

## RESEARCH ARTICLE

# A pan-cancer analysis of matrisome proteins reveals CTHRC1 and a related network as major ECM regulators across cancers

Keerthi Harikrishnan<sup>1</sup>\*, Srinivas Sheshagiri Prabhu<sup>1</sup>, Nagaraj Balasubramanian<sup>1</sup>\*

Indian Institute of Science Education and Research (IISER) Pune, Pashan, Pune, Maharashtra, India

\* These authors contributed equally to this work.

\* [dr.keerthiharikrishnan@gmail.com](mailto:dr.keerthiharikrishnan@gmail.com) (KH); [brandbombay@gmail.com](mailto:brandbombay@gmail.com) (NB)

## Abstract

The extracellular matrix in the tumour microenvironment can regulate cancer cell growth and progression. A pan-cancer analysis of TCGA data from 30 cancer types, identified the top 5% of matrisome genes with amplifications or deletions in their copy number, that affect their expression and cancer survival. A similar analysis of matrisome genes in individual cancers identified CTHRC1 to be significantly altered. CTHRC1, a regulator of collagen synthesis, was identified as the most prominently upregulated matrisome gene of interest across cancers. Differential gene expression analysis identified 19 genes whose expression is increased with CTHRC1. STRING analysis of these genes classified them as ‘extracellular’, involved most prominently in ECM organization and cell adhesion. KEGG analysis showed their involvement in ECM-receptor and growth factor signalling. Cytohubba analysis of these genes revealed 13 hub genes, of which MMP13, POSTN, SFRP4, ADAMTS16 and FNDC1 were significantly altered in their expression with CTHRC1 and seen to affect survival across cancers. This could in part be mediated by their overlapping roles in regulating ECM (collagen or fibronectin) expression and organisation. In breast cancer tumour samples CTHRC1 protein levels are significantly upregulated with POSTN and MMP13, further supporting the need to evaluate their crosstalk in cancers.

## OPEN ACCESS

**Citation:** Harikrishnan K, Prabhu SS, Balasubramanian N (2022) A pan-cancer analysis of matrisome proteins reveals CTHRC1 and a related network as major ECM regulators across cancers. *PLoS ONE* 17(10): e0270063. <https://doi.org/10.1371/journal.pone.0270063>

**Editor:** Nikos K Karamanos, University of Patras, GREECE

**Received:** February 4, 2022

**Accepted:** June 2, 2022

**Published:** October 3, 2022

**Copyright:** © 2022 Harikrishnan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting information files](#).

**Funding:** This study was supported by the Indian Council of Medical Research [35/03/2019-NANO/BMS] in the form of funds to NB. This study was also supported by the Department of Science and Technology, Ministry of Science and Technology [SR/WOS-A/LS-158/2017] in the form of funds to KH.

## Introduction

Extracellular matrix (ECM) is a dynamic interconnected mesh of macromolecules that provides structural support and also regulates cellular behaviour via mechanical and biochemical cues. It regulates several cellular processes including proliferation, differentiation, migration, invasion and survival [1]. ECM composition is tightly regulated by the cells and changes in ECM production, secretion, deposition and remodelling are reported in pathological diseases such as atherosclerosis, fibrosis, skeletal disorders, vascular disorders and cancer [2, 3]. ECM composition also varies between the tumour cells, tumour stroma and it is distinctly different across the metastatic sites [4, 5]. Recent studies characterizing changes in the composition of ECM in normal and tumour microenvironments have emphasized the importance of ECM and its contribution towards developing novel biomarkers and therapeutic targets [5–7].

**Competing interests:** NO authors have competing interests.

The “Matrisome” pioneered by *Naba et al.*, is an ensemble of genes that codes for core ECM proteins, ECM associated growth factors, ECM regulators and other ECM associated factors [6, 8]. It accounts for 4% of the human and mouse genome and reflects the composition of proteins in normal and tumour tissues [3, 9]. Since the publication of MatrisomeDB in 2012, the understanding of the role of ECM in cancers has significantly enhanced. Matrisome proteins like Insulin like growth factor binding protein 3, 4, 5 (IGFBP3, IGFBP4, IGFBP5), Cellular Communication Network factor (CCN) family, Thrombospondin2 (THBS2), Tenascin N (TNN) and Von Willebrand factor A 9 (VWA9) are detected primarily in cancer tissues [10] while Lysyl oxidase like 2 (LOXL2) [11], Cartilage oligomeric matrix protein (COMP), Periostin (POSTN) [12], Tenascin N (TNC) [13], Tenascin X (TNX) [13] and Fibronectin (FN) (EIIIA and EIIIB variant) [14] have been reported to be upregulated in cancers. OncoPrint analysis of core matrisome genes in the lung, gastric, ovarian and colon cancers show that a signature of 9 genes Collagen type XI alpha 1 chain (COL11A1), Secreted phosphoprotein 1 (SPP1), Microfibrillar associated protein 2 (MFAP2), Collagen type X alpha 1 chain (COL10A1), Biglycan (BGN), Cartilage oligomeric matrix protein (COMP), Agrin (AGRN) and Matrix remodelling associated 5 (MXRA5) is associated with poor survival and is involved in regulating cancer hallmarks such as epithelial to mesenchymal transition (EMT), and angiogenesis [15]. Data mining of 10 NSCLC microarray datasets has identified 29 ECM signature genes which were found to be consistently upregulated in patients with NSCLC and also predicts prognosis [16]. Analysis of 12 cancer types (lung, pancreas, prostate, kidney, stomach, colon, ovary, breast, liver, bladder and skin) shows that tumour matrisome index (TMI) is associated with disease progression and poor clinical outcome [17]. In addition, tumours with high TMI show enrichment for Mage family member A3 (MAGEA3) and CD8 positive T cells and also display high expression of B7-H3 which is negatively associated with clinical outcome in solid tumours [18].

Pan-Cancer analysis of transforming growth factor  $\beta$  (TGF $\beta$ ) associated ECM gene expression shows a set of matrisome genes to be upregulated in cancer and the expression is associated with a worse prognosis. This study also reveals an association of aberrant ECM expression with immunosuppression in cancers [17, 19]. Cell-cell adhesion, Forkhead box O (FOXO), Wnt pathways are found to control matrisome in most cancer types whereas tumour protein 53 (TP53), Notch and TGF $\beta$  signalling pathways regulate matrisome genes in some cancers [20].

Using a multi-omics approach and machine learning, several landmark matrisome genes have been identified from 74 clinical and molecular subtypes of cancers that show prognostic significance [21]. Bioinformatic analysis of the copy number alterations (CNA's) reveals that matrisome genes display a disproportionately high number of CNA's and mutations compared to the rest of the genome [22] across cancers. This increase in the genome alterations of matrisome was further predictive of prognosis across cancer types. Together, these findings highlight a significant role of ECM genes in cancer progression.

While recent studies have evaluated the role of matrix protein families or individual cancer matrisome signatures, our study was aimed at identifying matrisome genes that can act as vital ECM regulators in pan-cancer analysis and individual cancers. Our pan-cancer analysis of matrisome genes for copy number variation (amplification or deletion), relative expression and effect on cancer survival identified collagen triple helix repeat containing 1 (CTHRC1) as a major pan-cancer ECM regulator. Its role is further supported by CTHRC1 being identified as the most prominently regulated ECM protein across individual cancers. Further, network analysis reveals CTHRC1 could work with matrisome genes Periostin (POSTN), Matrix metalloproteinase 13 (MMP13), Secreted frizzled related protein 4 (SFRP4), Fibronectin type III domain containing 1 (FNDC1) and ADAM metalloproteinase with thrombospondin type 1 motif (ADAMTS16) in regulating the impact ECM has on cancers.

## Materials and methods

### Data sources

The list of matrisome genes was downloaded from the matrisome database [23]. TCGA Pan-Cancer copy number data calculated using the GISTIC2 threshold method was downloaded from the UCSC Xena browser [24]. TCGA and GTEx data was used to perform the expression, survival, correlation, cooccurrence disease stage and protein analysis. The analysis was restricted to 30 types of cancers from TCGA which include: Adrenocortical Carcinoma (ACC), Bladder Urothelial Carcinoma (BLCA), Breast Invasive Carcinoma (BRCA), Cervical Squamous Cell Carcinoma (CESC), Cholangiocarcinoma (CHOL), Colon Adenocarcinoma (COAD), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC), Esophageal Carcinoma (ESCA), Glioblastoma Multiforme (GBM), Head and Neck Squamous Cell Carcinoma (HNSC), Kidney Chromophobe (KICH), Kidney Renal Clear Cell Carcinoma (KIRC), Kidney Renal Papillary Cell Carcinoma (KIRP), Brain Lower Grade Glioma (LGG), Liver Hepatocellular Carcinoma (LIHC), Lung Adenocarcinoma (LUAD), Lung Squamous Cell Carcinoma (LUSC), Ovarian Serous Cystadenocarcinoma (OV), Pancreatic Adenocarcinoma (PAAD), Pheochromocytoma and Paranglioma (PCPG), Prostate Adenocarcinoma (PRAD), Rectum Adenocarcinoma (READ), Sarcoma (SARC), Skin Cutaneous Melanoma (SKCM), Stomach Adenocarcinoma (STAD), Testicular Germ Cell Tumours (TGCT), Thyroid Carcinoma (THCA), Thymoma (THYM), and Uterine Corpus Endometrial Carcinoma (UCEC), Uterine Carcinosarcoma (UCS).

### Copy number variation analysis

First genes were clustered based on their functionality (ECM genes, Proteoglycans etc.). These gene clusters were prepared as an excel file containing a single column with gene symbols as cell entries and this list represented the genes of interest whose copy number variations were to be analyzed. There were 10845 samples in total and the gene-level copy number estimate values of -2, -1, 0, 1, 2 represented deep deletion, shallow deletion, no change, amplification and gain respectively. The excel file retrieved after extracting the dataset was in the genomic Matrix format (ROWS (identifiers) x COLUMNS (samples)). Next, we wrote a code in Python to find the gene-level copy number estimate values of our genes of interest across all 10845 samples. To achieve this, we had to explore our list in the database retrieve the corresponding values for the samples and process the data to calculate the number of deep deletions and gain. The top 5% of the genes (n = 104) with amplification or deep deletion were then used for further analysis. The code used to perform the analysis can be made available upon written request.

### Expression, survival, correlation and disease stage analysis

RNA expression for the matrisome genes was analyzed using the GEPIA2 portal [25] which contains the expression data for 9736 tumour samples and 8587 normal samples and the data is processed using a standard processing pipeline. The data from 30 cancer types are represented in a box plot (log scale) as mean  $\pm$  standard deviation (S.D). We also evaluated the expression of the matrisome genes by pathological stage using the stage plot function in the database. The data from 30 cancer types are represented in a violin plot (log scale) which shows the distribution of data. p-value less than 0.05 was considered statistically significant. Survival analysis for all the matrisome genes was performed using TCGA data through the GEPIA2 portal. A custom cutoff was set for the survival analysis with the top 75 percentile being classified as “high” and the bottom 25 percentile classified as “low”. This classification

was used to evaluate the significance of the difference in effect on survival that exists between the “high” and “low” groups. Log-rank p values and the hazard ratio (HR) with 95% confidence interval were calculated using the GEPIA2 web portal and a p-value of less than 0.05 was considered statistically significant. This was used to classify individual cancers as affecting survival ( $p < 0.05$ ) or not ( $p > 0.05$ ).

Percentage effect on expression and survival was calculated as follows:

$$\text{Percentage affected} = \frac{\text{number of cancers significantly affected}}{\text{total number of cancers}(n = 30)}$$

Univariate and Multivariate Survival analysis for Figs 1B and 3B was performed using TCGA data with the survival analysis code in the TCGA2STAT package in R (version 3.6.3) Analysis 2 (univariate), 3 (multivariate) [26].

Correlation analysis for CTHRC1 along with its 10 hub genes across 30 cancer types was done using the Spearman correlation coefficient. Correlation where  $p < 0.05$  was considered statistically significant.

### Scoring of genes based on expression and effect on survival

Genes selected based on their copy number agreeing with mRNA expression (Fig 2A) were arranged in descending order based on their upregulation or downregulation across cancers. The data plotted in these graphs represent percentage alteration calculated as above. A score was assigned for each gene based on their position in the upregulated or downregulated graphs respectively. The top gene was assigned a score of 1 and this score increased by 1 point for the next gene. This was done for all genes in the graph.

All of the above genes (upregulated or downregulated) were further arranged based on their effect on survival across cancers in descending order. A score was assigned for each gene based on its position in the survival graph. The top gene was assigned a score of 1 and this score increased by 1 point for the next gene. This was done for all genes in the graph.

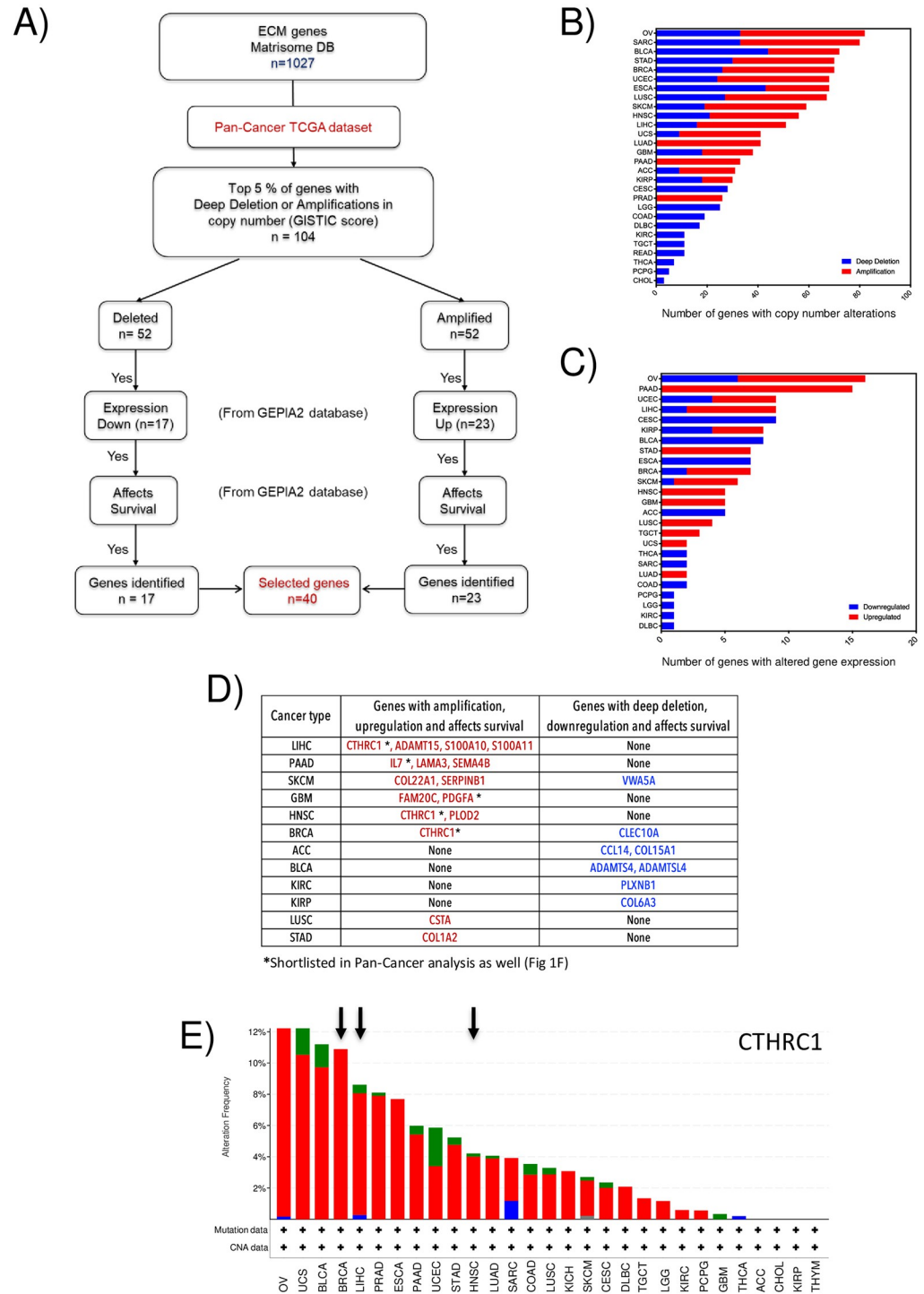
These two scores were added to obtain the final score for upregulated and downregulated genes. They were arranged in a table based on their score (low to high). A low score indicates a higher ranking in expression (up or down) and survival.

### Mutation and co-occurrence analysis

Mutation analysis for 8 shortlisted genes (based on expression, survival data as mentioned above) was analyzed using the cBioPortal database across 30 cancer types. It showed the mutational burden along with the copy number analysis in the patient samples. We also used this portal to identify the co-expressed genes of CTHRC1 (30 cancer types) in the cBioPortal using the cooccurrence-mutual exclusivity tab in the database.  $p$  less than or equal to 0.05 was considered statistically significant.

### Differential gene expression analysis

First, the data was preprocessed locally in R using the preprocessing script. The differential gene expression (DEG) analysis was performed (30 cancer types) in groups with high expression (75%) or low (25%) expression. Thresholds for DEG's was set at fold change  $> 2$ ,  $-2$  with a p-value of 0.05. The following source code from the GitHub repository was used to perform the preprocessing and the DEG analysis [27–29]



**Fig 1. Cancer type-specific analysis of matrisome genes.** **A)** Schematic shows the steps involved in shortlisting genes from the individual TCGA cancer dataset based on copy number (GISTIC score), expression and survival (GEPIA2 database). The number of genes that met the criteria for each stage and were eventually shortlisted are indicated in the box for genes identified. Their occurrence in two or more cancers was used for the final selection (n = 1). **B)** Based on their copy number, the number of deep deleted (BLUE bar) or amplified (RED bar) matrisome genes in individual cancers (n = 28) were arranged in the descending order based on total genes affected (BLUE+RED). **C)** Based on their mRNA expression, the top 5% of matrisome genes in individual cancers (n = 28) that are downregulated (BLUE bar) or upregulated (RED bar) were arranged in descending order based on total genes affected (BLUE+RED). **D)** Table lists the individual cancers with one or more amplified and upregulated (RED) or deleted and downregulated (BLUE) genes that

affect survival. Genes marked with an asterisk (\*) are also shortlisted in the pan-cancer analysis (Fig 2F). Genes marked in **bold** are affected by more than one cancer type. E) Graph represents the mutational (GREEN), copy number amplification (RED) and deletion (BLUE) analysis of CTHRC1 in 30 individual cancers. Arrows point to the individual cancers where CTHRC1 is also selected as detailed above (Fig 1D).

<https://doi.org/10.1371/journal.pone.0270063.g001>

### Venn diagram analysis for overlapping genes

Omics Box software and the Venn diagram tool [30] were used to identify the overlapping differentially expressed genes across cancer types. 20 genes that overlapped in 3 or more cancers were used for further analysis. Venn diagram tool was also used to identify the top hub genes that overlapped in survival, disease stage expression, correlation and co-occurrence analysis. 3 genes identified were then used to perform CPTAC analysis.

### Protein-protein interaction and functional enrichment analysis

STRING database [31] was used to identify the protein-protein interactions of CTHRC1 along with the differentially expressed genes across different cancer types. The list of proteins along with CTHRC1 was entered in the online database and a medium confidence interaction score of 0.4 was used to generate the full string network. The network was then exported as a high-resolution image. Functional enrichment analysis performed using the STRING database included biological process, molecular function, cellular compartment and KEGG pathways. FDR of < 0.05 was used to identify the gene ontology (GO) terms that were statistically significant.

### CytoHubba analysis

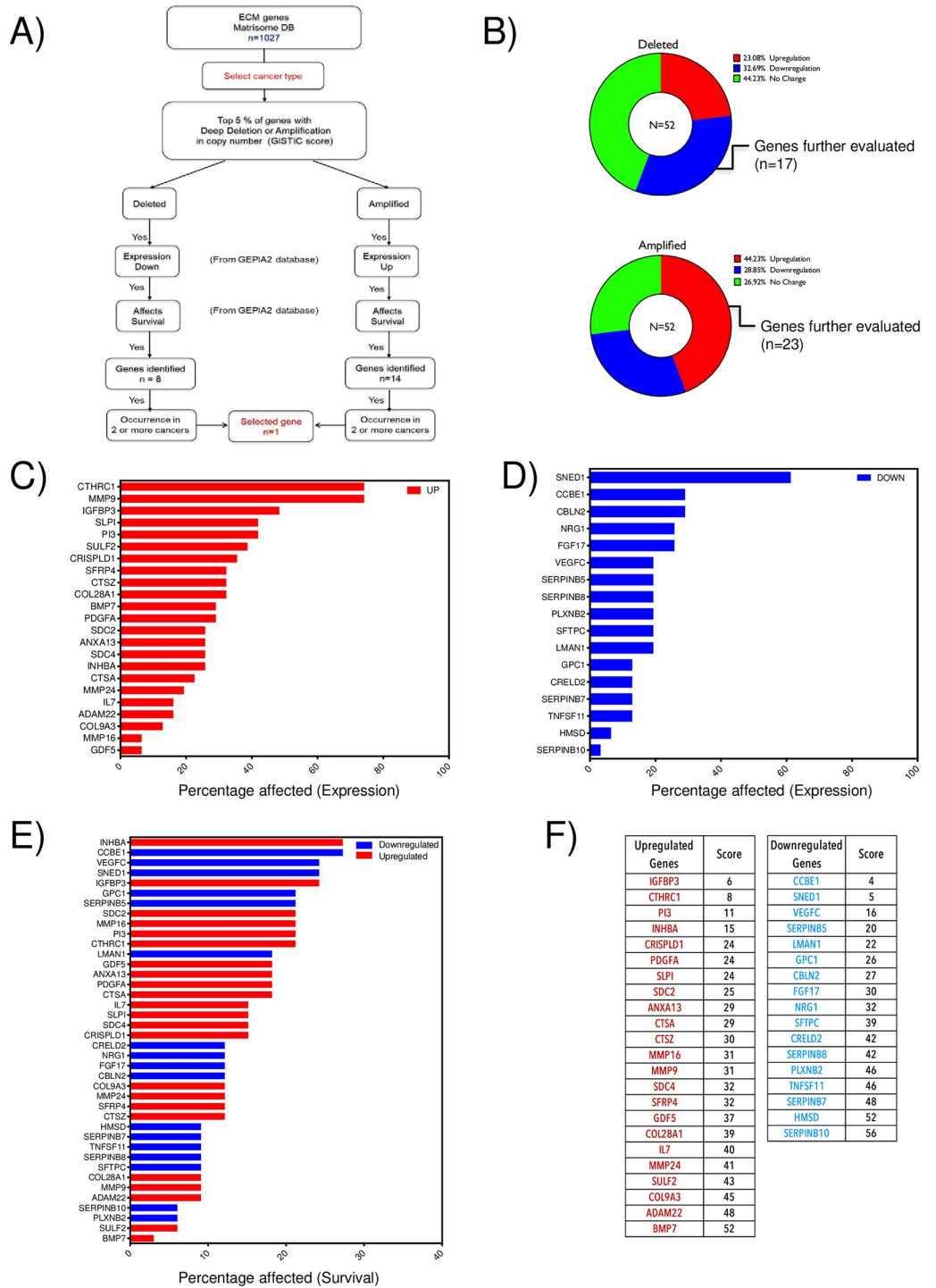
The PPI network constructed using the STRING database was sent to Cytoscape using the web link. Cytoscape (version 3.8.2) is an open-source software used for the visualization and analysis of protein-protein interaction networks. Using the cytoHubba tool in the software, we identified the top 14 hub genes based on the degree of connectivity. These hub genes were then used for further analysis.

### GeneMANIA PPI network analysis

GeneMANIA [32] was used to construct a PPI network of CTHRC1 along with its hub genes (POSTN, MMP13, FNDC1, SFRP4 and ADAMTS16) to identify if these genes are functionally related and discover genes that could be part of this functional network. The networks were generated using the following weighing methods: 1) Based on query genes and 2) Gene Ontology. In the gene ontology-based weighing method, Biological Process and Cellular component methods were used for building networks. GeneMANIA will build networks showing genetic interactions, protein-protein interactions, protein-DNA interactions, protein expression, similarity in protein domains, pathways and phenotypic screening profiles using the publicly available datasets.

### UALCAN analysis

UALCAN is a web-based tool for the analysis of CPTAC data from the TCGA cancer types. We used the UALCAN database [33] to analyze the protein levels of CTHRC1, POSTN, MMP13, SFRP4 and FNDC1 in breast cancer. Data were represented as mean  $\pm$  S.D and a p-value of less than 0.05 were considered statistically significant.



**Fig 2. Pan-cancer analysis of matrisome genes.** A) Schematic shows the steps involved in shortlisting genes from the pan-cancer TCGA dataset based on copy number (GISTIC score), expression and survival (GEPIA2 database). The number of genes shortlisted at each stage of the selection process is indicated in each box. B) Deleted (n = 52) and amplified (n = 52) genes were classified based on their mRNA expression as downregulated (BLUE), upregulated (RED) and no change (GREEN). The nested bar graph represents the percentage of each for deleted (top graph) and amplified (bottom graph) genes. C) Bar graph shows the percentage of cancers where the top 5% of amplified matrisome genes are also upregulated (n = 23). Genes are arranged in descending order (RED—represents upregulated genes). D) Bar graph shows the percentage of cancers where the top 5% of deleted matrisome genes are also downregulated (n = 17). Genes are arranged in descending order (BLUE—represents downregulated genes). E) Bar graph shows the percentage of survival in

cancers for the top 5% of matrisome genes shortlisted in C (RED bar) and D (BLUE bar). These are arranged in descending order. F) Tables list genes in the ascending order of their score calculated based on their position in the expression (C, D) and survival (E) graphs (as detailed in methods). A lower score is indicative of a higher position in these graphs. Upregulated genes are listed in RED and downregulated genes are listed in BLUE.

<https://doi.org/10.1371/journal.pone.0270063.g002>

## Results

### Pan-cancer analysis of altered matrisome genes

In this study, we first obtained the list of matrisome genes ( $n = 1027$ ) from the Matrisome database (DB). We then analyzed the copy number variation of these genes using the Pan-Cancer TCGA dataset from the UCSC cancer genome browser (Fig 2A). Overall, there were 10845 samples in this dataset which were scored based on their copy number variations (GISTIC score from UCSC Xena browser) (Fig 2A). The top 5% of genes with deep deletions or amplifications were identified and shortlisted (104 genes). Using the GEPIA2 portal, we evaluated the mRNA expression of these 104 genes across 30 cancer types. Of these deep deleted and amplified genes, 32.69% showed reduced mRNA expression (17 genes) and 44.23% showed increased mRNA expression (23 genes) respectively (Fig 2B). Upregulated (Red Graph) and downregulated (Blue Graph) genes were arranged in descending order as per their change in pan-cancer expression (Fig 2C and 2D). Similarly, the effect of these genes on pan-cancer survival was determined using the GEPIA2 portal and genes were arranged in descending order (Fig 2E) (the colour code used for expression of genes was retained in the graph). The relative position of genes in the expression and survival graphs was used to score them as detailed in the methods section. The final score thus obtained was used to arrange upregulated (red) and downregulated (blue) genes in the descending order (Fig 2F). A low score indicates a higher ranking for the gene in expression and survival.

### Individual cancer analysis of altered matrisome genes

Further, we analyzed the copy number, expression and effect on survival of matrisome genes in individual cancers (Fig 1A). The pooling of cancer types in the above pan-cancer analysis could be sensitive to tumours with significant copy number alterations that could skew the selection of gene(s) of interest. Combining the pan- and individual cancer analysis could be better at identifying matrisome gene(s) of consequence. TCGA data for 30 individual cancer types obtained from the UCSC cancer genome browser was used for copy number analysis. This showed THYM and KICH cancers to lack deep deletions or amplifications in any of the matrisome genes (*data not shown*). Of the 28 remaining cancers, 15 showed both deep deletions and amplifications in their matrisome genes, while 10 showed only deep deletions and 3 had only amplifications. The total number of genes with amplification (red bar) and deep deletions (blue bar) for individual cancers were represented in the graph and arranged in descending order (Fig 1B). In each of these 28 cancers, we selected the top 5% of genes whose copy number is altered (either deleted or amplified) (Fig 1A). Using the GEPIA2 portal, we evaluated the mRNA expression of these genes. Genes with deep deletion and reduced mRNA expression or amplification and increased mRNA expression were shortlisted for each cancer (Fig 1A). The number of so-identified upregulated (red bar) and downregulated (blue bar) genes for each cancer were plotted and they were arranged in descending order (Fig 1C). We further tested if these shortlisted genes in their respective cancers significantly affected survival (Fig 1A). This revealed 14 amplified (red) and 8 deleted (blue) genes to affect survival in 12 individual cancers, listed in the table for Fig 1D. Of these genes, CTHRC1, PDGFA, and IL7 were also shortlisted as matrisome genes of interest in the pan-cancer analysis (Figs 1D and



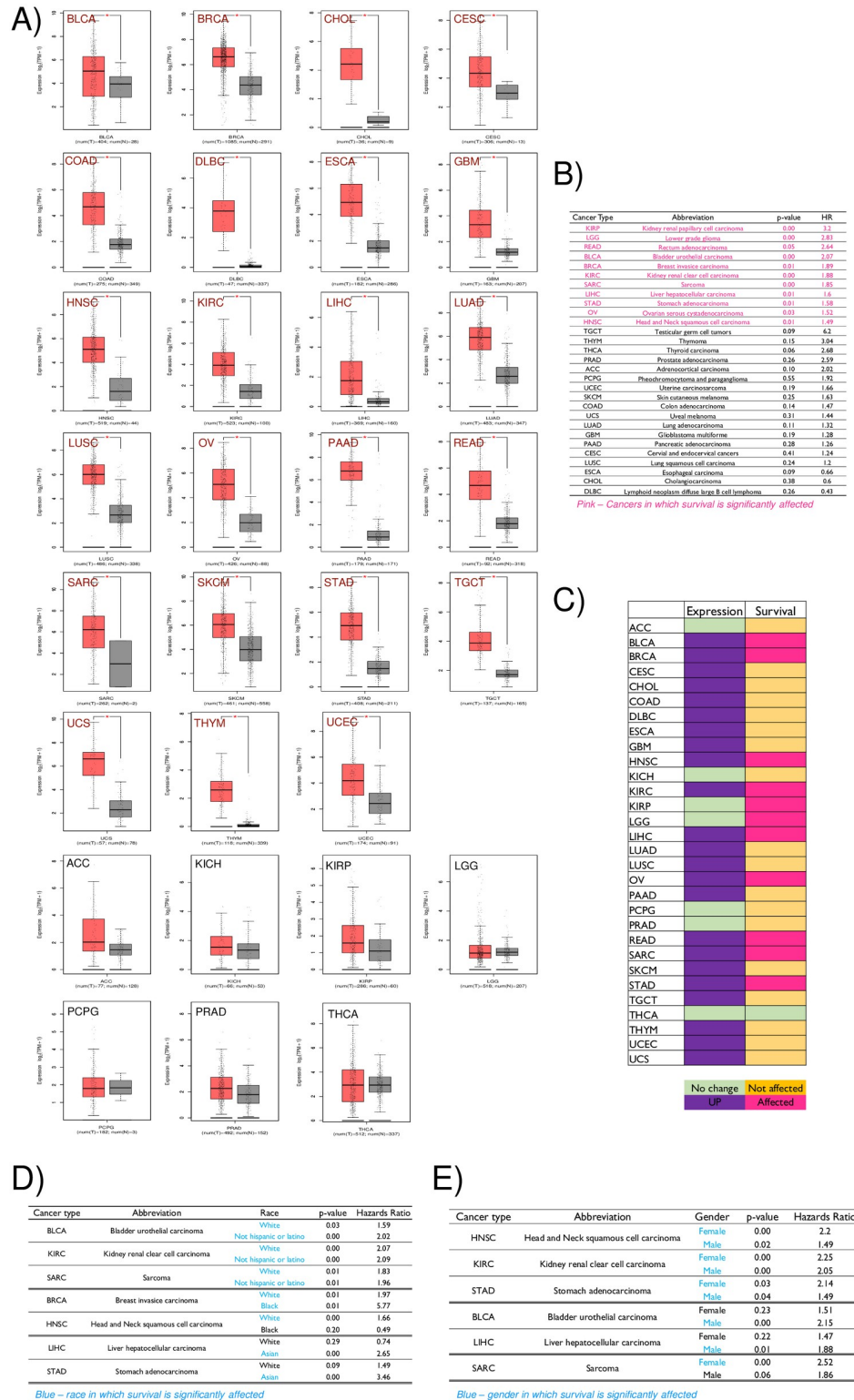
2F). CTHRC1 was however the only gene that was affected in more than one individual cancer type (BRCA, HNSC and LIHC) (Fig 1D). Copy number variation and mutation analysis for CTHRC1 across cancers find it to be prominently amplified in most cancers (Fig 1E) including BRCA, HNSC and LIHC (marked by arrow). This led us to choose CTHRC1 as the matrisome gene of interest for further evaluation.

### Detailed Pan-cancer analysis of CTHRC1

CTHRC1 (Collagen triple helix repeat containing-1) expression (evaluated in 30 individual cancers using GEPIA2 portal confirmed its overexpression in 23 cancers (Fig 3A—labels in red and Fig 3C—boxes coloured in purple). In 7 cancers no change in expression was observed (Fig 3A—labels in black and Fig 3C—boxes coloured in green). We also compared the effect CTHRC1 has on survival in 30 cancers using univariate and multivariate analysis. Using TCGA data univariate analysis was done to calculate the hazards ratio (HR), which captures the likelihood of CTHRC1 expression affecting survival in cancers (Fig 3B). The calculated significance showed CTHRC1 expression to significantly affect survival in 11 cancers (Fig 3B and 3C—boxes in pink). A comparison of the expression and survival data to identify cancers significantly affected in both reveals 9 individual cancers (Fig 3C). This includes the 3 CTHRC1 overexpressing individual cancers (BRCA, HNSC and LIHC) identified earlier (Fig 1D and 1E). The table in Fig 3C marks cancers with significant CTHRC1 expression in purple and their significant effect on survival in pink. Multivariate survival analysis for race (data available in 7 cancers) and gender (data available in 6 cancers) was also done. In these cancers with available data, a comparison for race between, 1) White vs Not Hispanic or Latino, 2) White vs Black and 3) White vs Asian revealed the following. In HNSC, survival was affected significantly for white but not black patients. In LIHC and STAD survival was significantly affected in Asian patients but not in white patients. In BRCA, black and white patients were both significantly affected though their hazard ratios were distinctly different (Black— 5.77 and White— 1.97) (Fig 3D). This suggests black patients with “high” CTHRC1 expression (top 75 percentile) are 3.8 times more likely to have poor survival compared to white patients in the top 75 percentile. Data for gender when compared revealed a significant effect on survival in only males in BLCA and LIHC and only females in SARC. Males and females were significantly affected in HNSC, KIRC and STAD (Fig 3E). Taken together, the multivariate analysis across CTHRC1 overexpressing cancers does not reveal any distinctly conserved effect on survival across race or gender. These effects when seen do seem to be limited to individual cancer types.

### Identifying possible genes involved in CTHRC1 dependent cancers

Differential gene expression (DEG) analysis of the 9 CTHRC1 dependent cancers, BLCA, BRCA, HNSC, KIRC, LIHC, OV, READ, SARC and STAD was done to identify genes that could be regulated by CTHRC1. The top 5% of these genes with a 2 fold increase or decrease ( $p < 0.05$ ) in expression in each of the above cancer types were selected for further analysis (Fig 4A, S1A - S1J Files in S1 File). This identified 19 genes that were upregulated in 3 or more cancers (Fig 4B and 4C), but none that were downregulated. The Venn diagram (Fig 4B) shows the overlapping upregulated genes. The 19 upregulated genes include Collagen type XI alpha 1 chain (COL11A1), Secreted Frizzled related protein 2 (SFRP2), Periostin (POSTN), Epiphygan (EPYC), Cartilage oligomeric matrix protein (COMP), Collagen type X alpha 1 chain (COL10A1), Osteomodulin (OMD), Leucine rich repeat containing 15 (LRR15), Secreted frizzled related protein 4 (SFRP4), Phosphatidic acid phosphatase type 1A (PPAPDC1A), ADAM metalloproteinase with thrombospondin type 1 motif (ADAMTS16), Osteoglycin



**Fig 3. Pan-cancer analysis of CTHRC1 expression and its effect on survival.** A) Graphs represent CTHRC1 expression data from the GEPIA2 portal in 30 different tumour types (T-RED bar) relative to normal (N-GREY bar). Cancers showing significant upregulation in CTHRC1 are listed first and labelled in RED. Those showing no significant change in expression are listed later and labelled in BLACK. Expression data are represented as mean ± standard deviation (S.D) on a log scale using a box plot. p-value of < 0.05 was determined as statistically

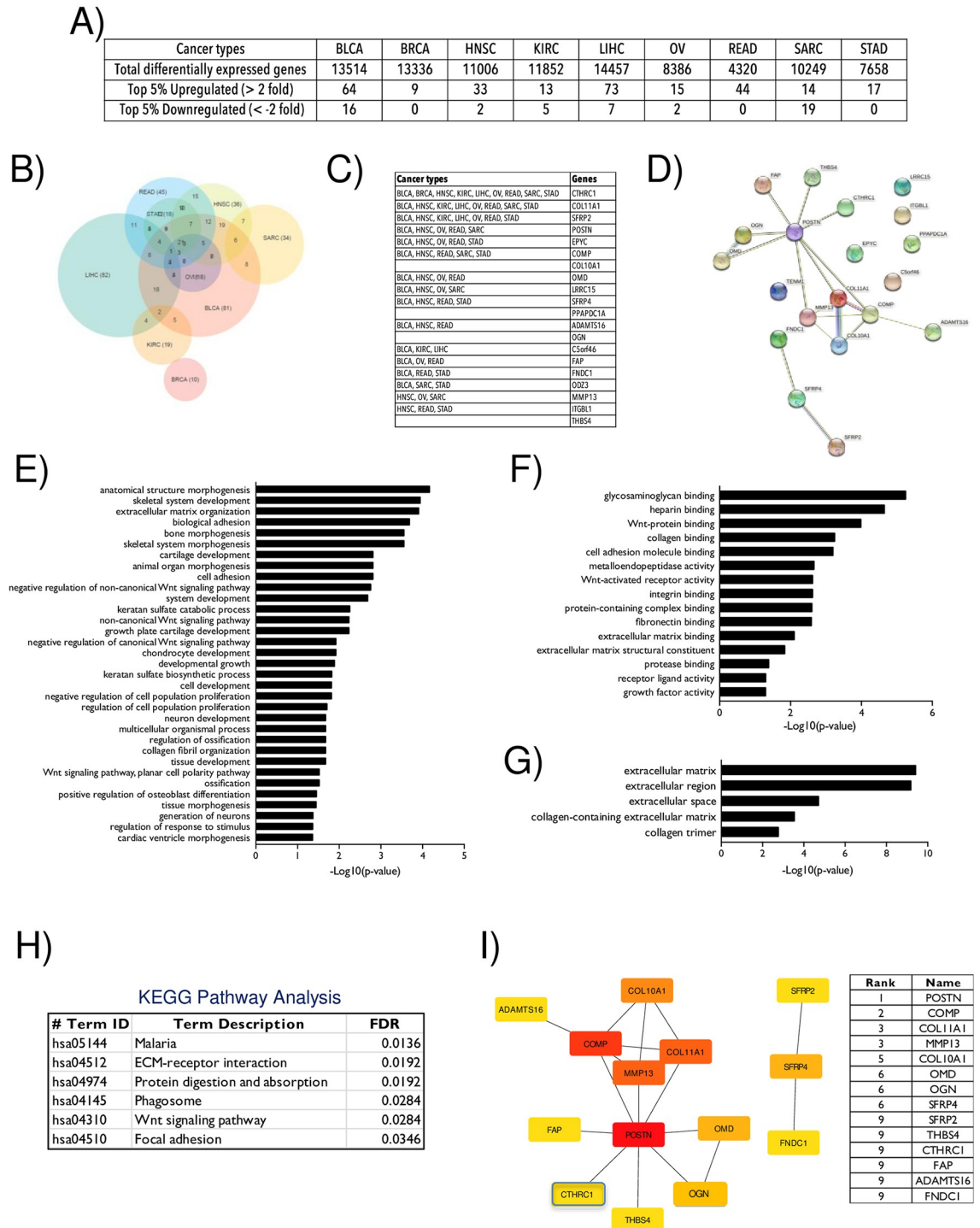
significant. **B)** Table lists the results of univariate analysis of CTHRC1 expression on survival across 30 individual cancers. It shows the significance values (p-value) for survival in patients with “high” vs “low” CTHRC1 expression and their hazards ratio (HR). Cancers with significance  $p \leq 0.05$  are listed in PINK and  $p > 0.05$  in BLACK in the descending order of their respective hazards ratio. **C)** Table shows CTHRC1 expression and survival data in 30 individual cancers. Upregulated (PURPLE), and comparable (GREEN) expression marked accordingly. Significant effect seen on survival is marked in PINK and lack thereof marked in ORANGE. **D-E)** Tables shows the multivariate survival analysis for CTHRC1 expression in the context of **(D)** race and **(E)** gender in selected cancers for which data is available. It shows the significance values (p-value) for survival in patients with “high” vs “low” CTHRC1 expression and their hazards ratio (HR) for comparison. Cancers with significance  $p < 0.05$  are listed in their descending order of significance.

<https://doi.org/10.1371/journal.pone.0270063.g003>

(OGN), Chromosome 5 open reading frame 46 (C5orf46), Fibroblast activation protein alpha (FAP), Fibronectin type III domain containing 1 (FNDC1), Teneurin transmembrane protein 1 (ODZ3/TENM1), Matrix metalloproteinase 13 (MMP13), Integrin subunit beta like 1 (ITGBL1) and Thrombospondin 4 (THBS4) (Fig 4C). They were used to construct a protein-protein interaction network using the STRING database (Fig 4D). Co-expression, database information and text mining are what largely contribute to the making of this network. There is hence a vital need to experimentally validate this network. Functional enrichment analysis of this network revealed ECM organization and adhesion to be among the top 5 biological processes regulated by these network genes (Fig 4E). At the molecular level, these proteins bind Fibronectin, Collagen, cell adhesion molecules and integrins among others (Fig 4F). Proteins in this network almost exclusively belong to the extracellular compartment (extracellular matrix, region and space) (Fig 4G). KEGG analysis of this network further shows their involvement in ECM-receptor interaction, Wnt and focal adhesion signalling (Fig 4H). cytoHubba tool identified 14 hub genes (with a score  $\geq 1$ ) from this predicted network which includes, POSTN, COMP, COL11A1, MMP13, COL10A1, OMD, OGN, SFRP4, SFRP2, THBS4, CTHRC1, FAP, ADAMTS16 and FNDC1 (Fig 4I). Of these POSTN ranks highest and has the most number of connections, which also includes CTHRC1.

To validate the significance of these hub genes with CTHRC1, we first analyzed the effect on pan-cancer survival of CTHRC1 and these 13 hub genes using the TCGA data through the GEPIA2 portal. Like CTHRC1, COL11A1, MMP13, COL10A1, POSTN, OGN, SFRP4, FAP, ADAMTS16 and FNDC1 were all seen to significantly affect survival in cancers with “high” expression (top 75 percentile) relative to “low” (bottom 25 percentile) (Fig 5A). THBS4 however significantly affects survival in cancers with “low” (bottom 25 percentile) relative to “high” (top 75 percentile) (Fig 5A). We used data from the GEPIA2 portal to determine if the change in expression of CTHRC1 and 13 hub genes is associated with tumour grade. Statistical analysis of this change across tumour grades by ANOVA showed all hub genes and CTHRC1 expression to indeed be tumour stage-dependent (Fig 5B). Spearman correlation analysis showed a positive correlation between all the 13 hub genes and CTHRC1 (Fig 5C). cBiportal analysis also showed a significant co-occurrence between CTHRC1 and OMD, POSTN, OGN, MMP13, COMP, SFRP4, ADAMTS16 and FNDC1 (Fig 5D). Based on the significant effect they have on survival, tumour grade, correlation and co-occurrence analysis, we identified five genes, POSTN, MMP13, SFRP4, ADAMTS16 and FNDC1 as the most likely mediators of CTHRC1 dependent function in cancers (Fig 5E—represented in the Venn diagram).

It would hence be of interest to look at the protein expression data in cancers for CTHRC1 and the now identified genes of interest, POSTN, MMP13, SFRP4, ADAMTS16 and FNDC1. Of the 9 cancers shortlisted in the pan-cancer analysis based on CTHRC1 overexpression and its effect on cancer survival (Fig 3A–3C), BRCA, HNSC and LIHC were also shortlisted in the individual cancer study based on their CTHRC1 expression and its effect on survival. Protein expression data for most of the network genes (4 out of 5) of interest was available for only



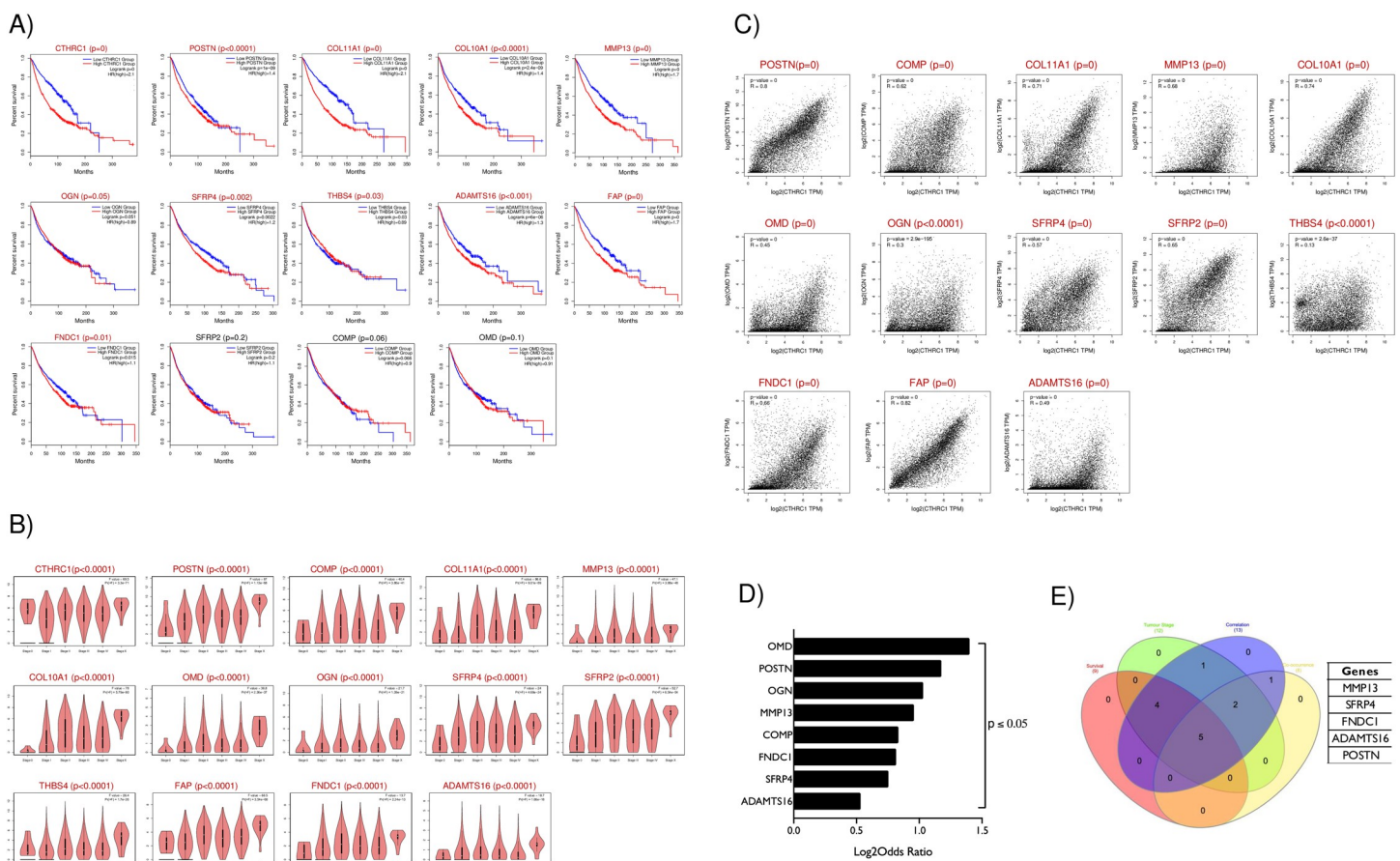
**Fig 4. Differential gene expression analysis of CTHRC1.** A) The table lists in the 9 selected cancers (BLCA, BRCA, HNSC, KIRC, LIHC, OV, READ, SARC and STAD—as detailed in Methods) the number of differentially expressed genes and the top 5% genes upregulated or downregulated with a 2 fold change (as detailed in Methods). B) This Venn diagram shows the overlap (if any) of the top 5% upregulated genes in the above listed 9 cancers. C) Table lists the 19 overlapping genes between 3 or more cancer types. D) Protein-protein interaction network constructed for CTHRC1 and its 19 differentially expressed genes using the STRING database. BLUE line marks predicted interactions from gene co-occurrence data, GREEN line marks predicted interactions based on gene neighbourhood evidence, PURPLE line marks experimentally determined known interactions, YELLOW line marks interactions based on text mining and the LIGHT BLUE line marks interactions based on database evidence. E-F) Functional enrichment for significant ( $p < 0.05$ ) (E) biological processes, (F) molecular functions and (G) cellular components in the STRING network analysis

are listed in their descending order of significance. (H) The table lists the pathways identified by KEGG analysis for the STRING network in the descending order of their significance (FDR). (I) Network of 13 hub genes identified using CytoHubba plugin in Cytoscape software. Colours of the hub genes are based on their rank which is also listed as a table (High to low).

<https://doi.org/10.1371/journal.pone.0270063.g004>

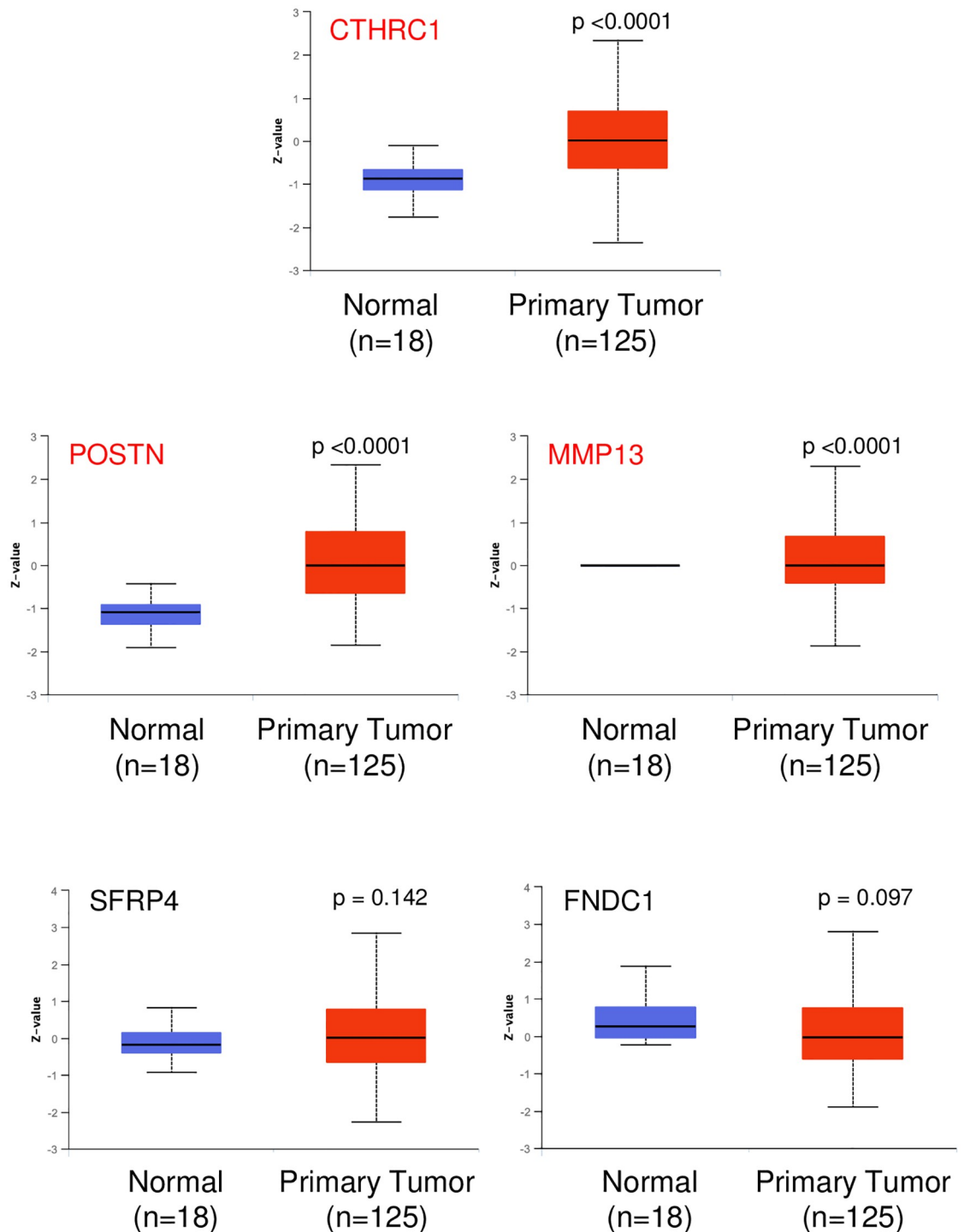
BRCA (breast cancer) in the UACLAN portal. Mass spectrometric data from 125 breast tumour samples showed CTHRC1, POSTN and MMP13 to all be significantly higher in breast tumour tissue relative to normal while SFRP4 and FNDC1 levels were unaffected (Fig 6A).

To further test if CTHRC1 and its related hub genes (POSTN, MMP13, SFRP4, ADAMTS 16 and FNDC1) share functional networks, we performed network analysis using GeneMANIA online database [34] and compared their co-expression and physical interaction networks. The co-expression network based on query-dependent weighting (Fig 7A) is seen to be rich in core matrisome genes, cell receptors and intracellular signalling molecules (S1A Fig). Co-expression and physical interaction networks based on Gene Ontology (GO) weighting for



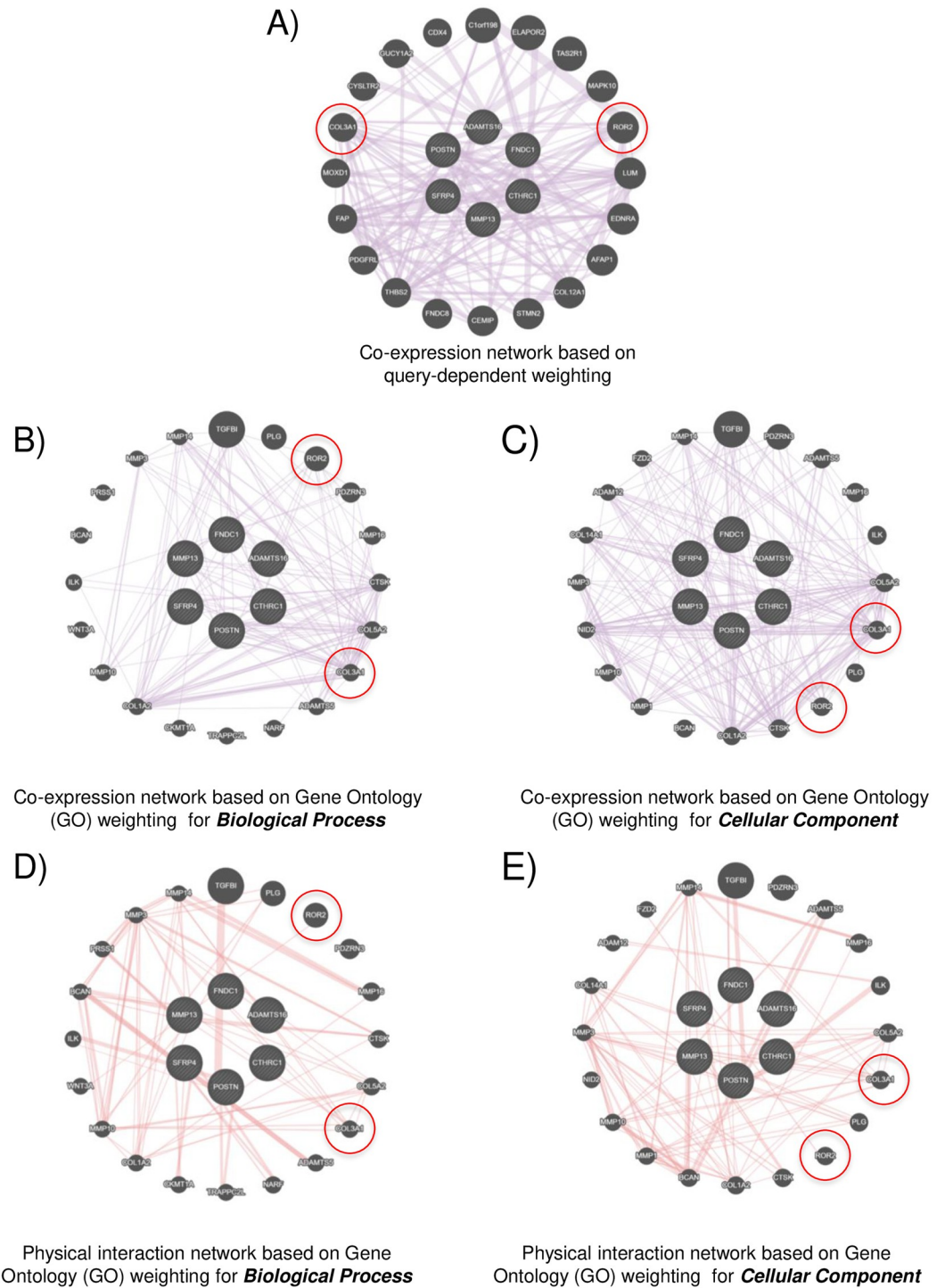
**Fig 5. Validation of CTHRC1 and its hub genes in cancers.** A) Graph represents percentage survival in 30 cancers with “high” (RED plot) vs “low” (BLUE plot) expression for CTHRC1 or each of its 13 hub genes (POSTN, COMP, COL11A1, MMP13, COL10A1, OMD, OGN, SFRP4, SFRP2, THBS2, FAP, FNDC1 and ADAMTS16) using GEPIA2 database. The significance of the difference in survival is listed above each graph. p values are as indicated above the graph. Genes with significance ( $p \leq 0.05$ ) are listed in RED and those lacking significance in BLACK. p values = 0 are representative of very high significance. B) Violin plot shows the expression of CTHRC1 and each of its 13 hub genes across pathological stages in 30 cancers analyzed using the GEPIA2 database. Differences across the stages of cancer for each gene of interest was calculated using the ANOVA test and significance was reported. p values are as indicated above the graph. Genes with significance ( $p \leq 0.05$ ) are listed in RED. C) Scatter plots show the Spearman correlation analysis for CTHRC1 and its 13 hub genes in 30 cancers using GEPIA2. p values are as indicated above the graph. Genes with significance ( $p \leq 0.05$  or  $p = 0$ ) are listed in RED. D) Bar graph shows log2 odds ratio from the cBioPortal for statistically significant co-occurrence between CTHRC1 and hub genes of interest (8 genes) in 30 cancer types. E) This Venn diagram shows the overlap of genes that significantly affects survival and tumour staging and are related in correlation and co-occurrence analysis in 30 cancer types. The table lists the 5 overlapping genes detected in this analysis.

<https://doi.org/10.1371/journal.pone.0270063.g005>



**Fig 6. Protein levels of CTHRC1 and network genes in breast cancer.** Graphs represent protein levels of CTHRC1 and shortlisted network genes (POSTN, MMP13, SFRP4 and FNDC1) in normal (BLUE) versus tumour tissue (RED) data from the UALCAN Portal. The box plot shows the median  $\pm$  standard deviation. p values are as indicated and calculated using the students t-test. Genes with significance ( $p < 0.05$ ) are listed in RED and those lacking significance in BLACK.

<https://doi.org/10.1371/journal.pone.0270063.g006>



**Fig 7. Functional Network analysis of CTHRC1 and its 5 hub genes.** Network analysis of CTHRC1 and its 5 hub genes (POSTN, MMP13, FNDC1, SFRP4 and ADAMTS16) identifies COL3A1 and ROR2 genes in both co-expression and physical interaction categories. **A)** Image shows co-expression network based on query genes, **B-C)** Images shows co-expression network based on **(B)** biological processes and **(C)** cellular component. **D-E)** Image shows physical interaction network based on **(D)** biological processes and **(E)** cellular component.

<https://doi.org/10.1371/journal.pone.0270063.g007>

biological process and cellular components were also evaluated (Fig 7B–7E). These networks were enriched in core matrisome genes and proteases (S1B–S1E Fig). Query and Gene Ontology-based networks when compared identified Collagen type III alpha chain 1 (COL3A1) and receptor tyrosine kinase like receptor 2 (ROR2) as two genes that were conserved among these networks and could hence be involved in regulating CTHRC1 and its hub gene-mediated regulation of the tumour matrisome (Fig 7A–7E –labelled with the red circle). In breast cancer (BRCA) CTHRC1 is overexpressed with hub genes POSTN and MMP13 (Fig 6), leading us to do evaluate their co-expression network based on query-dependent weighting and co-expression and physical interaction network based on Gene Ontology (GO) weighting for biological process and cellular components. These co-expression networks when compared confirmed COL3A1 and ROR2 to be genes conserved (Fig 8A–8C- labelled with the red circle), as seen earlier (Fig 7). Physical interaction networks however detected only COL3A1 (Fig 8D and 8E), suggesting it could be a gene of interest that CTHRC1, POSTN and MMP13 could use to regulate the tumour matrisome in breast cancers. Together they further emphasize the possible role overexpression of these matrisome genes and their crosstalk as part of a regulatory and/or functional network could have in driving the impact of the ECM through CTHRC1 in cancers.

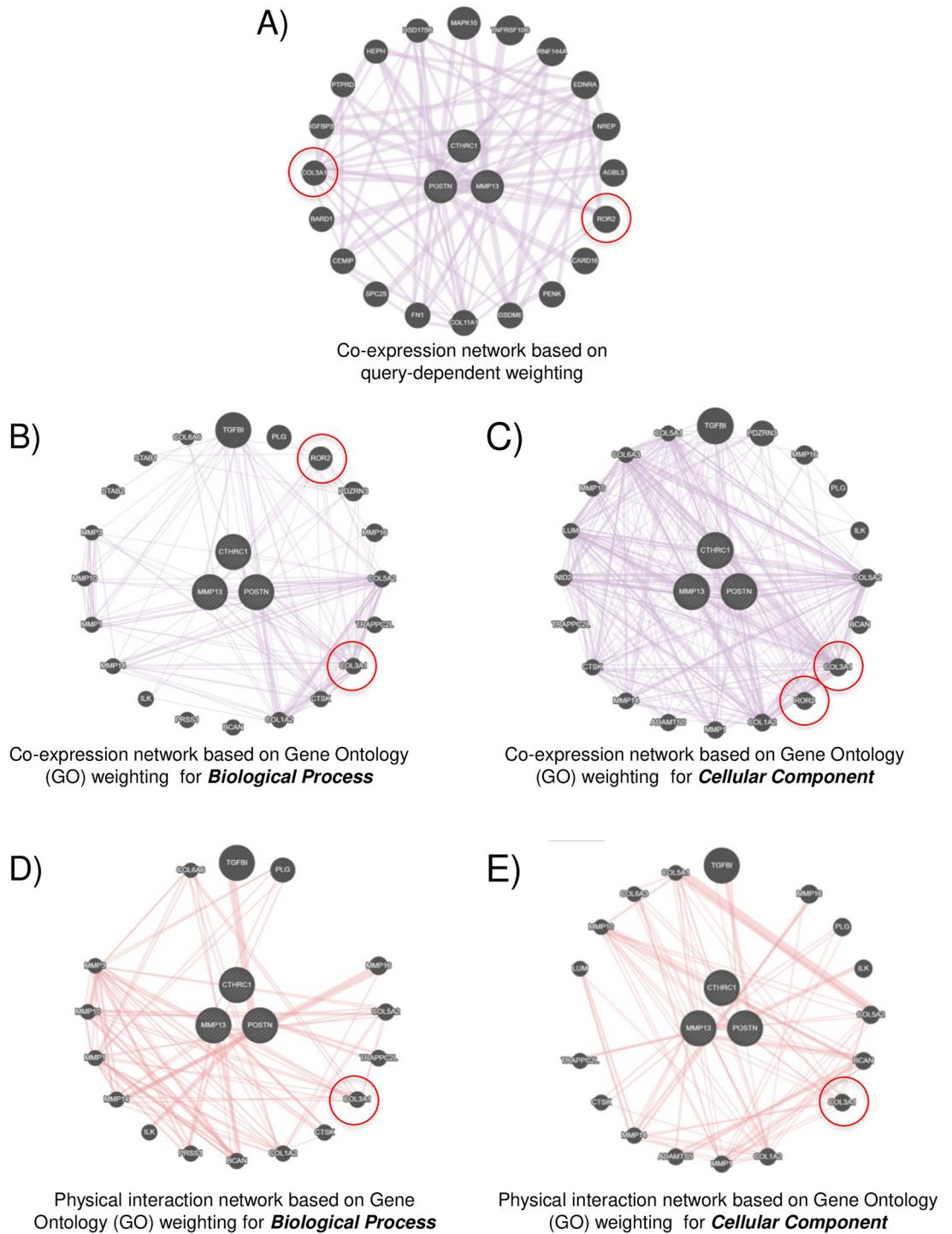
## Discussion

In the past decade, the ECM has emerged as a key player in the progression, diagnosis, and treatment of cancers. Changes in the ECM composition and structure have been known to promote migration and invasion of cancer cells [2, 22, 35, 36]. Altered ECM deposition is further associated with poor prognosis in multiple cancers [2, 3]. Recent studies have shown that CTHRC1 is overexpressed in cancers and associated with poor prognosis [37]. These studies have specifically implicated CTHRC1 with immune infiltration in Kidney and Brain cancers. Peng *et al* [38] reported that in addition to being upregulated in high-grade gliomas, CTHRC1 expression correlated with genes associated with the Wnt Signaling pathway (DVL3, DVL1, DVL2, ROR2, WNT3A, FZD6 and FZD5). Our study in evaluating 1027 matrisome genes across cancers, has identified a novel set of genes (MMP13, FNDC1, SFRP4 and ADAMTS16) that could work with CTHRC1 to regulate cancer progression.

Using TCGA data from 30 cancer types, we evaluated 1027 matrisome genes stringently analyzing their copy number, expression data with their effect on cancer survival and identified the top overexpressed ( $n = 23$ ) and downregulated ( $n = 17$ ) genes of interest across cancers. We further used a similar criterion to evaluate matrisome genes in individual cancers and identify genes affecting survival in 2 or more cancers. The intent here was to compare the pan- and individual cancer analysis to identify matrisome gene(s) that consistently show differential expression and affect cancer survival. This while eliminating possible false positives will also ensure that the genes eventually selected are strong candidates for a pan-cancer role.

This identified 3 matrisome genes, CTHRC1, PDGFA and IL7 to be prominent candidates of which only CTHRC1 was upregulated in pan-cancer and 3 individual cancers making it the matrisome gene of interest (Figs 1D and 2F, *table in red*). The expression and effect on survival of CTHRC1 were hence compared across 30 cancer types identifying 9 cancers where both were affected (BLCA, BRCA, HNSC, KIRC, LIHC, OV, READ, SARC and STAD). Differential gene expression analysis of these cancers to identify genes with a 2-fold change (increase/decrease) in 3 or more cancers led us to 19 matrisome genes that could work with CTHRC1. STRING analysis of these 19 genes further identified 13 hub genes. This network data is largely based on predictions from text mining and suggests CTHRC1 to primarily talk to POSTN which in turn communicates with other hub genes. Could the CTHRC1-POSTN connect be of significance to their role as part of the matrisome in cancers remains to be tested





**Fig 8. Functional Network analysis of CTHRC1, POSTN and MMP13.** Network analysis of CTHRC1, POSTN and MMP13 identifies COL3A1 in both co-expression and physical interaction categories. **A)** Image shows co-expression network based on query genes, **B-C)** Images shows co-expression network based on **(B)** biological processes and **(C)** cellular component. **(D-E)** Image shows physical interaction network based on **(D)** biological processes and **(E)** cellular component.

<https://doi.org/10.1371/journal.pone.0270063.g008>

experimentally. Additional hub genes MMP13, SFRP4, ADAMTS16 and FNDC1 also significantly affect survival with their expression across tumour grades comparable to CTHRC1. They also show a significant correlation and co-occurrence with CTHRC1 across cancers. Proteomics data for breast cancers in revealing CTHRC1 overexpression with POSTN and MMP13 further strengthens the need to look at this gene network in other cancers.

**CTHRC1** is a known regulator of collagen synthesis [39], shown to inhibit collagen type I and III transcripts [40], with KO mice showing reduced type I collagen levels [41]. Fibroblasts from CTHRC1 KO mice show significant downregulation of genes involved in ECM organization and collagen biosynthesis [42]. CTHRC1 is also known to regulate Wnt signalling and enhances the binding of Wnt3A with the Frizzled receptors [43]. Both POSTN and MMP13 are known to independently regulate the collagen [44, 45]. **POSTN** plays an important role in ECM structure and organization via its interaction with BMP1 to accelerate collagen cross-linking [12]. POSTN null mice exhibit aberrant collagen fibrillogenesis in the periosteum and a decrease in collagen cross-linking in the skin, tendons, and heart [44]. CTHRC1 and POSTN along with collagen are both highly expressed by the same cluster of cells in fibrotic lungs and hearts [42, 46]. **MMP13** binds collagen through its c terminal domain and cleaves collagen [45]. MMP13 KO mice implanted with mammary tumour cells show increased lung metastasis as a result of increased collagen synthesis and altered collagen structure and organization [47]. Similarly, **SFRP4** (a known Wnt antagonist) could also regulate collagen structure and organization by controlling Wnt signalling via beta-catenin [48]. Exogenous addition of SFRP4 decreases scar formation by regulating ECM deposition in infarcted hearts of mice [49].

Studies using cardiac fibroblasts have revealed the presence of crosstalk between TGF $\beta$  and Wnt signalling to regulate the fibrotic response [50]. CTHRC1 binds TGF $\beta$  and promotes its activation in colorectal cancer cells [51]. It also interacts with TGF $\beta$  receptor II and TGF $\beta$  receptor III to stabilize the ligand-receptor pathway and promote liver metastasis [51]. Hence, CTHRC1 could regulate collagen synthesis and organization by modulating both TGF $\beta$  and Wnt signalling pathways. Increased activation of TGF $\beta$  induces POSTN and CTHRC1 expression [52]. Like CTHRC1, **ADAMTS16** also physically binds TGF $\beta$  to promote its activation [53] and loss of ADAMTS16 in rats results in a reduced TGF $\beta$  activation [39]. In a mice model for heart failure, ADAMTS16 is upregulated, its expression correlating with collagen expression [53]. This TGF $\beta$  -ADAMTS16 feedback could regulate collagen synthesis, organization and degradation.

Similar to collagen, these CTHRC1 related network genes could also act to regulate fibronectin synthesis, organization and degradation. FN and collagen organization are interdependent [54, 55]. Cancer-associated fibroblasts secrete high levels of FN to form a highly organized FN rich ECM which promotes cancer cell migration and invasion [56, 57]. Such a fibronectin matrix promotes cancer cell survival during dormancy while its MMP mediated degradation promotes proliferation [58]. **CTHRC1** and FN are both overexpressed in melanomas where they are found to be localized in similar regions [59]. **POSTN** which binds FN also promotes its synthesis via the JNK pathway [60]. **MMP13** cleaves FN which also upregulates its expression using the same pathway [61]. **ADAMTS16** also cleaves FN inducing MMP3 expression to promote FN degradation [62]. Exogenous Wnt ligands can promote FN synthesis [63] that **SFRP4** dependent regulation of Wnt signalling [49, 64] could also regulate. During wound healing, CTHRC1, MMP13, ADAMTS16, SFRP4, POSTN and FNDC1 are all significantly upregulated by myofibroblasts to regulate ECM deposition [65]. POSTN binding to FN and collagen [66] could further regulate the ECM to drive tumour cell migration and invasion.

In breast cancer, CTHRC1 gene and protein expression is upregulated with MMP13 and POSTN. The UALCAN proteomic data while currently limited to 125 breast cancer samples

(of the total TCGA cohort of 1080), does support the same. Though CTHRC1 and POSTN upregulation in breast cancer is associated with poor prognosis [67], no direct interaction between them is reported. In breast cancer, tumour-associated collagen signatures are of prognostic significance. Aligned collagen bundles are negatively correlated with breast cancer survival [68]. CTHRC1 with POSTN and MMP13 could potentially affect ECM remodelling which in turn could aid in tumor cell migration and invasion [69, 70]. CTHRC1 secreted by cells could also independently bind collagen and regulate its assembly and organization. This could also modulate how secreted growth factors (i.e., TGF $\beta$  and Wnt) are sequestered by the matrix in breast cancer cells [71]. CTHRC1 expression is associated with metastasis to the bone [72]. In breast cancer patients with high periostin, the risk of bone metastases is enhanced by elevated CTHRC1 expression [68]. This when considered with the known role for MMP13 in bone metastasis of breast cancer [73, 74] strongly supports a role for CTHRC1-POSTN-MMP13 crosstalk in mediating the same. Understanding how this crosstalk facilitates bone metastasis remains a vital open question.

Further experimental validation of the predictive bioinformatics data will be vital to establish the regulatory and/or functional crosstalk between CTHRC1 and POSTN, MMP13, SFRP4, FNDC1 and ADAMTS16. CTHRC1 is susceptible to proteolysis and cleaved CTHRC1 has been reported to be a better inhibitor of collagen synthesis [40]. If MMP13 or ADAMTS16 could regulate CTHRC1 cleavage remains to be tested. CTHRC1, ADAMTS16 and POSTN are all reported to bind TGF $\beta$  [12, 39] which could as a point of convergence for their regulatory crosstalk. CTHRC1 and POSTN expression [53] are directly induced upon TGF $\beta$  activation which could be one of the key players in influencing the CTHRC1 related matrisome network. Whether FNDC1, SFRP4, ADAMTS16 and MMP13 are regulated by CTHRC1 independent of TGF $\beta$  remains to be evaluated.

Could the CTHRC1 network also regulate other growth factors (Wnt, EGF, FGF etc) remains to be verified. A joint role for these proteins in ECM remodelling (through collagen and fibronectin) could regulate tumour cell migration and invasion [69, 70]. The cross-linking of collagen fibers controls their density and packing order which could regulate ECM stiffness to further drive cancer progression [75–80] downstream of CTHRC1.

Our functional network analysis of CTHRC1 and its hub genes using GeneMANIA has identified COL3A1 and ROR2 to be a part of the co-expression and physical interaction networks. CTHRC1 has been shown to regulate type III collagen synthesis and binds ROR2 to regulate planar cell polarity during development [81]. Changes in ECM composition and remodelling drive tumour progression [82]. Collagen fibers are seen to be less dense, shorter, straighter, thinner, and more aligned with one another in breast cancer [83] which can predict their pathology and outcomes [84]. Both COL3A1 and ROR2 have been implicated in breast cancer pathogenesis [85–88]. ROR2 has been shown to act as an oncogene to promote breast cancer progression. COL3A1 is highly expressed by the tumour stroma and associated with increased survival in breast cancer patients [89]. Secreted COL3A1 causes a wavy collagen fiber orientation promoting tumour dormancy in breast cancer [90] possibly through a DDR pathway (Discoidin domain receptor—tyrosine kinase proteins activated by collagen) to limit metastasis [88, 90]. COL3A1 expression could have implications as a vital biomarker in breast cancer [88]. In Esophageal Cancer COL3A1 is overexpressed with POSTN [91] which further emphasizes the joint role they could have with CTHRC1. Thus, in identifying CTHRC1 and the network of genes it works with across cancers, this study not only helps reveal the possible role POSTN, MMP13, SFRP4, FNDC1 and ADAMTS16 could have in regulating the impact of the matrisome in cancers but also highlights the role such a network could have in sustaining the same.

## Supporting information

**S1 Fig. Protein categories of the functional network analysis of CTHRC1 and its 5 hub genes.** Nested bar graphs represent the percentage of proteins that belong to each category in (A) the network analysis based on query genes, (B) co-expression network based on biological processes or (C) cellular component and physical interaction network based on (D) biological processes or (E) cellular component. Each colour represents a distinct protein subcategory. (PDF)

**S1 File. S1A is the summary file of the top 5% genes for 9 cancers. S1B - S1J** are the raw data output files of the differential gene expression analysis of the 9 cancers (Fig 4A) where CTHRC1 is upregulated and affects survival. (ZIP)

## Author Contributions

**Conceptualization:** Keerthi Harikrishnan, Nagaraj Balasubramanian.

**Data curation:** Keerthi Harikrishnan, Srinivas Sheshagiri Prabhu.

**Formal analysis:** Keerthi Harikrishnan, Srinivas Sheshagiri Prabhu.

**Funding acquisition:** Keerthi Harikrishnan, Nagaraj Balasubramanian.

**Investigation:** Keerthi Harikrishnan, Srinivas Sheshagiri Prabhu.

**Methodology:** Keerthi Harikrishnan, Srinivas Sheshagiri Prabhu.

**Project administration:** Keerthi Harikrishnan, Nagaraj Balasubramanian.

**Resources:** Keerthi Harikrishnan, Nagaraj Balasubramanian.

**Software:** Keerthi Harikrishnan, Srinivas Sheshagiri Prabhu.

**Supervision:** Keerthi Harikrishnan, Nagaraj Balasubramanian.

**Validation:** Keerthi Harikrishnan, Nagaraj Balasubramanian.

**Visualization:** Keerthi Harikrishnan, Srinivas Sheshagiri Prabhu.

**Writing – original draft:** Keerthi Harikrishnan, Nagaraj Balasubramanian.

**Writing – review & editing:** Keerthi Harikrishnan, Nagaraj Balasubramanian.

## References

1. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.* 2014; 15(12):1243–53. <https://doi.org/10.15252/embr.201439246> PMID: 25381661
2. Socovich AM, Naba A. The cancer matrisome: From comprehensive characterization to biomarker discovery. *Seminars in Cell and Developmental Biology.* 2019; 89:157–66. <https://doi.org/10.1016/j.semcdb.2018.06.005> PMID: 29964200
3. Taha IN, Naba A. Exploring the extracellular matrix in health and disease using proteomics. *Essays in Biochemistry.* 2019; 63(3):417–32. <https://doi.org/10.1042/EBC20190001> PMID: 31462529
4. Naba A, Clauser KR, Lamar JM, Carr SA, Hynes RO. Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. *Elife.* 2014; 2014(3):1–23. <https://doi.org/10.7554/eLife.01308> PMID: 24618895
5. Hebert JD, Myers SA, Naba A, Abbruzzese G, Lamar JM, Carr SA, et al. Proteomic profiling of the ECM of xenograft breast cancer metastases in different organs reveals distinct metastatic niches. *Cancer Research.* 2020; 80(7):1475–85. <https://doi.org/10.1158/0008-5472.CAN-19-2961> PMID: 32019869

6. Shao X, Taha IN, Clauser KR, Gao Y (Tom), Naba A. MatrisomeDB: The ECM-protein knowledge database. *Nucleic Acids Research*. 2020; 48(D1):D1136–44. <https://doi.org/10.1093/nar/gkz849> PMID: 31586405
7. Harikrishnan K, Joshi O, Madangirikar S, Balasubramanian N. Cell Derived Matrix Fibulin-1 Associates With Epidermal Growth Factor Receptor to Inhibit Its Activation, Localization and Function in Lung Cancer Calu-1 Cells. *Frontiers in Cell and Developmental Biology*. 2020; 8(July). <https://doi.org/10.3389/fcell.2020.00522> PMID: 32719793
8. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: In silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular and Cellular Proteomics*. 2012; 11(4):M111.014647. <https://doi.org/10.1074/mcp.M111.014647> PMID: 22159717
9. Teuscher AC, Jongsma E, Davis MN, Statzer C, Gebauer JM, Naba A, et al. The in-silico characterization of the *Caenorhabditis elegans* matrisome and proposal of a novel collagen classification. *Matrix Biology Plus*. 2019; 1:100001. <https://doi.org/10.1016/j.mbplus.2018.11.001> PMID: 33543001
10. Socovich AM, Naba A. The cancer matrisome: From comprehensive characterization to biomarker discovery. *Seminars in Cell and Developmental Biology*. 2019; 89:157–66. <https://doi.org/10.1016/j.semcd.2018.06.005> PMID: 29964200
11. Lin S, Zheng L, Lu Y, Xia Q, Zhou P, Liu Z. Comprehensive analysis on the expression levels and prognostic values of LOX family genes in kidney renal clear cell carcinoma. *Cancer Medicine*. 2020; 9(22):8624–38. <https://doi.org/10.1002/cam4.3472> PMID: 32970930
12. González-González L, Alonso J. Periostin: A matricellular protein with multiple functions in cancer development and progression. *Frontiers in Oncology*. 2018; 8(JUN):1–15. <https://doi.org/10.3389/fonc.2018.00225> PMID: 29946533
13. Liot S, Aubert A, Hervieu V, El Kholti N, Schalkwijk J, Verrier B, et al. Loss of Tenascin-X expression during tumor progression: A new pan-cancer marker. *Matrix Biology Plus*. 2020; 6–7(1):6–7. <https://doi.org/10.1016/j.mbplus.2020.100021> PMID: 33543019
14. Lin TC, Yang CH, Cheng LH, Chang WT, Lin YR, Cheng HC. Fibronectin in Cancer: Friend or Foe. *Cells*. 2019; 9(1):1–37. <https://doi.org/10.3390/cells9010027> PMID: 31861892
15. Yuzhalin AE, Urbonas T, Silva MA, Muschel RJ, Gordon-Weeks AN. A core matrisome gene signature predicts cancer outcome. *British Journal of Cancer*. 2018; 118(3):435–40. <https://doi.org/10.1038/bjc.2017.458> PMID: 29360819
16. Lim S Bin Tan SJ, Lim WT, Lim CT. An extracellular matrix-related prognostic and predictive indicator for early-stage non-small cell lung cancer. *Nature Communications*. 2017; 8(1):1–10.
17. Chakravarthy A, Khan L, Bensler NP, Bose P, De Carvalho DD. TGF- $\beta$ -associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. *Nature Communications*. 2018; 9(1):1–10.
18. Lim S Bin, Chua MLK, Yeong JPS, Tan SJ, Lim WT, Lim CT. Pan-cancer analysis connects tumor matrisome to immune response. *npj Precision Oncology*. 2019; 3(1).
19. Korkut A, Zaidi S, Kanchi RS, Rao S, Gough NR, Schultz A, et al. A Pan-Cancer Analysis Reveals High-Frequency Genetic Alterations in Mediators of Signaling by the TGF- $\beta$  Superfamily. *Cell Systems*. 2018; 7(4):422–437.e7. <https://doi.org/10.1016/j.cels.2018.08.010> PMID: 30268436
20. Izzi V, Lakkala J, Devarajan R, Kääriäinen A, Koivunen J, Heljasvaara R, et al. Pan-Cancer analysis of the expression and regulation of matrisome genes across 32 tumor types. *Matrix Biology Plus*. 2019; 1:100004. <https://doi.org/10.1016/j.mbplus.2019.04.001> PMID: 33543003
21. Kääriäinen A, Pesola V, Dittmann A, Kontio J, Koivunen J, Pihlajaniemi T, et al. Machine learning identifies robust matrisome markers and regulatory mechanisms in cancer. *International Journal of Molecular Sciences*. 2020; 21(22):1–12. <https://doi.org/10.3390/ijms21228837> PMID: 33266472
22. Izzi V, Davis MN, Naba A. Pan-cancer analysis of the genomic alterations and mutations of the matrisome. *Cancers (Basel)*. 2020; 12(8):1–21. <https://doi.org/10.3390/cancers12082046> PMID: 32722287
23. Shao X, Taha IN, Clauser KR, Gao Y (Tom), Naba A. MatrisomeDB: the ECM-protein knowledge database. *Nucleic Acids Research*. 2020 Jan 8; 48(D1):D1136–44. <https://doi.org/10.1093/nar/gkz849> PMID: 31586405
24. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*. 2020 Jun 22; 38(6):675–8. <https://doi.org/10.1038/s41587-020-0546-8> PMID: 32444850
25. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Research*. 2019 Jul 2; 47(W1):W556–60. <https://doi.org/10.1093/nar/gkz430> PMID: 31114875
26. TCGA survival analysis. TCGAsurvival/survival.Rmd at master · mdozmorov/TCGAsurvival · GitHub
27. TCGA\_preprocessing. TCGA/preprocessing.R at master · ericaspada/TCGA · GitHub

28. TCGA\_DEG. TCGAsurvival/TCGA\_DEGs.Rmd at master · dozmorov/TCGAsurvival · GitHub
29. Cai Q, Dozmorov M, Oh Y. IGFBP-3/IGFBP-3 Receptor System as an Anti-Tumor and Anti-Metastatic Signaling in Cancer. *Cells*. 2020 May 20; 9(5):1261. <https://doi.org/10.3390/cells9051261> PMID: 32443727
30. Calculate and draw custom Venn diagrams. Draw Venn Diagram (ugent.be)
31. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*. 2009 Jan 1; 37(Database):D412–6. <https://doi.org/10.1093/nar/gkn760> PMID: 18940858
32. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*. 2010 Jul 1; 38(suppl\_2):W214–20. <https://doi.org/10.1093/nar/gkq537> PMID: 20576703
33. Chandrashekar DS, Karthikeyan SK, Korla PK, Patel H, Shovon AR, Athar M, et al. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia*. 2022 Mar; 25:18–27. <https://doi.org/10.1016/j.neo.2022.01.001> PMID: 35078134
34. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*. 2010 Jul 1; 38(suppl\_2):W214–20. <https://doi.org/10.1093/nar/gkq537> PMID: 20576703
35. Brechbuhl HM, Barrett AS, Kopin E, Hagen JC, Han AL, Gillen AE, et al. Fibroblast subtypes define a metastatic matrixome in breast cancer. *JCI Insight*. 2020; 5(4):1–16. <https://doi.org/10.1172/jci.insight.130751> PMID: 32045383
36. Costea DE, Hills A, Osman AH, Thurlow J, Kalna G, Huang X, et al. Identification of two distinct carcinoma-associated fibroblast subtypes with differential tumor-promoting abilities in oral squamous cell carcinoma. *Cancer Research*. 2013; 73(13):3888–901. <https://doi.org/10.1158/0008-5472.CAN-12-4150> PMID: 23598279
37. Zhou F, Shen D, Xiong Y, Cheng S, Xu H, Wang G, et al. CTHRC1 Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Kidney Renal Papillary Cell Carcinoma and Kidney Renal Clear Cell Carcinoma. *Front Oncol [Internet]*. 2021 Feb 8; 10:570819. Available from: <https://pubmed.ncbi.nlm.nih.gov/33628726> <https://doi.org/10.3389/fonc.2020.570819> PMID: 33628726
38. Peng D, Wei C, Zhang X, Li S, Liang H, Zheng X, et al. Pan-cancer analysis combined with experiments predicts CTHRC1 as a therapeutic target for human cancers. *Cancer Cell International [Internet]*. 2021; 21(1):566. Available from: <https://doi.org/10.1186/s12935-021-02266-3> PMID: 34702252
39. Gopalakrishnan K, Kumarasamy S, Abdul-Majeed S, Kalinoski AL, Morgan EE, Gohara AF, et al. Targeted disruption of Adamts16 gene in a rat genetic model of hypertension. *Proc Natl Acad Sci U S A*. 2012; 109(50):20555–9. <https://doi.org/10.1073/pnas.1211290109> PMID: 23185005
40. Pygay P, Heroult M, Wang Q, Lehnert W, Belden J, Liaw L, et al. Collagen triple helix repeat containing 1, a novel secreted protein in injured and diseased arteries, inhibits collagen expression and promotes cell migration. *Circulation Research*. 2005; 96(2):261–8. <https://doi.org/10.1161/01.RES.0000154262.07264.12> PMID: 15618538
41. Kimura H, Kwan KM, Zhang Z, Deng JM, Darnay BG, Behringer RR, et al. Cthrc1 is a positive regulator of osteoblastic bone formation. *PLoS ONE*. 2008; 3(9).
42. Ruiz-villalba A, Romero JP, Hernandez SC. Single-Cell RNA-seq Analysis Reveals a Crucial Role for Collagen Triple Helix Repeat Containing 1 (CTHRC1) Cardiac Fibroblasts after Myocardial Infarction. *Circulation*. 2020 Nov 10; 142(19):1831–1847.
43. Kelley MW. Leading Wnt down a PCP Path: Cthrc1 Acts as a Coreceptor in the Wnt-PCP Pathway. *Developmental Cell*. 2008; 15(1):7–8. <https://doi.org/10.1016/j.devcel.2008.06.008> PMID: 18606135
44. Norris R a, Damon B, Mironov V, Kasyanov V, Moreno-rodriguez R, Trusk T, et al. Periostin Regulates Collagen Fibrillogenesis. 2007; 101(3):695–711.
45. Knäuper V, Cowell S, Smith B, López-Otin C, O'Shea M, Morris H, et al. The role of the C-terminal domain of human collagenase-3 (MMP-13) in the activation of procollagenase-3, substrate specificity, and tissue inhibitor of metalloproteinase interaction. *Journal of Biological Chemistry*. 1997; 272(12):7608–16. <https://doi.org/10.1074/jbc.272.12.7608> PMID: 9065415
46. Tsukui T, Sun KH, Wetter JB, Wilson-Kanamori JR, Hazelwood LA, Henderson NC, et al. Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nature Communications*. 2020; 11(1):1–16.
47. Perry SW, Schueckler JM, Burke K, Arcuri GL, Brown EB. Stromal matrix metalloproteinase-13 knockout alters Collagen I structure at the tumor-host interface and increases lung metastasis of C57BL/6 syngeneic E0771 mammary tumor cells. *BMC Cancer*. 2013; 13. <https://doi.org/10.1186/1471-2407-13-411> PMID: 24010522

48. Mullin NK, Mallipeddi N V., Hamburg-Shields E, Ibarra B, Khalil AM, Atit RP. Wnt/ $\beta$ -catenin signaling pathway regulates specific lncRNAs that impact dermal fibroblasts and skin fibrosis. *Frontiers in Genetics*. 2017; 8(NOV):1–13. <https://doi.org/10.3389/fgene.2017.00183> PMID: 29209359
49. Matsushima K, Suyama T, Takenaka C, Nishishita N, Ikeda K, Ikada Y, et al. Secreted frizzled related protein 4 reduces fibrosis scar size and ameliorates cardiac function after ischemic injury. *Tissue Engineering—Part A*. 2010; 16(11):3329–41. <https://doi.org/10.1089/ten.TEA.2009.0739> PMID: 20528676
50. Działo E, Tkacz K, Błyszczuk P. Crosstalk between the TGF- $\beta$  and WNT signalling pathways during cardiac fibrogenesis. *Acta Biochimica Polonica*. 2018; 65(3):341–9. [https://doi.org/10.18388/abp.2018\\_2635](https://doi.org/10.18388/abp.2018_2635) PMID: 30040870
51. Zhang XL, Hu LP, Yang Q, Qin WT, Wang X, Xu CJ, et al. CTHRC1 promotes liver metastasis by reshaping infiltrated macrophages through physical interactions with TGF- $\beta$  receptors in colorectal cancer. *Oncogene*. 2021; 40(23):3959–73. <https://doi.org/10.1038/s41388-021-01827-0> PMID: 33986509
52. Bagordakis E, Sawazaki-Calone I, Macedo CCS, Carnielli CM, de Oliveira CE, Rodrigues PC, et al. Secretome profiling of oral squamous cell carcinoma-associated fibroblasts reveals organization and disassembly of extracellular matrix and collagen metabolic process signatures. *Tumor Biology*. 2016; 37(7):9045–57. <https://doi.org/10.1007/s13277-015-4629-y> PMID: 26762409
53. Yao Y, Hu C, Song Q, Li Y, Da X, Yu Y, et al. ADAMTS16 activates latent TGF- $\beta$ , accentuating fibrosis and dysfunction of the pressure-overloaded heart. *Cardiovascular Research*. 2020; 116(5):956–69. <https://doi.org/10.1093/cvr/cvz187> PMID: 31297506
54. Dzamba BJ, Wu H, Jaenisch R, Peters DM. Fibronectin binding site in type I collagen regulates fibronectin fibril formation. *Journal of Cell Biology*. 1993; 121(5):1165–72. <https://doi.org/10.1083/jcb.121.5.1165> PMID: 8501121
55. Sottile J, Hocking DC. Fibronectin Polymerization Regulates the Composition and Stability of Extracellular Matrix Fibrils and Cell-Matrix Adhesions. *Molecular Biology of the Cell*. 2002 Aug 6; 13(10):3546–59. <https://doi.org/10.1091/mbc.e02-01-0048> PMID: 12388756
56. Spada S, Tocci A, Di Modugno F, Nisticò P. Fibronectin as a multiregulatory molecule crucial in tumor matrisome: from structural and functional features to clinical practice in oncology. *Journal of Experimental and Clinical Cancer Research*. 2021; 40(1):1–14.
57. Erdogan B, Ao M, White LM, Means AL, Brewer BM, Yang L, et al. Cancer-associated fibroblasts promote directional cancer cell migration by aligning fibronectin. *Journal of Cell Biology*. 2017; 216(11):3799–816. <https://doi.org/10.1083/jcb.201704053> PMID: 29021221
58. Barney LE, Hall CL, Schwartz AD, Parks AN, Sparages C, Galarza S, et al. Tumor cell-organized fibronectin maintenance of a dormant breast cancer population. *Science Advances*. 2020; 6(11):1–13. <https://doi.org/10.1126/sciadv.aaz4157> PMID: 32195352
59. Eriksson J, Le Joncour V, Nummela P, Jahkola T, Virolainen S, Laakkonen P, et al. Gene expression analyses of primary melanomas reveal CTHRC1 as an important player in melanoma progression. *Oncotarget*. 2016; 7(12):15065–92. <https://doi.org/10.18632/oncotarget.7604> PMID: 26918341
60. Kim SS, Nikoloudaki GE, Michelsons S, Creber K, Hamilton DW. Fibronectin synthesis, but not  $\alpha$ -smooth muscle expression, is regulated by periostin in gingival healing through FAK/JNK signaling. *Scientific Reports*. 2019; 9(1):1–17.
61. Forsyth CB, Pulai J, Loeser RF. Fibronectin fragments and blocking antibodies to  $\alpha 2\beta 1$  and  $\alpha 5\beta 1$  integrins stimulate mitogen-activated protein kinase signaling and increase collagenase 3 (matrix metalloproteinase 13) production by human articular chondrocytes. *Arthritis and Rheumatism*. 2002; 46(9):2368–76.
62. Schnellmann R, Sack R, Hess D, Annis DS, Mosher DF, Apte SS, et al. A selective extracellular matrix proteomics approach identifies fibronectin proteolysis by a disintegrin-like and metalloprotease domain with thrombospondin type 1 motifs (ADAMTS16) and its impact on spheroid morphogenesis. *Molecular and Cellular Proteomics*. 2018; 17(7):1410–25. <https://doi.org/10.1074/mcp.RA118.000676> PMID: 29669734
63. Xiao L, Zhou D, Tan RJ, Fu H, Zhou L, Hou FF, et al. Sustained activation of Wnt/ $\beta$ -catenin signaling drives AKI to CKD progression. *Journal of the American Society of Nephrology*. 2016; 27(6):1727–40.
64. Pohl S, Scott R, Arfuso F, Perumal V, Dharmarajan A. Secreted frizzled-related protein 4 and its implications in cancer and apoptosis. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2015 Jan; 36(1):143–52. <https://doi.org/10.1007/s13277-014-2956-z> PMID: 25501511
65. Bergmeier V, Etich J, Pitzler L, Frie C, Koch M, Fischer M, et al. Identification of a myofibroblast-specific expression signature in skin wounds. *Matrix Biology*. 2018; 65:59–74. <https://doi.org/10.1016/j.matbio.2017.07.005> PMID: 28797711
66. Soikkeli J, Podlasz P, Yin M, Nummela P, Jahkola T, Virolainen S, et al. Metastatic outgrowth encompasses COL-I, FN1, and POSTN up-regulation and assembly to fibrillar networks regulating cell

- adhesion, migration, and growth. *American Journal of Pathology*. 2010; 177(1):387–403. <https://doi.org/10.2353/ajpath.2010.090748> PMID: 20489157
67. Kharashvili G, Cizkova M, Bouchalova K, Mgebrishvili G, Kolar Z, Bouchal J. Collagen triple helix repeat containing 1 protein, periostin and versican in primary and metastatic breast cancer: An immunohistochemical study. *Journal of Clinical Pathology*. 2011; 64(11):977–82. Available from: <https://jcp.bmj.com/content/64/11/977>. <https://doi.org/10.1136/jclinpath-2011-200106> PMID: 21742751
  68. Ouellette JN, Drifka CR, Pointer KB, Liu Y, Lieberthal TJ, Kao WJ, et al. Navigating the collagen jungle: The biomedical potential of fiber organization in cancer. *Bioengineering*. 2021; 8(2):1–19. <https://doi.org/10.3390/bioengineering8020017> PMID: 33494220
  69. Leclère L, Nir TS, Bazarsky M, Braitbard M, Schneidman-Duhovny D, Gat U, et al. Dynamic Evolution of the Cthrc1 Genes, a Newly Defined Collagen-Like Family. *Genome Biology and Evolution*. 2020; 12(2):3957–70. <https://doi.org/10.1093/gbe/evaa020> PMID: 32022859
  70. Vallet SD, Ricard-Blum S. Lysyl oxidases: From enzyme activity to extracellular matrix cross-links. *Essays in Biochemistry*. 2019; 63(3):349–64. <https://doi.org/10.1042/EBC20180050> PMID: 31488698
  71. Insua-Rodríguez J, Oskarsson T. The extracellular matrix in breast cancer. *Advanced Drug Delivery Reviews [Internet]*. 2016; 97:41–55. Available from: <https://www.sciencedirect.com/science/article/pii/S0169409X15300235> <https://doi.org/10.1016/j.addr.2015.12.017> PMID: 26743193
  72. Myngbay A, Manarbek L, Ludbrook S, Kunz J. The role of collagen triple helix repeat-containing 1 protein (Cthrc1) in rheumatoid arthritis. Vol. 22, *International Journal of Molecular Sciences*. MDPI AG; 2021. p. 1–16. <https://doi.org/10.3390/ijms22052426> PMID: 33670905
  73. IBARAGI S, SHIMO T, HASSAN NURMM, ISOWA S, KURIO N, MANDAI H, et al. Induction of MMP-13 Expression in Bone-metastasizing Cancer Cells by Type I Collagen through Integrin  $\alpha 1\beta 1$  and  $\alpha 2\beta 1$ -p38 MAPK Signaling. *Anticancer Research [Internet]*. 2011 Apr 1; 31(4):1307. Available from: <http://ar.iijournals.org/content/31/4/1307.abstract> PMID: 21508380
  74. Shah M, Huang D, Blick T, Connor A, Reiter LA, Hardink JR, et al. An MMP13-Selective Inhibitor Delays Primary Tumor Growth and the Onset of Tumor-Associated Osteolytic Lesions in Experimental Models of Breast Cancer. *PLOS ONE [Internet]*. 2012 Jan 11; 7(1):e29615-. Available from: <https://doi.org/10.1371/journal.pone.0029615> PMID: 22253746
  75. Candiello J, Balasubramani M, Schreiber EM, Cole GJ, Mayer U, Halfter W, et al. Biomechanical properties of native basement membranes. *The FEBS Journal*. 2007; 274(11):2897–908. <https://doi.org/10.1111/j.1742-4658.2007.05823.x> PMID: 17488283
  76. Shen ZL, Dodge MR, Kahn H, Ballarín R, Eppell SJ. Stress-Strain Experiments on Individual Collagen Fibrils. *Biophysical Journal*. 2008 Oct 15; 95(8):3956–63. <https://doi.org/10.1529/biophysj.107.124602> PMID: 18641067
  77. Yang L, van der Werf KO, Fitié CFC, Bennink ML, Dijkstra PJ, Feijen J. Mechanical Properties of Native and Cross-linked Type I Collagen Fibrils. *Biophysical Journal*. 2008 Mar 15; 94(6):2204–11. <https://doi.org/10.1529/biophysj.107.111013> PMID: 18032556
  78. Svensson RB, Hassenkam T, Grant CA, Magnusson SP. Tensile Properties of Human Collagen Fibrils and Fascicles Are Insensitive to Environmental Salts. *Biophysical Journal*. 2010 Dec 15; 99(12):4020–7. <https://doi.org/10.1016/j.bpj.2010.11.018> PMID: 21156145
  79. Aifantis KE, Shrivastava S, Odegard GM. Transverse mechanical properties of collagen fibers from nanoindentation. *Journal of Materials Science: Materials in Medicine*. 2011; 22(6):1375–81. <https://doi.org/10.1007/s10856-011-4320-9> PMID: 21556981
  80. Kohn JC, Lampi MC, Reinhart-King CA. Age-related vascular stiffening: causes and consequences. *Frontiers in Genetics*. 2015; 6:112. <https://doi.org/10.3389/fgene.2015.00112> PMID: 25926844
  81. Yamamoto S, Nishimura O, Misaki K, Nishita M, Minami Y, Yonemura S, et al. Cthrc1 selectively activates the planar cell polarity pathway of Wnt signaling by stabilizing the Wnt-receptor complex. *Dev Cell*. 2008 Jul; 15(1):23–36. <https://doi.org/10.1016/j.devcel.2008.05.007> PMID: 18606138
  82. Martino J Di, Nobre AR, Mondal C, Taha I, Farias E, Fertig E, et al. No Title. *Nature Portfolio*. 2021;
  83. Bodelon C, Mullooly M, Pfeiffer RM, Fan S, Abubakar M, Lenz P, et al. Mammary collagen architecture and its association with mammographic density and lesion severity among women undergoing image-guided breast biopsy. *Breast Cancer Res*. 2021 Nov; 23(1):105. <https://doi.org/10.1186/s13058-021-01482-z> PMID: 34753492
  84. Sprague BL, Vacek PM, Mulrow SE, Evans MF, Trentham-Dietz A, Herschorn SD, et al. Collagen Organization in Relation to Ductal Carcinoma In Situ Pathology and Outcomes. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2021 Jan; 30(1):80–8. <https://doi.org/10.1158/1055-9965.EPI-20-0889> PMID: 33082201



85. Menck K, Heinrichs S, Wlochowicz D, Sitte M, Noeding H, Janshoff A, et al. WNT11/ROR2 signaling is associated with tumor invasion and poor survival in breast cancer. *Journal of Experimental & Clinical Cancer Research*. 2021; 40(1):395. <https://doi.org/10.1186/s13046-021-02187-z> PMID: 34911552
86. Guo M, Ma G, Zhang X, Tang W, Shi J, Wang Q, et al. ROR2 knockdown suppresses breast cancer growth through PI3K/ATK signaling. *Aging*. 2020 Jul; 12(13):13115–27. <https://doi.org/10.18632/aging.103400> PMID: 32614787
87. Lucchetta M, da Piedade I, Mounir M, Vabistsevits M, Terkelsen T, Papaleo E, et al. Cystatin A suppresses tumor cell growth through inhibiting epithelial to mesenchymal transition in human lung cancer. *International Journal of Molecular Sciences*. 2021 Jan 1; 71(1):14084–98.
88. Di Martino JS, Nobre AR, Mondal C, Taha I, Farias EF, Fertig EJ, et al. A tumor-derived type III collagen-rich ECM niche regulates tumor cell dormancy. *Nature Cancer*. 2021; <https://doi.org/10.1038/s43018-021-00291-9> PMID: 35121989
89. Brisson BK, Mauldin EA, Lei W, Vogel LK, Power AM, Lo A, et al. Type III Collagen Directs Stromal Organization and Limits Metastasis in a Murine Model of Breast Cancer. *Am J Pathol*. 2015 May; 185(5):1471–86. <https://doi.org/10.1016/j.ajpath.2015.01.029> PMID: 25795282
90. Sun X, Wu B, Chiang HC, Deng H, Zhang X, Xiong W, et al. Tumour DDR1 promotes collagen fibre alignment to instigate immune exclusion. *Nature*. 2021 Nov; 599(7886):673–8. <https://doi.org/10.1038/s41586-021-04057-2> PMID: 34732895
91. Zhang SW, Zhang N, Wang N. Role of COL3A1 and POSTN on Pathologic Stages of Esophageal Cancer. *Technol Cancer Res Treat*. 2020; 19:1533033820977489–1533033820977489. <https://doi.org/10.1177/1533033820977489> PMID: 33280513