Cell Reports, Volume *26*

Supplemental Information

A Systematic Investigation of the Malignant Functions and Diagnostic Potential of the Cancer Secretome

Jonathan L. Robinson, Amir Feizi, Mathias Uhlén, and Jens Nielsen

Table S1. Definition of TCGA cancer type abbreviations and mapping to GTEx tissues. Related to STAR Methods.

TCGA cancer type	Description	GTEx tissue type
ACC	adrenocortical carcinoma	adrenal gland
BLCA	bladder urothelial carcinoma	bladder
BRCA	breast invasive carcinoma	breast
CESC	cervical squamous cell carcinoma and endocervical adenocarcinoma	cervix uteri
CHOL	cholangiocarcinoma	liver
COAD	colon adenocarcinoma	colon
DLBC	lymphoid neoplasm diffuse large B-cell lymphoma	--
ESCA	esophageal carcinoma	esophagus
GBM	glioblastoma multiforme	brain
HNSC	head and neck squamous cell carcinoma	salivary gland
KICH	kidney chromophobe	kidney
KIRC	kidney renal clear cell carcinoma	kidney
KIRP	kidney renal papillary cell carcinoma	kidney
LGG	brain lower grade glioma	brain
LIHC	liver hepatocellular carcinoma	liver
LUAD	lung adenocarcinoma	lung
LUSC	lung squamous cell carcinoma	lung
MESO	mesothelioma	--
OV	ovarian serous cystadenocarcinoma	ovary
PAAD	pancreatic adenocarcinoma	pancreas
PCPG	pheochromocytoma and paraganglioma	adrenal gland
PRAD	prostate adenocarcinoma	prostate
READ	rectum adenocarcinoma	colon
SARC	sarcoma	--
SKCM	skin cutaneous melanoma	skin
STAD	stomach adenocarcinoma	stomach
TGCT	testicular germ cell tumors	testis
THCA	thyroid carcinoma	thyroid
THYM	thymoma	--
UCEC	uterine corpus endometrial carcinoma	uterus
UCS	uterine carcinosarcoma	uterus
UVM	uveal melanoma	--

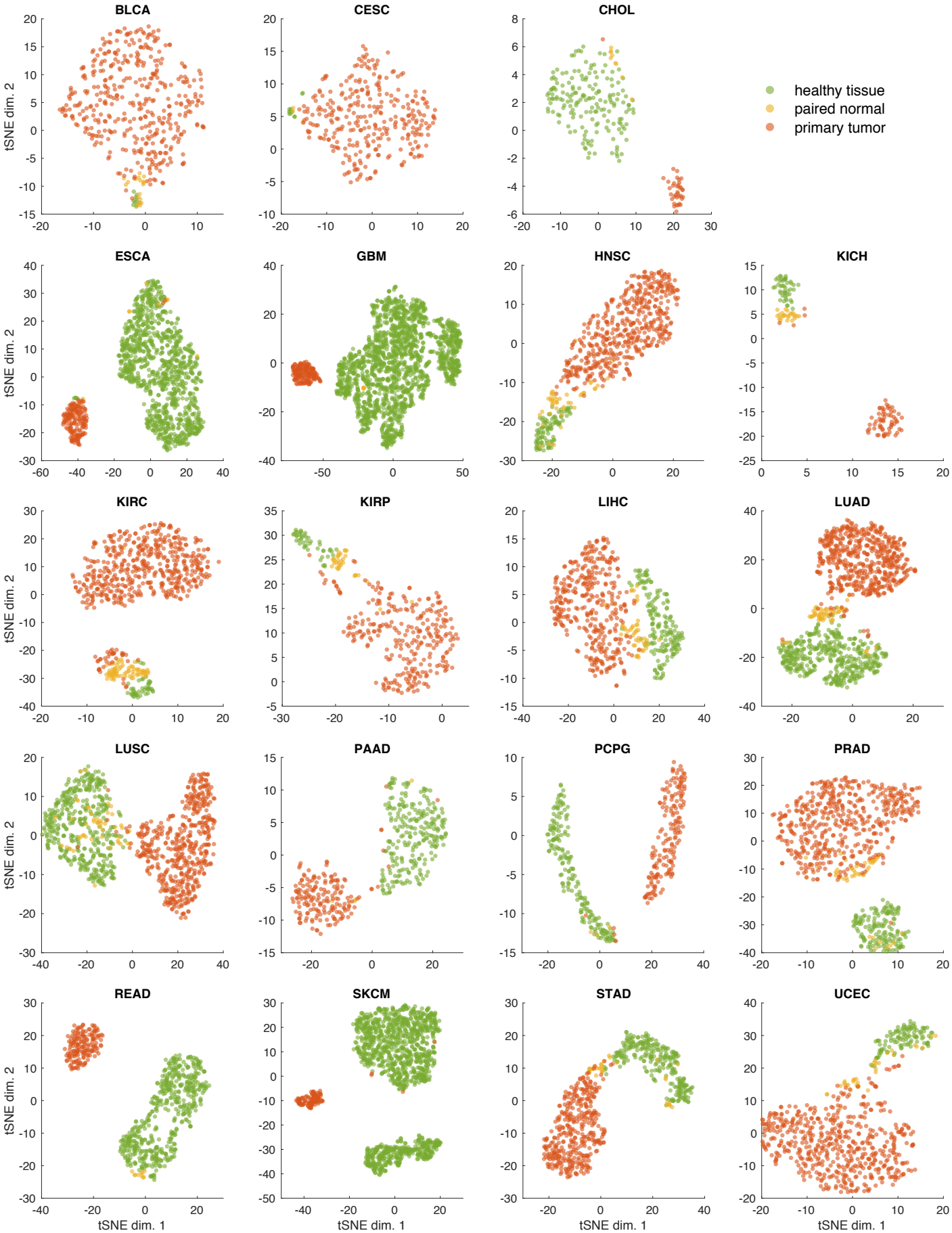

Figure S1. tSNE projections illustrating the separation of primary tumor (red), paired normal (yellow), and healthy tissue of origin (green) samples based on the abundance (log_{10}TPM) of the top 10 consensus-ranked genes for each cancer type. Related to Figure 1B. Only cancer types not presented in Fig. 1B, and with RNA-Seq data for at least one sample of each type, are shown.

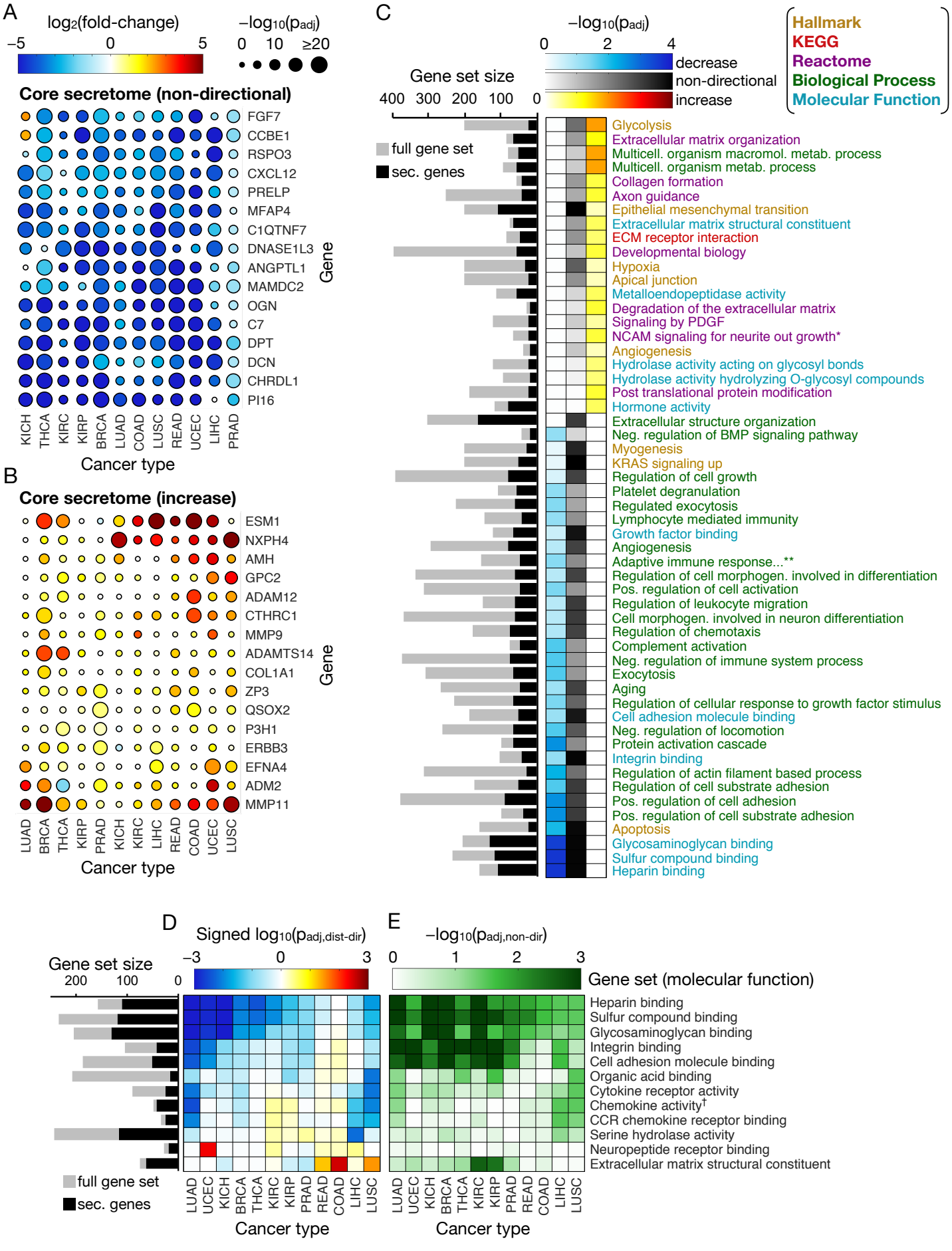
A
log2(fold-change)
−5
0
5
−log10(padj)
0
10
≥20
Core secretome (non-directional)
FGF7
CCBE1
RSPO3
CXCL12
PRELP
MFAP4
C1QTNF7
DNASE1L3
ANGPTL1
MAMDC2
OGN
C7
DPT
DCN
CHRDL1
PI16
Gene
KICH
THCA
KIRC
KIRP
BRCA
LUAD
COAD
LUSC
READ
UCEC
LIHC
PRAD
Cancer type
B
Core secretome (increase)
ESM1
NXPH4
AMH
GPC2
ADAM12
CTHRC1
MMP9
ADAMTS14
COL1A1
ZP3
QSOX2
P3H1
ERBB3
EFNA4
ADM2
MMP11
Gene
LUAD
BRCA
THCA
KIRP
PRAD
KICH
KIRC
LIHC
READ
COAD
UCEC
LUSC
Cancer type
C
−log10(padj)
0
2
4
decrease
non-directional
increase
Gene set size
400
300
200
100
0
full gene set
sec. genes
Hallmark
KEGG
Reactome
Biological Process
Molecular Function
Glycolysis
Extracellular matrix organization
Multicell. organism macromol. metab. process
Multicell. organism metab. process
Collagen formation
Axon guidance
Epithelial mesenchymal transition
Extracellular matrix structural constituent
ECM receptor interaction
Developmental biology
Hypoxia
Apical junction
Metalloendopeptidase activity
Degradation of the extracellular matrix
Signaling by PDGF
NCAM signaling for neurite out growth*
Angiogenesis
Hydrolase activity acting on glycosyl bonds
Hydrolase activity hydrolyzing O-glycosyl compounds
Post translational protein modification
Hormone activity
Extracellular structure organization
Neg. regulation of BMP signaling pathway
Myogenesis
KRAS signaling up
Regulation of cell growth
Platelet degranulation
Regulated exocytosis
Lymphocyte mediated immunity
Growth factor binding
Angiogenesis
Adaptive immune response...**
Regulation of cell morphogen. involved in differentiation
Pos. regulation of cell activation
Regulation of leukocyte migration
Cell morphogen. involved in neuron differentiation
Regulation of chemotaxis
Complement activation
Neg. regulation of immune system process
Exocytosis
Aging
Regulation of cellular response to growth factor stimulus
Cell adhesion molecule binding
Neg. regulation of locomotion
Protein activation cascade
Integrin binding
Regulation of actin filament based process
Regulation of cell substrate adhesion
Pos. regulation of cell adhesion
Pos. regulation of cell substrate adhesion
Apoptosis
Glycosaminoglycan binding
Sulfur compound binding
Heparin binding
D
Signed log10(padj,dist-dir)
−3
0
3
Gene set size
200
100
0
full gene set
sec. genes
LUAD
UCEC
KICH
BRCA
THCA
KIRC
KIRP
PRAD
READ
COAD
LIHC
LUSC
Cancer type
E
−log10(padj,non-dir)
0
1
2
3
Gene set (molecular function)
Heparin binding
Sulfur compound binding
Glycosaminoglycan binding
Integrin binding
Cell adhesion molecule binding
Organic acid binding
Cytokine receptor activity
Chemokine activity†
CCR chemokine receptor binding
Serine hydrolase activity
Neuropeptide receptor binding
Extracellular matrix structural constituent
LUAD
UCEC
KICH
BRCA
THCA
KIRC
KIRP
PRAD
READ
COAD
LIHC
LUSC
Cancer type

Figure S2. Constituents and functions of the "core" cancer secretome, and gene set analysis of the cancer secretome, using high-purity tumor samples. Related to Figures 2 and 3. The analyses were performed after removing all tumor samples with a consensus purity estimate (CPE) of < 80%. (**A**) A heat-scatter plot presenting the log_2FCs and corresponding significance (FDR-adjusted p-values) for the 16 genes comprising the top 1% of the non-directional core secretome. The color and size of the points correspond to the log_2FC and log-transformed p-values, respectively, from the DE analysis between tumor and paired normal samples. (**B**) The top 1% of the increased core secretome, obtained in the same manner as the non-directional set in (A), except the fold-change direction was incorporated to identify secretome genes exhibiting increased expression across many cancer types. (**C**) Gene sets found to be significantly enriched in the decreased (left column), non-directional (center column), or increased core secretome (right column), where the top 20 most significant sets from each directional class are shown. The intensity of the color in the heatmap indicates the enrichment significance of the gene set. Gene set names in (C) are colored according to the MSigDB collection from which they originate: Hallmark, KEGG, Reactome, GO biological process, and GO molecular function. Additional heatmaps illustrate the (**D**) directional and (**E**) non-directional GSA results for secretome genes based on the tumor vs. paired-normal fold-changes and significance in 17 different cancer types. Only the GO molecular function gene set collection (MSigDB) was evaluated in (C) and (D), and sets with <10 genes were excluded. A non-stacked bar plot to the left of the heatmaps shows the sizes of the original gene sets (grey bars) and of the filtered gene sets containing only secretome genes (black bars). *The "NCAM signaling for neurite outgrowth" gene set was identical to the "NCAM1 interactions" set after removing non-secretome genes, thus the latter set is not shown. **The full gene set name is "adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains". †The "chemokine activity" gene set was identical to the "chemokine receptor binding" gene set after removing non-secretome genes, thus the latter set is not shown.

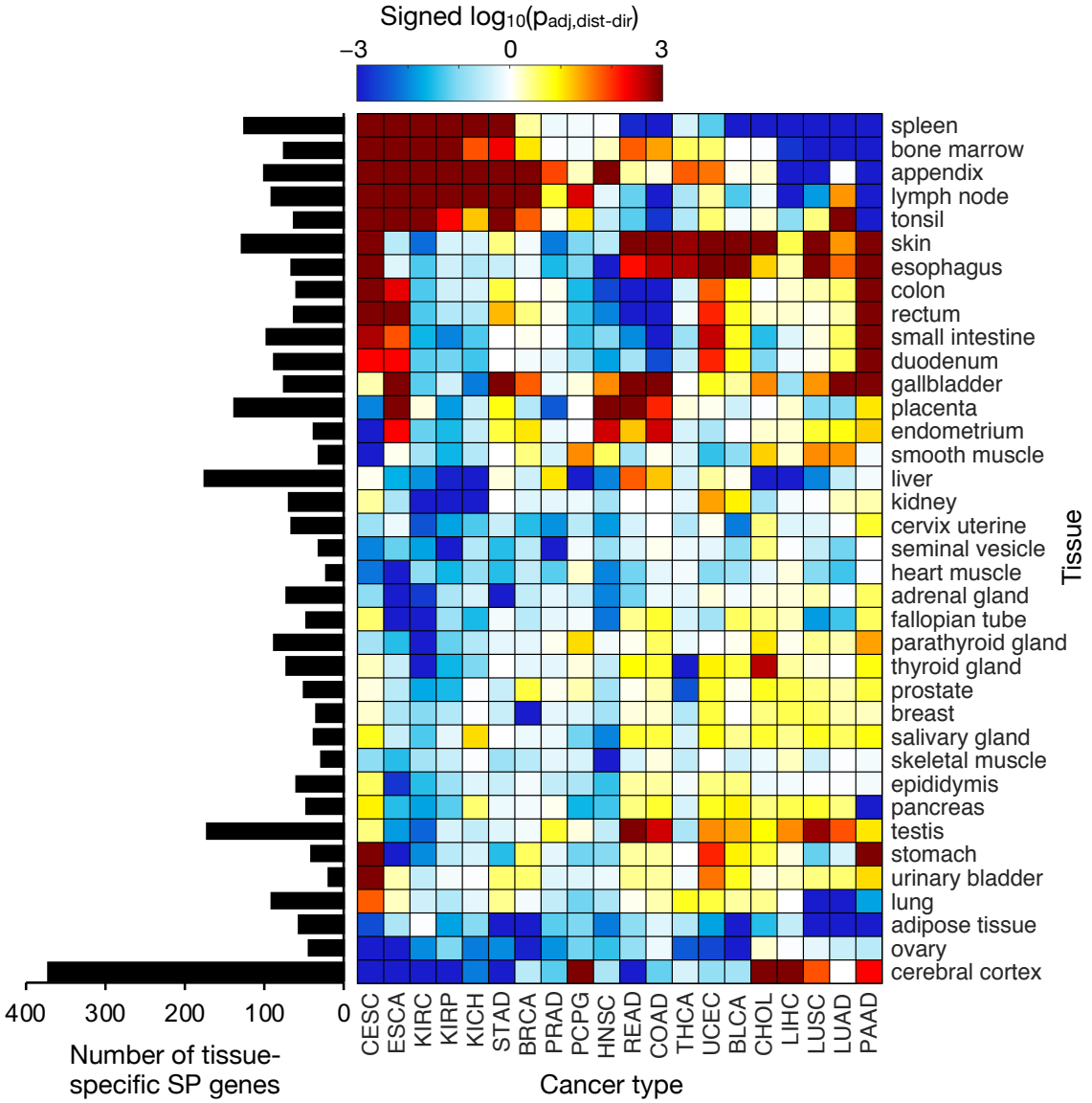

Figure S3. Tissue-specific expression changes in SP genes for all HPA tissue types. Related to Figure 4. The heatmap is similar to that in Fig. 4, but presents the significance of expression changes in tissue-specific SP genes for all analyzed tissues, and does not group cancer types by tissue of origin. The number of tissue-specific SP genes for each tissue type are shown as a bar plot to the left of the heatmap.

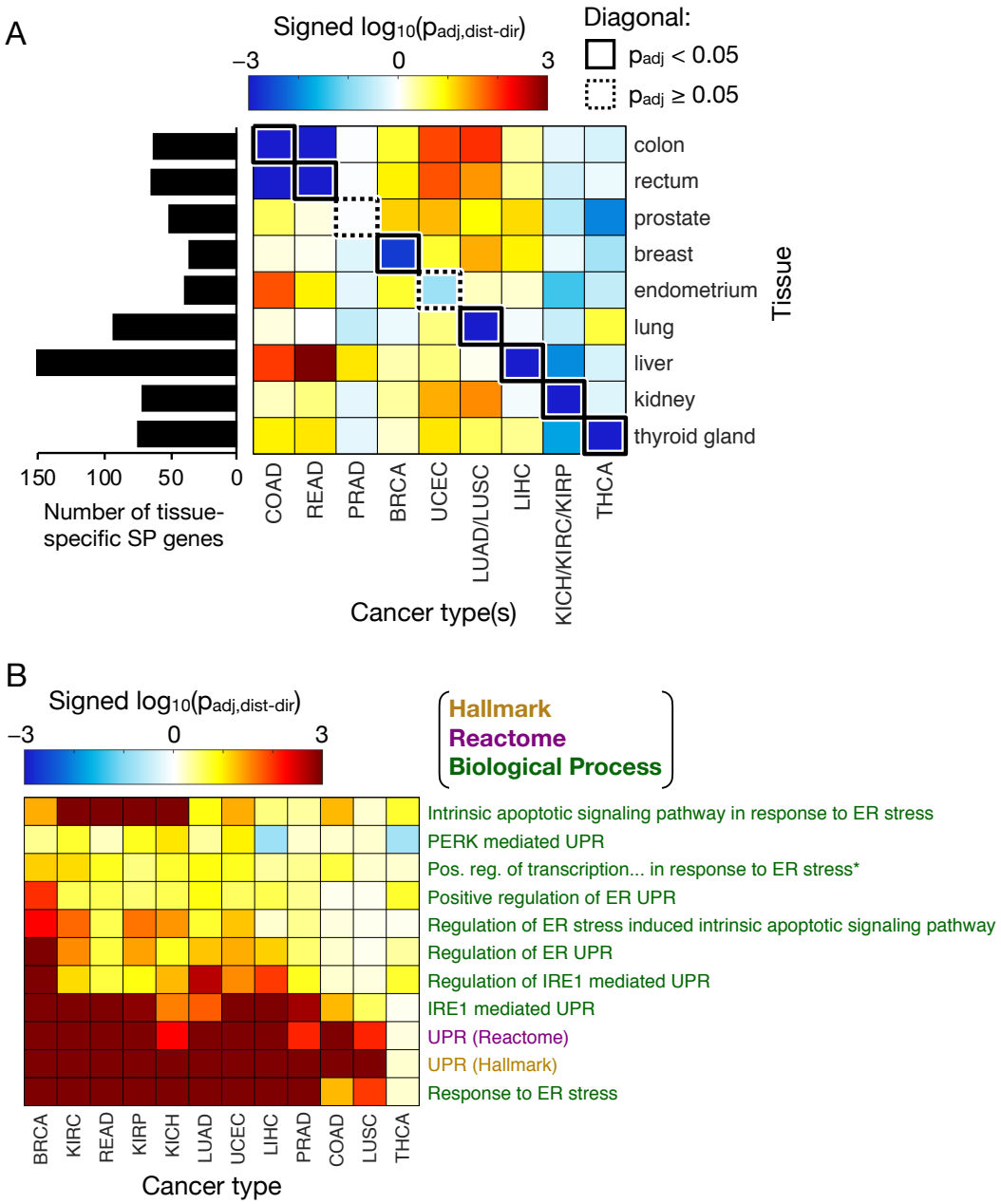

Figure S4. Related to Figures 4 and 5. (**A**) Tissue-specific expression changes in SP genes, when using only high-purity tumor samples. The figure is identical to Fig. 4, except the analysis was performed after removing all tumor samples with a CPE of < 80%. The heatmap shows the significance and direction of coordinated expression changes in SP genes classified as specific to various tissue types. Cancer and tissue types are organized such that entries along the diagonal represent cancer types paired with their tissue of origin, and are outlined in a solid box if there is a significant ($p_{adj} < 0.05$) coordinated expression decrease among the tissue-specific SP genes for that cancer type, or a dotted box otherwise. The number of tissue-specific SP genes for each tissue type are indicated in the bar plot to the left of the heatmap. **(B)** Coordinated expression changes in various UPR-related gene sets, when using only high-purity tumor samples. The heatmap presents the significance of increased (or decreased) coordinated expression changes in genes associated with the UPR and/or ER stress, where sets are colored according to the MSigDB gene set library from which they originate (Hallmark, Reactome, or GO Biological Process). *The full name of the abbreviated gene set is "positive regulation of transcription from RNA polymerase II promoter in response to ER stress".

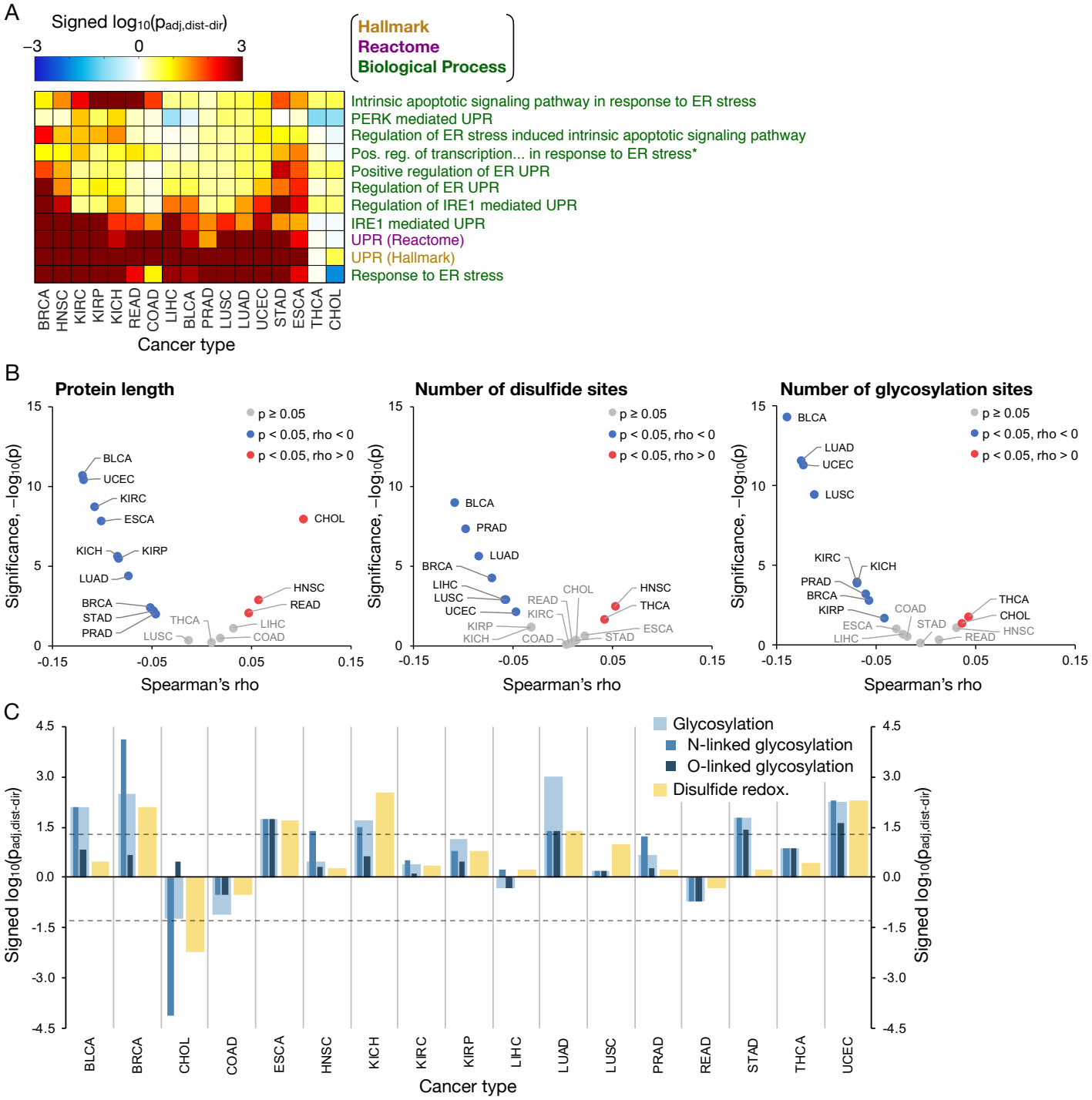

Figure S5. Related to Figures 5 and 6. (**A**) Coordinated expression changes in various UPR-related gene sets. Similar to Fig. 5, the heatmap presents the significance of increased (or decreased) coordinated expression changes in genes associated with the UPR, but includes additional UPR and ER stress related sets obtained from other gene set libraries as well (Hallmark, Reactome, and GO Biological Process). *The full name of the abbreviated gene set is "positive regulation of transcription from RNA polymerase II promoter in response to ER stress". (**B**) Correlation of SP gene expression fold-changes with individual terms of the SB score. A Spearman correlation was performed, comparing expression fold-changes (tumor vs. normal) of SP genes with their corresponding protein length (*left*), number of disulfide sites (*center*), and number of glycosylation sites (*right*). For each comparison and cancer type, the correlation significance ($-\log_{10}$p) is plotted as a function of the correlation coefficient (rho). Cancer types with a significantly ($p < 0.05$) negative correlation are colored blue, significantly positive cancers are colored red, and those with an insignificant correlation are in gray. (**C**) Coordinated expression changes in genes associated with PTM activities. Shown are the log-transformed directional p-values representing the significance of coordinated expression changes in genes associated with glycosylation (light blue) or disulfide oxidation/reduction (yellow), defined as those in the GO bioprocess "glycosylation" or GO molecular function "protein disulfide oxidoreductase activity" gene sets, respectively. Results for the more specific N- and O-linked subsets of glycosylation activity are also shown (blue and dark blue bars, respectively), as defined by the GO bioprocess "protein N-linked glycosylation" and "protein O-linked glycosylation" gene sets. Horizontal dashed lines correspond to $\pm\log_{10}$-transformed p-values of 0.05 (~1.3), where bars lying above the positive line or below the negative line are considered significant.

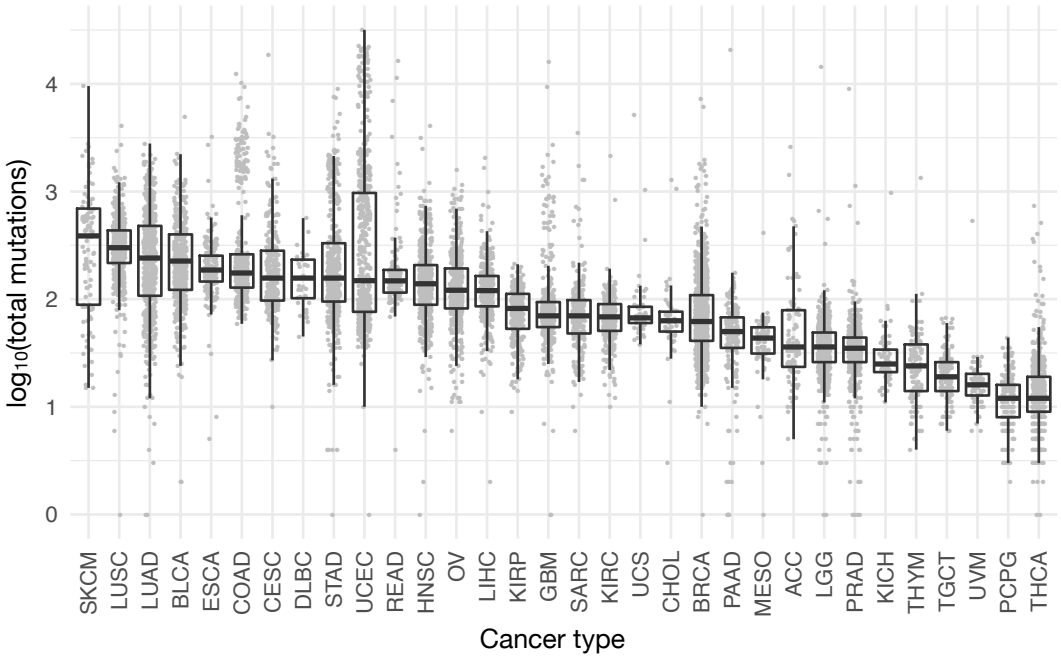

Figure S6. Somatic mutation burden of TCGA primary tumor samples for each cancer type. Related to Figures 5 and 6. The total number of somatic mutations per sample was determined from whole-exome sequencing data from TCGA, and are presented as box and whisker plots for each cancer type. The individual sample values are shown as gray points. Cancer types are sorted by descending median sample mutation burden.