BMC Cancer

Somatic mutations in collagens are associated with a distinct tumor environment and overall survival in gastric cancer

Alexander S. Brodsky[1,2,3*] , Jay Khurana[1], Kevin S. Guo[1], Elizabeth Y. Wu[1], Dongfang Yang[1], Ayesha S. Siddique[1], Ian Y. Wong[1,3,4], Ece D. Gamsiz Uzun[1,2] and Murray B. Resnick[1,5]

Abstract

Background: Gastric cancer is a heterogeneous disease with poorly understood genetic and microenvironmental factors. Mutations in collagen genes are associated with genetic diseases that compromise tissue integrity, but their role in tumor progression has not been extensively reported. Aberrant collagen expression has been long associated with malignant tumor growth, invasion, chemoresistance, and patient outcomes. We hypothesized that somatic mutations in collagens could functionally alter the tumor extracellular matrix.

Methods: We used publicly available datasets including The Tumor Cancer Genome Atlas (TCGA) to interrogate somatic mutations in collagens in stomach adenocarcinomas. To demonstrate that collagens were significantly mutated above background mutation rates, we used a moderated Kolmogorov-Smirnov test along with combination analysis with a bootstrap approach to define the background accounting for mutation rates. Association between mutations and clinicopathological features was evaluated by Fisher or chi-squared tests. Association with overall survival was assessed by Kaplan-Meier and the Cox-Proportional Hazards Model. Gene Set Enrichment Analysis was used to interrogate pathways. Immunohistochemistry and in situ hybridization tested expression of COL7A1 in stomach tumors.

Results: In stomach adenocarcinomas, we identified individual collagen genes and sets of collagen genes harboring somatic mutations at a high frequency compared to background in both microsatellite stable, and microsatellite instable tumors in TCGA. Many of the missense mutations resemble the same types of loss of function mutations in collagenopathies that disrupt tissue formation and destabilize cells providing guidance to interpret the somatic mutations. We identified combinations of somatic mutations in collagens associated with overall survival, with a distinctive tumor microenvironment marked by lower matrisome expression and immune cell signatures. Truncation mutations were strongly associated with improved outcomes suggesting that loss of expression of secreted collagens impact tumor progression and treatment response. Germline collagenopathy variants guided interpretation of impactful somatic mutations on tumors.

Conclusions: These observations highlight that many collagens, expressed in non-physiologically relevant conditions in tumors, harbor impactful somatic mutations in tumors, suggesting new approaches for classification and

*Correspondence: alexander_brodsky@brown.edu
[3] Joint Program in Cancer Biology, Brown University and Lifespan Cancer Institute, Providence, RI 02912, USA
Full list of author information is available at the end of the article

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 2 of 20

therapy development in stomach cancer. In sum, these findings demonstrate how classification of tumors by collagen mutations identified strong links between specific genotypes and the tumor environment.

Keywords: Collagen, Stomach cancer, Somatic mutations, Extracellular matrix

Background

Collagens are the most abundant proteins in extracellular matrix and are critical components and regulators of the tumor microenvironment [1, 2]. Increased collagen expression in many solid tumors has been associated with poor outcomes and resistance in multiple settings [3], likely through increased epithelial-to-mesenchymal transitions (EMT) and drug resistance [4]. The 28 members of the collagen family are expressed by 43 genes and defined by the common triple helix motif. Collagens are classified into families including fibrillar collagens (i.e. Collagen type I, II, III, V, XI, XIV), network collagens (i.e. Collagen type IV), membrane (i.e. type XVII) and other (type VII, XXVIII) [5] (Table S1). Although most studies have focused on the most abundant collagen, collagen type I, in cancer there is increasing awareness of the role of many minor collagens in cancer such as types X and XI [6, 7]. Minor collagens are defined by their low abundance compared to the major fibrillar collagen types such as type I, but nonetheless they have critical functions and large impacts on tissues. Collagen structures are very complex because of the tendency to form heterotrimers, interact with each other, post-translational modifications and regulation through crosslinking [5]. The breadth of mechanisms by which collagens mediate tumor progression is not yet understood and collagens could have context dependent functions in tumors analogous to their normal tissue specific expression and functions [4]. The cellular origin of collagens is not always clear as both cancer and stroma cells are known to secrete collagens as indicated both by in situ hybridization studies [8, 9] and recent proteomic studies also suggest tumor cells secrete collagens [10].

Worldwide, gastric cancer remains one of the top deadliest malignancies [11]. Advanced gastric tumors are treated with surgery and chemotherapy with 5-year survival rates above 50% if the disease has not spread, and < 10% if metastasis has occurred [12]. The connections between therapy outcomes, the stroma, and collagens remains uncertain in stomach cancer. Physical properties of collagen fibers have been associated with outcomes in gastric cancer (GC) and these observations are likely driven by the most abundant collagen, type I [13]. Collagen type I expression has been associated with metastasis in early onset gastric cancer [14].

To gain new insights into the function of collagens in cancer, we hypothesized that collagens are significantly mutated in tumors and that these mutations impact disease progression, therapy response, and patient outcomes. We further hypothesized that somatic mutations in collagens would resemble mutations observed in many collagenopathies, providing insights to function of collagens expressed and secreted from cancer cells [15, 16].. Patients with collagenopathies have both missense and truncation mutations that can be either dominant or recessive and demonstrate a range of penetrance depending on the mutation and collagen [17–20]. Notably, the lessons from collagenopathies highlight how collagen mutations impact tissue even in the presence of wild-type collagen and even when collagens are expressed at low levels in the tissue.

There has been limited analysis of somatic mutations in the matrisome. COL2A1 has been reported to harbor recurring mutations in chondrosarcoma [21]. Screening by MutSig2CV across 27 cancers identified COL11A1, COL13A1, COL19A1, COL1A2, and COL4A4 as borderline significantly mutated [22], and 2 collagens were significant in the TCGA stomach adenocarcinoma (STAD) (Table S2). Functional studies of these variants were not pursued in larger -omic screens in part due to "technical limitations", as stated by the authors [22]. COL14A1 was reported to have a nonsynonymous mutation rate of 4.4% in Microsatellite Stable (MSS) gastric tumors, [23]. Grouping genes either by network methods [24] or by careful examination of specific gene families and mutation has provided insights into splicing regulators [25], TGF-β signaling [26], and complement genes [27], for example. Because there is some redundancy and overlap in function of collagens, we applied this approach to consider the collagens as a group, and to identify sets of collagens that may be significantly mutated and impactful in stomach cancer. We focused on collagens, as opposed to the whole matrisome to ease interpretation of the mutations, evaluate combinations, and leverage insights from collagenopathies to interpret the impact of mutations and propose specific hypotheses for future testing. A recent study by Izzi et al. also suggested that matrisome genes, including collagens, are significantly mutated across many cancer types [28]. This study focused on reporting the existence of somatic mutations, enrichment of mutations in some protein domains, and identified individual genes associated with patient survival. However, Izzi et al. did not consider gene combinations, gene families, or using germline mutations to interpret the potential

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 3 of 20

impacts of the somatic mutations. We also present a new approach to consider significant association with overall survival by defining a background set of genes considering mutation rate to identify collagen genes likely not associated with overall survival by chance. By focusing on collagens and leveraging the insights from collagenopathies, we also highlight how association with protein domains does not tell the whole story of the impact of somatic mutations of ECM factors.

In this article, we use bioinformatics to elucidate how collagen mutations affect gastric tumor outcomes. First, we find that many expressed collagens harbor somatic missense and truncation mutations, at a higher rate than expected compared to the background mutation rate. Next, we show that collagen genes and combinations of these genes associate with differential patient outcomes. We further investigate how collagen mutations correlate with tumor hallmarks, extracellular matrix components, and immune infiltration. Together, these findings suggest that collagen mutations impact stomach tumors via distinct tumor microenvironments, and that many collagens have unexpected novel functions in stomach tumors.

Methods

Data sources

The Cancer Genome Atlas (TCGA) Pan-Cancer RNA-seq V2 normalized gene expression and clinical data was downloaded from Firebrowse in April 2018. TCGA somatic mutation data file, mc3.v0.2.8.PUBLIC.maf, was downloaded from the Genomic Data Commons (GDC) [29]. Microsatellite data was downloaded from Firebrowse (STAD.merged_only_auxillary_clin_format. txt). Immune gene sets input into GSEA were defined by Tamborero et al. [30]. Stroma scores, and overall mutation rates were derived from Table S1 by Thorsson et al. [31]. Hallmark [32] and NABA [33] gene sets were downloaded from MsigDB v7.0 [34]. For the analysis of collagen mutations in the ACRG, collagen mutation data was obtained from the supplemental data published by Cristescu et al. [35]. Truncation mutations include all variants predicted to cause a shorter protein or degrade the mRNA including nonsense mutations, frameshift mutations and mutations that affect splicing. Germline collagen variants for every collagen with > 15 variants were downloaded from the Leiden Open Variation Database (LOVD) [36], except for COL7A1. COL7A1 pathological mutations were obtained from a DEB mutation database [37]. LOVD is an open source tool and database of gene-centered DNA variants.

Software and statistical tests

Analyses were performed using R and python custom scripts. GSEA version 2.4 was run on either a Unix or MacOS system. Statistical tests were performed using the Lifelines v0.25.1 and SciPy v1.5.2.

libraries in Python. Moderated Kolmogorov-Smirnov test was adopted from Olcina et al. to assess significance of collagen somatic mutations relative to other genes [27]. Morpheus was used to generate the heatmaps [38]. Survival curves were generated using cBioPortal's oncoprinter web app and matplotlib v3.3.1. Lollipop plots were generated via the MutationMapper tool in cBioPortal.

Identifying collagen gene combinations

We aimed to identify sets of collagen genes significantly associated with overall survival, accounting for gene size and mutation rate. To correct for multiple combinations occurring by chance, we calculated a q value for a given subset of collagen genes. To determine background, genes were randomly chosen until the expected number of mutations were within 5 of the number of observed mutations in collagen genes. A survival analysis was performed on the subset of patients used in the collagen subset analysis where the indicator variable was based on whether a patient has a mutation in the randomly chosen subset in at least 5% of cases of the designated cohort. We considered subsets with collagen genes significantly expressed with an average RSEM > 200. Table S3 lists the average RSEM scores. If a combination of 2 collagens was identified, this combination was not considered in combinations of 3 collagens. We then counted the frequency of each collagen included in the subsets as an indication of the contribution of each collagen to overall survival risk and exclusivity with the other collagens.

Case selection for immunohistochemistry

With institutional review board approval, IRB #1070389–9, 10 cases of gastric adenocarcinoma diagnosed from 2010 to 2019 were retrieved from the archives of the Department of Pathology and Laboratory Medicine at Lifespan Academic Medical Center (Providence, RI).

Immunohistochemistry

Immunohistochemistry staining for COL7A1 was performed on 4-μm paraffin sections. After incubation at 60 °C for 30 min, the sections were deparaffinized and rehydrated with xylene and graded alcohols. Antigen retrieval was performed with Ready-to-Use Proteinase K (Agilent, Santa Clara, CA) incubating at 37 °C for 10 min. The slides were then incubated with anti-COL7A1 antibody (1:5000) for overnight at 4 °C. The immunoreactivity was detected by using the DAKO Envision + Dual Link System and the DAKO Liquid 3,3′-diaminobenzidine (DAB+) Substrate Chromagen System (Agilent, Santa Clara, CA). Immunohistochemistry was assessed by 2 pathologists (MR and EW).

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 4 of 20

In situ hybridization

mRNA expression was determined using ISH with the RNAscope Assay (Advanced Cell Diagnostics, Hayward, CA). The ISH staining for COL7A1 was performed on 4-μm paraffin sections. After baking slides at 60 °C for 1 h and deparaffinizing FFPE sections with xylene, RNAscope® 2.5 HD Reagent Kit was used for the ISH assay. All the steps were done according to the kit protocol. After pre-treating the sample with hydrogen peroxide solution, heat target retrieval and protease plus, COL7A1 probe was added for 2 h at 40 °C, sequentially hybridize with AMP 1, AMP 2, AMP 3, AMP 4, AMP 5, and AMP 6 reagents, for 30, 15, 30, 15, 60, 15 min, respectively. ISH signal was detected by the application of a chromogenic substrate. Tissue was counter-stained with haematoxylin. Scrambled negative control probes showed no signal.

Antibody sources

Rabbit polyclonal anti-COL7A1 targeting the human LH7.2 domain was a kind gift from Alexander Nystrom, University of Freiburg [39].

Results

Collagen mutations are prevalent in STAD

We evaluated the frequency of somatic mutations in the 43 human collagen genes. We observed a clear bias in the distribution of the frequency of mutations in collagens compared to other genes (p <1e-16, Wilcoxon Rank test) (Fig. 1A). Five individual collagen genes are mutated at frequencies larger than 8% (Fig. 1B). Frequently mutated genes include COL12A1, COL11A1, COL6A2, and COL7A1, representing a range of collagen families and functions (see Table S1 for collagen family information). To account for the range of mutation rates in stomach tumors, we evaluated the MSS and MSIH types separately and found frequent somatic mutations in collagens in both MSIH and MSS tumors (Fig. 1C). Some collagens such as COL12A1 and COL4A1 showed high mutation rates in both MSIH and MSS tumors, while others such as COL7A1 were frequently mutated in MSIH, but not MSS tumors. Every MSIH tumor has at least one mutation in a collagen gene. In MSS tumors, COL12A1 was the most frequently mutated at 8% with only 20 tumors harboring any collagen truncation mutation.

Because collagen somatic variants are relatively rare and have not previously been identified as significantly mutated by standard algorithms, we evaluated if collagen genes were significantly mutated relative to the background mutation rate using multiple approaches. By MutSigCV2, only 2 collagen genes were significantly mutated in the TCGA gastric cancer cohort while 2 other collagens had borderline q-values (Table S2). To determine the significance of somatic mutations in collagens relative to other genes, accounting for mutation rate, we applied a modified Kolmogorov-Smirnov (KS) test [27]. KS test analysis revealed that as a group, mutation rate of collagens, accounting for gene size, occurred significantly above background (Fig. 1D). Because some GC tumors are MSIH with high mutation rates compared to MSS tumors, we determined that collagens and subsets of collagen genes had significantly higher mutation rates in these more specific cohorts as well (Fig. 1D).

Collagens are mutated at similar rates in independent datasets

We examined independent datasets to assess if other GC cohorts harbor similar mutations rates of collagens. 52% of tumors have at least one somatic mutation in a collagen in the Pfizer/Hong Kong whole genome sequencing dataset in 100 cases collected in Hong Kong, including a 19% rate of truncation mutations [40], but patient survival data is not available. The Asian Cancer Research Group (ACRG) performed targeted sequencing of 251 gastric tumors including a selection of collagens including COL11A1, COL12A1, COL21A1, COL22A1, COL4A1, COL5A1, COL5A3, COL6A3, and COL6A5 [35]. Recurrent variants of the collagen genes tested were reported at frequencies slightly lower than observed in TCGA (Fig. S1). Tumors harboring at least one mutation in COL11A1, COL5A1, COL5A3, COL6A3, COL6A5, or COL4A1 were moderately associated with improved outcomes in both TCGA and ACRG (Fig. S1B). Major differences in the studies included that ACRG only sequenced patients of Asian ethnicity compared to TCGA including mostly Caucasians. Patients in the ACRG cohort had a longer overall survival (OS) than the TCGA patients (<50% survival vs. >60% survival at 5 years), and the MSI cases in the ACRG cohort showed a stronger association with improved outcomes compared to the TCGA cohort (Fig. S5A).

Collagen mutations associated with clinicopathological characteristics

Collagen mutations classify STAD tumors independently of clinicopathological characteristics including

(See figure on next page.)

Fig. 1 Collagens are significantly mutated in stomach adenocarcinoma in the TCGA dataset. **A**. Distribution of alteration frequencies for collagen genes (orange) compared to all other genes (blue) in the TCGA STAD cohort. *P*-value determined by Wilcoxon rank test comparing the distribution of collagen genes relative to all other genes. **B**. Alteration frequencies for each collagen gene in all TCGA STAD cases. **C**. and in MSS, MSIH, and MSIL STAD cases. **D**. Kolmogorov-Smirnov moderated tests suggest that collagen genes as a group are significantly mutated compared to gene sets of similar size and length in the whole TCGA STAD cohort and in both MSS and MSIH tumors
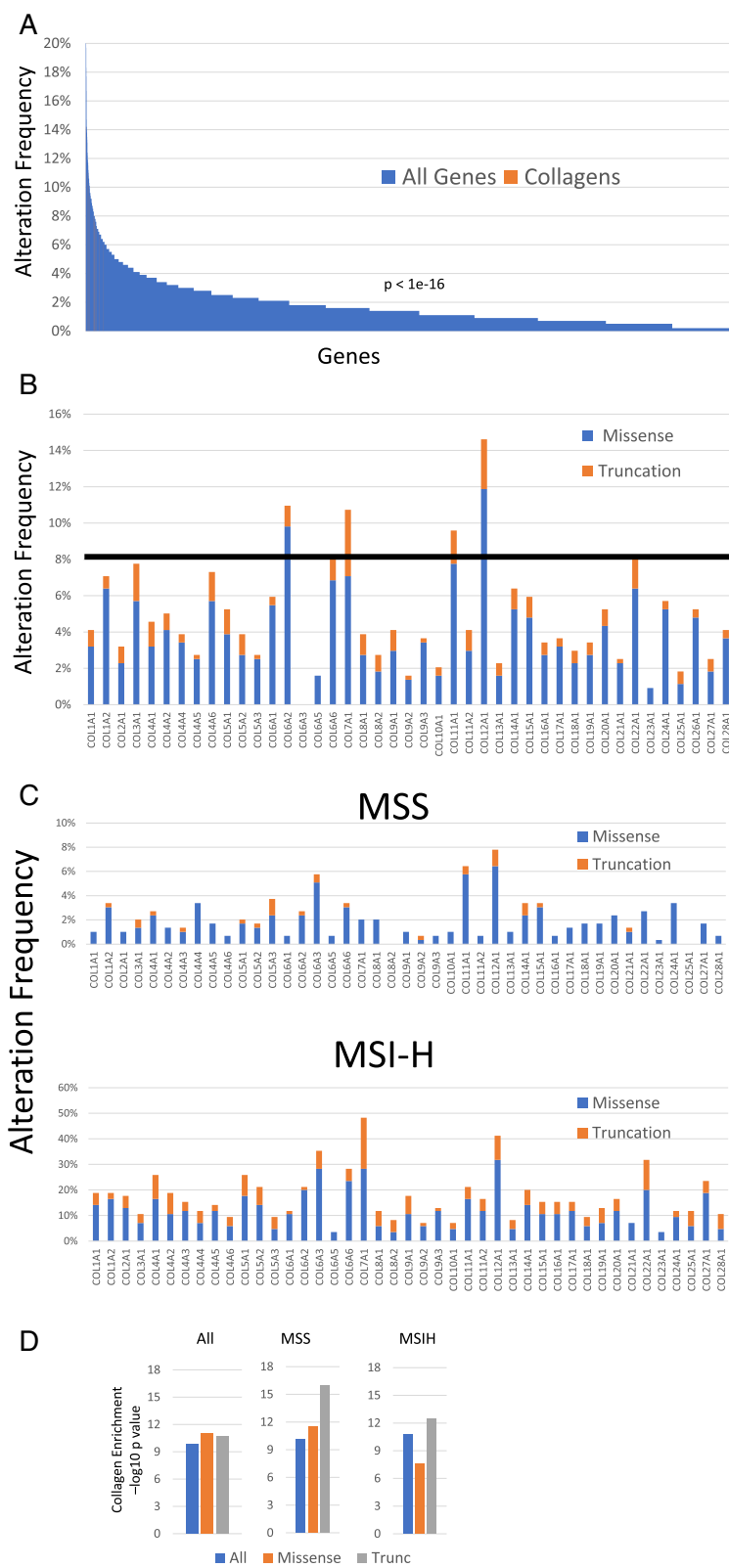
Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 5 of 20

Fig. 1 (See legend on previous page.)

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 6 of 20

stage, grade, MSI status, and mutation rate (Tables 1 and 2). Age at diagnosis was associated with missense mutations and the overall mutation rate, consistent with MSIH tumors' known association with older patients [41]. Previous STAD classification identifies 4 major groups: Epstein-Barr Virus (EBV), High Mutation (HM), Genomically Stable (GS), and chromosome instability (CIN). Almost every tumor in HM, characterized by high mutation rates, has at least one mutation in a collagen gene, but also, 56% of the CIN and 36% of the GS groups have collagen mutations even though the mutation rates are much lower in these groups. Neither EBV nor *H. pylori* status was associated with collagen mutations, or COL7A1 mutations (Table 1). These TCGA defined classifications were not associated with patient outcomes.

Collagen mutations associated with patient survival

We evaluated the association of tumors harboring a somatic nonsynonymous missense or truncation mutation in any collagen with OS by Kaplan-Meier analysis in the TCGA STAD dataset (Fig. 2A). Tumors with at least one mutation in any collagen were not associated with OS, but tumors with at least one truncation mutation in any collagen were significantly associated with longer OS (Fig. 2A). Only COL5A2, COL11A1, and COL19A1 were significantly associated with longer OS when considered as individual genes (Fig. S1). COL23A1 was associated with shorter survival but is mutated in only 4 cases and is not significantly expressed in STAD (Table S3).

To address the potential redundancy of collagen functions and since many collagens were not mutated in sufficient number of cases for survival analysis, we undertook a combinatorics approach to identify sets of tumors with mutated collagen genes associated with OS more significantly than a bootstrap defined background accounting for number of patients, mutation rate and number of genes (Fig. 2B). We evaluated all combinations of tumors with at least one mutation in 2 or 3 expressed collagen genes (Table S4) and tested their association with OS. The combinatorics approach identified collagen gene sets associated with OS at $q \leq 0.05$ (Fig. 2C). COL5A2, COL4A1, COL11A1, COL15A1, and COL16A1 were the most frequently included collagen genes in the combinations (Fig. S3A), and each at least trended with longer OS on their own (Fig. S2A). Truncation mutations were particularly strongly associated with OS with no combinations identified associated with shorter OS at these thresholds (Fig. 2A). Of the 104 tumors with at least one truncation mutation, 65% were in MSIH tumors (Table 1). We identified combinations of 2 or 3 collagen genes with truncation mutations strongly associated with OS and COL12A1 was the most collagen gene

most frequently included in these combinations (Fig. 2C, Table S4).

Collagen genes classify MSIH tumors by overall survival

Because the majority of collagen mutations were in MSIH cases, and because of the differences in MSIH and MSS stomach tumors in treatment response, we evaluated each of these groups separately. For simplicity and to observe differences between MSIH and MSS differences more clearly, we removed the MSIL annotated tumors to avoid complications from these tumors with moderate mutation burden that are clinically treated as MSS tumors. Even in just MSIH cases, most truncation mutations in collagens were associated with longer survival including in COL1A1, COL5A2, COL11A1, and COL15A1 (Fig. 2D). A few collagens were associated with shorter survival; including most notably, COL5A3. All patients with either a COL1A1 or COL5A2 truncation in MSIH tumors were associated with longer OS (Fig. 2D). In MSS tumors, COL5A3, COL6A2, COL11A1, and COL24A1 were associated with longer OS (Fig. S3C).

Combining the top truncation variants identified by combinatorics defined a group of tumors strongly associated with longer OS in MSS and MSIH tumors (Figs. S3E and S4). These observations suggest that loss of expression of many collagens, especially those involved in collagen type I expression and formation (see Table S1), such as COL12A1, was associated with increased OS in both MSS and MSIH tumors. On the other hand, the loss of function of some collagen type I regulating collagens, such as COL5A3 and COL14A1, is detrimental to patients with MSIH tumors. These observations can be explained by considering that COL5A3 and COL14A1 are negative regulators of collagen type I fiber size. LOF mutations in these two collagens lead to gain of function of collagen type I. COL14A1 is a Fibril Associated Collagens with Interrupted Triple helices (FACIT) collagen that regulates collagen fibrillogenesis such that absence of COL14A1 leads to larger fibers in mice [42–44]. Increasing collagen type I fiber width has been associated with poor outcomes in stomach cancer [13]. Analogous to the observations in stomach tumors, mutations of collagen α3(V) chains have phenotypes distinct from collagen α1(V) and α2(V) chains in mice [45]. Germline mutations in COL5A1 and COL5A2 cause Ehlers-Danlos-like phenotypes, while mutations in COL5A3 affect adiposity [45]. On the other hand, LOF mutations in other regulators of fiber formation such as COL11A1 [46] were associated with longer OS (Fig. S2). These observations suggest that regulation of COL1A1 and fibrillogenesis, when unchecked, leads to even worse survival in MSIH cases. These examples demonstrate

Table 1 Association of collagen mutation status with clinicopathological characteristics

Factor	MSI Status (MSI-H, MSI-L, MSS)	All Collagen Mutations	All Collagens Missense	All Collagens Truncation	All COL7A1 Mutations	Mutation Rate (High/Low)
Microsatellite		**P<1e-3**	**P<1e-3**	**P<1e-3**	**P<1e-3**	**P<1e-3**
MSI-H		100% (85/85)	100% (85/85)	81% (69/85)	41% (35/85)	100% (84/84)
MSI-L		63% (40/63)	59% (37/63)	13% (8/63)	0% (0/63)	59% (37/63)
MSS		48% (143/295)	46% (135/295)	8% (23/295)	2% (6/295)	33% (98/293)
Grade	P=0.5	P=0.9	P=0.9	P=0.9	P=0.7	P=0.3
G1	33% (4/12) MSIH 8% (1/12) MSIL 58% (7/12) MSS	67% (8/12)	67% (8/12)	33% (4/12)	8% (1/12)	67% (8/12)
G2	17% (26/156) MSIH 15% (23/156) MSIL 69% (107/156) MSS	63% (98/156)	61% (95/156)	22% (35/156)	7% (11/156)	54% (82/153)
G3	21% (54/259) MSIH 15% (39/259) MSIL 64% (166/259) MSS	61% (157/259)	58% (149/259)	23% (60/259)	11% (28/259)	47% (121/259)
GX	11% (1/9) MSIH 0% (0/9) MSIL 89% (8/9) MSS	56% (5/9)	56% (5/9)	11% (1/9)	11% (1/9)	67% (6/9)
TNM Stage	P=0.5	P=0.4	P=0.4	P=0.3	P=0.4	P=0.10
I	26% (15/57) MSIH 18% (10/57) MSIL 56% (32/57) MSIH	70% (40/57)	68% (39/57)	28% (16/57)	7% (4/57)	57% (32/56)
II	21% (27/131) MSIH 15% (20/131) MSIL 64 % (84/131) MSS	61% (80/131)	56% (74/131)	24% (31/131)	9% (12/131)	56% (73/131)
III	16% (29/186) MSIH 13% (24/186) MSIL 72% (133/186) MSS	58% (108/186)	56% (105/186)	18% (34/186)	9% (16/186)	45% (83/186)
IV	16% (7/44) MSIH 16% (7/44) MSIL 68% (30/44) MSS	59% (26/44)	57% (25/44)	27% (12/44)	16% (7/44)	40% (17/42)
Gender	**P=2e-3**	P=0.5	P=0.4	**P=0.02**	P=0.11	P=0.05
Male	14% (40/279) MSIH 15% (43/279) MSIL 70% (196/279) MSS	63% (175/279)	61% (169/279)	19% (54/279)	8% (21/279)	46% (129/278)
Female	29% (45/157) MSIH 13% (23/157) MSIL 59% (92/157) MSS	59% (93/157)	56% (88/157)	30% (46/157)	13% (20/157)	57% (88/155)
Age	**P=2e-3**	**P=1e-3**	**P<1e-3**	P=0.2	P=0.5	**P<1e-3**
≤65	12% (23/192) MSIH 16% (30/192) MSIL 72% (139/192) MSS	101/192 (53%)	95/192 (38%)	38/192 (20%)	7% (16/192)	40% (76/192)
>65	26% (61/239) MSIH 13% (32/239) MSIL 61% (146/239) MSS	163/239 (68%)	158/239 (66%)	61/239 (26%)	11% (26/239)	59% (139/237)
Immune Subtype	**P<1e-3**	**P=1e-3**	**P=4e-3**	P=0.07	P=0.05	**P<1e-3**
C1	14% (18/129) MSIH 19% (24/129) MSIL 67% (87/129) MSS	66% (85/129)	61% (79/129)	21% (27/129)	7% (9/129)	50% (65/129)
C2	27% (57/210) MSIH 12% (25/210) MSIL 61% (128/210) MSS	65% (135/210)	62% (131/210)	28% (58/210)	13% (28/210)	59% (123/210)
C3	0% (0/36) MSIH 6% (2/36) MSIL 94% (34/36) MSS	31% (11/36)	31% (11/36)	6% (2/33)	0% (0/33)	8% (3/36)
C4	11% (1/9) MSIH 22% (2/9) MSIL 67% (6/9) MSS	78% (7/9)	78% (7/9)	22% (2/9)	0% (0/9)	44% (4/9)
C5	0	0% (0/0)	0% (0/0)	0% (0/0)	0% (0/0)	0% (0/0)
C6	0% MSIH 0% MSIL 100% (7/7) MSS	57% (4/7)	43% (3/7)	14% (1/7)	0% (0/7)	0% (0/7)
NA	16% (8/49) MSIH 20% (10/49) MSIL 63% (31/49) MSS	49% (24/49)	49% (24/49)	18% (9/49)	6% (3/49)	49% (24/49)
TCGA Subtype	**P<1e-3**	**P<1e-3**	**P<1e-3**	**P<1e-3**	**P<1e-3**	**P<1e-3**
EBV	0 MSIH 20% (4/30) MSIL 87% (26/30) MSS	57% (17/30)	53% (16/30)	7% (2/30)	0% (0/30)	43% (13/30)
HM	93% (74/80) MSIH 3% (2/80) MSIL 5% (4/80) MSS	99% (79/80)	99% (79/80)	79% (63/80)	41% (33/80)	100% (80/80)
GS	0 MSIH 4% (2/50) MSIL 96% (48/50) MSS	36% (18/50)	32% (16/50)	6% (2/50)	0% (0/50)	14% (7/50)
CIN	0.4% (1/223) MSIH 20% (44/223) MSIL 80% (178/223) MSS	56% (124/223)	52% (116/223)	9% (20/223)	1% (3/223)	41% (91/223)

Table 2 Univariate and multivariate analysis by cox proportional hazards analysis. Multivariate survival analysis of all variables with $p < 0.05$ in univariate analysis by cox proportional hazards analysis

Group	OS HR	OS P
All Collagens; all mutations	0.8 (0.6-1.1)	0.2
All Collagens; missense only	0.9 (0.7-1.2)	0.6
All Collagens; truncation only	0.5 (0.4-0.8)	**0.002**
COL5A2; all mutations	0.3 (0.1-0.8)	**0.02**
Grade	1.3 (1.0-1.7)	0.07
Stage	**1.7 (1.4-2.0)**	**<1e-3**
MSI status (MSS vs. MSIH)	0.7 (0.4-1.1)	0.09
Age at diagnosis	1.0 (1.0-1.0)	**0.03**
Nonsilent mutation rate	1.0 (1.0-1.0)	0.06
Stromal Fraction	**2.4 (1.1-5.3)**	**0.04**
Leukocyte Fraction	1.5 (0.5-4.6)	0.5
TGF-beta Response	**1.1 (1.0-1.2)**	**0.004**
MSIH only (N=68)		
COL14A1/COL5A3; all mutations	**3.2 (1.4-7.3)**	**0.006**
COL7A1/COL11A1; all mutations	**0.4 (0.2-0.9)**	**0.02**
Stage	1.7 (1.0-2.9)	0.06
Stroma	1.8 (0.2-20)	0.6
TGF-beta Response	**1.3 (1.0-1.5)**	**0.02**
Age at diagnosis	1.1 (1.0-1.1)	0.007
Nonsilent mutation rate	1.0((1.0-1.0)	0.4
MULTIVARIATE		
All cases (N=388)		
Stage	**2.0 (1.5-2.5)**	**<1e-3**
All Collagens; truncation only	**0.5 (0.3-0.9)**	**0.02**
Stromal Fraction	1.8 (0.7 – 4.3)	0.2
MSIH only (N=68)		
TGF-beta Response	**10.1 (2.8-37.5)**	**0.007**
Missense COL5A3	**1.3 (1.1-1.6)**	**<1e-3**
MSS Only (N=289)		
Stage	**1.7 (1.3-2.1)**	**<1e-3**
COL11A1, COL16A1, COL4A1, COL6A2 missense	**0.37 (0.2-0.8)**	**0.02**

(See figure on next page.)

Fig. 2 Identification of collagen genes mutations associated with overall survival. **A**. Patients with tumors that harbor at least one mutation in a collagen gene, have significantly better outcomes in the STAD TCGA cohort. Patients with tumors with at least one collagen mutation of the type indicated in red. Wild-type tumors in blue. Log-rank test *p*-values shown. Truncation mutations in any collagen gene were associated with better outcomes while nonsynonymous missense mutations were not associated with overall survival. Both missense and truncation mutations in COL5A2 were associated with longer overall survival. **B**. Schematic of approach to identify tumors with combinations of mutated collagens associated survival more significantly relative to background accounting for mutation rate, gene size and number of patients. **C**. Frequency of the inclusion of each collagen gene with a truncation mutation in a combination significantly associated with overall survival. A representative combination of collagen genes strongly associated with overall survival curve and the oncoprint. **D**. Identification of collagen genes with truncation mutations in MSIH tumors most strongly associated with overall survival. Frequency of the inclusion of each collagen in subsets consisting of 2 and 3 collagen genes with truncation only mutations in MSIH tumors
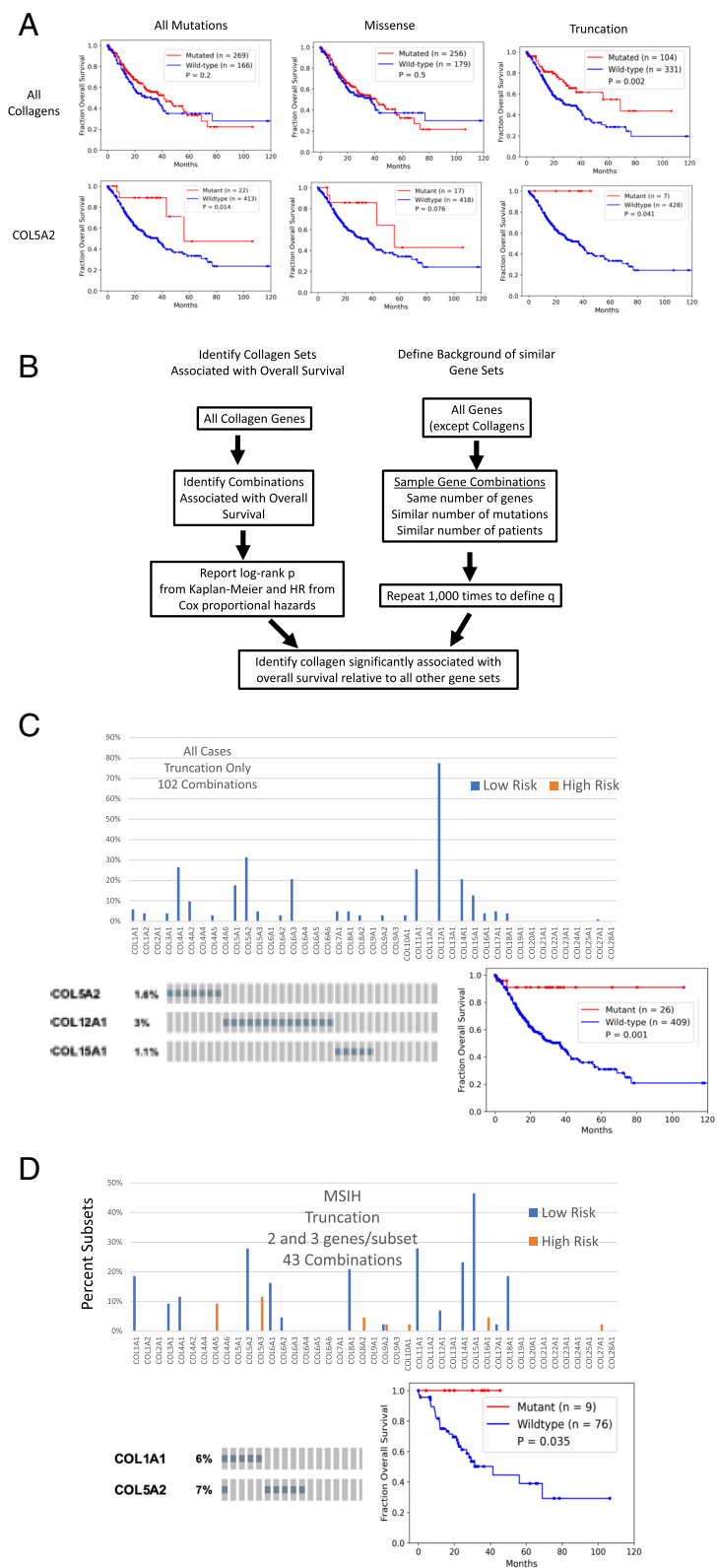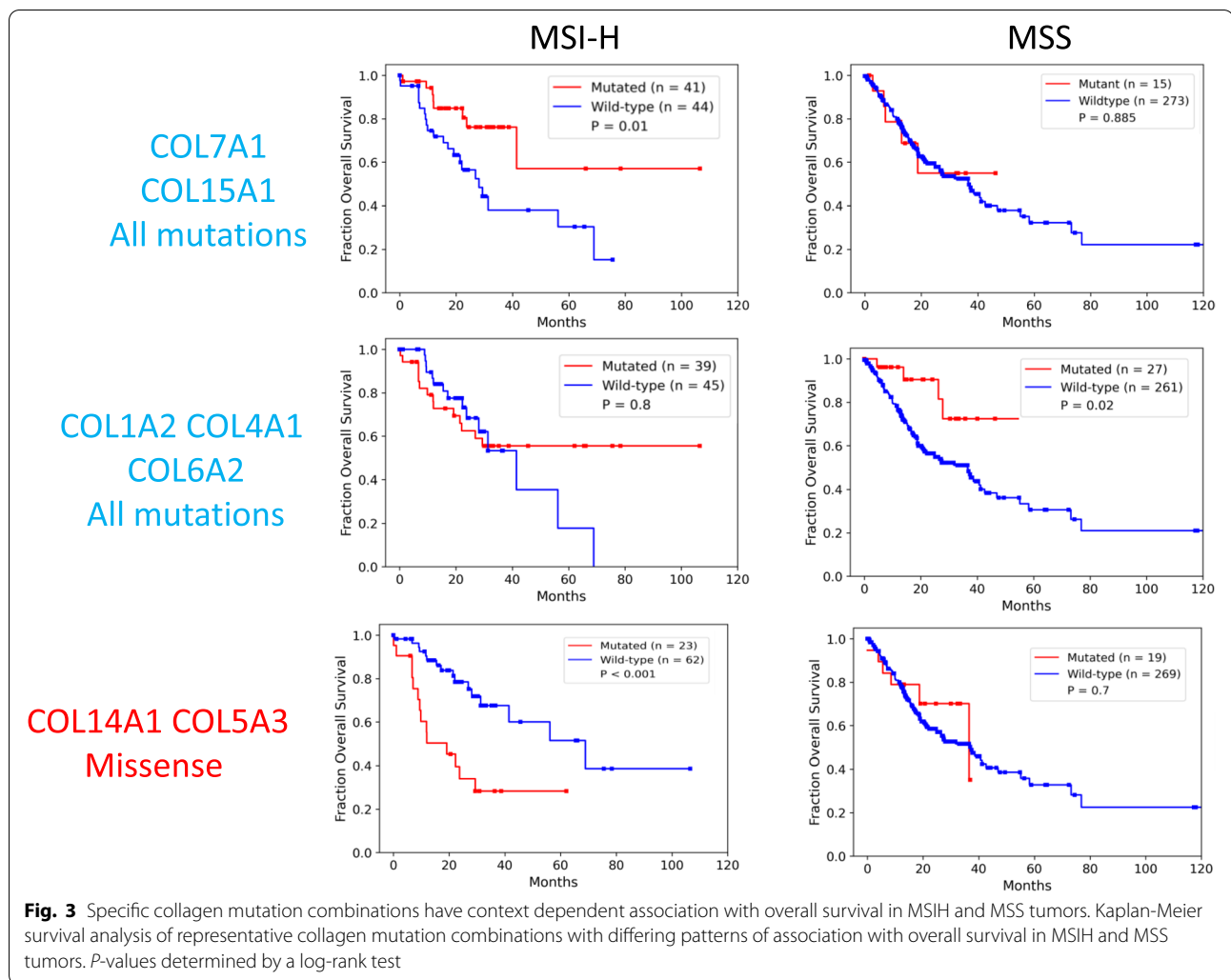
Fig. 2 (See legend on previous page.)

Brodsky *et al. BMC Cancer* (2022) 22:139

Page 10 of 20

Fig. 3 Specific collagen mutation combinations have context dependent association with overall survival in MSIH and MSS tumors. Kaplan-Meier survival analysis of representative collagen mutation combinations with differing patterns of association with overall survival in MSIH and MSS tumors. *P*-values determined by a log-rank test

how we can leverage observations of collagen biochemistry and collagenopathies to interpret the impact of mutations and generate novel hypotheses to begin to explain the range of treatment responses and patient survival.

Many collagen gene combinations' associations with OS were specific for MSIH or MSS tumors (Table S4, Fig. S4, Fig. 3). Representative sets were associated with OS in either MSIH or MSS tumor, but not both (Fig. 3). In particular, even though COL5A3 and COL14A1 were mutated at similar levels, these collagens showed MSI status dependent associations with OS.

We applied the cox proportional hazards model to test the relationship of collagen mutations with other common survival associated characteristics including age and stage. Multivariate analysis showed that collagen mutations were independent predictors of OS compared to other clinicopathological characteristics including stage (Table 2). Neither mutation rate nor MSI status were

associated with OS, despite including many collagen mutations. These findings suggest that tumors with collagen mutations specifically define a class of tumors with distinct properties and treatment responses.

Collagen mutations impact on stomach tumors

To gain insight into how collagens could be affecting STAD tumors, we used pre-ranked Gene Set Enrichment Analysis (GSEA) of TCGA normalized RSEM scores to identify biological processes associated with tumors that harbor a collagen mutation compared to tumors without collagen mutations. We first evaluated the 50 MSigDB hallmark gene sets [34] (Fig. 4). There was high similarity of the impact of collagen mutations on expression of cancer hallmarks highlighted by the higher expression of cell cycle drivers including E2F targets and MYC gene sets (Fig. 4A, B). On the other hand, EMT, KRAS and myogenesis gene sets were expressed higher in wild-type compared to collagen mutant tumors (Fig. 4A, B). Lower

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 11 of 20

expression of the EMT hallmark in tumors with collagen mutations is consistent with reduced collagen function and an altered ECM that leads to more epithelial features in these tumors [47].

To test if the ECM was different in tumors with collagen mutations, we evaluated the NABA ECM gene sets. The NABA defined matrisome gene sets are a group of curated gene sets based on a combination of protein domains and secreting signals to ensure these are matrisome components [33]. The NABA gene sets were expressed lower in tumors with mutations compared to wild-type tumors when considering the full TCGA STAD cohort, consistent with a disrupted ECM in the collagen mutated tumors relative to the wild-type tumors (Fig. 4B). Total collagen expression has been associated with patient outcomes in many cancers [48]. Although collagens as a group were expressed lower in mutant tumors, only rarely was the expression directly associated with mutation within that collagen gene (Table S3). This may be because many of the collagens are expressed from both cancer and stroma cells, obscuring the relationship between collagen mutation and expression in bulk RNAseq data.

Collagen mutations associated with distinct tumor microenvironments

Collagens can mediate the migration and infiltration of immune cells in tumors [49]. To evaluate association between collagen mutations and immune cell infiltration, we evaluated immune cell signatures [30] with pre-ranked GSEA. MSIH tumors have higher expression of most of the immune cell expression signatures compared to MSS tumors, consistent with more immune cell infiltration in MSIH tumors (Fig. S5). Together with lower expression of the NABA ECM gene sets in MSIH tumors (Fig. S5), these observations suggest that the MSIH and MSS tumors differ in their tumor microenvironments. Immune cell gene expression signatures were expressed higher in MSIH tumors compared to MSS tumors, consistent with a more inflammatory environment and higher tumor mutation burden in MSIH compared to MSS tumors. Because of these large differences in the tumor environments of MSS and MSIH tumors, considering all the stomach tumors together may be obfuscating

impacts. We therefore evaluated the impact of collagen mutations in the whole cohort as well in MSIH and MSS tumors separately.

Combinations of collagen mutations had a more consistent impact on the expression of ECM and immune cell gene signature in MSS tumors, compared to more variable associations in MSIH tumors (Fig. 4 and Figs. S6-S10). Figure 4 shows representative combinations and pre-ranked GSEA of all combinations listed in Table S4 are shown in Figs. S6-S10. The consistent nature of the tumors with collagen mutations suggests that changes in EMT, expression in basement membranes, and many immune cells, are common features of tumors with collagen mutations in both MSS and MSIH tumors. This consistency could be because we selected for tumors with collagen mutations associated with overall survival and these tumors have similar mechanisms of impact in mediating EMT and expression of the basement membrane.

In MSS tumors, tumors with collagen mutations had consistently lower expression of extracellular matrix gene sets and the majority of the immune cell gene sets (Fig. 4C). While in MSIH tumors, the ECM NABA and immune cell signature gene sets were split into 2 groups (Fig. 4D). Notably, tumors with COL14A1 and COL5A3 mutations were associated with higher expression of NABA gene sets in MSIH tumors and shorter OS (Fig. 4D). On the other hand, other fibril associated collagens such as COL11A1 and COL5A2 were associated with longer OS (Fig. S3, Table S4). COL11A1, COL5A1, and COL5A2 promote fibril formation and loss of function mutations of these collagens have been associated with smaller collagen type I fibers [45]. Mutations in COL1A1, COL11A1, COL5A1, and COL5A2 were all associated with lower expression of the EMT hallmark gene set compared to wild type in MSIH tumors, while mutations in COL14A1 and COL5A3 were associated with higher expression of EMT expression signature in MSIH tumors (Fig. 4D). These observations predict that tumors with mutations in collagen types XIV and Vα3 would have thicker fibers, promoting cell migration, and subsequent tumor cell escape. On the other hand, tumors with in COL1A1, COL11A1, COL5A1, and COL5A2 are predicted to have thinner or fewer collagen type I fibers

(See figure on next page.)

Fig. 4 Tumors with collagen mutations have distinct expression of cancer hallmarks and tumor environments. **A**. Representative enrichment plots from pre-rank GSEA suggest upregulation of E2F regulated transcripts and down-regulation of the expression of the matrisome in tumors with collagen mutations. TCGA stomach tumors were classified by collagen mutation status and pre-ranked GSEA revealed associations with the indicated gene set. **B**. Heat map of normalized enrichment scores of the cancer hallmark, immune cell, and NABA ECM gene sets. Red indicates higher expression in mutant tumors and blue indicates higher expression in wild-type tumors. Nonsignificant and modest enrichment scores between −1.5 and 1.5 are in white. In the full STAD TCGA cohort, tumors with any collagen mutation or with only mutations in COL7A1 or COL11A1 showed similar patterns. **C**. Heatmap of gene sets in MSS only tumors. **D**. Heatmaps of gene sets in MSIH only tumors reveal a more diverse pattern of enrichment. All heatmaps generated in Morpheus [38]
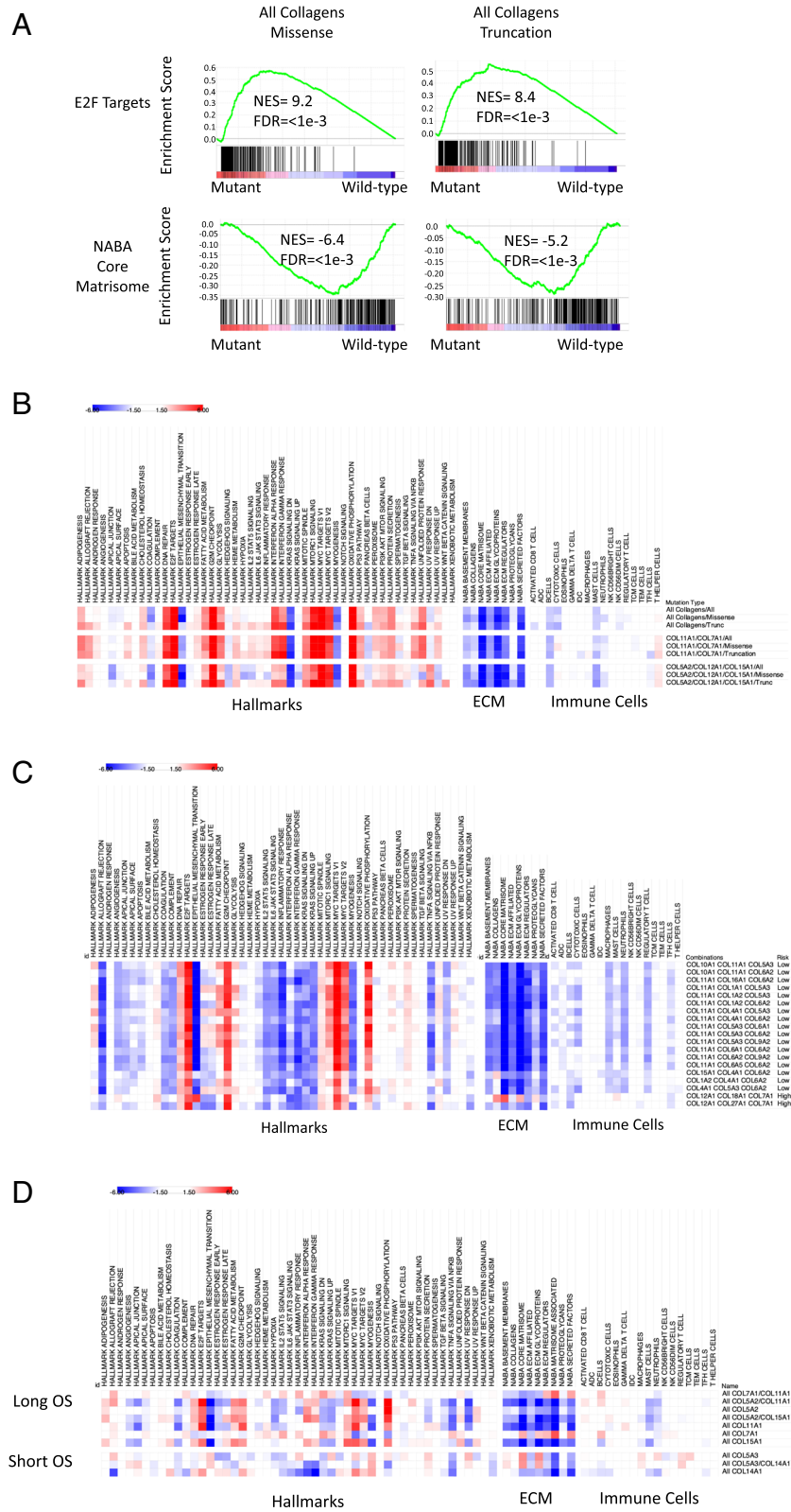
Brodsky *et al. BMC Cancer* (2022) 22:139

Page 12 of 20

Fig. 4 (See legend on previous page.)

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 13 of 20

leading to less migration, lower mesenchymal properties, less metastasis, and higher sensitivity to treatments.

Immunoenvironment

We used the immune cell gene sets reported by Tamborero et al. to evaluate the immunoenvironment by pre-rank GSEA [30]. Across the whole cohort, B and mast cells were lower in tumors with a variety of mutant collagens, while T helper cells were modestly increased in some mutant collagen tumors (Fig. 4A). In MSS tumors, Mast cells, macrophages, neutrophils were lower in mutant tumors associated with longer survival, and higher in wildtype tumors (Fig. 4B). Mast cells and neutrophils have been associated with short OS in STAD [50, 51]. When considering just the MSIH tumors, neutrophils and mast cells were lower in mutant tumors, but no other clear patterns emerged across the cohort. Immunosuppressive cell types including macrophages and regulatory T cells are higher in many of the shorter OS mutation combinations in both MSS and MSIH tumors. Tumors with COL14A1 mutations for example had higher levels of macrophages, neutrophils and mast cells Mast cells have been associated with shorter OS in stomach cancer [50]. These observations, based on molecular signatures, suggest changes to the immunoenvironment in tumors with collagen mutations.

COL7A1 mutations

We hypothesized that insight into the functional impact of mutations can be gained by comparing the pattern and type of mutation to those observed in collagenopathies. As an example, we focused on COL7A1 which is the mutational cause of Dystrophic Epidermolysis Bullosa (DEB) and had significant associations with patient outcomes in MSIH tumors (Fig. 2B). COL7A1 germline mutations were downloaded from a DEB mutation database [37]. The distribution of germline and stomach somatic mutations were very similar (Fig. 5B). A Kruskal-Wallis test ($P = 0.3$) suggested that the two distributions were not significantly different. COL7A1 mutations in STAD were slightly more biased towards the N-terminus of the protein compared to DEB. The largest exon, exon 73, was most frequently mutated in both the germline and cancer mutations. The recurring nonsense mutation at position 2029 is the same hotspot observed in DEB patients (Fig. 5A and B). The distribution of somatic mutations resembled the distribution of DEB germline mutations suggesting that no unusual tumor specific mutation pattern is prevalent in stomach tumors. Moreover, because the type of mutation is similar as observed in DEB, we can infer the function of the somatic mutations in tumors. For example, variants in the collagen domains, especially mutations changing the Gly of the G-X-X

repeat are typically dominant [52]. Meanwhile, variants in the N-terminal NC1 domain often reduce COL7A1 expression [54].
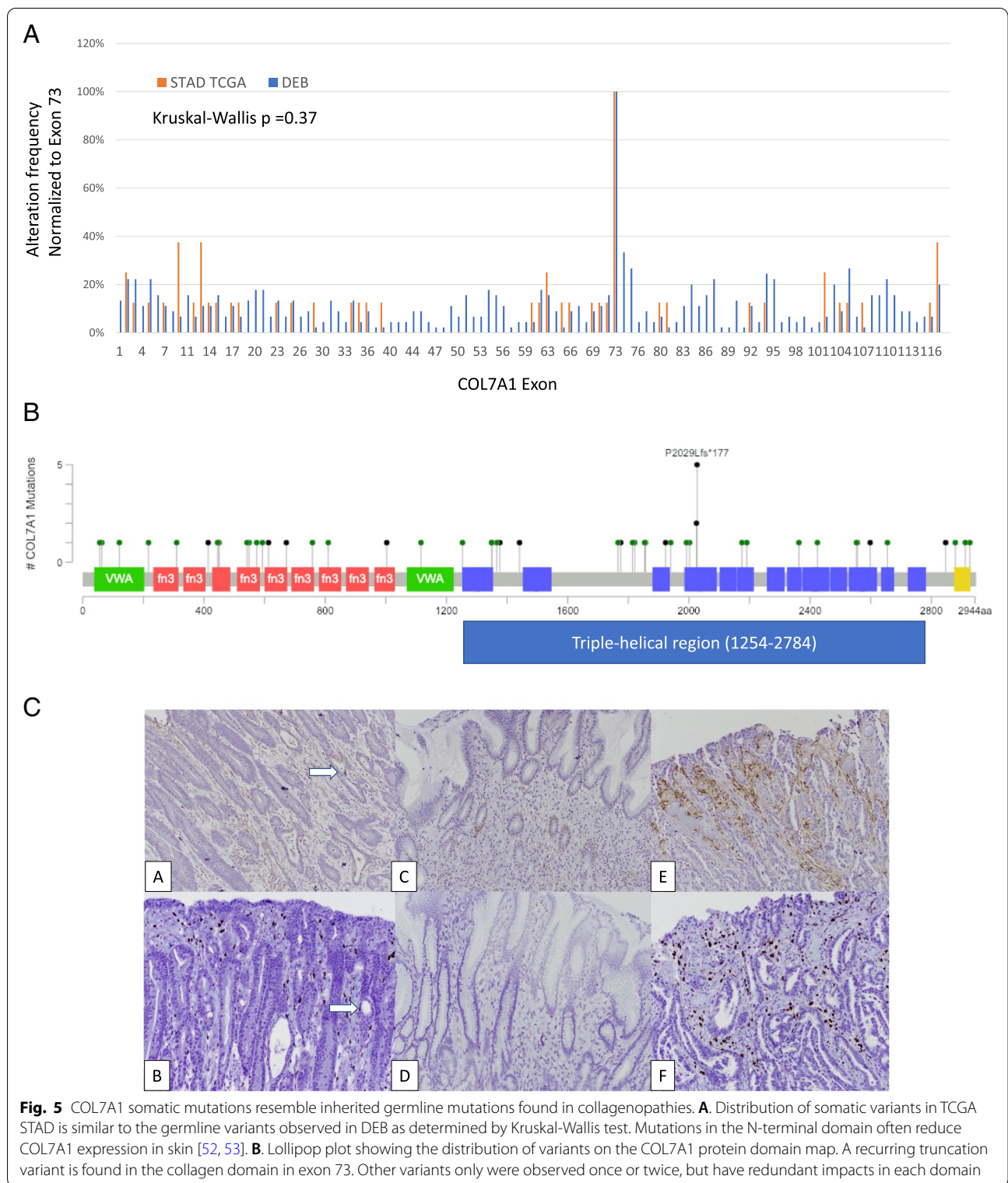
Expression of COL7A1 in stomach tumors

Many minor collagens have a high tissue specificity including COL7A1 [55]. Because COL7A1 is not known to be expressed in normal stomach, or anywhere in the gastrointestinal tract, we wanted to confirm that COL7A1 protein was expressed in stomach tumors, and importantly determine if COL7A1 was expressed in cancer cells and not just in the stroma. We evaluated COL7A1 protein expression by immunohistochemistry in a set of 10 stomach tumors from patients treated at Rhode Island Hospital (Fig. 5C and S11). COL7A1 was expressed in the stroma (4/10), in the epithelium (3/10) or was not detectable (3/10) (Table S5). In the epithelium, COL7A1 was expressed in the cytoplasm, similar to skin cells highly expressing COL7A1 (Fig. S11C), suggesting that these tumor cells are also expressing and secreting high levels of COL7A1. Expression from tumor cells was confirmed by in situ hybridization using RNAscope, controlling for non-specific antibody staining (Fig. 5C). Other collagens have been shown to be expressed in tumor cells by in situ hybridization such as collagen type IV [9]. Hynes and colleagues have suggested that matrisome components secreted from tumors cells in pancreatic tumors are more impactful than those originating from the stroma [56]. Larger studies are needed to evaluate any connection between COL7A1 expression patterns and mutation status.

Discussion

This work demonstrates that collagen mutations are significant relative to background, stratify patients in meaningful ways, and are likely impactful in STAD. Although this is an association study, we believe that this report will inspire additional investigation into the function of collagens in tumors and specifically in stomach cancer.

In collagenopathies, many causative collagen mutations are heterozygotes and genetically dominant because the missense mutation forms destabilized triple helices. A second class of missense mutations in non-collagen domains and truncation mutations both reduce or eliminate expression of the collagen. Compared to tumors secreting wild-type versions of these collagens, stomach tumors with reduced levels of these collagens respond better to treatment and have distinct expression of multiple cancer hallmarks.

Izzi and co-workers recently reported on somatic mutations in the matrisome including collagens PanCancer [28]. Their PanCancer approach focused on identifying common features across multiple cancer types while this

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 14 of 20

Fig. 5 COL7A1 somatic mutations resemble inherited germline mutations found in collagenopathies. **A**. Distribution of somatic variants in TCGA STAD is similar to the germline variants observed in DEB as determined by Kruskal-Wallis test. Mutations in the N-terminal domain often reduce COL7A1 expression in skin [52, 53]. **B**. Lollipop plot showing the distribution of variants on the COL7A1 protein domain map. A recurring truncation variant is found in the collagen domain in exon 73. Other variants only were observed once or twice, but have redundant impacts in each domain

study focused on the presence and potential impacts of mutations in stomach cancer. They also identified COL5A2 and COL15A1 associated with longer OS in STAD but did not consider combinations or accounting for MSI status. This study reports the presence of collagen mutations in both high and low mutation burden tumors

and their differential impacts. Unlike Izzi et al., in the case of collagens in STAD, we did not observe a correlation between RNA expression levels and mutations. This is likely because of the contribution to collagen expression from myriad cell types in tumors. We also leverage the vast knowledge from germline mutations to interpret the tumor somatic mutations. Together, the report from Izzi et al. and this report, using different methods to account for mutation burden, provide data emphasizing the importance of somatic mutations in ECM components.

We hypothesized that comparing somatic mutations in tumors to germline mutations in collagenopathies would be informative and aide interpretation of the somatic mutations including the recessive or dominant nature of the mutation and the mechanisms that the mutation impacts collagen structure, subsequent tumor hallmarks and response to treatment. For example, COL12A1 was the collagen gene with the highest truncation mutation frequency and most often strongly associated with shorter OS in the combinatorics analysis (Fig. 2C). COL12A1 is a FACIT homotrimer collagen that mediates interactions between collagen type I and the rest of the ECM. COL12A1 is reported to be expressed by both stroma and tumor cells in STAD [57–60] and is expressed in gastric cancer cell lines [61]. Germline mutations, including rare splicing mutation truncation variants, in COL12A1 cause a Ehlers-Danlos/Bethlem-like myopathy syndrome [62, 63]. Together, these observations suggest that COL12A1 is a critical determinant of STAD disease progression and therapy response and exemplifies how comparison of somatic and germline mutations aids interpretation and provides new insights into tumors.

There are two general patterns of collagenopathy mutations also observed in STAD: mutations that disrupt the collagen triple helix such as Glycine mutations and mutations outside the protein domains known to cause reduced expression of the protein including both missense and truncation variants. These are the major mechanisms that cause multiple collagenopathies. As highlighted, these types of mutations are observed in both COL7A1 and COL12A1 (Fig. S12). We further highlight mutations in 3 types of collagens frequently mutated in STAD and linked with collagenopathies: collagen types I, IV, and V (Fig. S12). For example, many pathological variants of collagen type V, associated with Ehlers-Danlos Syndromes, are non-functional [64]. For collagen type I, truncation variants lead to haploinsufficiency because of lower expression in recessive disease while structural mutations including glycine variants impair collagen function and have a dominant negative effect [65]. These observations in collagenopathy variants suggest that somatic variants likely have similar impacts. For collagen type IV, another emerging theme is that missense variants

of COL4A1 and COL4A2, accumulate in cells inducing ER stress responses [66]. In all these diseases, similar to the case with COL7A1, the specific impacts of Gly missense variants are variable. These observations all suggest that the similar landscape between pathological germline and somatic variants observed in STAD provide guidance on interpretation. Loss of expression leads to disease and disrupted ECMs which is consistent with the strong association with OS for collagen truncation variants in STAD (Fig. 2). Larger cohorts for association studies along with mechanistic studies may further elucidate the impact of the missense somatic collagen variants.

Collagens form two of the major structures in the ECM: the basement membrane and the interstitial ECM. Both of these ECM components are impacted by collagen mutations. The loss of integrity of the basement membrane in tumors suggests a disorganized, more porous structure that could cause increased inflammation, analogous to COL7A1 mutations in DEB, or with collagen type IV and type VI variants [16]. Collagen type I, expressed by the COL1A1 and COL1A2 genes, plays a critical role in forming the ECM and organizing cell-cell interactions and mechanical properties in tumors [67]. Increased collagen type I has been associated with worse outcomes in many cancers including stomach [13, 68]. Both COL1A1 and COL1A2 had modest mutation frequencies with only weak association with OS in STAD (Fig. 1; Fig. S2). However, truncation mutations of COL1A1 were associated with longer survival and most truncation mutations of COL1A2 were also associated with longer survival except for 1 case, TCGA-HU-A4GQ-01, which was reported to have deceased at 0 months, and therefore may be reflecting other causes of death. Collagen type I missense variants were not associated with OS, perhaps because the majority of collagen type I originates from the stroma and therefore any impact of a mutated collagen type I originating from tumor cells may be diluted. Collagens that interact with collagen type I and regulate fiber size and structure including all 3 collagen type V genes, COL11A1, COL12A1, and COL14A1 have significant mutation rates and association with OS (Fig. 2). These observations further support the concept that loss of collagen type I from the tumor cells, or dysregulation of the network that forms collagen type I dependent structures affects patient outcomes.

Altogether, these observations suggest the regulation of collagen type I and the basement membrane by a panoply of cancer cell secreted collagens are critical for tumor fate. Collagens, as one of the dominant structural proteins in the ECM play myriad functions in regulating cancer hallmarks [69]. Minor collagens such as COL7A1 and COL12A1 form structural links between the collagen type I fiber network and/or the basement membrane zone. These data support a model where a local ECM

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 16 of 20

Fig. 6 Model of impact of cancer cell secreted collagens on tumors. Collagens originate from either the cancer or stroma cells. Truncation and missense collagen mutants reorganize the tumor microenvironment decreasing multiple processes that increase drug sensitivity and reduce metastasis risk including reduced EMT, less local collagen around the cancer cells, a more disorganized collagen structure, and increased infiltration of cytotoxic immune cells and drugs

derived from components secreted from the cancer cells, reshape the local ECM and are critical for tumor phenotypes, including EMT, drug response, the immunoenvironment and overall disease progression (Fig. 6). In tumors with wild-type collagens, EMT is higher, collagen type I fibers are wider, and higher expression of the matrisome including the basement membrane compared to tumors with mutant collagens. On the other hand, some mutant MSIH tumors, associated with shorter OS, exemplified by missense COL5A3 and COL14A1 which regulate COL1A1, have higher expression of mesenchymal genes, wider collagen type I fibers, and a different immune cell infiltration pattern (Fig. 4D). Linking hallmarks and pathways to dysregulated ECM caused by collagen mutations may lead to new opportunities to refine drug targeting and development.

Conclusions

In conclusion, we find a high frequency of individual collagen genes and sets of collagen genes harboring somatic mutations compared to background in both microsatellite stable (MSS), and microsatellite instable (MSIH) stomach adenocarcinomas in TCGA and comparable datasets. Overall, combinations of somatic mutations are predictive of patient survival,

and truncation mutations associate with improved survival. We further associate these combinations with distinctive tumor microenvironments based on lower matrisome expression, cell cycle and EMT, as well as immune cell infiltration. Interestingly, stomach cells express COL7A1, normally associated with skin ECM, and somatic mutations in COL7A1 predict improved overall survival. It should be noted that this study is limited by the dependence on genotype-phenotype correlations in patients so that there is risk in oversimplifying rare mutations. Some of this risk is ameliorated by the combinatorial approach and the interpretation of the mutations based on similar variants observed in collagenopathies. Nevertheless, many of these missense mutations resemble the loss of function (LOF) mutations in collagenopathies, which give some insight into their potential role in tumor progression. Overall, this study suggests the further testing of collagen mutations in stomach cancer is promising and collagen mutations could be incorporated into strategies to classify cancer patients.

Research highlights

Collagen mutations are prevalent in stomach cancer.

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 17 of 20

Collagen somatic missense mutations resemble collagenopathy mutations.

Collagen mutations associate with overall survival in stomach cancer.

Tumors with collagen mutations have distinct molecular pathways and tumor microenvironments.

Abbreviations

ACRG: Asian Cancer Research Group; CIN: Chromosome Instability; DEB: Dystrophic Epidermolysis Bullosa; EBV: Epstein-Barr Virus; ECM: Extracellular Matrix; EMT: Epithelial mesenchymal transition; FACIT: Fibril Associated Collagens with Interrupted Triple helices; GDC: Genomic Data Commons; GSEA : Gene Set Enrichment Analysis; GS: Genomically Stable; HM: High Mutation; MSIH: Microsatellite Instable; MSS: Microsatellite stable; OS: Overall Survival; RSEM: RNA-seq by Expectation Maximum; STAD: Stomach Adenocarcinoma; TCGA: The Cancer Genome Atlas.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-021-09136-1.

Additional file 1 : **Table S1.** Collagen family for collagens of interest in stomach cancer.

Additional file 2 : **Table S2.** MutSig 2CV v3.1 analysis of significantly mutated collagen genes in STAD TCGA cohort. Data downloaded from Firebrowse.

Additional file 3 : **Table S3.** Average expression level of each collagen gene in the STAD TCGA cohort. Values are RSEM.

Additional file 4 : **Table S4.** Collagen gene combinations identified from combinatorics approach.

Additional file 5 : **Table S5.** Summary of COL7A1 protein expression in stomach tumors from Rhode Island Hospital as assessed by immunohistochemical staining.

Additional file 6 : **Figure S1**. Alteration frequencies of collagens in ACRG and HK/Pfizer datasets. **A.** Alteration frequencies of sequenced collagens in other stomach cancer cohorts. **B.** Kaplan-Meier analysis of COl11A1, COL5, COL4, COL6 mutations in the ACRG targeted sequencing dataset compared to the same set of collagen genes in the TCGA cohort. **Figure S2.** Survival analysis of somatic mutations in each collagen gene. **A.** Kaplan Maier analysis of tumors with any type of mutation in each collagen gene across the whole STAD TCGA cohort. Tumors with the designated collagen mutation are in red. Wild-type tumors are in blue. *P*-values determined by log-rank test. **B.** Kaplan Maier analysis of tumors with truncation mutations in each collagen gene across the whole STAD TCGA cohort. **C.** Kaplan Maier analysis of tumors with any type of mutation in each collagen gene in MSIH cases. **D.** Kaplan Maier analysis of tumors with truncation mutation in each collagen gene in MSIH cases. **Figure S3.** Identification of combinations of collagens genes associated with overall survival relative to background. **A.** All mutations across the whole TCGA cohort. **B.** Representative examples of combinations of 2 collagens associated with overall survival. **C.** Combinations of all mutations in MSS tumors only. **D.** Combinations of all mutations in MSIH tumors only. **E.** Example of collagen genes with truncation mutations most frequently associated with overall survival when combined, classify MSIH tumors into high and low overall survival risk. **Figure S4.** Collagen mutations have MSIH and MSS context dependent differences in overall survival. Mutations in COL5A3 and COL14A1 have different associations in MSIH and MSS tumors even though the total number of mutations is similar. **Figure S5.** MSIH and MSS tumors have distinct microenvironments in TCGA. **A.** MSI status was associated with outcome in ACRG but not in TCGA. **B.** Comparison of MSIH and MSS stomach tumors by

pre-ranked GSEA reveals differences in expression. Each heatmap plots the Normalized Enrichment Scores (NES) from the GSEA. NABA ECM gene sets were expressed higher in MSS tumors compared to MSIH tumors. Many immune cell expression signatures including cytotoxic cells were expressed higher in MSIH tumors compared to MSS tumors. B cells were expressed higher in MSS tumors. The majority of cancer hallmark expression signatures were expressed significantly higher in MSIH tumors compared to MSS tumors. **Figure S6.** Pre-ranked GSEA of collagen mutation combinations in Table S4 for the whole TCGA STAD cohort shows consistent impact for each mutation combination. **A.** Hallmarks for combinations with both missense and truncation mutations. **B.** The NABA and immune signature genes sets for combinations with both missense and truncation mutations. **C.** Hallmark, NABA, and immune signature gene sets for combinations with just truncation mutations. **D.** Clustering of hallmark gene sets for tumors with missense mutations only in the whole TCGA cohort showed significant difference for the EMT hallmark relative to overall survival. *P*-value calculated by Kolmogorov-Smirnov. **Figure S7.** In MSS cases, pre-ranked GSEA of tumors with either a missense or truncation mutation combination as listed in Table S4 show impact of collagen mutations on pathways some of which are correlated with overall survival. **A.** For all mutations in MSS cases, some hallmarks such as EMT were associated with overall survival as shown in the heat map and box plot. P-value calculated by Kolmogorov-Smirnov. **B.** NABA ECM and immune signature gene sets in MSS tumors. Basement membrane and macrophage signature gene sets were among the gene sets most associated with overall survival, showing consistent downregulation in tumors with mutant collagens and higher expression in wild-type tumors. P-value calculated by Kolmogorov-Smirnov. **Figure S8.** In MSIH cases, pre-ranked GSEA of tumors with either a missense or truncation mutation combination as listed in Table S4 show impact of collagen mutations on pathways some of which are correlated with overall survival. **A.** Clustering of hallmark gene sets partitions tumors with collagen combinations by overall survival. Box plot shows the significant difference in the EMT hallmark as defined by combinations associated with high or low risk of overall survival. **B.** NABA gene sets showing large differences in Basement Membrane and ECM Affiliated gene sets relative to overall survival. **C.** Immune cell signature gene sets showing large difference in Tregs and Macrophage expression signatures. P-value calculated by Kolmogorov-Smirnov. **Figure S9**. In MSIH cases, pre-ranked GSEA of tumors with only missense mutation combinations as listed in Table S4 show impact of collagen mutations on pathways some of which are correlated with overall survival. **A.** Hallmark gene sets. **B.** NABA ECM sets. **C.** Immune cell gene signatures. P-value calculated by Kolmogorov-Smirnov. **Figure S10**. In MSIH cases, pre-ranked GSEA of tumors with only truncation mutation combinations as listed in Table S4 show impact of collagen mutations on pathways some of which are correlated with overall survival. **A.** Hallmark gene sets. **B.** NABA ECM and immune cell signature gene sets. P-value calculated by Kolmogorov-Smirnov. **Figure S11.** COL7A1 is expressed in some tumor cells in STAD. Representative images of COL7A1 protein and RNA expression in stomach adenocarcinoma. **A.** Immunohistochemistry (A, C, E) and in situ hybridization (B, D, F) for COL7. Stromal localization in C, E, D, and F, and mixed stromal and carcinoma localization (at white arrows; A, B). **B.** Higher magnification of panels A and B from S7A showing expression by IHC in panel A and ISH in panel B of COL7A1 in epithelial regions. The arrow shows ISH signal in tumor cells. **C.** Representative images at higher power of COL7A1 protein expression by IHC in the epithelium and stroma. **D.** Representative image of COL7A1 protein expression in normal human skin. Note the line of expression in the ECM between the dermal and epidermal layers (red arrow). Cells expressing COL7A1 show cytoplasmic signal as they are overexpressing COL7A1 to be secreted to form the "anchorage" line. **Figure S12.** Comparison of pathological germline and somatic mutations in STAD in three collagens. Comparison of the distribution of mutations across each gene and a lollipop plot mapping the somatic mutations to the protein domains. **A.** COL1A1 and COL1A2. **B.** COL4A1 and COL4A2. **C.** COL5A1 and COL5A2.

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 18 of 20

Declarations

Ethics approval and consent to participate
Use of patient material was approved by the Lifespan institutional review board approval, IRB #1070389–9. All procedures were performed in accordance with the relevant guidelines and regulations.

Consent for publication
Not applicable.

Competing interests
All the authors declare that we have no competing interests.

Author details
[1]Department of Pathology and Laboratory Medicine, Rhode Island Hospital, Warren Alpert Medical School at Brown University, Providence, RI 02903, USA. [2]Center for Computational Molecular Biology, Brown University, Providence, RI 02903, USA. [3]Joint Program in Cancer Biology, Brown University and Lifespan Cancer Institute, Providence, RI 02912, USA. [4]School of Engineering, Center for Biomedical Engineering, Brown University, Providence, RI 02912, USA. [5]Currently at PathAI, 1325 Boylston St, Boston, MA 02215, USA.

References
1. Egeblad M, Nakasone ES, Werb Z. Tumors as organs: complex tissues that interface with the entire organism. Dev Cell. 2010;18(6):884–901. https://doi.org/10.1016/j.devcel.2010.05.012 Epub 2010/07/16. PubMed PMID: 20627072; PMCID: PMC2905377.
2. Nissen NI, Karsdal M, Willumsen N. Collagens and cancer associated fibroblasts in the reactive stroma and its relation to cancer biology. J Exp Clin Cancer Res. 2019;38(1):115. https://doi.org/10.1186/s13046-019-1110-6 Epub 2019/03/08. PubMed PMID: 30841909; PMCID: PMC6404286.
3. Becht E, de Reynies A, Giraldo NA, Pilati C, Buttard B, Lacroix L, et al. Immune and stromal classification of colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. Clin Cancer Res. 2016;22(16):4057–66. https://doi.org/10.1158/1078-0432.CCR-15-2879 PubMed PMID: 26994146.
4. Fang M, Yuan J, Peng C, Li Y. Collagen as a double-edged sword in tumor progression. Tumour Biol. 2014;35(4):2871–82. https://doi.org/10.1007/s13277-013-1511-7 PubMed PMID: 24338768; PMCID: PMC3980040.
5. Ricard-Blum S. The collagen family. Cold Spring Harb Perspect Biol. 2011;3(1):a004978. https://doi.org/10.1101/cshperspect.a004978 PubMed PMID: 21421911; PMCID: PMC3003457.
6. Brodsky AS, Xiong J, Yang D, Schorl C, Fenton MA, Graves TA, et al. Identification of stromal ColXalpha1 and tumor-infiltrating lymphocytes as putative predictive markers of neoadjuvant therapy in estrogen receptor-positive/HER2-positive breast cancer. BMC Cancer. 2016;16(1):274. https://doi.org/10.1186/s12885-016-2302-5 PubMed PMID: 27090210; PMCID: PMC4835834.
7. Jia D, Liu Z, Deng N, Tan TZ, Huang RY, Taylor-Harding B, et al. A COL11A1-correlated pan-cancer gene signature of activated fibroblasts for the prioritization of therapeutic targets. Cancer Lett. 2016;382(2):203–14. https://doi.org/10.1016/j.canlet.2016.09.001 PubMed PMID: 27609069.
8. Soini Y, Hurskainen T, Hoyhtya M, Oikarinen A, Autio-Harmainen H. 72 KD and 92 KD type IV collagenase, type IV collagen, and laminin mRNAs in breast cancer: a study by in situ hybridization. J Histochem Cytochem. 1994;42(7):945–51. https://doi.org/10.1177/42.7.8014478 Epub 1994/07/01. PubMed PMID: 8014478.
9. Li N, Sun H, Wang X, Zhang Z, Zhou Y, Anderson C, et al. Extracellular matrix gene expression and cytotoxic T lymphocyte infiltration in the tumor microenvironment in non-small cell lung cancer. In: AACR annual meeting 2019; 2019 mar 29-Apr 3 2019. Atlanta: AACR; 2019.
10. Tian C, Clauser KR, Ohlund D, Rickelt S, Huang Y, Gupta M, et al. Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. Proc Natl Acad Sci U S A. 2019. https://doi.org/10.1073/pnas.1908626116 Epub 2019/09/06. PubMed PMID: 31484774.
11. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 2019;144(8):1941–53. https://doi.org/10.1002/ijc.31937 Epub 2018/10/24. PubMed PMID: 30350310.
12. Song Z, Wu Y, Yang J, Yang D, Fang X. Progress in the treatment of advanced gastric cancer. Tumour Biol. 2017;39(7):1010428317714626. https://doi.org/10.1177/1010428317714626 PubMed PMID: 28671042.
13. Zhou ZH, Ji CD, Xiao HL, Zhao HB, Cui YH, Bian XW. Reorganized collagen in the tumor microenvironment of gastric cancer and its association with prognosis. J Cancer. 2017;8(8):1466–76. https://doi.org/10.7150/jca.18466 PubMed PMID: 28638462; PMCID: PMC5479253.
14. Chen D, Chen G, Jiang W, Fu M, Liu W, Sui J, et al. Association of the Collagen Signature in the tumor microenvironment with lymph node metastasis in early gastric cancer. JAMA Surg. 2019:e185249. https://doi.org/10.1001/jamasurg.2018.5249 Epub 2019/01/31. PubMed PMID: 30698615.
15. Zankl A, Neumann L, Ignatius J, Nikkels P, Schrander-Stumpel C, Mortier G, et al. Dominant negative mutations in the C-propeptide of COL2A1 cause platyspondylic lethal skeletal dysplasia, torrance type, and define a novel subfamily within the type 2 collagenopathies. Am J Med Genet A. 2005;133A(1):61–7. https://doi.org/10.1002/ajmg.a.30531 PubMed PMID: 15643621.
16. Jobling R, D'Souza R, Baker N, Lara-Corrales I, Mendoza-Londono R, Dupuis L, et al. The collagenopathies: review of clinical phenotypes and molecular correlations. Curr Rheumatol Rep. 2014;16(1):394. https://doi.org/10.1007/s11926-013-0394-3 PubMed PMID: 24338780.
17. Christiano AM, Ryynanen M, Uitto J. Dominant dystrophic epidermolysis bullosa: identification of a Gly-->Ser substitution in the triple-helical domain of type VII collagen. Proc Natl Acad Sci U S A. 1994;91(9):3549–53 PubMed PMID: 8170945; PMCID: PMC43617.
18. Kuivaniemi H, Tromp G, Prockop DJ. Mutations in collagen genes: causes of rare and some common diseases in humans. FASEB J. 1991;5(7):2052–60 PubMed PMID: 2010058.
19. Spranger J, Winterpacht A, Zabel B. The type II collagenopathies: a spectrum of chondrodysplasias. Eur J Pediatr. 1994;153(2):56–65 PubMed PMID: 8157027.
20. Vikkula M, Metsaranta M, Ala-Kokko L. Type II collagen mutations in rare and common cartilage diseases. Ann Med. 1994;26(2):107–14 PubMed PMID: 8024727.

Brodsky *et al. BMC Cancer*   (2022) 22:139

Page 19 of 20

21. Tarpey PS, Behjati S, Cooke SL, Van Loo P, Wedge DC, Pillay N, et al. Frequent mutation of the major cartilage collagen gene COL2A1 in chondrosarcoma. Nat Genet. 2013;45(8):923–6. https://doi.org/10.1038/ng.2668 PubMed PMID: 23770606; PMCID: PMC3743157.

22. Kim E, Ilic N, Shrestha Y, Zou L, Kamburov A, Zhu C, et al. Systematic functional interrogation of rare cancer variants identifies oncogenic alleles. Cancer Discov. 2016;6(7):714–26. https://doi.org/10.1158/2159-8290.CD-16-0160 PubMed PMID: 27147599; PMCID: PMC4930723.

23. Li X, Wu WK, Xing R, Wong SH, Liu Y, Fang X, et al. Distinct subtypes of gastric cancer defined by molecular characterization include novel mutational signatures with prognostic capability. Cancer Res. 2016;76(7):1724–32. https://doi.org/10.1158/0008-5472.CAN-15-2443 PubMed PMID: 26857262.

24. Leiserson MD, Vandin F, Wu HT, Dobson JR, Eldridge JV, Thomas JL, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat Genet. 2015;47(2):106–14. https://doi.org/10.1038/ng.3168 PubMed PMID: 25501392; PMCID: 4444046.

25. Zhou B, Wang GZ, Wen ZS, Zhou YC, Huang YC, Chen Y, et al. Somatic mutations and splicing variants of focal adhesion kinase in non-small cell lung cancer. J Natl Cancer Inst. 2018;110(2). https://doi.org/10.1093/jnci/djx157 PubMed PMID: 29087503.

26. 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 Syst. 2018;7(4):422–37 e7. https://doi.org/10.1016/j.cels.2018.08.010 Epub 2018/10/01. PubMed PMID: 30268436.

27. Olcina MM, Balanis NG, Kim RK, Aksoy BA, Kodysh J, Thompson MJ, et al. Mutations in an innate immunity pathway are associated with poor overall survival outcomes and hypoxic signaling in cancer. Cell Rep. 2018;25(13):3721–32 e6. https://doi.org/10.1016/j.celrep.2018.11.093 Epub 2018/12/28. PubMed PMID: 30590044.

28. Izzi V, Davis MN, Naba A. Pan-cancer analysis of the genomic alterations and mutations of the matrisome. Cancers. 2020;12(8):2046. https://doi.org/10.3390/cancers12082046 PMID - 32722287.

29. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, et al. Scalable Open Science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst. 2018;6(3):271–81 e7. https://doi.org/10.1016/j.cels.2018.03.002 Epub 2018/03/30. PubMed PMID: 29596782; PMCID: PMC6075717.

30. Tamborero D, Rubio-Perez C, Muinos F, Sabarinathan R, Piulats JM, Muntasell A, et al. A Pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations. Clin Cancer Res. 2018;24(15):3717–28. https://doi.org/10.1158/1078-0432.CCR-17-3509 Epub 2018/04/19. PubMed PMID: 29666300.

31. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. Immunity. 2018;48(4):812–30 e14. https://doi.org/10.1016/j.immuni.2018.03.023 Epub 2018/04/10. PubMed PMID: 29628290; PMCID: PMC5982584.

32. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417–25. https://doi.org/10.1016/j.cels.2015.12.004 Epub 2016/01/16. PubMed PMID: 26771021; PMCID: PMC4707969.

33. 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. Mol Cell Proteomics. 2012;11(4):M111 014647. https://doi.org/10.1074/mcp.M111.014647 Epub 2011/12/14. PubMed PMID: 22159717; PMCID: PMC3322572.

34. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–50. https://doi.org/10.1073/pnas.0506580102 PubMed PMID: 16199517; PMCID: 1239896.

35. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21(5):449–56. https://doi.org/10.1038/nm.3850 PubMed PMID: 25894828.

36. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat. 2011;32(5):557–63. https://doi.org/10.1002/humu.21438 Epub 2011/04/27. PubMed PMID: 21520333.

37. Wertheim-Tysarowska K, Sobczynska-Tomaszewska A, Kowalewski C, Skronski M, Swieckowski G, Kutkowska-Kazmierczak A, et al. The COL7A1 mutation database. Hum Mutat. 2012;33(2):327–31. https://doi.org/10.1002/humu.21651 Epub 2011/11/08. PubMed PMID: 22058051.

38. Morpheus. Available from: https://software.broadinstitute.org/morpheus.

39. Kuhl T, Mezger M, Hausser I, Handgretinger R, Bruckner-Tuderman L, Nystrom A. High local concentrations of intradermal MSCs restore skin integrity and facilitate wound healing in dystrophic epidermolysis bullosa. Mol Ther. 2015;23(8):1368–79. https://doi.org/10.1038/mt.2015.58 Epub 2015/04/11. PubMed PMID: 25858020; PMCID: PMC4817872.

40. Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. Nat Genet. 2014;46(6):573–82. https://doi.org/10.1038/ng.2983 PubMed PMID: 24816253.

41. Ratti M, Lampis A, Hahne JC, Passalacqua R, Valeri N. Microsatellite instability in gastric cancer: molecular bases, clinical perspectives, and new treatment approaches. Cell Mol Life Sci. 2018;75(22):4151–62. https://doi.org/10.1007/s00018-018-2906-9 Epub 2018/09/03. PubMed PMID: 30173350; PMCID: PMC6182336.

42. Ansorge HL, Meng X, Zhang G, Veit G, Sun M, Klement JF, et al. Type XIV collagen regulates Fibrillogenesis: premature collagen fibril growth and tissue dysfunction in null mice. J Biol Chem. 2009;284(13):8427–38. https://doi.org/10.1074/jbc.M805582200 Epub 2009/01/13. PubMed PMID: 19136672; PMCID: PMC2659201.

43. Young BB, Zhang G, Koch M, Birk DE. The roles of types XII and XIV collagen in fibrillogenesis and matrix assembly in the developing cornea. J Cell Biochem. 2002;87(2):208–20. https://doi.org/10.1002/jcb.10290 Epub 2002/09/24. PubMed PMID: 12244573.

44. Tao G, Levay AK, Peacock JD, Huk DJ, Both SN, Purcell NH, et al. Collagen XIV is important for growth and structural integrity of the myocardium. J Mol Cell Cardiol. 2012;53(5):626–38. https://doi.org/10.1016/j.yjmcc.2012.08.002 Epub 2012/08/22. PubMed PMID: 22906538; PMCID: PMC3472103.

45. Mak KM, Png CY, Lee DJ. Type V collagen in health, disease, and fibrosis. Anat Rec (Hoboken). 2016;299(5):613–29. https://doi.org/10.1002/ar.23330 PubMed PMID: 26910848.

46. Grassel S, Bauer RJ. Collagen XVI in health and disease. Matrix Biol. 2013;32(2):64–73. https://doi.org/10.1016/j.matbio.2012.11.001 Epub 2012/11/15. PubMed PMID: 23149016.

47. Jung HY, Fattet L, Yang J. Molecular pathways: linking tumor microenvironment to epithelial-mesenchymal transition in metastasis. Clin Cancer Res. 2015;21(5):962–8. https://doi.org/10.1158/1078-0432.CCR-13-3173 Epub 2014/08/12. PubMed PMID: 25107915; PMCID: PMC4320988.

48. Martins Cavaco AC, Damaso S, Casimiro S, Costa L. Collagen biology making inroads into prognosis and treatment of cancer progression and metastasis. Cancer Metastasis Rev. 2020;39(3):603–23. https://doi.org/10.1007/s10555-020-09888-5 Epub 2020/05/25. PubMed PMID: 32447477.

49. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. Nat Rev Mol Cell Biol. 2014;15(12):786–801. https://doi.org/10.1038/nrm3904.

50. Lv YP, Peng LS, Wang QH, Chen N, Teng YS, Wang TT, et al. Degranulation of mast cells induced by gastric cancer-derived adrenomedullin prompts gastric cancer progression. Cell Death Dis. 2018;9(10):1034. https://doi.org/10.1038/s41419-018-1100-1 Epub 2018/10/12. PubMed PMID: 30305610; PMCID: PMC6180028.

51. Hiramatsu S, Tanaka H, Nishimura J, Sakimura C, Tamura T, Toyokawa T, et al. Neutrophils in primary gastric tumors are correlated with neutrophil infiltration in tumor-draining lymph nodes and the systemic inflammatory response. BMC Immunol. 2018;19(1):13. https://doi.org/10.1186/s12865-018-0251-2 Epub 2018/04/18. PubMed PMID: 29661142; PMCID: PMC5902874.

52. Has C, Nystrom A, Saeidian AH, Bruckner-Tuderman L, Uitto J. Epidermolysis bullosa: molecular pathology of connective tissue components in the cutaneous basement membrane zone. Matrix Biol. 2018;71-72:313–29. https://doi.org/10.1016/j.matbio.2018.04.001 Epub 2018/04/09. PubMed PMID: 29627521.

53. Chung HJ, Uitto J. Type VII collagen: the anchoring fibril protein at fault in dystrophic epidermolysis bullosa. Dermatol Clin. 2010;28(1):93–105. https://doi.org/10.1016/j.det.2009.10.011 Epub 2009/12/01. PubMed PMID: 19945621; PMCID: PMC2791403.

54. Varki R, Sadowski S, Uitto J, Pfendner E. Epidermolysis bullosa. II. Type VII collagen mutations and phenotype-genotype correlations in the

Brodsky *et al. BMC Cancer*   *(2022) 22:139*

Page 20 of 20

dystrophic subtypes. J Med Genet. 2007;44(3):181–92. https://doi.org/10.1136/jmg.2006.045302 Epub 2006/09/15. PubMed PMID: 16971478; PMCID: PMC2598021.

55. Bornert O, Nystrom A. Cloning and mutagenesis strategies for large collagens. Methods Mol Biol. 1944;2019:3–15. https://doi.org/10.1007/978-1-4939-9095-5_1 Epub 2019/03/07. PubMed PMID: 30840231.

56. Tian C, Ohlund D, Rickelt S, Lidstrom T, Huang Y, Hao L, et al. Cancer-cell-derived matrisome proteins promote metastasis in pancreatic ductal adenocarcinoma. Cancer Res. 2020. https://doi.org/10.1158/0008-5472.CAN-19-2578 Epub 2020/02/08. PubMed PMID: 32029550.

57. Duan S, Gong B, Wang P, Huang H, Luo L, Liu F. Novel prognostic biomarkers of gastric cancer based on gene expression microarray: COL12A1, GSTA3, FGA and FGG. Mol Med Rep. 2018;18(4):3727–36. https://doi.org/10.3892/mmr.2018.9368 Epub 2018/08/15. PubMed PMID: 30106150; PMCID: PMC6131538.

58. Jiang X, Wu M, Xu X, Zhang L, Huang Y, Xu Z, et al. COL12A1, a novel potential prognostic factor and therapeutic target in gastric cancer. Mol Med Rep. 2019. https://doi.org/10.3892/mmr.2019.10548 Epub 2019/08/23. PubMed PMID: 31432110.

59. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. Science. 2017;357(6352). https://doi.org/10.1126/science.aan2507 Epub 2017/08/19. PubMed PMID: 28818916.

60. Human Protein Atlas available from http://www.proteinatlas.org.

61. Xiang Z, Li J, Song S, Wang J, Cai W, Hu W, et al. A positive feedback between IDO1 metabolite and COL12A1 via MAPK pathway to promote gastric cancer metastasis. J Exp Clin Cancer Res. 2019;38(1):314. https://doi.org/10.1186/s13046-019-1318-5 Epub 2019/07/19. PubMed PMID: 31315643; PMCID: PMC6637527.

62. Hicks D, Farsani GT, Laval S, Collins J, Sarkozy A, Martoni E, et al. Mutations in the collagen XII gene define a new form of extracellular matrix-related myopathy. Hum Mol Genet. 2014;23(9):2353–63. https://doi.org/10.1093/hmg/ddt637 Epub 2013/12/18. PubMed PMID: 24334769.

63. Malfait F, Francomano C, Byers P, Belmont J, Berglund B, Black J, et al. The 2017 international classification of the Ehlers-Danlos syndromes. Am J Med Genet C Semin Med Genet. 2017;175(1):8–26. https://doi.org/10.1002/ajmg.c.31552 Epub 2017/03/18. PubMed PMID: 28306229.

64. Malfait F, De Paepe A. Molecular genetics in classic Ehlers-Danlos syndrome. Am J Med Genet C Semin Med Genet. 2005;139C(1):17–23. https://doi.org/10.1002/ajmg.c.30070 Epub 2005/11/10. PubMed PMID: 16278879.

65. Ben Amor IM, Glorieux FH, Rauch F. Genotype-phenotype correlations in autosomal dominant osteogenesis imperfecta. J Osteoporos. 2011;2011:540178. https://doi.org/10.4061/2011/540178 PubMed PMID: 21912751; PMCID: PMC3170785.

66. Kuo DS, Labelle-Dumais C, Gould DB. COL4A1 and COL4A2 mutations and disease: insights into pathogenic mechanisms and potential therapeutic targets. Hum Mol Genet. 2012;21(R1):R97–110. https://doi.org/10.1093/hmg/dds346 PubMed PMID: 22914737; PMCID: PMC3459649.

67. Xu S, Xu H, Wang W, Li S, Li H, Li T, et al. The role of collagen in cancer: from bench to bedside. J Transl Med. 2019;17(1):309. https://doi.org/10.1186/s12967-019-2058-1 Epub 2019/09/16. PubMed PMID: 31521169.

68. Ohno S, Tachibana M, Fujii T, Ueda S, Kubota H, Nagasue N. Role of stromal collagen in immunomodulation and prognosis of advanced gastric carcinoma. Int J Cancer. 2002;97(6):770–4 PubMed PMID: 11857352.

69. 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 PubMed PMID: 25381661.

Publisher's Note