JBUON 2018; 23(6): 1699-1710
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
E-mail: editorial_office@jbuon.com

ORIGINAL ARTICLE

Development of novel real-time RT-qPCR methodologies for quantification of the *COL11A1* mRNA general and C transcripts and evaluation in non-small cell lung cancer specimens

Tatiana Rizou[1]*, Fotis Perlikos[2,3]*, Maria Lagiou[1], Makrina Karaglani[1], Sotirios Nikolopoulos[4], Ioannis Toumpoulis[2], Christos Kroupis[1]

[1]Department of Clinical Biochemistry, Attikon University General Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece; [2]Department of Cardiovascular Surgery, Attikon University General Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece; [3]First Department of Critical Care and Pulmonary Services, Evangelismos General Hospital, Athens, Greece; [4]Center for Molecular Analysis and Research S.A., Agrinio, Greece

*These authors contributed equally to this study.

Summary

Purpose: *The purpose of this study was the development of new quantitative methodologies for the general (total) COL11A1 gene and the C transcript (RT-qPCR methods for A and E transcripts have already been developed by our group previously), the quantification of all COL11A1 transcripts and the investigation for the first time of their potential association with histopathological prognostic factors in lung cancer.*

Methods: *Real-time RT-qPCR methodologies with dual hybridization probes were developed on the Light Cycler 1.5 platform (Roche,Germany). All COL11A1 transcripts were measured in 27 cDNA lung tissue specimens in a blinded fashion (8 control and 19 non-small cell lung cancer (NSCLC) tissues with known histopathological data). Statistical analysis was performed with the IBM SPSS program.*

Results: *The novel real-time RT-qPCR methodologies were appropriately validated. All 19 NSCLC samples were positive for the general COL11A1 transcript (range 11.2-1198.0 copies/μg total RNA, while 5 out of 8 control samples were negative: mean values were also statistically significantly different ($p<0.001$). In 4 tumor samples (21%), no specific COL11A1 transcript was detected. Transcript C was detected in only 3 tumor samples. Regarding transcripts A and E, 13 out of 19 tumor samples were positive for either one (68%) and 11 for both (58%).*

Conclusions: *No other statistically significant association of the specific transcripts with histopathological data was observed, most probably due to the limited number of samples. As the number of general COL11A1 transcripts/μg exceeds the sum of A+E+C transcripts in all samples, there is opportunity for discovery and identification of other transcripts as well.*

Key words: *COL11A1, lung cancer, mRNA variant transcripts, real-time RT-qPCR*

Abbreviations: *BAC: bronchio-alveolar carcinoma; bp: base pair; CAFs: cancer-associated fibroblasts; Cq: Quantification cycle; CV: Coefficient of variation; ECM: extracellular matrix; NGS: Next Generation Sequencing; NSCLC: non-small cell lung cancer; RT-qPCR: Reverse Transcription quantitative Polymerase Chain Reaction; SD: Standard Deviation; Tm: Melting point temperature*

Introduction

Lung cancer, and especially NSCLC, remains one of the most frequent and lethal malignancies worldwide, despite efforts in improving diagnostic and therapeutic strategies in recent years [1]

Correspondence to: Fotis Perlikos, MD, MSc. National and Kapodistrian University of Athens, Medical School, Attikon University General Hospital, Rimini 1, 12462 Haidari, Greece.
Tel: +30 6947 590763, E-mail: perlykos@yahoo.com
Received: 25/03/2018; Accepted: 21/04/2018

Lung cancer is the leading cause of cancer death among males in both more and less developed countries, and has surpassed breast cancer as the leading cause of cancer death among females in more developed countries [2]. Despite advances in surgical and chemotherapy treatments, lung cancer patients suffer from disease recurrence in a large proportion, affecting quality of life and ultimately survival. Around 90% of all cancer deaths are the result of metastases, rather than of the primary tumors [3]. The majority of patients with lung cancer are diagnosed at late stages, when already beyond the possibility of treatment by surgery or radiotherapy, because of local invasion or distant metastases. These facts highlight the need for further understanding of the underlying pathways and molecular mechanisms of lung cancer metastasis, in order to develop novel therapeutic approaches [4].

Tumor invasion and metastasis are a hallmark of the process of carcinogenesis and a major cause of treatment failure and patient death [5]. The basic stages of invasion and metastasis involve the detachment of tumor cells from the extracellular matrix (ECM), invasion of surrounding tissues and basal lamina, intravasation into the blood stream, survival and transport through the blood stream, migration, arrest and extravasation at a distant site and formation of a metastatic lesion. These steps require fundamental mechanisms such as degradation of ECM barriers, disruption of cell-cell and cell-ECM adhesion and induction of cellular motility [6]. The role of ECM in cancer development and progression is increasingly recognized [7]. Some of the most important components of ECM, the collagens, have been implicated in various molecular mechanisms of cancer development [8]. It has been shown that the expression of collagen changes in malignancies, especially that of collagen XI. This would affect the ability of cancer cells to invade the stroma and metastasize [9].

The collagens are categorized in three main groups, according to their supramolecular assemblies: fibrillar collagens, non-fibril forming collagens and fibril-associated collagens [10]. Collagen type XI is a minor fibril-forming collagen, whose primary physiological role is the regulation of the diameter and function of major collagen fibrils in connective tissue, since it helps in the formation of the core of type II collagen, and it is responsible for the formation of thin fibrils of either developing or under-remodeling tissues. It is a heterotrimeric protein that consists of α1, α2 and α3 chains and they are organized into a triple helix. Both a1 and a2 genes of XI are unique products, however a3 gene is a hyperglycosylated version of the collagen a1 (II) chain [11]. The α1 chain of collagen type XI is encoded by the *COL11A1* gene at the genomic locus 1p21.1. Initially, collagen type XI is synthesized as procollagen; then its C and N termini are cleaved with proteolysis as soon as it is secreted from the cell [12]. The remaining a1 chain contains a large globular N-terminal domain that has been suggested to reduce tensile strength through steric hindrance with the major fibrillar collagens preventing their assembly into the heterotypic fibrils [13-15]. Decreased expression of this chain can lead to abnormally thickened fibrils [16].

The *COL11A1* gene consists of 67 exons. Due to alternative splicing of three separate *COL11A1* exons (6,7 and 8), there exist possibilities of production of at least 8 different protein isoforms [17-19]. Four different splicing variants of *COL11A1* mRNA (termed A, B, C and E) have already been deposited in GenBank and are known to differ in their propensity for N-terminal domain proteolysis and possibly the arrangement of ECM [18-21].

Collagen type XI (α1) is found mainly in cartilage issues, but has also been traced in skeletal muscles tissue, lung tissue, placenta, odontoblasts, trabecular bone and brain neoepithelium [22]. Multiple studies have recently highlighted the important role collagen XI(α1) plays in many aspects of cancer growth, invasion and metastasis [9]. Collagen type XI(α1) is a particularly attractive biomarker for malignancy, because of its restricted expression in normal tissues and non-cancer conditions like fibrosis or inflammation [23]. Under physiological conditions, the 5'-UTR that controls collagen type XI(α1) expression is highly methylated, leading to low levels of *COL11A1* expression [24]. However, deregulation and overexpression of collagen type XI(a1) has been identified in bladder [25], breast [26], colorectal [27,28], gastric [29,30], gliomas [31], squamous cell head and neck [32], ovarian [33,34] and pancreatic cancer [35]. In NSCLC, overexpression of collagen type XI(α1) has been correlated with the pathological staging, lymph node metastasis and poor prognosis [36,37].

None of the aforementioned studies has investigated the *COL11A1* variants. In our previous work, we have developed accurate quantitative RT-qPCR methods for the *COL11A1* A and E transcripts and have obtained promising results for their association with breast cancer [38]. Now with this present study, we expand into the development of the general (total) *COL11A1* transcript and the C transcript quantitative methodologies and then, measure all *COL11A1* transcripts and examine their potential association with histopathological prognostic factors in NSCLC lung tissue samples.

Methods

Patients

Tissue samples were collected from 27 patients, 19 from patients diagnosed with NSCLC (mean age±SD, 65±6.2 y) and 8 from cancer-free patients (controls) who underwent thoracic surgery for other reasons, in the Department of Cardiovascular Surgery of Attikon University Hospital, during the 2008-2012 period. The study was approved by both Bioethics and Scientific committees of Attikon Hospital. Patients provided their informed consent and samples were sent to our laboratory in a blinded fashion, stripped from any name or clinical information. Samples were either available material from diagnostic biopsies from controls and patients with non-resectable NSCLC or surgically removed tumor specimens in early-stage NSCLC; surgical resection is available as an option in IASLC classification NSCLC stages IA, IB, IIA, IIB and IIIA [39]. The presence of lymph node involvement or metastasis was recorded for those patients with follow-up data. Histopathological and clinical characteristics of the patient population are summarized in Table 1.

Table 1. Clinical and histopathological characteristics of the 19 tumor samples

Characteristics	*n (%)*
Sex	
Male	16 (84.2)
Female	3 (15.8)
Tumor histology	
Squamous	6 (31.6)
Adenocarcinoma	9 (47.4)
Large cell	2 (10.5)
Carcinosarcoma	1 (5.3)
BAC	1 (5.3)
Stage	
IA	3 (12.0)
IIA	9 (36.0)
IB	2 (8.0)
IIB	5 (20.0)
IIIA	4 (16.0)
IV	2 (8.0)
Lymph node status	
Negative	10 (52.6)
Positive	9 (47.4)
Metastasis during follow-up	
Yes	17 (89.5)
No	2 (10.5)
Death	
Yes	4 (21.0)
No	15 (79.0)

RNA isolation and cDNA synthesis

After resection, all tissue samples where immediately placed in RNA stabilization buffer (RNAlater, Qiagen, Germany) for 24-48 hrs at 4°C and then stored at -80°C until RNA extraction.

For RNA isolation, after removal of RNAlater, tissue specimens (~100 mg) were digested in 50 μl of Proteinase K buffer (20 mg/ml, Genaxxon Bioscience, Germany) for 4 hrs and then homogenized with the addition of 1 ml TRI REAGENT RT (MRC, USA) and vortexing, according to manufacturer's instructions. Briefly, 50 μl of bromoanisole were added to each vial, followed by centrifugation at 12000 *g* for 15 min at 4°C. The upper fraction containing the RNAs was separated and RNAs were precipitated after the addition of equal amount of isopropanol, incubation for 10 min at room temperature and centrifugation at 12000 *g* for 8 min at 4°C. The resulted RNA pellets were washed twice with 75% ethanol and centrifuged at 6000 *g* for 5 min, at 4°C. After ethanol removal by decantation and brief air-drying, RNAs were reconstituted in nuclease-free H_2O. RNA concentration and quality indication was accessed by spectrophotometry. All RNA samples were stored at -80°C.

For cDNA synthesis, each RNA sample (1 μg) was treated with DNase I (1u, ThermoFisher Scientific, USA) for 30 min at 37°C, in a 10 μl reaction, in order to digest any DNA molecules that might be co-precipitated during RNA isolation process. DNase I enzyme was deactivated after incubation for 10 min at 65°C in the presence of EDTA. Each DNA-free RNA sample was added in a 20 μl reaction containing the appropriate volumes of cDNA synthesis buffer, random primers, RNAse inhibitor (40 u) and Maxima reverse transcriptase (200 u, ThermoFisher Scientific) and incubated at 50°C for 1 hr, followed by incubation at 85°C for 5 min. All cDNA samples were stored at -20°C till further *COL11A1* gene expression analysis.

Real-time qPCR methodologies for the general COL11A1 and C transcripts

For the quantification of the general *COL11A1* and C transcripts, suitable pairs of primers and hybridization sets of dual probes (labelled with fluorescein donor and LC-Red 640 or 705 acceptor dyes) were designed in the CLC Free Workbench v.4 program (QIAGEN Bioinformatics, Germany) in order to select the most suitable regions for their binding. For the general transcript, the forward primer binds to the exon junction between exons 48 and 49, while the reverse binds to exon 51. This area is present in all the reported *COL11A1* transcripts so far. The C transcript is characterized by the absence of the exons 7, 8 and 9. Its primers are bound to the exons 5 and 10 respectively, whereas the probes are designed for the detection of exon 6. Sequences of primers and probes synthesized by TIB MOLBIOL (Germany) are shown in Table 2.

Real-time qPCR was performed in Light Cycler 1.5 platform (Roche Applied Science, Germany) in glass capillaries in a total volume of 10 μl. For the general *COL11A1* transcript, 1 μl of the sample cDNA was added to the 0.1 μl of forward primer (*COL11A1* F with final

Table 2. Sequences of primers and probes for the novel general *COL11A1* and *COL11A1* C transcript RT-qPCR methodologies

	Name	*Oligonucleotide sequence*
General transcript forward primer	F	5' AATGGAGCTGATGGACCACA
General transcript reverse primer	R	5' TCCTTTGGGACCGCCTAC
C transcript forward primer	VARAC F	5' TGTGAGCATTATAGTCCAGACTGTGA
C transcript reverse primer	VARC R	5' CCATGGCCATTTATCTCCGT
General transcript Sensor probe	COL11A1 FL	5' TCCCTgCTTCTCCAggTTCACCC-FL*
General transcript Anchor probe	COL11A1 LC	5' LC 640-TTTCTCCAACACCACCAACTgAACCAAC
C transcript Sensor probe	VARAC FL	5' TCCTCAGTTACAGTGGGTCCCTCTGTTAC-FL*
C transcript Anchor probe	VARAC LC	5' LC 705-CTTTCAGCCTCTTTATACTCTGCTTCCCCA

*FL denotes fluorescein

concentration: 0.6 pmol/μl), 0.3 μl of the reverse primer (*COL11A1* R with final concentration: 0.2 pmol/μl), 1 μl of the sensor probe (*COL11A1* FL with final concentration: 0.3 μM), 1 μl of the anchor probe (*COL11A1* LC with final concentration: 0.3 μM), 1.2 μl $MgCl_2$ (Roche, final concentration 3mM), 1μl LightCycler FastStart DNA Master HybProbe 10X (with final concentration 1X) and ddH_2O to the final volume.

For the transcript C, 1 μl of the sample cDNA was added to the 0.3 μl of the forward primer (VARAC F with final concentration 0.6 pmol/μl), 0.1 μl of the reverse primer (VARC R with final concentration 0.2 pmol/μl), 0.6 μl of the sensor probe (VARAC FL with final concentration 0.18 μM), 0.6 μl of the anchor probe (VARC LC with final concentration 0.18 μM), 1.2 μl $MgCl_2$ (Roche, final concentration 3 mM), 1 μl LightCycler FastStart DNA Master HybProbe 10X (with final concentration 1X), 0.6 μl DMSO and ddH_2O to the final volume.

All reactions were initiated with a 10-min denaturation at 95°C and terminated with a 30 sec cooling step at 40°C. The cycling protocol consisted of denaturation step at 95°C for 10 sec, annealing at 52°C for 30 sec and extension at 72°C for 30 sec, repeated for 42 cycles. This protocol was followed by both variants, namely the general *COL11A1* and the C transcript. In each preparation for the real-time qPCR, together with the unknown samples, standards and blank samples were included. The blank samples did not contain cDNA. Also, melting curve analysis was performed in the Light Cycler after the amplification process in order to assess amplicon specificity with the temperature melting point (Tm). The protocol that was followed was raising the temperature at 95°C, cooling at 55°C for 15 sec and slow heating to 95°C at a rate of 0.1°C/sec, while the fluorescence was continuously monitored. The real-time qPCR products were additionally checked with electrophoresis on 2% w/v agarose gel for their proper size and purity.

For absolute quantification of unknown samples in both assays, an external standard curve was obtained by using quantified pure PCR amplicon standards and plotting the log number of copies corresponding to each standard versus its quantification cycle (Cq). In order to prepare a highly concentrated calibrator (standard), several PCR products were united and then purified by the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Germany) followed by measuring the concentration by the Quant-iT dsDNA Broad range Assay kit in the Qubit 1.0 Fluorometer (Invitrogen/Thermo Fisher, USA). The copies/μl was calculated as described previously [40]. The highly concentrated calibrator was serially diluted (1:10) and a standard curve was obtained.

The reproducibility (calculated as coefficients of variation, CVs of Cqs and Tms) and the efficiency of the PCR reaction (expressed as $E= 10^{-1/slope}$) were also determined in order to complete the validation file of the novel methodologies with the established MIQE guidelines [41].

COL11A1 transcripts A and E in the tissue samples were detected with RT-qPCR methods that have already been reported previously by our group [38]. For normalization in our tissue samples, all *COL11A1* results in copies/μl were converted to copies/μg total RNA.

Restriction Fragment Analysis

For the confirmation of *COL11A1* general PCR products (132 bp each), *HaeIII* and *MnlI* restriction enzymes were used. For each restriction digest 0.5 μl of the enzyme, 1.5 μl of NEB4 10X buffer and 13 μl of PCR product were mixed and incubated in 37°C for 2 hrs. After that, the products of the restriction procedure were identified by electrophoresis on a high resolution 3% w/v agarose gel (2:1 Seakem:Nusieve agarose, Lonza, USA).

DNA sequencing

For the *COL11A1* C transcript, the products were also checked for proper DNA composition. The Sanger DNA sequencing methodology was performed after PCR product column clean-up (High Pure PCR Cleanup Micro Kit, Roche, Applied Science, Germany) and a cycle sequencing reaction employing the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, USA). PCR products of cycle sequencing were freed from unincorporated nucleotides with the NucleoSEQ kit (Macherey-Nagel, Germany). All the above procedures were held according to the manufacturers' instructions. The electrophoregram in the ABI Prism 310 Genetic Analyzer (Applied Biosystems, USA) were manually base-called

with the Chromas Lite 2.01 software (Technelysium Pty, Tewantin, Australia) and compared with the expected sequence with the BLAST tool of PubMed.

Statistics

Statistical analyses were performed with the SPSS v.22 (IBM, USA) in a qualitative way (presence or absence of a transcript) with Pearson x^2 or Fisher's exact test and in a quantitative way (transcript copies/μg total RNA) with t-test or Mann Whitney U test (depending on normality assessed with the Kolmogorov-Smirnoff test). Spearman's correlations were examined between *COL11A1* transcripts and when correlated, linear regression was applied. Logistic regression was used in order to assess statistical models including the presence of any *COL11A1* transcript and histopathological parameters for prediction of either lymph node status or metastasis. The association of *COL11A1* transcript variants with metastasis was analyzed with the Kaplan-Meier method and survival curves were compared with the log-rank test. For all tests performed, a two-sided p value of <0.05 was considered significant.

Results

Clinical samples

The majority of the NSCLC tumors were adenocarcinoma (47.4%) and the rest were squamous (31.6%), large cell (10.5%), carcinosarcoma (3.7%) and BAC (bronchio-alveolar carcinoma, adenosubcategory, 3.7%). The NSCLC patients were followed up for a median time-period of 39 months and any metastasis during this period was recorded (clinical and histopathological characteristics are summarized in Table 1).

Development of novel general COL11A1 and COL11A1 C transcript assays

To establish linear, reproducible and specific real-time quantitative assays, we performed extensive optimization of primers, probes and $MgCl_2$ concentrations as well as of the reaction temperatures and cycles (data not shown). DMSO addition was essential for transcript C. The same temperature profile and the assymetric PCR format in the PCR mix composition was followed in both novel assays.

The measuring range for the standard curve for the general *COL11A1* transcript was between 4.71×10^8 copies/μl to 4.71×10^2 copies/μl. A characteristic curve with 7 calibrators is shown in Figure 1 that produces a linear equation [$\log C_{general}$= 3.457xCq + 40.13] and the PCR efficiency is calculated as 1.95 (ideal value 2.0). A 4.71x10 copies/μl calibrator is also detected but is not included in the

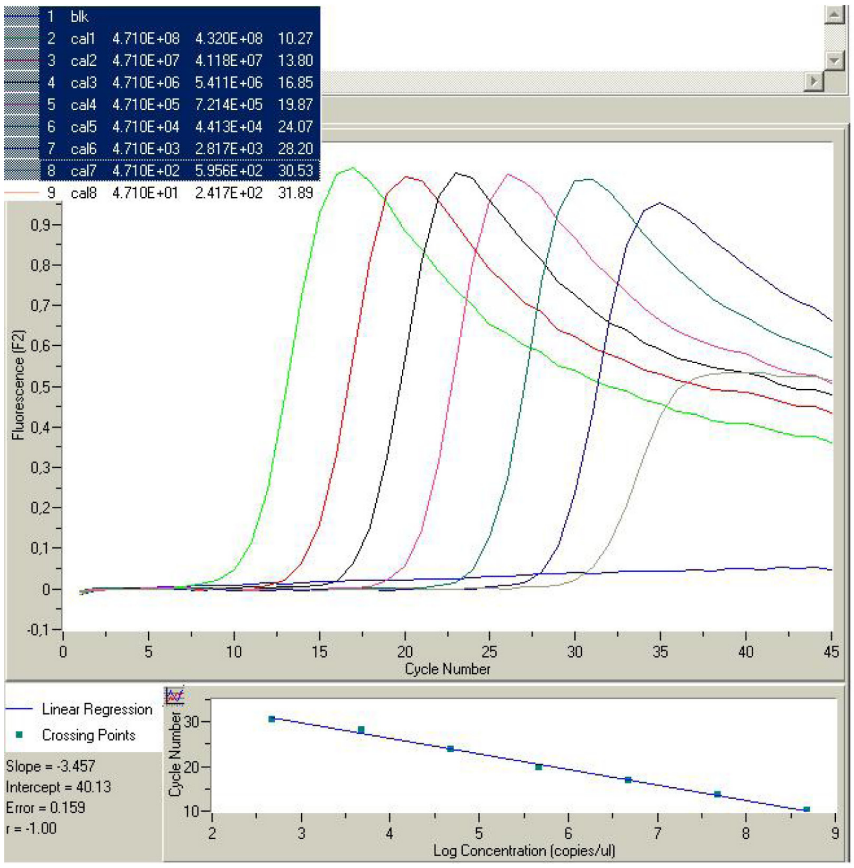

Figure 1. Standard curve for the general *COL11A1* transcript.

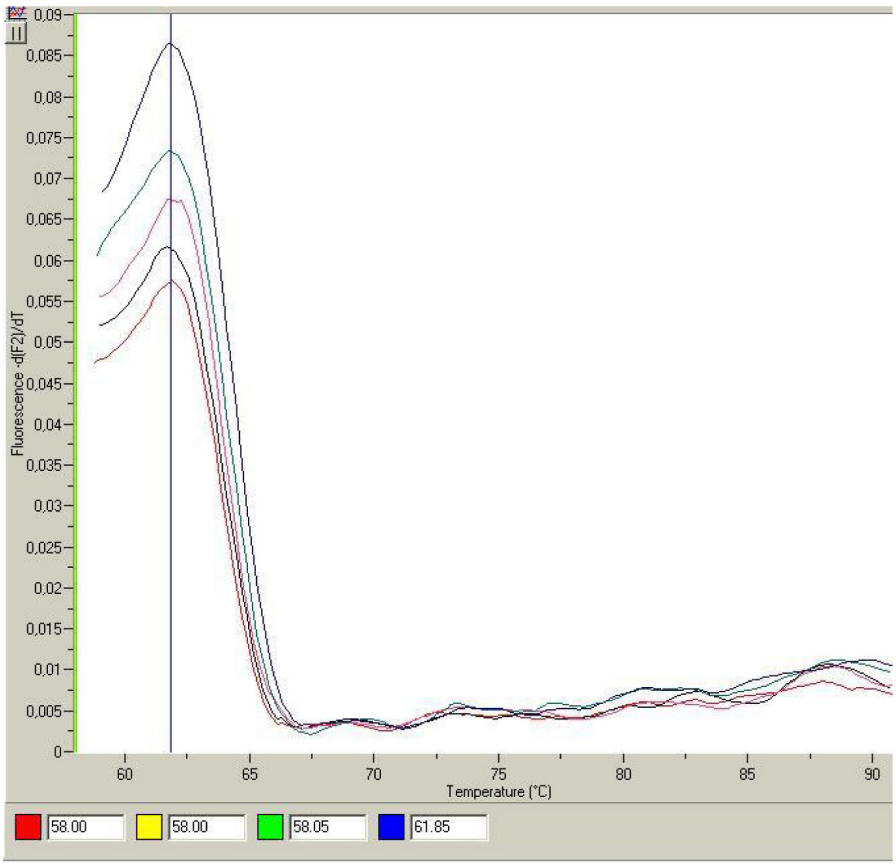

Figure 2. Tm analysis for the *COL11A1* general transcript.

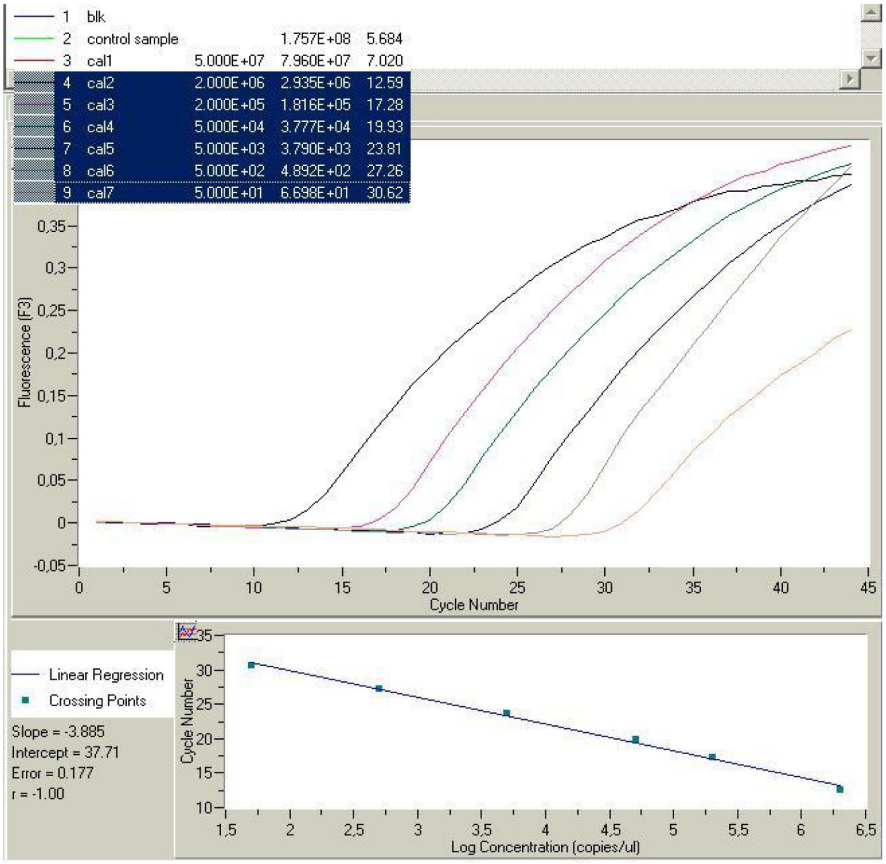

Figure 3. Standard curve for the *COL11A1* C transcript.

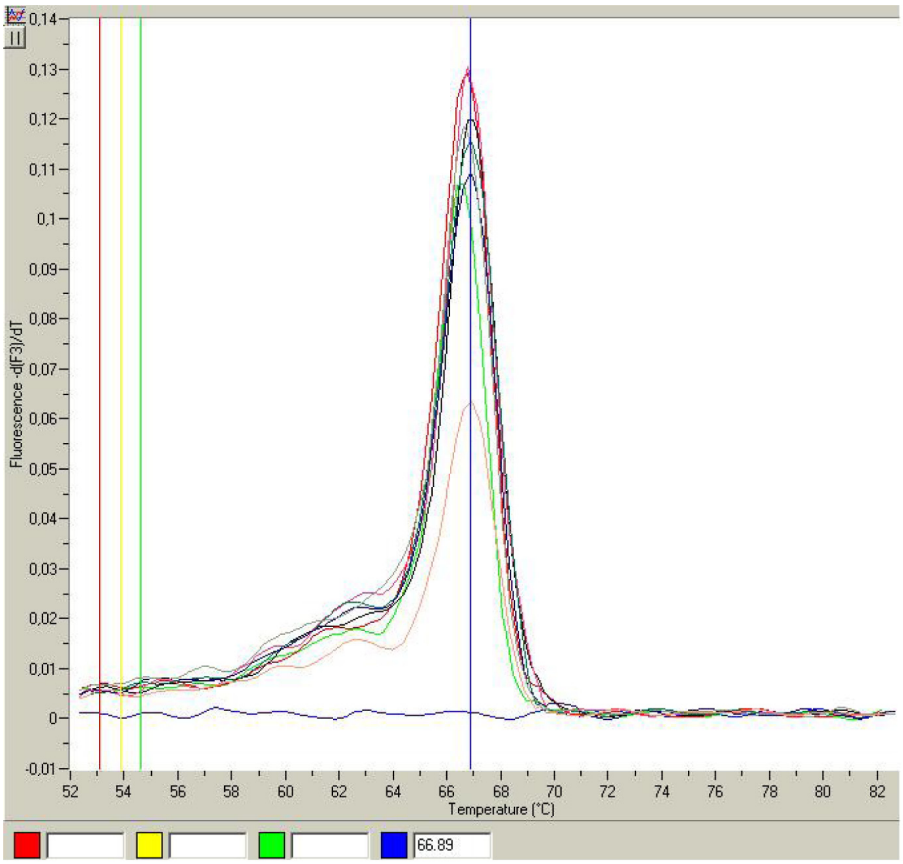

Figure 4. Tm analysis for the *COL11A1* C transcript.

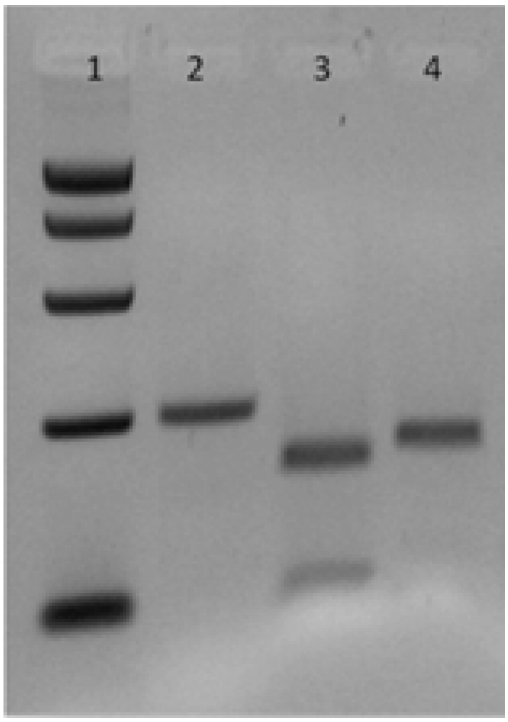

Figure 5. Agarose gel electrophoresis showing the restriction fragment analysis of the general *COL11A1* transcript (Position 1: Molecular Weight Marker ladder (PCR Marker, NEB, USA) showing bands at 766, 500, 300 and 150 bp; Position 2: *COL11A1* PCR product; Position 3: *HaeIII* restriction digest; Position 4: *MnlI* restriction digest).

standard curve. All CVs of Cqs of the standards were <2.5% in all standard curves that were run (n=5). The Tm of the general transcript amplicon is 61.85 ± 1.03°C (Figure 2).

The measuring range for the standard curve for the *COL11A1* C transcript was between 2×10^6 copies/µl to 5×10^1 copies/µl. A characteristic curve with 6 calibrators is shown in Figure 3 that produces a linear equation ($logC_{transcript}$= 3.885xCq + 35.71) and the PCR efficiency is calculated as 1.81. All CVs of Cqs of the standards were <2.5% in all standard curves that were run (n=5). The Tm of the C transcript amplicon is 66.89 ±0.15°C (Figure 4).

COL11A1 transcripts A and E in our tissue samples were detected with RT-qPCR methods that have already been reported previously by our group [38]. The expected size of the PCR products for *COL11A1* general, A, C and E transcripts were 132, 596, 248 and 479 bp, respectively and confirmed by agarose electrophoresis by inverting the amplicons in the glass capillaries into Eppendorf tubes. General *COL11A1* amplicon size was also confirmed by restriction enzyme digests, as it is shown in Figure 5 with *HaeIII* (producing two fragments of 96 bp and 36 bp) and *MnlI* (producing two fragments of 107 bp and 25 bp). Transcript's

C products were additionally confirmed by DNA sequencing, using Chromas Lite and BLAST tools (Figure 6).

COL11A1 tissue results and statistical analysis

Tissue *COL11A1* results are presented in Table 3. All 19 NSCLC samples were positive for the general *COL11A1* transcript (range 11.2-1198.0 copies/μg total RNA), while 5 out of 8 control samples were negative (x^2, $p<0.05$): mean values were also statistically significantly different (197 *vs.* 7 copies/μg total RNA, $p<0.001$). In 4 tumor samples (21%), no specific *COL11A1* transcript was detected. Transcript C was detected in only 3 tumor samples. These 3 tumors contained additionally A and E transcripts as well. Regarding *COL11A1* A and E transcripts, 13 out of 19 tumor samples were positive for either one (68%) and 11 for both (58%). *COL11A1* A and E transcripts mRNA levels were significantly correlated (rho=0.62, p=0.01); in most tissues, A levels were higher than E levels (linear regression: transcriptA=2.7*transcriptE +3.2). The sum of A, C and E transcripts correlated highly with the result of the total (general) COL11A transcript (rho=0.57, p=0.002, general =8.29*sum +14.22). When trying to compare *COL11A1* general, A and E mean and/or median levels (and percentages in A and E) between tumor subgroups (squamous *vs.* adenocarcinomas, lymph node positive *vs.* negative, metastasis *vs.* no metastasis) no statistically significant difference was detected. Also logistic regression and survival statistical analysis failed to reveal any important correlations, most probably due to the limited number of samples.

Discussion

The tumor stroma is consisted of cancer-associated fibroblasts (CAFs), tumor-associated macrophages, tumor-associated endothelial cells and tumor-directed lymphocytes [42]. Alteration in normal stromal properties, including CAFs activation, ECM protein disposition, angiogenesis and infiltration by immune cells is known as desmoplasia [43]. Tumor stromal cells not only produce inflammatory, growth, angiogenic and proteolytic factors that enhance motility and invasion, but they also generate a tumor-supportive microenvironment by leading to an excessive production of ECM proteins like collagen or fibronectin [44]. The conversion of normal fibroblasts to CAFs is known as tumor stromatogenesis, is controlled by cancer-cell derived

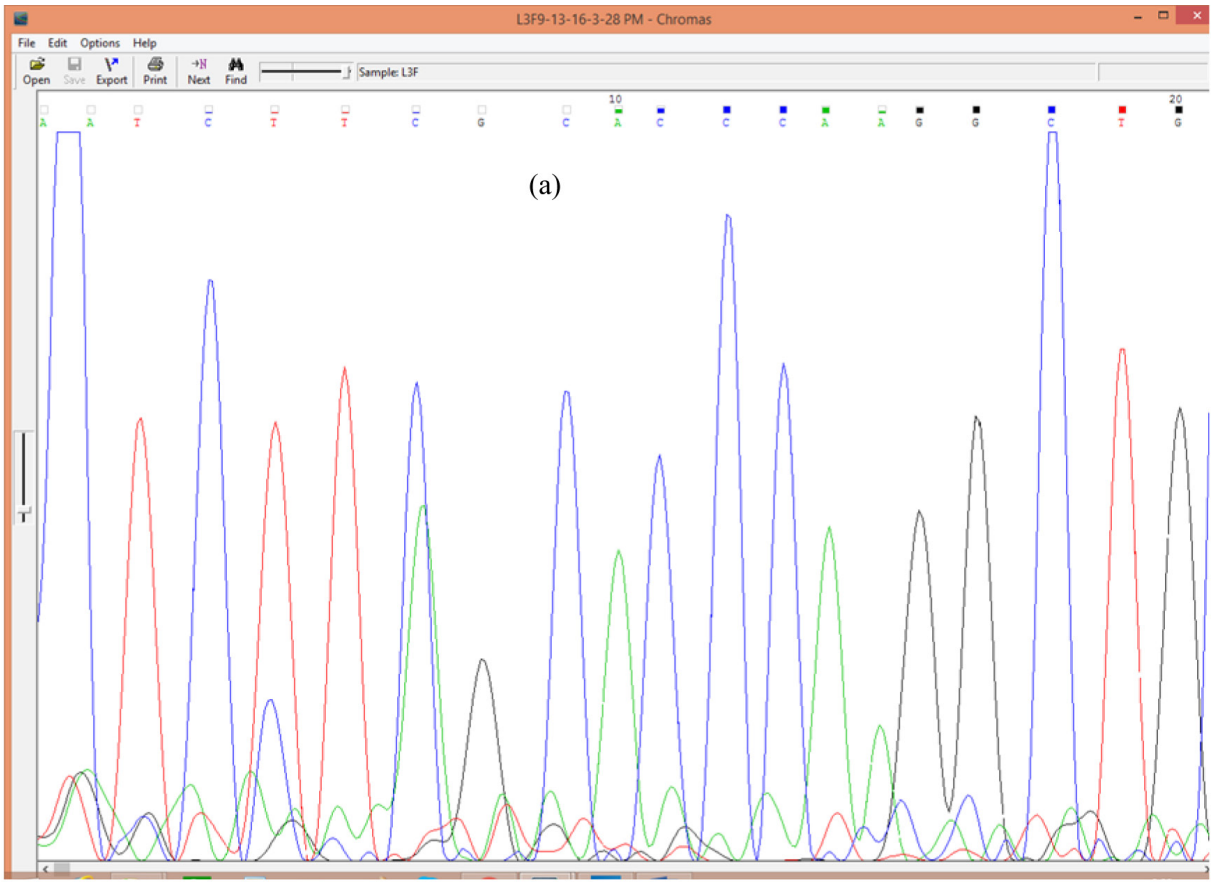

Figure 6. DNA sequencing electrophoregram of the *COL11A1* C transcript.

Table 3. Patient sample type and *COL11A1* general and A,E and C transcripts quantitative results in copies/μg total RNA. The sum of A,C and E transcripts is provided in the last column of the Table

ID	*Type*	*COL11A1 general*	*A transcript*	*C transcript*	*E transcript*	*Sum of A+C+E*
L1	Tumor	38.5	6.8	0	2.4	9.2
L2	Tumor	11.5	4.4	0	0.8	5.2
L5	Tumor	42.3	14.3	16.9	11.4	42.6
L6	Tumor	11.2	4.1	0	0.9	5.0
L7	Control	17.9	1.5	0	0.6	2.1
L10	Control	0	0	0	0	0
L13	Tumor	12.9	5.3	0	1.8	7.1
L14	Tumor	23.3	0	0	0.2	0.2
L16	Tumor	65.7	7.3	4.5	1.8	13.6
L18	Control	28.9	14.3	0	0.9	15.2
L25	Tumor	25.4	0	0	0	0
L26	Control	0	0	0	0	0
L27	Tumor	223.1	0	0	1.1	1.1
L28	Tumor	1198.0	114	0	23.1	137.1
L30	Tumor	203.9	4.8	0	2.8	7.6
L32	Tumor	84.8	12.5	0	0	12.5
L34	Control	0	0	0	0	0
L35	Tumor	748.5	23.8	5.5	24.0	53.3
L36	Tumor	500.0	60.1	0	5.1	65.2
L37	Control	44.5	15.9	0	0	15.9
L38	Tumor	96.5	0	0	0	0
L42	Tumor	125.5	7.4	0	2.2	9.6
L45	Tumor	228.3	0	0	0	0
L48	Control	0	0	0	0	0
L52	Tumor	29.1	5.0	0	0	5.0
L53	Control	0	0	0	1.4	1.4
L54	Tumor	83.8	0	0	0	0

TGFβ and leads to overproduction of growth factors like TGFβ, VEGF and FGF2 [45-49].

COL11A1 gene has been shown to be the most overexpressed gene in a metastasis-associated gene signature identified in stromal tissues of different cancers, including colon, breast, ovarian and NSCLC [50,51]. This gene expression corresponds to a stromal desmoplastic reaction produced by CAFs, usually in advanced cancer stages [9,52]. *COL11A1* has been suggested as a specific marker of CAFs-associated desmoplastic events in pancreatic cancer [35], in head and neck squamous cell cancers [32], allowing the distinction from normal benign fibroblasts. The fact that *COL11A1* mRNA expression correlates with cancer staging in multiple cancers, and the absence of its expression in normal stromal cells, makes *COL11A1* expression an early marker of invasive potential [9,53].

Although the sequence of 4 variant *COL11A1* mRNA transcripts has already been deposited in GenBank (A, B, C, E), no other research group has investigated them as tumor biomarkers so far. In our previous study in 90 breast cancer specimens, we have detected and quantified A and E *COL11A1* transcripts and *COL11A1* E variant correlated with lymph node positivity and metastasis [38]. No *COL11A1* B and C transcript variants were detected, thus, we were unable to validate the corresponding methodologies. With this study, we expand our variant *COL11A1* investigation in a pilot study for lung cancer (19 NSCLC lung tissue samples and 8 control samples). In 3 tumor samples, the *COL11A1* C transcript was detected for the first time and we were able to validate RT-qPCR for its accurate measurement (the C transcript was also confirmed with DNA sequencing). No *COL11A1* B transcript was detected, while regarding A and E transcripts that were measured according to our previous study [38], 13 out of 19 tumor samples were positive for either one (68%) and 11 for both (58%). Most probably, the limited number of samples prevented the finding of any statistically sig-

nificant correlation of the detected specific A,C and E *COL11A1* transcripts with patient characteristics (histology, stage, lymph node positivity, metastasis, death) in either a qualitative or a quantitative way.

In order to investigate the possibility of the existence of other transcripts as well, the development and validation of a quantitative method for the measurement of the total *COL11A1* mRNA burden was a necessity. The method was designed to detect solely RNA since one of the primers binds to an exon junction area in the corresponding C-terminal part which is a common area in all of the deposited transcript variants. To further ensure proper RNA preservation and amplification, all tissue samples were stored in appropriate stabilization reagent at -80°C until RNA isolation and then, they were DNAse-treated. The developed novel RT-qPCR method for the general *COL11A1* was adequately validated in the Light Cycler 1.5 real-time PCR platform for linearity and reproducibility. The specificity of the general *COL11A1* PCR amplicon was confirmed with agarose gel electrophoresis (proper size and purity), Tm reproducibility and further, with restriction fragment analysis.

All 19 NSCLC samples were positive for the general (total) *COL11A1* transcript while 5 out of 8 control samples were negative, a finding that corroborates all the aforementioned studies that have shown that *COL11A1* expression is a reliable biomarker in the majority of cancer types where it has been investigated. Towards this direction, our novel quantitative assay could become a valuable tool. The mean value of total *COL11A1* expression in the 3 positive control samples is significantly less than the corresponding value in the tumor samples ($p<0.001$).

In 4 tumor samples (21%), no specific *COL11A1* transcript was detected. This finding along with the fact that the sum of the detected A+C+E transcripts lags significantly from the number of transcripts/μg of the general *COL11A1* transcript in all of our lung tumor samples (at least 8 times more general transcript than what the sum, in only 1 sample is equal), indicate the existence of other unknown transcripts, besides A, B, C and E, that have not officially been deposited but could also have prognostic significance. It should also be noted that the existence of transcript variants lacking the C-terminal area - that is detected by our general *COL11A1* method - cannot be excluded as well.

In general, the presence of multiple aberrant gene transcripts could show a defect in the cancer cell splicing machinery but in many cases, there also exists a specific transcript effect e.g full-length survivin and the aberrant survivin-ΔEx3 transcripts have been shown to act as inhibitors of apoptosis while another transcript, survivin-2b presents pro-apoptotic functions by dimerizing with wild type survivin and reducing its anti-apoptotic effects [54]. Regarding collagen XI transcripts, *COL11A1* E variant correlated with lymph node positivity and metastasis in our previous study in breast cancer [38].

With the advent of Next Generation Sequencing techniques (NGS) a more thorough transcriptome analysis could provide the missing information. So far, a limited number of FDA-approved gene expression panels exist for analyzing tumor tissues in few types of cancer (Oncotype Dx for breast, colon, and prostate, Mammaprint and PAM50 Prosigna for breast). In lung cancer, most tumor-driven analyses have been focused on *KRAS* and *EGFR* somatic DNA mutational analysis and, in some instances, in *ALK* rearrangements (and/or *EML4-ALK* RT-qPCR mRNA analysis). According to results from other studies [36,37] but also from our pilot study regarding the general transcript, *COL11A1* gene could be a candidate gene for inclusion in prognostic expression gene panels for NSCLC in the future. In a 2018 study, *COL11A1* was one of the three genes that were suitable as prognostic predictors in colon cancer in a 65-gene expression panel [28]. It remains to be seen whether *COL11A1* variants could be included as well. Besides analyzing tumor genomic profiles, tumor burden shedding in blood provides an opportunity for non-invasive biomarker testing and towards this goal, liquid-biopsy testing of cell-free DNA with NGS or digital PCR - such as Guardant360, exosomeDx - is promising in cancer therapy monitoring [55,56]. In the future, it would be intriguing to investigate *COL11A1* expression in plasma exosome analysis as well.

Conflict of interests

The authors declare no conflict of interests.

References

1. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. Lancet Oncol 2012;13:790-801.

2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87-108.
3. Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD. Epidemiology of lung cancer: Diagnosis and management of lung cancer (3rd Edn): American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013;143:e1S-e29S.
4. Jemal A, Ward EM, Johnson CJ et al. Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. J Natl Cancer Inst 2017;109. doi 10.1093/jnci/djx030.
5. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.
6. Perlikos F, Harrington KJ, Syrigos KN. Key molecular mechanisms in lung cancer invasion and metastasis: a comprehensive review. Crit Rev Oncol Hematol 2013;87:1-11.
7. Barkan D, Green JE, Chambers AF. Extracellular matrix: a gatekeeper in the transition from dormancy to metastatic growth. Eur J Cancer 2010;46:1181-8.
8. Hynes RO. The extracellular matrix: not just pretty fibrils. Science 2009;326:1216-9.
9. Raglow Z, Thomas SM. Tumor matrix protein collagen XIalpha1 in cancer. Cancer Lett 2015;357:448-53.
10. Ricard-Blum S. The collagen family. Cold Spring Harb Perspect Biol 2011;3:a004978.
11. Morris NP, Bachinger HP. Type XI collagen is a heterotrimer with the composition (1 alpha, 2 alpha, 3 alpha) retaining non-triple-helical domains. J Biol Chem 1987;262:11345-50.
12. Thom JR, Morris NP. Biosynthesis and proteolytic processing of type XI collagen in embryonic chick sterna. J Biol Chem 1991;266:7262-9.
13. Gelse K, Poschl E, Aigner T. Collagens--structure, function, and biosynthesis. Adv Drug Deliv Rev 2003;55:1531-46.
14. Hansen U, Bruckner P. Macromolecular specificity of collagen fibrillogenesis: fibrils of collagens I and XI contain a heterotypic alloyed core and a collagen I sheath. J Biol Chem 2003;278:37352-9.
15. Toumpoulis IK, Oxford JT, Cowan DB et al. Differential expression of collagen type V and XI alpha-1 in human ascending thoracic aortic aneurysms. Ann Thorac Surg 2009;88:506-13.
16. Hida M, Hamanaka R, Okamoto O et al. Nuclear factor Y (NF-Y) regulates the proximal promoter activity of the mouse collagen alpha1(XI) gene (Col11a1) in chondrocytes. In Vitro Cell Dev Biol Anim 2014;50:358-66.
17. Yoshioka H, Inoguchi K, Khaleduzzaman M, Ninomiya Y, Andrikopoulos K, Ramirez F. Coding sequence and alternative splicing of the mouse alpha 1(XI) collagen gene (Col11a1). Genomics 1995;28:337-40.
18. Zhidkova NI, Justice SK, Mayne R. Alternative mRNA processing occurs in the variable region of the pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains. J Biol Chem 1995;270:9486-93.
19. Oxford JT, Doege KJ, Morris NP. Alternative exon splicing within the amino-terminal nontriple-helical domain of the rat pro-alpha 1(XI) collagen chain generates multiple forms of the mRNA transcript which exhibit tissue-dependent variation. J Biol Chem 1995;270:9478-85.
20. Rousseau JC, Farjanel J, Boutillon MM, Hartmann DJ, van der Rest M, Moradi-Ameli M. Processing of type XI collagen. Determination of the matrix forms of the alpha1(XI) chain. J Biol Chem 1996;271:23743-8.
21. Medeck RJ, Sosa S, Morris N, Oxford JT. BMP-1-mediated proteolytic processing of alternatively spliced isoforms of collagen type XI. Biochem J 2003;376:361-8.
22. Yoshioka H, Iyama K, Inoguchi K, Khaleduzzaman M, Ninomiya Y, Ramirez F. Developmental pattern of expression of the mouse alpha 1 (XI) collagen gene (Col11a1). Dev Dyn 1995;204:41-7.
23. Jia D, Liu Z, Deng N et al. A COL11A1-correlated pan-cancer gene signature of activated fibroblasts for the prioritization of therapeutic targets. Cancer Lett 2016;382:203-14.
24. Cortese R, Hartmann O, Berlin K, Eckhardt F. Correlative gene expression and DNA methylation profiling in lung development nominate new biomarkers in lung cancer. Int J Biochem Cell Biol 2008;40:1494-1508.
25. Ewald JA, Downs TM, Cetnar JP, Ricke WA. Expression microarray meta-analysis identifies genes associated with Ras/MAPK and related pathways in progression of muscle-invasive bladder transition cell carcinoma. PLoS One 2013;8:e55414.
26. Vargas AC, McCart Reed AE, Waddell N et al. Gene expression profiling of tumour epithelial and stromal compartments during breast cancer progression. Breast Cancer Res Treat 2012;135:153-65.
27. Fischer H, Salahshor S, Stenling R et al. COL11A1 in FAP polyps and in sporadic colorectal tumors. BMC Cancer 2001;1:17.
28. Jiang H, Du J, Gu J, Jin L, Pu Y, Fei B. A 65gene signature for prognostic prediction in colon adenocarcinoma. Int J Mol Med 2018;41:2021-7.
29. Zhao Y, Zhou T, Li A et al. A potential role of collagens expression in distinguishing between premalignant and malignant lesions in stomach. Anat Rec (Hoboken) 2009;292:692-700.
30. Li A, Li J, Lin J, Zhuo W, Si J. COL11A1 is overexpressed in gastric cancer tissues and regulates proliferation, migration and invasion of HGC-27 gastric cancer cells in vitro. Oncol Rep 2017;37:333-40.
31. Chernov AV, Baranovskaya S, Golubkov VS et al. Microarray-based transcriptional and epigenetic profiling of matrix metalloproteinases, collagens, and related genes in cancer. J Biol Chem 2010;285:19647-59.
32. Sok JC, Lee JA, Dasari S et al. Collagen type XI alpha1 facilitates head and neck squamous cell cancer growth and invasion. Br J Cancer 2013;109:3049-56.
33. Cheon DJ, Tong Y, Sim MS et al. A collagen-remodeling gene signature regulated by TGF-beta signaling is associated with metastasis and poor survival in serous ovarian cancer. Clin Cancer Res 2014;20:711-23.
34. Wu YH, Chang TH, Huang YF, Huang HD, Chou CY. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. Oncogene 2014;33:3432-40.

35. Garcia-Pravia C, Galvan JA, Gutierrez-Corral N et al. Overexpression of COL11A1 by cancer-associated fibroblasts: clinical relevance of a stromal marker in pancreatic cancer. PLoS One 2013;8:e78327.
36. Chong IW, Chang MY, Chang HC et al. Great potential of a panel of multiple hMTH1, SPD, ITGA11 and COL11A1 markers for diagnosis of patients with non-small cell lung cancer. Oncol Rep 2006;16:981-8.
37. Shen L, Yang M, Lin Q, Zhang Z, Zhu B, Miao C. COL11A1 is overexpressed in recurrent non-small cell lung cancer and promotes cell proliferation, migration, invasion and drug resistance. Oncol Rep 2016;36:877-85.
38. Karaglani M, Toumpoulis I, Goutas N et al. Development of novel real-time PCR methodology for quantification of COL11A1 mRNA variants and evaluation in breast cancer tissue specimens. BMC Cancer 2015;15:694.
39. Detterbeck FC, Boffa DJ, Tanoue LT. The new lung cancer staging system. Chest 2009;136:260-71.
40. Kroupis C, Stathopoulou A, Zygalaki E, Ferekidou L, Talieri M, Lianidou ES. Development and applications of a real-time quantitative RT-PCR method (QRT-PCR) for BRCA1 mRNA. Clin Biochem 2005;38:50-7.
41. Bustin SA, Benes V, Garson JA et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 2009;55:611-22.
42. Hofmeister V, Schrama D, Becker JC. Anti-cancer therapies targeting the tumor stroma. Cancer Immunol Immunother 2008;57:1-17.
43. Demarchi LM, Reis MM, Palomino SA et al. Prognostic values of stromal proportion and PCNA, Ki-67, and p53 proteins in patients with resected adenocarcinoma of the lung. Mod Pathol 2000;13:511-20.
44. Morra L, Moch H. Periostin expression and epithelial-mesenchymal transition in cancer: a review and an update. Virchows Arch 2011;459:465-75.
45. De Wever O, Mareel M. Role of tissue stroma in cancer cell invasion. J Pathol 2003;200:429-47.
46. Giatromanolaki A, Sivridis E, Koukourakis MI. The pathology of tumor stromatogenesis. Cancer Biol Ther 2007;6:639-45.
47. Karvonen HM, Lehtonen ST, Sormunen RT, Lappi-Blanco E, Skold CM, Kaarteenaho RL. Lung cancer-associated myofibroblasts reveal distinctive ultrastructure and function. J Thorac Oncol 2014;9:664-74.
48. Horie M, Saito A, Noguchi S et al. Differential knockdown of TGF-beta ligands in a three-dimensional co-culture tumor- stromal interaction model of lung cancer. BMC Cancer 2014;14:580.
49. Massague J. TGFbeta signalling in context. Nat Rev Mol Cell Biol 2012;13:616-30.
50. Kim H, Watkinson J, Varadan V, Anastassiou D. Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1. BMC Med Genomics 2010;3:51.
51. Navab R, Strumpf D, Bandarchi B et al. Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer. Proc Natl Acad Sci U S A 2011;108:7160-5.
52. Anastassiou D, Rumjantseva V, Cheng W et al. Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. BMC Cancer 2011;11:529.
53. Galvan JA, Garcia-Martinez J, Vazquez-Villa F et al. Validation of COL11A1/procollagen 11A1 expression in TGF-beta1-activated immortalised human mesenchymal cells and in stromal cells of human colon adenocarcinoma. BMC Cancer 2014;14:867.
54. Pavlidou A, Kroupis C, Goutas N, Dalamaga M, Dimas K. Validation of a Real-Time Quantitative Polymerase Chain Reaction Method for the Quantification of 3 Survivin Transcripts and Evaluation in Breast Cancer Tissues. Clin Breast Cancer 2014;14:122-31.
55. Barata PC, Koshkin VS, Funchain P et al. Next-generation sequencing (NGS) of cell-free circulating tumor DNA and tumor tissue in patients with advanced urothelial cancer: a pilot assessment of concordance. Ann Oncol 2017;28:2458-63.
56. Rossi G, Mu Z, Rademaker AW et al. Cell-Free DNA and Circulating Tumor Cells: Comprehensive Liquid Biopsy Analysis in Advanced Breast Cancer. Clin Cancer Res 2018;24:560-8.