ONLINE DATA SUPPLEMENT

GENE EXPRESSION PROFILE OF EPITHELIAL-MESENCHYMAL TRANSITION IN TUMORS OF PATIENTS WITH NSCLC: THE INFLUENCE OF COPD

Yingchen Xia*, Jianhua Zha, Víctor Curull, Albert Sánchez-Font, Maria Guitart, Alberto Rodríguez-Fuster, Rafael Aguiló, and Esther Barreiro

METHODS

Study population

All patients were prospectively and consecutively recruited from the Lung Cancer Clinic at Hospital del Mar (Barcelona, Spain). All the patients were part of the *Lung Cancer Mar Cohort*. For this observational investigation, 50 patients with LC were consecutively recruited during the years 2018-2020. Patients were further subdivided according to the presence or absence of COPD: n=30 patients with COPD (LC-COPD group) and n=20 patients with no COPD (LC control group).

In all cases, pre-operative staging was performed using chest and upper abdomen Computed Tomography (CT) scan and fluoro-deoxy-glucose positron emission tomography/computed tomography (PET) body-scan. When suspected mediastinal lymph-node involvement, a fiberoptic bronchoscopy with endo-bronchial ultra-sound (EBUS) and trans-tracheal biopsy of the suspected nodes were performed. In case of negative results, a surgical exploration of the mediastinum: cervical video-assisted mediastinal lymphadenectomy (VAMLA) and/or anterior mediastinotomy were performed, the latter depending on the location of the suspected nodes. Notwithstanding, in all surgical cases, intra-operative systematic hilar and mediastinal lymphadenectomy (at least, ipsilateral paratracheal, subcarinal, and ipsilateral pulmonary ligament) was performed as previously recommended [1,2].

Standard clinical guidelines were used to establish the selection of patients and contraindications for thoracic surgery as previously described [2]. Decisions on the best therapeutic approach were always made during the weekly meetings of the Multidisciplinary Lung Cancer Committee. Candidates for tumor resection underwent pulmonary surgery (video-assisted thoracoscopic surgery, VATS) prior to administration of any sort of adjuvant therapy. LC diagnosis and staging were

established by histological confirmation and classified according to currently available guidelines for the diagnosis and management of LC [3,4]. TNM (tumor, node, and metastasis) staging was defined as stated in the 8th edition Lung Cancer Stage Classification [5].

Exclusion criteria were: small cell lung cancer (SCLC), chronic cardiovascular disease, metabolic or clot system disorders, signs of severe inflammation and/or bronchial infection (bronchoscopy), current or recent invasive mechanical ventilation, or long-term oxygen therapy. The presence/absence of these diseases was confirmed using standard clinical tests: exercise capacity electrocardiogram, clinical examination, blood tests, bronchoscopy and echocardiography.

This was a prospective controlled clinical investigation, in which the World Medical Association guidelines for research on human beings (Seventh revision of Declaration of Helsinki, Fortaleza, Brazil, 2013) were followed. The institutional Ethics Committee on Human investigations (protocol #2008/3390 /I, February 4th 2008, at Hospital del Mar-IMIM, Barcelona, Spain) approved all the procedures and study protocol. All patients invited to participate in the study signed their written informed consent.

Clinical assessment

In all patients, lung function parameters were assessed following standard procedures. Diagnosis and severity of patients with COPD were determined according to currently available guidelines [5,6]. Nutritional evaluation included the assessment of body mass index (BMI) and nutritional blood parameters from all the patients.

Sample collection and preservation

Lung samples were obtained from tumors following standard technical procedures during VATS in the surgery room. The fresh samples were carefully transported to the Pathology Department, located at a very short distance (less than 5 minutes). In all

patients, the expert pulmonary pathologists selected tumor lung specimens of approximately 10x10 mm^2 area from the fresh samples. For all the recruited patients, fragments of tumor specimens were immediately snap-frozen and stored at -80ºC until further use in the laboratory experiments. As patients were recruited consecutively, sample collection procedures were equally applied to all. In the same vein, non-tumor specimens were also collected as far distal to the tumor margins as possible (average >7 cm). Fragments of non-tumor specimens were also immediately snap-frozen and stored at -80ºC until further use. The pathologists were not aware of the presence or absence of COPD in the study patients. Thus, no differences in sample preservation were applied between LC and LC-COPD patients. Driver mutations were analyzed by the pathologists on the tumor specimens of all the study patients. Pilot experiments were conducted in order to test whether any differences were observed in non-tumor samples between LC-COPD and LC control patients. No significant differences were observed in non-tumor samples between LC-COPD and LC patients. Thus, for the sake of clarity the results obtained from the tumor samples in both study groups are the ones depicted in the figures.

Biological experiments

Figure 1 illustrates the flow of the signaling markers analyzed in the investigation. The sequence of experiments and results description follow this chart. *SMAD4* can form homologous complexes by itself or heterologous complexes with other activating SMAD family members, translocate to the nucleus, and act synergistically with other transcription factors to affect the EMT pathway [7,8]. *ZEB2* inhibits the expression of E-cadherin, activates matrix metalloproteinases (MMP), induces the occurrence of EMT, and promotes cell proliferation and migration [9,10]. In EMT pathway, increased expression of *CDH2*, a downstream protein of Twist1, may render tumor cells more

aggressive [11,12]. The most important function of Snail1 is to induce EMT in tumor cells by inhibiting the transcription of E-cadherin, while favoring the expression of *VIM* [13]. Cells with low expression of *ICAM1* or *MMP9* will enter EMT [14,15].

RNA isolation. RNA was isolated from 30-50 microgram frozen tumor samples using 500 microL TRIzol reagent (Cat. 15596026, Thermo Fisher Scientific, Waltham, MA, USA). After incubation of the samples at room temperature for 10 minutes to achieve complete dissociation of nucleoprotein complexes, 200 microL chloroform were added, and samples were then centrifuged at 13,500 rpm at 4ºC for 15 minutes. The aqueous phase was recovered and the RNA was precipitated with 600 microL isopropanol. Subsequently, samples were incubated at 4ºC for 30 minutes and were then cooled down to -20ºC overnight. After thawing the samples at room temperature, they were centrifuged at 13,500 rpm at 4ºC for 10 minutes, and the supernatant was removed. The remaining pellet was then washed using one mL solution of 75% ethanol to be subsequently centrifuged at 9,000 rpm at 4ºC for five minutes. The RNA containing pellet was air-dried for 30 minutes and was then dissolved in 20 microL RNase-free water. To assess the quality and purity of the isolated RNA, concentrations of total RNA were determined using NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions.

RNA reverse transcription (RT). Invitrogen® cDNA Synthesis Kit (Cat.18018044, Thermo Fisher Scientific, Carlsbad, CA, USA) was used to prepare cDNA templates following the manufacturer's instructions. Total RNA isolated samples were manipulated to add oligo (dT), dNTP mix, DTT, reverse transcriptase and buffer. Initially, 20 microL reaction mix (4 microL buffer, 1 microL 0.1 M DTT, 1 microL reverse transcriptase and 1 microL oligo (DT), 1 microL dNTP mix, 12 microL RNase-free water) was mixed with 100 nanogram of each sample for all the samples and

singlets. The mixture was then incubated in a thermal cycler (Biometra Tone 96, Analytik Jena, Jena, Germany) to perform the synthesis reaction at 50ºC for 60 minutes. This step was followed by incubation at 70ºC for 15 minutes to stop the reaction and samples were finally kept at -80ºC up until the performance of the real-time polymerase chain reaction (PCR) procedures.

Quantitative real time-PCR amplification (qRT-PCR). Real-time PCR was performed using commercially gene expression assays for human studies. The probes corresponding to the following genes involved in signaling of EMT were detected: SMAD family member 3 (*SMAD3*, Hs00969210_m1, Life Technologies), SMAD family member 4 (*SMAD4*, Hs00929647_m1, Life Technologies), zinc finger E-box binding homeobox 2 (*ZEB2*, Hs00207691_m1, Life Technologies), twist family transcription factor 1 (*TWIST1*, Hs00361186_m1, Life Technologies), snail family transcriptional repressor 1 (*SNAIL1*, Hs00195591_m1, Life Technologies), intercellular adhesion molecule 1 (*ICAM1*, Hs00164932_m1, Life Technologies), vimentin (*VIM*, Hs00185584_m1, Life Technologies), cadherin-2 (*CDH2*, Hs00983056_m1, Life Technologies), matrix metallopeptidase 1 (*MMP-1*, Hs00899658_m1, Life Technologies), matrix metallopeptidase 9 (*MMP-9*, Hs00957562_m1, Life Technologies) and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*, Hs99999905_m1, Life Technologies). As previously described [16], *GAPDH* was used as an exogenous control in order to normalize the RNA amplification in all the study samples. Samples from both groups of patients were always run simultaneously in duplicates for comparison purposes. Briefly, 4.5 microL of the resulting cDNA samples were mixed with 0.5 microL of each specific probe and 5 microL Taqman universal master mix no AmpErase UNG^{TM} (Cat. 4440044, Thermo Fisher Scientific). The samples were run in a thermal cycler (QuantStudio Real-time PCR system, Thermo

Fisher Scientific). The first step was the enzyme activation, achieved at 95ºC for 20 seconds and was followed by 40 combined cycles of denaturation (95ºC for one second), and final annealing (60ºC for 20 seconds) as also previously described in studies assessing gene expression of EMT markers [16,17]. As duplicates from all the patient samples were run, the average value was calculated for each marker in each patient. Two replicas of these experiments were performed for all the target genes. No expression could be detected for a few of the patients despite the sample duplicates and the replicas of the experiments. Patients with no expression have been represented as no expression samples in a separate table in the results section. Appropriate statistical analyses have been conducted in order to assess potential differences in the number of no-expression samples between the two study groups. The results obtained from the experiments were collected and analyzed using the Expression Suite Software version 1.0.4 from Applied Biosystems (Thermo Fisher Scientific), in which the comparative CT method (2-ΔΔCT) for relative quantification was used as also previously described [18].

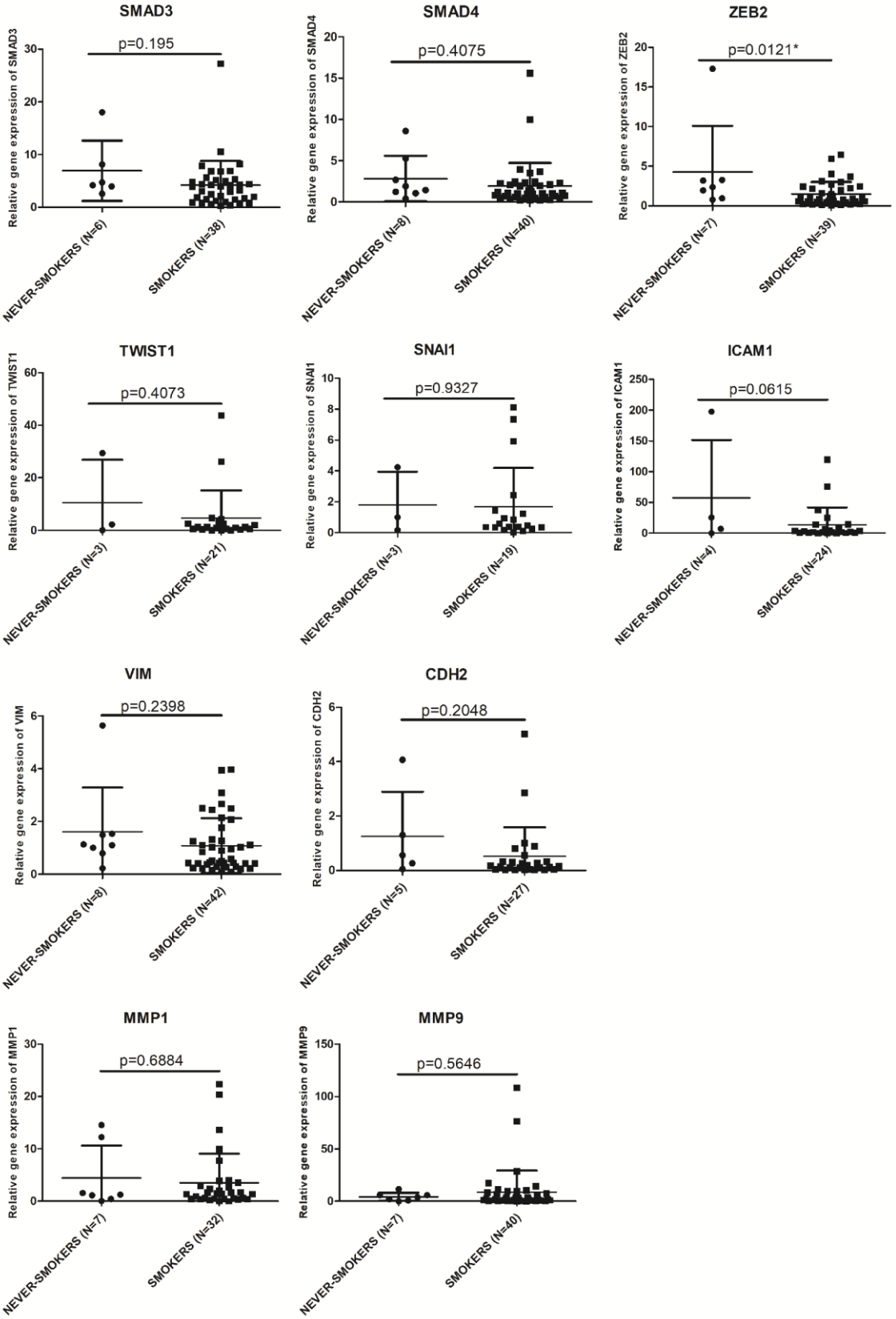
Statistical analysis

Sample size was calculated on the basis of four target markers (CDH2, ZEB2, SMAD4, and SMAD3). Accepting an alpha risk of 0.05 and a beta risk of 0.2 in a two-sided test: 16, 16, 13, and 16 subjects were required in each group to identify a statistically significant difference greater than or equal to 0.5, 2, 1.7, and 3 units in the mean value and a standard deviation of 0.5, 2, 1.5, and 3 in the expression of the genes *CDH2*, ZEB2, SMAD4, and SMAD3, respectively. As 30 and 20 patients were included in LC-COPD and LC control group of patients, the total number of patients was sufficient to attain an 80% power.

The normality of the study variables was examined using the Shapiro-Wilk test. For an initial descriptive analysis of clinical parameters, qualitative variables were described as frequencies (number and percentage) and quantitative variables as mean and standard deviation. Differences between LC and LC-COPD and between smokers and never-smokers as a whole (with no distinction between COPD and non-COPD patients) were assessed using Student's T-test. Chi-square test was used to assess differences between the two groups for the categorical variables including the driver mutations and expression or no gene expression for the markers: *SMAD3*, *SMAD4*, *ZEB2*, *TWIST1*, *SNAI1*, *ICAM1*, *VIM* and *CDH2*, *MMP1*, *MMP9*.

Potential correlations between clinical and biological variables were explored using the Pearson's correlation coefficient. All the statistical analyses were performed using the Statistical Package for the Social Sciences (Portable SPSS, PASW statistics 22.0 version for Windows, SPSS Inc., Chicago, IL, USA). Correlations are displayed in graphical correlation matrixes, obtained from R package corrplot (https://cran.r-project.org/web/packages/corrplot/index.html), in different colors: blue for positive correlations and red for negative ones. Statistical significance was established at $p \leq 0.05$ for all the comparisons.

176 **Figure S1**



177

178 **Figure S1**. Gene expression of EMT markers (*SMAD3*, *SMAD4*, *ZEB2*, *TWIST1*,
179 *SNAI1*, *ICAM1*, *VIM*, *CDH2*, *MMP1*, *MMP9*) between never smokers and smokers in

the overall patients (n=50). Student's t test was used to assess potential significant differences, *P < .05.

Figure S2

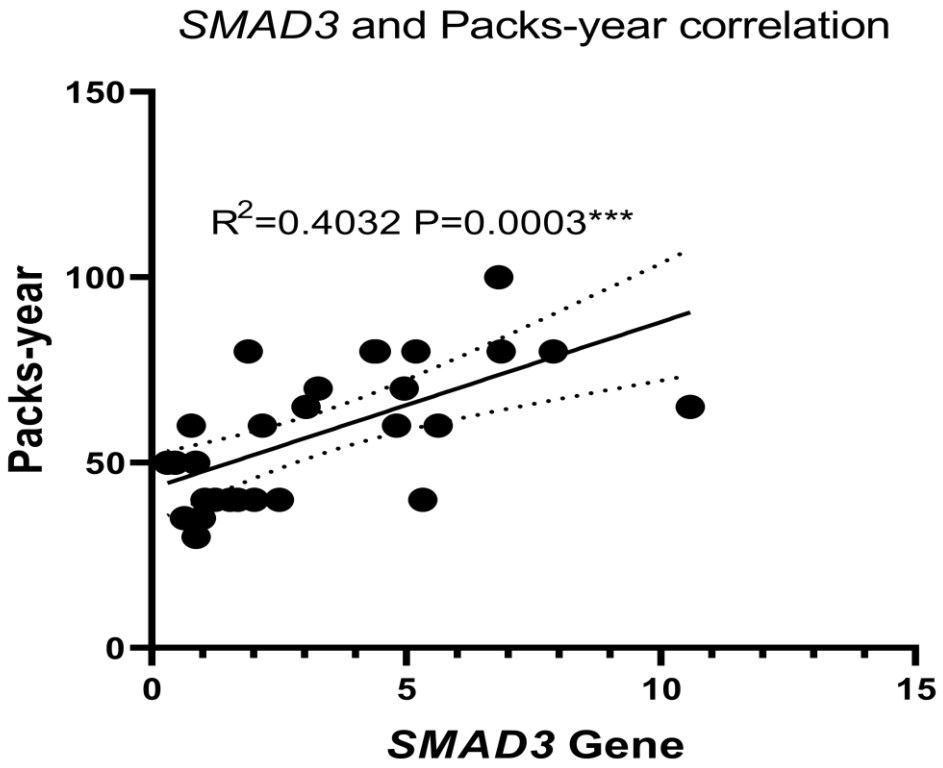


Figure S2. Linear regression plot between *SMAD3* expression and the number of packs-years among LC-COPD patients. SMAD3 gene expression for each patient is represented in the X-axis, while the number of packs-year is represented in the Y-axis. Twenty-eight patients are represented, since two patients did not show any expression of *SMAD3* in their lung tumors.

REFERENCES

1. Justyna W. Epithelial-Mesenchymal Plasticity: A central regulator of cancer progression. *Physiol. Behav.* 2017; 176: 139–148.

2. Armstrong P, Congleton J, Fountain SW, et al. Guidelines on the selection of patients with lung cancer for surgery. *Thorax* 2001; 56: 89–108.

3. De leyn P, Dooms C, Kuzdzal J, et al. Revised ests guidelines for preoperative mediastinal lymph node staging for non-small-cell lung cancer. *Eur. J. Cardio-thoracic Surg.* 2014; 45: 787–798.

4. Slatore CG, Horeweg N, Jett JR, et al. An Official American Thoracic Society research statement: A research framework for pulmonary nodule evaluation and management. *Am. J. Respir. Crit. Care Med.* 2015; 192: 500–514.

5. Detterbeck FC, Boffa DJ, Kim AW, et al. The Eighth Edition Lung Cancer Stage Classification. *Chest* Elsevier; 2017; 151: 193–203.

6. Soler-Cataluña JJ, Piñera P, Trigueros JA, et al. Spanish COPD Guidelines (GesEPOC) 2021 Update Diagnosis and Treatment of COPD Exacerbation Syndrome. *Arch. Bronconeumol. (English Ed.* Elsevier; 2021; .

7. Dardare J, Witz A, Merlin JL, et al. SMAD4 and the TGFB pathway in patients with pancreatic ductal adenocarcinoma. *Int. J. Mol. Sci.* 2020; 21.

8. Cicenas J, Kvederaviciute K, Meskinyte I, et al. KRAS, TP53, CDKN2A, SMAD4, BRCA1, and BRCA2 mutations in pancreatic cancer. *Cancers (Basel).* 2017; 9.

9. Sánchez-Tilló E, Siles L, de Barrios O, et al. Expanding roles of ZEB factors in tumorigenesis and tumor progression. *Am. J. Cancer Res.* Am J Cancer Res; 2011; 1: 897–912.

10. Vandamme N, Denecker G, Bruneel K, et al. The EMT Transcription Factor ZEB2 Promotes Proliferation of Primary and Metastatic Melanoma While Suppressing an Invasive, Mesenchymal-Like Phenotype. *Cancer Res.* Cancer Res; 2020; 80: 2983–2995.

11. Na YR, Lee JS, Lee SJ, et al. Interleukin-6-induced Twist and N-cadherin enhance melanoma cell metastasis. *Melanoma Res.* Melanoma Res; 2013; 23: 434–443.

12. Chaves LP, Melo CM, Saggioro FP, et al. Epithelial-Mesenchymal Transition Signaling and Prostate Cancer Stem Cells: Emerging Biomarkers and Opportunities for Precision Therapeutics. *Genes (Basel).* Genes (Basel); 2021; 12.

13. Tang X, Sui X, Weng L, et al. SNAIL1: Linking Tumor Metastasis to Immune Evasion. *Front. Immunol.* Frontiers Media SA; 2021; 12.

14. Santarosa M, Maestro R. The Autophagic Route of E-Cadherin and Cell Adhesion Molecules in Cancer Progression. *Cancers (Basel).* Cancers (Basel); 2021; 13: 6328.

15. Nowak E, Bednarek I. Aspects of the Epigenetic Regulation of EMT Related to Cancer Metastasis. *Cells* Cells; 2021; 10: 3435.

16. Paradowska-Gorycka A, Wajda A, Romanowska-Próchnicka K, et al. Th17/Treg-Related Transcriptional Factor Expression and Cytokine Profile in Patients With Rheumatoid Arthritis. *Front. Immunol.* Frontiers Media S.A.; 2020; 11.

17. Ma L, Andrieu T, McKinnon B, et al. Epithelial-to-mesenchymal transition contributes to the downregulation of progesterone receptor expression in endometriosis lesions. *J. Steroid Biochem. Mol. Biol.* Elsevier Ltd; 2021; 212: 105943.

18. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* Academic Press Inc.; 2001; 25: 402–408.