Indian Journal of Experimental Biology
Vol. 62, March 2024, pp. 179-186
DOI: 10.56042/ijeb.v62i03.2203

Systematic evaluation of collagen family gene transcripts: Implications for prognostic stratification in lung cancer patients

Sedef Hande Aktas[1,2]*

[1]Vocational School of Health Services, Eskisehir Osmangazi University, Eskisehir, Turkey

[2]Department of Biotechnology and Biosafety, Graduate School of Natural and Applied Sciences, Eskisehir Osmangazi University, Eskisehir, Turkey

Supplementary Data

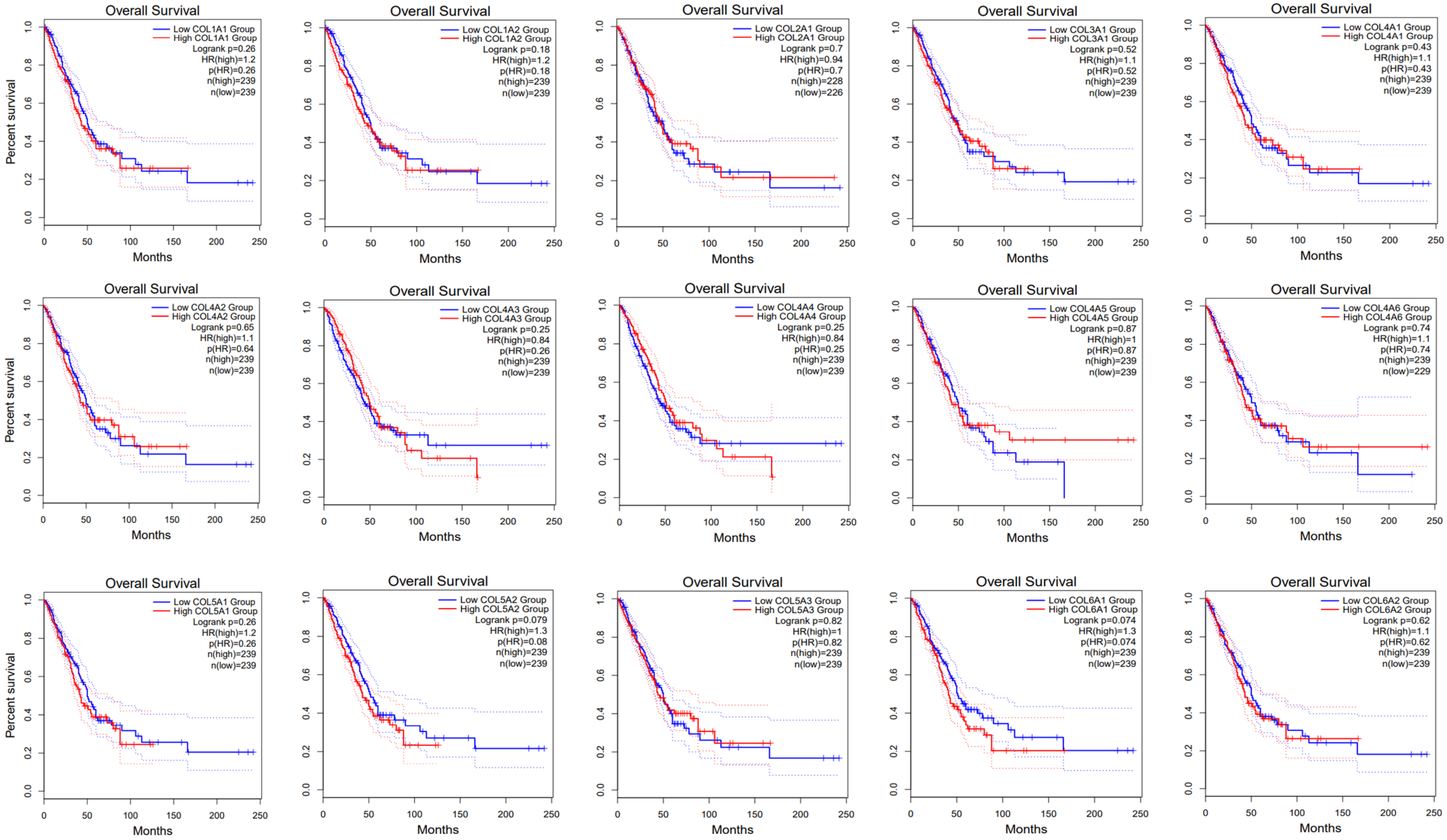
Suppl. Table S1 — Demographic and clinical data of the LUAD and LUSC patients

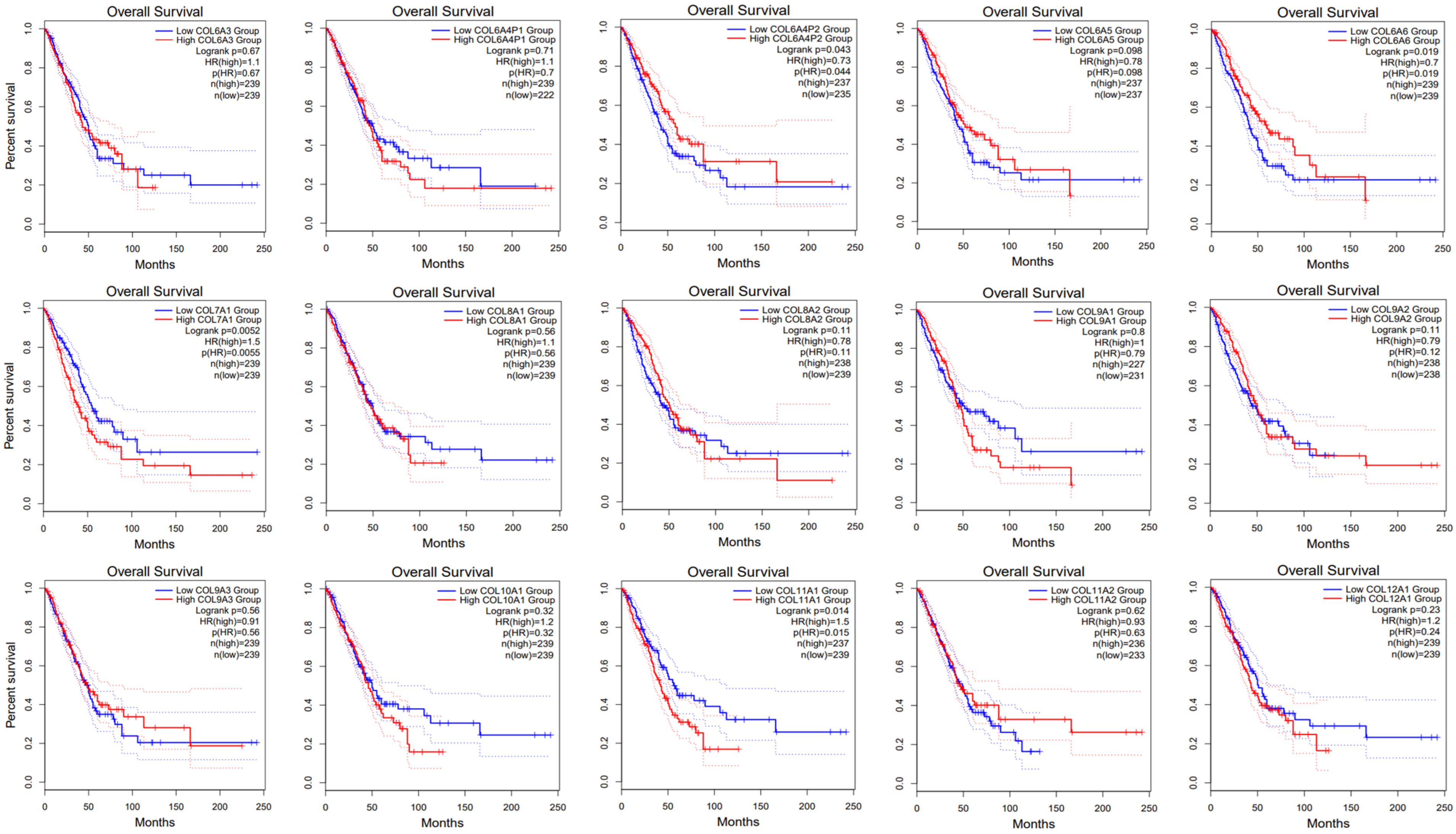
Characteristic	Patient data n:483 (%)
LUAD patients	
Gender	
Male	221 (45.7%)
Female	262 (54.2%)
Diagnosis median age, years	60.5 (range, 33-88)
Overall Survival Status	
Surviving	301 (62,3%)
Death	182 (37.6%)
Primary Diagnosis	
Adenocarcinoma, NOS	328 (67.9%)
Adenocarcinoma with mixed subtypes	108 (22.3%)
Bronchiolo-alveolar carcinoma, non-mucinous	19 (3.9%)
Bronchiolo-alveolar adenocarcinoma, NOS	3 (0.6%)
Papillary adenocarcinoma, NOS	22 (4.5%)
Micropapillary carcinoma, NOS	3 (0.6%)
Tissue or organ of origin	
Upper lobe, lung	282 (58.3%)
Middle lobe, lung	21 (4.3%)
Lower lobe, lung	160 (33.1%)
Overlapping lesion of lung	4 (0.8%)
Lung, NOS	14 (2.8%)
Main bronchus	2 (0.41%)
AJCC Pathologic_M	
M0	330 (68.3%)
MX	125 (25.8%)
M1	24 (4.9%)
ND	4 (0.8%)
AJCC Pathologic_N	
N0	311 (64.3%)
NX	11 (2.2%)
N1	89 (18.4%)
N2	69 (14.2%)
N3	2 (0.41%)
ND	1 (0.2%)
AJCC Pathologic_stage	
Stage IA	125 (25.8%)
Stage IB	131 (27.1%)
Stage IIA	46 (9.5%)
Stage IIB	64 (13.2%)
Stage IIIA	67 (13.8%)
Stage IIIB	11 (2.2%)
Stage IV	25 (5.1%)
ND	14 (2.8%)
AJCC Pathologic_T	
TX	3 (0.6%)

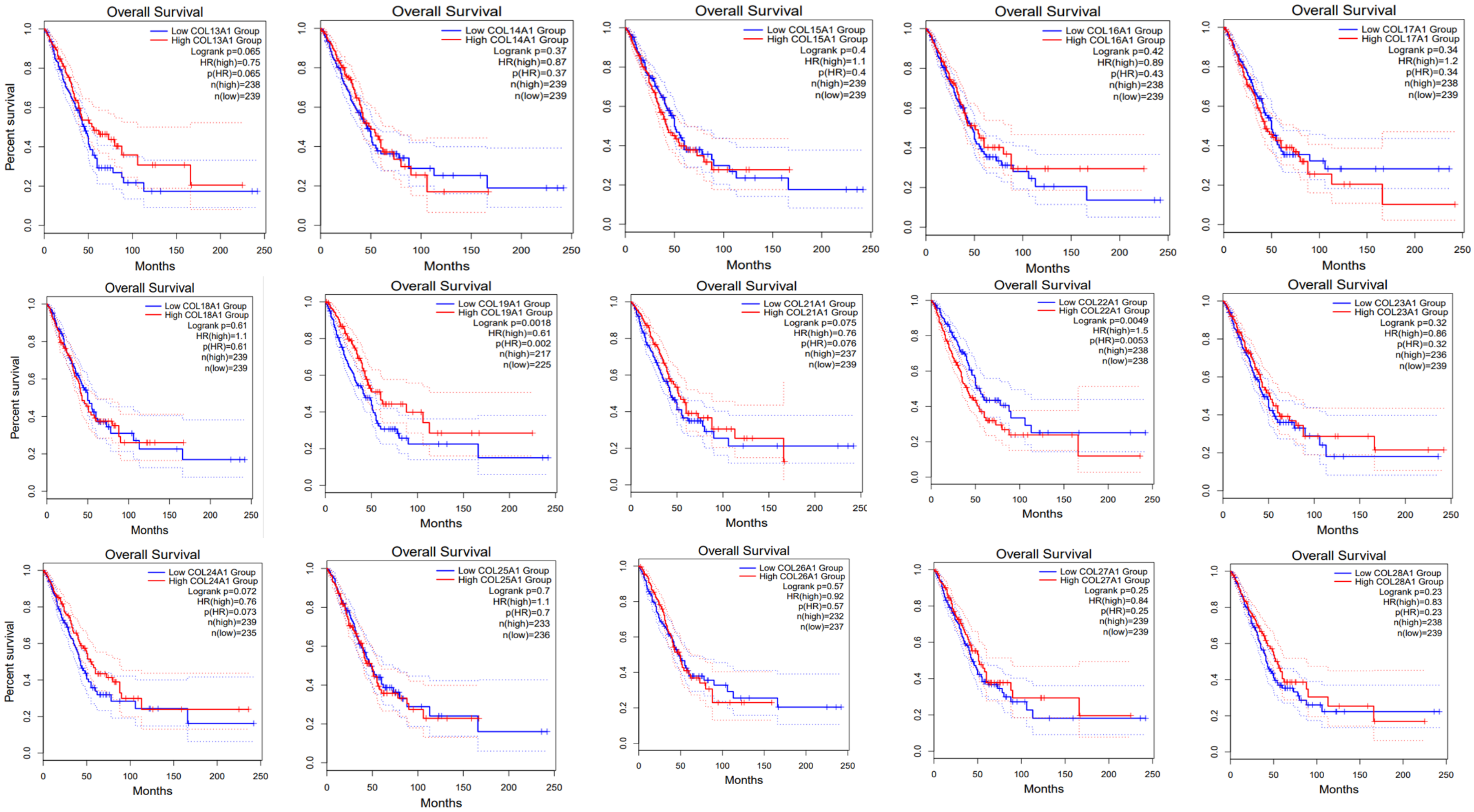
T1	162 (33.5%)
T2	258 (53.4%)
T3	41 (8.4%)
T4	19 (3.9%)
LUSC patients [Patient data n:*489 (%)]	
Gender	
Male	361
Female	128
Diagnosis median age, years	64.5 (range, 39-90)
Overall Survival Status	
Surviving	276
Death	213
Primary Diagnosis	
Squamous cell carcinoma, NOS	455
Squamous cell carcinoma, keratinizing, NOS	12
Squamous cell carcinoma, large cell, nonkeratinizing, NOS	3
Squamous cell carcinoma, small cell, nonkeratinizing	1
Basaloid squamous cell carcinoma	14
Papillary squamous cell carcinoma	4
Tissue or organ of origin	
Upper lobe, lung	255
Middle lobe, lung	15
Lower lobe, lung	169
Overlapping lesion of lung	8
Lung, NOS	35
Main bronchus	7
AJCC Pathologic_M	
M0	401
MX	77
M1	7
ND	4
AJCC Pathologic_N	
N0	312
NX	6
N1	128
N2	38
N3	5
AJCC Pathologic_stage	
Stage I	3
Stage IA	87
Stage IB	148
Stage II	3
Stage IIA	64
Stage IIB	91
Stage III	3
Stage IIIA	60
Stage IIIB	19
Stage IV	7
ND	4
AJCC Pathologic_T	
T1	48
T1a	24
T1b	39
T2	167
T1Ia	84
T1Ib	34
T3	69
T4	24

[ND: Not defined; NOS: Not otherwise specified.

*Survival analysis were performed with available 486 LUSC patient data from 489 LUSC patients given above which GEPIA exracts from TGCA database]

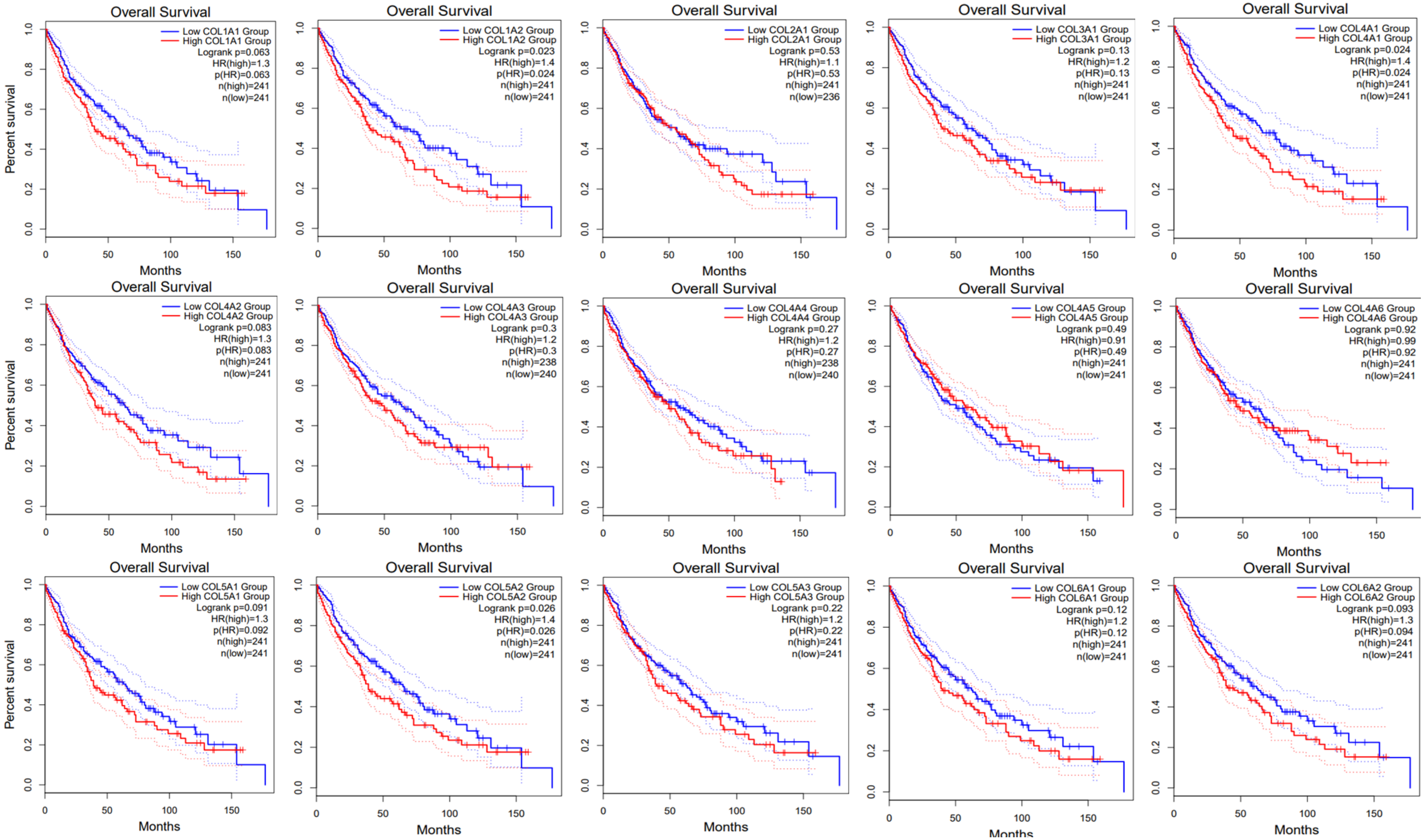

Overall Survival
Low COL1A1 Group
High COL1A1 Group
Logrank p=0.26
HR(high)=1.2
p(HR)=0.26
n(high)=239
n(low)=239
Overall Survival
Low COL1A2 Group
High COL1A2 Group
Logrank p=0.18
HR(high)=1.2
p(HR)=0.18
n(high)=239
n(low)=239
Overall Survival
Low COL2A1 Group
High COL2A1 Group
Logrank p=0.7
HR(high)=0.94
p(HR)=0.7
n(high)=228
n(low)=226
Overall Survival
Low COL3A1 Group
High COL3A1 Group
Logrank p=0.52
HR(high)=1.1
p(HR)=0.52
n(high)=239
n(low)=239
Overall Survival
Low COL4A1 Group
High COL4A1 Group
Logrank p=0.43
HR(high)=1.1
p(HR)=0.43
n(high)=239
n(low)=239
Overall Survival
Low COL4A2 Group
High COL4A2 Group
Logrank p=0.65
HR(high)=1.1
p(HR)=0.64
n(high)=239
n(low)=239
Overall Survival
Low COL4A3 Group
High COL4A3 Group
Logrank p=0.25
HR(high)=0.84
p(HR)=0.26
n(high)=239
n(low)=239
Overall Survival
Low COL4A4 Group
High COL4A4 Group
Logrank p=0.25
HR(high)=0.84
p(HR)=0.25
n(high)=239
n(low)=239
Overall Survival
Low COL4A5 Group
High COL4A5 Group
Logrank p=0.87
HR(high)=1
p(HR)=0.87
n(high)=239
n(low)=239
Overall Survival
Low COL4A6 Group
High COL4A6 Group
Logrank p=0.74
HR(high)=1.1
p(HR)=0.74
n(high)=239
n(low)=229
Overall Survival
Low COL5A1 Group
High COL5A1 Group
Logrank p=0.26
HR(high)=1.2
p(HR)=0.26
n(high)=239
n(low)=239
Overall Survival
Low COL5A2 Group
High COL5A2 Group
Logrank p=0.079
HR(high)=1.3
p(HR)=0.08
n(high)=239
n(low)=239
Overall Survival
Low COL5A3 Group
High COL5A3 Group
Logrank p=0.82
HR(high)=1
p(HR)=0.82
n(high)=239
n(low)=239
Overall Survival
Low COL6A1 Group
High COL6A1 Group
Logrank p=0.074
HR(high)=1.3
p(HR)=0.074
n(high)=239
n(low)=239
Overall Survival
Low COL6A2 Group
High COL6A2 Group
Logrank p=0.62
HR(high)=1.1
p(HR)=0.62
n(high)=239
n(low)=239
Percent survival
0.0
0.2
0.4
0.6
0.8
1.0
Months
0
50
100
150
200
250

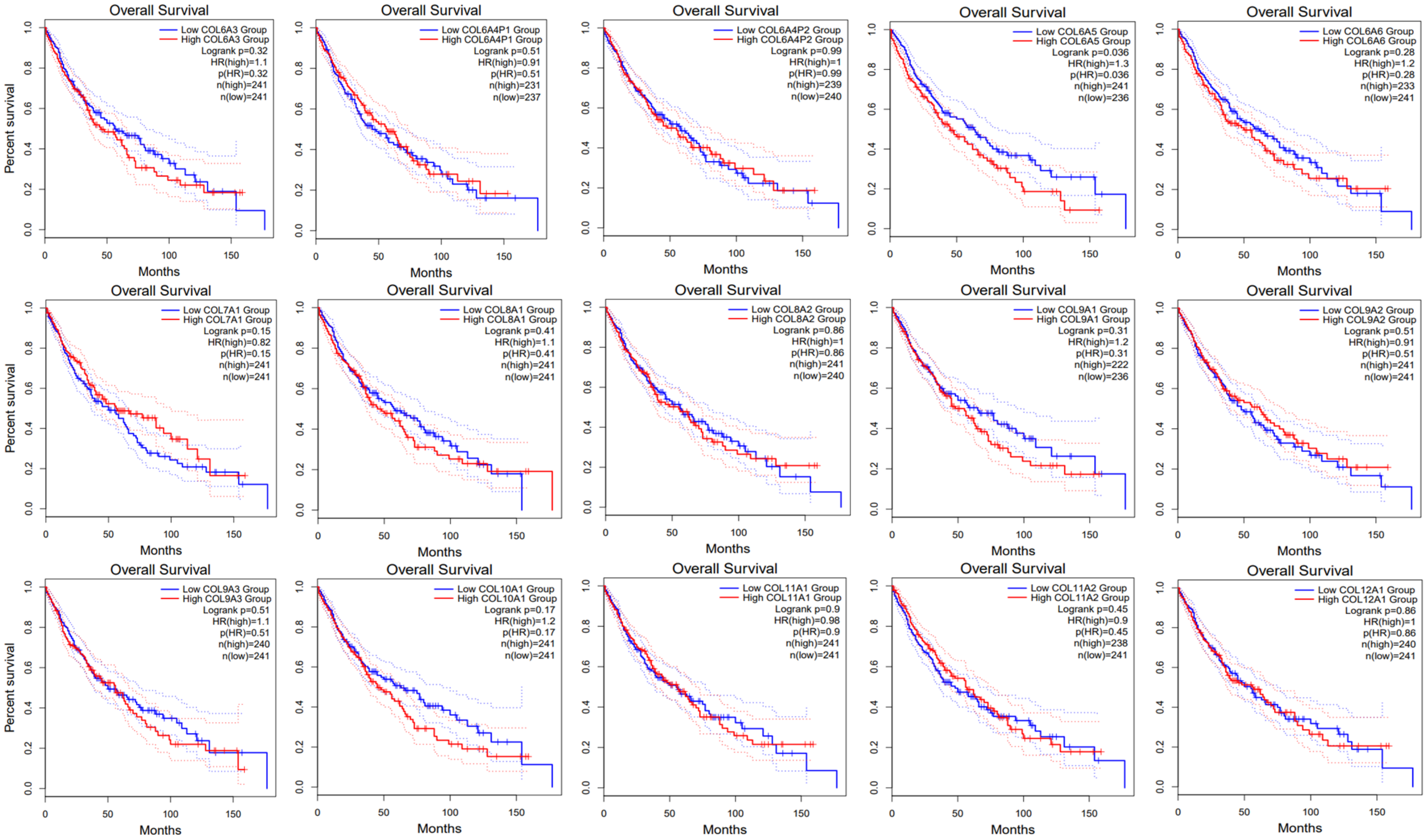

Overall Survival
Low COL6A3 Group
High COL6A3 Group
Logrank p=0.67
HR(high)=1.1
p(HR)=0.67
n(high)=239
n(low)=239
Overall Survival
Low COL6A4P1 Group
High COL6A4P1 Group
Logrank p=0.71
HR(high)=1.1
p(HR)=0.7
n(high)=239
n(low)=222
Overall Survival
Low COL6A4P2 Group
High COL6A4P2 Group
Logrank p=0.043
HR(high)=0.73
p(HR)=0.044
n(high)=237
n(low)=235
Overall Survival
Low COL6A5 Group
High COL6A5 Group
Logrank p=0.098
HR(high)=0.78
p(HR)=0.098
n(high)=237
n(low)=237
Overall Survival
Low COL6A6 Group
High COL6A6 Group
Logrank p=0.019
HR(high)=0.7
p(HR)=0.019
n(high)=239
n(low)=239
Overall Survival
Low COL7A1 Group
High COL7A1 Group
Logrank p=0.0052
HR(high)=1.5
p(HR)=0.0055
n(high)=239
n(low)=239
Overall Survival
Low COL8A1 Group
High COL8A1 Group
Logrank p=0.56
HR(high)=1.1
p(HR)=0.56
n(high)=239
n(low)=239
Overall Survival
Low COL8A2 Group
High COL8A2 Group
Logrank p=0.11
HR(high)=0.78
p(HR)=0.11
n(high)=238
n(low)=239
Overall Survival
Low COL9A1 Group
High COL9A1 Group
Logrank p=0.8
HR(high)=1
p(HR)=0.79
n(high)=227
n(low)=231
Overall Survival
Low COL9A2 Group
High COL9A2 Group
Logrank p=0.11
HR(high)=0.79
p(HR)=0.12
n(high)=238
n(low)=238
Overall Survival
Low COL9A3 Group
High COL9A3 Group
Logrank p=0.56
HR(high)=0.91
p(HR)=0.56
n(high)=239
n(low)=239
Overall Survival
Low COL10A1 Group
High COL10A1 Group
Logrank p=0.32
HR(high)=1.2
p(HR)=0.32
n(high)=239
n(low)=239
Overall Survival
Low COL11A1 Group
High COL11A1 Group
Logrank p=0.014
HR(high)=1.5
p(HR)=0.015
n(high)=237
n(low)=239
Overall Survival
Low COL11A2 Group
High COL11A2 Group
Logrank p=0.62
HR(high)=0.93
p(HR)=0.63
n(high)=236
n(low)=233
Overall Survival
Low COL12A1 Group
High COL12A1 Group
Logrank p=0.23
HR(high)=1.2
p(HR)=0.24
n(high)=239
n(low)=239
Percent survival
0.0
0.2
0.4
0.6
0.8
1.0
Months
0
50
100
150
200
250

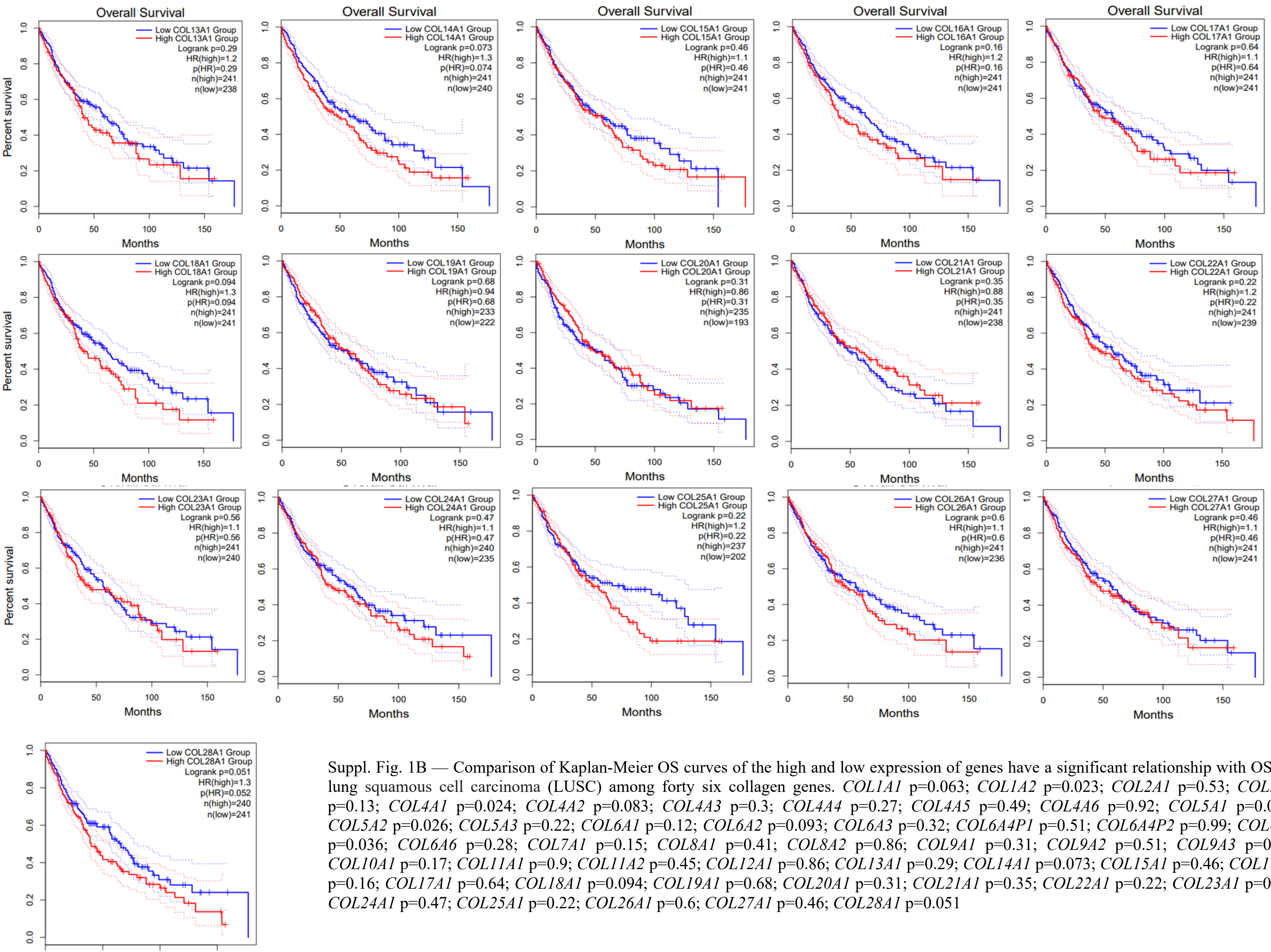


Suppl. Fig. 1A — Comparison of Kaplan-Meier OS curves of the high and low expression of genes have a significant relationship with OS for lung adenocarcinoma (LUAD) among forty five collagen genes*. *COL1A1* p=0.26; *COL1A2* p=0.18; *COL2A1* p=0.7; *COL3A1* p=0.52; *COL4A1* p=0.43; *COL4A2* p=0.65; *COL4A3* p=0.25; *COL4A4* p=0.25; *COL4A5* p=0.87; *COL4A6* p=0.74; *COL5A1* p=0.26; *COL5A2* p=0.079; *COL5A3* p=0.82; *COL6A1* p=0.074; *COL6A2* p=0.62; *COL6A3* p=0.67; *COL6A4P1* p=0.71; *COL6A4P2* p=0.043; *COL6A5* p=0.098; *COL6A6* p=0.019; *COL7A1* p=0.0052; *COL8A1* p=0.56; *COL8A2* p=0.11; *COL9A1* p=0.8; *COL9A2* p=0.11; *COL9A3* p=0.56; *COL10A1* p=0.32; *COL11A1* p=0.014; *COL11A2* p=0.62; *COL12A1* p=0.23; *COL13A1* p=0.065; *COL14A1* p=0.37; *COL15A1* p=0.4; *COL16A1* p=0.42; *COL17A1* p=0.34; *COL18A1* p=0.61; *COL19A1* p=0.0018; *COL21A1* p=0.075; *COL22A1* p=0.0049; *COL23A1* p=0.32; *COL24A1* p=0.072; *COL25A1* p=0.7; *COL26A1* p=0.57; *COL27A1* p=0.25; *COL28A1* p=0.23

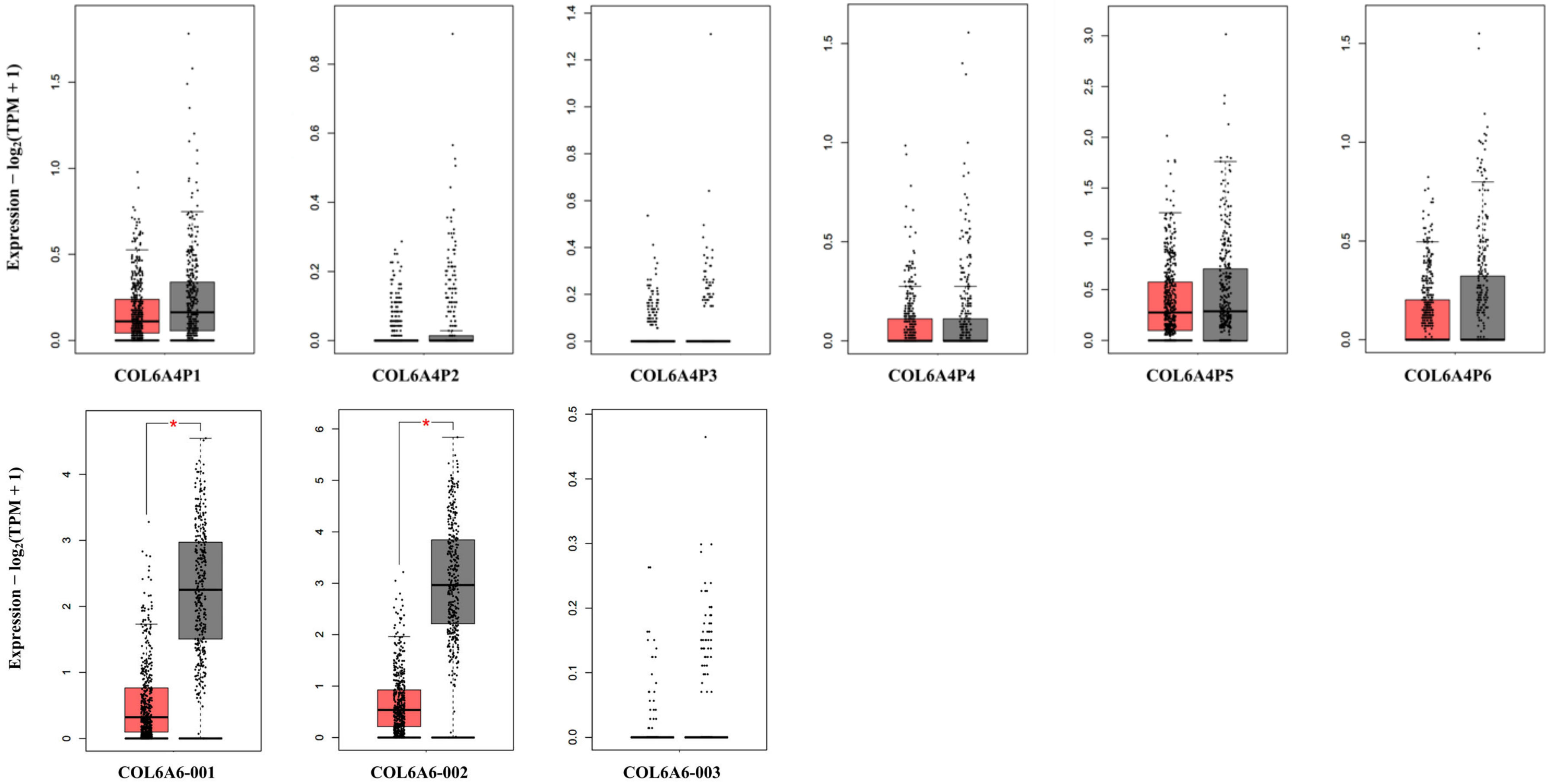
[*Since *COL20A1* gene isn’t significant in terms of survival, GEPIA2 does not form a graph. Therefore, it is not included in the analysis]

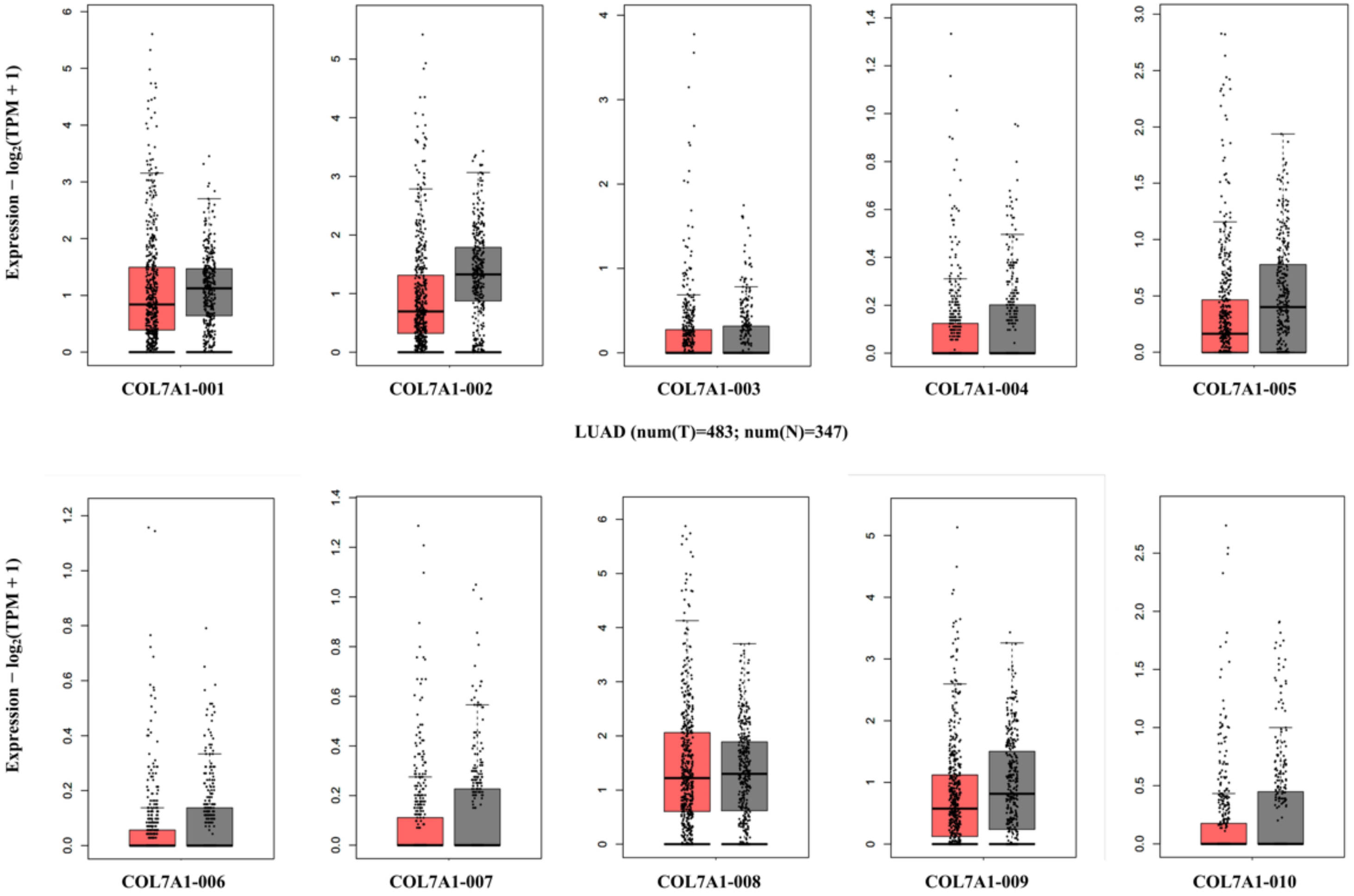

Overall Survival
Low COL1A1 Group
High COL1A1 Group
Logrank p=0.063
HR(high)=1.3
p(HR)=0.063
n(high)=241
n(low)=241
Overall Survival
Low COL1A2 Group
High COL1A2 Group
Logrank p=0.023
HR(high)=1.4
p(HR)=0.024
n(high)=241
n(low)=241
Overall Survival
Low COL2A1 Group
High COL2A1 Group
Logrank p=0.53
HR(high)=1.1
p(HR)=0.53
n(high)=241
n(low)=236
Overall Survival
Low COL3A1 Group
High COL3A1 Group
Logrank p=0.13
HR(high)=1.2
p(HR)=0.13
n(high)=241
n(low)=241
Overall Survival
Low COL4A1 Group
High COL4A1 Group
Logrank p=0.024
HR(high)=1.4
p(HR)=0.024
n(high)=241
n(low)=241
Overall Survival
Low COL4A2 Group
High COL4A2 Group
Logrank p=0.083
HR(high)=1.3
p(HR)=0.083
n(high)=241
n(low)=241
Overall Survival
Low COL4A3 Group
High COL4A3 Group
Logrank p=0.3
HR(high)=1.2
p(HR)=0.3
n(high)=238
n(low)=240
Overall Survival
Low COL4A4 Group
High COL4A4 Group
Logrank p=0.27
HR(high)=1.2
p(HR)=0.27
n(high)=238
n(low)=240
Overall Survival
Low COL4A5 Group
High COL4A5 Group
Logrank p=0.49
HR(high)=0.91
p(HR)=0.49
n(high)=241
n(low)=241
Overall Survival
Low COL4A6 Group
High COL4A6 Group
Logrank p=0.92
HR(high)=0.99
p(HR)=0.92
n(high)=241
n(low)=241
Overall Survival
Low COL5A1 Group
High COL5A1 Group
Logrank p=0.091
HR(high)=1.3
p(HR)=0.092
n(high)=241
n(low)=241
Overall Survival
Low COL5A2 Group
High COL5A2 Group
Logrank p=0.026
HR(high)=1.4
p(HR)=0.026
n(high)=241
n(low)=241
Overall Survival
Low COL5A3 Group
High COL5A3 Group
Logrank p=0.22
HR(high)=1.2
p(HR)=0.22
n(high)=241
n(low)=241
Overall Survival
Low COL6A1 Group
High COL6A1 Group
Logrank p=0.12
HR(high)=1.2
p(HR)=0.12
n(high)=241
n(low)=241
Overall Survival
Low COL6A2 Group
High COL6A2 Group
Logrank p=0.093
HR(high)=1.3
p(HR)=0.094
n(high)=241
n(low)=241
Percent survival
0.0
0.2
0.4
0.6
0.8
1.0
Months
0
50
100
150

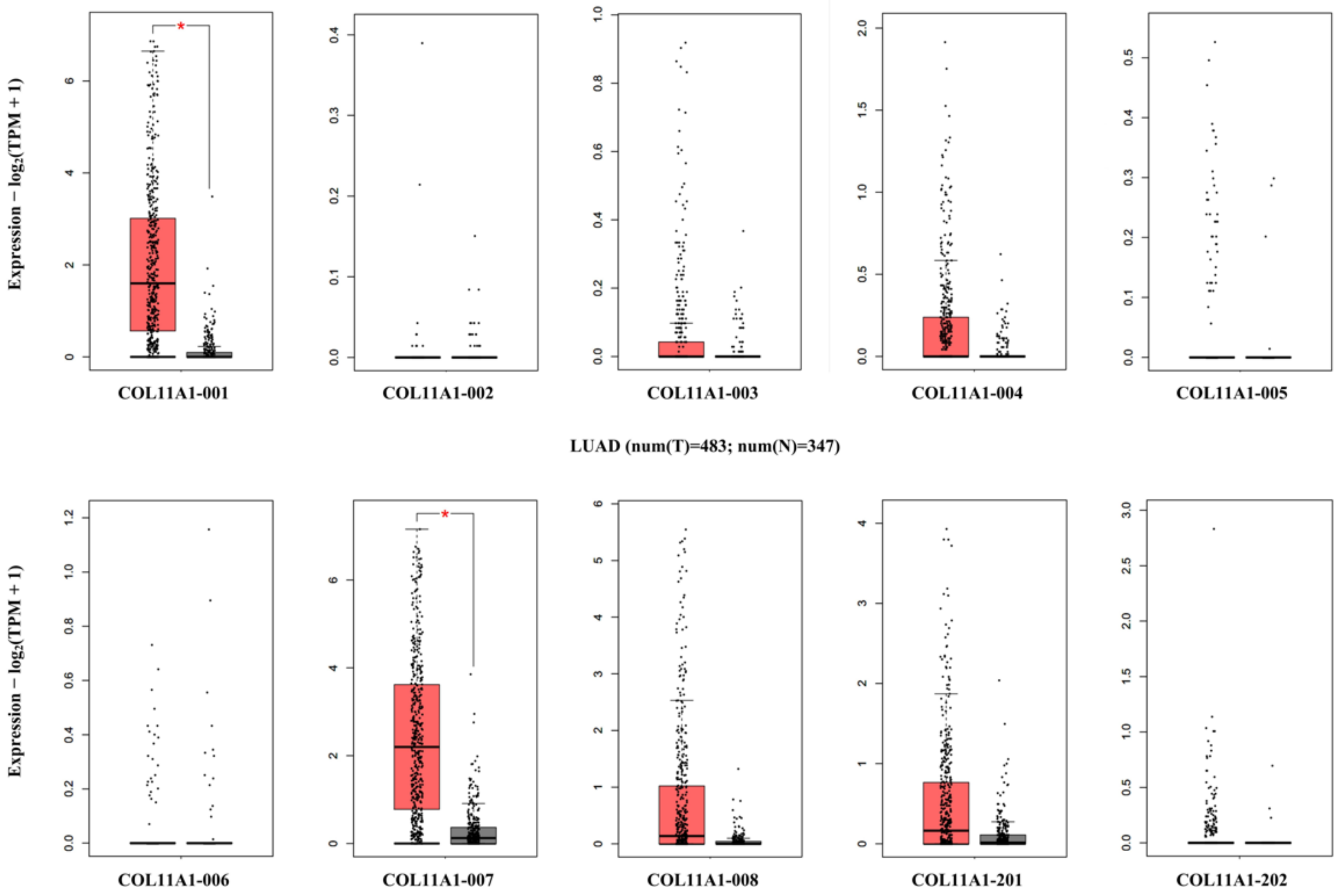

Overall Survival
Low COL6A3 Group
High COL6A3 Group
Logrank p=0.32
HR(high)=1.1
p(HR)=0.32
n(high)=241
n(low)=241
Overall Survival
Low COL6A4P1 Group
High COL6A4P1 Group
Logrank p=0.51
HR(high)=0.91
p(HR)=0.51
n(high)=231
n(low)=237
Overall Survival
Low COL6A4P2 Group
High COL6A4P2 Group
Logrank p=0.99
HR(high)=1
p(HR)=0.99
n(high)=239
n(low)=240
Overall Survival
Low COL6A5 Group
High COL6A5 Group
Logrank p=0.036
HR(high)=1.3
p(HR)=0.036
n(high)=241
n(low)=236
Overall Survival
Low COL6A6 Group
High COL6A6 Group
Logrank p=0.28
HR(high)=1.2
p(HR)=0.28
n(high)=233
n(low)=241
Overall Survival
Low COL7A1 Group
High COL7A1 Group
Logrank p=0.15
HR(high)=0.82
p(HR)=0.15
n(high)=241
n(low)=241
Overall Survival
Low COL8A1 Group
High COL8A1 Group
Logrank p=0.41
HR(high)=1.1
p(HR)=0.41
n(high)=241
n(low)=241
Overall Survival
Low COL8A2 Group
High COL8A2 Group
Logrank p=0.86
HR(high)=1
p(HR)=0.86
n(high)=241
n(low)=240
Overall Survival
Low COL9A1 Group
High COL9A1 Group
Logrank p=0.31
HR(high)=1.2
p(HR)=0.31
n(high)=222
n(low)=236
Overall Survival
Low COL9A2 Group
High COL9A2 Group
Logrank p=0.51
HR(high)=0.91
p(HR)=0.51
n(high)=241
n(low)=241
Overall Survival
Low COL9A3 Group
High COL9A3 Group
Logrank p=0.51
HR(high)=1.1
p(HR)=0.51
n(high)=240
n(low)=241
Overall Survival
Low COL10A1 Group
High COL10A1 Group
Logrank p=0.17
HR(high)=1.2
p(HR)=0.17
n(high)=241
n(low)=241
Overall Survival
Low COL11A1 Group
High COL11A1 Group
Logrank p=0.9
HR(high)=0.98
p(HR)=0.9
n(high)=241
n(low)=241
Overall Survival
Low COL11A2 Group
High COL11A2 Group
Logrank p=0.45
HR(high)=0.9
p(HR)=0.45
n(high)=238
n(low)=241
Overall Survival
Low COL12A1 Group
High COL12A1 Group
Logrank p=0.86
HR(high)=1
p(HR)=0.86
n(high)=240
n(low)=241
Percent survival
0.0
0.2
0.4
0.6
0.8
1.0
Months
0
50
100
150

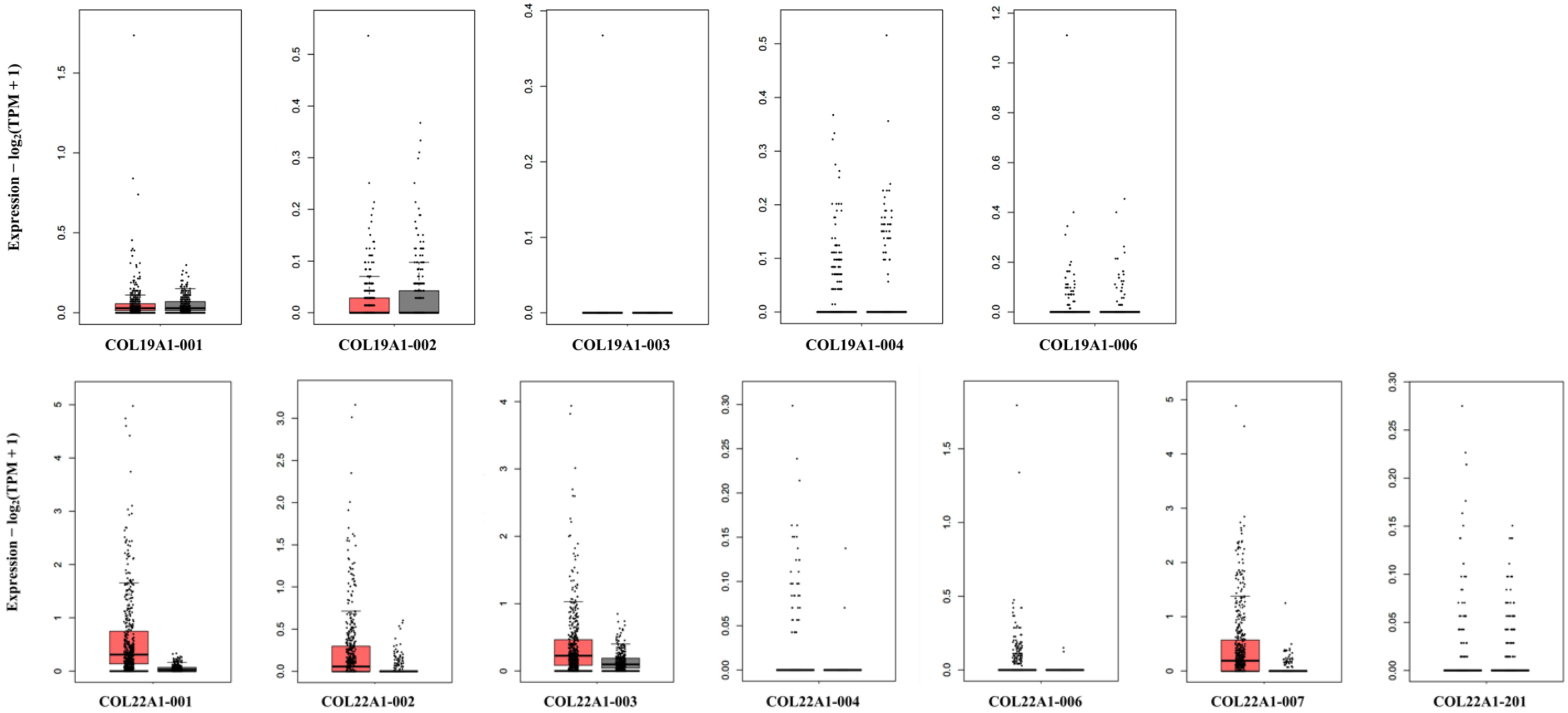


Suppl. Fig. 1B — Comparison of Kaplan-Meier OS curves of the high and low expression of genes have a significant relationship with OS for lung squamous cell carcinoma (LUSC) among forty six collagen genes. *COL1A1* p=0.063; *COL1A2* p=0.023; *COL2A1* p=0.53; *COL3A1* p=0.13; *COL4A1* p=0.024; *COL4A2* p=0.083; *COL4A3* p=0.3; *COL4A4* p=0.27; *COL4A5* p=0.49; *COL4A6* p=0.92; *COL5A1* p=0.091; *COL5A2* p=0.026; *COL5A3* p=0.22; *COL6A1* p=0.12; *COL6A2* p=0.093; *COL6A3* p=0.32; *COL6A4P1* p=0.51; *COL6A4P2* p=0.99; *COL6A5* p=0.036; *COL6A6* p=0.28; *COL7A1* p=0.15; *COL8A1* p=0.41; *COL8A2* p=0.86; *COL9A1* p=0.31; *COL9A2* p=0.51; *COL9A3* p=0.51; *COL10A1* p=0.17; *COL11A1* p=0.9; *COL11A2* p=0.45; *COL12A1* p=0.86; *COL13A1* p=0.29; *COL14A1* p=0.073; *COL15A1* p=0.46; *COL16A1* p=0.16; *COL17A1* p=0.64; *COL18A1* p=0.094; *COL19A1* p=0.68; *COL20A1* p=0.31; *COL21A1* p=0.35; *COL22A1* p=0.22; *COL23A1* p=0.56; *COL24A1* p=0.47; *COL25A1* p=0.22; *COL26A1* p=0.6; *COL27A1* p=0.46; *COL28A1* p=0.051

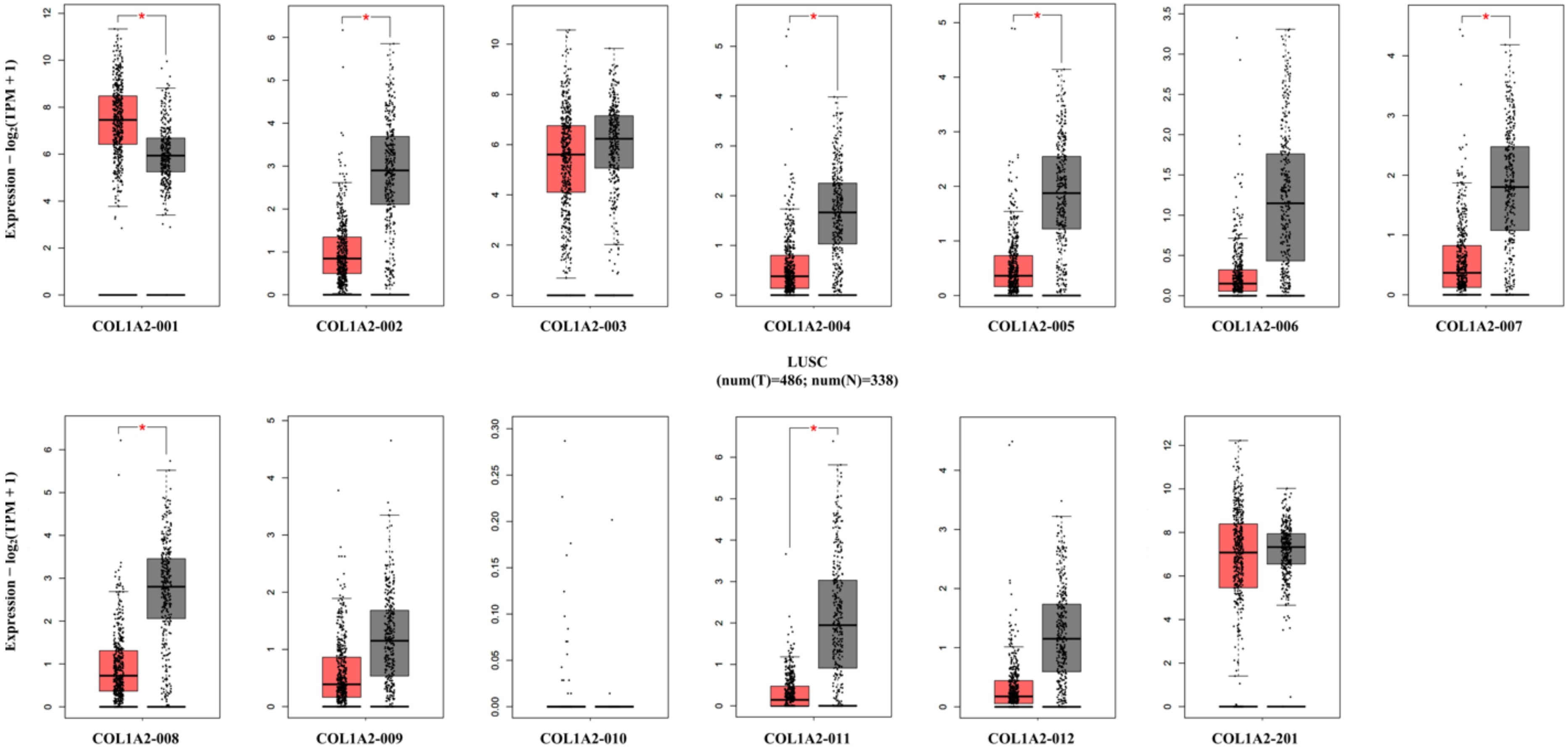

Expression − log2(TPM + 1)
0.0
0.5
1.0
1.5
COL6A4P1
0.0
0.2
0.4
0.6
0.8
COL6A4P2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
COL6A4P3
0.0
0.5
1.0
1.5
COL6A4P4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
COL6A4P5
0.0
0.5
1.0
1.5
COL6A4P6
Expression − log2(TPM + 1)
0
1
2
3
4
*
COL6A6-001
0
1
2
3
4
5
6
*
COL6A6-002
0.0
0.1
0.2
0.3
0.4
0.5
COL6A6-003

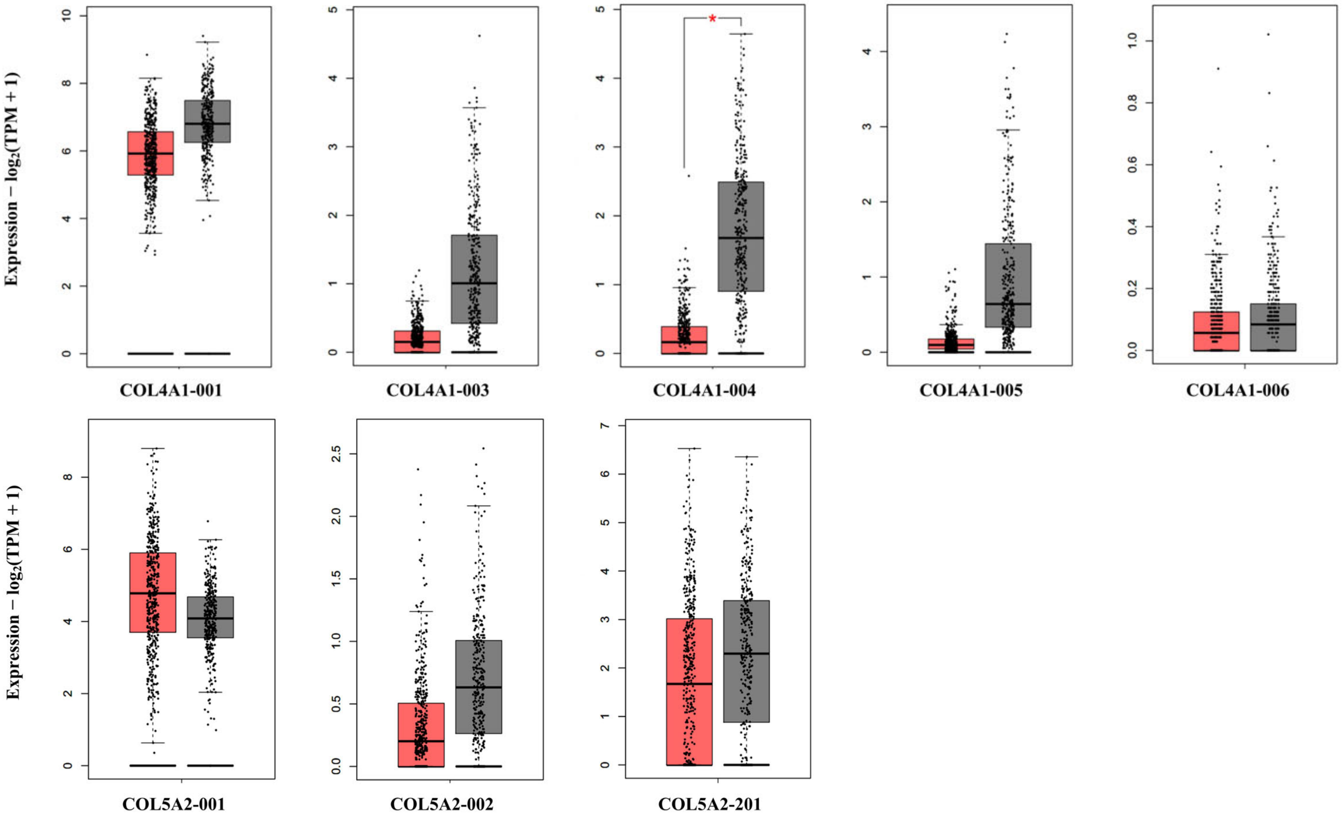

Expression − log2(TPM + 1)
COL7A1-001
COL7A1-002
COL7A1-003
COL7A1-004
COL7A1-005
LUAD (num(T)=483; num(N)=347)
Expression − log2(TPM + 1)
COL7A1-006
COL7A1-007
COL7A1-008
COL7A1-009
COL7A1-010

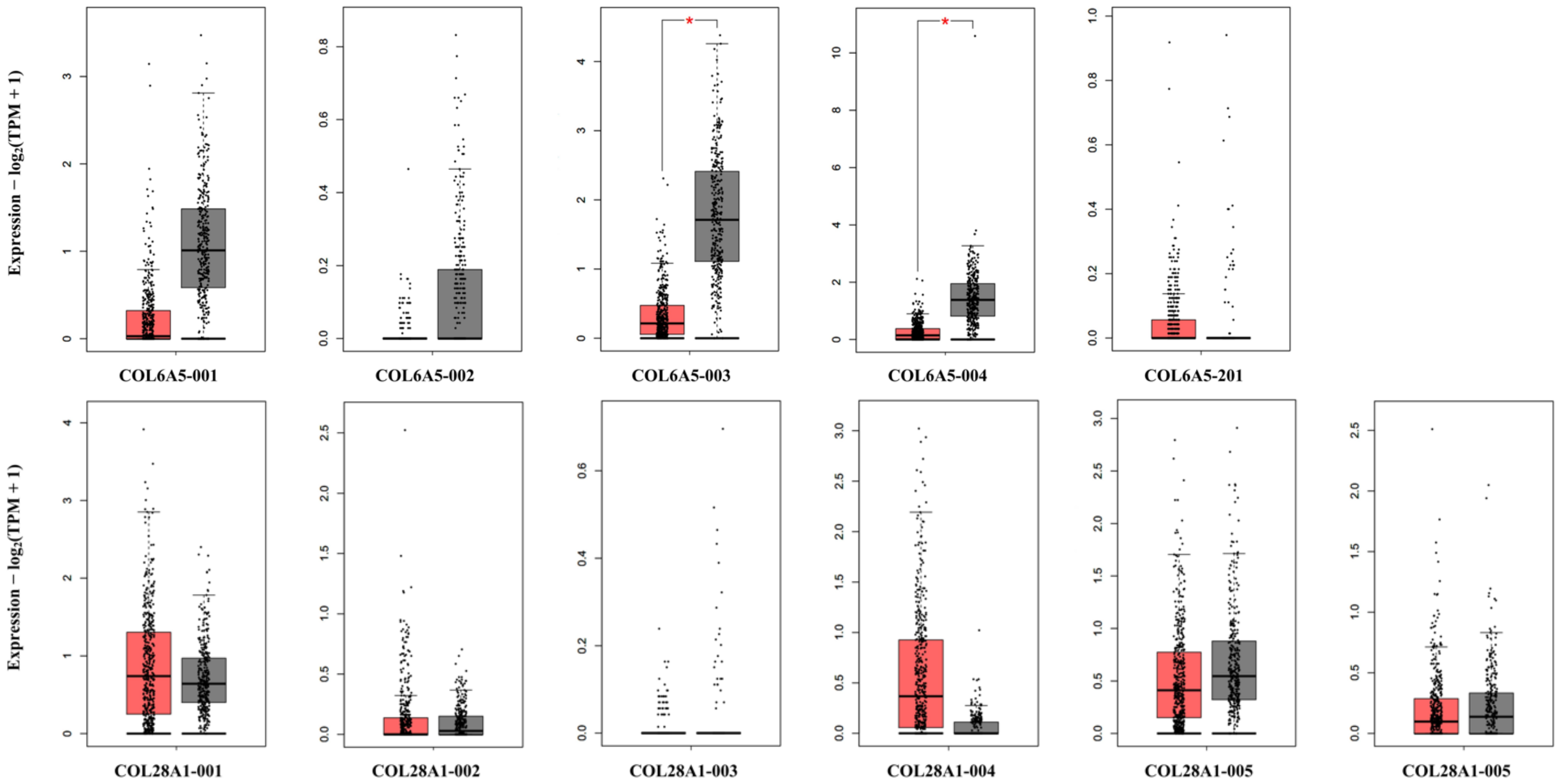

Expression − $\log_2$(TPM + 1)
COL11A1-001
COL11A1-002
COL11A1-003
COL11A1-004
COL11A1-005
LUAD (num(T)=483; num(N)=347)
Expression − $\log_2$(TPM + 1)
COL11A1-006
COL11A1-007
COL11A1-008
COL11A1-201
COL11A1-202
*
*



Suppl. Fig. 2A — Comparative analysis of the tissue-specific differential expression of *COL6A4P*, *COL6A6*, *COL7A1*, *COL11A1*, *COL19A1* and *COL22A1* transcripts in lung adenocarcinoma (LUAD)* using GEPIA2. [*Fourty one box plots are based on 483 LUAD samples (marked in red) and 347 normal samples (marked in grey). (*P <0.05)]


Expression − log$_2$(TPM + 1)
*
COL1A2-001
COL1A2-002
COL1A2-003
COL1A2-004
COL1A2-005
COL1A2-006
COL1A2-007
LUSC
(num(T)=486; num(N)=338)
Expression − log$_2$(TPM + 1)
COL1A2-008
COL1A2-009
COL1A2-010
COL1A2-011
COL1A2-012
COL1A2-201


Expression − $\log_2$(TPM + 1)
0
2
4
6
8
10
COL4A1-001
0
1
2
3
4
5
COL4A1-003
0
1
2
3
4
5
*
COL4A1-004
0
1
2
3
4
COL4A1-005
0.0
0.2
0.4
0.6
0.8
1.0
COL4A1-006
Expression − $\log_2$(TPM + 1)
0
2
4
6
8
COL5A2-001
0.0
0.5
1.0
1.5
2.0
2.5
COL5A2-002
0
1
2
3
4
5
6
7
COL5A2-201



Suppl. Fig. 2B — Comparative analysis of the tissue-specific differential expression of *COL1A2*, *COL5A2*, *COL6A5* and *COL28A1* transcripts in lung squamous cell carcinoma (LUSC)* using GEPIA2. [*Fourty one box plots are based on 486 LUSC samples (marked in red) and 338 normal samples (marked in grey). (*P <0.05)]