Prognostic and biologic relevance of collagen biosynthesis pathway in prostate cancer

Supplementary material, transcriptome part

2023-07-27

# Supplementary Tables

Table 1: Expression of the collagen pathway genes in the malignant and benign tissue compared by paired T test with Cohen's d effect size statistic. P values were corrected for multiple testing with the false discovery rate method. log2-transformed expression values are presented as medians with interquartile ranges (IQR) and ranges. The table is available as a supplementary Excel file.

|  |
| --- |
|  |

Table 2: Expression of the cluster-defining collagen pathway genes in the collagen clusters of prostate cancer. Statistical significance was assessed by one-way ANOVA with eta-squared effect size statistic. P values were corrected for multiple testing with the false discovery rate method. log2-transformed expression values are presented as medians with interquartile ranges (IQR) and ranges. The table is available as a supplementary Excel file.

|  |
| --- |
|  |

Table 3: Clinical characteristic of the collagen clusters. Numeric variables are presented as medians with interquartile ranges (IQR) and ranges. Nominal variables are presented as percentages and counts of categories within the cluster.

| **Cohort** | **Variablea** | **Collagen low** | **Collagen intermediate** | **Collagen high** | **Significanceb** | **Effect sizeb** |
| --- | --- | --- | --- | --- | --- | --- |
| GSE16560 | Age at diagnosis, years | 74 [IQR: 72 - 79] range: 57 - 91 n = 52 | 74 [IQR: 69 - 80] range: 51 - 91 n = 206 | 74 [IQR: 69 - 79] range: 57 - 89 n = 23 | ns (p = 0.79) | η² = -0.0055 |
| Gleason sum score | 7 [IQR: 6 - 7.2] range: 6 - 10 n = 52 | 7 [IQR: 6 - 8] range: 6 - 10 n = 206 | 9 [IQR: 7 - 9] range: 6 - 10 n = 23 | p = 0.014 | η² = 0.026 |
| 5 - 6: 33% (n = 17) 7: 42% (n = 22) 8+: 25% (n = 13) n = 52 | 5 - 6: 30% (n = 61) 7: 44% (n = 91) 8+: 26% (n = 54) n = 206 | 5 - 6: 22% (n = 5) 7: 17% (n = 4) 8+: 61% (n = 14) n = 23 | p = 0.014 | V = 0.15 |
| GSE40272 | Age at diagnosis, years | 62 [IQR: 54 - 64] range: 46 - 71 n = 17 | 62 [IQR: 57 - 65] range: 43 - 73 n = 60 | 61 [IQR: 58 - 68] range: 50 - 73 n = 7 | ns (p = 0.8) | η² = -0.014 |
| PSA at diagnosis | 4.8 [IQR: 3.8 - 5.7] range: 3 - 10 n = 15 | 5.6 [IQR: 4.3 - 7.5] range: 2.1 - 44 n = 60 | 4.4 [IQR: 4 - 7.1] range: 2.8 - 16 n = 7 | ns (p = 0.8) | η² = -0.0074 |
| Pathological tumor stage | T2: 69% (n = 11) T3: 31% (n = 5) T4: 0% (n = 0) n = 16 | T2: 78% (n = 47) T3: 20% (n = 12) T4: 1.7% (n = 1) n = 60 | T2: 71% (n = 5) T3: 29% (n = 2) T4: 0% (n = 0) n = 7 | ns (p = 0.85) | V = 0.091 |
| Pathological node stage | N0: 94% (n = 15) N1: 6.2% (n = 1) n = 16 | N0: 97% (n = 58) N1: 3.3% (n = 2) n = 60 | N0: 71% (n = 5) N1: 29% (n = 2) n = 7 | ns (p = 0.13) | V = 0.29 |
| Pathological metastasis stage | M0: 94% (n = 16) M1: 5.9% (n = 1) n = 17 | M0: 100% (n = 60) M1: 0% (n = 0) n = 60 | M0: 100% (n = 7) M1: 0% (n = 0) n = 7 | ns (p = 0.41) | V = 0.22 |
| Gleason sum score | 7 [IQR: 7 - 7] range: 6 - 9 n = 16 | 7 [IQR: 7 - 7] range: 6 - 9 n = 60 | 7 [IQR: 6.5 - 7] range: 6 - 9 n = 7 | ns (p = 0.8) | η² = -0.011 |
| Positive surgical margins | 18% (n = 3) n = 17 | 13% (n = 8) n = 60 | 0% (n = 0) n = 7 | ns (p = 0.8) | V = 0.13 |
| Gleason sum score | 5 - 6: 6.2% (n = 1) 7: 81% (n = 13) 8+: 12% (n = 2) n = 16 | 5 - 6: 17% (n = 10) 7: 72% (n = 43) 8+: 12% (n = 7) n = 60 | 5 - 6: 29% (n = 2) 7: 57% (n = 4) 8+: 14% (n = 1) n = 7 | ns (p = 0.8) | V = 0.11 |
| GSE70768 | Age at diagnosis, years | 62 [IQR: 58 - 65] range: 48 - 73 n = 33 | 62 [IQR: 56 - 64] range: 41 - 73 n = 62 | 62 [IQR: 55 - 66] range: 47 - 69 n = 16 | ns (p = 0.94) | η² = -0.017 |
| PSA at diagnosis | 7 [IQR: 5.7 - 10] range: 4 - 17 n = 33 | 7.8 [IQR: 6.1 - 9.7] range: 3.6 - 18 n = 61 | 9.4 [IQR: 6.4 - 12] range: 3.2 - 24 n = 16 | ns (p = 0.94) | η² = -0.0065 |
| Clinical stage | T1: 62% (n = 20) T1/T2: 0% (n = 0) T2: 22% (n = 7) T2N0M0: 3.1% (n = 1) T3: 12% (n = 4) n = 32 | T1: 56% (n = 34) T1/T2: 1.6% (n = 1) T2: 31% (n = 19) T2N0M0: 0% (n = 0) T3: 11% (n = 7) n = 61 | T1: 44% (n = 7) T1/T2: 0% (n = 0) T2: 25% (n = 4) T2N0M0: 0% (n = 0) T3: 31% (n = 5) n = 16 | ns (p = 0.94) | V = 0.19 |
| Pathological tumor stage | T2: 42% (n = 14) T3: 58% (n = 19) T4: 0% (n = 0) n = 33 | T2: 27% (n = 17) T3: 73% (n = 45) T4: 0% (n = 0) n = 62 | T2: 19% (n = 3) T3: 75% (n = 12) T4: 6.2% (n = 1) n = 16 | ns (p = 0.21) | V = 0.2 |
| Pathological node stage | N0: 92% (n = 23) N1: 8% (n = 2) n = 25 | N0: 90% (n = 46) N1: 9.8% (n = 5) n = 51 | N0: 93% (n = 13) N1: 7.1% (n = 1) n = 14 | ns (p = 0.94) | V = 0.038 |
| Pathological metastasis stage | M0: 100% (n = 2) M1: 0% (n = 0) n = 2 | M0: 100% (n = 4) M1: 0% (n = 0) n = 4 | M0: 0% (n = 0) M1: 100% (n = 1) n = 1 | ns (p = 0.17) | V = 1 |
| Gleason sum score | 7 [IQR: 7 - 7] range: 6 - 8 n = 33 | 7 [IQR: 7 - 7] range: 6 - 8 n = 62 | 7 [IQR: 7 - 7] range: 6 - 9 n = 17 | ns (p = 0.94) | η² = -0.013 |
| Positive surgical margins | 15% (n = 5) n = 33 | 26% (n = 16) n = 62 | 29% (n = 5) n = 17 | ns (p = 0.94) | V = 0.13 |
| Extracapsular extension | 70% (n = 23) n = 33 | 69% (n = 43) n = 62 | 65% (n = 11) n = 17 | ns (p = 0.94) | V = 0.037 |
| Gleason sum score | 5 - 6: 12% (n = 4) 7: 82% (n = 27) 8+: 6.1% (n = 2) n = 33 | 5 - 6: 18% (n = 11) 7: 74% (n = 46) 8+: 8.1% (n = 5) n = 62 | 5 - 6: 12% (n = 2) 7: 76% (n = 13) 8+: 12% (n = 2) n = 17 | ns (p = 0.94) | V = 0.074 |
| GSE70769 | PSA at diagnosis | 6.9 [IQR: 5 - 11] range: 3.5 - 35 n = 24 | 8.2 [IQR: 6.4 - 11] range: 2.2 - 120 n = 51 | 7.5 [IQR: 6.2 - 18] range: 1.5 - 42 n = 15 | ns (p = 0.67) | η² = -0.0084 |
| Clinical stage | T1: 54% (n = 13) T2: 33% (n = 8) T3: 12% (n = 3) n = 24 | T1: 46% (n = 23) T2: 44% (n = 22) T3: 10% (n = 5) n = 50 | T1: 33% (n = 5) T2: 60% (n = 9) T3: 6.7% (n = 1) n = 15 | ns (p = 0.68) | V = 0.12 |
| Pathological tumor stage | T2: 72% (n = 18) T3: 28% (n = 7) n = 25 | T2: 46% (n = 22) T3: 54% (n = 26) n = 48 | T2: 40% (n = 6) T3: 60% (n = 9) n = 15 | ns (p = 0.12) | V = 0.25 |
| Pathological node stage | N0: 75% (n = 3) NK: 25% (n = 1) n = 4 | N0: 100% (n = 11) NK: 0% (n = 0) n = 11 | N0: 100% (n = 4) NK: 0% (n = 0) n = 4 | ns (p = 0.2) | V = 0.46 |
| Pathological metastasis stage | M0: 100% (n = 7) M1: 0% (n = 0) n = 7 | M0: 94% (n = 15) M1: 6.2% (n = 1) n = 16 | M0: 57% (n = 4) M1: 43% (n = 3) n = 7 | ns (p = 0.11) | V = 0.48 |
| Gleason sum score | 7 [IQR: 6 - 7] range: 5 - 9 n = 26 | 7 [IQR: 7 - 7] range: 5 - 10 n = 50 | 7 [IQR: 7 - 8] range: 6 - 9 n = 15 | ns (p = 0.11) | η² = 0.053 |
| Positive surgical margins | 46% (n = 12) n = 26 | 44% (n = 23) n = 52 | 47% (n = 7) n = 15 | ns (p = 0.98) | V = 0.021 |
| Extracapsular extension | 27% (n = 7) n = 26 | 52% (n = 26) n = 50 | 60% (n = 9) n = 15 | ns (p = 0.12) | V = 0.25 |
| Gleason sum score | 5 - 6: 35% (n = 9) 7: 58% (n = 15) 8+: 7.7% (n = 2) n = 26 | 5 - 6: 20% (n = 10) 7: 64% (n = 32) 8+: 16% (n = 8) n = 50 | 5 - 6: 6.7% (n = 1) 7: 60% (n = 9) 8+: 33% (n = 5) n = 15 | ns (p = 0.19) | V = 0.2 |
| TCGA | Age at diagnosis, years | 61 [IQR: 56 - 66] range: 46 - 78 n = 153 | 61 [IQR: 55 - 66] range: 41 - 77 n = 234 | 62 [IQR: 58 - 66] range: 44 - 75 n = 108 | ns (p = 0.51) | η² = -0.00078 |
| PSA at diagnosis | 0.1 [IQR: 0.03 - 0.1] range: 0 - 320 n = 134 | 0.1 [IQR: 0.03 - 0.1] range: 0 - 37 n = 209 | 0.1 [IQR: 0.03 - 0.34] range: 0 - 32 n = 95 | ns (p = 0.48) | η² = 9.9e-05 |
| Pathological tumor stage | T2: 49% (n = 71) T3: 51% (n = 73) n = 144 | T2: 39% (n = 86) T3: 61% (n = 135) n = 221 | T2: 20% (n = 19) T3: 80% (n = 75) n = 94 | p < 0.001 | V = 0.21 |
| Pathological node stage | N0: 81% (n = 100) N1: 19% (n = 23) n = 123 | N0: 85% (n = 172) N1: 15% (n = 30) n = 202 | N0: 74% (n = 72) N1: 26% (n = 25) n = 97 | ns (p = 0.12) | V = 0.11 |
| Pathological metastasis stage | M0: 99% (n = 138) M1: 0.72% (n = 1) n = 139 | M0: 100% (n = 221) M1: 0.45% (n = 1) n = 222 | M0: 99% (n = 94) M1: 1.1% (n = 1) n = 95 | ns (p = 0.83) | V = 0.029 |
| Gleason sum score | 7 [IQR: 7 - 8] range: 6 - 10 n = 153 | 7 [IQR: 7 - 9] range: 6 - 9 n = 234 | 8.5 [IQR: 7 - 9] range: 6 - 10 n = 108 | p < 0.001 | η² = 0.055 |
| 5 - 6: 10% (n = 16) 7: 57% (n = 87) 8+: 33% (n = 50) n = 153 | 5 - 6: 9.4% (n = 22) 7: 53% (n = 125) 8+: 37% (n = 87) n = 234 | 5 - 6: 6.5% (n = 7) 7: 31% (n = 34) 8+: 62% (n = 67) n = 108 | p < 0.001 | V = 0.16 |
| aPSA: prostate-specific antigen. | | | | | | |
| bCategorical variables: χ² test with Cramer's V effec size statistic. Numeric variables: Kruskal-Wallis test with η² effect size statistic. P values corrected for multiple testing with the false discovery rate method. | | | | | | |

Table 4: Non-malignant cell numbers predicted for the collagen Non-maignant cell fractions predicted for the collagen Statistical significance was assessed by Kruskal-Wallis test with eta-squared effect size statistic. P values were corrected for multiple testing with the false discovery method. The table is available as a supplementary Excel file.

|  |
| --- |
|  |

Table 5: clusters by the MCP counter algorithm. clusters by the xCell algorithm. Statistical significance was assessed by Kruskal-Wallis test with eta-squared effect size statistic. P values were corrected for multiple testing with the false discovery method. The table is available as a supplementary Excel file.

|  |
| --- |
|  |

Table 6: Gene set variation analysis with the Reactome pathway gene signatures. Differences between the collagen intermediate or high clusters versus collagen low cancers were investigated by one-way ANOVA with eta-squared effect size statistic and linear modeling. Results for signatures significantly regulated with moderate-to-large effect size (eta-squared at leat 0.06) in at least four cohorts are presented. P values were corrected for multiple testing with the false discovery rate method (FDR). The table is available as a supplementary Excel file.

|  |
| --- |
|  |

Table 7: Genes differentially expressed in the collagen intermediate or high cluster as compared with collagen low cancers were identified by one-way ANOVA and linear modeling with the 1.25-fold regulation cutoff P values were corrected for multiple testing with the false discovery rate method (FDR). The table is available as a supplementary Excel file.

|  |
| --- |
|  |

Table 8: Signaling pathway activity in the collagen clusters investigated by the SPIA algorithm. Resulat for signaling pathways significantly activated or inhibited in at least four cohorts are shown. The table is available as a supplementary Excel file.

|  |
| --- |
|  |

Table 9: Biochemical reactions predicted to be significantly activated in the collagen high or collagen low cluster as compared with collagen low cancers. Statistical significance was determined by Monte Carlo simulation. P values were corrected for multiple testing with the false discovery rate method. The table is available as a supplementary Excel file.

|  |
| --- |
|  |

Table 10: Results of enrichment analysis for significantly activated and inhibited biochemical reaction within the Recon metabolism subsystem. Statistical significance was determined by Fisher's exact test corrected for multiple testing with the false discovery rate method (FDR). The table is available as a supplementary Excel file.

|  |
| --- |
|  |

# Supplementary Figures

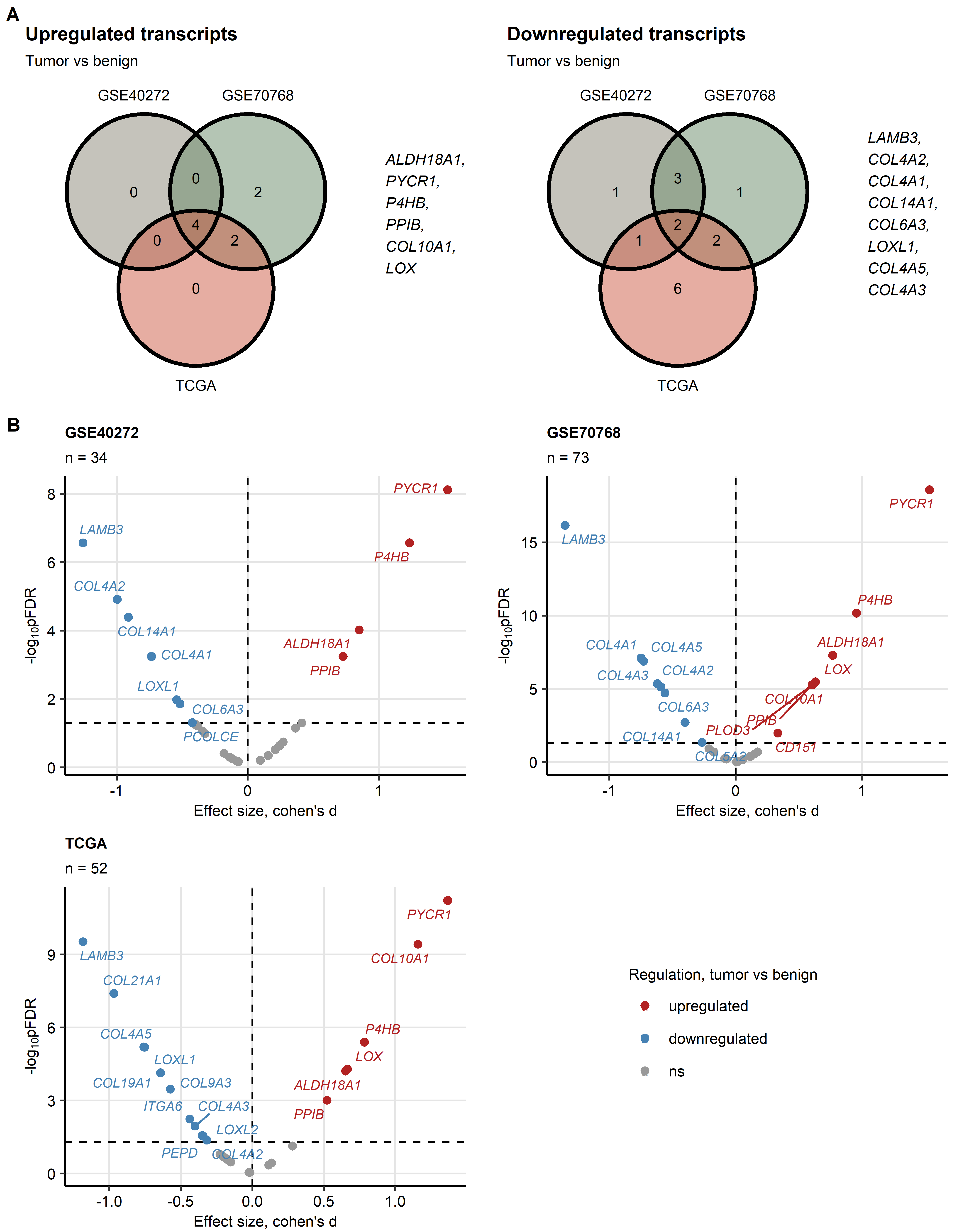


Figure 1: Expression of collagen pathway genes in the normal prostate and prostate cancer tissue.

**Supplementary Figure S1. Expression of collagen pathway genes in the normal prostate and prostate cancer tissue.**

*Differences in -transformed expression of 28 genes related to collagen metabolism between donor-matched pairs of the prostate cancer and benign tissue were assessed by paired T test with Cohen’s effect size statistic. P values were corrected for multiple testing with the false discovery rate (FDR) method.*

*(A) Numbers of significantly up- and downregulated genes in the tumor tissue as compared with the benign tissue in the investigated cohorts presented in Venn plots. Genes regulated in at least two cohorts are listed next to the plots.*

*(B) Gene expression regulation significance and effect size presented in Volcano plots. Points represent single genes. Significantly regulated genes are labeled with their symbols. Number of analyzed tissue pairs are displayed in the plot captions.*

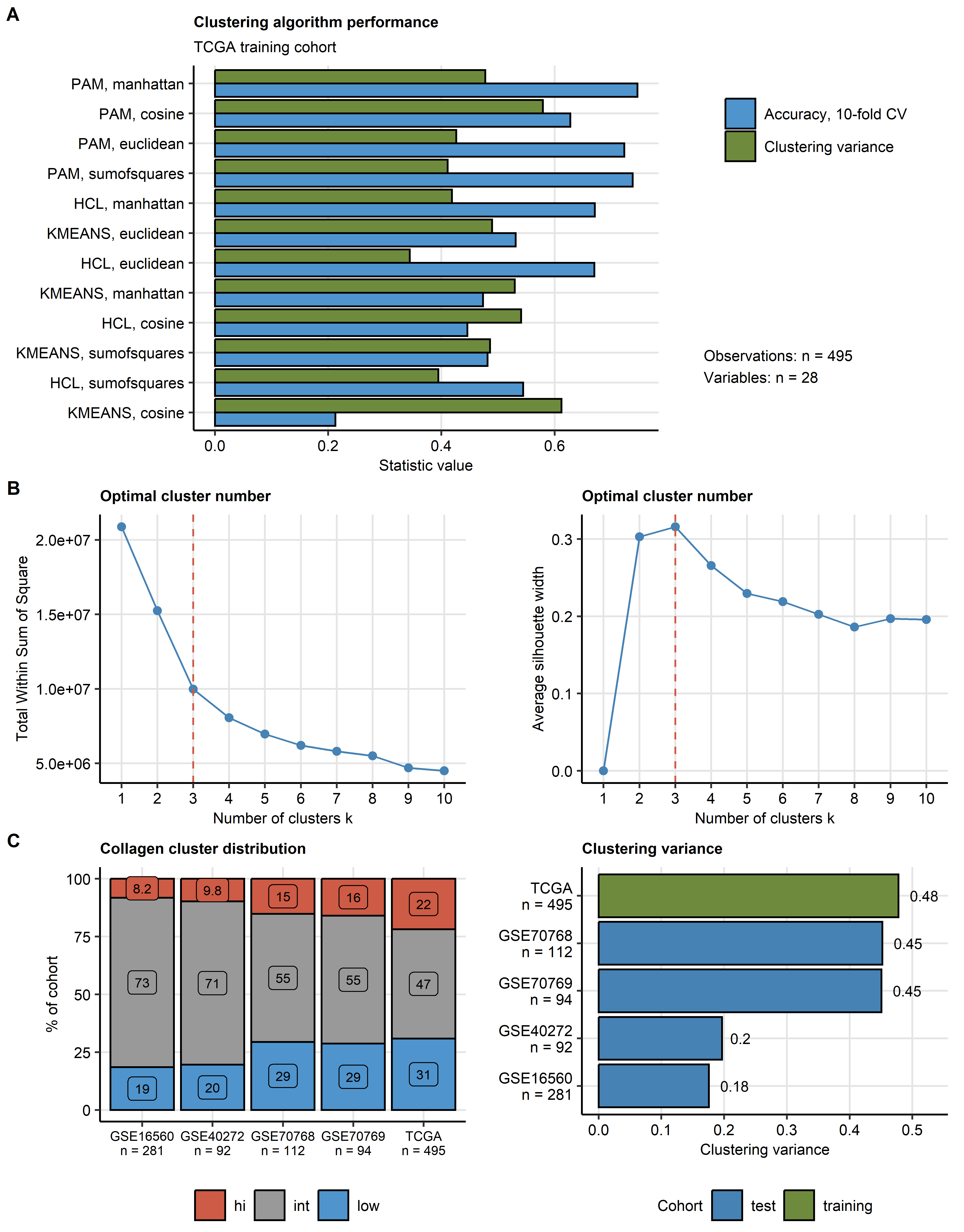


Figure 2: Semi-supervised clustering of prostate cancer samples in respect to expression of the collagen pathway genes.

**Supplementary Figure S2. Semi-supervised clustering of prostate cancer samples in respect to expression of the collagen pathway genes.**

*Tumor samples in the TCGA training cohort were clustered in respect to normalized, -transformed expression levels of the collagen pathway genes with the PAM (partition around medoids) algorithm with Manhattan distance metric. Three clusters were defined: Collagenlow, Collagenintermediate and Collagenhigh. Assignment of the tumor samples from the training GSE16560, GSE40272, GSE70768 and GSE70769 collectives to the collagen clusters was accomplished by an inverse distance-weighted k-nearest neighbor classifier.*

*(A) Comparison of performance of several clustering algorithms (PAM: partition around medoids, HCL: hierarchical clustering/Ward D2 and KMEANS) and distance metrics (Euclidean, Manhattan, sum-of-squares and cosine) in the training TCGA cohort was measured by the amount of ‘explained’ clustering variance (ratio of the between-cluster sum of squares to the total sum of squares) and rate of correct cluster assignment in 10-fold cross-validation (CV). Number of complete observations and the clustering factors are shown next to the plot.*

*(B) Determination of the cluster number for PAM/Manhattan distance algorithm in the TCGA cohort by the bend of the within-cluster sum of squares (WSS) (B) and the peak of the mean silhouette statistic (C).*

*(C) Distribution of the collagen clusters (left panel) and performance of the PAM/Manhattan distance algorithm in semi-supervised clustering in the training cohort and the test collectives measured by the ‘explained’ clustering variance statistic (right panel). Numbers of complete observations are presented in the plot axes..*



Figure 3: Expression of the collagen pathway genes in the collagen clusters of prostate cancers.

**Supplementary Figure S3. Expression of the collagen pathway genes in the collagen clusters of prostate cancers.**

*Tumor samples were subjected to semi-supervised clustering in respect to the collagen pathway gene expression with the PAM algorithm with Manhattan distance metric (training: TCGA, test cohorts: GSE16560, GSE40272, GSE70768 and GSE70769). Statistical significance of differences in expression of normalized, -transformed expression levels (Z-scores) between the collagen clusters were investigated by Kruskal-Wallis test with effect statistic. P values were corrected for multiple testing with the false discovery rate method. Mean expression Z scores in the Collagenlow, Collagenintermediate (int) and Collagenhigh (hi) clusters are presented as thick lines. The tinted ribbons represent 2 SEM (standard error of the mean) intervals. The collagen pathway genes were grouped according to their biological function as proline synthesis pathway, extracellular matrix (ECM) component and ECM processing protein genes. Significantly regulated genes are highlighted in bold. Numbers of samples in the clusters are indicated in the plot captions.*

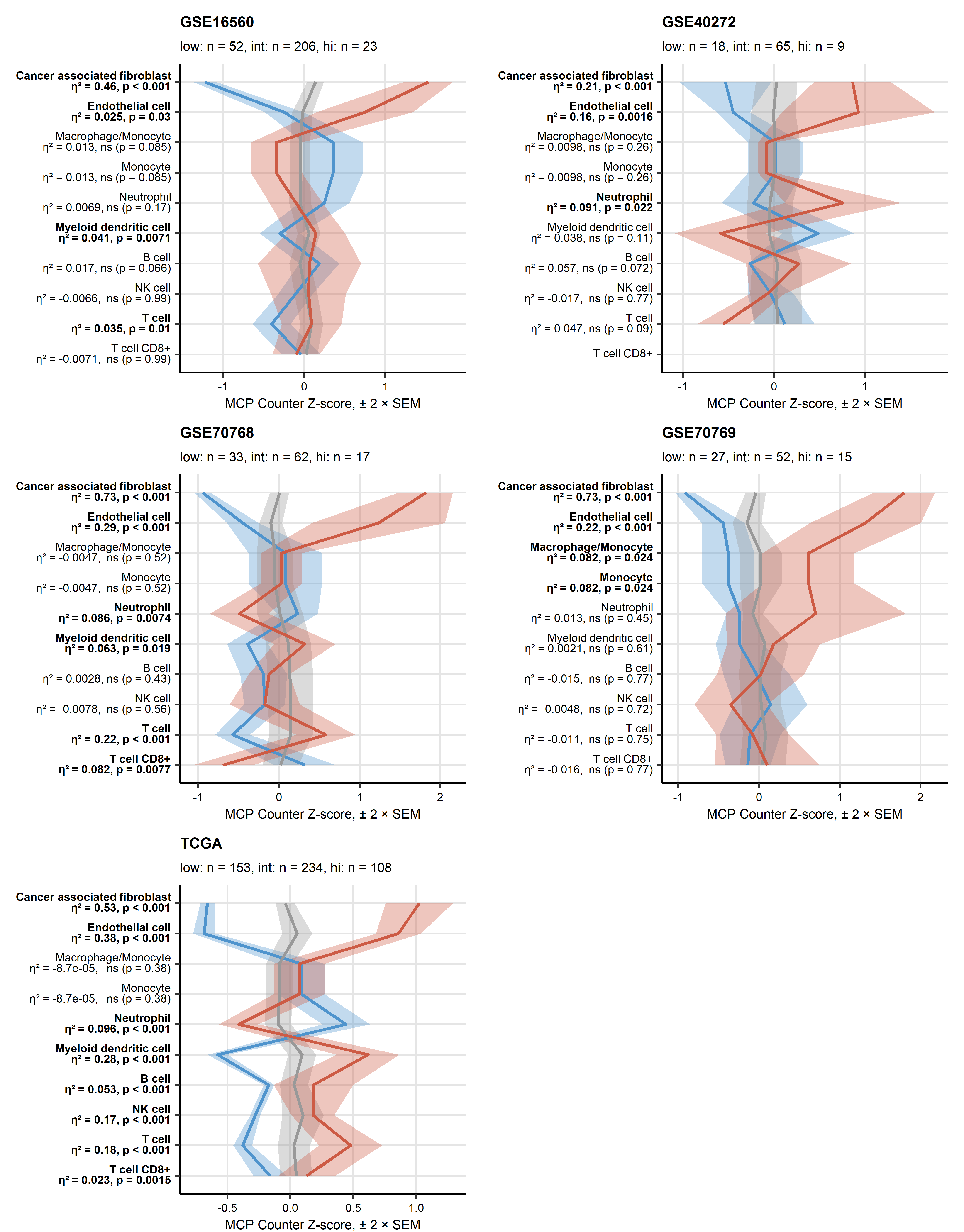


Figure 4: MCP Counter estimates of non-malignant cell content in the collagen clusters.

**Supplementary Figure S4. MCP Counter estimates of non-malignant cell content in the collagen clusters.**

*Counts of non-malignant cells (CAF: cancer-associated fibroblast, EC: endothelial cells, T cells, CD8+ T cells, NK cells: natural killer cells, B cells, macrophages, monocytes, neutrophils and mDC: myeloid dendritic cells) in the prostate cancer samples were predicted by the MCP counter algorithm. Differences in cell content between the Collagenlow, Collagenintermediate (in) and Collagenhigh (hi) clusters were analyzed with Kruskal-Wallis test with effect size statistic. P values were corrected for multiple testing with the false discovery rate method. Means of normalized cell counts in the collagen clusters are depicted as thick lines. Tinted ribbons represent the 2 SEM (standard error of the mean) intervals. Effect sizes and p values are shown in the Y axes. Significant effects are highlighted in bold. Numbers of samples assigned to the clusters are indicated in the plot captions.*

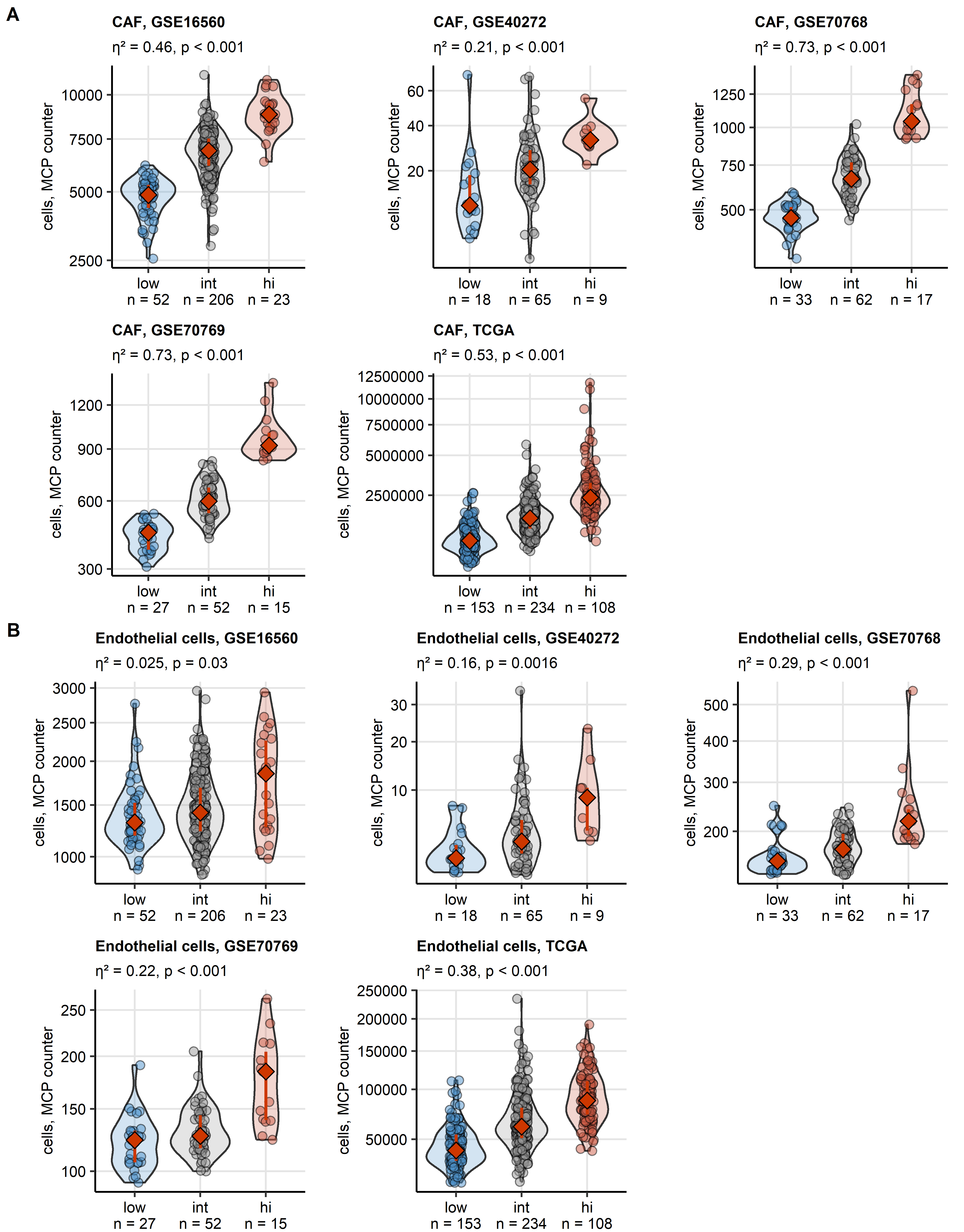


Figure 5: Counts of cancer-associated fibroblast and endothelial cell in tumor samples in the collagen clusters predicted by the MCP counter algorithm.

**Supplementary Figure S5. Counts of cancer-associated fibroblast and endothelial cell in tumor samples in the collagen clusters predicted by the MCP counter algorithm.**

*Counts of cancer associated fibroblasts (CAF, A) and endothelial cells (EC, B) in collagenlow, collagenintermediate (int) and collagenhigh (hi) prostate cancers were predicted by the MCP counter algorithm. Differences between the collagen clusters were assessed by Kruskal-Wallis test with effect size statistic. Cell counts were presented in violin plots with single observations visualized as points. Red diamonds and whiskers represent medians with interquartile ranges. Effect sizes and p values are displayed in the plot captions. Numbers of samples in the clusters are indicated in the X axes.*

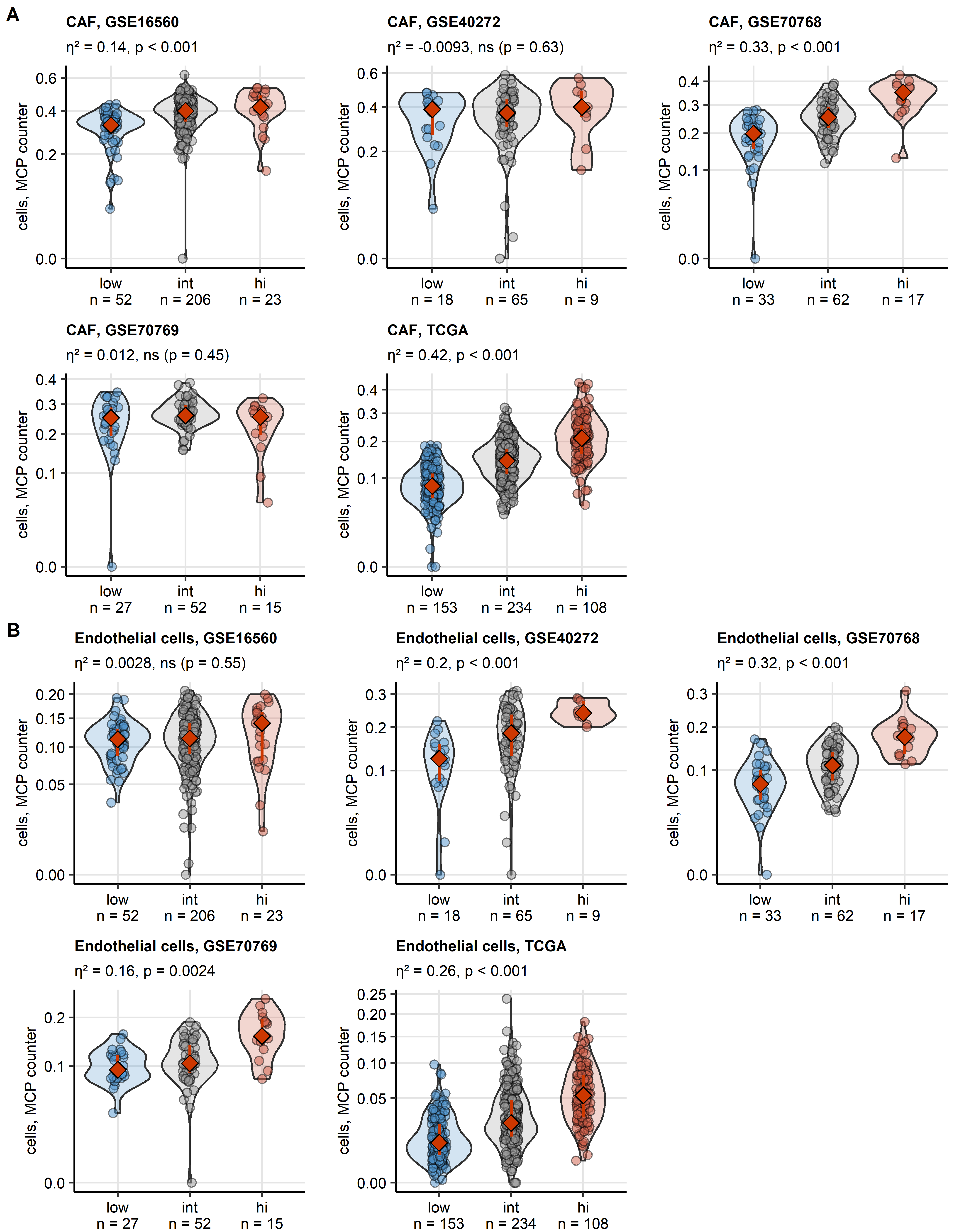


Figure 6: Fractions of cancer-associated fibroblast and endothelial cell in tumor samples in the collagen clusters predicted by the xCell algorithm.

**Supplementary Figure S6. Fractions of cancer-associated fibroblast and endothelial cell in tumor samples in the collagen clusters predicted by the xCell algorithm.**

*Fractions of cancer associated fibroblasts (CAF, A) and endothelial cells (EC, B) in the malignant tissue of collagenlow, collagenintermediate (int) and collagenhigh (hi) prostate cancers were predicted by the xCell counter algorithm. Differences between the collagen clusters were assessed by Kruskal-Wallis test with effect size statistic. Cell counts were presented in violin plots with single observations visualized as points. Red diamonds and whiskers represent medians with interquartile ranges. Effect sizes and p values are displayed in the plot captions. Numbers of samples in the clusters are indicated in the X axes.*

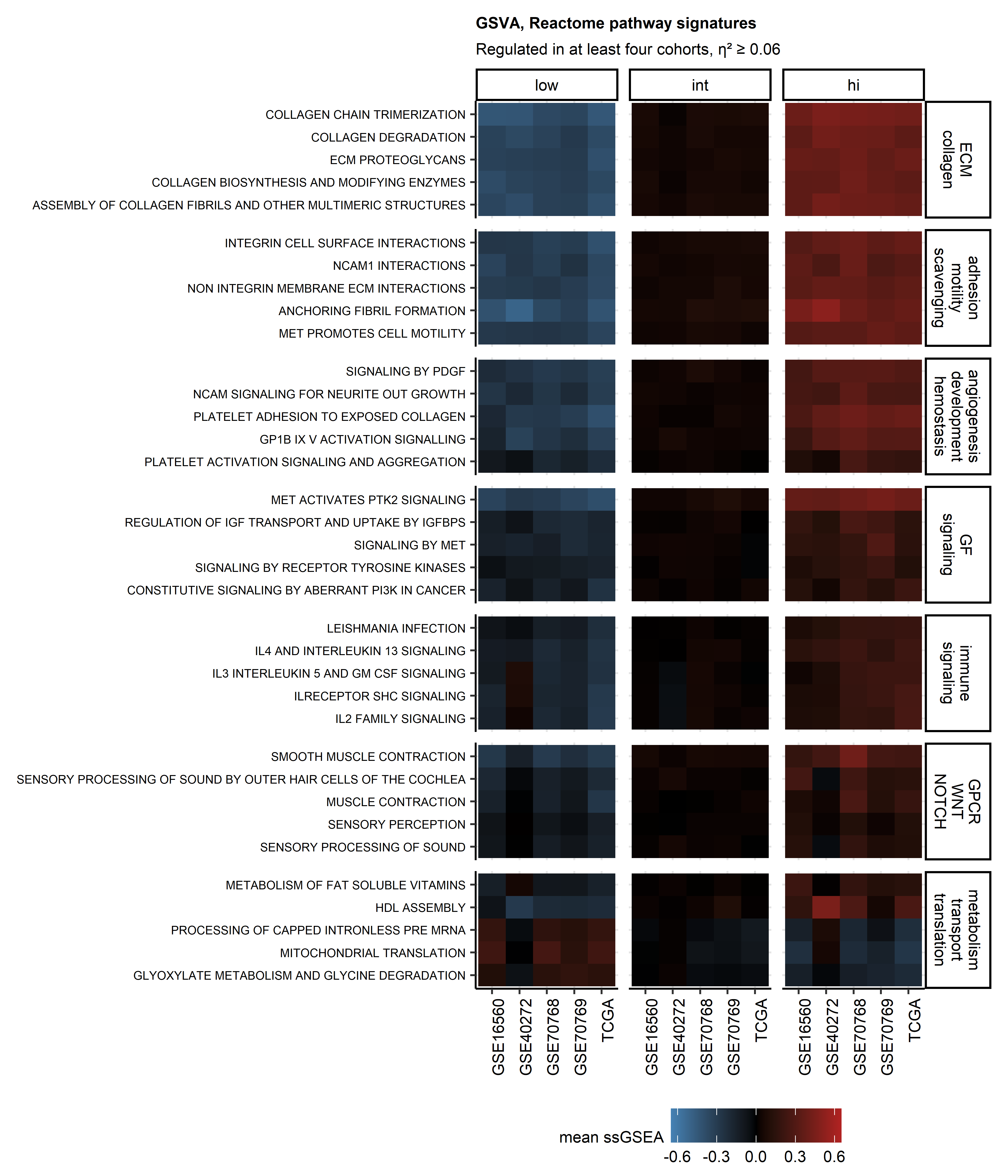


Figure 7: Gene set variation analysis results for representative Reactome pathway gene signatures differentiating between the collagen clusters.

**Supplementary Figure S7. Gene set variation analysis results for representative Reactome pathway gene signatures differentiating between the collagen clusters.**

*Gene set variation analysis of the collagen clusters was performed with the 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. Common regulated signatures were defined as signatures significantly regulated with at least moderate effect size ( 0.6) in at least four cohorts (Supplementary Table S6). Mean ssGSEA scores in the clusters and cohorts for the top five strongest regulated signatures for each functional category (extracellular matrix [ECM], collagen; adhesion, motility, scavenging; angiogenesis, development, hemostasis; growth factor [GF] signaling; immune signaling; G protein coupled receptor, WNT, NOTCH signaling; metabolism, transport, translation) are presented in a heat map.*

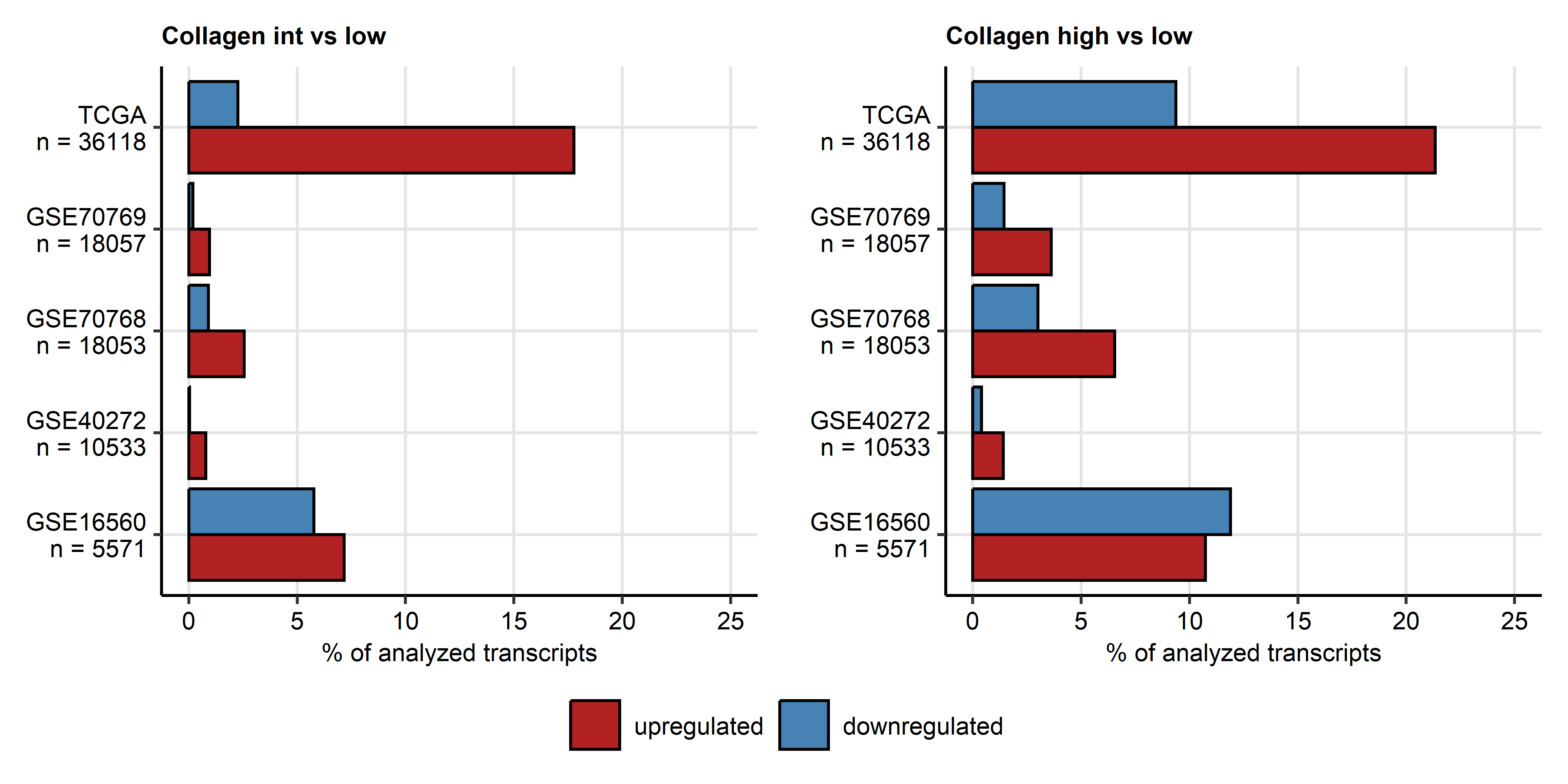


Figure 8: Percentages of the analyzed transcriptome significantly up- and downregulated in the collagen high and collagen low clusters.

**Supplementary Figure S8. Percentages of the analyzed transcriptome significantly up- and downregulated in the collagen high and collagen low clusters.**

*Genes significantly differentially regulated in the collagenhigh (hi) and collagenintermediate (int) clusters as compared with collagenlow prostate cancers were identified by false discovery rate corrected ANOVA (pFDR < 0.05) and linear modeling (fold-regulation > 1.25). Percentages of significantly up- and downregulated genes are presented in bar plots. Total numbers of analyzed genes are indicated in the Y axes.*

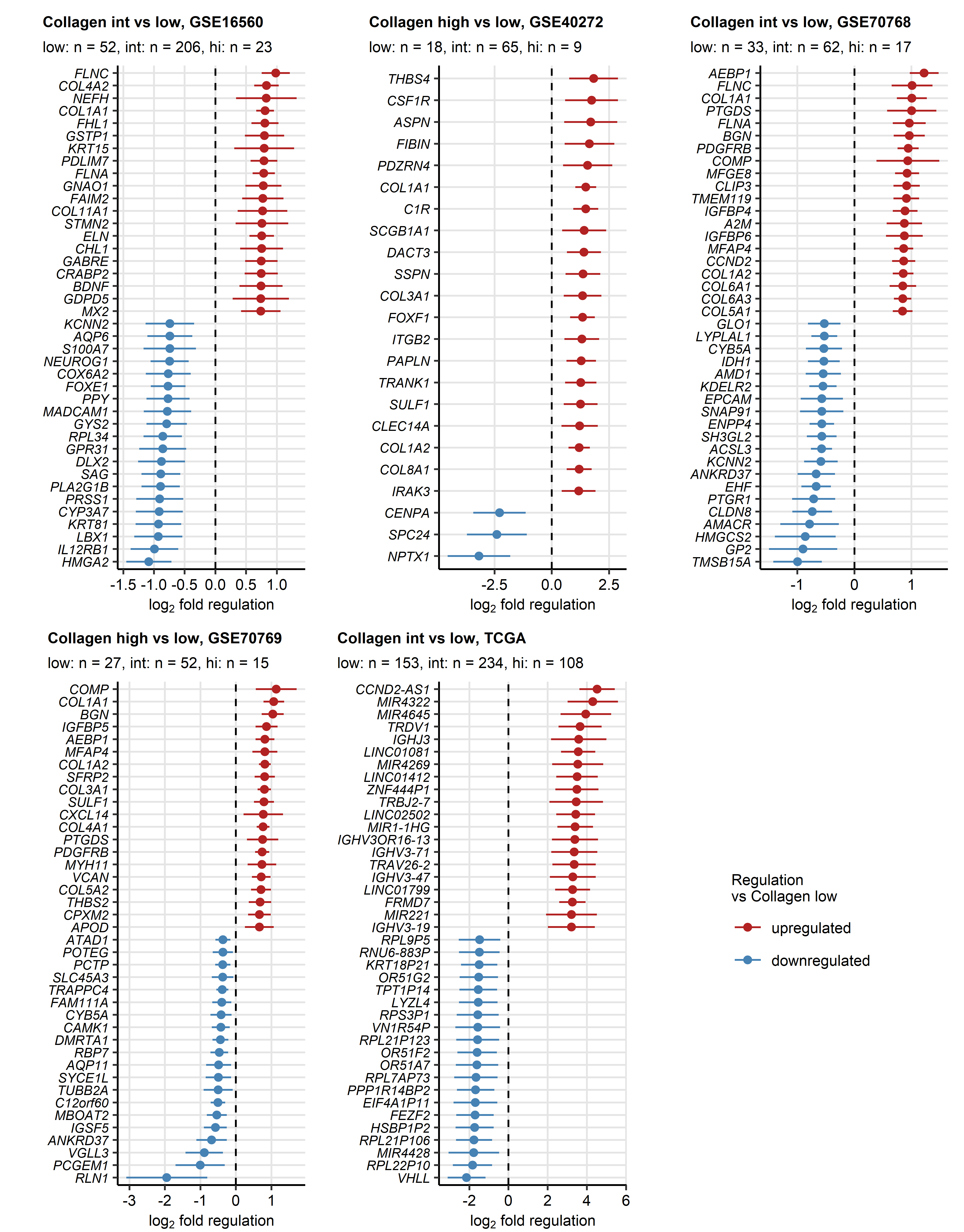


Figure 9: Top strongest differentially expressed genes in the collagen intermediate cluster as compared with collagen low cancers.

**Supplementary Figure S9. Top strongest differentially expressed genes in the collagen intermediate cluster as compared with collagen low cancers.**

*fold-regulation estimates of differences in expression between collagenintermediate (int) and collagenlow cancers with 95% confidence intervals for top 20 strongest up- and downregulated genes were presented in Forest plots. Numbers of samples in the clusters are displayed in the plot captions.*

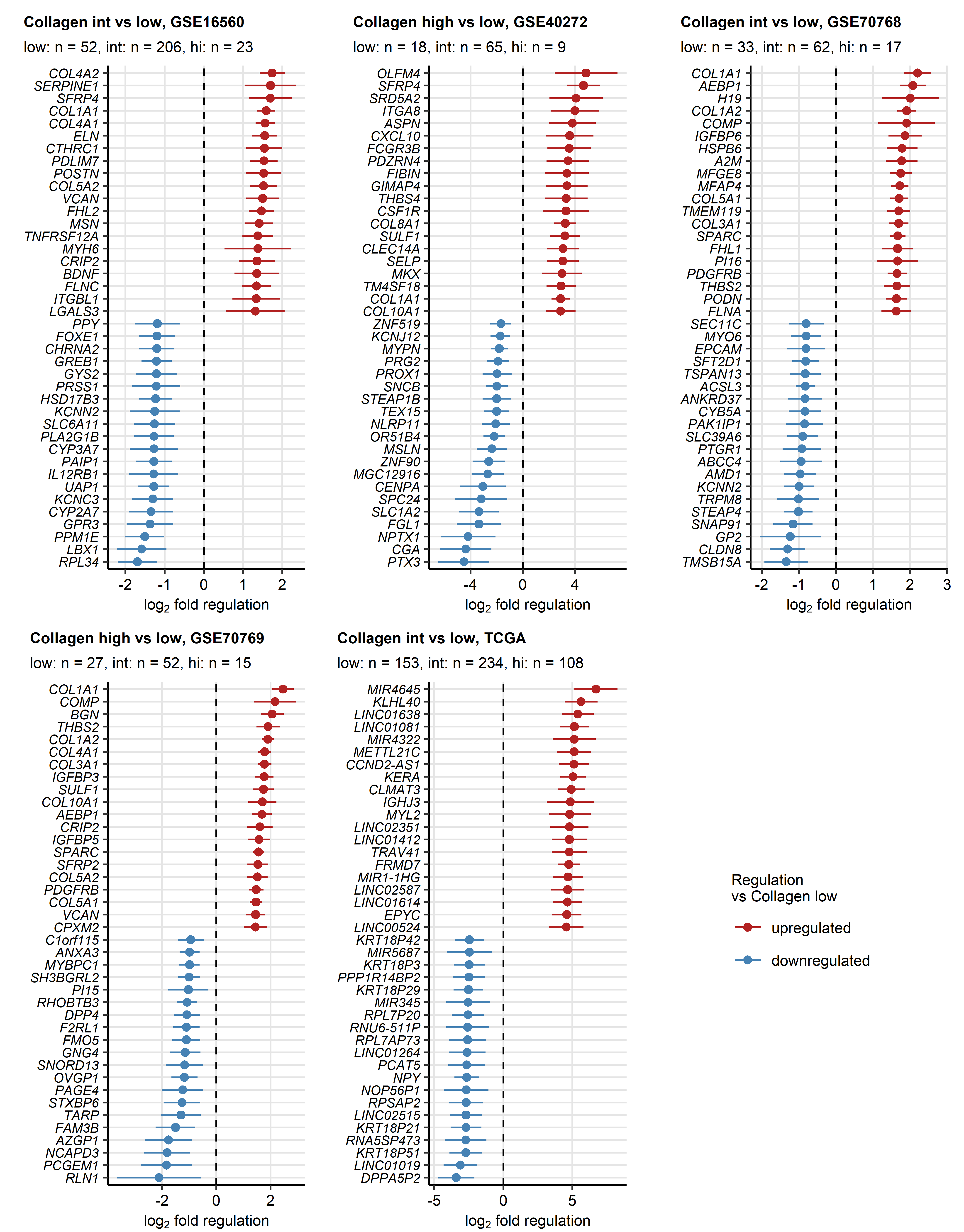


Figure 10: Top strongest differentially expressed genes in the collagen high cluster as compared with collagen low cancers.

**Supplementary Figure S10. Top strongest differentially expressed genes in the collagen high cluster as compared with collagen low cancers.**

*fold-regulation estimates of differences in expression between collagenhigh (high) and collagenlow cancers with 95% confidence intervals for top 20 strongest up- and downregulated genes were presented in Forest plots. Numbers of samples in the clusters are displayed in the plot captions.*



Figure 11: Numbers of significantly activated and inhibited biochemical reactions predicted for the collagen high and collagen intermediate clusters.

**Supplementary Figure S11. Numbers of significantly activated and inhibited biochemical reactions predicted for the collagen high and collagen intermediate clusters.**

*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 whole genome differential gene expression estimates. Statistical significance was determined by Monte Carlo simulation with n = 1000 draws from distributions of the gene expression regulation standard errors. P values were corrected for multiple testing with the false discovery rate method. Percentages of predicted significantly activated and inhibited metabolic reactions are shown in bar plots. Total reaction numbers are displayed in the plot captions.*

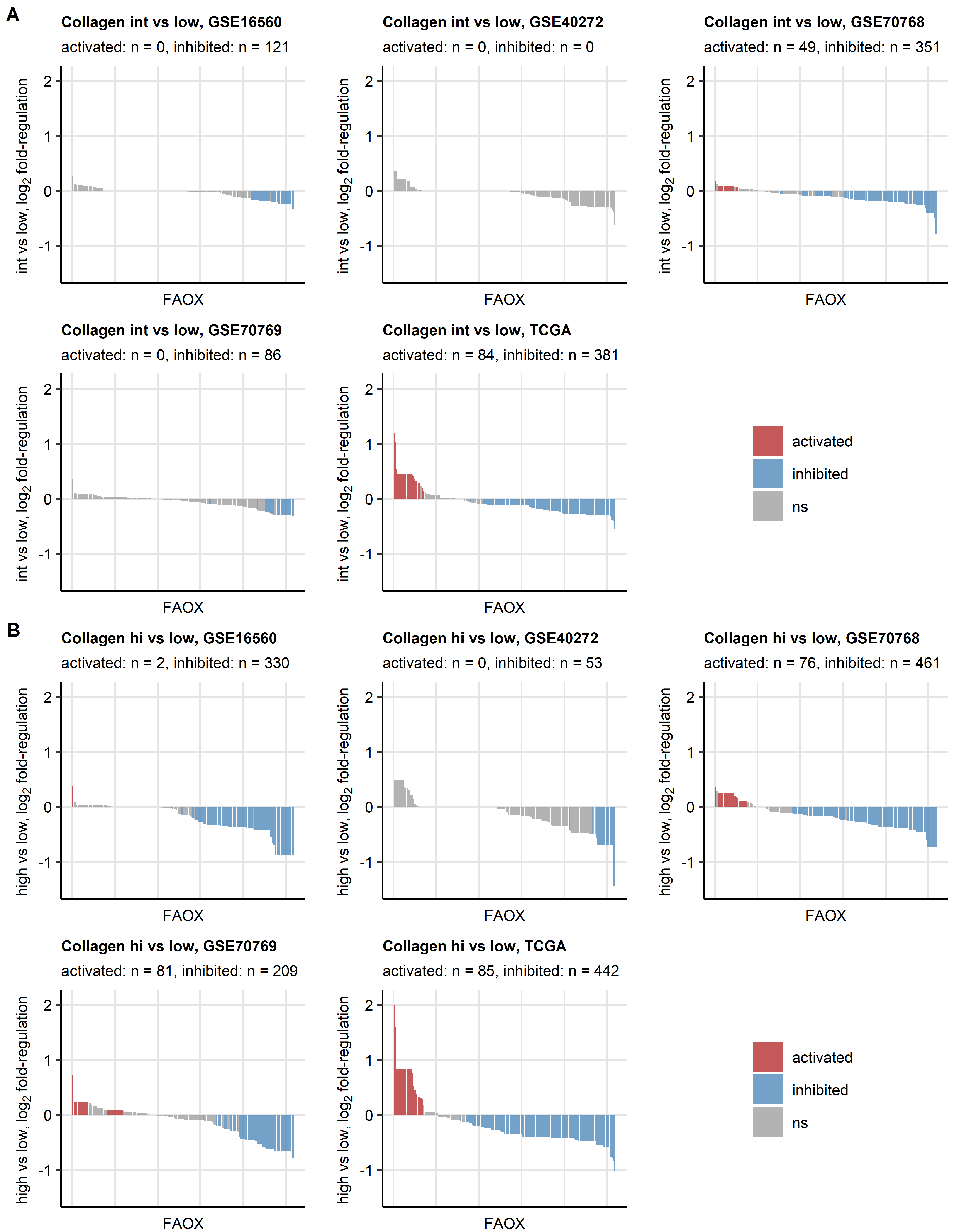


Figure 12: Activity of extracellular transport reactions predicted for the collagen high and collagen intermediate clusters.

**Supplementary Figure S12. Activity of extracellular transport reactions predicted for the collagen high and collagen intermediate clusters.**

*Regulation of the fatty acid oxidation (FAOX) Recon subsystem reactions (total reactions: n = 869) in the collagenhigh (hi, A) or collagenintermediate (hi, B) as compared with collagenlow tumors was predicted by the BiGGR and biggrExtra algorithms based on whole genome differential gene expression estimates. Statistical significance was determined by Monte Carlo simulation with n = 1000 draws from distributions of the gene expression regulation standard errors. P values were corrected for multiple testing with the false discovery rate method. Estimates of fold-regulation of reaction activity are presented in bar plots. Activity status is color coded. Numbers of significantly activated and inhibited reactions are displayed in the plot captions.*

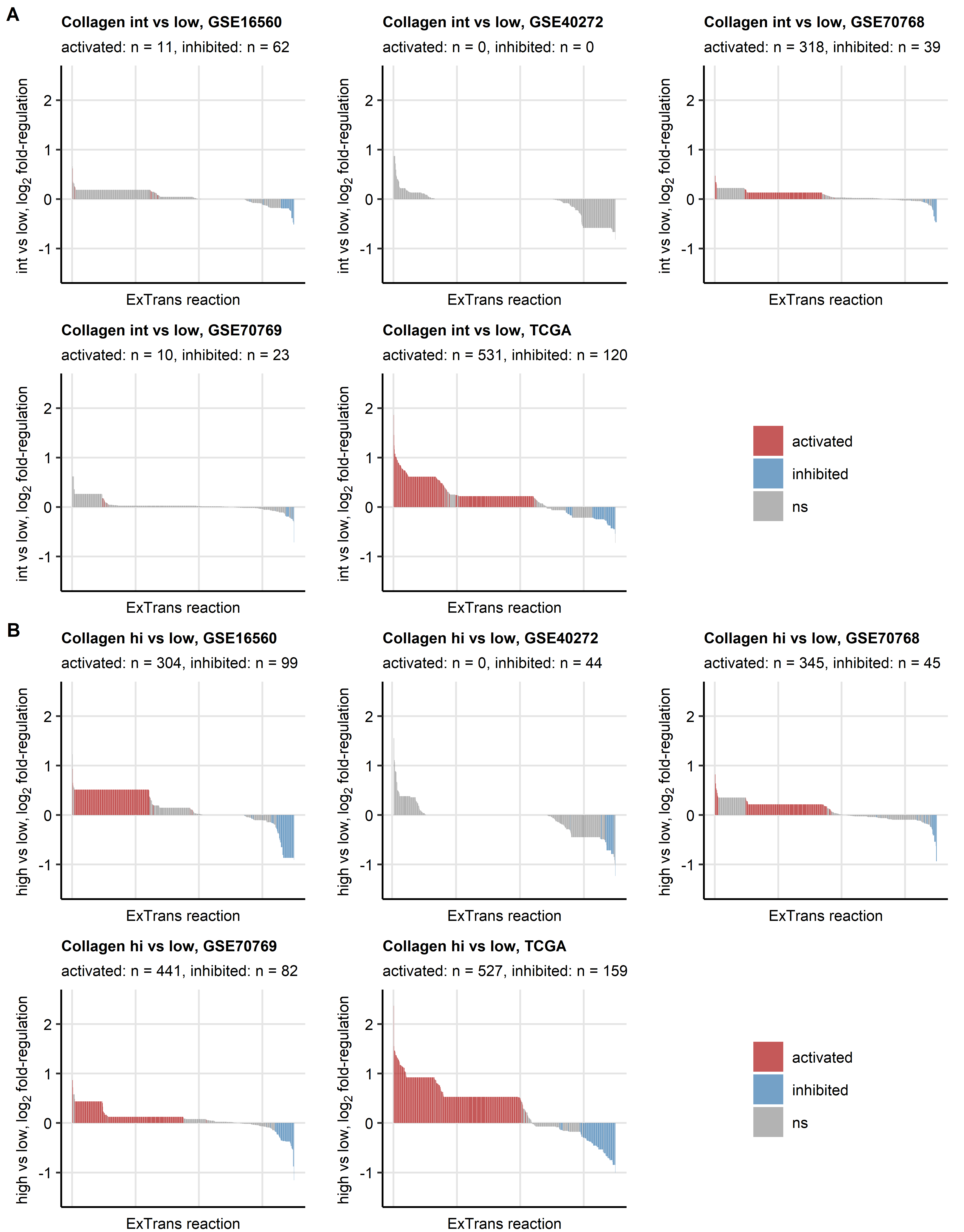


Figure 13: Activity of extracellular transport reactions predicted for the collagen high and collagen intermediate clusters.

**Supplementary Figure S13. Activity of extracellular transport reactions predicted for the collagen high and collagen intermediate clusters.**

*Regulation of the extracellular transport (ExTrans) Recon subsystem reactions (total reactions: n = 1550) in the collagenhigh (hi, A) or collagenintermediate (hi, B) as compared with collagenlow tumors was predicted by the BiGGR and biggrExtra algorithms based on whole genome differential gene expression estimates. Statistical significance was determined by Monte Carlo simulation with n = 1000 draws from distributions of the gene expression regulation standard errors. P values were corrected for multiple testing with the false discovery rate method. Estimates of fold-regulation of reaction activity are presented in bar plots. Activity status is color coded. Numbers of significantly activated and inhibited reactions are displayed in the plot captions.*

# References