Fucosyltransferase 4 shapes oncogenic glycoproteome to drive metastasis of lung adenocarcinoma

Hsuan-Hsuan Lu[1], Shu-Yung Lin[2,3], Rueyhung Roc Weng[1], Yi-Hsiu Juan[1], Yen-Wei Chen[4], Hsin-Han Hou[1], Zheng-Ci Hung[1], Giovanni Audrey Oswita[4], Yi-Jhen Huang[4], Jin-Yuan Shih[1], Chong-Jen Yu[1,5*], Hsing-Chen Tsai[1,4*]

[1]Department of Internal Medicine, National Taiwan University Hospital, Taipei, 10002, Taiwan.
[2]National Taiwan University Hospital Jin-Shan Branch, New Taipei City, 20844, Taiwan.
[3]Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, 10051, Taiwan.
[4]Graduate Institute of Toxicology, College of Medicine, National Taiwan University, Taipei, 10051, Taiwan.
[5]Department of Internal Medicine, College of Medicine, National Taiwan University, Taipei, 10051, Taiwan.
*Contributed equally

Correspondence: Hsing-Chen Tsai, MD, PhD., htsai@ntu.edu.tw, Graduate Institute of Toxicology, College of Medicine, National Taiwan University, No.1 Jen Ai Road Section 1, Rm1424, Taipei, Taiwan 10051, Phone: +886-2-23123456#88797 or Chong-Jen Yu, MD, PhD. jefferycjyu@ntu.edu.tw; Department of Internal Medicine, National Taiwan University Hospital, No. 7, Zhongshan S Rd, Zhongzheng District, Taipei City, 10002, Taiwan. Phone: +886-23123456#62905

Conflict of interest statement: The authors have declared that no conflict of interest exists.

Abstract

Aberrant fucosylation plays critical roles in lung cancer progression. Identification of the key fucosyltransferase as a therapeutic target may refine lung cancer management. Here, we identified a terminal α1,3-fucosyltransferase, FUT4, as the key prognostic predictor for lung adenocarcinoma through a transcriptomic screen in lung cancer cohorts. Overexpression of FUT4 promoted lung cancer invasion, migration, epithelial-to-mesenchymal transition and cell adhesion *in vitro*, which can be reversed by genetic depletion of the enzyme. Notably, knockdown of FUT4 markedly curtailed lung colonization and distant metastases of lung adenocarcinoma cells in mouse xenograft models. Moreover, immunoprecipitation-mass spectrometry with anti-Lewis x, a major fucosylated glycan generated by FUT4, revealed increased fucosylation on cascade proteins of multiple oncogenic signalings including epidermal growth factor (EGF) and transforming growth factor-β (TGF-β) pathways with concomitant transcriptional activation. The malignant phenotype provoked by FUT4-mediated fucosylproteomic networks can be pharmacologically diminished as treating FUT4-high expressing cells with EGFR inhibitors showed reduced metastatic capacity *in vivo*. Collectively, FUT4 represents a promising therapeutic target in lung cancer metastasis. Our data highlight the potential for integration of glycomics into precision medicine-based therapeutics.

Introduction

Non-small cell lung cancer is a disease with genetic and gene regulatory complexities. Molecular analysis of individual tumors is essential for clinical choice of therapeutic strategies in the era of precision medicine. Despite the growing dimensions of lung cancer-associated molecular abnormalities, current practices for lung cancer therapy rely primarily on pathological, mutational, and immunological characterizations. Abundant evidence indicates that aberrant glycosylation plays critical roles in fundamental steps of tumor development and progression, including cell-cell/cell-matrix interactions, metastasis, cancer metabolism as well as immune surveillance[1-3]. In particular, patterns of fucosylation, which adds a fucose (6-deoxy-L-galactose) residue to surface oligosaccharides/proteins catalyzed by the fucosyltransferase (FUT) gene family, are frequently altered in various types of cancer[4,5]. Depending on the site of oligosaccharide chain to which the fucose is added, two types of fucosylation— core and terminal fucosylation—were defined[4]. So far, there are 13 FUTs identified in human, each of which catalyzes the synthesis of fucosylated glycans with designated glycosidic linkages and targets different substrate proteins/glycans in a tissue-specific manner.

In search for the key fucosyltransferases underlying aberrant fucosylation patterns specific for lung cancer progression, we and others have previously discovered that fucosyltransferase 8 (FUT8), the only enzyme responsible for the core fucosylation with α1,6-linkage, mediated the malignant phenotypes of non-small cell lung cancer[6,7]. On the other hand, clinical significance of fucosyltransferases involved in terminal fucosylation (α1,3- or α1,4- linkage) in lung cancer remains controversial. Past studies

have demonstrated aberrant expression of terminal fucosylated epitopes such as Lewis antigens in non-small cell lung cancer tissues[8-10]. Nevertheless, the prognostic values of various Lewis antigens appeared to differ. Expressions of sialyl Lewis x (sLe^x) and Lewis x (Le^x) were associated with shortened survival times[10]. In contrast, Lewis y (Le^y) seemed to predict better survival or confer limited clinical significance[8,9]. Furthermore, while some groups found that terminal fucosyltransferases such as FUT4 or FUT7 may promote lung cancer progression[11-13], another group demonstrated that FUT4- or FUT6-mediated fucosylation of epidermal growth factor receptor (EGFR) could suppress EGFR dimerization and activation[14]. As multiple fucosyltransferases (FUT3-7, 9-11) are involved in terminal fucosylation and Lewis antigen synthesis, a systemic approach on all α1,3- or α1,4- fucosyltransferases with large scale clinical correlation in different subtypes of non-small cell lung cancers, followed by in-depth mechanistical studies would be needed to identify the key enzyme that accounts for malignant phenotype of lung cancer and decipher the complex molecular networks involved in cancer progression.

In the present study, our group examined 81 surgically-resected tumor tissues from patients with non-small cell lung cancer for molecular and prognostic correlations on all terminal α1,3- or α1,4- fucosyltransferases, and independently validated our results with the TCGA lung cancer cohorts for cross-ethnic generalization. We identified fucosyltransferase 4 (FUT4) as the main indicator for poor clinical outcome in lung adenocarcinoma patients. We conducted mechanistic and functional studies by altering FUT4 expression levels in lung cancer cells with tetracycline-controlled transcriptional regulatory system. The enhancement of metastatic capacity of lung cancer cells by

FUT4 was further corroborated by multiple *in vivo* xenograft mouse experiments. Moreover, through integrated glycoproteomic and transcriptomic analyses, we found that FUT4-mediated cancer progression was tightly linked to activated oncogenic regulatory networks and may be reversed by pharmacological inhibitions. Our study pinpointed a clinically-significant fucosylation enzyme for the malignant phenotype of lung adenocarcinoma, and delineated its complex glycoproteomic networks for developing preventive therapeutics for cancer progression in the future.

Results

Higher levels of FUT4 is associated with poor prognosis in lung adenocarcinoma

Terminal fucosyltransferases, such as FUT 3-7 and 9-11, were known to catalyze fucosylation with α1,3- and α1,4- linkage. In order to decipher clinicopathological impact of these fucosyltransferases, we examined expression of FUT 3-7 and 9-11 in the genome-wide RNA-seq data of 81 primary lung cancer tissues (adenocarcinoma, n=44; squamous cell carcinoma, n=37) from National Taiwan University Hospital (NTUH) (GEO: GSE120622). We observed high expression of FUT3, 4, 6, 10, 11 in lung cancer tissues, while little expressions of FUT5, 7 and 9 were detected (Fig. 1A). Interestingly, similar patterns were observed in another cohort of 517 lung adenocarcinoma and 502 squamous cell patients from the The Cancer Genome Atlas (TCGA) despite the ethnic differences and distinct mutational profiles between the two cohorts (Fig. 1A)[15-18]. To determine the clinical significance of individual fucosyltransferases in terms of overall survival, we divided the patient cohorts into high-expressing and low-expressing groups

by the median value of individual fucosyltransferases and performed a survival analysis on the two groups using Cox proportional hazard model (Table 1). The analysis in lung adenocarcinoma patients demonstrated a negative prognostic effect of FUT4 (*, hazard ratio 2.172 ± 0.217; P = 0.0003, Cox regression model), and protective effects of FUT1 (†, hazard ratio = 0.497 ± 0.220, P = 0.0015) and FUT10 (†, hazard ratio = .649 ± 0.217, P = 0.0047) on overall survivals (Fig. 1B). Kaplan-Meier curves also showed significantly worse overall survival of the $FUT4^{High}$ group (P = 0.00068; Fig. 1B). Notably, we did not observe a significant association between FUT4 expression and overall survival in lung squamous cell carcinoma (P = 0.23; Fig. 1B and Supplemental Table 1), indicating a distinct glycobiological mechanism may be involved in the progression of this cancer subtype. Moreover, to investigate potential functional coordination between different FUTs, we performed correlation analysis based on mRNA levels of FUTs in 517 lung adenocarcinoma tissues. We found little or modest concordance between the expressions of FUT4 and other α1,3- or α1,4-fucosyltransferases (Pearson correlation coefficient between -0.1 to 0.22, Fig. 1C). This suggests the unique role of FUT4 plays in promoting the aggressive phenotypes of lung adenocarcinoma. In contrast, while not having prognostic values, expressions of FUT3 and FUT6 showed a strong correlation (Pearson correlation coefficient, R = 0.73; Fig. 3C), in consistent with a previous study showing a functional coordination between these two enzymes in colorectal cancers[19].

FUT4 promotes aggressive phenotypes of human lung cancer cells

To decipher the functional mechanisms of FUT4 that account for poor prognosis in lung cancer patients, we conducted a series of *in vitro* and *in vivo* mechanistic studies using human lung cancer cell lines. First, we examined expression levels of FUT4 in a panel of 17 human non-small cell lung cancer lines. Higher expressions of FUT4 were observed in 13 of the 17 cell lines as compared with immortalized BEAS-2B lung epithelial cells (Supplementary Fig. S1A). Subsequently, we stably transfected two human lung adenocarcinoma cell lines, A549 (Supplemental Fig. S1B) and CL1-0 cells (Supplemental Fig. S1C), with FUT4. For A549 cells, we selected two clones with different levels of FUT4 expression—A549_FUT4^{high} and A549_FUT4^{med} – to investigate the dose-responsive relationship of FUT4 protein. We found that A549_FUT4^{high}, A549_FUT4^{med}, CL1-0_FUT4 cells showed markedly increased invasion abilities through Matrigel®-coated Boyden chamber (Fig. 2A and Supplementary Fig. S1D). Moreover, single cell migration assays in a live-cell imaging and tracking system revealed increased migration abilities of A549_FUT4^{high} and A549_FUT4^{med} cells. (~8-fold upregulation in A549_FUT4^{high} and ~7-fold upregulation in A549_FUT4^{med}) (Fig. 2B). Similarly, there was a trend towards increased cell migration velocity for CL1-0 cells (Supplementary Fig. S1E). Interestingly, A549_FUT4^{med} cells showed similar invasion and migration abilities with A549_FUT4^{high} cells, which indicates that moderate elevation of FUT4 protein level was sufficient to cause significant functional changes.

Subsequently, to determine if FUT4 is required for aggressiveness in lung cancer, we knocked down FUT4 using an shRNA approach in CL1-5 lung adenocarcinoma cells, which has relatively high expression of FUT4 (Supplementary

Fig. S1A). Migration and invasion abilities of cancer cells appeared to be reduced in CL1-5_shFUT4#751 cells which contained the lowest expression level of FUT4 among all shRNA clones (#751, #753 and #792) compared to the vector control (Fig. 2 C-D). The other two knockdown clones (#753 and #792) also showed a consistent statistical trend towards significance. These data suggested that FUT4 is a major player in promoting invasion and metastases of lung adenocarcinoma cells.

FUT4 triggers epithelial-mesenchymal transition of human lung cancer cells.

Cancer cells of epithelial origin often undergo epithelial-to-mesenchymal transition (EMT) to achieve a more malignant phenotype during the process of cancer progression and metastases[20]. It is increasingly acknowledged that EMT comprises a series of intermediate states between the epithelial and mesenchymal phenotypes, rather than a binary cellular decision[21]. The transitions between these intermediate states often take place in response to various signaling cues and involve complex regulatory networks. Thus, the process tends to be dynamic, reversible, and reflective of cellular plasticity[21]. While the roles of fucosyltransferases such as FUT4 or FUT8 in EMT have been suggested in several cancer types[11,22,23], the dynamics between fucosyltransferases and EMT remains incompletely understood. To track the temporal course of FUT4-mediated EMT transition and its reversibility in lung adenocarcinoma cells, we established tetracycline-controlled transcriptional repressive (tet-off) system in A549 lung cancer cells. We found that addition of doxycycline at 0.5 ng/mL induced a significant decrease in FUT4 protein level starting at 24 hours and reaching maximum reduction at 7 days (data not shown). Using immunofluorescent microscopy, we

confirmed that FUT4 overexpression lead to morphological changes characteristic of EMT, which was manifested by a prominent reorganization of the cytoskeleton toward a more mesenchymal phenotype in A549_FUT4^{med}, A549_FUT4^{high} (Fig. 2E) and CL1-0_FUT4 cells (Supplementary Fig. S2B). Down-regulations of CDH1 (E-cadherin) and CLDN1 (claudin-1), two essential molecules for cell-cell adhesion, were evident in A549 cells in a FUT4 level-dependent manner. The changes were accompanied by minimal increases of mesenchymal proteins, including VIM (vimentin) and CDH2 (N-cadherin) (Supplementary Fig. S2A). Concordantly, western blot analyses of A549_FUT4^{med} cells and A549_FUT4^{high} cells revealed profound downregulation of epithelial proteins, CDH1 and CLDN1, along with significant increases of EMT-inducing transcription factors—SNAI1 (a.k.a SNAIL) and SNAI2 (a.k.a SLUG), which are known to bind to the CDH1 promoter and repress expression of CDH1 gene [24,25](Supplementary Fig. S2C). This observation indicates FUT4 induces partial EMT as A549_FUT4 ^{med} and A549_FUT4^{high} cells appeared to be in the intermediate states between the epithelial and mesenchymal phenotypes. Similar findings were observed for CL1-0_FUT4, which moved farther along the EMT spectrum with a marked increase of VIM, an intermediate filament protein expressed in the mesenchymal state (Supplementary Fig. S2D).

Notably, this FUT4-mediated EMT phenomenon was reversible as we turned off FUT4 expression in A549_FUT4^{high} cells using the Tet-off system. The phenotypic reversal took approximately up to 7 days to be evident. We observed reappearance of CDH1 proteins and reorganization of cytoskeletons back to the epithelial phenotype in A549_FUT4^{high} cells following the addition of 0.5, 1, and 2 ng/ml doxycycline to shut down FUT4 expression in a dose-dependent manner (Fig. 2F). This plasticity and

reversibility of glycosylation-mediated EMT process may of critical importance when exploring therapeutic opportunities in the metastatic journey of cancer cells. Since dynamic epithelial and mesenchymal inter-transitions are required for successful metastatic initiation (EMT) and distant colonization (MET), modulation of FUT4-mediated fucosylation at the right timing represents a potential strategy in preventing lung cancer metastasis.

FUT4 enhances adhesion and colonization ability of lung cancer cells.

Fucosyltransferases generate or modify glycans on the cell surface, which regulate cell-cell interaction, cell adhesion, cell trafficking, and signal transduction[4,26,27]. To confirm whether FUT4 high-expressing cells may lead to a stronger cell adhesion—an early step in the metastatic cascade, we performed *in vitro* adhesion assays with several common adhesion molecules within human tissues and blood vessels, including collagen IV, E-selectin (on endothelial cells), L-selectin (on lymphocytes) and P-selectin (on platelets). Lung cancer cells bound to the adhesion molecules pre-coated on 96-well plates were stained and counted under an immunofluorescent microscope. Consistent with recent reports[12,31,32], our data showed that higher levels of FUT4 protein facilitated binding of lung cancer cells to E-selectin and L-selectin, but not P-selectin (on platelets) or collagen IV (Fig. 3A). In light of this, we further performed an *in vivo* assay to investigate the effects of FUT4 on organotropic extravasation ability to model the metastatic process in which the cells travel and lodge themselves at distant sites. We pre-labeled lung cancer cells with CYTO-ID® red tracer dye and injected the cells into the right ventricle of C57BL6 mice. Thirty minutes later, mice were sacrificed and

perfused with normal saline intracardially to remove cells not adhered to the pulmonary vasculature (Fig. 3B). A marked retention of A549_FUT4^{high} and CL1-0_FUT4 cells in the lungs were observed under a microscope as opposed to their respective vector controls (Fig. 3C). On the contrary, knock-down of FUT4 significantly decreased the extravasation and retention ability of CL1-5 cells in the lungs (Fig. 3D).

Genetic depletion of FUT4 reduces metastasis of lung cancer *in vivo*

To substantiate the role of FUT4 on lung cancer metastasis *in vivo*, we investigated whether aberrant FUT4 expression led to spontaneous distant metastasis using several mouse xenograft models. We subcutaneously injected A549_FUT4^{med}, A549_FUT4^{high} cells and A549_vector into NOD/SCID immunocompromised mice. When mice were sacrificed 56 days post-injection, we observed that mice carrying subcutaneous tumors of A549_FUT4^{high} cells and A549_FUT4^{med} cells had higher frequencies of metastatic lung nodules. Moreover, higher FUT4 levels were correlated with higher metastatic potentials (Fig. 4A-B). The FUT4-mediated metastasis was also evaluated in a second mouse model of metastasis where we performed tail vein injection of lung cancer cells into nude mice. We observed a rapid lung homing phenomenon enhanced by FUT4. As short as one day post tail vein injection, more A549_FUT4^{med} and A549_FUT4^{high} tumor cells were trapped in the lung areas shown by the IVIS® Spectrum imaging system compared to A549_vector cells (Fig. 4C). When we sacrificed mice 28 days post-injections, we found that mice in the FUT4^{high} group had a much greater number of metastatic lung nodules compared to those in the FUT4^{med} and vector groups ($p < 0.0005$) (Fig. 4D). On the other hand, when we silenced FUT4 using

shRNA in CL1-5 lung cancer cells which have spontaneous *in vivo* metastatic ability[33], we found that FUT4 silencing markedly abolished metastasis of lung cancer cells (Fig. 4E).

FUT4-mediated fucosylation modulates multiple oncogenic signaling pathways

Fucosylation has been shown to play an important regulatory role in protein structure and functions, and the effects of FUTs on signaling pathways can be either stimulatory[6,23] or inhibitory[14]. To investigate molecular and functional gene networks altered by FUT4-induced aberrant fucosylation, we performed transcriptomic analysis on A549_FUT4^{high} and CL1-0_FUT4 cells using genome-wide RNA-seq technologies (GEO: GSE120622). We observed that there were 193 and 291 genes significantly upregulated, and 183 and 330 genes significantly downregulated by more than 1.4-fold ($p < 0.05$) between FUT4-overexpressing cells and parental cells in A549 and CL1-0, respectively. For genes commonly regulated in both cell lines, Gene Set Enrichment Analysis[34] revealed alterations of multiple pathways related to oncogenesis, cancer metastasis and immune signaling (Supplementary Fig. S3). In particular, EGF and TGFβ signaling pathways were among the top enriched gene sets (Fig. 5A). To investigate whether this FUT4-mediated signaling modulatory effect was also shown in primary tumor tissues, we examined transcriptomic data in the TCGA cohort of 517 lung adenocarcinomas. We found that high FUT4-expressing tumors displayed significantly enhanced activities in oncogenic signaling networks such as EGF and TGFβ pathways as opposed to low FUT4-expressing tumors (Fig. 5B and C). Notably, FUT4-mediated activation of oncogenic signaling was observed in both EGFR wild-type and mutant

tumors. Higher levels of FUT4 expression correlated with EGFR pathway gene activations in the EGFR wild-type tumors (Fig. 5C). These data suggest that FUT4 can be an onco-signaling modulator in lung cancer via non-mutational mechanisms.

To decipher how FUT4-mediated fucosylation modulates oncogenic pathways, we went on to probe acceptor proteins fucosylated by FUT4 using immunoprecipitation (IP) with anti-Le^x antibody followed by tandem mass spectrometry (LC-MS/MS) as we showed that Le^x antigen is the major glycan antigen synthesized by FUT4 in our system (Supplementary Fig. S4). We identified 3,348 and 1,936 proteins bearing Le^x antigen in A549_FUT4^{high} and CL1-0_FUT4 cells, and 1,187 and 1,012 proteins in their respective vector controls (Supplementary Fig. S5A and B)(MassIVE: MSV000083028). We found that FUT4 synthesized Le^x on a wide variety of intracellular proteins involved in broad molecular and cellular processes (Supplementary Fig. S5B). Remarkably, many FUT4 acceptor proteins interacted physically and formed an intricate network containing key mediator proteins downstream of major cancer metastatic pathways (Supplementary Fig. S5C). These relationships were further illustrated in our integrated analysis of transcriptome and glycoproteome (Fig. 5D), which showed that FUT4 altered glycosylation of signaling cascade proteins at multiple levels downstream of EGF and TGFβ signaling pathways (Fig. 5E). For EGF signaling, mediators downstream of EGF receptor signaling such as AKT, mTOR, ERK, and SRC appeared to be cytoplasmic targets of FUT4. FUT4 also had profound effects on TGFBR signaling through fucosylating Smad2/ Smad3/ Smad4, which are the three principal mediators downstream of TGFβ signaling. Increased activity of AKT or TGFβ/Smad signaling has been shown to be crucial in driving the EMT process[21,35]. Taken together, these data

demonstrated that FUT4 alters multiple fucosyl-proteomic networks to promote the malignant phenotype of lung cancer.

FUT4-enhanced oncogenic signaling can be pharmacologically diminished

Having demonstrated that FUT4 promotes malignant phenotype through activating several major oncogenic networks including EGF and TGFBR signaling, we then attempted to elucidate which pathway plays a leading role in FUT4-mediated effects, and whether pharmacologically inhibiting the pathway may provide a therapeutic opportunity against metastasis of FUT4-high-expressing tumors. We performed *in vivo* lung cancer metastasis assay via tail vein injection of A549_vector or A549_FUT4^{high} cells into nude mice, which subsequently received intraperitoneal treatment of either 0.25 mg/kg afatinib (an EGFR inhibitor) or 15 mg/kg LY2157299 (a TGF-β inhibitor) three times a week (Fig. 6A). We found that, while both drugs significantly diminished lung metastasis in A549_FUT4^{high} cells, afatinib demonstrated differential inhibitory effects between A549_FUT4^{high} and A549_vector cells (Fig. 6B). The data indicate EGFR signaling is a main effector of FUT4 upregulation, thereby directing FUT4-mediated lung cancer progression. Indeed, the data were further supported by the *in vitro* time-series experiment on individual pathway proteins. we observed increased expression of Le^{x} antigen on EGF receptor through Immunoprecipitation-western blot analysis in A549_FUT4^{high} cells (Fig. 6C). Moreover, we found that EGF receptor and its downstream proteins in FUT4 high-expressing cells exhibited enhanced sensitivity to exogenous EGF at 10ng/mL as demonstrated by an increase of phospho-EGFR (pEGFR) and phospho-ERK (pERK) at 0.5 to 2 hours post ligand-mediated receptor

activation, followed by a later increase of TGFBR pathway genes including phosphor-Smad2 (pSmad2) at 2 hours to 6 hours in A549_FUT4^{high} compared to the vector control (Fig. 6D). This suggest that activation of EGFR signaling precedes TGFBR signaling in the FUT4-high setting.

As anticipated, this signaling augmentation effect of FUT4 could be abolished by afatinib. Upon exogenous EGF stimulation, FUT4-transfected A549 cells had higher expression of SNAI1 protein indicating mesenchymal transition, which was completely reversed by afatinib (Fig. 6E). On the other hand, although increased expression of Le^{x} antigen on TGFβ receptor was noted in A549_FUT4^{high} cells (Supplementary Fig. S6A), and addition of 5ng/mL TGFβ triggered activation of TGFβ pathway (Supplementary Fig. S6B), the exogenous TGFβ led to upregulation of SNAI1 in both A549_vector and A549_FUT4^{high} cells. This induced mesenchymal phenotype appeared to be reversed with LY2157299 treatment in both cells (Supplementary Fig. S6C). Therefore, TGFBR signaling may be inherently activated in A549 lung cancer cells regardless of FUT4 status. Taken together, the data further support our observation that FUT4-mediated lung cancer progression primarily via the EGF signaling pathway.

Discussion

Aberrant expression of fucosylated epitopes such as Lewis antigens have been widely reported in various types of cancer[4]. These Lewis antigens are mainly synthesized by α1-3/ α1-4 fucosyltransferases in a tissue-specific manner. In the current study, we holistically evaluated all α1-3/ α1-4 fucosyltransferases and identified FUT4

as a prognostic biomarker in Asian and Western cohorts of lung cancer patients through large scale transcriptomic analysis. Asian lung cancer is known to be a molecularly and etiologically distinct identity composed of a much higher percentage of never smokers and EGFR mutations[17,18] as compared to the western counterpart[15,16]. Yet, comparative transcriptomic analyses of Eastern and Western cohorts revealed striking similarities in the expression patterns of α1-3 terminal fucosyltransferases in non-small cell lung cancers despite mutational status and ethnicities. Among all α1-3 FUTs, FUT3, 4, 6, 10 and 11 appeared to be highly expressed in both lung adenocarcinoma and squamous cell carcinoma. Moreover, despite the general conception that FUTs may facilitate cell proliferation or invasion in cancers[19,36-38], our genome-wide analysis in large patient cohorts revealed that FUTs may have either negative or protective prognostic values in lung cancer. While FUT4 is correlated with poor survival, FUT1 and FUT10 appeared to confer survival benefit in lung cancer. It is interesting that FUT1 has been shown to promote proliferation and spreading potential of ovarian carcinoma[39,40]. This suggests that the same fucosyltransferase may act on different sets of acceptor proteins in different tumor types. While we delineated the role of FUT4 in lung adenocarcinoma, detailed mechanistic and clinical studies may be required to elucidate the precise roles of other FUTs in individual tumor types.

Several groups have demonstrated that FUT4 may promote EMT in cell line models of breast[22], lung cancer[11], and etc. In our study, we extended these findings by investigating the fluidity and reversibility of FUT4-modulated mesenchymal transition via a tet-off transcriptional repression system. We demonstrated that FUT4-mediated fucosylation induced a partial EMT state, which can be rapidly reversed by turning off

FUT4. This indicates a therapeutic potential for metastasis prevention by targeting FUT4. Nevertheless, it is increasingly acknowledged that EMT can be dynamic and heterogeneous within a tumor. Studies have shown that mesenchymal tumor cells disseminated in the circulation need to be partially reversed to the epithelial state for outgrowth at metastatic sites[41,42]. This raises an important question regarding the optimal timing and magnitude of FUT4 depletion in preventing lung cancer metastasis. Although targeting FUT4 may reverse EMT and prevent invasion and migration of cancer cells at primary sites, if FUT4 depletion happens to occur only when cells are reaching a distant site, it may provide an unwanted boost for metastatic colonization. Thus, caution should be exercised when developing targeted therapy against FUT4 or other FUTs for clinical applications.

In addition, abundant evidence has shown that fucosylation of surface receptors, including EGFR, may either sensitize receptor activation or suppress receptor dimerization in a context- and enzyme-dependent manner[6,14]. Nevertheless, attempts to map intracellular targets of FUTs for a deeper mechanistic interrogation have been fairly limited due to transient interactions between FUTs and their acceptor proteins as they fold and pass through the Golgi apparatus. Most previous molecular and mechanistic studies took a candidate-pathway approach and focused on one or two pathways involved in fucosylation-mediated malignant phenotype in various cancer types[11,19,43,44], such as EGFR/MAPK, NF-κB/p65, and TGF-beta. A global view on the complex network crosstalks has been lacking. Here, we not only performed genome-wide RNA-seq to profile FUT4 transcriptome, but also took a glimpse into candidate FUT4 acceptor proteins through global proteomics approach by tandem mass spectrometry

(LC-MS/MS) based on the presence of Le^x, a predominant product synthesized by FUT4 in lung cancer cells. We discovered an intricate network of Le^x-decorated intercellular proteins that exert activating effects and coordinate the augmentation of signaling pathways. Consistent with prior studies, we identified EGFR and TGFBR pathways upregulated in FUT4 high-expressing cells (Figure 5)[11,19]. We also discovered other pathways such as WNT, NOTCH or immune signaling that may participate in FUT4-mediated tumor progression (Supplementary Fig. S4). The tangled crosstalks between intracellular proteins downstream of EGFR and TGFBR, along with involvement of other pathways indicates that FUT4-mediated effects are far more complex than previously identified, and require further investigation and clarification with state-of-the-art technologies. In addition, our finding brings to light the additional biological role of Le^x or other carbohydrate epitopes on cytosolic proteins aside from being merely a surface antigen, which may open up research on whether and how altered fucosylated glycans modulate the activities of intracellular proteins. Further studies such as glycosylation/fucosylation site-directed mutagenesis are warranted to decode how Le^x can alter protein functions and stability.

Amongst all research efforts on clinical implications of cancer-specific fucosylation, our *in vitro* and *in vivo* studies demonstrated a prime role of FUT4 during the process of lung cancer metastasis, which highlights the potential of FUT4 as an attractive therapeutic target. Indeed, knockdown of FUT4 in our study significantly diminished lung metastases in multiple mouse xenograft models without apparent toxicities (Fig. 4E). Currently, glycosylation-related therapeutics is under active development, ranging from modulating stability of therapeutic proteins via glycosylation

to glycoprotein-based biodrugs[45]. Multiple attempts have been taken to identify inhibitors of sialyl- and fucosyl-transferases via high-throughput screening[46-48]. Non-selective metabolic inhibitors such as synthetically-derived fucose-related molecules have been shown to inhibit global fucosylation thereby diminishing neutrophil extravasation or xenograft tumor growth in immunocompromised mice[49]. As our data showed that individual fucosyltransferases may have opposite effects on patient prognosis, we believe that development of FUT4-specific inhibitors may be necessary for preventing lung cancer metastases, and at the same time minimize potential off-target effects due to non-selective ablation of global fucosylation.

In addition, our combined transcriptomic and glycoproteomic approach delineated major signaling pathways triggered by FUT4. This indicates that pharmacological targeting FUT4-mediated pathways may provide similar therapeutic benefits as FUT4 depletion therapy. The question would be which signaling pathway to target. Our data showed that, while EGFR and TGFβ pathways were the two major pathways activated upon FUT4 upregulation, high FUT4-expressing lung cancer cells showed markedly enhanced susceptibility in the tail vein metastasis mouse model to an EGFR inhibitor, afatinib, compared to the vector control. On the other hand, the inhibitory effect of TGFβ inhibitor, LY2157299, was less significant and appeared to inhibit the vector control as well. Furthermore, we showed that FUT4-mediated EGFR activation can occur in lung tumors not carrying EGFR activating mutations (Fig. 5D). Collectively, our data suggest the possibility for the use of EGFR-targeted therapy in a subset of EGFR-wild type lung cancers with medium-to-high FUT4 expressions. As many EGFR inhibitors have been FDA-approved in treating lung cancers with activating

EGFR mutations, their safety profiles have been extensively evaluated in the clinic. It would be rather straightforward to design future clinical trials and test the use of EGFR-targeted therapy in patients with wild-type EGFR lung cancers expressing higher levels of FUT4.

In summary, our results suggest that FUT4, a terminal α1,3-fucosyltransferase, is an important prognostic indicator in lung adenocarcinoma patients. FUT4 promotes lung cancer progression at multiple steps of metastatic process — epithelial-mesenchymal transition, increased migration/invasion, vascular adhesion, extravasation, and tumor metastasis at distant organs. These profound cellular and functional effects can be attributable to the augmentation of oncogenic signaling pathways by FUT4-mediated aberrant fucosylation, which may be reversed by targeting FUT4 directly or pathway-specific inhibitors (Fig. 6). Our study may pave way to development of novel glycol-based therapeutic strategies for clinical management for lung cancer metastases.

Materials and methods

Human tissues and subject follow-up data

Surgically-resected lung tumors and adjacent uninvolved lung tissues were obtained from patients with non-small cell lung cancer from National Taiwan University Hospital (NTUH, Taipei, Taiwan) from 1994 to 2010. Clinical information including gender, histological type, stage, and overall survival was collected. Median follow-up time was 51 months (range 2-164 months). The staging of lung cancer was based on the American Joint Committee on Cancer (AJCC) TNM system. This study has been

approved by the Institutional Review Board (IRB) of National Taiwan University Hospital (Approval No.: 201701010RINB).

TCGA data acquisition and analysis

Normalized RNA-seq expression datasets and EGFR mutation status of The Cancer Genome Atlas (TCGA) lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) cohorts were acquired from Firehose data repository using the R/Bioconductor package “RTCGAToolbox” (version 2.10.0, run data 2016-01-28)[50]. We used log-rank test to compare between-group survival differences. Patients were censored at the last follow-up or death, with post-surgery recurrence (>3 months) classified as an event for relapse-free survival. As for overall survival, patients were censored at the last follow-up, and death was classified as an event. Heatmaps of gene expression profiles were created by R package “pheatmap” version 1.0.10.

RNA purification and reverse transcription

Total RNA was extracted from monolayer cell cultures and lung cancer tissues using TRIzol™ reagent (ThermoFisher Scientific) according to the manufacturer's instructions. All RNA was treated with 10 U/μL DNase I (Invitrogen). Total RNA (1 μg per 25μL reaction) was reverse- transcribed using 200 U Moloney murine leukemia virus (M-MLV) reverse transcriptase primed with 0.5 μg random hexamer oligonucleotides (Promega).

RNA-sequencing and data analysis

Total RNAs were isolated from human lung cancer tissues from National Taiwan University Hospital and lung cancer cell lines. RNA integrity was evaluated with an Agilent 2100 Bioanalyzer, and were subjected to quality check with a Bioanalyzer 2100 using an RNA 6000 nano kit (Agilent Technologies). RNA sequencing was performed using Illumina HiSeq 4000 ® (for lung cancer tissue RNAs) and Illumina NextSeq 500 ® systems (for cell line RNAs). Raw reads in FASTAQ format were processed using Trimmomatic software version 0.33 for adapter trimming and quality filtering [51]. The processed sequencing reads were mapped to the human reference genome (hg19) using bowtie v2.2.6 with parameters: --forward-prob 0 --output- genome-bam for cell lines and clinical samples [52,53]. The mapped reads in BAM files were annotated with GENCODE Release 25 (GRCh37), using GenomicFeatures (version 1.32.2) and GenomicAlignments (version 1.16.0) packages. Finally, DESeq2 package (version 1.20.0) was used to generate a raw-counts matrix followed by FPKM normalization and differential expression analysis [54]. Gene Set Enrichment Analysis (GSEA) was performed using the official software javaGSEA version 3.0 with the Molecular Signatures Database (MSigDB, version 6.1) containing BioCarta, Reactome and KEGG gene sets [34,55]. The *signal-to-noise metric* to rank differentially-expressed genes in GSEA. Number of permutations were set at 1000, and enrichment statistic were set as weighted. The resulting gene sets with nominal p value < 0.05 were considered as enriched. All RNA-seq data have been deposited in GEO: GSE120622.

Human lung cancer cell lines

CL1-0 and CL1-5 human lung adenocarcinoma cell lines were kindly provided by Prof. Pan- Chyr Yang (Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan). The CL-1 cell line was established from a 64-year-old man with a poorly differentiated adenocarcinoma. The CL1-5, a subclone derived from CL1-0, possesses higher invasiveness and metastatic ability *in vitro* and *in vivo*. A549 human lung adenocarcinoma cell line was obtained from ATCC. Cell line authentication was performed by STR analysis for all cell lines used in this study. Cells were maintained in complete RPMI-1640 media with 10% (v/v) FBS, 1% (v/v) L-glutamine, and 1% (v/v) Penicillin-Streptomycin (Gibco™) at 37°C and 5% CO_2. Detection of mycoplasma in cell culture was performed every season in our lab to avoid mycoplasma contamination. For signaling transduction assays, cells were cultured in medium containing 0.5% (v/v) FBS for 16 hours for cell cycle synchronization. Then cells were treated with EGF (10 ng/mL; AF-100-15, PeproTech) or TGF-β1 (5 ng/mL; 100-21, PeproTech) for 1, 2, 6 hours, respectively. To block the effects of growth factors, cells were pre-treated with Afatinib (2 μM; T2303, TargetMol) or LY2157299 (10 μM; T2510, TargetMol) for 1 hour, followed by treatment with EGF or TGF-β for 24 hours.

Tet-off system

The human FUT4 coding sequence was obtained from CL1-5 lung adenocarcinoma cell line and was inserted into the pcTRE2hyg vector (BD Biosciences). A549 cells were transfected with 5 **μ**g of target gene expression vector (FUT4) or empty vector (Vector control) in 100-mm dishes using Lipofectamine™ 3000 Transfection Reagent (Invitrogen, L3000) according to the manufacturer's instruction. After 4 weeks of

hygromycin B (1600 μg/ml) selection, the A549 cell lines stably expressing FUT4 (FUT4) or empty vector (Vector) were established. For tet-off assays, cells were treated with doxycycline (0~2 ng/mL) to turn down FUT4 expression.

RNA interference

FUT4 expression was knockdown using shERWOOD UltramiR lentiviral inducible shFUT4 system. pZIP-Mcmv-shFUT4 plasmids (15 μg) were transfected into cells (5*106 cells) using Lipofectamine™ 3000 transfection reagents (45 μL). Cells were incubated for 16 hours and selected by fresh RPMI medium with 2 μg/mL puromycin for 24 hours. The effects of shRNA knockdown will be evaluated using proliferation assays, migration assays, and invasion assays.

Invasion/migration assay

Cell invasion assays were performed using 24-well transwell units with permeable polycarbonate filter with pore sizes of 8 μm. 24-well transwell units were pre-coated with 2 mg/mL BD Matrigel® (BD Biosciences), and incubated at 37oC for 1 hour. Each lower compartment of the transwell unit contained 10% FBS as chemoattractant. $5*10^4$ Cells in 0.5 ml RPMI with 0.5% BSA were added into the upper compartment of the transwell unit and were incubated at 37oC overnight. On the next day, the Matrigel® coated on the filter was removed with cotton swabs, and the filter was gently separated from the transwell unit. The cells attached to the lower surface of the filter membrane were fixed with 4% paraformaldehyde in PBS, and stained with 300nM DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride). The number of cells was counted at 400X magnification

under a fluorescent microscope. For each sample, the average number of cells from eight high-power fileds (HPF) was recorded. All experiments were performed in triplicate. Cell migration assays were performed using the same 24-well transwell units without the Matrigel® coating.

Single cell tracking

Cells were seeded on the 6-well plate overnight. On the next day, time-lapse imaging of the cells was performed using an automated inverted microscope (Leica DMI 6000B) for 24 hours. Data were analyzed by MetaMorph® software. For each sample, at least 20 cells were selected and tracked for cell motility calculation.

Immunofluorescence staining

For immunolocalization, samples were blocked with 1% (wt/vol) BSA in PBS for 1 hour, followed by incubation at 4°C with primary antibodies for 16 h and with secondary antibodies (donkey anti-mouse IgG Alexa 488 or donkey anti-goat IgG Alexa 488, 1:250 dilution) for 1 hour. Cells were counterstained with 300 nM DAPI for 10 min before mounted with ProLong Gold Antifade Mountant (ThermoFisher Scientific). Images were obtained with the Zeiss LSM 880 confocal microscope, and analyzed with ZEN microscope software. Antibodies used in this study were listed in Supplemental Table 2.

Adhesion assay

The uncharged polystyrene 96-well plate were coated with different recombinant proteins- Collagen IV (Merck Millipore), P-Selectin, E-Selectin or L-selectin (PeproTech), and incubated at room temperature for 2 hours. The coated 96-well plates were blocked with 5% BSA/PBS at 37oC for 1 hour. Subsequently, 5 x 10^4 cells in serum-free medium were added onto the coated 96-well plate. After incubation at 37oC for 1 hour, non-adherent cells were gently washed away twice with 1x PBS. Adherent cells were fixed with 4% paraformaldehyde/PBS for 20 minutes, and stained with 300 nM DAPI. Cell images were obtained with MD ImageXpress Micro XL High-Content Analysis System (Molecular Devices). Numbers of adherent cell were calculated with the MetaMorph software (Molecular Devices).

Mouse experiments

4-6 weeks old male C57BL/6JNarl, NOD.CB17-*Prkdc*^{scid}/JNarl (NOD SCID) and BALB/cAnN.Cg-*Foxn1nu*/CrlNarl (NUDE) mice were purchased from National Laboratory Animal Center (Taiwan) and maintained under standard pathogen free conditions. *Spontaneous metastasis of subcutaneous tumor xenografts:* Tumor xenografts were established in 6-week-old NOD SCID mice via subcutaneous injection into the dorsal region of each mouse with 1 x 10^7 cells/0.15 ml Hank's balanced salt solution containing Matrigel® (BD Biosciences). Experiments were performed at five mice in each group (vector control, FUT4^{med} and FUT4^{high}). Mice were then sacrificed at 8 weeks. Distant metastases in the lungs were evaluated by gross visualization of nodules on the lung surfaces, and by microscopic imaging of lung tissue sections. *Tail Vein Assay for Cancer Metastases:* A549 (FUT4^{high} and vector control) and CL1-5

(shFUT4 and scramble control) human lung cancer cells were pre- stained with Cyto-ID™ long-term dyes (Enzo Life Sciences), and were injected into the tail vein of 6-week-old nude mice with $5*10^5$ cells per mouse. In vivo bioluminescent imaging to monitor cancer cell metastasis was performed using In Vivo Imaging System (IVIS) at day 1 and day 7 after tail vein injection. Mice were euthanized with carbon dioxide at 28 days. All organs were removed and fixed in 10% formalin. The lung nodules were counted through gross inspection and under microscopic examination. The number of mice used in each group (n = 6) was based on the goal of having 98% power to detect a 2-fold difference in nodule numbers between groups at $p < 0.05$. *In vivo adhesion/extravasation assays:* A549 ($FUT4^{high}$ and vector control) and CL1-5 (shFUT4 and scramble control) human lung cancer cells were pre-stained with Cyto-ID™ long-term dyes (Enzo Life Sciences), and intracardially injected into the right ventricle of 6-week-old C57BL/6JNarl mice (5 mice per group; $5*10^5$ cells/mouse). Mice were kept warm with electric heating blankets for 20 minutes. Mice were then sacrificed and intravenously perfused with phosphate-buffered saline to remove blood cells and unattached human cancer cells. The lungs were removed and fixed with 1% agarose on a 35mm imaging μ-Dish with a high glass bottom (ibidi, 81158). The lungs were examined by the Axio Observer Inverted Microscope System (Zeiss). *Drug treatment assay:* A549 ($FUT4^{high}$ and vector control) human lung cancer cells were injected into the tail vein of 6-week-old nude mice with $5*10^5$ cells per mouse to establish lung cancer metastatic model. Mice were intraperitoneally injected with DMSO (mock control), afatinib (0.25 mg/Kg BW) (TargetMol TM-T2510) and LY2157299 (15 mg/Kg BW) (TargetMol TM-T2303) three times per week from week 2 to week 7 post-injection

to treat metastatic lung tumor. Mice were euthanized with carbon dioxide at 49 days. Lung were removed and fixed in 10% formalin. The lung nodules were counted through gross inspection and under microscopic examination. The number of mice used in each group (n = 6) was based on the goal of having 98% power to detect a 2-fold difference in nodule numbers between groups at $p < 0.05$.

Immunoprecipitation

Lewis and TβRI immunoprecipitations were performed using PierceTM Protein L Magnetic Beads (ThermoFisher Scientific, #88849) and PierceTM Protein A/G Magnetic Beads (ThermoFisher Scientific, #88802) respectively, according to the manufacturer's protocol. A549 and CL1-0 cells (both vector control and FUT4-overexpressed cells) were lysed with RIPA buffer containing a protease inhibitor cocktail (GOAL Bio), and incubated on ice for 20 minutes. Cell lysate was centrifuged for 15 min at 4°C, and an aliquot of the supernatant was kept aside on ice as input. 1 mg of cell lysate was incubated with 40 µg of primary antibody per IP at 4oC on a shaker overnight. The following antibodies were used for IP: TGF beta receptor I (Abcam, ab31013) and Lewis X (Abcam, ab3358). Next day, the magnetic beads were loaded in the cell lysate for 4 hours at room temperature and washed according to the manufacturer's protocol. For Western blotting and mass spectrometry, beads were resuspended in the SDS sample buffer, boiled for 5 min, and loaded onto an 8% acrylamide gel.

Mass spectrometry

The post-IP samples were subjected to SDS-PAGE and the subsequent in-gel trypsin digestion as described (Lin et al., 2003). The peptides were dried using the CentriVap Benchtop Vacuum Concentrator (Labconco Corp.), dissolved in 0.1% trifluoroacetic acid (TFA), and desalted with C18 Ziptips (Millipore) according to the manufacturer's protocol. NanoLC−nanoESI-MS/MS identification was performed on an Ultimate system 3000 nanoLC system (ThermoFisher Scientific) connected to the Orbitrap Fusion Lumos mass spectrometer equipped with NanoSpray Flex ion source (ThermoFisher Scientific). Briefly, peptide mixtures conditioned in 0.1% formic acid were injected into the HPLC instrument and enriched on a C18 Acclaim PepMap NanoLC reverse phase column of 25 cm length, 75 µm internal diameter and 2 µm particle size with a pore of 100 Å (ThermoScientific Scientific). The mobile phase consisted of aqueous solvent (A) 0.1% formic acid, and organic solvent (B) 0.1% formic acid in acetonitrile. Peptides were separated by 90 min of the segmented gradient from 2% to 40% solvent B at a flow rate of 300 nL/min. The mass spectra were acquired using one full MS scan followed by data-dependent MS/MS of the most intense ions in 3 sec. Full MS scans were recorded with an automatic gain control (AGC) target set at 500,000 ions and a resolution of 120,000 at m/z=200. For MS/MS spectra, selected parent ions within a 1.4 Da isolation window were fragmented by high-energy collision activated dissociation (HCD) with normalized collision energies of 32 eV. HCD-MS/MS with a resolution of 15,000, AGC at 50,000, and potential charge status of 2+ to 7+ were performed. A dynamic exclusion duration for the selection of parent ions was set at 180 sec with a repeat count. Spectra were collected by Xcalibur tool version 4.1 (ThermoScientific Scientific). The resulting RAW files were directly processed and analyzed by using MaxQuant software version

1.6.0.16. MS/MS peaks were searched in the Andromeda peptide search engine against the forward and reverse (as a decoy database) of reviewed UniProtKB/Swiss-Prot human proteome database. The common contaminants list attached in MaxQuant was also provided during the search (Tyanova et al., 2016). Peptides with minimum of 6 amino acids and maximum trypsin missed cleavages of 3 were considered. Variable modifications including methionine oxidation, N-terminal acetylation, Serine/Threonine/Tyrosine phosphorylation, and Asparagine/Glutamine deamidation, and the fixed modification of carbamidomethyl cysteine were applied. Initial parent peptide and fragment mass tolerance were set to 4.5 and 20 ppm, respectively. False discovery rate (FDR) filtration of peptide-spectrum match and protein levels were applied at 0.05 and 0.01, respectively. Finally, proteins identified as reverse, common contaminations or matched with only one unique peptide were excluded.

Statistics

Statistical analysis of expression data was performed by R for Mac (Version 3.5.0) and GraphPad Prism 7 for Mac software. Patient demographics and survival information were gathered from database registry, electronic medical record and Taiwan Cancer Registry (for NTUH cohort). Cox proportional hazard modeling was used to examine the effect of expression levels of various fucosyltransferases on overall survival (time until death). The correlation between all members in the FUTs family were analyzed using Pearson correlation coefficients. Statistical analysis of the experimental data, including migration/invasion assays, drug treatments, etc, were performed using Mann-Whitney

test or one-way ANOVA with Dunnett's test. Data are presented as mean ± standard errors. p value < 0.05 is considered statistically significant.

Study approval

The study on primary tumor tissues from lung cancer patients was approved by the Institutional Review Board (IRB) of National Taiwan University Hospital (Approval No.: 201701010RINB). Written informed consent was obtained from all participants prior to inclusion in the study. Every participant was assigned a number and the data were analyzed at the group level for patient de-identification. All mice experiments were performed in compliance with an experimental protocol (20160126) approved by the NTU College of Medicine Institutional Animal Care and Use Committee (IACUC).

Data availability

All RNA-seq data are available at the National Center for Biotechnology Information Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo) under the accession number GSE120622. All proteomics data are available at the MassIVE repository (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) under the accession number: MSV000083028

Author contributions

Conceptualization, H.-C.T. and C.-J.Y.; Methodology; Y.-H.J. and H.-H.L.; Software and Formal Analysis, Y.-W.C., S.-Y.L., G.A.O, and R.R.W.; Resources, H.-C.T., J.-Y.S. and C.-J.Y.; Investigation, H.-H.L., Y.-H.J., H.-H.H., Z.-C.H., and Y.-J.H.; Writing – Original

Draft, H.-H.L. and H.-C.T.; Writing – Review & Editing, H.-C.T. and C.-J.Y.; Funding Acquisition, H.-C.T. and C.-J.Y.; Supervision, H.-C.T. and C.-J.Y.

Acknowledgments

We acknowledge the service provided by the Flow Cytometric Analyzing and Sorting Core of the First Core Laboratory and Laboratory Animal Center of National Taiwan University College of Medicine, Instrumentation Center, National Taiwan University, as well as Confocal Microscopy Core Facility, Institute of Biomedical Sciences, Academia Sinica. We also thank Ms. Min-Chuan Hsu and Ms. Pei-Chun Tsai for their administrative assistance. This study was supported by Ministry of Science and Technology (MOST 107-2314-B-002 -241, MOST 105-2628-B-002 -040), and National Health Research Institutes (NHRI-EX106-10610BC).

References

1 Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867, doi:10.1016/j.cell.2006.08.019 (2006).

2 Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540-555, doi:10.1038/nrc3982 (2015).

3 Oliveira-Ferrer, L., Legler, K. & Milde-Langosch, K. Role of protein glycosylation in cancer metastasis. *Semin Cancer Biol* **44**, 141-152, doi:10.1016/j.semcancer.2017.03.002 (2017).

4 Blanas, A., Sahasrabudhe, N. M., Rodriguez, E., van Kooyk, Y. & van Vliet, S. J. Fucosylated Antigens in Cancer: An Alliance toward Tumor Progression, Metastasis, and Resistance to Chemotherapy. *Front Oncol* **8**, 39, doi:10.3389/fonc.2018.00039 (2018).

5 Miyoshi, E., Moriwaki, K. & Nakagawa, T. Biological function of fucosylation in cancer biology. *J Biochem* **143**, 725-729, doi:10.1093/jb/mvn011 (2008).

6 Chen, C. Y. *et al.* Fucosyltransferase 8 as a functional regulator of nonsmall cell lung cancer. *Proc Natl Acad Sci U S A* **110**, 630-635, doi:10.1073/pnas.1220425110 (2013).

7 Honma, R. *et al.* Expression of fucosyltransferase 8 is associated with an unfavorable clinical outcome in non-small cell lung cancers. *Oncology* **88**, 298-308, doi:10.1159/000369495 (2015).

8 Ogawa, J., Sano, A., Inoue, H. & Koide, S. Expression of Lewis-related antigen and prognosis in stage I non-small cell lung cancer. *Ann Thorac Surg* **59**, 412-415 (1995).

9 Tanaka, F. *et al.* Lewis Y antigen expression and postoperative survival in non-small cell lung cancer. *Ann Thorac Surg* **66**, 1745-1750 (1998).

10 Kadota, A., Masutani, M., Takei, M. & Horie, T. Evaluation of expression of CD15 and sCD15 in non-small cell lung cancer. *Int J Oncol* **15**, 1081-1089 (1999).

11 Tian, L. *et al.* Ginsenoside Rg3 inhibits epithelial-mesenchymal transition (EMT) and invasion of lung cancer by down-regulating FUT4. *Oncotarget* **7**, 1619-1632, doi:10.18632/oncotarget.6451 (2016).

12 Jassam, S. A., Maherally, Z., Ashkan, K., Pilkington, G. J. & Fillmore, H. L. Fucosyltransferase 4 and 7 mediates adhesion of non-small cell lung cancer cells to brain-derived endothelial cells and results in modification of the blood-brain-barrier: in vitro investigation of CD15 and CD15s in lung-to-brain metastasis. *J Neurooncol*, doi:10.1007/s11060-019-03188-x (2019).

13 Liang, J. X., Gao, W. & Cai, L. Fucosyltransferase VII promotes proliferation via the EGFR/AKT/mTOR pathway in A549 cells. *Onco Targets Ther* **10**, 3971-3978, doi:10.2147/OTT.S140940 (2017).

14 Liu, Y. C. *et al.* Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells. *Proc Natl Acad Sci U S A* **108**, 11332-11337, doi:10.1073/pnas.1107385108 (2011).

15 Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543-550, doi:10.1038/nature13385 (2014).

16 Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519-525, doi:10.1038/nature11404 (2012).

17 Lan, Q. *et al.* Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat Genet* **44**, 1330-1335, doi:10.1038/ng.2456 (2012).

18 Shi, Y. *et al.* A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). *J Thorac Oncol* **9**, 154-162, doi:10.1097/JTO.0000000000000033 (2014).

19 Hirakawa, M. *et al.* Fucosylated TGF-beta receptors transduces a signal for epithelial-mesenchymal transition in colorectal cancer cells. *Br J Cancer* **110**, 156-163, doi:10.1038/bjc.2013.699 (2014).

20 Tsai, J. H. & Yang, J. Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev* **27**, 2192-2206, doi:10.1101/gad.225334.113 (2013).

21 Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. Emt: 2016. *Cell* **166**, 21-45, doi:10.1016/j.cell.2016.06.028 (2016).

22 Yang, X., Liu, S. & Yan, Q. Role of fucosyltransferase IV in epithelial-mesenchymal transition in breast cancer cells. *Cell Death Dis* **4**, e735, doi:10.1038/cddis.2013.241 (2013).

23 Tu, C. F., Wu, M. Y., Lin, Y. C., Kannagi, R. & Yang, R. B. FUT8 promotes breast cancer cell invasiveness by remodeling TGF-beta receptor core fucosylation. *Breast Cancer Res* **19**, 111, doi:10.1186/s13058-017-0904-8 (2017).

24 Bolos, V. *et al.* The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. *J Cell Sci* **116**, 499-511 (2003).

25 Puisieux, A., Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol* **16**, 488-494, doi:10.1038/ncb2976 (2014).

26 Hou, R. *et al.* Lewis(y) antigen promotes the progression of epithelial ovarian cancer by stimulating MUC1 expression. *Int J Mol Med* **40**, 293-302, doi:10.3892/ijmm.2017.3009 (2017).

27 Liu, J. *et al.* Lewis y antigen promotes the proliferation of ovarian carcinoma-derived RMG-I cells through the PI3K/Akt signaling pathway. *J Exp Clin Cancer Res* **28**, 154, doi:10.1186/1756-9966-28-154 (2009).

28 Nakayama, F. *et al.* CD15 expression in mature granulocytes is determined by alpha 1,3-fucosyltransferase IX, but in promyelocytes and monocytes by alpha 1,3-fucosyltransferase IV. *J Biol Chem* **276**, 16100-16106, doi:10.1074/jbc.M007272200 (2001).

29 Ogawa, J., Inoue, H. & Koide, S. Expression of alpha-1,3-fucosyltransferase type IV and VII genes is related to poor prognosis in lung cancer. *Cancer Res* **56**, 325-329 (1996).

30 Azuma, Y., Ito, M., Taniguchi, A. & Matsumoto, K. Expression of cell surface Lewis X and Y antigens and FUT4 mRNA is increased in Jurkat cells undergoing apoptosis. *Biochim Biophys Acta* **1672**, 157-163, doi:10.1016/j.bbagen.2004.03.006 (2004).

31 Jassam, S. A. *et al.* TNF-alpha enhancement of CD62E mediates adhesion of non-small cell lung cancer cells to brain endothelium via CD15 in lung-brain metastasis. *Neuro Oncol* **18**, 679-690, doi:10.1093/neuonc/nov248 (2016).

32 Jassam, S. A. *et al.* CD15s/CD62E Interaction Mediates the Adhesion of Non-Small Cell Lung Cancer Cells on Brain Endothelial Cells: Implications for Cerebral Metastasis. *Int J Mol Sci* **18**, doi:10.3390/ijms18071474 (2017).

33 Chu, Y. W. *et al.* Selection of invasive and metastatic subpopulations from a human lung adenocarcinoma cell line. *Am J Respir Cell Mol Biol* **17**, 353-360, doi:10.1165/ajrcmb.17.3.2837 (1997).

34 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

35 Larue, L. & Bellacosa, A. Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways. *Oncogene* **24**, 7443-7454, doi:10.1038/sj.onc.1209091 (2005).

36 Yin, X., Rana, K., Ponmudi, V. & King, M. R. Knockdown of fucosyltransferase III disrupts the adhesion of circulating cancer cells to E-selectin without affecting hematopoietic cell adhesion. *Carbohydr Res* **345**, 2334-2342, doi:10.1016/j.carres.2010.07.028 (2010).

37 Liang, L. *et al.* miR-125a-3p/FUT5-FUT6 axis mediates colorectal cancer cell proliferation, migration, invasion and pathological angiogenesis via PI3K-Akt pathway. *Cell Death Dis* **8**, e2968, doi:10.1038/cddis.2017.352 (2017).

38 Yang, X. S. *et al.* Overexpression of fucosyltransferase IV promotes A431 cell proliferation through activating MAPK and PI3K/Akt signaling pathways. *J Cell Physiol* **225**, 612-619, doi:10.1002/jcp.22250 (2010).

39 Yan, L. M. *et al.* Enhancement of the adhesive and spreading potentials of ovarian carcinoma RMG-1 cells due to increased expression of integrin alpha5beta1 with the Lewis Y-structure on transfection of the alpha1,2-fucosyltransferase gene. *Biochimie* **92**, 852-857, doi:10.1016/j.biochi.2010.02.012 (2010).

40 Liu, J. J. *et al.* Lewis(y) antigen stimulates the growth of ovarian cancer cells via regulation of the epidermal growth factor receptor pathway. *Oncol Rep* **23**, 833-841 (2010).

41 Beerling, E. *et al.* Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep* **14**, 2281-2288, doi:10.1016/j.celrep.2016.02.034 (2016).

42 Ocana, O. H. *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* **22**, 709-724, doi:10.1016/j.ccr.2012.10.012 (2012).

43 Giordano, G. *et al.* Cancer-related CD15/FUT4 overexpression decreases benefit to agents targeting EGFR or VEGF acting as a novel RAF-MEK-ERK kinase downstream regulator in metastatic colorectal cancer. *J Exp Clin Cancer Res* **34**, 108, doi:10.1186/s13046-015-0225-7 (2015).

44 Shan, X. *et al.* Ginsenoside Rg3 suppresses FUT4 expression through inhibiting NF-kappaB/p65 signaling pathway to promote melanoma cell death. *Int J Oncol* **47**, 701-709, doi:10.3892/ijo.2015.3057 (2015).

45 Sola, R. J. & Griebenow, K. Glycosylation of therapeutic proteins: an effective strategy to optimize efficacy. *BioDrugs* **24**, 9-21, doi:10.2165/11530550-000000000-00000 (2010).

46 Rillahan, C. D., Brown, S. J., Register, A. C., Rosen, H. & Paulson, J. C. High-throughput screening for inhibitors of sialyl- and fucosyltransferases. *Angew Chem Int Ed Engl* **50**, 12534-12537, doi:10.1002/anie.201105065 (2011).

47 Ahsen, O. *et al.* A miniaturized high-throughput screening assay for fucosyltransferase VII. *Anal Biochem* **372**, 96-105, doi:10.1016/j.ab.2007.08.029 (2008).

48 Rillahan, C. D. *et al.* Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome. *Nat Chem Biol* **8**, 661-668, doi:10.1038/nchembio.999 (2012).

49 Okeley, N. M. *et al.* Development of orally active inhibitors of protein and cellular fucosylation. *Proc Natl Acad Sci U S A* **110**, 5404-5409, doi:10.1073/pnas.1222263110 (2013).

50 Samur, M. K. RTCGAToolbox: a new tool for exporting TCGA Firehose data. *PLoS One* **9**, e106397, doi:10.1371/journal.pone.0106397 (2014).

51 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

52 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

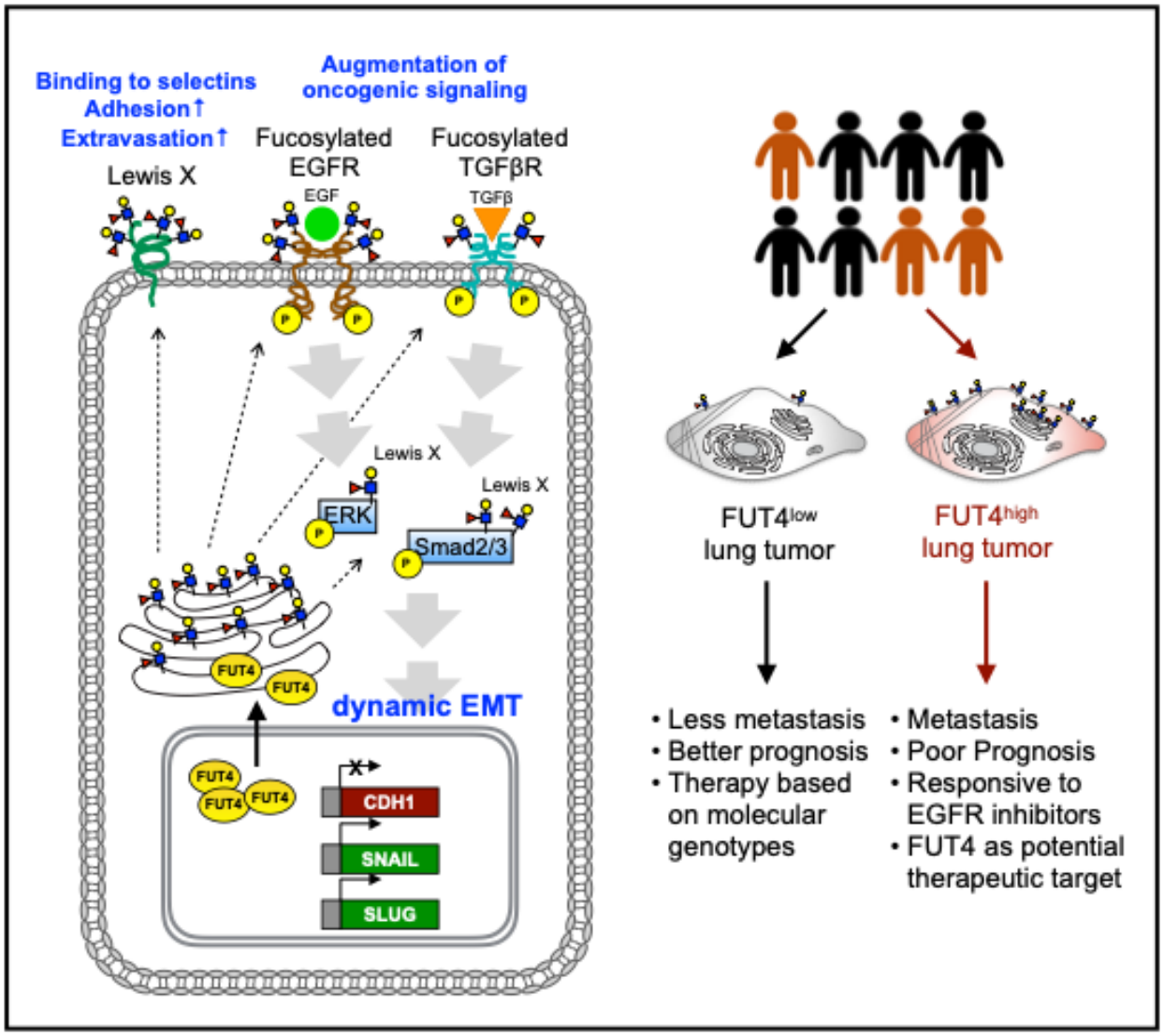
53 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).

54 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

55 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).

Figure legends

Graphic abstract



Fucosyltransferase 4 is a major prognostic indicator in lung adenocarcinoma and drives lung cancer metastasis via fucosylation-mediated intensification of oncogenic signaling. We identified a terminal α1,3-fucosyltransferase, FUT4, as a key prognostic predictor for lung adenocarcinoma patients. *Mechanistically*, FUT4 mediates the metastatic phenotype via aberrant

fucosylation of signaling cascade proteins and augments specific oncogenic signaling through epidermal growth factor (EGF) and transforming growth factor-β (TGF-β) pathways. As a result, FUT4 provokes cancer metastasis through promoting cell invasion, migration, epithelial-to-mesenchymal transition, and binding of selectins. *Therapeutically*, FUT4 high-expressing cells demonstrated enhanced sensitivity to the targeted therapy. Depletion of FUT4 or treatment with EGFR inhibitors mitigates lung cancer metastasis. Thus, FUT4 represents an oncogenic signaling amplifier and a potential therapeutic target in lung cancer metastasis.

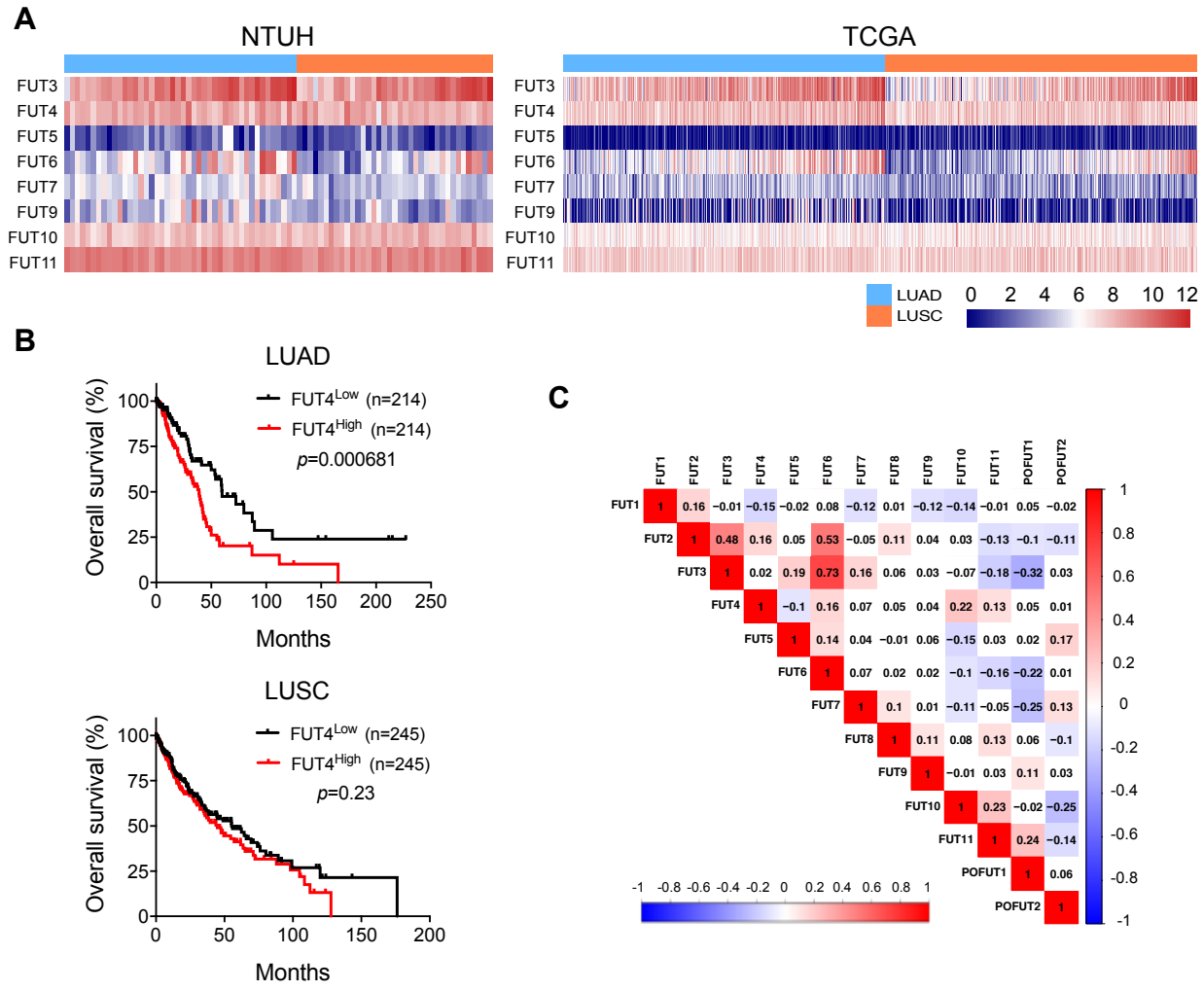


Figure 1. High FUT4 expression in lung adenocarcinoma (LUAD) is associated with poor prognosis. **A,** Heatmap of RNA-seq transcriptomic analysis for eight α-(1,3)-fucosyltransferases in the surgically-resected lung cancer tissues at National Taiwan University Hospital (LUAD, N=44 and LUSC, N=37) and from TCGA database (LUAD, N=517 and LUSC, N=502), respectively. LUAD: lung adenocarcinoma. LUSC: lung squamous cell carcinoma. **B,** Kaplan-Meier curves of overall survival for patients with lung adenocarcinoma (LUAD) or squamous cell carcinoma (LUSC). Patients were divided into two groups based on the median value of FUT4 expressions. *p* value was calculated by log-rank test. **C,** The correlation matrix of all members in the FUT family

based on the RNA-seq data of the TCGA lung adenocarcinoma samples. Numbers in the colored box represent Pearson correlation coefficients. Red and blue colors denote positive and negative correlations with statistical significance ($p < 0.05$), respectively.

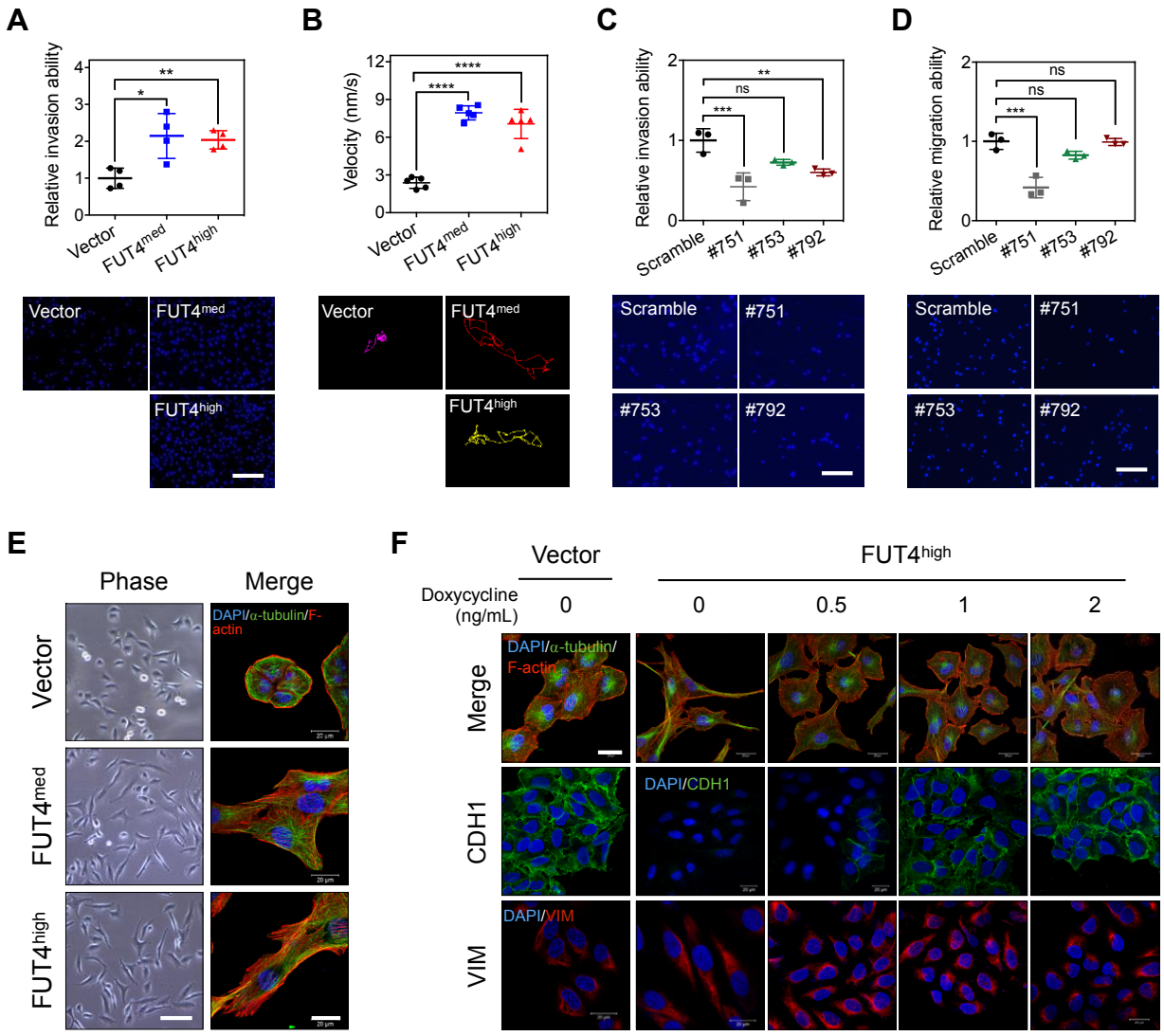


Figure 2. FUT4 triggers epithelial-mesenchymal transition (EMT) and promotes aggressive phenotypes of human lung cancer cells. **A,** Dot plots showing relative invasion ability of A549 cells measured by matrigel-based transwell invasion assays. Representative images of cells with DAPI nuclear stains in the lower chambers are shown in the lower panels. Relative invasion ability was calculated using the cell number in the lower chamber of transwell system for each clone compared to that of the vector control. **B,** Dot plots showing migration velocity of A549 cells measured by single cell tracking assays under fluorescence microscope. Data were analyzed using

Metamorph® software. The paths of cell migration were delineated in the lower panels using pseudo-colors. **C,** Dot plots showing relative invasion ability of CL1-5 cells with FUT4 knockdowns measured by matrigel-based transwell invasion assays. Representative images of cells with DAPI nuclear stains in the lower chambers are shown. **D,** Dot plots showing relative migration ability of CL1-5 cells with FUT4 knockdowns. Relative migration ability was calculated using the cell number in the lower chamber of the transwell system for each clone compared to that of the scramble control. Representative images of cells with DAPI nuclear stains in the lower chambers are shown. Scale bar in **A ~ D**, 100 μm. **E,** Immunofluorescent imaging analysis of EMT markers in A549 lung cancer cells with various levels of FUT4 expression (A549_vector, A549_FUT4^{med}, and A549_FUT4^{high}). Epithelial markers: CLDN1 (claudin-1) and CDH1 (E-cadherin). Mesenchymal markers: CDH2 (N-cadherin) and VIM (vimentin). Scale bar: 50 μm (first column) and 20 μm. **F,** Immunofluorescent imaging analysis of A549 cells carrying a tetracycline-controlled transcriptional repression (Tet-off) system of FUT4 expression at day 7 following the addition of doxycycline at concentrations of 0, 0.5, 1, 2 ng/mL. FUT4 overexpression induced epithelial-to-mesenchymal transition (0 ng/mL doxycycline), which could be reversed when FUT4 expression was turned off by doxycycline (0.5, 1, 2 ng/mL). Scale bar: 20 μm. *p* value was calculated by one-way ANOVA with Dunnett's test. All experiments were performed in triplicates and presented as mean ± standard errors. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0005$, **** $p < 0.0001$. ns: not significant.

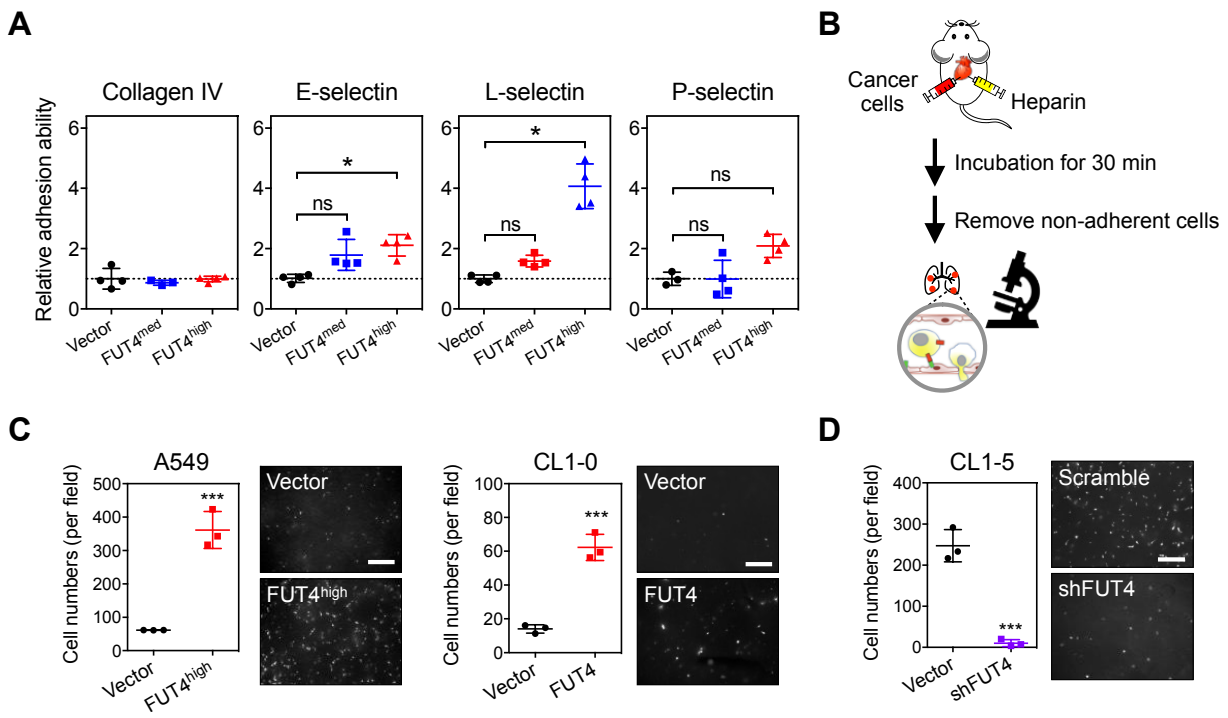


Figure 3. FUT4 enhances adhesion and colonization ability of lung cancer cells. **A,** Dot plots showing relative adhesion abilities of FUT4-overexpressing A549 lung cancer cells to collagen IV, E-selectin, L-selectin and P-selectin. Cells bound to the adhesion molecules precoated on 96-well plates were stained by DAPI and counted using Metamorph® software. Relative adhesion ability refers to the numbers of adhering cells divided by those of the vector control. **B,** Diagram of *in vivo* extravasation and lung colonization assay following injection of lung cancer cells into the right ventricle of C57BL/6 mice. Mice were sacrificed 30 mins after intracardial injection. Whole lung perfusion with normal saline was performed to remove blood and cells not adhered to the pulmonary vasculature. **C and D,** Numbers of lung cancer cells with FUT4 over-expression (A549_FUT4^{high}, CL1-0_FUT4)(**C**) or knockdowns (CL1-5_shFUT4)(**D**) adhered to the vascular walls or retained in the lung tissues following intracardial

injection were visualized under Zeiss Axio Observer microscope. All experiments were done in three biological replicates. Data are presented as mean ± standard errors. Representative images are shown on the right. Scale bar: 5 mm. *p* value was calculated by Mann-Whitney test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0005$. ns: not significant.

A

	Numbers of spontaneous metastasis		
Experiment	Vector	FUT4^{med}	FUT4^{high}
1st	0/3 (0%)	0/3 (0%)	2/4 (50%)
2nd	0/4 (0%)	2/4 (50%)	3/4 (75%)
3rd	0/2 (0%)	1/2 (50%)	2/2 (50%)
Total	0/9 (0%)	3/9 (33%)	7/10 (70%)

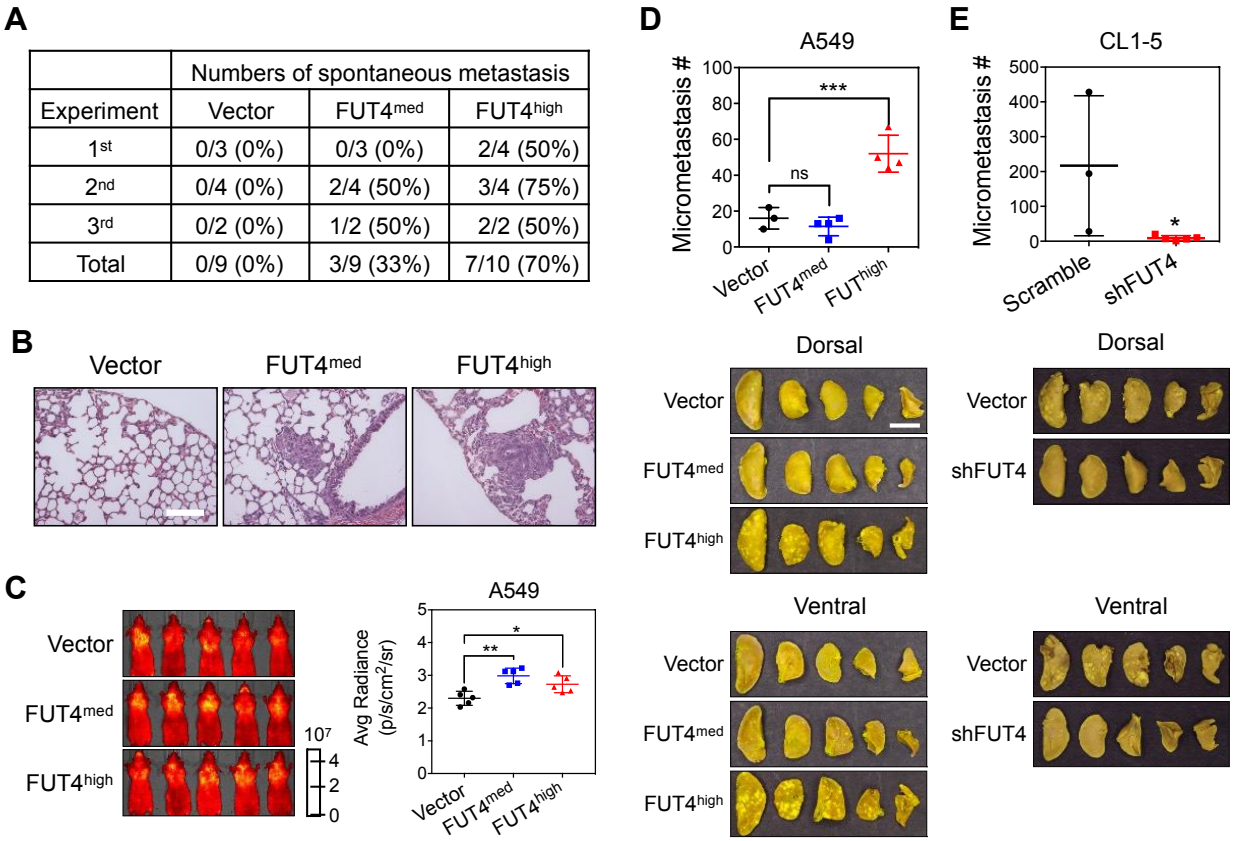


Figure 4. Depletion of FUT4 decreases lung cancer metastasis *in vivo*. **A,** Numbers (percentages) of mice with spontaneous lung metastases in NOD/SCID mice bearing subcutaneous tumors of A549_vector, A549_FUT4^{med} and A549_FUT4^{high} lung cancer cells. Data from three independent experiments are shown. **B,** Representative hematoxylin & eosin (H&E) stained images of lung tissues from mice bearing subcutaneous xenograft lung tumors from A. Scale bar: 0.1 mm. **C,** *In vivo* whole animal imaging of tumor metastasis in nude mice at 24 hours following tail vein injections of A549 lung cancer cells. Mice were intravenously injected with lung cancer cells stained with Cyto-ID™ long-term dyes and imaged with the IVIS® Spectrum. Quantification of signal radiances of metastatic foci in the dorsal and ventral sides of animals was

graphed on the right. **D and E,** Numbers of metastatic foci in the lungs of nude mice at 28 days after tail vein injection of lung cancer cells with FUT4 overexpression (A549_FUT4^{med}, A549_FUT4^{high})(**D**) or knockdown (CL1-5_shFUT4)(**E**). Pictures of lung nodules are also shown. All data in the dot plots are presented as mean ± standard errors. *p* value was calculated by one-way ANOVA with Dunnett's test or unpaired t test. * $p < 0.05$, ** $p < 0.01$ *** $p < 0.0005$. ns: not significant.

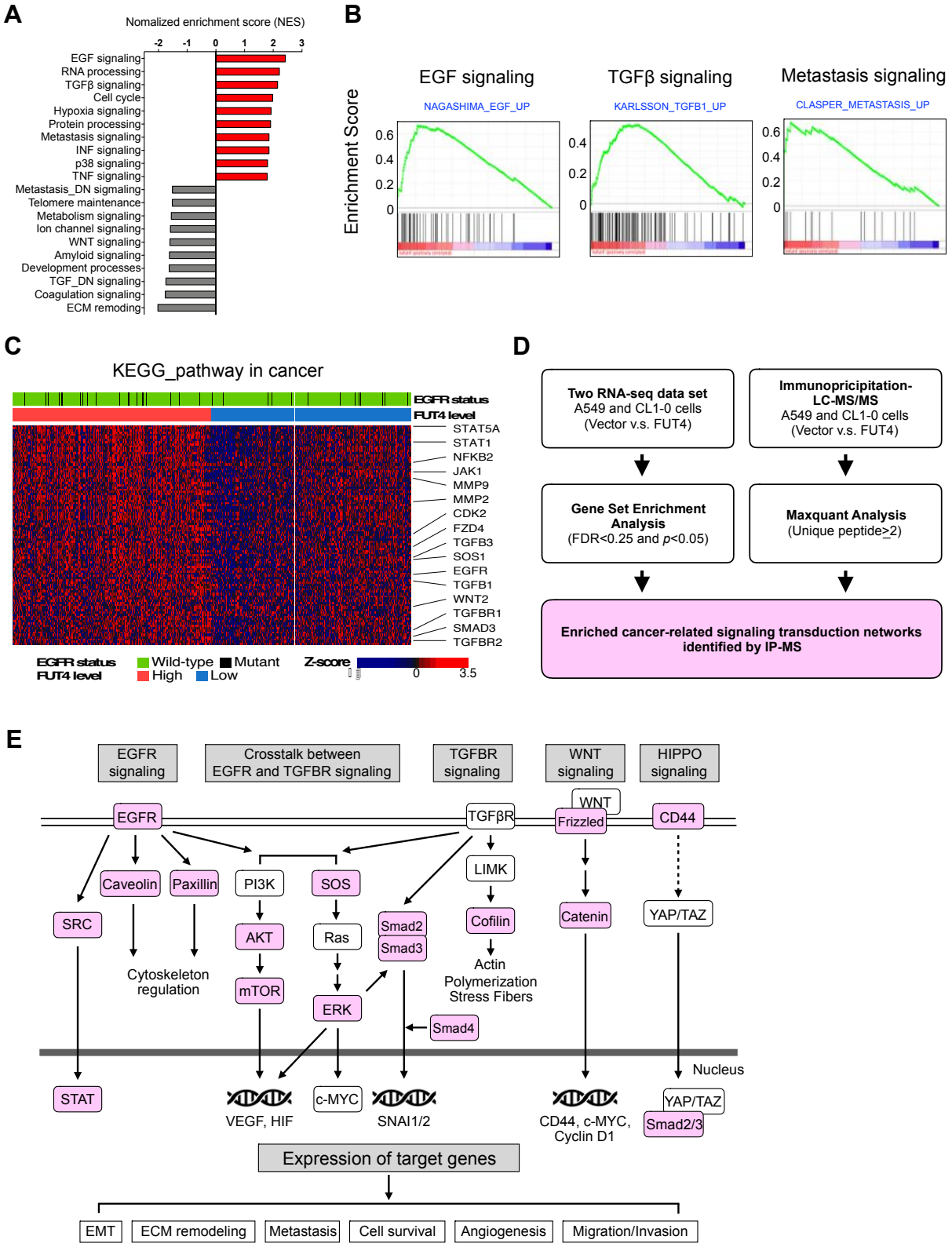


Figure 5. EGF and TGF-β signaling cascade proteins are major molecular acceptors of

FUT4. **A,** Top 10 positively and negatively enriched gene sets in genome-wide RNA-seq data of A549_FUT4^{high} and CL1-0_FUT4 cells. Data were derived from Gene Set Enrichment Analysis (GSEA) with normalized enrichment score (NES) greater than 2 and false discovery rate (FDR) less than 0.25. **B,** GSEA enrichment plots of epidermal growth factor (EGF) signaling, transforming growth factor beta (TGFβ) signaling and metastases gene sets in A549_FUT4^{high} and CL1-0_FUT4 cells. **C,** Heatmap of RNA-seq transcriptomes for the leading edge genes of KEGG_pathway_in_cancer from GSEA analyses of TCGA lung adenocarcinoma. Top sidebars denote mutation status of EGFR and expression levels of FUT4 grouped with the median value cut-off. **D,** Flow chart for integrative analysis of RNA-seq transcriptome and immuno-precipitation mass spectrometry. **E,** Immunoprecipitation (IP) with anti-Le^{x} antibody followed by tandem mass spectrometry (LC-MS/MS) in A549_FUT4^{high} and CL1-0_FUT4 cells revealed key mediator proteins (in pink) in oncogenic signaling cascades including EGF and TGFβ pathways.

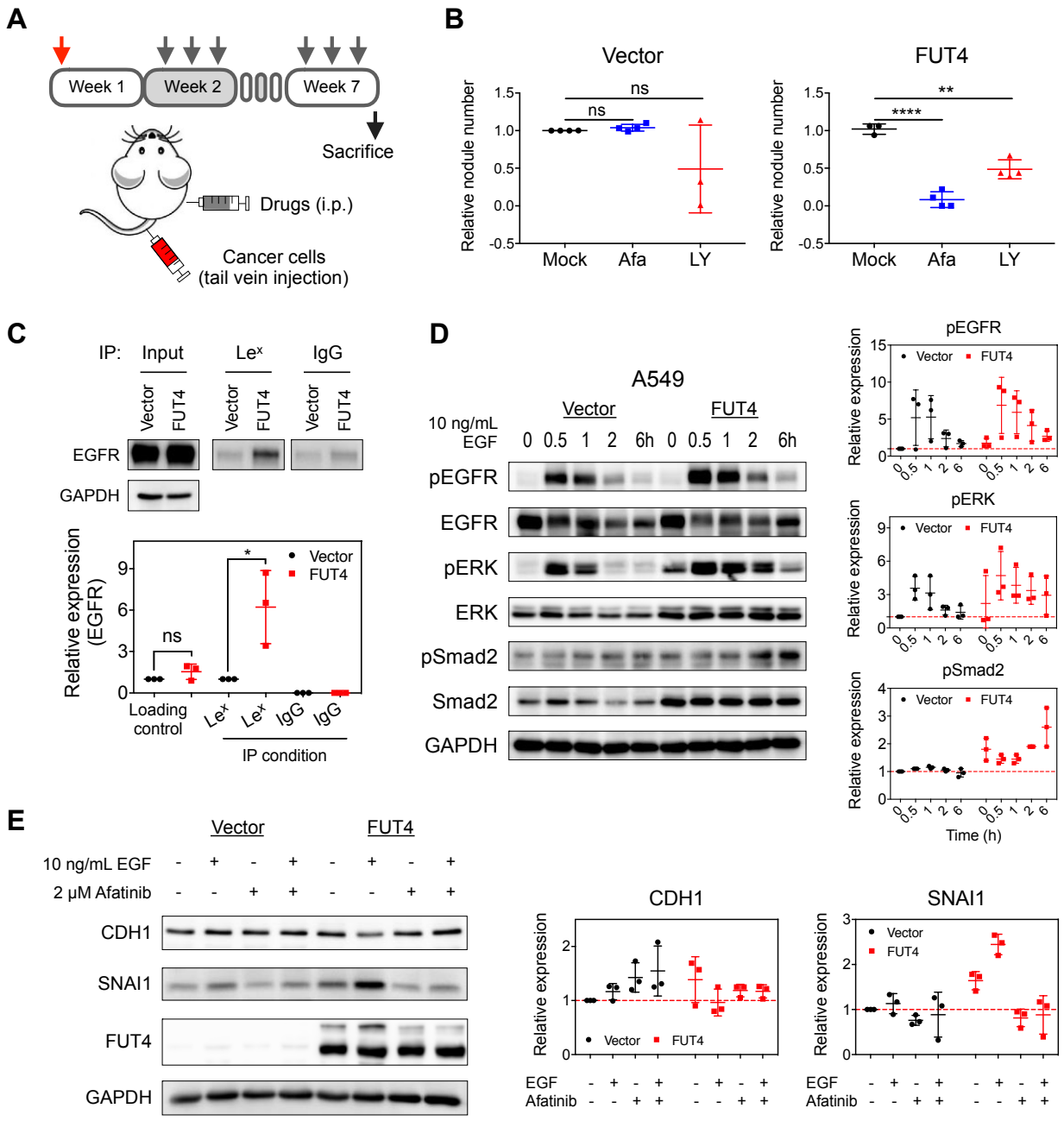


Figure 6. Inhibition of EGF signaling attenuates FUT4-mediated metastasis in lung cancer. **A,** Diagram of *in vivo* lung cancer metastasis assay via tail vein injection in nude mice subject to treatments with EGFR or TGF-β inhibitors. **B,** Dot plots showing relative metastatic ability of A549_vector or A549_FUT4^{high} lung cancer cells in nude mice receiving 0.25 mg/kg Afatinib or 15 mg/kg LY2157299 treatment. Numbers of metastatic

foci in the lungs were counted for the mice in each treatment group and normalized to the numbers in the mock (normal saline)-treated groups. *p* value was calculated by one-way ANOVA with Dunnett's test. * $p < 0.05$, ** $p < 0.01$ *** $p < 0.0005$. ns: not significant. **C,** Immunoprecipitation-western blot analysis of Lewis X (Le^x) and EGFR in A549_FUT4^{high} cells as compared to A549_vector cells. Quantifications of EGFR in anti-Lewis X pull-down lysates from both cell lines are shown in dot plots. **D,** Western blot analyses of signaling cascade proteins in the EGFR pathway in A549_FUT4^{high} versus A549_vector cells at 0, 0.5, 1, 2, and 6 hrs following the addition of 10 ng/mL EGF. Quantifications of signal intensities on western blots for phospho-EGFR, phospho-ERK and phospho-Smad2 from three biological replicates are graphed on the right. **E,** Western blot analyses of EMT marker proteins— CDH1 (epithelial marker) and SNAI1(mesenchymal marker) in A549_vector and A549_FUT4^{high} cells 24 hrs following addition of 10 ng/mL EGF and treatment with 2μM afatinib. Dot plots show quantifications of signal intensities on western blots for CDH1 and SNAI1 from three biological replicates. The data show that Afatinib treatment reverses FUT4-mediated epithelial-mesenchymal transition in A549_FUT4^{high} cells.

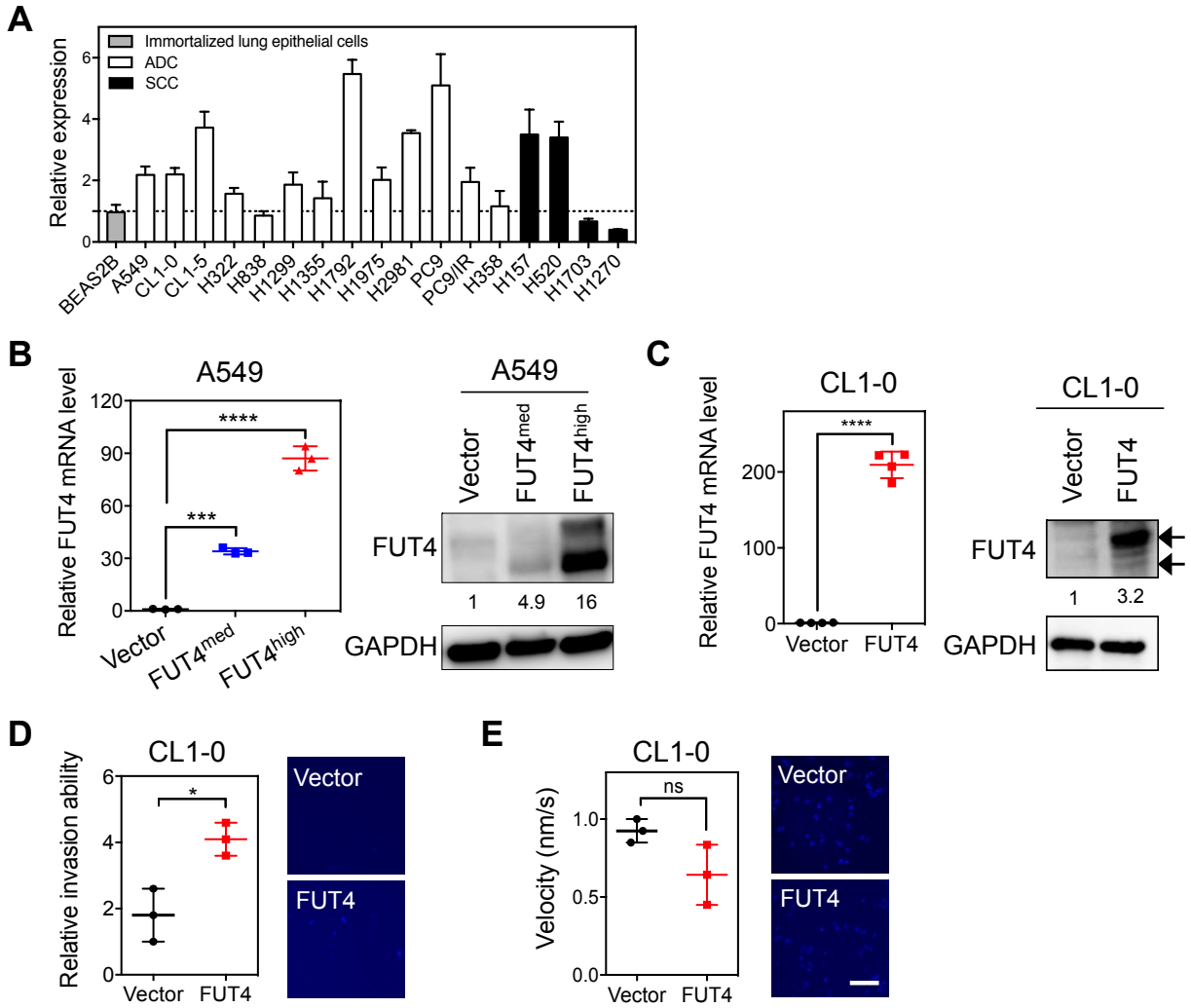
Table

Table 1. Relationships between expression levels of individual fucosyltransferases and overall survival in patients with lung adenocarcinoma. (N=428, *High vs. Low expressions, cutoff=median; Cox proportional hazard models*)

	Hazard Ratio	**Standard Error**	***p* value**
FUT1	0.4970	0.2205	0.0015†
FUT2	0.8936	0.2401	0.6495
FUT3	1.1454	0.2641	0.6072
FUT4	2.1722	0.2167	0.0003*
FUT5	1.2299	0.2106	0.3259
FUT6	0.8077	0.2792	0.4444
FUT7	0.6469	0.2265	0.0545
FUT8	1.1338	0.2109	0.5516
FUT9	0.8167	0.2150	0.3463
FUT10	0.6491	0.2172	0.0047†
FUT11	1.1472	0.2217	0.5357
POFUT1	1.2308	0.2172	0.3390
POFUT2	0.9102	0.2207	0.6699

Supplemental Figures and legends

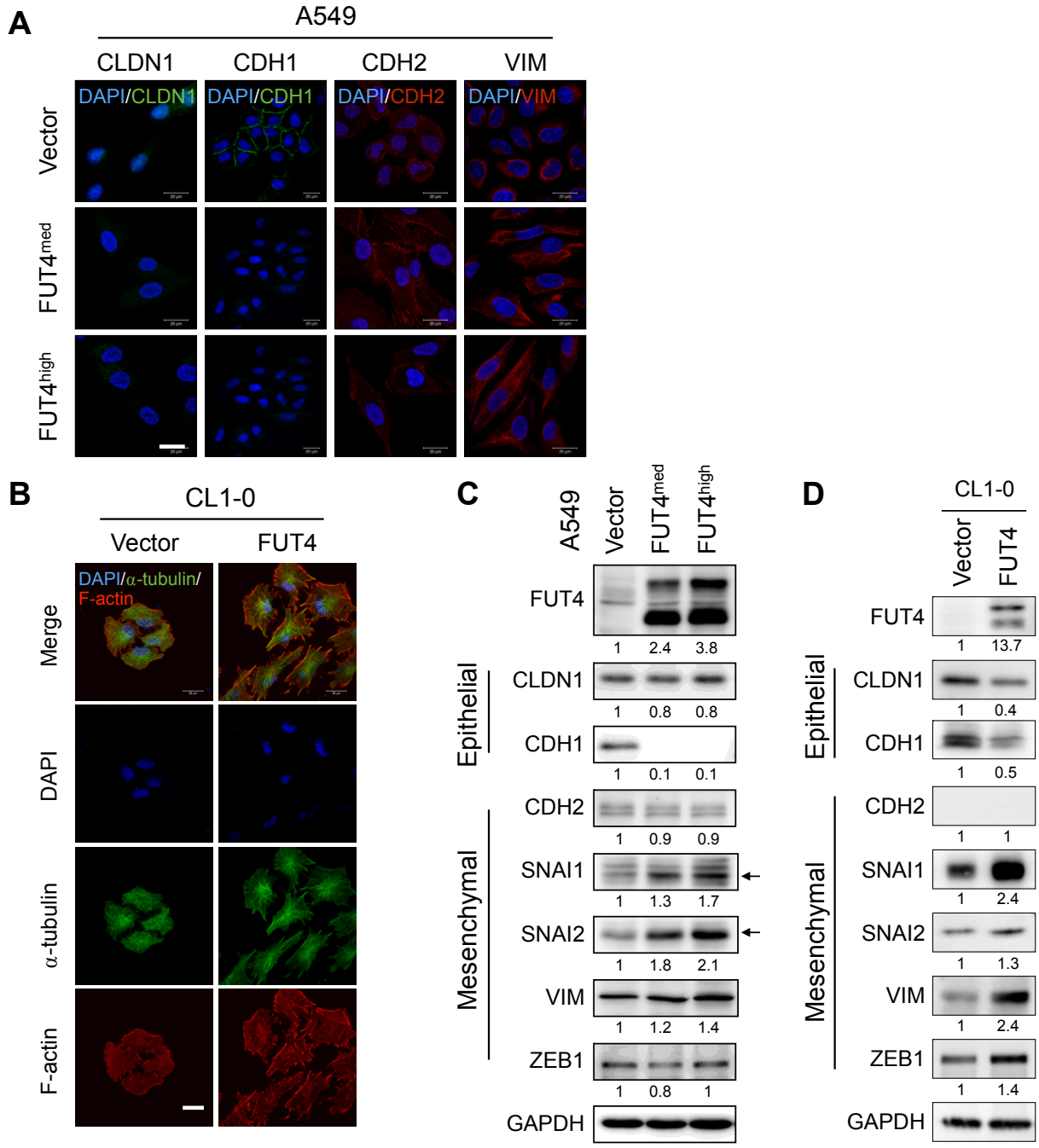
Supplemental Figure 1



Supplemental Figure 1. FUT4 enhances invasiveness and motility and in CL1-0 human lung cancer cells. **A,** Expressions of FUT4 in various human lung cancer cell lines. FUT4 mRNA expressions in human lung cancer cell lines analyzed by quantitative realtime PCR. All experiments were performed in triplicates and presented as mean ± standard errors. ADC: adenocarcinoma. SCC: squamous cell carcinoma. **B and C,** FUT4 mRNA (left panel) and protein expression (right panel) in A549 (**B**) and CL1-0 (**C**)

human lung cancer cell line over-expressed with FUT4 measured by quantitative real-time PCR and western blot analyses. Multiple bands of FUT4 proteins due to post-translational modifications are shown on western blots. GAPDH: loading control. **D and E,** Matrigel-based transwell invasion assays (**D**) and transwell migration assays (**E**) of CL1-0 lung cancer cells over-expressed with FUT4. Relative invasion or migration ability was calculated using the cell number in the lower chamber of transwell system for each sample compared to that of the vector control. Lower panels are representative images of cells with DAPI nuclear stains in the lower chambers. Scale bar, 100 μm. All experiments were performed in triplicates and presented as mean ± standard errors. *p* value was calculated by Mann–Whitney U test * $p < 0.05$ ** $p < 0.01$, *** $p < 0.0005$, **** $p < 0.0001$.

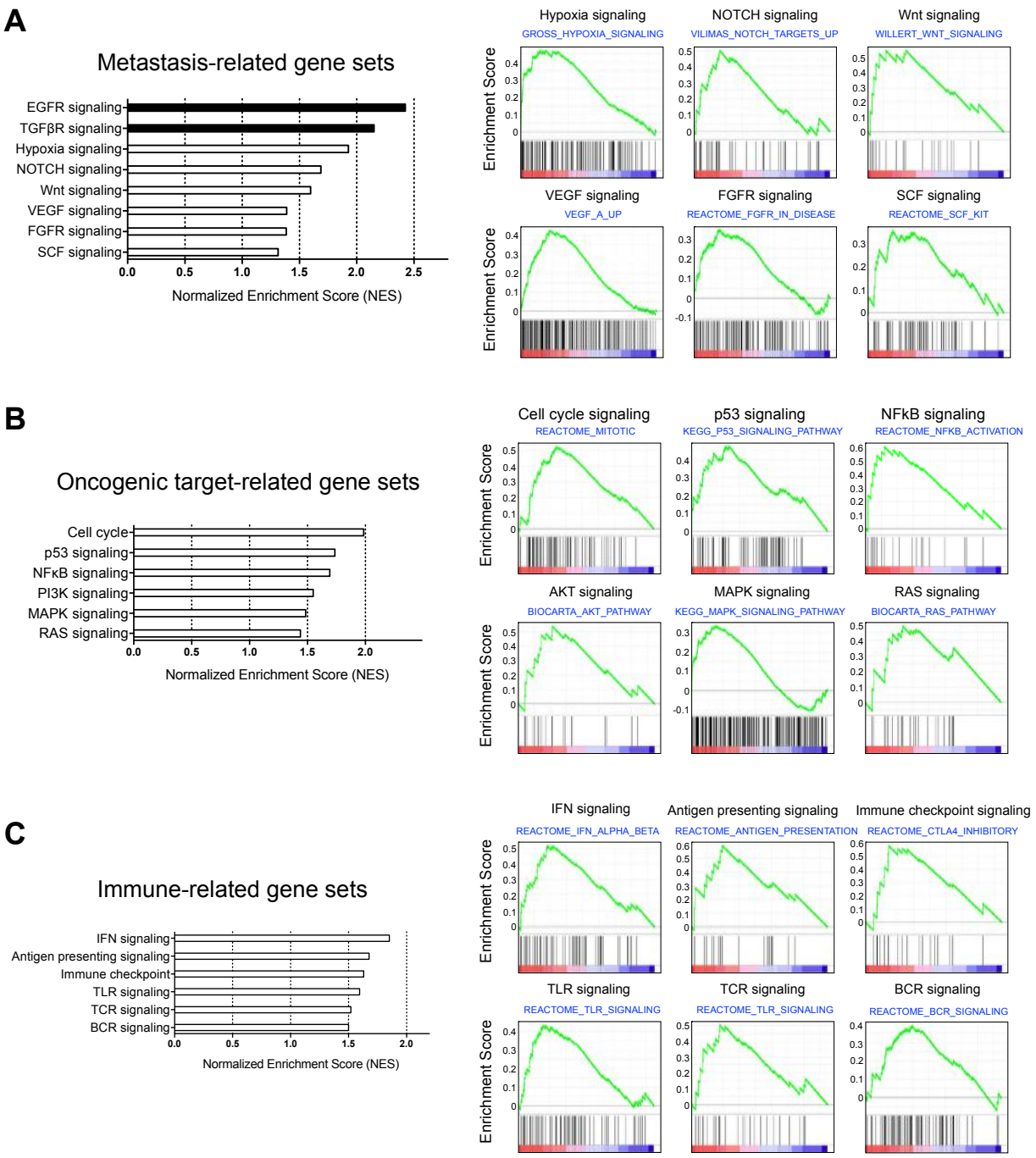
Supplemental Figure 2



Supplemental Figure 2. Immunofluorescent imaging and EMT protein analyses of lung cancer cells with FUT4 over-expression. **A,** Immunofluorescent imaging analysis of EMT markers in A549 lung cancer cells with various levels of FUT4 expression

(A549_vector, A549_FUT4^{med}, and A549_FUT4^{high}). Epithelial markers: CLDN1 (claudin-1) and CDH1 (E-cadherin). Mesenchymal markers: CDH2 (N-cadherin) and VIM (vimentin). Scale bar: 50 μm (first column) and 20 μm. **B,** Immunofluorescent imaging analysis of CL1-0 lung cancer cells with FUT4 over-expression (CL1-0_FUT4) compared to vector control (CL1-0_vector). Blue fluorescence: DAPI (4',6-diamidino-2-phenylindole). Green fluorescence: alpha-tubulin. Red fluorescence: F-actin. Scar bar: 20 μm. **C and D,** Western blot analyses of EMT marker proteins in A549 (**C**) and CL1-0 (**D**) lung cancer cell lines with FUT4 overexpression. Epithelial markers: CLDN1 (claudin-1) and CDH1 (E-cadherin). Mesenchymal markers: CDH2 (N-cadherin), SNAI1 (snail), SNAI2 (slug), VIM (vimentin) and ZEB1 (Zinc Finger E-Box Binding Homeobox 1).

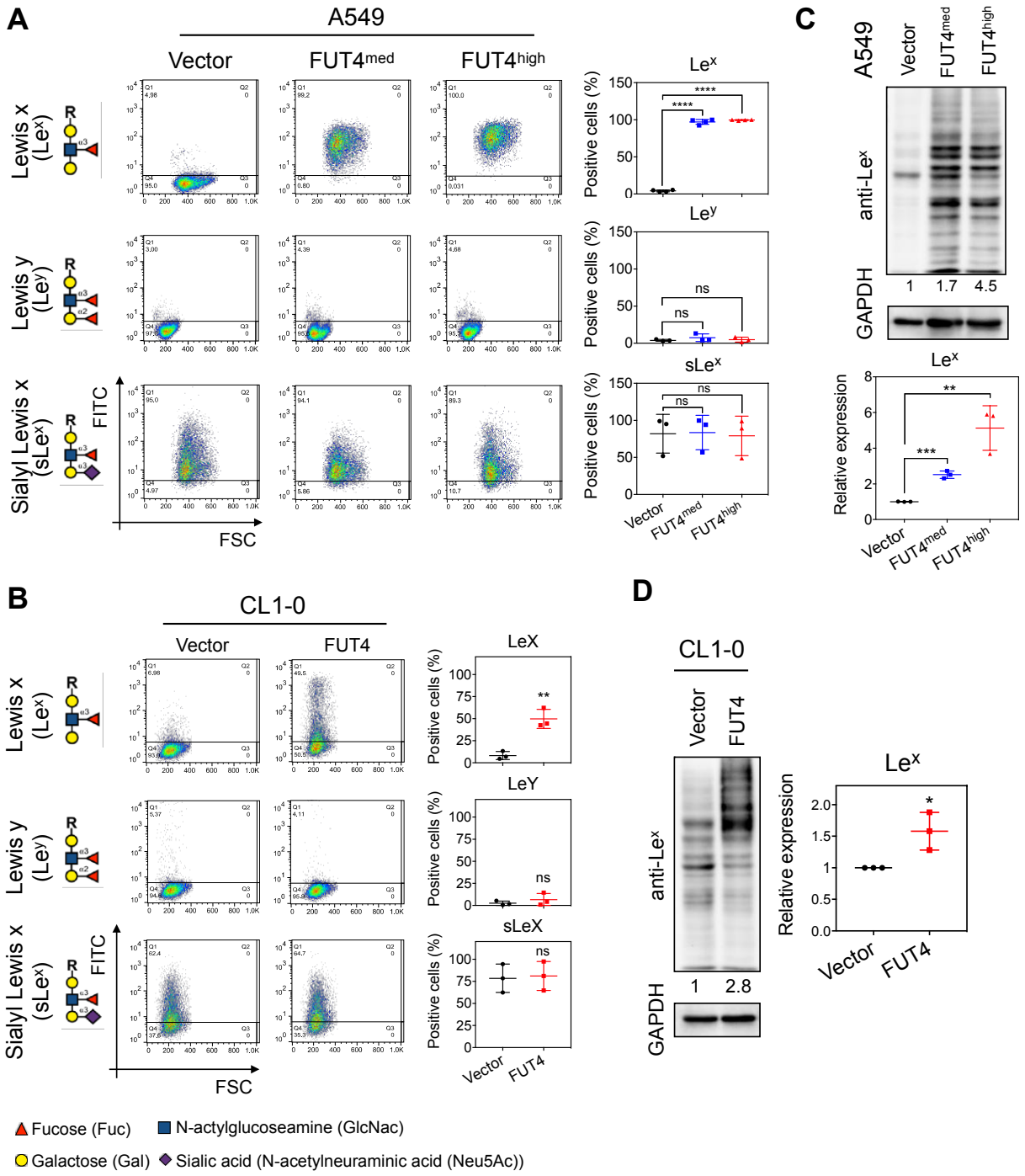
Supplemental Figure 3



Supplemental Figure 3. Top enriched gene sets in A549 and CL1-0 lung cancer cells over-expressed with FUT4. **A,** Top 8 positively-enriched gene sets related to cancer metastasis in A549_FUT4^{high} and CL1-0_FUT4 cells (*left panel*) and GSEA enrichment

plots for individual gene sets (*right panel*). **B,** Top 6 positively-enriched gene sets related to oncogenic targets in A549_FUT4^{high} and CL1-0_FUT4 cells (*left panel*) and GSEA enrichment plots for individual gene sets (*right panel*). **C,** Top 6 positively-enriched gene sets related to immune response in A549_FUT4^{high} and CL1-0_FUT4 cells (*left panel*) and GSEA enrichment plots for individual gene sets (*right panel*)

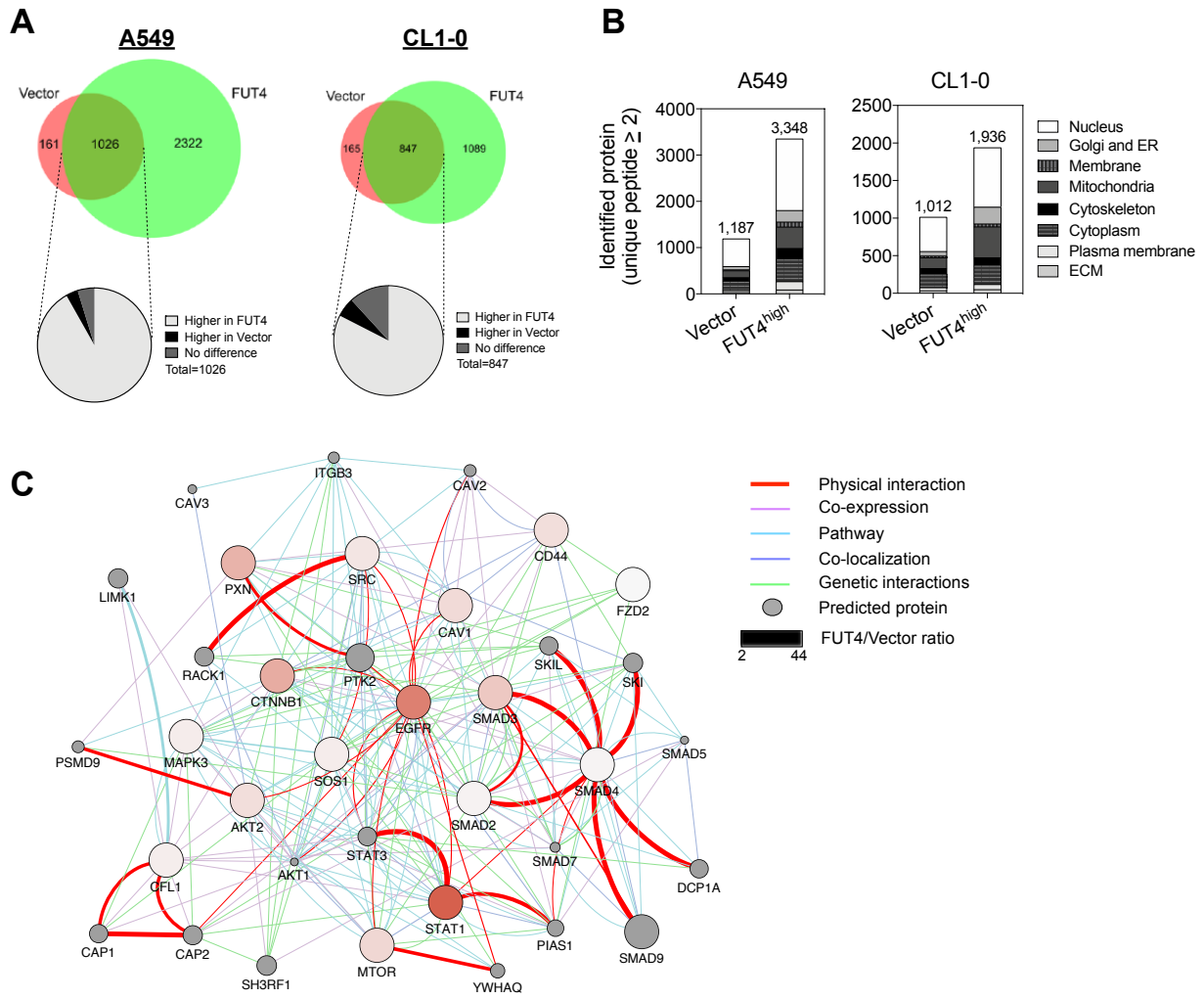
Supplemental Figure 4



Supplemental Figure 4. FUT4 increased expression of Lewis x (Le^x) in human lung cancer cells. **A and B,** Flow cytometric analysis of cell surface glycans, Lewis x (Le^x), Lewis y (Le^y) and sialyl Lewis x (sLe^x), in FUT4-overexpressing A549 (**A**) and CL1-0 (**B**)

lung cancer cells. *Left panels*, representative flow cytometric dot plots Le^y and sLe^x expression are shown. *Right panels*, dot plots showing percentages of cells expressing individual surface glycans in three biological replicates. **C and D,** Western blot analysis of Le^x antigen-reactive fucosylated glycans in A549 (**C**) and CL1-0 (**D**) cells with FUT4 overexpression. Bar graphs showing relative expressions of Le^x-reactive fucosylated glycans in FUT4-overexpressed A549 and CL1-0_FUT4 cells as compared to those of the vector control in three biological replicates.

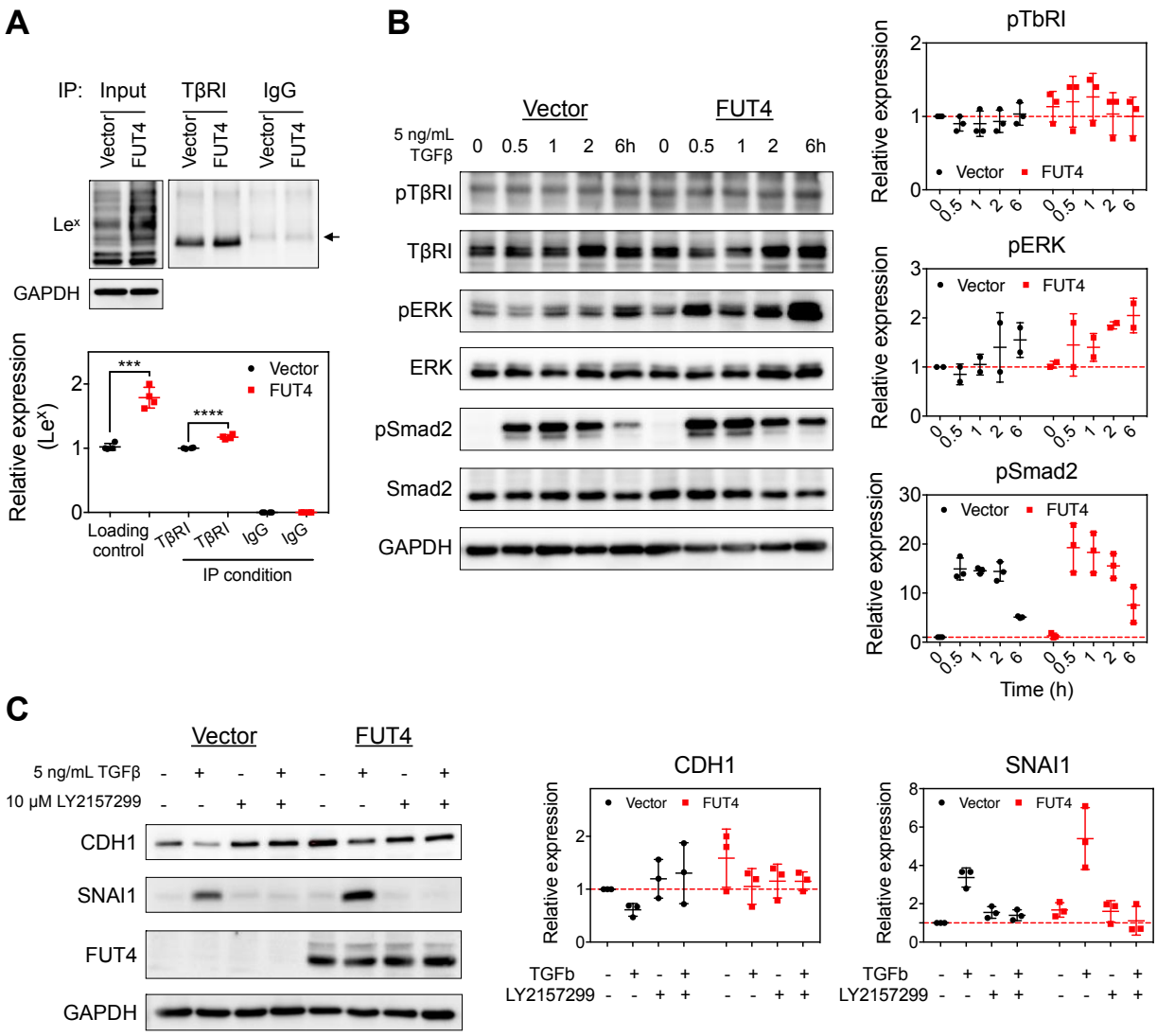
Supplemental Figure 5



Supplemental Figure 5. Acceptor proteins of FUT4-mediated fucosylation identified by anti-Le^x-immunoprecipitation (IP) followed by liquid chromatography-mass spectrometry (LC/MS). **A,** Numbers of identified proteins in FUT4-overexpressing cells vs vector controls in A549 (left panel) and CL1-0 (right panel) cells. Pie charts show percentages of up- or down-regulated proteins commonly identified in both FUT4-overexpressin and vector control cells. **B,** Subcellular localizations of FUT4 acceptor proteins bearing higher levels of Le^x antigens in A549_FUT4^{high} and CL1-0_FUT4 cells compared to their respective vector controls. Y axis denotes numbers of identified protein in each

subcellular location. **C,** Protein networks showing protein-protein interactions between FUT4 acceptor proteins bearing higher levels of Le^x antigens in A549_FUT4^{high} cells relative to A549_vector cells. The data were analyzed by Cytoscape (ver. 3.6.1) using GeneMania application (ver. 3.4.1).

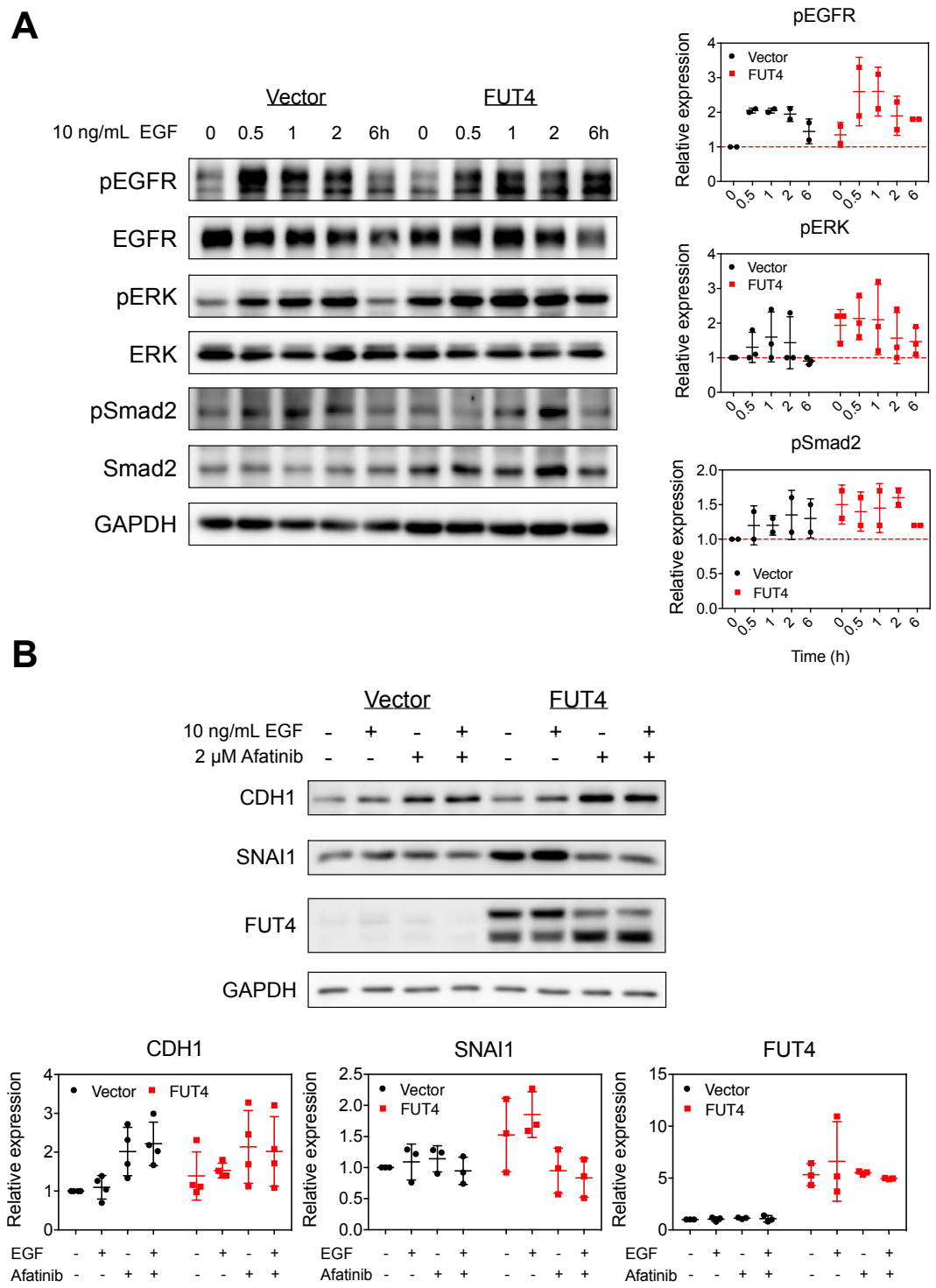
Supplemental Figure 6



Supplemental Figure 6. FUT4 augments transforming growth factor beta (TGF-β) signaling in human lung cancer cells. **A,** Immunoprecipitation-western blot analysis of TGF-β receptor I (TβRI) and Le^x in A549_FUT4high cells as compared to A549_vector cells. Quantifications of Le^x expression in anti-TβRI pull-down lysates are shown in dot plots below. *p* value was calculated by one-way ANOVA with Dunnett's test. * p <0.05 ** p < 0.01, *** p <0.0005, **** p <0.0001. **B,** Western blot analyses of signaling cascade proteins in the TGF-β pathway in A549_vector and A549_FUT4^{high} cells at 0, 0.5, 1, 2, and 6 hrs following the addition of 5 ng/mL TGF-β. Quantifications of signal intensities

on western blots for phospho-TβRI, phospho-ERK and phospho-Smad2 from three biological replicates are graphed on the right. **C,** Western blot analyses of EMT marker proteins— CDH1 (epithelial marker) and SNAI1(mesenchymal marker) in A549_vector and A549_FUT4^{high} cells 24 hours following addition of 5 ng/mL TGF-β and treatment with 10μM LY2157299 (a TGF-β receptor I inhibitor). Dot plots show quantifications of signal intensities on western blots for CDH1 and SNAI1 from three biological replicates. All experiments were performed in three biological replicates and presented as mean ± standard errors.

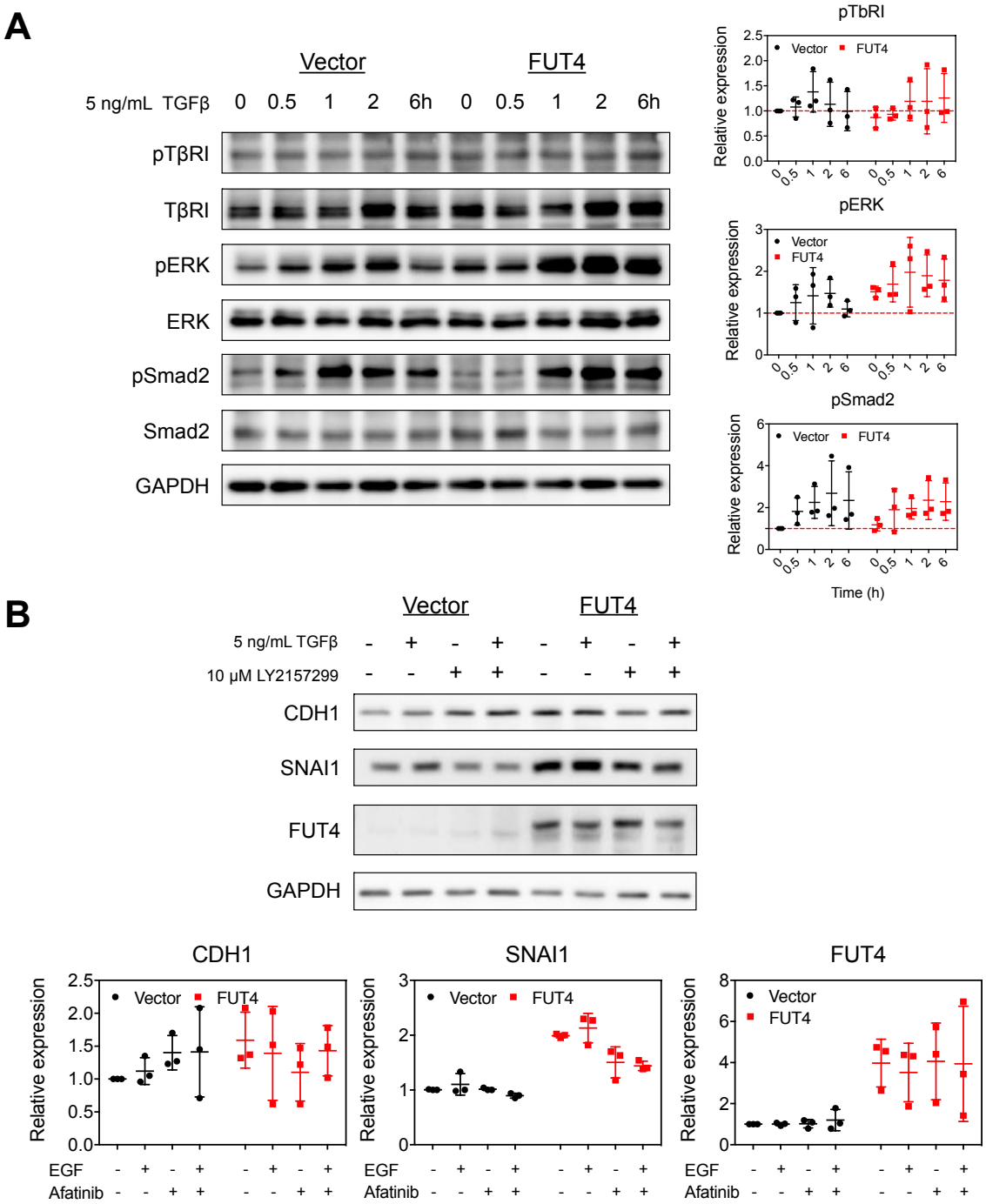
Supplemental Figure 7



Supplemental Figure 7. FUT4 augments epidermal growth factor (EGF) signaling in CL1-0 human lung cancer cells. **A,** Western blot analyses of signaling cascade proteins in the EGFR pathway in CL1-0_vector and CL1-0_FUT4 cells at 0, 0.5, 1, 2, and 6 hrs

following the addition of 10 ng/mL EGF. Quantifications of signal intensities on western blots for phospho-EGFR, phospho-ERK and phospho-Smad2 from three biological replicates are graphed on the right. **B,** Western blot analyses of EMT marker proteins—CDH1 (epithelial marker) and SNAI1(mesenchymal marker) in CL1-0_vector and CL1-0_FUT4 cells 24 hrs following addition of 10 ng/mL EGF and treatment with 2μM afatinib (an EGFR inhibitor). Bar graphs show quantifications of signal intensities on western blots for CDH1, SNAI1 and FUT4 from three biological replicates.

Supplemental Figure 8



Supplemental Figure 8. FUT4 augments transforming growth factor beta (TGF-β) signaling in CL1-0 human lung cancer cells. **A,** Western blot analyses of signaling cascade proteins in the TGF-β pathway in CL1-0_vector and CL1-0_FUT4 cells at 0,

0.5, 1, 2, and 6 hrs following the addition of 5ng/mL TGF-β. Quantifications of signal intensities on western blots for phospho-TβRI, phospho-ERK and phospho-Smad2 from three biological replicates are graphed on the right. **B,** Western blot analyses of EMT marker proteins— CDH1 (epithelial marker) and SNAI1(mesenchymal marker) in A549_vector and A549_FUT4^{high} cells 24 hrs following addition of 5ng/mL TGF-β and treatment with 10μM LY2157299 (a TGF-β receptor I inhibitor). Bar graphs show quantifications of signal intensities on western blots for CDH1, SNAI1 and FUT4 from three biological replicates. All experiments were performed in triplicates and presented as mean ± standard errors.

Supplemental Tables

Supplemental Table 1. Overall survival and expression levels of individual fucosyltransferases in patients with lung squamous cell carcinoma (N=491, High vs. Low expressions, cutoff=median; Cox proportional hazard model)

	Hazard Ratio	Standard Error	*p* value
FUT1	1.10	0.20	0.580
FUT2	1.03	0.21	0.885
FUT3	0.88	0.26	0.660
FUT4	1.23	0.22	0.244
FUT5	0.91	0.16	0.594
FUT6	1.22	0.33	0.459
FUT7	1.03	0.18	0.883
FUT8	0.69	0.13	0.042
FUT9	1.08	0.18	0.652
FUT10	0.70	0.12	0.035
FUT11	1.44	0.26	0.041
POFUT1	0.66	0.12	0.025
POFUT2	0.96	0.17	0.829

Supplemental Table 2. List of antibodies for western blots, immunofluorescence assays and flow cytometric analysis

Antibody	Company	Catalog Number
Rabbit polyclonal anti-human ACTB	GeneTex	Cat# GTX109639
Rabbit polyclonal anti-AKT (phospho)	GeneTex	Cat# GTX59559,
Rabbit polyclonal anti-AKT	Cell Signaling Technology	Cat# 9272
Mouse monoclonal anti-CD15	BioLegend	Cat# 912901
Mouse monoclonal anti-CD15 (HI98)	Santa Cruz Biotechnology	Cat# sc-19649
Rabbit polyclonal anti-β-Catenin	Cell Signaling Technology	Cat# 8480
Rabbit polyclonal anti-Claudin-1	Cell Signaling Technology	Cat# 4933S
Rabbit monoclonal anti-E-Cadherin	Cell Signaling Technology	Cat# 3195
EGFR (phospho)	Cell Signaling Technology	Cat# 2236
EGFR	Cell Signaling Technology	Cat# 2646
anti-ERK (phospho)	Cell Signaling Technology	Cat# 4376
ERK	Cell Signaling Technology	Cat# 4695
Mouse monoclonal anti-FUT4	Abcam	Cat# ab181461
Mouse monoclonal anti-GAPDH	Santa Cruz Biotechnology	Cat# sc-365062
Mouse monoclonal anti-Lewis Y	Abcam	Cat# ab3359
Rabbit monoclonal anti-N-Cadherin	Cell Signaling Technology	Cat# 13116
Rabbit polyclonal anti-TGF beta Receptor I (Ser165)	Abcam	Cat# ab112095
Rabbit polyclonal anti-TGF beta Receptor I	Abcam	Cat# ab31013
Rabbit polyclonal anti-TGF beta Receptor II (Ser225/250)	Abcam	Cat# ab111564
Rabbit polyclonal anti-TGF beta Receptor II	Abcam	Cat# ab186838
Mouse monoclonal anti-Sialyl Lewis X	BioLegend	Cat# 368102
Rabbit polyclonal anti-Slug	Cell Signaling Technology	Cat# 9585P
Rabbit polyclonal anti-Smad2 (Ser465/467)	Cell Signaling Technology	Cat# 3101
Mouse monoclonal anti-Smad2	Santa Cruz Biotechnology	Cat# sc-101153
Rabbit monoclonal anti-Smad3 (Ser423/425)	Cell Signaling Technology	Cat# 9520
Mouse monoclonal anti-Smad3	Santa Cruz Biotechnology	Cat# sc-101154
Rabbit polyclonal anti-Snail	Cell Signaling Technology	Cat# 3879S
Rabbit monoclonal anti-TCF8/ZEB1	Cell Signaling Technology	Cat# 3396
Mouse monoclonal anti-alpha Tubulin	GeneTex	Cat# GTX628802
Rabbit monoclonal anti-Vimentin	Cell Signaling Technology	Cat# 5741
Rabbit monoclonal anti-ZO-1	Cell Signaling Technology	Cat# 8193S

Supplemental Table 3: Primers, shRNA sequence and constructs.

	Gene	Direction	Primer sequence
mRNA qPCR	*FUT4*	Forward	5'-GGTGCCCGAAATTGGGCTCCTGCACAC-3'
		Reverse	5'-CAACGCGTTGCGCCAGAGCTTCTC-3'
	TBP	Forward	5'- CACGAACCACGGCACTGATT-3'
		Reverse	5'- TTTTCTTGCTGCCAGTCTGGA C-3'

Gene construct	Plasmid	Sequence
FUT4 sh#751	pZIP-mCMV-ZsGreen-Puro	TGCTGTTGACAGTGAGCGACCTGGCAAGTAACCTC TTCAATAGTGAAGCCACAGATGTATTGAAGAGGTT ACTTGCCAGGCT GCCTACTGCCTCGGA
FUT4 sh#753	pZIP-mCMV-ZsGreen-Puro	TGCTGTTGACAGTGAGCGATCGGACGTCTTTGTGC CTTATTAGTGAAGCCACAGATGTAATAAGGCACAA AGACGTCCGAGTGCCTACTGCCTCGGA
FUT4 sh#792	pZIP-mCMV-ZsGreen-Puro	TGCTGTTGACAGTGAGCGATACTACCACCAACTGA GCCAATAGTGAAGCCACAGATGTATTGGCTCAGTT GGTGGTAGTAGTGCCTACTGCCTCGGA

pcDNA3.1(+)-FUT4
5'-ATGAGGCGCTTGTGGGGCGCGGCCCGGAAGCCCTCGGGCGCGGGCTGGGAG AAGGAGTGGGCGGAGGCGCCGCAGGAGGCTCCCGGGGCCTGGTCGGGCCGGC TGGGCCCCGGGCGCAGTGGAAGAAAGGGACGGGCGGTGCCCGGTTGGGCGTC CTGGCCAGCTCACCTTGCCCTGGCGGCTCGCCCCGCCCGGCACTTGGGAGGAGC AGGGCAGGGCCCGCGGCCTTTGCATTCTGGGACCGCCCCCTTCCATTCCCGGGC CAGCGGCGAGCGGCAGCGACGGCTGGAGCCGCAGCTACAGCATGAGAGCCGGT GCCGCTCCTCCACGCCTGCGGACGCGTGGCGAGCGGAGGCAGCGCTGCCTGTTC GCGCCATGGGGGCACCGTGGGGCTCGCCGACGGCGGCGGCGGGCGGGCGGCGC GGGTGGCGCCGAGGCCGGGGGCTGCCATGGACCGTCTGTGTGCTGGCGGCCGCC GGCTTGACGTGTACGGCGCTGATCACCTACGCTTGCTGGGGGCAGCTGCCGCCG CTGCCCTGGGCGTCGCCAACCCCGTCGCGACCGGTGGGCGTGCTGCTGTGGTGG GAGCCCTTCGGGGGGCGCGATAGCGCCCCGAGGCCGCCCCCTGACTGCCGGCTG CGCTTCAACATCAGCGGCTGCCGCCTGCTCACCGACCGCGCGTCCTACGGAGAG

GCTCAGGCCGTGCTTTTCCACCACCGCGACCTCGTGAAGGGGCCCCCCGACTGG
CCCCCGCCCTGGGGCATCCAGGCGCACACTGCCGAGGAGGTGGATCTGCGCGTG
TTGGACTACGAGGAGGCAGCGGCGGCGGCAGAAGCCCTGGCGACCTCCAGCCC
CAGGCCCCCGGGCCAGCGCTGGGTTTGGATGAACTTCGAGTCGCCCTCGCACTC
CCCGGGGCTGCGAAGCCTGGCAAGTAACCTCTTCAACTGGACGCTCTCCTACCG
GGCGGACTCGGACGTCTTTGTGCCTTATGGCTACCTCTACCCCAGAAGCCACCC
CGGCGACCCGCCCTCAGGCCTGGCCCCGCCACTGTCCAGGAAACAGGGGCTGG
TGGCATGGGTGGTGAGCCACTGGGACGAGCGCCAGGCCCGGGTCCGCTACTAC
CACCAACTGAGCCAACATGTGACCGTGGACGTGTTCGGCCGGGGCGGGCCGGG
GCAGCCGGTGCCCGAAATTGGGCTCCTGCACACAGTGGCCCGCTACAAGTTCT
ACCTGGCTTTCGAGAACTCGCAGCACCTGGATTATATCACCGAGAAGCTCTGGC
GCAACGCGTTGCTCGCTGGGGCGGTGCCGGTGGTGCTGGGCCCAGACCGTGCCA
ACTACGAGCGCTTTGTGCCCCGCGGCGCCTTCATCCACGTGGACGACTTCCCAA
GTGCCTCCTCCCTGGCCTCGTACCTGCTTTTCCTCGACCGCAACCCCGCGGTCTA
TCGCCGCTACTTCCACTGGCGCCGGAGCTACGCTGTCCACATCACCTCCTTCTGG
GACGAGCCTTGGTGCCGGGTGTGCCAGGCTGTACAGAGGGCTGGGGACCGGCCC
AAGAGCATACGGAACTTGGCCAGCTGGTTCGAGCGGTGA-3’

*Construct sequences were confirmed by standard sequencing.