Prognostic and biologic relevance of collagen biosynthesis pathway in prostate cancer

Bioinformatic methods, figures and tables

2023-07-27

# Tables

Table 1: Characteristic of the analyzed cohorts. Numeric variables are presented as medians with interquartile ranges (IQR) and ranges. Qualitative variables are presented as percentages of categories within the complete observation set.

| **Variablea** | **GSE16560** | **GSE40272** | **GSE70768** | **GSE70769** | **TCGA** |
| --- | --- | --- | --- | --- | --- |
| Age at diagnosis, years | 74 [IQR: 69 - 79] range: 51 - 91 n = 281 | 62 [IQR: 56 - 65] range: 43 - 73 n = 84 | 62 [IQR: 56 - 65] range: 41 - 73 n = 112 |  | 61 [IQR: 56 - 66] range: 41 - 78 n = 495 |
| PSA at diagnosis |  | 5.1 [IQR: 4.2 - 7.4] range: 2.1 - 44 n = 82 | 7.8 [IQR: 5.9 - 10] range: 3.2 - 24 n = 111 | 8 [IQR: 5.9 - 11] range: 1.5 - 120 n = 90 | 0.1 [IQR: 0.03 - 0.11] range: 0 - 320 n = 438 |
| Clinical stage |  |  | T1: 56% (n = 61) T1/T2: 0.92% (n = 1) T2: 28% (n = 30) T2N0M0: 0.92% (n = 1) T3: 15% (n = 16) n = 109 | T1: 46% (n = 41) T2: 44% (n = 39) T3: 10% (n = 9) n = 89 |  |
| Pathological tumor stage |  | T2: 76% (n = 63) T3: 23% (n = 19) T4: 1.2% (n = 1) n = 83 | T2: 31% (n = 34) T3: 68% (n = 76) T4: 0.9% (n = 1) n = 111 | T2: 52% (n = 46) T3: 48% (n = 42) n = 88 | T2: 38% (n = 176) T3: 62% (n = 283) n = 459 |
| Pathological node stage |  | N0: 94% (n = 78) N1: 6% (n = 5) n = 83 | N0: 91% (n = 82) N1: 8.9% (n = 8) n = 90 | N0: 95% (n = 18) NK: 5.3% (n = 1) n = 19 | N0: 82% (n = 344) N1: 18% (n = 78) n = 422 |
| Pathological metastasis stage |  | M0: 99% (n = 83) M1: 1.2% (n = 1) n = 84 | M0: 86% (n = 6) M1: 14% (n = 1) n = 7 | M0: 87% (n = 26) M1: 13% (n = 4) n = 30 | M0: 99% (n = 453) M1: 0.66% (n = 3) n = 456 |
| Gleason sum score | 7 [IQR: 6 - 8] range: 6 - 10 n = 281 | 7 [IQR: 7 - 7] range: 6 - 9 n = 83 | 7 [IQR: 7 - 7] range: 6 - 9 n = 113 | 7 [IQR: 7 - 7] range: 5 - 10 n = 91 | 7 [IQR: 7 - 9] range: 6 - 10 n = 495 |
| 6: 30% (n = 83) 7: 42% (n = 117) 8: 9.6% (n = 27) 9: 17% (n = 49) 10: 1.8% (n = 5) n = 281 | 6: 16% (n = 13) 7: 72% (n = 60) 8: 4.8% (n = 4) 9: 7.2% (n = 6) n = 83 | 6: 15% (n = 17) 7: 76% (n = 86) 8: 7.1% (n = 8) 9: 1.8% (n = 2) n = 113 | 5: 2.2% (n = 2) 6: 20% (n = 18) 7: 62% (n = 56) 8: 5.5% (n = 5) 9: 9.9% (n = 9) 10: 1.1% (n = 1) n = 91 | 6: 9.1% (n = 45) 7: 50% (n = 246) 8: 13% (n = 63) 9: 28% (n = 137) 10: 0.81% (n = 4) n = 495 |
| Positive surgical margins |  | 13% (n = 11) n = 84 | 23% (n = 26) n = 112 | 45% (n = 42) n = 93 |  |
| Extracapsular extension |  |  | 69% (n = 77) n = 112 | 46% (n = 42) n = 91 |  |
| Death | 73% (n = 206) n = 281 |  |  |  | 1.6% (n = 8) n = 495 |
| Overall survival, months | 100 [IQR: 52 - 150] range: 6 - 270 n = 281 |  |  |  | 17 [IQR: 5.6 - 33] range: 0.033 - 150 n = 495 |
| Relapse |  | 23% (n = 19) n = 82 | 17% (n = 19) n = 112 | 48% (n = 45) n = 93 | 14% (n = 58) n = 427 |
| Relapse-free survival, months |  | 30 [IQR: 16 - 53] range: 0.033 - 92 n = 82 | 30 [IQR: 17 - 49] range: 1 - 65 n = 111 | 58 [IQR: 19 - 80] range: 0.36 - 100 n = 92 | 17 [IQR: 7.1 - 33] range: 0.92 - 150 n = 418 |
| aPSA: prostate-specific antigen. | | | | | |

Table 2: Collagen genes of interest and their classification.

| **Gene symbol** | **Entrez ID** | **Gene groupa** |
| --- | --- | --- |
| *ALDH18A1* | 5832 | proline pathway |
| *PYCR1* | 5831 | proline pathway |
| *PEPD* | 5184 | proline pathway |
| *COL3A1* | 1281 | ECM component |
| *COL5A2* | 1290 | ECM component |
| *COL1A2* | 1278 | ECM component |
| *COL1A1* | 1277 | ECM component |
| *LAMB3* | 3914 | ECM component |
| *COL4A2* | 1284 | ECM component |
| *COL4A1* | 1282 | ECM component |
| *COL4A5* | 1287 | ECM component |
| *COL4A3* | 1285 | ECM component |
| *MMP7* | 4316 | ECM processing |
| *COL14A1* | 7373 | ECM component |
| *COL10A1* | 1300 | ECM component |
| *COL6A3* | 1293 | ECM component |
| *ITGA6* | 3655 | ECM component |
| *CD151* | 977 | ECM component |
| *COL9A3* | 1299 | ECM component |
| *PCOLCE* | 5118 | ECM processing |
| *LOXL2* | 4017 | ECM processing |
| *LOXL1* | 4016 | ECM processing |
| *LOX* | 4015 | ECM processing |
| *COL21A1* | 81578 | ECM component |
| *COL19A1* | 1310 | ECM component |
| *PLOD3* | 8985 | ECM processing |
| *P4HB* | 5034 | ECM processing |
| *PPIB* | 5479 | ECM processing |
| aECM: extracellular matrix. | | |

# Figures

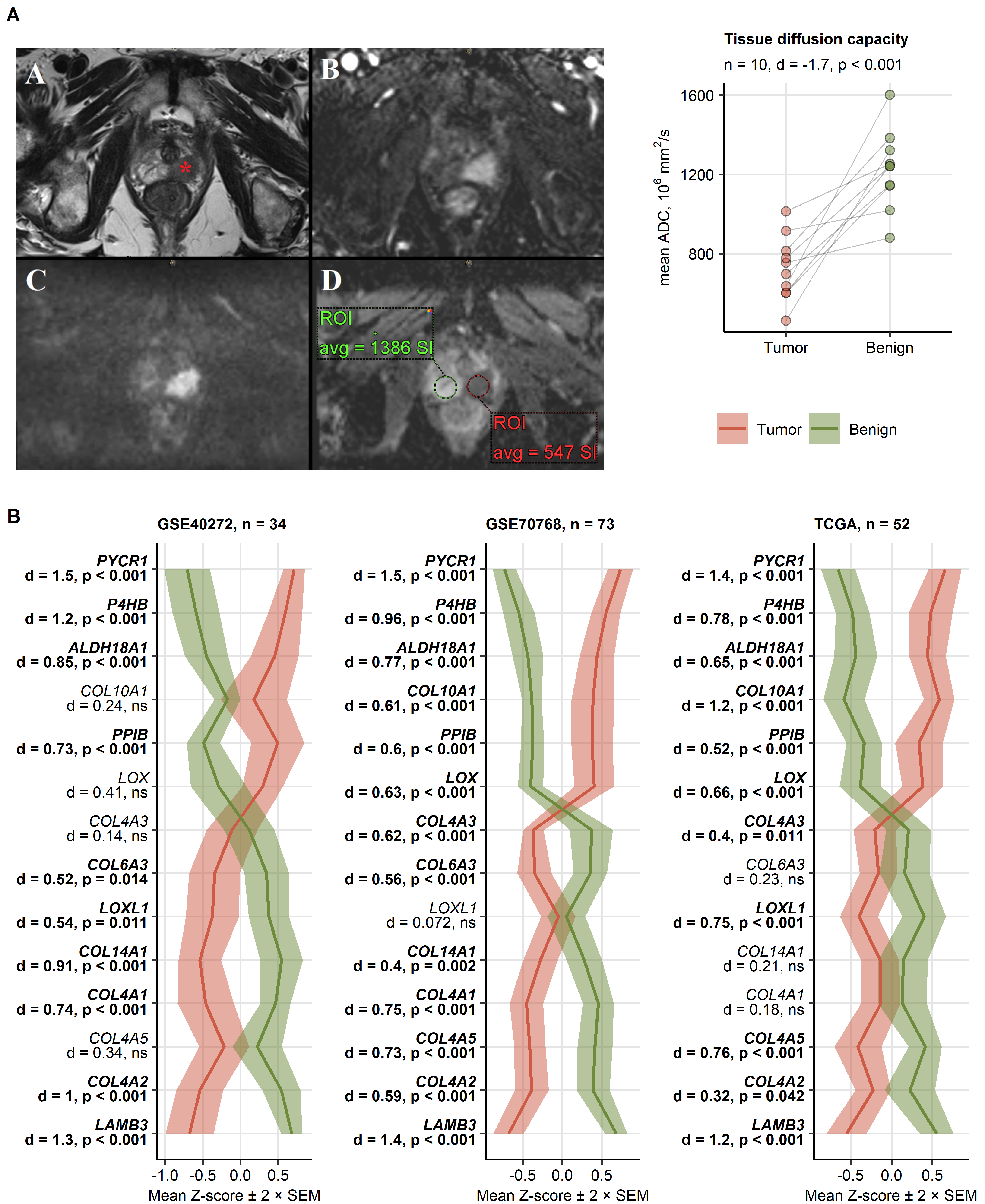


Figure 1: Differences in tissue diffusion capacity and expression of the collagen pathway genes between the prostate cancer and benign tissue.

**Figure 1. Differences in tissue diffusion capacity and expression of the collagen pathway genes between the prostate cancer and benign tissue.**

*(A) Representative MRI axial sequences of the prostate of an 83-year-old patient with prostate cancer (left panel). The T2-weighted images (a) show a 23mm suspicious hypo-intense lesion in the left peripheral zone (red star) of the prostate, which has an early enhancement in the contrast-enhanced T1-weighted image (b). The DWI scan (c) and ADC map (d) show a strong diffusion restriction (black and white character) of the tumor area suggesting a PI-RADS 5 lesion. Statistical significance for differences in diffusion capacity in patient-matched tumor and benign tissue measured by ADC was determined by paired T test with Cohen’s d effect size statistic (right panel). Single ADC values are visualized as points, grey lines connect measurements of the same donors. The number of measurement pairs, effect size and p value are displayed in the plot caption.*

*(B) Differences in -transformed expression levels of 28 genes related to collagen metabolism between donor-matched tumor and benign prostate tissue were investigated by paired T test with Cohen’s d effect size statistic in three published cohorts. P values were corrected for multiple testing with the false discovery method. Results for gene significantly regulated in at least two cohorts are presented. Lines represent means, 2 SEM (standard error of the mean) intervals are visualized as tinted ribbons. Gene symbols, effect sizes and p values are shown in the Y axes. Significant effects are highlighted in bold. Numbers of tissue pairs are displayed in the plot captions.*

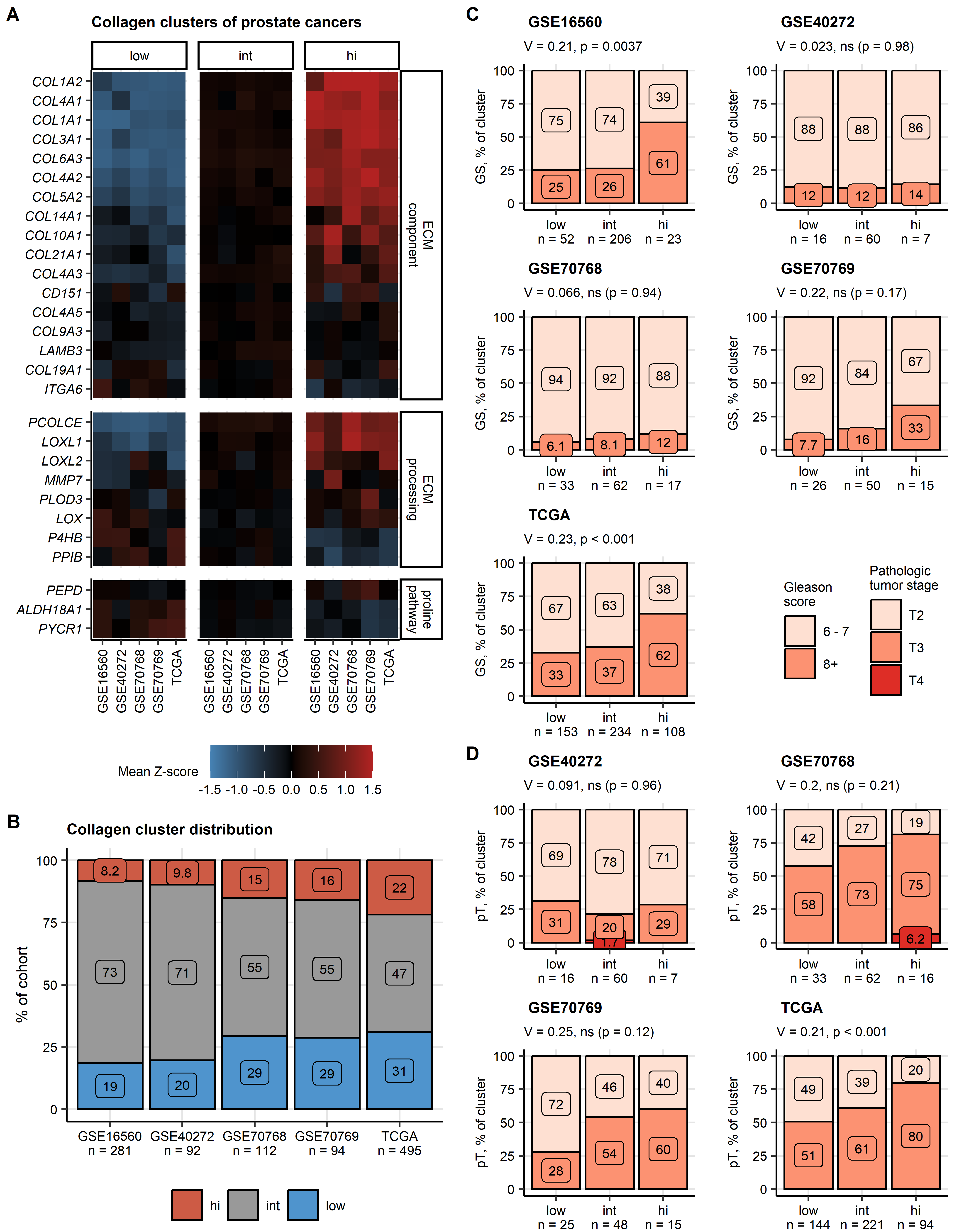


Figure 2: Collagen clusters of prostate carcinoma and their clinical characteristic.

**Figure 2. Collagen clusters of prostate carcinoma and their clinical characteristic.**

*Prostate cancers in the training TCGA cohort were clustered in respect to normalized -transformed expression levels of the collagen pathway genes of interest by the PAM algorithm with Manhattan distance between the observations. Three clusters, collagenlow (low), collagenintermediate (int) and collagenhigh (hi) were defined. Cancer samples from the test cohorts (GSE16560, GSE40272, GSE70768, GSE70769) were assigned to the collagen clusters with an inverse distance weighted k-nearest neighbor classifier.*

*(A) Mean normalized -transformed expression levels of the collagen pathway genes in the collagen clusters and cohorts presented in a heat map.*

*(B) Distribution of the collagen clusters expressed as percentage of the cohort cancer samples visualized in a stack plot. Total numbers of cancer samples are displayed in the X axis.*

*(C, D) Distribution of Gleason scores (GS, C) and pathological tumor stages (pT, D) in the collagen clusters. Statistical significance was determined by test with Cramer’s V effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Gleason score and tumor stage distributions are shown in stack plots. Effect sizes and p values are displayed in the plot captions. Numbers of observations in the clusters are indicated in the X axes.*

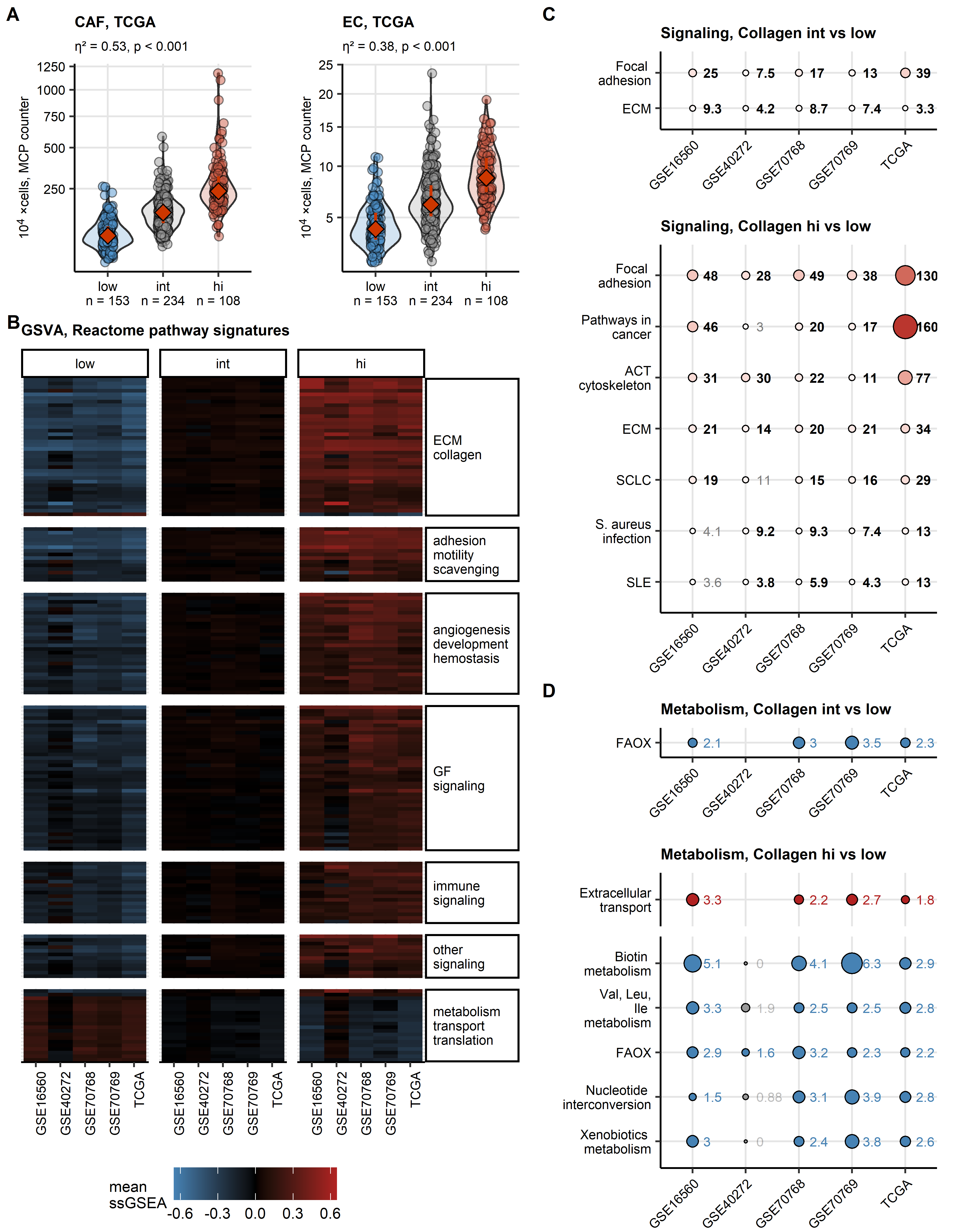


Figure 3: Infiltration, biological processes, signaling and metabolic pathways differentially regulated in the collagen clusters.

**Figure 3. Infiltration, biological processes, signaling and metabolic pathways differentially regulated in the collagen clusters.**

*(A) Counts of cancer-associated fibroblasts (CAF) and endothelial cells (EC) estimated by the MCP counter algorithm were found to differ significantly between the collagen clusters in all investigated collectives. Statistical significance was assessed by Kruskal-Wallis test with effect size statistic. P values were corrected for multiple comparison with the false discovery rate method. Counts of fibroblasts and endothelial cells in the collagen clusters in the TCGA cohort are shown in violin plots. Red diamonds with whiskers represented medians with interquartile ranges. Single tumor samples are visualized as points. Effect sizes and p values are displayed in the plot captions. Numbers of observations in the clusters are indicated in the X axes. Full analysis results are presented in Supplementary Table S4.*

*(B) Gene set variation analysis (GSVA) with Reactome pathway gene signatures. Statistical significance for differences in signature single sample gene set enrichment analysis scores (ssGSEA) between the collagen clusters was determined by one-way ANOVA with effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Mean ssGSEA scores in the clusters and cohorts for signatures significantly regulated in at least four cohorts are presented in a heat map. Full analysis results are presented in Supplementary Table S6.*

*(C) Modulation of KEGG-listed signaling pathways in the collagenhigh (hi) or collagenintermediate (hi) as compared with collagenlow tumors was predicted by the SPIA algorithm based on differential gene expression estimates. Predicted magnitude of pathway modulation (tA parameter) for pathways significantly modulated in at least four cohorts was visualized in bubble plots (tA < 0: inhibition, blue; tA > 0, activation, red). Points are labeled with their tA values. Significant effects are highlighted in bold. Full analysis results are presented in Supplementary Table S8.*

*(D) Modulation of Recon2 model metabolic reactions in the collagenhigh (hi) or collagenintermediate (hi) as compared with collagenlow tumors was predicted by the BiGGR and biggrExtra algorithms based on differential gene expression estimates. Enrichment of significantly activated and significantly inhibited reactions in the collagenhigh or collagenintermediate clusters within the Recon model metabolic subsystems was investigated by Fisher’s exact test. Enrichment p values were corrected for multiple testing with the false discovery rate method. Enrichment odds ratios for activated (red) and inhibited reaction (blue) for significantly enriched metabolic subsystems shared by at least four cohorts are presented in bubble plots. Points are labeled with their odds ration values. Full analysis results are presented in Supplementary Table S10.*

# References