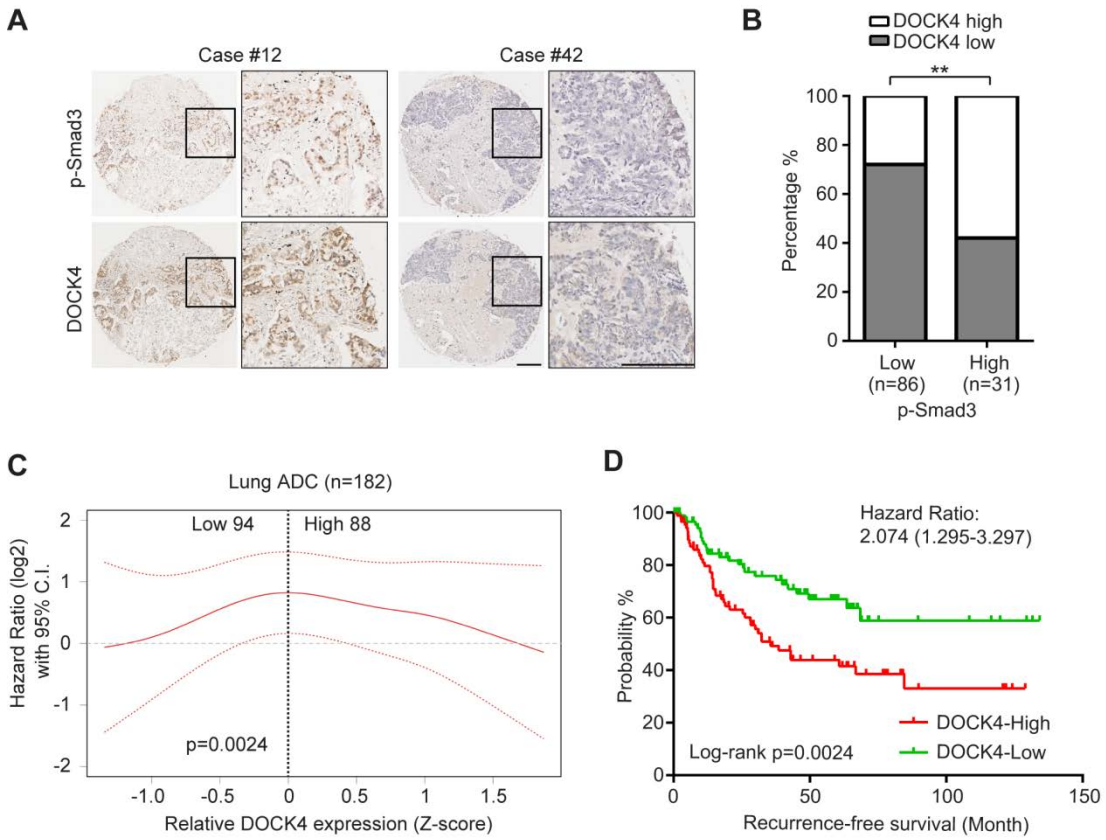

Figure 3. DOCK4 expression is correlated with activity of TGF-β signaling and recurrence-free survival in lung ADC. (*A*) Representative images of immunohistochemical (IHC) stainings for p-Smad3 and DOCK4 in human lung ADC tissue microarrays. Nuclei were counterstained with hematoxylin. Boxed regions are enlarged and shown on the right. Scale bars, 250 μm. (*B*) Percentage of human lung ADC samples displaying low or high DOCK4 expression in low or high p-Smad3 expression group. $^{**}P < 0.01$ by Fisher's exact test. (*C*) Hazard ratio plot in function of DOCK4 expression based on gene expression and recurrence-free survival data for a cohort of 182 lung ADC patients. The dotted line indicates the cut-off that yields the highest hazard ratio with confidence interval 95% to define low/high DOCK4 expression groups. (*D*) Kaplan-Meier survival curve for low/high DOCK4 expression groups indicated in (*C*). The *P* value was calculated by log-rank test.

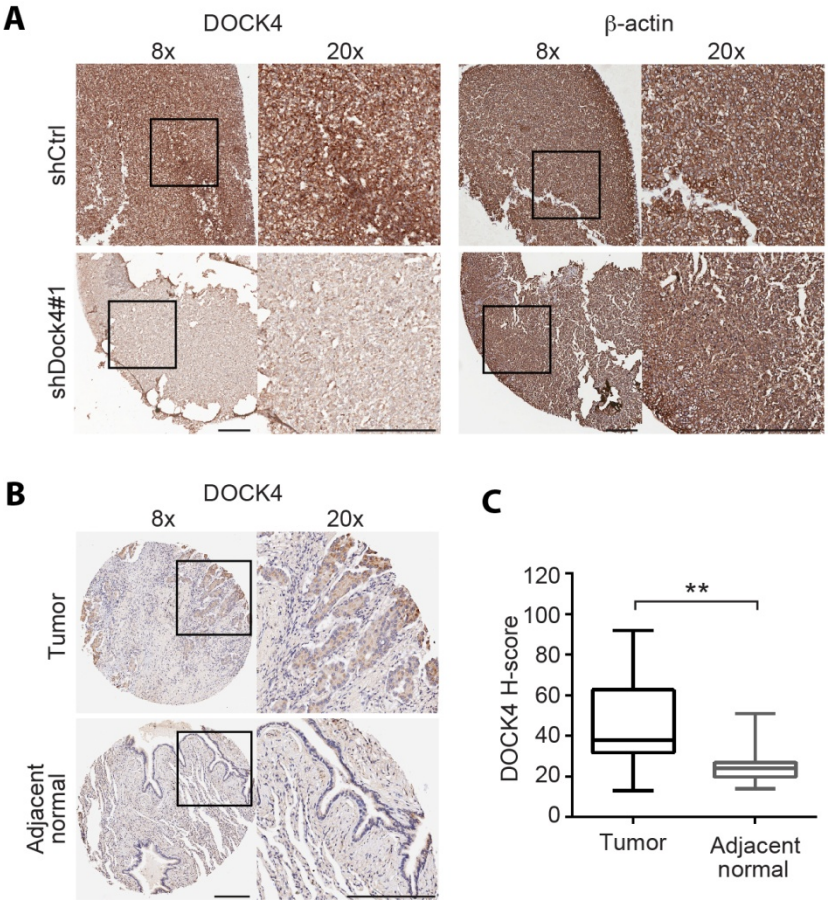

Figure 4. Validation of anti-DOCK4 antibody for immunohistochemical (IHC) staining, and DOCK4 expression in primary lung ADC and tumor adjacent normal tissues. *(A)* shCtrl- and shDock4#1-expressing A549 cell pellets processed with acidic formalin fixation and paraffin embedding, and stained with antibodies against DOCK4 (left) and β-actin (right). Boxed regions are enlarged and shown on the right. Scale bars, 250 μm. *(B)* Representative images of DOCK4 IHC staining of a tissue microarray containing human lung ADC tissues and tumor adjacent normal tissues. Nuclei were counterstained with hematoxylin. Boxed regions are enlarged and shown on the right. Scale bars, 250 μm. *(C)* H-scores of DOCK4 expression in 16 pairs of matched human lung ADC tissues and their adjacent normal tissues that contain sufficient epithelium area for IHC quantification. $^{**}P < 0.01$ by a paired Student's *t*-test.

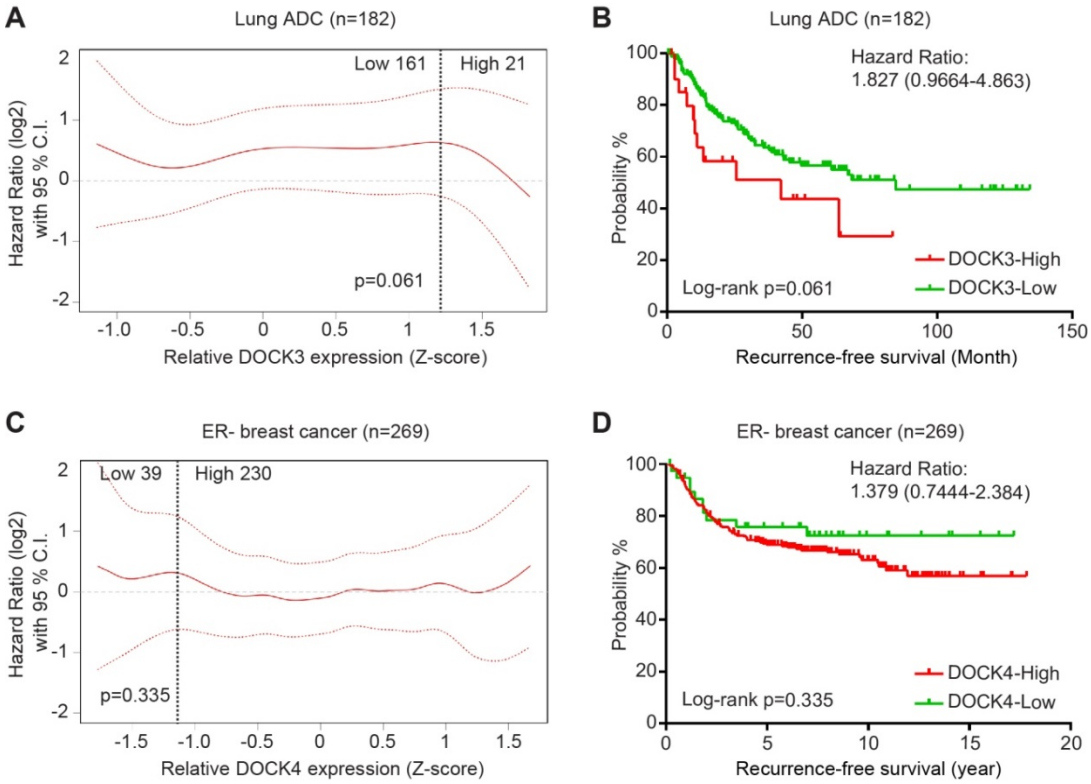

Figure 5. DOCK3 and DOCK4 are not significantly associated with recurrence-free survival in lung ADC and in estrogen-receptor (ER)-negative breast cancer patients, respectively. *(A)* Hazard ratio plot in function of DOCK3 expression based on gene expression and recurrence-free survival data for a cohort of 182 lung ADC patients. *(B)* Kaplan-Meier survival curve for low and high DOCK3 expression groups indicated in *(A)*. *(C)* Hazard ratio plot in function of DOCK4 expression based on gene expression and recurrence-free survival data for a cohort of 269 ER-negative (ER-) breast cancer patients. Of note, publicly available breast cancer datasets were filtered to include only patients with ER-negative breast cancer, as TGF-β activity was found to be correlated with distant metastasis in ER-negative, but not ER-positive, breast tumors (Padua et al. 2008). *(D)* Kaplan-Meier survival curve for low and high DOCK4 expression groups indicated in *(C)*. Dotted lines in A and C indicate the cut-offs at the points for the highest hazard ratio with confidence interval 95% to define low/high expression groups.

	Univariate analysis					Multivariate analysis			
Covariates	**HR**	**Low 95% CI**	**Up 95% CI**	**P-value**		**HR**	**Low 95% CI**	**Up 95% CI**	**P-value**
Age									
< 61	1								
> 61	1.508	0.911	2.495	0.11					
DOCK3 expression (Z-score)									
Low (<1.15)	1								
High (>1.15)	1.827	0.966	4.863	0.061					
DOCK4 expression (Z-score)									
Low (<0)	1					1			
High (>0)	2.074	1.295	3.297	**0.0024**		1.907	1.18	3.018	**0.0084**
Final stage									
< III	1					1			
> III	2.46	1.537	3.95	**0.00011**		2.342	1.46	3.755	**0.00041**
Gender									
Male	1								
Female	1.359	0.8552	2.159	0.193					
Smoke									
Yes	1								
No	0.8507	0.465	1.556	0.599					
Race									
Asian	1								
Caucasian	0.90824	0.2836	2.909	0.871					
African American	1.80796	0.3646	8.966	0.469					
Hispanic	1.075	0.2567	4.502	0.921					

Table 1: Cox proportional univariate and multivariate analyses of recurrence free survival in a cohort of 182 lung ADC patients. (HR = hazard ratio; CI = confidence interval)

		DOCK4 level	
		Low	High
Recurrence	Yes	26 (28%)	46 (52%)
	No	68 (72%)	42 (48%)
	Fisher's exact test:	*****p<0.001**	

		DOCK4 level	
		Low	High
Tumor stage	< III	73 (78%)	56 (64%)
	≥ III	21 (22%)	32 (36%)
	Fisher's exact test:	***p<0.05**	

Table 2: Contingency analysis table demonstrating the correlation between DOCK4 expression level, frequency of recurrence, and tumor stage in a cohort of 182 lung ADC patients.

Chapter 3

Evidence that DOCK4 mediates TGF-β driven lung ADC metastasis

3.1 Establishment of a xenograft model for studying TGF-β driven lung ADC metastasis *in vivo*

Based on our above findings, and previous reports correlating increased TGF-β expression with tumor progression and metastasis in lung ADC (Hasegawa et al. 2001; Nemunaitis et al. 2009; Vazquez et al. 2013), we next asked whether DOCK4 mediates the pro-metastatic effects of TGF-β in lung ADC. To address this, we first established an experimental model for the analysis of lung ADC metastasis. Because lung ADCs typically metastasize to multiple organs (including bone, adrenal gland, brain, and liver) (Nguyen et al. 2009), with lung ADC cells released from the primary site traveling via the arterial circulation to distant organ sites, we opted to use an intracardiac injection model of experimental metastasis, in which cancer cells are injected into the left cardiac ventricle of NOD/SCID-IL2γ (NSG) mice (Fig. 6A). Using this model, we assessed the metastatic potential of both A549 (KRAS-mutant) and HCC4006 (EGFR-mutant) cells, and evaluated whether pre-exposure of these cells to TGF-β prior to their introduction into the arterial circulation (to "mimic" the source of TGF-β that tumor cells normally experience within the primary tumor microenvironment) increases their metastatic potential. Specifically, A549 and HCC4006 cells engineered to stably express firefly luciferase (A549-luc and HCC4006-luc) were either left untreated or treated with TGF-β for 24 h, and placed in the arterial circulation of NSG mice by intracardiac injection. Mice were subsequently monitored for multi-organ metastasis by bioluminescence imaging. While both untreated A549-luc and HCC4006-luc cells formed some metastases in multiple organs, with A549-luc cells preferentially colonizing the liver and bone and HCC4006-luc cells adrenal glands, we noted that

the metastatic burden in animals was markedly increased when the A549-luc and HCC4006-luc cells were pretreated with TGF-β prior to injection (Fig. 6B), indicating that TGF-β stimulation enhances the metastatic potential of both lung ADC cell lines. Metastases in adrenal gland, bone, and liver were confirmed by *ex vivo* bioluminescence imaging and histology (Fig. 6C).

3.2 DOCK4 is essential for TGF-β driven lung ADC metastasis *in vivo*

Having established an experimental model for lung ADC metastasis, we next assessed the requirement of DOCK4 in TGF-β-driven lung ADC metastasis. To approach this, we generated a retroviral vector that co-expresses EGFP and a miR-30-based short-hairpin RNA (shRNA) targeting the 3' untranslated region (UTR) of *DOCK4* mRNA (shDock4#1). This shRNA substantially reduced DOCK4 protein levels when stably introduced into A549-luc or HCC4006-luc cells, whereas a control shRNA (shCtrl) had no effect. Moreover, and importantly, shDock4#1 largely blunted induction of DOCK4 expression by TGF-β (Fig. 6D, and Fig. 7A). Before assessing the metastatic potential of the shRNA-expressing cells, we first scrutinized their proliferative properties, especially since DOCK4 had been reported to display tumor-suppressive activity in osteosarcoma cells (Yajnik et al. 2003). We found that DOCK4 knockdown did not affect the growth rate of A549 or HCC4006 lung ADC cells, regardless of them being pretreated with TGF-β (Fig. 6E, and Fig. 7B). Moreover, when shRNA-expressing cells, pretreated or not pretreated with TGF-β, were implanted into the lungs of NSG mice via intrathoracic injection, no difference in tumor growth rate was observed among the 4 experimental groups (Fig. 6F), indicating that DOCK4 knockdown does not affect the ability of these cells to form primary pulmonary tumors. Also, DOCK4 knockdown did not affect the survival of these cells in either monolayer or suspension culture (Fig. 7C,D, E, and data not shown).

We then injected the shRNA-expressing cells, pretreated or not pretreated with TGF-β, into the arterial circulation of NSG mice, and monitored the mice for metastases. Strikingly, DOCK4 knockdown markedly blunted the pro-metastatic effects of TGF-β in both A549 and HCC4006 cells. Compared to animals injected with TGF-β-pretreated shCtrl-expressing cells, the metastatic burden in animals injected with TGF-β-pretreated shDock4#1-expressing cells was markedly reduced (Fig. 6G,H). Importantly, rescue experiments using DOCK4 cDNA that lacks the 3'UTR and is therefore resistant to Dock4 shRNA-mediated RNA interference (Fig. 6D) demonstrated that the effects of DOCK4 RNAi were specific (Fig. 6G; Rescue). Expression of DOCK4 alone (to levels similar to those induced by TGF-β) did not alter the metastatic potential or growth properties of lung ADC cells, and accordingly did not affect the metastatic burden in animals (Fig. 8A-C), implying that a pathway(s) parallel to DOCK4 contribute(s) to TGF-β's pro-metastatic effects (see further below). Of note, we also saw a slight decrease in metastatic burden in animals injected with TGF-β-untreated shDock4#1-expressing cells as compared to animals injected with untreated shCtrl-expressing cells (Fig. 6G,H). We presume that the lung ADC cells, when placed in circulation, can still become exposed to TGF-β signals provided by platelets in the bloodstream (Labelle et al. 2011). Clearly though, compared to TGF-β-pretreated lung ADC cells, the metastatic efficiency of untreated cells is much lower, supporting the notion that pre-exposure of cells to TGF-β ("mimicking" the source of TGF-β at the primary site) primes these cells for efficient metastasis as they enter the circulation. Together, these data unveil an essential role for DOCK4 in mediating the metastasis-promoting activity of TGF-β in lung ADC cells from circulation to distant organ sites.

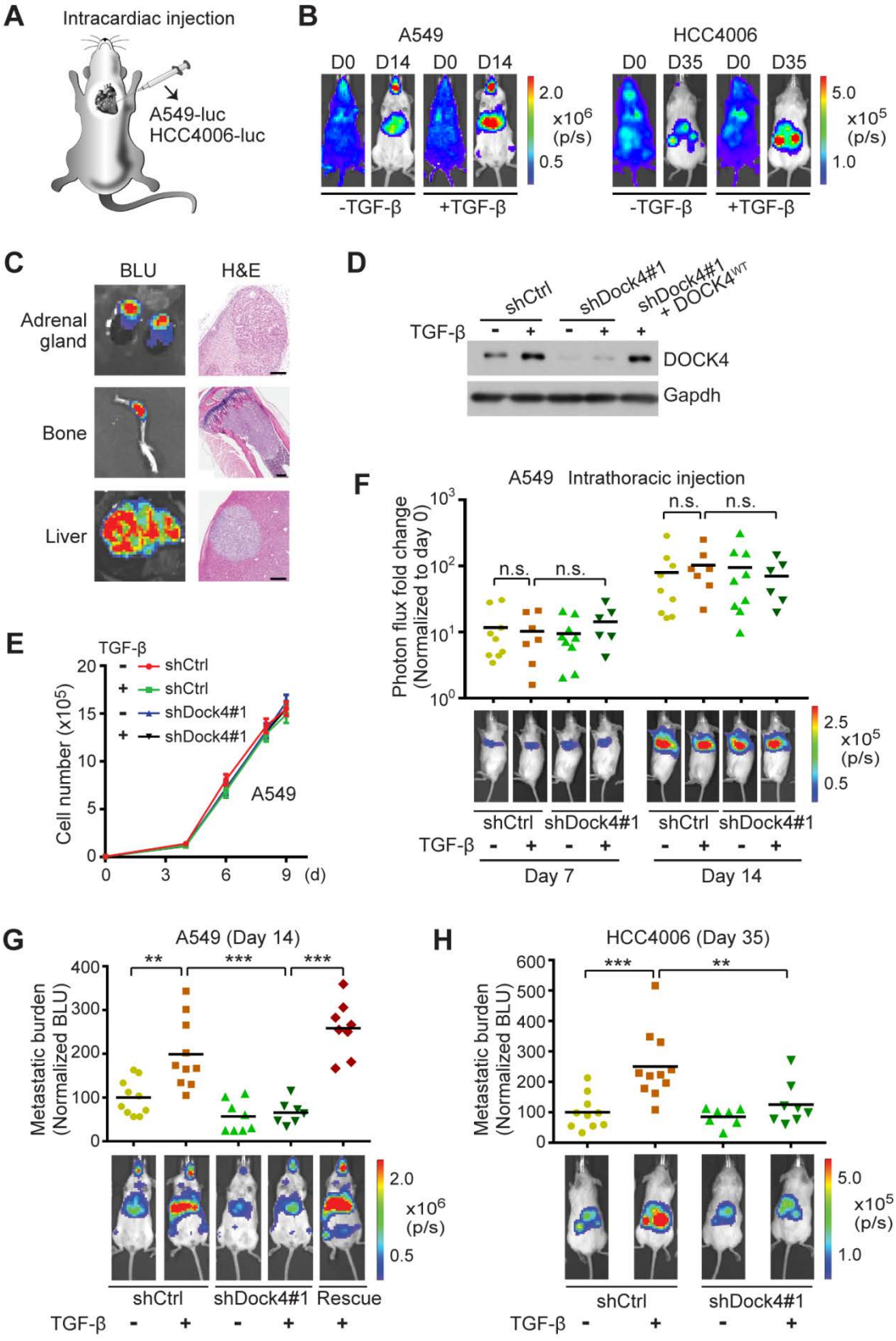
A
Intracardiac injection
A549-luc
HCC4006-luc
B
A549
HCC4006
D0
D14
D35
2.0
x10^6 (p/s)
0.5
5.0
x10^5 (p/s)
1.0
-TGF-β
+TGF-β
C
BLU
H&E
Adrenal gland
Bone
Liver
D
shCtrl
shDock4#1
shDock4#1 + DOCK4^WT
TGF-β
DOCK4
Gapdh
E
TGF-β
shCtrl
shDock4#1
A549
Cell number (x10^5)
(d)
F
A549 Intrathoracic injection
Photon flux fold change (Normalized to day 0)
n.s.
2.5
x10^5 (p/s)
0.5
Day 7
Day 14
G
A549 (Day 14)
Metastatic burden (Normalized BLU)
**
***
2.0
x10^6 (p/s)
0.5
Rescue
H
HCC4006 (Day 35)
5.0
x10^5 (p/s)
1.0

Figure 6. DOCK4 is required for TGF-β-driven lung ADC metastasis. (*A*) Schematic drawing of intracardiac injection. (*B*) Bioluminescent (BLU) images of NSG mice intracardially injected with TGF-β pretreated (24 h, +) or untreated (-) A549 and HCC4006 cells, taken at indicated days post-injection. (*C*) BLU images and H&E staining of metastases in indicated organs harvested from mice in (*B*). Scale bars, 200 μm. (*D*) Western blot analysis of DOCK4 in A549 cells expressing indicated constructs. (*E*) Growth curve of TGF-β pretreated (24 h, +) or untreated (-) A549 cells stably expressing shCtrl or shDock4#1. Data represent mean ± SD (n = 3). (*F*) Analysis of lung tumor growth in NSG mice intrathoracically injected with TGF-β pretreated (24 h, +) or untreated (-) A549 cells expressing shCtrl or shDock4#1. Top, dot plots of lung photon flux at days 7 and 14. BLU signals were normalized to day 0. Bottom, representative images of mice with lung tumors. n = 6-9 mice/condition. (*G*, *H*) Analysis of metastatic burden in NSG mice intracardially injected with TGF-β pretreated (24 h, +) or untreated (-) A549 (*G*) or HCC4006 (*H*) cells expressing indicated constructs. Top, dot plots of metastatic burden at day 14 (A549) and day 35 (HCC4006). BLU signals were normalized to day 0, and then to a value of 100 for control condition (shCtrl, TGF-β -). Bottom, representative images of mice with metastases. n = 7-10 (*G*) and 7-11 (*H*) mice/condition. P values in *F*, *G* and *H* were calculated by an unpaired Mann-Whitney test. $^{**}P < 0.01$, $^{***}P < 0.001$, n.s., $P \geq 0.05$.

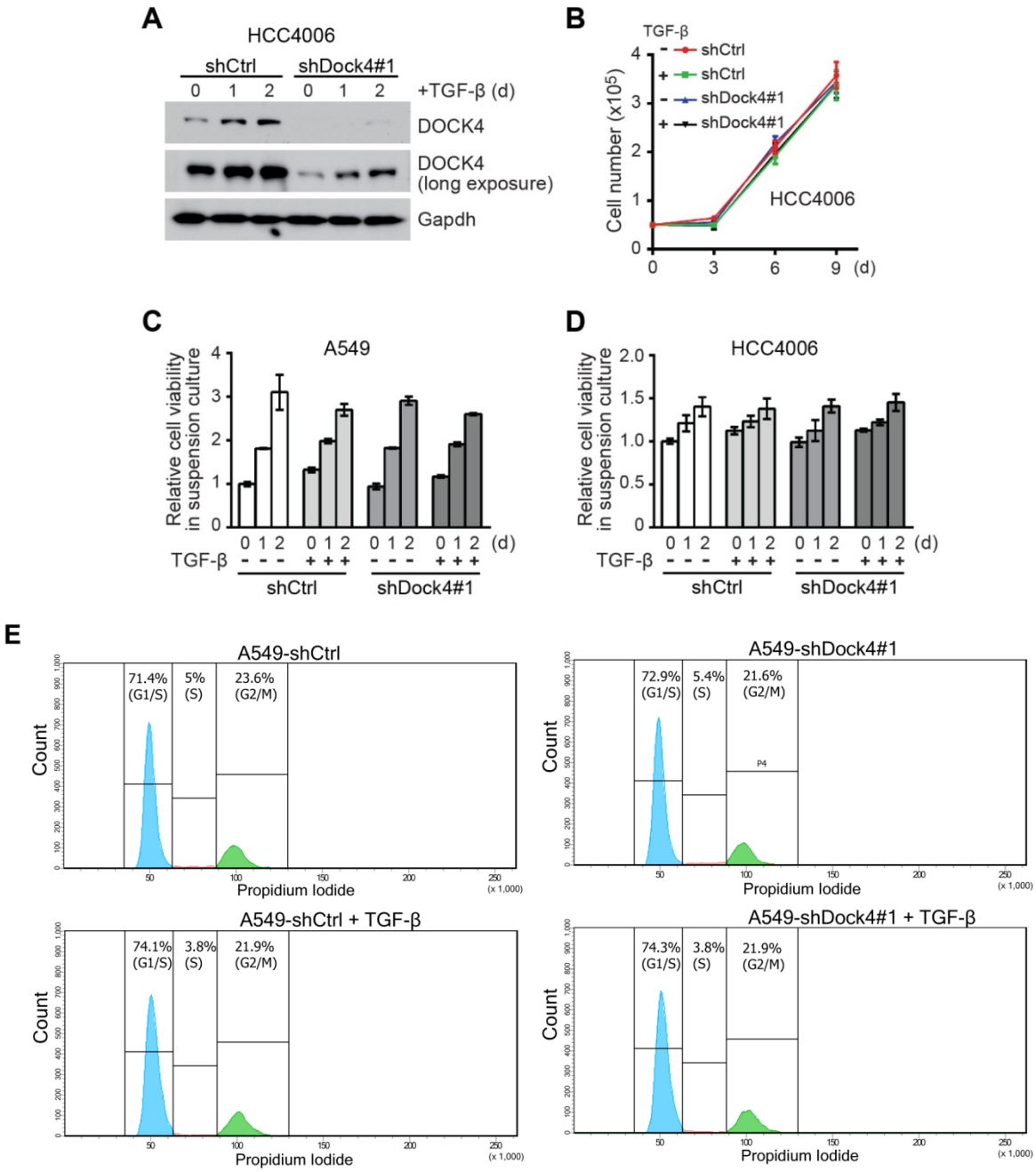

Figure 7. DOCK4 is dispensable for lung ADC cell proliferation in 2D culture and viability in suspension culture. (*A*) Western blot analysis of DOCK4 in shCtrl- and shDock4#1-expressing HCC4006 cells treated with 2 ng/ml TGF-β over a 2-d time window. Gapdh was used as a loading control. (*B*) Growth curve of TGF-β pretreated (24 h, +) and untreated (-) shCtrl- and shDock4#1-expressing HCC4006 cells over a 9-d time window. (*C*,*D*) MTT cell viability assay of TGF-β pretreated (24 h, +) and untreated (-) shCtrl- and shDock4#1-expressing A549 (*C*) and HCC4006 (*D*) cells in suspension culture over a 2-d time window. Data in B, C and D represent mean ± SD (n = 3). (*E*) Cell cycle analysis following PI staining of TGF-β pretreated (24 h, +) and untreated (-) shCtrl- and shDock4#1-expressing A549 cells.

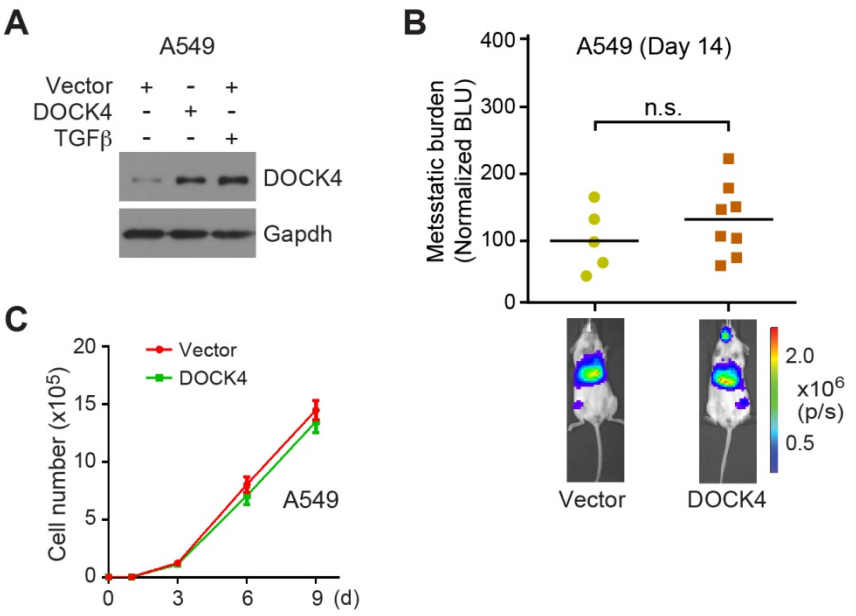

Figure 8. Ectopic expression of DOCK4 does not alter the metastatic potential or growth properties of lung ADC cells. (*A*) Western blot analysis of DOCK4 expression in vector- and DOCK4-expressing A549 cells, and in vector-expressing A459 cells treated with 2 ng/ml TGF-β for 24 h. (*B*) Analysis of metastatic burden in NSG mice intracardially injected with vector- or DOCK4-expressing A549 cells. Top, dot plots of metastatic burden at day 14. BLU signals were normalized to day 0, and then to a value of 100 for vector control condition. Bottom, representative images of mice with metastases. $n = 5$ (vector) and 8 (DOCK4) mice. *P* value was calculated by an unpaired Mann-Whitney test. n.s., $P \geq 0.05$. (*C*) Growth curve of A549 cells stably expressing vector or DOCK4 over a 9-d time window. Data represent mean ± SD ($n = 3$).

3.3 DOCK4 depletion inhibits TGF-β driven tumor cell extravasation but does not affect cell proliferation and tumor growth at both primary and metastatic sites

For circulating cancer cells to establish distant metastases, they must leave the circulation in a process called extravasation at distant organ sites, and then after infiltrating the new tissue they must acquire the ability to survive and proliferate in this new microenvironment in order to form macroscopic metastases (Valastyan and Weinberg 2011; Scott et al. 2012). To explore whether depletion of DOCK4 in the *in vitro*TGF-β primed lung ADC cells impact their extravasation capabilities and/or their ability to survive/proliferate at the distant organ sites, we began by scoring the number and size of metastatic nodules formed in livers of mice that were intracardially injected with TGF-β-pretreated or untreated shCtrl- or shDock4#1-expressing A549-luc cells (Fig. 9A). In line with our above findings, pretreatment of shCtrl-expressing cells, but not of shDock4#1-expressing cells, with TGF-β prior to injection led to a marked increase in the number of metastatic nodules (Fig. 9A). Interestingly, though, the size of the metastatic nodules that developed in the livers was not significantly different among the 4 experimental groups (Fig. 9B), implying that TGF-β pretreatment and, importantly, DOCK4 depletion do not alter the capacity of lung ADC cells to grow in the new microenvironment. To substantiate this, we first checked that the liver metastases formed from shDock4#1-expressing cells were not attributable to proliferation of tumor cells that lost shRNA expression. To this end, we extracted EGFP-labeled tumor cells from liver metastases originating from TGF-β-pretreated shCtrl- and shDock4#1-expressing A549-luc cells (referred to as liver mets-derived cells), and assessed DOCK4 levels by Western blot analysis. We found that DOCK4 levels were still efficiently knocked down in these cells (Fig. 9C). Moreover, when we assessed their proliferative properties *in vitro*, we did not detect any differences between the growth rates of shDock4#1- and shCtrl-

expressing liver mets-derived cells (Fig. 9D). In addition, when we injected shDock4#1- or shCtrl-expressing A549-luc cells, pretreated or not pretreated with TGF-β, into livers of NSG mice and evaluated tumor growth 7 and 14 d following injection, we found that the growth rate of the tumors was similar among all the experimental groups (Fig. 9E). Thus, DOCK4 depletion in *in vitro* TGF-β primed lung ADC cells does not affect their ability to survive/proliferate in a new microenvironment.

Hence, we next assessed whether TGF-β/DOCK4 signaling influences the capacity of lung ADC cells to extravasate into distant target organs. To this end, we isolated livers from NSG mice that were sacrificed 20 h after intracardiac injection with TGF-β-pretreated or untreated shCtrl- or shDock4#1-expressing A549-luc cells. Of note, we chose the liver because it is a relatively large and highly vascularized organ. Livers were perfused and liver sections immunostained for GFP and CD31 to visualize tumor cells and liver vasculature, respectively (Fig. 9F). The percentage of tumor cells inside (intravascular) and outside (extravascular) the blood vessels were then quantified (Fig. 9G). We found that the fraction of tumor cells that extravasated out of the liver vasculature was markedly increased in the TGF-β-pretreated shCtrl group, as compared to the untreated shCtrl group, supporting the notion that TGF-β promotes tumor cell extravasation. No such increase, however, was observed in the TGF-β-pretreated shDock4#1 group, where similar to the untreated shDock4#1 and shCtrl groups, the majority of tumor cells remained in the blood vessels (Fig. 9G), indicating that DOCK4 function is required for TGF-β's enhancing effect on tumor cell extravasation. Noteworthy, DOCK4 levels remained high in the TGF-β-pretreated cells for at least 48 h following TGF-β removal (Fig. 10), which is well within the time period needed for these cells to extravasate following intracardiac injection.

Combined, these data establish a critical role for the TGF-β/DOCK4 signaling axis in the regulation of lung ADC cell extravasation and metastasis *in vivo*.

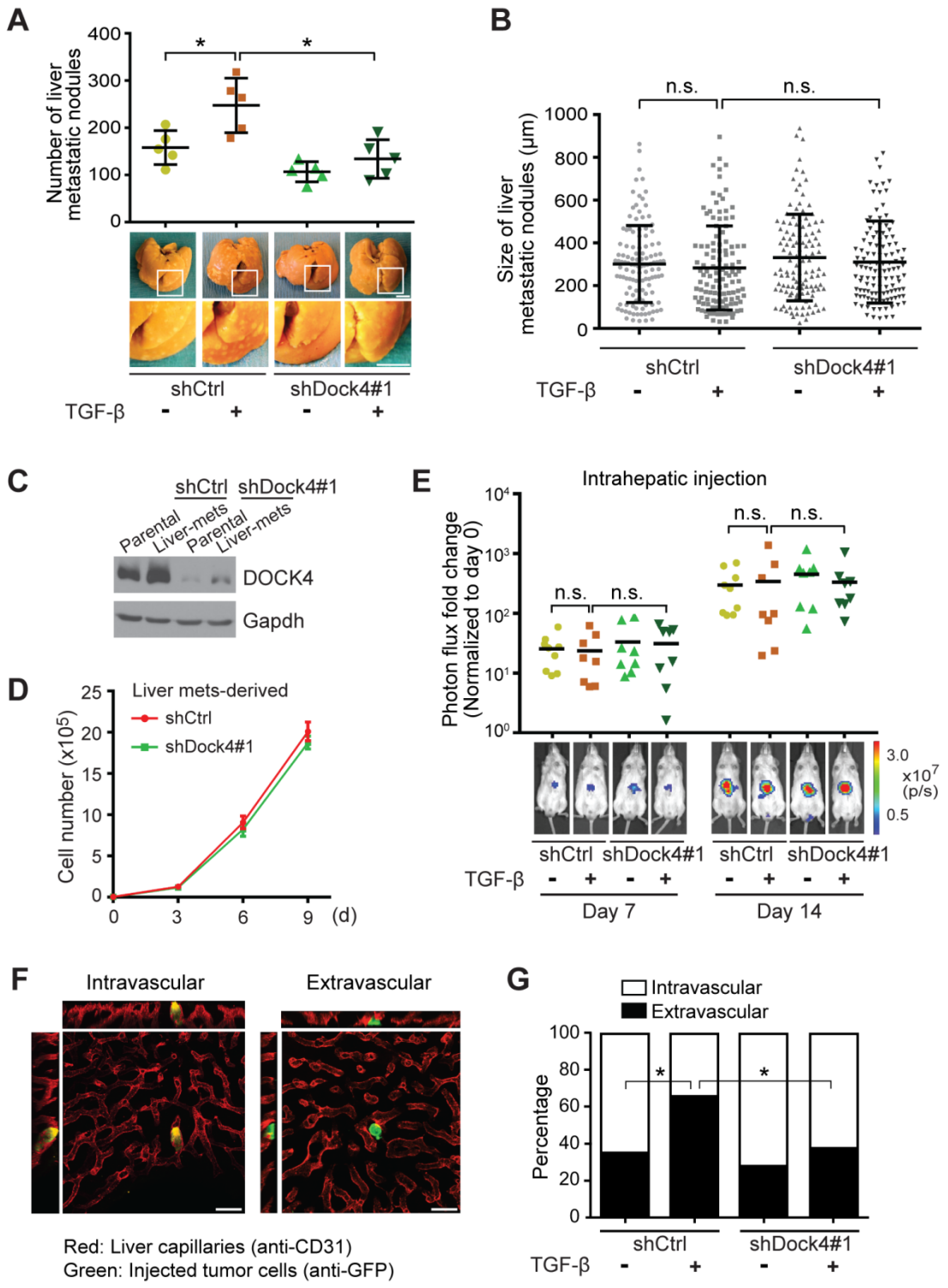

Figure 9. DOCK4 depletion inhibits TGF-β-driven lung ADC cell extravasation, but does not affect ability of lung ADC cells to grow in distant organs. (*A*) Quantification of number of metastatic nodules in liver after intracardiac injection of TGF-β pretreated (24 h, +) or untreated

(-) A549 cells expressing shCtrl or shDock4#1. Top, number of metastatic nodules on liver surface (n = 5 livers/condition). Bottom, representative images of livers. Boxed regions are enlarged and shown in the bottom row. Scale bars, 1 cm. (*B*) Quantification of size (by measuring diameter) of metastatic nodules in H&E stained liver sections from mice in (*A*). n = 120/condition. Data in *A* and *B* represent mean ± SD. (*C*) Western blot analysis of DOCK4 protein expression in A549 parental and liver-mets derived cells. Gapdh was used as a loading control. (*D*) Growth curve of shCtrl- and shDock4#1-expressing A549 liver-mets derived cells. Data represent mean ± SD (n = 3). (*E*) Analysis of tumor growth in livers of NSG mice intrahepatically injected with TGF-β pretreated (24 h, +) or untreated (-) A549 cells expressing shCtrl or shDock4#1. Top, dot plots of liver photon flux at days 7 and 14, normalized to day 0. Bottom, representative images of mice with liver tumors. n = 8-9 mice/condition. P values in *A, B,* and *E* were calculated by an unpaired Mann-Whitney test. *$P < 0.05$, n.s., $P \geq 0.05$. (*F*) Representative confocal images of liver sections depicting tumor cell inside (intravascular) or outside (extravascular) blood vessel, obtained from NSG mice 20 h after intracardiac injection with cells in *A.* Scale bars, 20 μm. (*G*) Percentage of intravascular or extravascular A549 cells from the experiment shown in (*F*). n = 31-40 cells/condition. *$P < 0.05$ by Fisher's exact test.

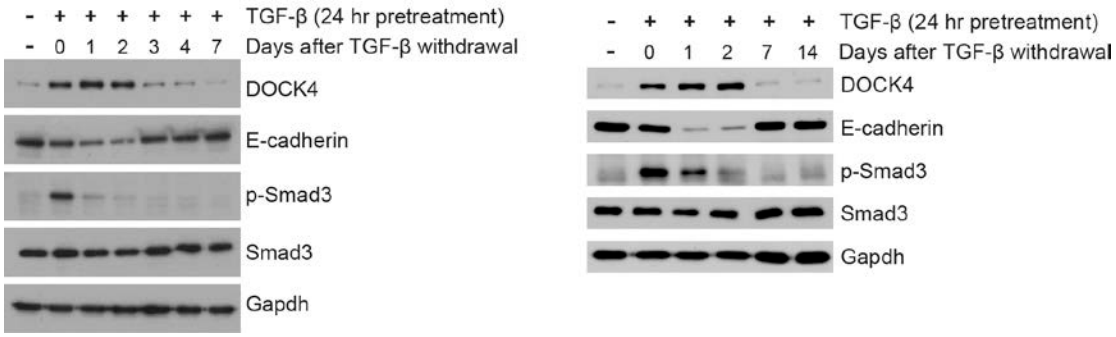

Figure 10. Downregulation of TGF-β signaling activity, and reversal of E-cadherin and DOCK4 expression following the removal of TGF-β. Western blot analyses of DOCK4, E-cadherin, p-Smad3, and Smad3 in A549 cells treated with TGF-β for 24 h followed by TGF-β withdrawal at indicated time points over a 7-d (left) and 14-d time window (right). Gapdh was used as a loading control.

3.4 DOCK4 mediates TGF-β's enhancing effects on lung ADC cell protrusive activity, motility, and invasion, but not EMT, via Rac1 activation

We next sought to gain further insight into the cellular mechanisms by which DOCK4 mediates TGF-β's enhancing effect on tumor cell extravasation and metastasis. While the cellular underpinnings of tumor cell extravasation remain still poorly understood, increasing evidence indicates it to be a dynamic process involving not only changes of the vascular endothelium, but also of the tumor cells during their intravascular transit to the sites of metastasis, with tumor cells undergoing changes in cell shape and migratory behavior (Strell and Entschladen 2008; Stoletov et al. 2010; Reymond et al. 2013). TGF-β has been implicated in most of these processes (Giampieri et al. 2009; Valastyan and Weinberg 2011), and, interestingly, recent studies reported that TGF-β-induced epithelial-mesenchymal transition (EMT) not only facilitates tumor cells to intravasate, but also helps them to extravasate (Stoletov et al. 2010; Labelle et al. 2011; Tsai and Yang 2013; Yu et al. 2013). Based on these findings, we first explored whether DOCK4 affects TGF-β-induced EMT in A549 and HCC4006 lung ADC cells. To this end, we examined the impact of DOCK4 knockdown on the expression levels of TGF-β-responsive genes known to be involved in EMT (including E-cadherin, Vimentin, Snail, Slug, Twist1, Zeb1/2) (Valastyan and Weinberg 2011). In both cell lines, TGF-β induced a downregulation and/or cytoplasmic translocation of E-cadherin, and an upregulation of Vimentin, Snail and Slug, as well as an EMT phenotype (Fig. 11A, and Fig. 12A-D). Knockdown of DOCK4 did not affect any of the TGF-β-induced changes in EMT markers, nor did it prevent the acquisition of a mesenchymal-like phenotype (Fig. 12A-D). Also, ectopic expression of DOCK4 did not alter any of these properties (Fig. 12E, and data not shown). These data indicate that

DOCK4 is dispensable for TGF-β-induced EMT in A549 and HCC4006 lung ADC cells, and further imply that TGF-β drives EMT and DOCK4 induction via parallel pathways.

We next assessed whether DOCK4 knockdown influences the migratory and invasive behavior of the lung ADC cells. Since the majority of circulating tumor cells appears to consist of single cells (Yu et al. 2013), we tracked the movement of single cells, pretreated or not pretreated with TGF-β. We observed that TGF-β pretreatment of shCtrl-expressing A549 or HCC4006 cells greatly enhanced the motility of these cells, and that this enhancement was abrogated upon inhibition of the TGF-β/Smad pathway (Fig. 11B,C, and Fig. 13B-E). Importantly, while DOCK4 knockdown, using two independent shRNAs (shDock4#1 and shDock4#2), did not affect the basal levels of A549 or HCC4006 cell motility, it impeded, similarly as seen for TGF-β/Smad inhibition, the TGF-β-induced increase in cell motility (Fig. 11B,C and Fig. 13A-E). We further examined the invasive potential of these cells using a Matrigel-coated Boyden chamber assay, given that tumor cells must invade the basement membrane surrounding the blood vessels to enter the parenchyma of their target organs (Reymond et al. 2013). While TGF-β treatment of shCtrl-expressing A549 cells resulted in a robust increase in invasion through Matrigel, only a very modest increase was observed when shDock4#1- or shDock4#2-expressing cells were treated with TGF-β (Fig. 11D). Thus, DOCK4 function is essential for TGF-β's stimulating effects on lung ADC cell motility and invasion. Of note, a role for DOCK4 in the migration of distinct cell types had been reported before (Hiramoto et al. 2006; Kawada et al. 2009; Kobayashi et al. 2014). Interestingly, a closer scrutiny of the morphology of the shRNA-expressing cells in our live cell imaging experiments revealed that TGF-β-pretreated shCtrl-expressing cells displayed higher membrane protrusive activity than the untreated shCtrl-expressing cells, with TGF-β-pretreated cells extending a large

forward protrusion (Fig. 11E,G). In contrast, while shDock4#1- and shDock4#2 expressing cells did undergo EMT when exposed to TGF-β, they hardly extended forward protrusions (Fig. 11E,G and data not shown), indicating that DOCK4's function is important for TGF-β-promoted mesenchymal cancer cell protrusive activity.

Finally, we asked whether DOCK4 exerts its cellular effects by acting on the Rac1 signaling pathway or potentially other pathways. On the one hand, Rac1 activation has been implicated in cell motility and protrusion extension during trans-endothelial migration (Reymond et al. 2013); on the other hand, however, we found that TGF-β induces DOCK4 expression via the canonical Smad pathway and so far Rac1 has been mainly linked to TGF-β via non-canonical pathway(s) (Zhang 2009). Hence, we first determined whether Rac1 activation is mediated via the TGF-β/Smad/DOCK4 pathway in lung ADC cells. While TGF-β triggered a robust increase in Rac1 activity in shCtrl-expressing A549 cells, we found that this increase was greatly reduced in both shSmad4#2- and shDock4#1-expressing A549 cells (Fig. 14A,B,D). Of note, DOCK4 knockdown did not affect TGF-β-induced activation of Rap1 or Cdc42 in A549 cells (Fig. 14C). Furthermore, we found that the DHR2 domain of DOCK4, which is conserved among all DOCK180 family members and catalyzes the exchange of GDP for GTP on Rac1 (Cote and Vuori 2002; Meller et al. 2005), is essential for TGF-β-elicited Rac1 activation in A549 cells. Indeed, introduction of $DOCK4^{WT}$, but not a $DOCK4^{\Delta DHR2}$ mutant lacking 77 amino acids within the DHR2 domain (Kawada et al. 2009), in shDock4#-expressing cells restored the levels of TGF-β-induced Rac1 activation to that seen in control cells (Fig. 14A,B). Thus, DOCK4 links the canonical TGF-β/Smad pathway to Rac1 activation in lung ADC cells. These findings prompted us to investigate whether DOCK4 exerts its cellular effects via activation of the Rac1 pathway. We first examined the ability of $DOCK4^{\Delta DHR2}$ to rescue the impaired protrusive activity

and cell motility observed in the TGF-β-treated shDock4#1-expressing cells. While DOCK4^{WT} was able to fully rescue these phenotypes, DOCK4$^{\Delta DHR2}$ failed to do so (Fig. 11E-G), indicating that DOCK4's Rac-GEF activity is essential for its function in mediating TGF-β's effects on cell motility and protrusion formation. Of note, expression of DOCK4^{WT} alone did not affect the motility or protrusive activity of these cells (Fig. 13F), in consistency with our findings that it does not affect their metastatic potential. We next examined whether concomitant expression of an activated mutant form of Rac1 (Rac1^{Q61L}) with shDock4#1 could rescue the DOCK4 RNAi-produced phenotypes, and found that this is indeed the case. Cells co-expressing shDock4#1 and Rac1^{Q61L} extended protrusions and displayed increased cell motility upon TGF-β treatment (Fig. 11E-G). Finally, we investigated whether knockdown of Rac1 phenocopies the effects of DOCK4 downregulation on TGF-β-promoted cell motility and protrusive activity. To this end, we took advantage of two previously described shRNAs (shRac1#1 and shRac1#2) (Akunuru et al. 2011), with shRac1#1 being more effective than shRac1#2 (Fig. 14E). We found that both shRac1#1 and shRac1#2 interfered with TGF-β-induced increase in cell motility and protrusive activity, with as expected shRac1#1 being more effective than shRac1#2 (Fig. 14F and data not shown). Thus, DOCK4 links the canonical TGF-β/Smad pathway to the Rac1 pathway in the regulation of lung ADC cell shape and migratory behavior, processes fundamental to tumor cell extravasation and metastasis.

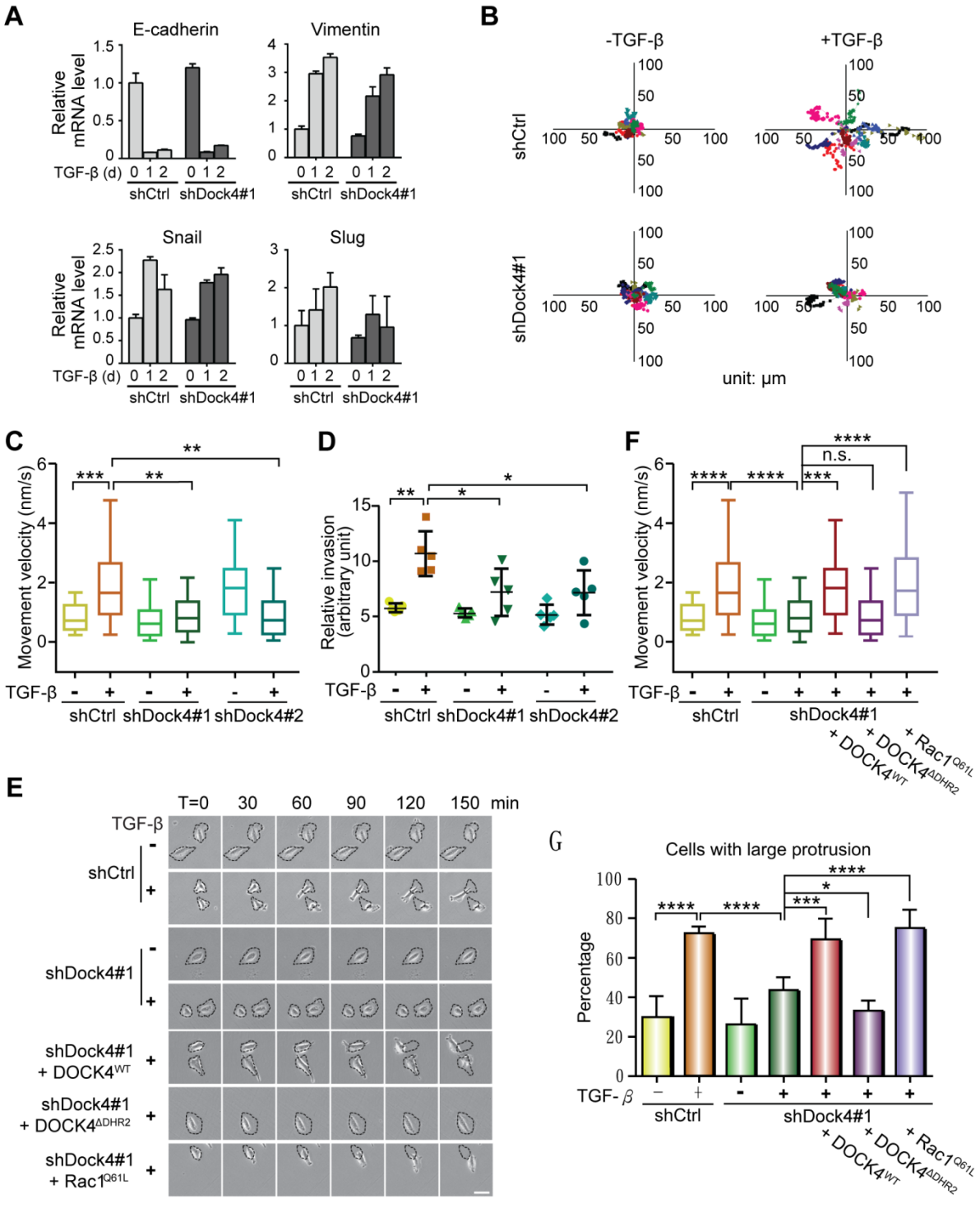

Figure 11. DOCK4 mediates TGF-β's enhancing effects on lung ADC cell protrusive activity, motility, and invasion, but not EMT, via Rac1 activation. (*A*) qPCR analysis of mRNAs of EMT markers in shCtrl- and shDock4#1-expressing A549 cells treated with TGF-β. Data represent mean ± SD (n = 3). (*B*) Representative movement trajectories of single shCtrl- or shDock4#1-

expressing A549 cells left untreated or pretreated (for 24 h) with TGF-β obtained over 4.5 h. (*C*) Quantification of movement velocity of shDock4#1- and shDock4#2-expressing A549 cells. n = 33-44 cells/condition. (*D*) Quantification of Matrigel invasion assays of shDock4#1- and shDock4#2-expressing A549 cells. Data represent mean ± SD (n = 5 transwells/group). (*E*) Representative images of single A549 cells expressing indicated constructs, left untreated or pretreated with TGF-β for 24 h, obtained from live cell imaging at indicated time points. Scale bars, 50 μm. (*F*) Quantification of movement velocity of cells in (*E*). n = 36-56 cells/condition. Data in *C* and *F* are presented as Tukey box plots. *P* values in *C*, *D*, and *F* were calculated by an unpaired Mann-Whitney test. (*G*) Percentage of cells with a large protrusion over a 30-min time interval (n = 6 fields containing 104-151 cells/condition). Data represent mean ± SD. *P* values were calculated using an unpaired two-tailed Student's *t*-test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, **** $P < 0.0001$, n.s., $P \geq 0.05$.

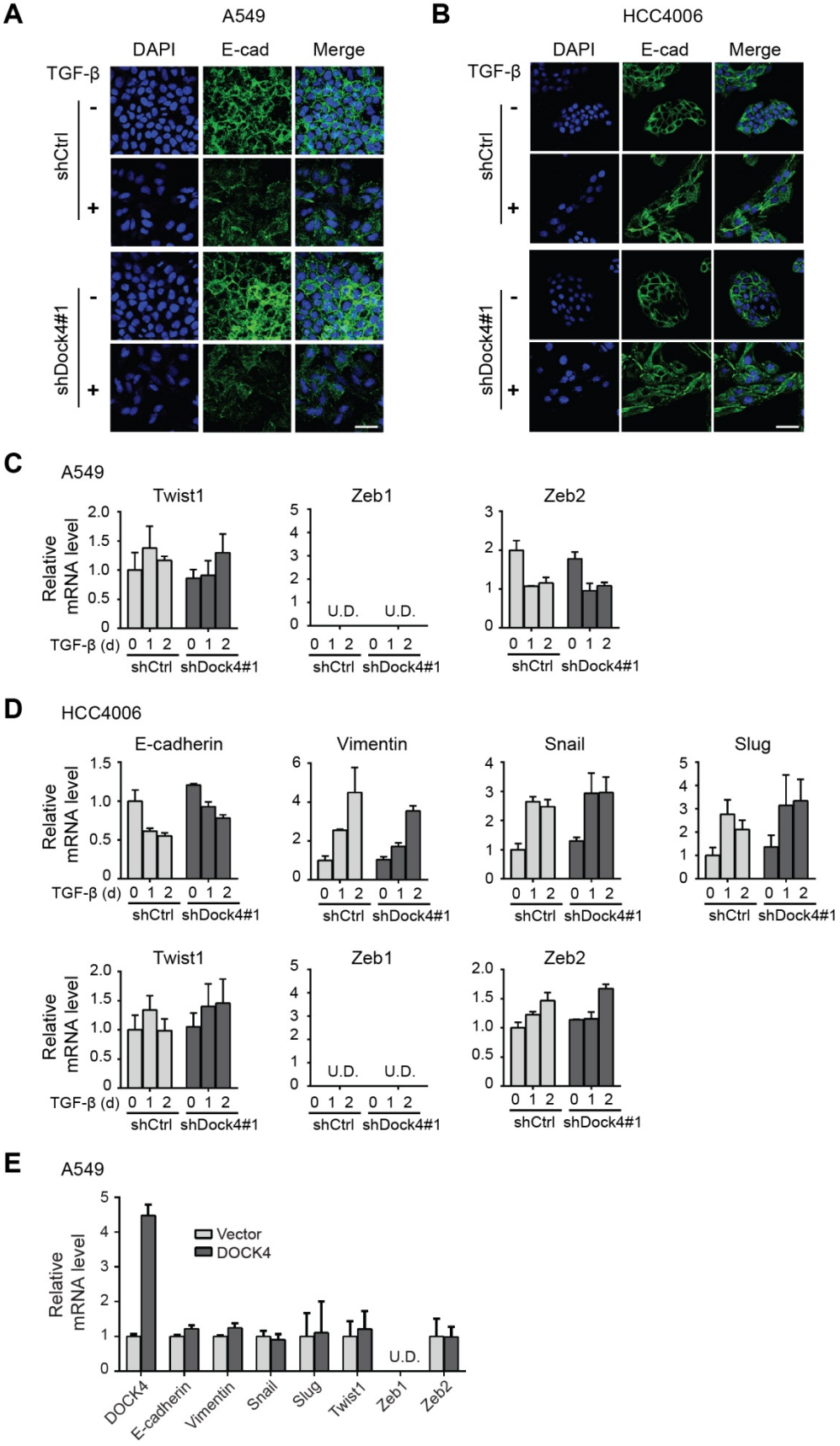
A
A549
DAPI
E-cad
Merge
TGF-β
shCtrl
shDock4#1
B
HCC4006
C
A549
Twist1
Zeb1
Zeb2
Relative mRNA level
U.D.
TGF-β (d) 0 1 2 0 1 2
shCtrl shDock4#1
D
HCC4006
E-cadherin
Vimentin
Snail
Slug
E
A549
Vector
DOCK4
DOCK4
E-cadherin
Vimentin
Snail
Slug
Twist1
Zeb1
Zeb2

Figure 12. DOCK4 is dispensable for TGF-β induced EMT. (*A*,*B*) Representative confocal images of TGF-β treated (24 h, +) and untreated (-) shCtrl- and shDock4#1-expressing A549 (*A*) and HCC4006 (*B*) cells immunostained for E-cadherin (green). Nuclei were counterstained with DAPI (blue). Scale bars, 50 μm. (*C*, *D*) qPCR analysis of indicated EMT markers mRNAs in shCtrl- and shDock4#1-expressing A549 (*C*) and HCC4006 (*D*) cells treated with TGF-β over a 2-d time window. Data represent mean ± SD (n = 3). U.D, undetectable. (*E*) qPCR analysis of indicated EMT markers mRNAs in vector- and DOCK4-expressing A549 cells. Data represent mean ± SD (n = 3). U.D, undetectable.

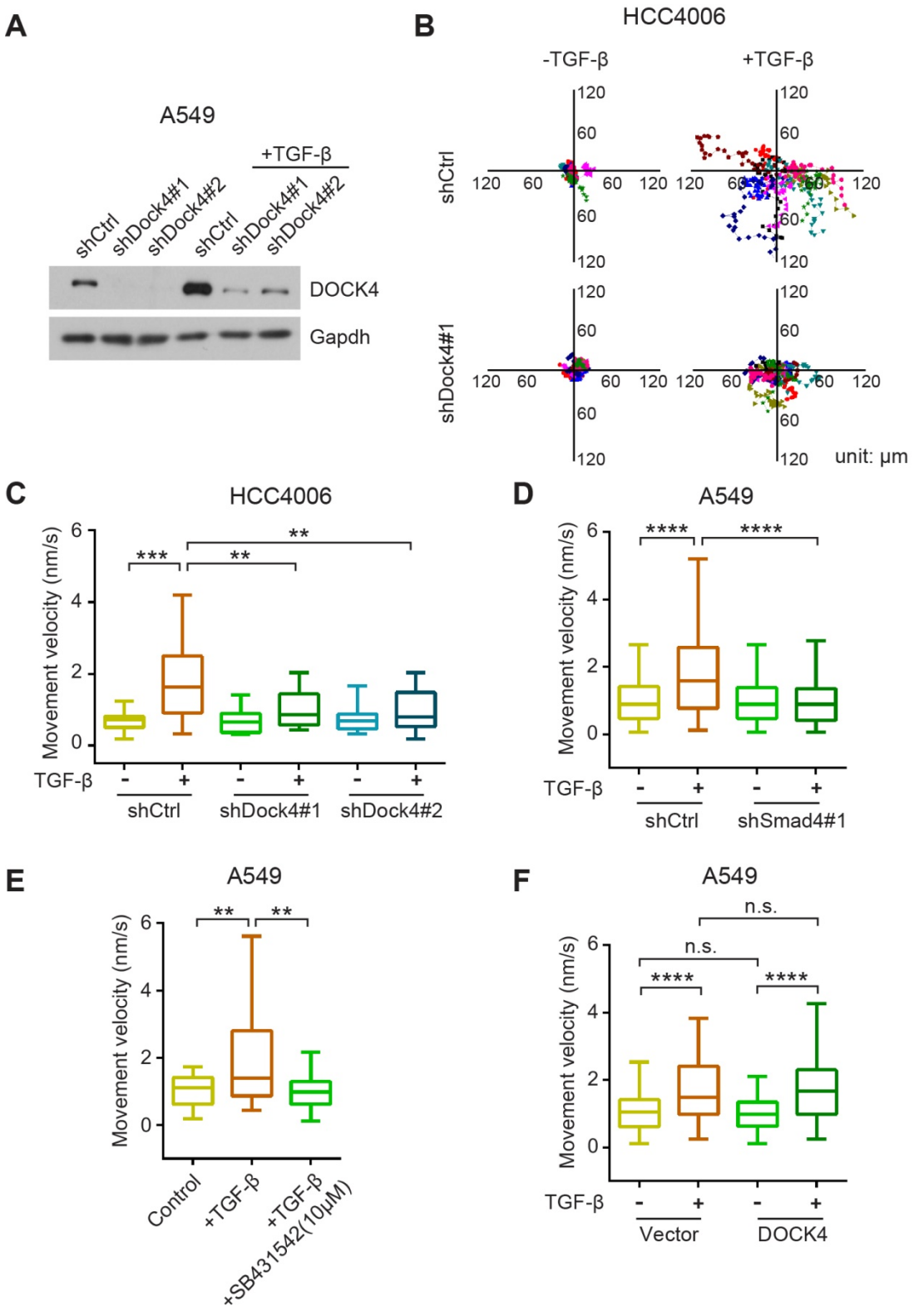

Figure 13. DOCK4 and the Smad signaling pathway mediate TGF-β-induced single cell motility, and ectopic expression of DOCK4 alone does not enhance single cell motility. (*A*) Western blot analysis of endogenous DOCK4 expression in A549 cells expressing shCtrl, shDock4#1, or

shDock4#2 left untreated or treated with 2 ng/ml TGF-β for 24 h. Gapdh was used as a loading control. (*B*) Representative movement trajectories of single shCtrl- or shDock4#1-expressing HCC4006 cells, left untreated or pretreated (24 h) with 2 ng/ml TGF-β, obtained by live-cell imaging over a 4.5-h time window. Each trajectory represents the movement of a single cell, and individual dots designate a frame of 6 min. (*C*) Quantification of velocity of movement of shCtrl-, shDock4#1- or shDock4#2-expressing HCC4006 cells in *B*. n = 14-26 cells/condition. (*D*) Quantification of velocity of movement of shCtrl- or shSmad4#1-expressing A549 cells, left untreated or pretreated (24 h) with 2 ng/ml TGF-β. n = 82-152 cells/condition. (*E*) Quantification of velocity of movement of A549 cells left untreated, pretreated (24 h) with 2 ng/ml TGF-β, or pretreated (24 h) with 2 ng/ml TGF-β and 10 μM SB431542. n = 31-72 cells/condition. (*F*) Quantification of velocity of movement of vector- or DOCK4-expressing A549 cells left untreated or pretreated (24 h) with 2 ng/ml TGF-β. n = 64-129 cells/condition. Data in C, D, E, and F are presented as Tukey box plots, representing the upper quartile + 1.5 interquartile range (IQR; upper end of whisker), upper quartile (top of box), mean (band in box), lower quartile (bottom of box), and lower quartile - 1.5 IQR (lower end of whisker). $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, n.s., $P \geq 0.05$ by an unpaired Mann-Whitney test.

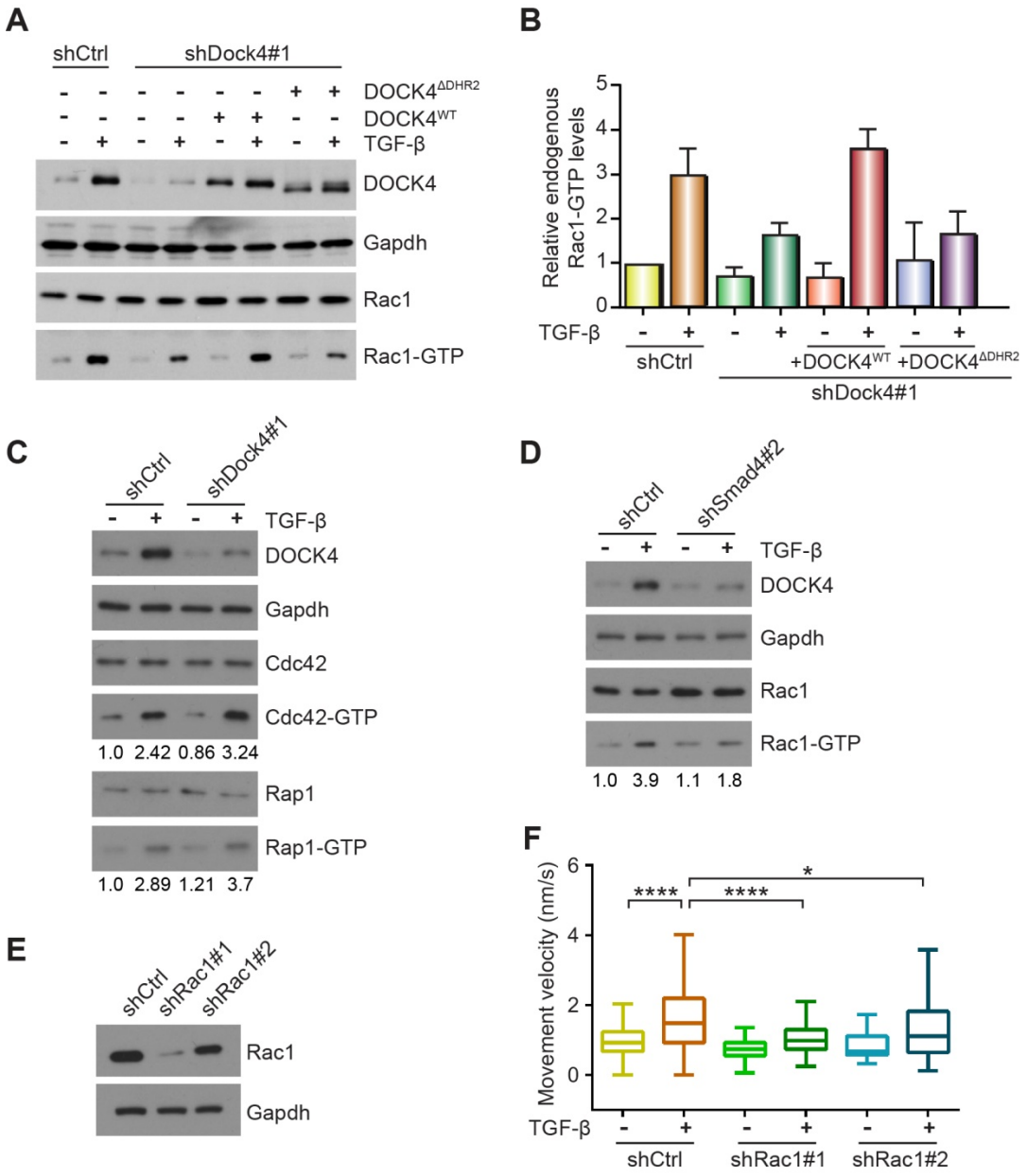

Figure 14. DOCK4 mediates TGF-β-induced Rac1, but not Rap1 or Cdc42, activation. (*A*) Activation of Rac1 measured with p21-activated kinase Rac/Cdc42-binding domain (PBD) pull-down assays. A549 cells expressing shCtrl, shDock4#1, shDock4#1 + DOCK4^{WT}, or shDock4#1 + DOCK4$^{\Delta DHR2}$ were treated with TGF-β (24 h, +) or left untreated (-), and GTP-bound Rac1 was precipitated from detergent extracts with GST-PBD. GST-PBD-bound Rac1-GTP, total Rac1, and DOCK4 levels in cell lysates were detected by immunoblotting with anti-Rac1 and anti-DOCK4 antibodies. Gapdh was included as a loading control. (*B*) Quantification of the ratio of Rac1-GTP versus total Rac1 from 3 independent experiments. Data were normalized to control condition (shCtrl, TGF-β -) and presented as mean ± SD. (*C*) Activation of Rap1 and Cdc42 measured with RalGDS-Rap-binding domain (RBD) and PBD pull-down assays, respectively. A549 cells expressing shCtrl or shDock4#1 were treated with TGF-β (24 h, +) or

left untreated (-), and GTP-bound Rap1 or Cdc42 was precipitated from detergent extracts with GST-RalGDS-RBD or GST-PBD, respectively. GST-RalGDS-RBD-bound Rap1-GTP, total Rap1, GST-PBD-bound Cdc42-GTP, total Cdc42, and DOCK4 levels in cell lysates were detected by immunoblotting with anti-Rap1, anti-Cdc42 and anti-DOCK4 antibodies. The relative amount of Cdc42-GTP and Rap1-GTP in extracts (compared to shCtrl, TGF-β -) is indicated in the bottom panels. (*D*) Activation of Rac1 measured as described in (*A*) in shCtrl- or shSmad4#2-expressing A549 cells left untreated or treated (24 h) with 2 ng/ml TGF-β. GST-PBD-bound Rac1-GTP, total Rac1, and DOCK4 levels in cell lysates were detected by immunoblotting with anti-Rac1 and anti-DOCK4 antibodies. The relative amount of Rac1-GTP in extracts (compared to shCtrl, TGF-β -) is indicated in the bottom panel. Data shown in *C* and *D* are representative of three independent experiments. (*E*) Western blot analysis of endogenous Rac1 protein in A549 cells expressing shCtrl, shRac1#1, or shRac1#2. (*F*) Quantification of velocity of movement of shCtrl-, shRac1#1-, or shRac1#2-expressing A549 cells left untreated or pretreated (24 h) with 2 ng/ml TGF-β. n = 48-77 cells/condition. Data are presented as Tukey box plots, representing the upper quartile + 1.5 interquartile range (IQR; upper end of whisker), upper quartile (top of box), mean (band in box), lower quartile (bottom of box), and lower quartile - 1.5 IQR (lower end of whisker). *P < 0.05, ****P < 0.0001 by an unpaired Mann-Whitney test.

3.5 DOCK4 shares high homology with DOCK3 but does not interact with NEDD9

Noteworthy, a previous study reported that DOCK3 forms a complex with NEDD9, and that this complex regulates Rac1 activation to drive mesenchymal movement in melanoma cells (Sanz-Moreno et al. 2008). We found however that DOCK4, while effectively activating Rac1, does not interact with NEDD9 (Fig. 15A,B), further supporting the notion that these two proteins have distinct modes of regulation and likely serve distinctive cell-type specific functions.

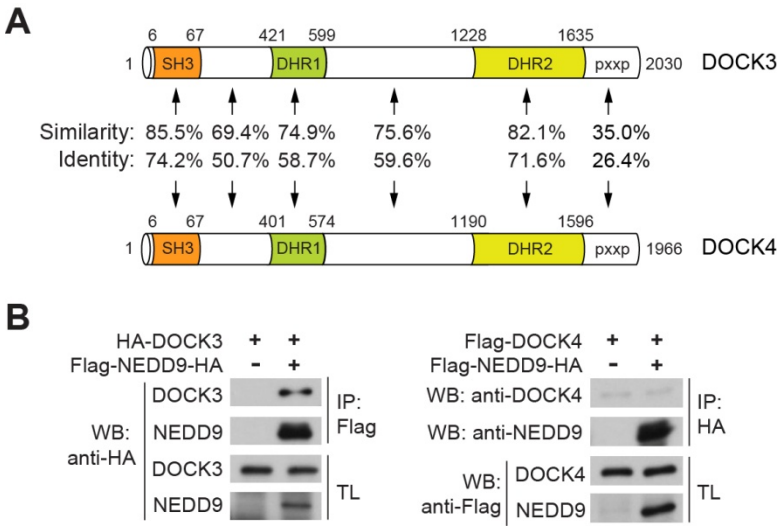

Figure 15. DOCK4 shares high protein sequence homology with DOCK3, but does not interact with NEDD9 (*A*) Protein sequence similarity and identity between DOCK3 and DOCK4 proteins. (*B*) Lysates from HEK-293T cells transiently expressing the indicated constructs were immunoprecipitated (IP) with an antibody to Flag (left panel) or HA (right panel) and analyzed by Western blot (WB) analysis with antibodies against HA (left panel) or DOCK4, NEDD9, or Flag (right panel). TL, total lysate.

Chaper 4

Concluding remarks and future perspectives

4.1 Significance

In this study, we identified DOCK4 as a novel and direct target gene of the TGF-β/Smad signaling pathway. We further showed DOCK4 is essential for TGF-β driven metastasis of lung ADC and DOCK4 does so by specifically regulating cell extravasation via the activation of Rac1. Further, we found DOCK4 expression is correlated with TGF-β signaling activity in primary lung ADC and anti-correlated with patient recurrence-free survival. In addition to the evidence of experimental validation, there are a number of interesting implications found in this study. First, Rac1 plays a central role in actin-based cytoskeletal remodeling and controls cell protrusive activity and motility. During solid tumor metastasis, disseminated tumor cells have to move across the tissue to intravasate and within the vasculature to extravasate. TGF-β primed EMT facilitates these cellular features but the mechanistic link between TGF-β and Rac1 has been long missing. Our study showed a direct mechanism as how TGF-β controls Rac1 activity (via DOCK4) during EMT. More interestingly, this molecular paradigm appears to be crucial in lung but not breast cancer cells, suggesting the tissue specific regulation of DOCK4. In fact, only one DOCK180 family homolog was found in *C. elegans* (CED5) and in *D. melanogaster* (Myoblast city), respectively. However, in mammals, there are 11 DOCK180 family members (DOCK1 to 11) and in the context of TGF-β, DOCK4 plays an essential, non-redundant role found in lung cells. Additionally, although highly conserved, DOCK4 differs from DOCK3 in the undefined protein domains of its C-terminus and does not reside in the integrin-β3/NEDD9 complex. These observations shed a new light in the evolutionarily divergent roles of DOCK180 proteins and the complexity of mammalian signaling transduction and tissue specificity. Other specific points

regarding the details of regulation of DOCK4 and future directions/perspectives are discussed below.

4.2 The context-dependent regulation of DOCK4

Metastasis from lung adenocarcinoma (ADC), the most common subtype of lung cancer, typically occurs rapidly to multiple organs (Hoffman et al. 2000; Provencio et al. 2011). A key factor reported to drive lung ADC metastasis is the cytokine TGF-β (Lund et al. 1991; Hasegawa et al. 2001; Nemunaitis et al. 2009; Toonkel et al. 2010; Vazquez et al. 2013); however the genes and mechanisms that mediate the pro-metastatic effects of TGF-β remain largely unknown. Here we identify DOCK4 as a novel, key target of the TGF-β/Smad signaling pathway that promotes lung ADC metastasis by enhancing the competence of lung ADC cells to extravasate into distant organs. We further present evidence that DOCK4 does so at least in part by stimulating the protrusive activity and motility of mesenchymal lung ADC cells via activation of the Rac1 pathway.

DOCK4 is a member of the DOCK180 family of GEFs, of which in total 11 mammalian family members have been identified (Cote and Vuori 2002; Meller et al. 2005). While all members, with the exception of DOCK2, are expressed in lung ADC cells, we found that the TGF-β/Smad pathway selectively upregulates the expression of DOCK4, but not other DOCK-family members, supporting the notion that these proteins exhibit different modes of regulation (Laurin and Cote 2014). Notably, high DOCK4 expression levels correlate with activated TGF-β signaling and poor prognosis in human lung ADC. Interestingly, the regulation of DOCK4 by TGF-β appears to be tumor-type dependent. Indeed, no induction of DOCK4 by TGF-β was observed in breast cancer and melanoma cells. Also, binding of p-Smad3 to Smad binding elements within DOCK4's first intron was detected in lung ADC cells, but not in breast cancer

cells. Finally, no correlation between DOCK4 expression and disease relapse was found in estrogen receptor (ER)-negative breast cancer patients. One possible explanation for these context-dependent observations is that the epigenetic status of lung ADC cells is different from that of breast cancer and melanoma cells. Also, a cell-type specific Smad co-factor(s) critical to the regulation of DOCK4 expression in response to TGF-β signaling could be present in lung ADC cells, but not in breast cancer and melanoma cells. Future studies will be required to decipher the precise cell/tumor-type specific regulation of DOCK4 expression in response to TGF-β/Smad signaling.

4.3 The role of DOCK4 in the metastatic cascade

Using xenograft mouse models, we showed that DOCK4 plays a critical role in mediating TGF-β-driven lung ADC metastasis. An intriguing finding from our studies is that while DOCK4 expression is already induced by TGF-β in primary human lung ADC (as indicated by our IHC stainings of human lung ADC TMAs), DOCK4 appears to exert its effects at a later time by enhancing the extravasation capabilities of lung ADC cells. Indeed, we found that DOCK4 knockdown in TGF-β primed lung ADC cells that enter circulation impedes their ability to form metastatic foci at the distant organ sites, and that DOCK4 exerts this effect without affecting the growth properties or survival of the metastasizing cells. Moreover, we observed that the fraction of TGF-β primed lung ADC cells that extravasated out of the liver vasculature was markedly reduced in the DOCK4 knockdown group, compared to the control group. These findings reinforce the view that TGF-β signals produced within the primary tumor microenvironment can influence later stages of metastasis, and that a protein induced by TGF-β in tumor cells at the primary site may also act at a later step of the metastatic process (Padua et al. 2008; Calon et al.

2012; Yuan et al. 2014). Our data, however, do not exclude that DOCK4 could also play a role in mediating the enhancing effect of TGF-β on lung ADC cell intravasation. Future studies will be required to determine whether DOCK4 not only acts on the late, but also early, steps of lung ADC metastasis.

Padua *et al* previously showed that TGF-β in the breast tumor microenvironment primes tumor cells for metastasis to the lungs by driving the expression of angiopoietin-like 4 (ANGPTL4) (Padua et al. 2008). Interestingly, while ANGPTL4 facilitates tumor cell extravasation in a non-cell-autonomous manner by disrupting endothelial junctions at the metastatic sites, our data indicate that DOCK4 does so in a cell-autonomous manner by promoting the protrusive activity and motility of lung ADC cells. Noteworthy, a recent *in vivo* study showed that tumor cells with high protrusive activity can migrate and navigate through narrow vessel lumen openings and vessel branch points; a process that could allow them to find optimal sites for extravasation (Stoletov et al. 2010). Combined, these findings indicate that tumor cell extravasation is a highly dynamic and coordinated process, involving contributions of both cell-extrinsic and intrinsic factors. With regard to DOCK4, it should be noted that an additional factor(s) acting in parallel to DOCK4 in mediating TGF-β-promoted lung ADC cell extravasation and metastasis likely come(s) into play. While depletion of DOCK4 impaired TGF-β-induced cell protrusive activity/motility and extravasation of lung ADC cells, ectopic expression of DOCK4 (at levels similar to those induced by TGF-β) did not enhance the protrusive activity/motility of these cells nor their metastatic potential. In light of our findings that DOCK4 is dispensable for TGF-β-induced EMT in lung ADC cells, we envision that TGF-β drives the activation of genes required for EMT and in parallel the induction of DOCK4 expression, and that the former is a prerequisite for DOCK4's subsequent enhancing effects on

the protrusive activity/motility and extravasation potential of lung ADC cells. In the two lung ADC cell lines we tested, Snail and Slug appeared to be the only upregulated canonical EMT-inducing transcription factors. However, ectopic expression of Snail and/or Slug in A549 cells was not able to induce EMT (data not shown), suggesting the requirement of EMT induction differs in cell types. To dissect the EMT and DOCK4 pathways, further studies are needed to identify the molecules regulating lung ADC cell EMT in parallel of the Smad/DOCK4 axis.

4.4 DOCK4 links TGF-β to Rac1- a new perspective on potential therapeutic approaches

Our data unveil that TGF-β/Smad-induced DOCK4 promotes the protrusive activity and motility of mesenchymal lung ADC cells via the activation of the Rac1 GTPase. While Rac1 has been implicated before in cell motility and protrusion formation in different tumor cell types (Reymond et al. 2013), so far Rac1 has been mainly linked to TGF-β via non-canonical pathway(s) (Zhang 2009) and the mechanism underlying TGF-β receptor-coupled Rac activation is unknown. Here we showed that blockage of the canonical Smad pathway greatly reduces TGF-β-induced activation of Rac1, similarly as depletion of DOCK4. Thus our findings unveil a previously unrecognized link between TGF-β/Smad and Rac1 signaling, and identify DOCK4 as a key player bridging the two pathways. Our results indicate that once primed by TGF-β tumor cells may stably maintain the invasive phenotype for a period of time even entering a new microenvironment devoid of TGF-β. Interestingly, while multiple Rac-GEFs, including other DOCK-family members, are expressed in lung ADC cells, our findings imply that they do not compensate for DOCK4 function, as depletion of DOCK4 alone was sufficient to blunt TGF-β-promoted lung ADC cell protrusive activity and motility. Thus, DOCK4 likely provides specificity in signaling to Rac1 activation downstream of TGF-β to control lung ADC cell protrusion and motility. This is of particular interest, given that global and long-term inhibition

of Rac1 is well known to exert anti-proliferative effects not only on tumor cells but also normal cells. Therefore the development of small molecules that specifically inhibit DOCK4's Rac-GEF activity or abrogate the interaction between DOCK4 and Rac1 could present a valid therapeutic strategy in the treatment of lung ADC metastasis. For example, screening for small molecular inhibitors of the DHR2 domain of DOCK4 could be carried out to specifically disrupt the GEF activity of DOCK4. It would particularly and potentially benefit the patients who have yet developed metastasis and underwent surgical resection of primary tumors as a preventive strategy against tumor cell extravasation and metastasis.

4.5 Transcriptional regulation of DOCK4

Although the ChIP-seq and ChIP-qPCR data clearly demonstrated that upon TGF-β stimulation p-Smad3 directly binds to the two putative enhancer SBEs in the first intron of DOCK4, it remained unclear how this process in precisely regulated. The Smad complex can certainly bind to a minimal four nucleotide AGAC SBE sequence. However, this interaction is of low affinity and can be stabilized with addition Smad binding partners through the MH1 domain of Smad proteins. The MH1 domain mediates highly versatile protein-protein interactions between Smads and other transcriptional co-activators or co-repressors (Mullen et al. 2012). Thus, the Smad co-factors are crucial for the transcriptional output downstream of TGF-β and the Smad co-factor(s) that co-localize at the DOCK4 SBEs remain to be identified in lung ADC cells. In addition, DOCK4 has a relatively large first intron (~200kb) and the two putative enhancer SBEs are localized distant from TSS at +45kb and +125kb, respectively. Thus, it may be potentially interesting to take advance of the thriving new technologies such as CRISPR/Cas and 4C-seq to investigate the long range chromatin interaction between the two SBEs and DOCK4 proximal promoter region to fully decipher the transcriptional mechanisms of TGF-β-

driven expression of DOCK4. Alternatively, given that TGF-β driven upregulation of DOCK4 was only seen in lung but not breast cancer cells, it would also be interesting to take advantage of either candidate gene and proteomic approaches to compare the difference between Smad3/4 binding partners in lung and breast cells, which could potentially lead to discovery of cell type or lineage specific transcription factors that modulate the Smad3/4 occupancy at the DOCK4 intronic SBEs.

Chapter 5

Material and methods

5.1 Cell lines

Lung adenocarcinoma (ADC) cell lines used in this study were: A549, HCC4006, H441, PC9, H1975, and H1793 (gift from R. Sordella, CSHL). All lung ADC cell lines were grown in RPMI 1640, GlutaMax^{TM} (Invitrogen) supplemented with 5% fetal bovine serum (FBS) and 1% penicillin/streptomycin (Invitrogen). The following breast cancer cell lines were used: MDA-MB-134VI, MDA-MB-231, MDA-MB-435S, SK-BR-3, and Hs578T (gift from M. Wigler, CSHL). MDA-MB-134-VI, MDA-MB-231, and MDA-MB-435S were grown in Leibovitz's L-15 medium (ATCC) supplemented with 10% FBS and 1% penicillin/streptomycin. SK-BR-3 cells were grown in McCoy's 5a (ATCC), and Hs578T in DMEM (Invitrogen), each of them supplemented with 10% FBS (Invitrogen) 0.01 mg/ml insulin (Invitrogen), and 1% penicillin/streptomycin. The known somatic mutations for common oncogenic/tumor suppressive pathways in the above cell lines are listed in Table 3. For TGF-β treatment, a 1:1 mixture of human recombinant TGF-β1 and TGF-β2 (R&D systems) was added to the culture media at a final concentration of 2 ng/ml.

Lung adenocarcinoma cell lines	Genes	Protein sequence alteration
A549	KRAS CDKN2A	G12S Deletion
H441	KRAS TP53	G12V R158L
HCC4006	EGFR	L747-E749 deletion, A750P
PC9	EGFR TP53	L746-A750 deletion R248Q
H1975	EGFR TP53 CDKN2A PIK3CA	L858R, T790M R273H E69* G118D
H1793	CDKN2A TP53	Deletion R209*; R273H

Breast cancer cell lines	Genes	Protein sequence alteration
MDA-MB-134-VI	TP53	E285K
MDA-MB-231	KRAS TP53 BRAF CDKN2A NF2	G13D R280K G464V Deletion E231*
MDA-MB-435S	TP53	G266E
SK-BR-3	TP53	R175H
Hs578T	HRAS TP53 CDKN2A PIK3R1	G12D V157F Deletion N453_T454insN

Table 3: List of cell lines used in this study.

5.2 Plasmids, shRNAs, and viral transduction

MLP-PGK-GFP-puro or MLP-PGK-DsRed-neo vectors (gift from S.W. Lowe) using *XhoI* and *EcoRI* restriction sites. The shRNAs targeting human Smad4 (shSmad4#1 and shSmad4#2) were purchased from Addgene (#15724 and #37046). The Rac1 shRNAs (shRac1#1 and shRac1#2) were gifts from Y. Zhang (Cincinnati Children's Hospital). A shRNA targeting Renilla luciferase was used as a negative control and was previously described (Zuber et al. 2011). Cell lines stably expressing cDNAs or shRNAs were generated by retroviral or lentiviral transduction. For the production of viral particles, retroviral vectors were co-transfected with pCL-Ampho and pVSV-G into Phoenix- Ampho cells, and lentiviral vectors were co-transfected with pCMVΔR8.91, pcREV and pVSV-G into HEK-293T cells. The virus-containing medium was collected 48 h after transfection and the target cells were spin infected. Infected cells were selected in media containing 2 μg/ml puromycin for 2 days or 100 μg/ml neomycin for 4 days.

5.3 Immunoblotting

For Western blot analyses, cells were lysed in RIPA buffer containing protease (Roche) and phosphatase (Sigma) inhibitors. Proteins were separated by SDS-PAGE and blotted onto PVDF membranes (Bio-Rad). Membranes were probed with the following primary anti-human antibodies: Gapdh (1:5,000, Novartis), DOCK4 (1:2,000, gift from V. Yajnik), phospho-Smad3 (Ser423/425) (1:1,000, Cell Signaling), Smad3 (1:1,000, Cell Signaling), Smad4 (1:1,000, Cell Signaling), E-cadherin (1:10,000, BD Biosciences), phospho-β-catenin (Ser33/Ser37/Thr41) (1:1,000, Cell Signaling), β-catenin (1:5,000, Cell Signaling), Rap1 (1:1000, BD Biosciences), Cdc42 (1:1000, Cytoskeleton), and Rac1 (1:1,000, Cytoskeleton), followed by HRP-conjugated secondary antibodies (1:5,000, Bio-Rad).

5.4 Co-immunoprecipatation

pCL-neo-HA-DOCK3 (gift from C. Marshall, ICR) or pcDNA3-Flag-DOCK4 was co-transfected with LZRS-hNEDD9-IRES-GFP, which contains an N-terminal Flag and a C-terminal HA epitope tag (#21962, purchased from Addgene), into HEK-293T cells. 24 h post-transfection, cells were lysed in lysis buffer containing 200 mM Tris pH 8.0, 150 mM NaCl, 0.5% NP-40, with protease and phosphatase inhibitors. 1 mg total protein was incubated with 1 µg anti-Flag M2 (Sigma) or 1 µg anti-HA (Covance) antibody for 1 h at 4°C, followed by incubation with protein A/G Sepharose beads (GE Healthcare) for 1 h at 4°C. Beads were washed 3 times with lysis buffer. Immunoprecipitates were then resolved by SDS-PAGE and immunoblotted with antibodies against HA (1:5,000, Covance), Flag M2 (1:2,000, Sigma), NEDD9 (1:2,000, Abcam), or DOCK4 (1:2,000, gift from V. Yajnik).

5.5 Rac1/Cdc/42/Rap1-GTP pulldown assays

To assay Rac1, Cdc42, and Rap1 activity, GST-PBD and GST-RalGDS-RBD pull-down assays were performed as described before (Govek et al. 2004; Boettner et al. 2000), using A549 cells expressing indicated constructs.

5.6 Immunofluorescence and confocal image acquisition

For immunostaining of cultured lung ADC cells, the cells were grown on coverslips and fixed for 15 min at RT in 4% paraformaldehyde (PFA) in TBS. For immunostaining of liver tissue sections, livers were dissected from mice 20 h after intracardiac injection of lung ADC cells. Animals were deeply anesthetized with isoflurane and perfused transcardially with 4% PFA. Livers were post-fixed in 4% PFA overnight, embedded in 3% agarose, and cut into 50-µm thick sections using a vibratome (Leica VT1000S). Cultured lung ADC cells and liver

sections were permeabilized and blocked with 0.1% Triton X-100 and 1% BSA in TBS for 5 min (for cells) or 1 h (for sections) at RT, followed by incubation with primary antibodies diluted in TBS with 1% BSA at 4°C overnight. We used the following primary antibodies: mouse anti-human E-cadherin (1:400, BD Biosciences); rat anti-mouse CD31 (clone MEC13.3, 1:200, BD Pharmingen); and chicken anti-GFP (1:1,000, Aves Labs). The secondary antibodies used were: goat anti-chicken Alexa Fluor 488, goat anti-rat Alexa Flour 594, and goat anti-mouse Alexa Fluor 647 (1:1,000, Molecular Probes). Fluorescence images of lung ADC cells and liver sections were acquired using Zeiss LSM 510 and 780 confocal microscopes with a 63x or 40x oil-immersion objective. In the case of liver sections, images were taken at z-sections (15-30 sections) of 1-μm intervals, and images were reconstructed using Volocity image processing software.

5.7 Tissue microarray and immunohistochemistry

Human lung adenocarcinoma tissue microarrays (TMA LC706 and TMA LC10013) were purchased from US Biomax Inc. Immunohistochemical staining with anti-p-Smad3 (#9520, 1:25, Cell Signaling), anti-DOCK4 (#ab56743, 1:400, Abcam), or anti-β-actin (#ab6276, 1:500, Abcam) antibody was performed using DAB chromogen kit (Vector labs) according to the manufacturer's protocol. For quantification of p-Smad3 IHC staining, the Aperio IHC Nuclear Image Analysis algorithm was used. This algorithm detects the staining intensity of nuclei stained with a specific chromagen (brown). Nuclei are also identified by size and shape. The values for the analysis are set by the pathologist (J.E.W). A minimal level for positive staining is set and the nuclear staining is classified as 0, 1+, 2+ and 3+ based on nuclear staining intensity. A nucleus is classified 0 when it has no nuclear staining. 1+ nuclei have weak but positive staining. A nucleus is classified 2+ when it has moderate nuclear staining and 3+ when there is

intense staining. Based on the percentages of 0, 1+, 2+ and 3+ nuclei, the final H-score is determined (1 X % weak staining) + (2 X % moderate staining) + (3 X % strong staining). For analysis, 10-20 areas of each core containing only tumor cells were traced. Only these areas were scanned for the analysis. For quantification of DOCK4 IHC staining, the Aperio Cytoplasm Analysis algorithm was used. This algorithm measures the intensity of the stain (brown) in the cytoplasm. The default values for nuclear staining (hematoxylin) and cytoplasmic staining (brown) were used. Cytoplasmic segmentation was included in the analysis and the distance from the nucleus to the cytoplasm was set using a visual check of this value by the pathologist (J.E.W). This parameter defines how far from the nucleus the cytoplasm can be reported as cytoplasm that surrounds the nucleus. For analysis, 10-20 areas containing only tumor or adjacent normal epithelial cells of each core were traced. Only these areas were scanned for the analysis. The intensity of the staining was recorded as 0, 1+, 2+ and 3+ based on cytoplasmic staining intensity. A cytoplasm is classified 0 when it has no cytoplasmic staining. 1+ cytoplasm has weak but positive staining. Cytoplasmic staining is classified 2+ when it has moderate cytoplasmic staining and 3+ when there is intense staining. The thresholds were set by the pathologist based on the most intense staining in the slide and the lowest intensity deemed positive. Based on the percentages of 0, 1+, 2+ and 3+ cytoplasm, the final H-score is determined (1 X % weak staining) + (2 X % moderate staining) + (3X % strong staining). The H-scores for p-Smad3 and DOCK4 were further converted to a 0-6 scale. The samples with raw H- scores $< 5^{th}$ percentile and $> 95^{th}$ percentile were designated scores of 0 and 6, respectively. In the samples between 5^{th} and 95^{th} percentile, the maximal raw H-score was set as 100%. The samples with raw H-scores between 80-100%, 60-80%, 40-60%, 20-40%, and 0-20% were then designated scores of 5, 4, 3, 2, and 1, respectively. Based on the

0-6 scale, the samples were classified into p-Smad3 or DOCK4 high (score 3-6) or low (score 0-2) groups for contingency analysis. The *P* value was calculated by Fisher's exact test.

5.8 RNA isolation and quantitative PCR (qPCR)

Total RNA was isolated from cells using TRIzol reagent (Invitrogen), purified by LiCl precipitation and then reverse transcribed using the TaqMan reverse transcription kit (Invitrogen). The resulting cDNA was amplified by qPCR using Power SYBR Green PCR Master Mix (Invitrogen). qPCR and data collection were performed on an Applied Biosystems (ABI) 7900HT Fast Real-Time PCR system. The primers used in this study are listed in Table 4.

DOCK4	
Forward:	5'-GCATCTCTCGCTGGTTTGAAG-3'
Reverse:	5'-CAGGCACATAGTCAGGGGATT-3'
E-cadherin	
Forward:	5'-CGAGAGCTACACGTTCACGG-3'
Reverse:	5'-GGGTGTCGAGGGAAAAATAGG-3'
Vimentin	
Forward:	5'-GACGCCATCAACACCGAGTT-3'
Reverse:	5'-CTTTGTCGTTGGTTAGCTGGT-3'
Twist	
Forward:	5'-GTCCGCAGTCTTACGAGGAG-3'
Reverse:	5'-GCTTGAGGGTCTGAATCTTGCT-3'
Snail	
Forward:	5'-TCGGAAGCCTAACTACAGCGA-3'
Reverse:	5'-AGATGAGCATTGGCAGCGAG-3'
Slug	
Forward:	5'-CGAACTGGACACACATACAGTG-3'
Reverse:	5'-CTGAGGATCTCTGGTTGTGGT-3'
Zeb1	
Forward:	5'-GATGATGAATGCGAGTCAGATGC-3'
Reverse:	5'-ACAGCAGTGTCTTGTTGTTGT-3'
Zeb2	
Forward:	5'-CAAGAGGCGCAAACAAGCC-3'
Reverse:	5'-GGTTGGCAATACCGTCATCC-3'
Gapdh	
Forward:	5'-ACAACTTTGGTATCGTGGAAGG-3'
Reverse:	5'-GCCATCACGCCACAGTTTC-3'
DOCK4-SBE1	
Forward:	5'-GGAAGTGTAGCTTTCTATTAGG-3'
Reverse:	5'-GGCCAGACACATAGTAATGG-3'
DOCK4-SBE2	
Forward:	5'-CTGTGTGTCTCCTCCAAACC -3'
Reverse:	5'-CGAAACAGGAAGAAGGAACC-3'

Table 4: List of qPCR primers used in this study

5.9 Chromatin immunoprecipitation (ChIP)

A549 or MDA-MB-231 cells (~3 × 10_6) in a 10 cm dish were left untreated or treated with 2 ng/ml TGF-β for 5 h. Chromatin immunoprecipitation (ChIP) assays were performed using a ChIP-IT Express Enzymatic Magnetic Chromatin Immunoprecipitation Kit & Enzymatic Shearing Kit according to the manufacturer's instructions (#53009, Active Motif). The antibodies used in the ChIP assays were normal rabbit IgG (sc-2027, Santa Cruz) and p-Smad3 (Ser423/425) (#9520, Cell Signaling). 4 μg of each antibody and ~50 μg of total chromatin were used in each ChIP reaction. The primers used in ChIP-qPCR analysis are listed in Table 4. Enrichment was quantified as the IP/Input ratio using SYBR green reagent on the ABI7900HT.

5.10 Cell proliferation and anoikis assays

For 2D cell proliferation assays, 5,000 cells were suspended in regular culture medium and plated into 6-well plates. Cell numbers were counted using a Nexcelom Cellometer Auto T4 Cell Counter at the indicated time points. For anoikis assays, 10,000 cells were suspended in culture media and plated into 96-well plates pre-coated with poly-HEMA (poly 2-hydroxyethyl methacrylate) to prevent cell attachment. Cells were incubated at 37°C for the indicated time points, and cell viability was measured using a MTT assay (Invitrogen) following the manufacturer’s protocol.

5.11 Time lapse microscopy and single cell movement analysis

5,000-10,000 cells were plated into 6 or 12-well plates in culture media supplemented with or without TGF-β for 24 h prior to imaging. Cells were imaged every 6 min over a total period of 4.5 h using a Zeiss Observer automated imaging system. The movement of single cells was tracked and quantified using Zeiss Axiovision 6.0 software. The coordinate graphs were

plotted using Origin 8 software; each cell was plotted with time 0 at the origin of the grid. Single cells were randomly chosen for each condition.

5.12 Matrigel invasion assay

Invasion assays were performed using 24-well PET Transwell inserts (Costar, 8.0 μm pore size) coated with 20 μl of Matrigel (BD Biosciences) at 37°C for 20 h. 10,000 cells resuspended in culture medium containing 0.2% FBS were plated in the upper chamber of the Transwell insert. Culture medium containing 10% FBS was used as a chemoattractant in the lower chamber. 20 h after plating, cells in the upper chamber were removed with a cotton swab. Cells that had migrated through the filters were fixed in 10% methanol and stained with 0.5% crystal violet. The filters were photographed and cell density was calculated using an Odyssey CLx infrared imaging system.

5.13 Animal studies

All animals were maintained in a specific pathogen-free facility, and all studies were conducted under protocols approved by the CSHL Institutional Animal Care and Use Committee. 5-7 week old NOD/SCID/IL2γ (NSG) mice (National Cancer Institute) were used for all xenografting studies. Lung ADC cells were engineered to stably express firefly luciferase. Prior to all injections, lung ADC cells were harvested at a concentration of $5x10^6$ cells/ml in PBS containing 1 mM EDTA pH 8.0, and in all cases animals were anesthetized with 2% isoflurane before injection. For intracardiac injections, a small skin incision was made on the left side of the chest, and $5x10^5$ cells were injected into the left ventricle using a 26-gauge needle attached to a 1-ml insulin syringe (BD biosciences). For intrathoracic injections, a small skin incision was made to the left chest wall. A 30- gauge needle attached to a 0.5-ml insulin syringe was

inserted directly through the intercostal space into the lung to a depth of about 3-5 mm, and $1x10^5$ cells were injected into the lung parenchyma. In both cases the wounds were closed using surgical wound clips (Fine Science Tool) and wound clips were removed 7 d later. For intrahepatic injections, a 2 cm incision was made in the upper abdomen through the peritoneum. The liver was carefully exposed and $1x10^5$ cells resuspended in 20 µl of a 1:1 mixture of Matrigel and PBS were injected intrahepatically using a 30-gauge needle attached to a 0.3-ml insulin syringe (Terumo). The wound was closed with a 6-0 silk suture (CP Medical Inc.). Animals were monitored until recovery after surgery for at least 30 min. For bioluminescence imaging, animals were injected intraperitoneally with 100 µl of 30 mg/ml D-luciferin (GoldBio) in PBS and anesthetized with 2% isofluorane using a XGI-8 gas anesthesia system (Xenogen). Bioluminescence images were acquired 10-15 minutes after injection using the IVIS-200 Imaging System (Xenogen). For metastasis assays, bioluminescent signals were quantified at the indicated time points and normalized to day 0 from live animals. The mean value of the control group was normalized to 100 for data presentation. Metastases were confirmed by necropsy and histology using *ex vivo* bioluminescent imaging and Hematoxylin & Eosin (H&E) staining. Growth rate in lung and liver was measured as function of photon flux normalized to day 0 in live animals.

5.14 Bioinformatics

For gene expression data analysis (Fig. 2A), a dataset comprising gene expression profiles of TGF-β treated A549 cells was downloaded from Gene Expression Omnibus (GEO) with accession number GSE17708 (Ranganathan et al. 2007). Probes against all 83 Rho family GEFs were selected. Gene expression levels were normalized to the 0 h time point and then converted to log2 scale. The data were processed and a heat map was generated using Expander

6.0. For DOCK3/4 homology analysis, protein sequences for DOCK3 (accession number: AAP80572.1) and DOCK4 (accession number: AAO73565.1) were downloaded from NCBI. The boundaries of the identified functional domains were defined based on UniProtKB/Swiss-Prot simulations. Sequence homology was analyzed using VectorNTI 11.0 software.

5.15 ChIP-seq data analysis

Short Read Archives (SRAs) for ChIP-Seq analysis were downloaded from GEO database (GSM1246720 and GSM1246721). After converting SRA to FASTQ format, the sequence reads were mapped to the reference genome assembly NCBI36/hg18 using Bowtie following these criteria: -m1, -v2. To identify ChIP-Seq peaks, we used the MACS version 1.4.0beta (Model based Analysis of ChIP-Seq) peak finding algorithm. A p value threshold of enrichment of 1e-5, and a false discovery rate (FDR) of less than 5% was adjusted to define a list of confident peaks.

5.16 Lung ADC and ER-negative breast cancer clinical data analysis

For lung ADC clinical association analysis, a publicly available dataset (GSE41271) was downloaded from NCBI Gene Expression Omnibus (GEO). It contains microarray-based gene expression data of 182 human lung adenocarcinomas and clinical follow-up information, including recurrence status, recurrence-free survival intervals, tumor stage, age, gender, smoking history, and race. The probe IDs representing DOCK3 and DOCK4 are ILMN_1723440 and ILMN_1801044, respectively. For ER-negative breast cancer clinical association analysis, a cohort of 269 primary ER-negative breast cancer samples with corresponding recurrence-free survival data was extracted from five publicly available datasets (GSE2034, GSE2603, GSE5327, GSE4922, and GSE7390) downloaded from NCBI GEO. The probe ID

representing DOCK4 is 205003_at. To assess the predictive power of DOCK3 and DOCK4 for tumor recurrence, we computed the normalized expression values by Z-scores across all patients (log2 transformation and RMA normalization in package "affy") and established a univariate Cox proportional hazards model (R-package "survival"). We obtained a smooth estimate of the relationship between DOCK3 or DOCK4 expression levels and recurrence hazard ratio, and defined the optimal cut-offs that yield the highest hazard ratio between low and high expression groups (customized cut-off finder; Budczies et al. 2012). We generated Kaplan-Meier survival curves for patients with low and high DOCK3 or DOCK4 expression using GraphPad Prism. Statistical significance was assessed by introducing the expression value as a continuous covariate in the Cox model. We applied the univariate Cox proportional hazards model on all available clinical annotations, DOCK3 and DOCK4 expression groups for the lung ADC cohort. Statistically significant covariates from the univariate model were included in a multivariate Cox model (R-package "survival"), which are DOCK4 expression and tumor stage. *P* values were calculated by log-rank test. For the correlation between high DOCK4 expression, frequency of recurrence, and tumor stage, the contingency analyses were performed using GraphPad Prism. *P* values were calculated by Fisher's exact test.

5.17 Statistical analysis

The two-tailed Student's *t*-test or Mann-Whitney test was used to compare continuous variables between two groups with parametric or non-parametric distributions, respectively. A Fisher's exact test was used to compare dichotomous variables. A Log- rank test was used for survival analysis. All data are presented as the mean ± SD, unless otherwise noted. Statistical significance was defined as $P < 0.05$ for all tests.

References

Akunuru S, Palumbo J, Zhai QJ, Zheng Y. 2011. Rac1 targeting suppresses human non-small cell lung adenocarcinoma cancer stem cell activity. *PLoS ONE* **6**: e16951.

American Cancer Society. Cancer Treatment and Survivorship Facts & Figures 2014-2015. Atlanta: American Cancer Society; 2014.

Boettner B, Govek EE, Cross J, Van Aelst L. 2000. The junctional multidomain protein AF-6 is a binding partner of the Rap1A GTPase and associates with the actin cytoskeletal regulator profilin. *Proc Natl Acad Sci* **97**: 9064-9069.

Budczies J, Klauschen F, Sinn BV, Gyrffy B, Schmitt WD, Darb-Esfahani S, Denkert C. 2012. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PloS ONE* **7**: 1-7.

Calon A, Espinet E, Palomo-Ponce S, Tauriello DV, Iglesias M, Cespedes MV, Sevillano M, Nadal C, Jung P, Zhang XH et al. 2012. Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. *Cancer Cell* **22**: 571-584.

Cote JF, Vuori K. 2002. Identification of an evolutionarily conserved superfamily of DOCK180-related proteins with guanine nucleotide exchange activity. *J Cell Sci* **115**: 4901-4913.

Derynck R, Akhurst RJ, Balmain A. 2001. TGF-beta signaling in tumor suppression and cancer progression. *Nat Genet* **29**: 117-129.

Elliott RL, Blobe GC. 2005. Role of transforming growth factor Beta in human cancer. *J Clin Oncol* **23**: 2078-2093.

Feng H, Hu B, Jarzynka MJ, Li Y, Keezer S, Johns TG, Tang CK, Hamilton RL, Vuori K, Nishikawa R et al. 2012. Phosphorylation of dedicator of cytokinesis 1 (Dock180) at

tyrosine residue Y722 by Src family kinases mediates EGFRvIII-driven glioblastoma tumorigenesis. *Proc Natl Acad Sci* **109**: 3018-3023.

Gadea G, Sanz-Moreno V, Self A, Godi A, Marshall CJ. 2008. DOCK10-mediated Cdc42 activation is necessary for amoeboid invasion of melanoma cells. *Curr Biol* **18**: 1456-1465.

Gaunt SJ, George M, Paul YL. 2013. Direct activation of a mouse Hoxd11 axial expression enhancer by Gdf11/Smad signalling. *Dev Biol* **383**: 52-60.

Giampieri S, Manning C, Hooper S, Jones L, Hill CS, Sahai E. 2009. Localized and reversible TGFbeta signalling switches breast cancer cells from cohesive to single cell motility. *Nat Cell Biol* **11**: 1287-1296.

Gibbons DL, Lin W, Creighton CJ, Rizvi ZH, Gregory PA, Goodall GJ, Thilaganathan N, Du L, Zhang Y, Pertsemlidis A et al. 2009. Contextual extracellular cues promote tumor cell EMT and metastasis by regulating miR-200 family expression. *Genes Dev* **23**: 2140-2151.

Govek EE, Newey SE, Akerman CJ, Cross JR, Van der Veken L, Van Aelst L. 2004. The X-linked mental retardation protein oligophrenin-1 is required for dendritic spine morphogenesis. *Nat Neurosci* **7**: 364-372.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017-1018.

Gregory PA, Bracken CP, Smith E, Bert AG, Wright JA, Roslan S, Morris M, Wyatt L, Farshid G, Lim YY et al. 2011. An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Mol Biol Cell* **22**: 1686-1698.

Hasegawa Y, Takanashi S, Kanehira Y, Tsushima T, Imai T, Okumura K. 2001. Transforming growth factor-beta1 level correlates with angiogenesis, tumor progression, and prognosis in patients with nonsmall cell lung carcinoma. *Cancer* **91**: 964-971.

Hiramoto K, Negishi M, Katoh H. 2006. Dock4 is regulated by RhoG and promotes Rac-dependent cell migration. *Exp Cell Res* **312**: 4205-4216.

Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MD, Niu B, McLellan MD, Uzunangelov V et al. 2014. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**: 929-944.

Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D, Li L, Place C et al. 2012. A landscape of driver mutations in melanoma. *Cell* **150**: 252-263.

Hoffman PC, Mauer AM, Vokes EE. 2000. Lung cancer. *Lancet* **355**: 479-485.

Isogaya K, Koinuma D, Tsutsumi S, Saito RA, Miyazawa K, Aburatani H, Miyazono K. 2014. A Smad3 and TTF-1/NKX2-1 complex regulates Smad4-independent gene expression. *Cell Res* **24**: 994-1008.

Jakowlew SB. 2010. Transforming Growth Factor-beta in Lung Cancer, Carcinogenesis, and Metastasis. *Springer*, The Tumor Microenvironment, Cancer Drug Discovery and Development: 633-671.

Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, Yamamoto S, Tatsuno K, Katoh H, Watanabe Y et al. 2014. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet* **46**: 583-587.

Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, Guise TA, Massague J. 2003. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**: 537-549.

Kawada K, Upadhyay G, Ferandon S, Janarthanan S, Hall M, Vilardaga JP, Yajnik V. 2009. Cell migration is regulated by platelet-derived growth factor receptor endocytosis. *Mol Cell Biol* **29**: 4508-4518.

Kennedy BA, Deatherage DE, Gu F, Tang B, Chan MW, Nephew KP, Huang TH, Jin VX. 2011. ChIP-seq defined genome-wide map of TGFbeta/SMAD4 targets: implications with clinical outcome of ovarian cancer. *PLoS ONE* **6**: e22606.

Kissil JL, Walmsley MJ, Hanlon L, Haigis KM, Bender Kim CF, Sweet-Cordero A, Eckman MS, Tuveson DA, Capobianco AJ, Tybulewicz VL et al. 2007. Requirement for Rac1 in a K-ras induced lung cancer in the mouse. *Cancer Res* **67**: 8089-8094.

Kobayashi M, Harada K, Negishi M, Katoh H. 2014. Dock4 forms a complex with SH3YL1 and regulates cancer cell migration. *Cell Signal* **26**: 1082-1088.

Labelle M, Begum S, Hynes RO. 2011. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. *Cancer Cell* **20**: 576-590.

Laurin M, Cote JF. 2014. Insights into the biological functions of Dock family guanine nucleotide exchange factors. *Genes Dev* **28**: 533-547.

Laurin M, Huber J, Pelletier A, Houalla T, Park M, Fukui Y, Haibe-Kains B, Muller WJ, Cote JF. 2013. Rac-specific guanine nucleotide exchange factor DOCK1 is a critical regulator of HER2-mediated breast cancer metastasis. *Proc Natl Acad Sci* **110**: 7434-7439.

Lund LR, Romer J, Ronne E, Ellis V, Blasi F, Dano K. 1991. Urokinase-receptor biosynthesis, mRNA level and gene transcription are increased by transforming growth factor beta 1 in human A549 lung carcinoma cells. *EMBO J* **10**: 3399-3407.

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG et al. 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**: 2129-2139.

Ma C, Rong Y, Radiloff DR, Datto MB, Centeno B, Bao S, Cheng AW, Lin F, Jiang S, Yeatman TJ et al. 2008. Extracellular matrix protein betaig-h3/TGFBI promotes metastasis of colon cancer by enhancing cell extravasation. *Genes Dev* **22**: 308-321.

Massague J, Seoane J, Wotton D. 2005. Smad transcription factors. *Genes Dev* **19**: 2783-2810.

Massague J. 2008. TGFbeta in Cancer. *Cell* **134**: 215-230.

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

Meller N, Merlot S, Guda C. 2005. CZH proteins: a new family of Rho-GEFs. *J Cell Sci* **118**: 4937-4946.

Michl P, Ramjaun AR, Pardo OE, Warne PH, Wagner M, Poulsom R, D'Arrigo C, Ryder K, Menke A, Gress T et al. 2005. CUTL1 is a target of TGF(beta) signaling that enhances cancer cell motility and invasiveness. *Cancer Cell* **7**: 521-532.

Morikawa M, Koinuma D, Tsutsumi S, Vasilaki E, Kanki Y, Heldin CH, Aburatani H, Miyazono K. 2011. ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. *Nucleic Acids Res* **39**: 8712-8727.

Mullen AC, Orlando DA, Newman JJ, Loven J, Kumar RM, Bilodeau S, Reddy J, Guenther MG, DeKoter RP, Young RA. 2011. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**: 565-576.

Nemunaitis J, Nemunaitis M, Senzer N, Snitz P, Bedell C, Kumar P, Pappen B, Maples PB, Shawler D, Fakhrai H. 2009. Phase II trial of Belagenpumatucel-L, a TGF-beta2 antisense gene modified allogeneic tumor vaccine in advanced non small cell lung cancer (NSCLC) patients. *Cancer Gene Ther* **16**: 620-624.

Nguyen DX, Chiang AC, Zhang XH, Kim JY, Kris MG, Ladanyi M, Gerald WL, Massague J. 2009. WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. *Cell* **138**: 51-62.

Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. 2013. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**: 548-551.

Padua D, Massague J. 2009. Roles of TGFbeta in metastasis. *Cell Res* **19**: 89-102.

Padua D, Zhang XH, Wang Q, Nadal C, Gerald WL, Gomis RR, Massague J. 2008. TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133**: 66-77.

Provencio M, Isla D, Sanchez A, Cantos B. 2011. Inoperable stage III non-small cell lung cancer: Current treatment and role of vinorelbine. *J Thorac Dis* **3**: 197-204.

Ranganathan P, Agrawal A, Bhushan R, Chavalmane AK, Kalathur RK, Takahashi T, Kondaiah P. 2007. Expression profiling of genes regulated by TGF-beta: differential regulation in normal and tumour cells. *BMC genomics* **8**: 98.

Reymond N, d'Agua BB, Ridley AJ. 2013. Crossing the endothelial barrier during metastasis. *Nat Rev Cancer* **13**: 858-870.

Sahai E, Marshall CJ. 2002. RHO-GTPases and cancer. *Nat Rev Cancer* **2**: 133-142.

Sakata-Yanagimoto M, Enami T, Yoshida K, Shiraishi Y, Ishii R, Miyake Y, Muto H, Tsuyama N, Sato-Otsubo A, Okuno Y et al. 2014. Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat Genet* **46**: 171-175.

Sanz-Moreno V, Gadea G, Ahn J, Paterson H, Marra P, Pinner S, Sahai E, Marshall CJ. 2008. Rac activation and inactivation control plasticity of tumor cell movement. *Cell* **135**: 510-523.

Schlenner SM, Weigmann B, Ruan Q, Chen Y, von Boehmer H. 2012. Smad3 binding to the foxp3 enhancer is dispensable for the development of regulatory T cells with the exception of the gut. *J Exp Med* **209**: 1529-1535.

Scott J, Kuhn P, Anderson AR. 2012. Unifying metastasis--integrating intravasation, circulation and end-organ colonization. *Nat Rev Cancer* **12**: 445-446.

Sethi N, Dai X, Winter CG, Kang Y. 2011. Tumor-derived JAGGED1 promotes osteolytic bone metastasis of breast cancer by engaging notch signaling in bone cells. *Cancer Cell* **19**: 192-205.

Shibue T, Brooks MW, Weinberg RA. 2013. An integrin-linked machinery of cytoskeletal regulation that enables experimental tumor initiation and metastatic colonization. *Cancer Cell* **24**: 481-498.

Siebert N, Xu W, Grambow E, Zechner D, Vollmar B. 2011. Erythropoietin improves skin wound healing and activates the TGF-beta signaling pathway. *Lab Invest* **91**: 1753-1765.

Stoletov K, Kato H, Zardouzian E, Kelber J, Yang J, Shattil S, Klemke R. 2010. Visualizing extravasation dynamics of metastatic tumor cells. *J Cell Sci* **123**: 2332-2341.

Strell C, Entschladen F. 2008. Extravasation of leukocytes in comparison to tumor cells. *Cell Commun Signal* **6**: 10.

Toonkel RL, Borczuk AC, Powell CA. 2010. Tgf-beta signaling pathway in lung adenocarcinoma invasion. *J Thorac Oncol* **5**: 153-157.

Travis WD, Brambilla E, Riely GJ. 2013. New pathologic classification of lung cancer: relevance for clinical practice and clinical trials. *J Clin Oncol* **31**: 992-1001.

Tsai JH, Yang J. 2013. Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev* **27**: 2192-2206.

Valastyan S, Weinberg RA. 2011. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147**: 275-292.

Vazquez PF, Carlini MJ, Daroqui MC, Colombo L, Dalurzo ML, Smith DE, Grasselli J, Pallotta MG, Ehrlich M, Bal de Kier Joffe ED et al. 2013. TGF-beta specifically enhances the metastatic attributes of murine lung adenocarcinoma: implications for human non-small cell lung cancer. *Clin Exp Metastasis* **30**: 993-1007.

Yajnik V, Paulding C, Sordella R, McClatchey AI, Saito M, Wahrer DC, Reynolds P, Bell DW, Lake R, van den Heuvel S et al. 2003. DOCK4, a GTPase activator, is disrupted during tumorigenesis. *Cell* **112**: 673-684.

Yamaguchi H, Lorenz M, Kempiak S, Sarmiento C, Coniglio S, Symons M, Segall J, Eddy R, Miki H, Takenawa T et al. 2005. Molecular mechanisms of invadopodium formation: the role of the N-WASP-Arp2/3 complex pathway and cofilin. *J Cell Biol* **168**: 441-452.

Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM et al. 2013. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**: 580-584.

Yuan JH, Yang F, Wang F, Ma JZ, Guo YJ, Tao QF, Liu F, Pan W, Wang TT, Zhou CC et al. 2014. A long noncoding RNA activated by TGF-beta promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell* **25**: 666-681.

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

Zuber J, Shi J, Wang E, Rappaport AR, Herrmann H, Sison EA, Magoon D, Qi J, Blatt K, Wunderlich M et al. 2011. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**: 524-528.