The Burden of Post-Translational Modification (PTM)—Disrupting Mutations in the Tumor Matrisome

Elisa Holstein [1,†], Annalena Dittmann [1,†], Anni Kääriäinen [1], Vilma Pesola [1], Jarkko Koivunen [1], Taina Pihlajaniemi [1], Alexandra Naba [2,3] and Valerio Izzi [1,4,5,*]

1   Faculty of Biochemistry and Molecular Medicine, University of Oulu, FI-90014 Oulu, Finland; elisa.holstein@gmx.at (E.H.); annalena.dittmann@oulu.fi (A.D.); anni.kaariainen@oulu.fi (A.K.); vilma.pesola@oulu.fi (V.P.); jarkko.koivunen@oulu.fi (J.K.); taina.pihlajaniemi@oulu.fi (T.P.)
2   Department of Physiology and Biophysics, University of Illinois at Chicago, Chicago, IL 60612, USA; anaba@uic.edu
3   University of Illinois Cancer Center, Chicago, IL 60612, USA
4   Faculty of Medicine, University of Oulu, FI-90014 Oulu, Finland
5   Finnish Cancer Institute, 00130 Helsinki, Finland
*   Correspondence: valerio.izzi@oulu.fi
†   These authors contributed equally to this work.

Simple Summary: Mutations are the driving force of the oncogenic process, altering regulatory pathways and leading to uncontrolled cell proliferation. Understanding the occurrence and patterns of mutations is necessary to identify the sequence of events enabling tumor growth and diffusion. Yet, while much is known about mutations in proteins whose actions are exerted inside the cells, much less is known about the extracellular matrix (ECM) and ECM-associated proteins (collectively known as the "matrisome") whose actions are exerted outside the cells. In particular, while post-translational modifications (PTMs) are critical for the functions of many proteins, both intracellular and in the matrisome, there are no studies evaluating the mutations impacting known PTM sites within matrisome proteins. Here we report on a large Pan-Cancer cohort spanning 32 tumor types and demonstrate the specificities of matrisome PTM-affecting mutations over the rest of the genome, also evidencing features and findings that might be relevant for prognostication and mechanistic understanding of the supportive role of the tumor microenvironment in the tumorigenic process.

Abstract: Background: To evaluate the occurrence of mutations affecting post-translational modification (PTM) sites in matrisome genes across different tumor types, in light of their genomic and functional contexts and in comparison with the rest of the genome. Methods: This study spans 9075 tumor samples and 32 tumor types from The Cancer Genome Atlas (TCGA) Pan-Cancer cohort and identifies 151,088 non-silent mutations in the coding regions of the matrisome, of which 1811 affecting known sites of hydroxylation, phosphorylation, N- and O-glycosylation, acetylation, ubiquitylation, sumoylation and methylation PTM. Results: PTM-disruptive mutations (PTM^{mut}) in the matrisome are less frequent than in the rest of the genome, seem independent of cell-of-origin patterns but show dependence on the nature of the matrisome protein affected and the background PTM types it generally harbors. Also, matrisome PTM^{mut} are often found among structural and functional protein regions and in proteins involved in homo- and heterotypic interactions, suggesting potential disruption of matrisome functions. Conclusions: Though quantitatively minoritarian in the spectrum of matrisome mutations, PTM^{mut} show distinctive features and damaging potential which might concur to deregulated structural, functional, and signaling networks in the tumor microenvironment.

Keywords: extracellular matrix (ECM); matrisome; mutations; post-translational modifications (PTM); pan-cancer

1. Introduction

Propelled by next-generation sequencing (NGS) techniques and the work of large, interdisciplinary consortia such as The Cancer Genome Atlas (TCGA), cancer research has made a gigantic leap forward in understanding the molecular details of tumorigenesis and cancer-causing mutations, aberrations, and pathways and in elucidating the events that lead to supporting and promoting tumor growth and metastatization [1].

While most of the cancer research effort has been devoted to the study of tumor cell-intrinsic processes, recent years have seen a resurgence of focus on the tumor microenvironment (TME) [2–6] and, in particular, on the tumor "matrisome" [7,8]. Being a framework defining structural component of the extra cellular matrix (ECM), including collagens, proteoglycans and various glycoproteins, and ECM-associated proteins, including ECM-remodeling enzymes, proteins structurally or functionally related to ECM components, as well as cytokines, chemokine, and growth factors [9,10], the matrisome represents the majority of structural and functional moieties of the acellular part of the TME and enables systems biology-level studies on its regulation and alteration in cancer [11–13]. This, in turn, enables understanding how the two sides of a tumor cell's plasma membrane "worlds" are connected, as all 10 hallmarks of cancers proposed by Weinberg and Hanahan [14] are directly controlled by signals coming from the ECM [13,15,16] and the TME reflects, at least partially, the genetic and molecular setup of a given tumor [17–19].

Protein post-translational modifications (PTM) encompass both reversible and irreversible modifications of proteins that follow synthesis and that are absolutely crucial for protein maturation and function [20–22]. For example, phosphorylation of a target protein by a kinase (and its subsequent activation or deactivation)—probably the most well-known of "message passing" among cell biologists—is a form of PTM [23]. Similarly, glycosylation—the most widespread form of PTM among eukaryotes [24,25]—is widely known for its crucial role in the functions and structures of the proteins, be it the triggering of a response in an effector cell after binding to the Fc portion of an antibody [23], the building of cellulose (a β-Linked homopolymer of glucose), one of the most abundant organic molecules on the planet [26], or the mediation of an enormous amount of biological signals at the physiological and the pathological level [26,27]. It is no surprise, then, that alterations in PTMs or enzymes mediating these modifications are linked to various diseases, ranging from neurodegenerative and skeletal diseases to cancer [28–30].

PTMs are all the more important for the ECM and the matrisome, conferring critical structural and functional features to the proteins they target [31,32]. For example, during or after collagen biosynthesis, proline and lysine residues are enzymatically converted to 3- and 4-hydroxyproline and 5-hydroxylysine, creating hydrogen bonds which stabilize collagen molecules and confer thermal stability to the nascent collagen fibrils as well as enable recognition by integrins and DDR receptors, at least as nucleation sites [33]. Similarly, glycosylation of collagens, fibronectins, laminins, and proteoglycans (and their eventual supramolecular assemblies) is crucial for recognition by, e.g., integrins and fundamental for cell adhesion and movement across a large span of physiological and developmental processes [34,35].

Altered PTM patterns are a hallmark of several diseases, including cancer [28–30]. Defective PTM is, in example, crucial in the deregulation of tumor-suppressor proteins such as p53, Rb, and PTEN, in the activation of oncogenic proteins such as the estrogen receptor α (ER-α) and the androgen receptor (AR), as well as in the transactivation of a plethora of tumorigenic pathways relying on chromatin remodeling and transcription factor activation [36–38]. Likewise, altered matrisome PTM participate to cancer and metastasis by affecting the regulation of stiffness within the TME, ablating or enabling ECM-receptor recognition, ectopically activating or inactivating functions within the matrisome proteins, enabling or suppressing the production of cryptic active domains (such as the matricryptins and matrikines) buried within larger ECM proteins and producing or hiding antigens and neoantigens whose role in cancer is still largely unknown [31,32,39].

In a recent pan-cancer analysis of mutations and copy number alterations (CNA), we evidenced several specificities of the matrisome with respect to the rest of the genome, such as a general higher frequency of mutations and alterations, the acquisition or the lack of mutations on a domain-specific basis and a robust association of specific mutations and alterations with patient survival [40]. Given the peculiar frequency of matrisome mutations in cancer [40,41] and the tremendous impact that PTM has on matrisome functions, and considering that no systematic analysis has yet produced a catalogue of disruptive mutations affecting experimentally validated PTM loci (PTM^{mut}) in the matrisome, we have devised a Pan-Cancer approach to enumerate and investigate PTM^{mut} across 32 different tumor types and more than 9000 patients.

Our results show that, unlike other mutations, PTM^{mut} are poorly tolerated by the matrisome and account for only approx. 1800 mutations in total. Despite their low counts, these mutations frequently affect the functional protein domains involved in message passing, homo- and heterotypic interactions, and catalytic, enzymatic, and structural roles, hinting at a significant impact on the overall structure and function of the ECM meshwork they reside in and providing foundations to further, targeted studies on the specific molecular disturbances they induce, their role in tumorigenesis and pathogenesis and their potential as therapeutic targets.

2. Materials and Methods

2.1. Source Data

TCGA Pan-Cancer Atlas mutation data were sourced from the University of California, Santa Cruz UCSC Xena Browser hub (http://xenabrowser.net/, TCGA Unified Ensemble "MC3" mutation calls, Xena identifier: mc3.v0.2.8.PUBLIC.xena). The following PTM were considered in the study: phosphorylation, acetylation, hydroxylation, N- and O-glycosylation, methylation, sumoylation, and ubiquitylation. Experimentally validated PTM site coordinates were downloaded from Uniprot (http://www.uniprot.org/), with the exception of phosphorylation sites sourced from Phosphosite database (https://www.phosphosite.org/homeAction.action) and N-glycosylation sites from the N-Glycosite-Atlas (http://glycositedb.biomarkercenter.org/accounts/login/). Protein regions, domains, and functions were obtained from Uniprot. The matrisome gene list was obtained from the matrisome project website (http://matrisomeproject.mit.edu/) and the conversion of matrisome gene names to protein IDs was performed in Uniprot. Gene length was obtained from the package "goseq" in R. Protein–protein interactions were sourced from BioGrid (http://thebiogrid.org/). All repositories were accessed between March and August 2020 and further finally confirmed on 20 September 2020. Mutations in the matrisome of healthy individuals from solid tissues, blood or bone marrow samples were sourced from the NCI-GDC Data Portal (https://portal.gdc.cancer.gov/) on 25 February 2021.

2.2. Statistical Analysis

All analyses were performed in The R Project for Statistical Computing (R). All data presented in the manuscript refer to non-silent, somatic mutations (SNP + INDELS) from the following tumor types: adrenocortical carcinoma (ACC), urothelial bladder carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endo-cervical adenocarcinoma (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), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), sarcoma (SARC), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), testicular germ cell tumors (TGCT), thyroid carcinoma (THCA), thymoma (THYM),

uterine corpus endometrial carcinoma (UCEC), and uterine carcinosarcoma (UCS). Uveal Melanoma (UVM) results are included in the first part of the manuscript, but later dropped as the few cases made their interpretation inconsistent. Quantitative categorical differences were tested using a two-sided Chi-square test, while quantitative numerical differences were tested using a two-sided Mann–Whitney U test, unless otherwise specified. Linear correlation (fitting) was tested using Pearson correlation. In all analyses, a *p* value < 0.05 was chosen as the threshold for reporting significant results. Detailed descriptions of data manipulation, filtering, analytical methods, and results are provided with the code (see the Data Availability Statement).

3. Results

3.1. Genomic Features of Overall Mutations and PTM-Affecting Mutations (PTM^{mut}) in the Tumor Matrisome and in the Rest of the Genome

Our analysis starts from approx. 2.3 million non-silent mutations in 9075 patients and 32 tumor types sourced from The Cancer Genome Atlas (TCGA) database [42,43]. Of these, the whole set of matrisome mutations is a sheer minority (151,088 out of 2,277,979; approx. 6.6%) (Figure 1A). Still, considering that the matrisome includes 1027 genes (out of 21,255; approx. 5% of the whole genome), this translates into a higher mean frequency of mutations/gene in the matrisome vs. rest of the genome (approx. 147 vs. 105) (Figure 1B), as already reported [40]. Further breakdown of total mutations by their expected effect shows, globally, no major differences between matrisome and rest of the genome irrespective of the algorithm used to predict the effect of the mutation (Figure 1C and Figure S1), again as already reported [40].

On this basis, we next focused on disruptive mutations affecting known PTM sites within proteins (PTM^{mut}), including typical ECM PTMs such as acetylation, hydroxylation, methylation, N- and O-glycosylation, phosphorylation, sumoylation, and ubiquitylation [31,32,39].

PTM^{mut} are a small portion of all the mutations, totaling 42,733 (approx. 1.88% of all mutations) from 6303 patients across 32 tumor types (Figure 1D and Table S1). Of these, only approx. 1811 are found in matrisome genes, thus setting the PTM^{mut}/all mutations ratio at 1.19% for the matrisome and 1.92% for the rest of the genome. Interestingly, the average ratio of PTM^{mut} normalized by gene length is also slightly smaller in the matrisome than in the rest of the genome (1.05×10^{-3} vs. 1.31×10^{-3}, respectively), in contrast with mutations not belonging to the PTM^{mut} group (non-PTM^{mut}) which are conversely more frequent (5.06×10^{-2} vs. 4.12×10^{-2}) [40]. At a more granular level, however, the comparison across different cancer types shows no significant deviation of the PTM^{mut}/length ratio from the total mutation/length ratio in any tumor (Figure S2A,B, respectively).

The total quantity of transitions and transversions in matrisome and rest of the genome is also similar for both PTM^{mut} and all mutations (Figure S3A,B, respectively) even breaking down by mutation effect (Figure S3C), marking no major differences in intracellular and extracellular proteomes. There is, however, a small and not statistically significant lack of transversions among PTM^{mut}, which derives mostly from splice-site mutations in ECM-affiliated proteins (Figure S3D and Table S1).

Owing to the differences in the number and types of PTM sites that exist already at the baseline level between matrisome and non-matrisome proteins (Figure S4A) and that might influence the observed distribution of PTM^{mut} (Figure S4B), we have devised a novel measure—the *burden*—that allows for a more robust comparison between the matrisome and the rest of the genome. This measure, in fact, quantifies the % of known PTM sites that, in each protein and for the different types of PTMs, are affected by PTM^{mut} per cancer (local burdens) and the results are then averaged across the Pan-Cancer cohort to obtain a global burden.
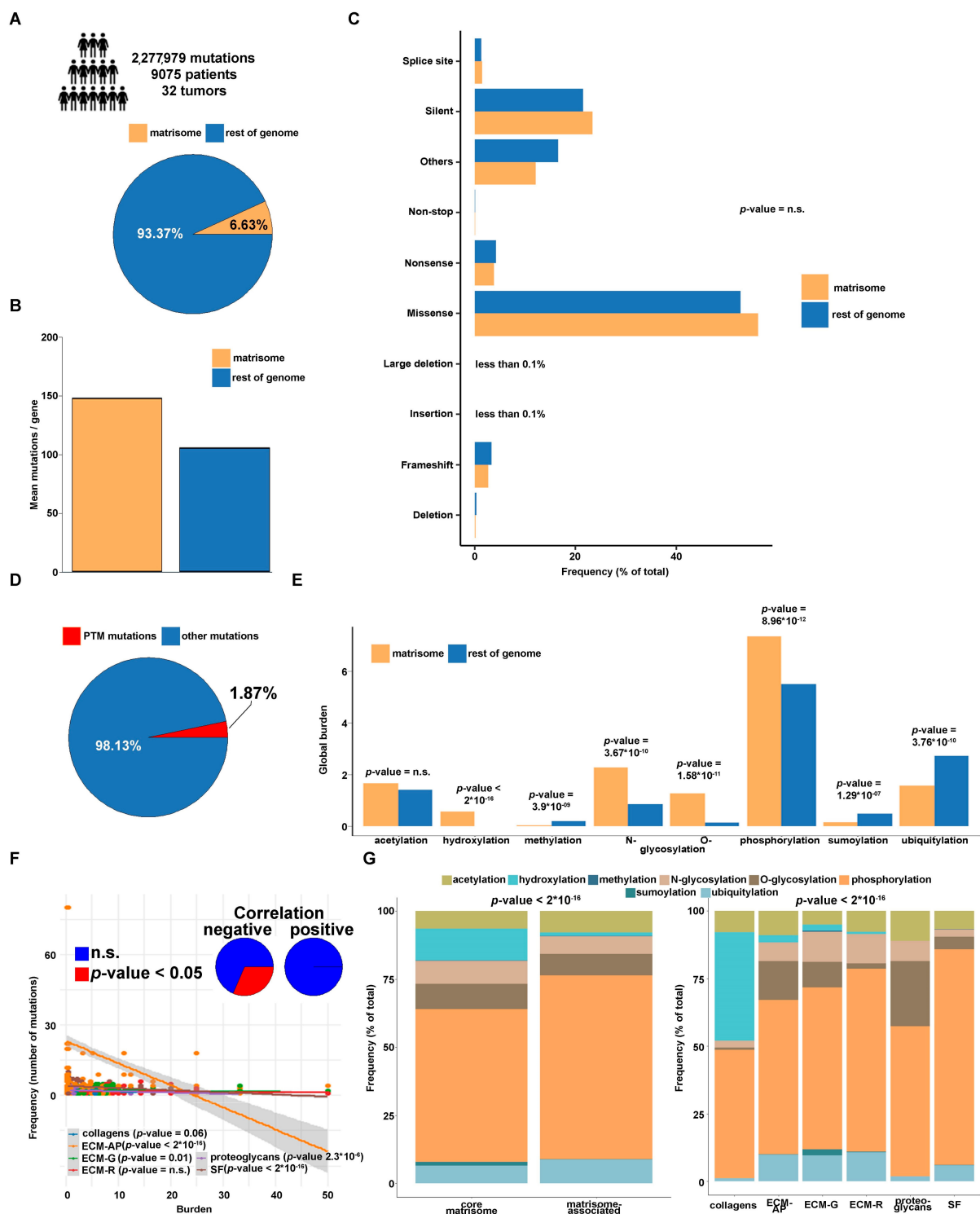
Figure 1. The genomic landscape of PTM-disrupting mutations (PTM^{mut}) in the tumor matrisome and in the rest of the genome. Dimensions of the analysis (**A**) and comparison between the tumor matrisome and rest of the tumor genome for overall mutation frequency (**B**) and type of mutation (**C**). PTM^{mut} are a minority of all matrisome mutations (**D**) with significant enrichments or depletions of PTM types affected (**E**). The amount of PTM^{mut} across tumors and divided by matrisome family (**F**) is negatively correlated with preservation of the PTM^{mut} (indicating poor preservation of a given PTM^{mut} in multiple patients (or cancer types) and (**G**) varies considerably across the different types of PTM investigated. Abbreviations: n.s., not significant; ECM-AP, ECM-affiliated proteins; ECM-G, ECM glycoproteins; ECM-R, ECM regulators; SF, secreted factors. *p*-values are from Chi-square tests.

Interestingly, the global burden of PTM^{mut} is equivalent for acetylation, while it is significantly higher in matrisome than in the rest of the genome for phosphorylation, N- and O-glycosylation (approx. 0.3, 1.6 and 8.5 folds, respectively, Figure 1E), and of course, for hydroxylation which has no correspondences outside the matrisome in the databases assessed to map baseline PTM sites (Figure 1E). Conversely, PTM^{mut} burden in matrisome is significantly devoid of sumoylation and ubiquitylation (approx. −0.7 and −0.5 folds, respectively, Figure 1E), in line with the global frequencies of the different PTMs in the matrisome. Reassuringly, these findings hold true even when comparing the matrisome against 100 random picks of the genome, in sets of the same size as the matrisome (X-square test, 0.0013 min to 0.02 max). Also, we notice that the local (tumor-specific) burdens (Figure S4C), once analyzed across each matrisome family (collagens, ECM-affiliated proteins, ECM glycoproteins, ECM regulators, proteoglycans and secreted factors) [9,16], correlate negatively with the amount of PTM^{mut} for that family (Figure 1F), especially for ECM-affiliated proteins, ECM glycoproteins, proteoglycans, and secreted factors indicating that—apart from a few recurrent ones—each PTM^{mut} occurs only one or twice in the whole dataset.

3.2. PTM^{mut} in the Tumor Matrisome

The abundance of PTM^{mut} across matrisome categories varies sensibly (Table S2), depending (as expected) on the prevalence of certain types of PTMs in the different categories [31–35]. For example, hydroxylation- and sumoylation-affecting PTM^{mut} are almost exclusively found in the core matrisome, where they mostly hit collagens and ECM glycoproteins, respectively (Figure 1G). On the other hand, acetylation- and phosphorylation-affecting PTM^{mut} are more abundant in matrisome-associated proteins (Figure 1G), in line with the regulatory functions of these proteins and the PTMs affected [9] and the presence of proteins with dual intracellular and extracellular localization (and functions) in this category.

The serial breakdown of PTM^{mut} counts by type, tumor, and matrisome category shows very similar patterns of mutation across the Pan-Cancer cohort, with skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), and uterine corpus endometrial carcinoma (UCEC) accounting for the majority of PTM^{mut} throughout the various matrisome categories (Figure 2 and Table S2) and no clear cell- or tissue-of-origin effects. Local differences exist, however, depending on the matrisome category (Figure S5A). In example, for collagens, lung adenocarcinoma and squamous cell carcinoma (LUAD and LUSC, respectively) show as many phosphorylation-affecting PTM^{mut} as STAD or more, and as double hydroxylation-affecting PTM^{mut}. Keeping with the example of collagens, the 10 most frequent hydroxylation-affecting PTM^{mut} target largely the same genes (*COL1A1*, *COL3A1*, *COL5A1* and *COL14A1*) across the Pan-Cancer cohort with the noticeable addition of *COL12A1* in LUSC and UCEC and the lack of *COL5A1* in LUAD; conversely, phosphorylation-affecting PTM^{mut} show a distinctive cluster of mutations in collagen VI genes (*COL6A2*, *COL6A3*, *COL6A5*, *COL6A6*) in UCEC, LUAD, and LUSC, of collagen III, IV, VII, and XXVII (*COL3A1*, *COL4A2*, *COL7A1*, *COL27A1*) in SCKM and of the fibril-associated collagen XXII (*COL22A1*) in all of them (Figure S5B).

The ten most frequent targets of PTM^{mut} per tumor type are shown in Figure 3. Unsurprisingly, the most frequently mutated matrisome genes across the Pan-Cancer cohort [40]—namely mucin 16 (*MUC16*) and filaggrin (*FLG*)—top the list of PTM^{mut}-affected genes too. There are, however, significant differences between PTM^{mut} and overall mutations, with e.g., hornerin (*HRNR*, a paralog of *FLG*) being a top-10 PTM^{mut} gene in 14/31 tumor types and versican (*VCAN*), collagen III (*COL3A1*) and XIV (*COL14A1*) all mutated in 9/31 tumor types.
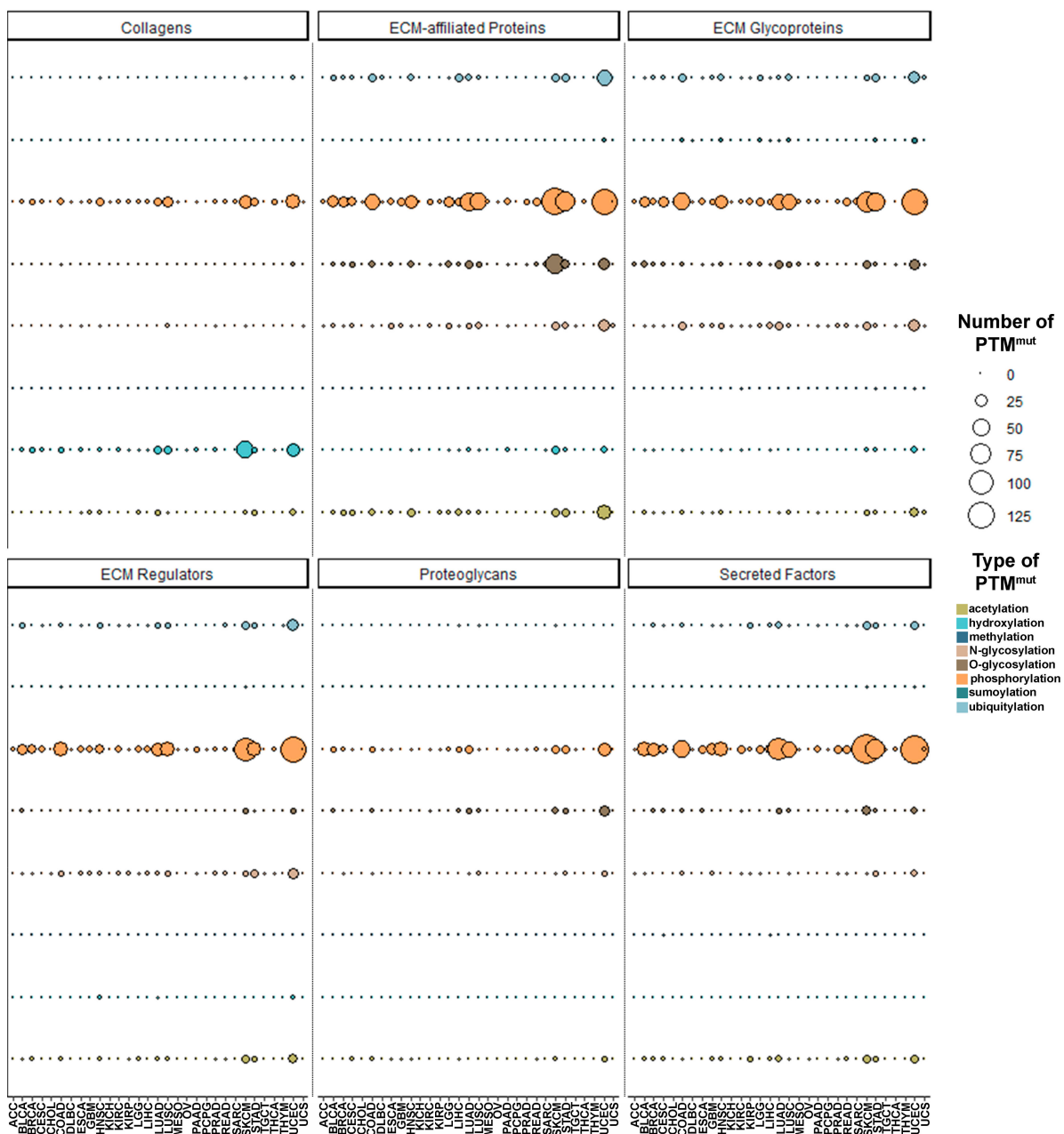
Figure 2. Pan-Cancer abundance and type of PTM^{mut} across the different matrisome families. The abundance (total number) of PTM^{mut}, divided by the different families composing the matrisome and the different types of PTM investigated, is shown.

In line with overall matrisome mutations, again, we observe a scarce conservation of PTM^{mut} which, mostly, occur only once in the whole dataset. 19 PTM^{mut}, however, occur at least once in three or more different cohorts, candidating to a role as potential "hotspots" (Figure S6 and Table S3). Of these, 16 (16/19, approx. 84% of total) are phosphorylation-PTM^{mut}, with the exception of PTM^{mut} affecting sumoylation or ubiquitylation of fibrillin 2 (*FBN2*) at lysine position 1078 and acetylation of fibroblast growth factor 14 (*FGF14*) at lysine position 245. Also, 17/19 of these "hotspot" genes (approx. 90% of total) are present in the list only once, the only exceptions being *FBN2* (though the two mutations just discussed appear at the same position) and mucin 16 (*MUC16*) enlisted thrice with PTM^{mut} affecting phosphorylation at serines in position 70, 496, and 3694. Interestingly, none of the PTM^{mut} types enriching the matrisome (hydroxylation, N- and O-glycosylation) occur frequently, suggesting that mutations at the positions where these PTMs occur are

poorly tolerated or, in the case of ubiquitylation-affecting mutations in protein with dual localization, that these mutations might alter the intracellular metabolism of the proteins involved. Analysis of SIFT and Polyphen results by PTM type, however, do only show minor differences between matrisome and non-matrisome PTM^{mut} (Figure S7A–D), with the interesting exception of O-glycosylation sites which are less frequently affected by damaging mutations in the matrisome than in the rest of the genome.
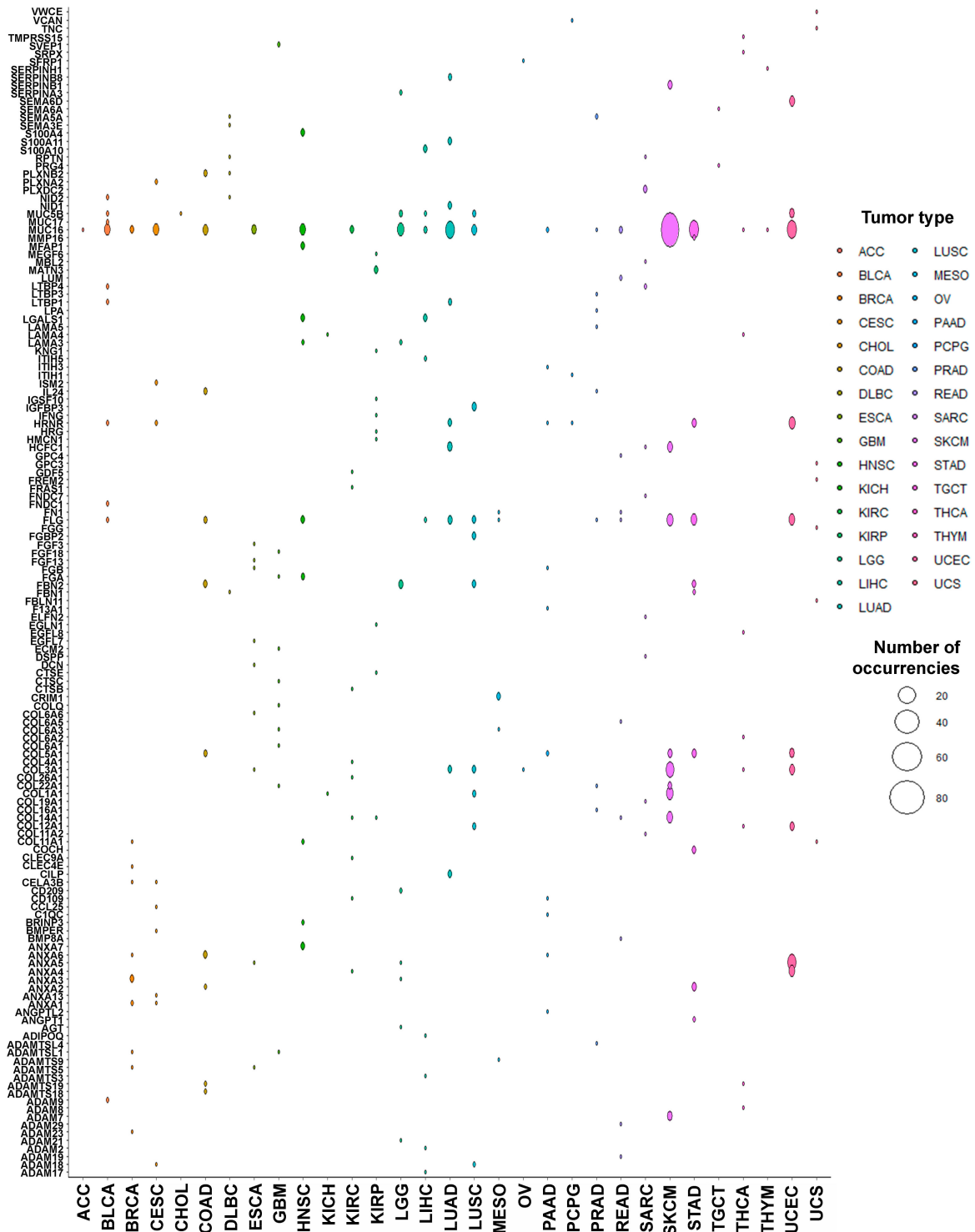

Figure 3. Genes most frequently affected by PTM^{mut}. The abundance (total number) of PTM^{mut} for the ten most-frequently affected genes, across the different tumor types, is shown.

The high ratio of mutations of unknown impact in both the algorithms prompted us to also calculate the ratio of non-silent to silent mutations (dN/dS), an indicator of negative selection pressure on mutations [44]. To this aim, per each gene, we calculated the dN/dS ratio of PTM^{mut} and of non-PTM^{mut} deriving a new measure (r$^{dN/dS}$) that represents the fraction obtained dividing the dN/dS value of PTM^{mut} by the dN/dS value of non-PTM^{mut}. Comparing the r$^{dN/dS}$ values by type of PTM (Figure S7E), it is evident that the matrisome is subject to more selection (lower r$^{dN/dS}$ values) for PTM^{mut} than the rest of the genome, in line with our previous results. Also we notice that, as compared to rest of the genome, the r$^{dN/dS}$ value of N-glycosylation and phosphorylation are particularly low (Figure S7F), suggesting that these categories are under a stringent selection for PTM^{mut} which might partially explain the poor presence of hotspot mutations.

3.3. Functional Characterization of Matrisome PTM^{mut}

To gain further understanding in the molecular and pathogenic effects of the PTM^{mut}, we evaluated all the PTM^{mut} identified so far in the context of the protein domains they reside in, comparing the frequency of these mutations to non-PTM mutations in the same domains to obtain a domain-specific PTM^{mut} ratio (Table S4). As the structure of ECM proteins is mostly modular, these domains mediate protein–protein interactions and are thus critical for proper protein functions [45]. Hence, we expect selection for or against PTM^{mut} to be critical at this level and, where possible, we compared the PTM^{mut} ratios of matrisome and non-matrisome genes by PTM type in the same domain. Results show that the frequency of PTM^{mut} varies significantly domain-wise in matrisome vs. rest of the genome, with 128 protein domains enriched for PTM^{mut} in the tumor matrisome (128/3373 domains from the Pfam database, approx. 3.79% of total, Table S5). In addition, also the comparison of PTM^{mut}/total mutation ratio in those 128 domains shows that PTM^{mut} are less tolerated than they are when the domain does not belong to a matrisome protein, especially in the case of O-glycosylation and acetylation (Figure S8), further suggesting that PTM^{mut} might have a strong negative impact on the function of the matrisome proteins they affect or their intracellular processing steps and eventual interactions, and that this might reflect in a lower fitness of the neoplastic clones harboring such mutations. On the other hand, the introduction of PTM^{mut} might, in fewer cases, alter protein functions and confer selective advantages to the harboring clones. Though this case is impossible to define here due to the lack of position-specific mapping of PTM functions in proteins, we suggest a few lines of evidence in support of the functional consequences of PTM^{mut}, focusing on those domains whose PTM^{mut} frequency is at least twice in matrisome than in rest of the genome or vice versa.

First, there are differences in the enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways that point specifically to peroxisome proliferator-activated receptor (PPAR) signaling and drug metabolism in the matrisome domains targeted by PTM^{mut} (Figure S9), in line with the recent evidence [46–51], suggesting possible gains for the clones harboring these mutations in the matrisome.

Furthermore, mapping the matrisome PTM^{mut} locations into the regions and sub-regions annotated for function by Uniprot we show that, of 921 mutations annotated with region information (no information available for the reminder 986), 286 (286/921, 31%) fall in one such a region (Table S6), again suggesting a functional consequence for matrisome PTM^{mut}. At an even finer scale, we attempted mapping PTM^{mut} and non-PTM mutations into the sequence context they belonged to, focusing on both specific motifs mediating matrisome interactions with receptors and other proteins and, more in general, on the sequence features of the mutated area. For the specific motifs (GPP, GVD, RGD, LDV, GFPGER, and GLPGER) [52,53] we applied a stringent cutoff of max ±3 amino acids from the mutated residue, while for the overall sequence we fetched ±10 amino acids from the mutated residue. Interestingly, both PTM^{mut} and non-PTM mutations are unlikely to occur in any of the motifs scanned, with only 11 PTM^{mut} from 7 genes mapping (Table S7) and 2 non-PTM mutations from 2 genes mapping. Though extremely low (speaking again

in favor of poor conservation of matrisome mutations disrupting functional loci), these numbers are still robustly different (X-square test $< 2.2 \times 10^{-16}$), suggesting that these few PTM^{mut} are tolerated and might favor the clones harboring them. In addition, the larger sequence context also likely influences the selection in favor or against PTM^{mut}, for example requiring almost exclusively a polar or hydrophobic amino acid two residues after the mutation site while showing no signs of selection for the same position in the same genes in the case of non-PTM mutations (Figure S10). As the N-glycosylation process shows evidence of stereotactic preferences for amino acids' charge and polarity in cancer [54], it is likely that matrisome proteins—undergoing N-glycosylation more massively than the rest of the proteome—might be more subject to protein-level selection, though this cannot be determined on the sole basis of these data.

Finally, of the 437 genes harboring PTM^{mut} in protein domains whose PTM^{mut} frequency is at least twice in matrisome than in the rest of the genome or vice versa (Table S8), 230 (230/437, approx. 53%) are involved in reciprocal homo- and/or heterotypic interactions according to BioGRID (Table S9), further supporting functional consequences for PTM^{mut} once acquired. In this context, we notice that 39 (39/230, approx. 17% of total) of the interacting genes affected by PTM^{mut} are also among those whose mutations map into known functional regions (Table S6 and Figure 4), again sustaining the hypothesis that, once gained, these mutations might disturb the integrity of the matrisome interaction network. Our results also shed a new light on the set of mutations affecting these genes, as the evidence provided by the assessment of their mutational effect is almost exclusively "Missense mutations" and that by SIFT and PolyPhen is otherwise largely of "unknown" type. In this context, we suggest that PTM^{mut}, especially in "hub" proteins [55] within the larger matrisome network, might be an overlooked class of mutations whose real effect is on the structure and stability of the network they participate to rather than the protein *per se*, though further studies will be needed to test this hypothesis.
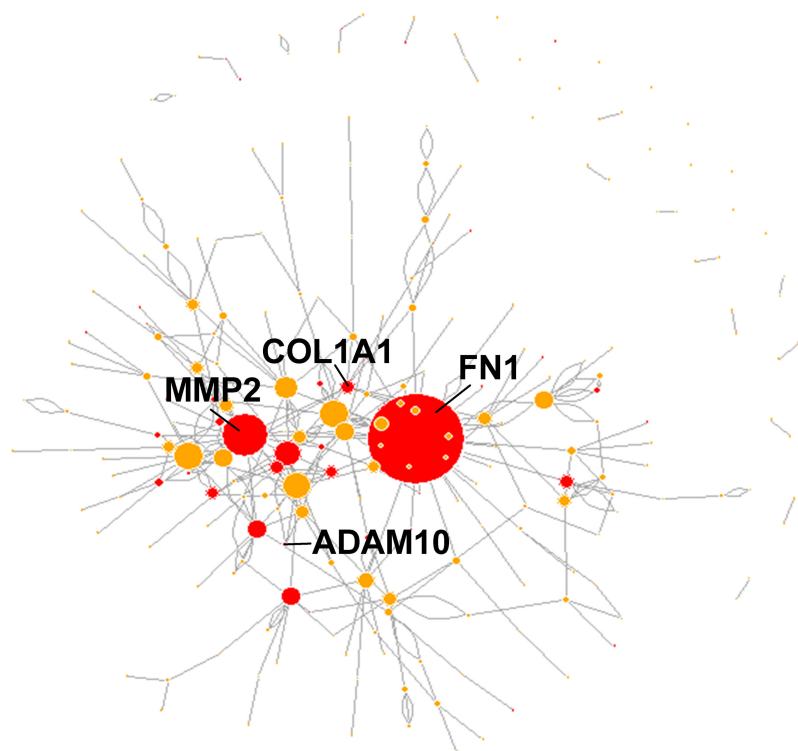

Figure 4. The interaction network of matrisome proteins with functional PTM^{mut}. Matrisome genes that harbored PTM^{mut} and whose proteins interacted reciprocally (according to BIOGRID) were isolated and the interaction network was drawn with node size proportional to the total degree of each node. In red, matrisome genes whose PTM^{mut} mapped in in regions endowed with specific protein functions (according to UniProt).

4. Discussion

In recent years, the development of "omics" techniques has opened the field of ECM and matrisome to systematic cancer-specific and Pan-Cancer investigations, which have shone a new light on the roles played by the matrisome in the TME and its importance in oncogenic and pathogenic mechanisms [3,8,9,12,13,16,40,56–62].

While much is known about the relative expression of matrisome genes and proteins in cancer, sensibly less is known about the mutations in the tumor matrisome. In particular, apart from a few focused studies [36,63–65], our recent effort at characterizing matrisome mutations Pan-Cancer [40] remains the only systematic example. In this study, we focused on a class of mutations, those affecting PTM, which might potentially have a great impact on the functional and structural roles of the tumor matrisome due to the paramount importance that PTM already have in the physiological and homeostatic functions of the matrisome.

It should be noted here that TCGA data derive from bulk but sufficiently pure tumor samples [66], ruling out a confounder role for the stromal and immune content of the samples in confidently mapping the mutations to the neoplastic cells within [40].

On this basis, we have analyzed 9075 patients and 32 tumor types from TCGA Pan-Cancer cohort and identified 151,088 non-silent mutations in the coding regions of the matrisome, of which 1811 affecting known sites of hydroxylation, phosphorylation, N- and O-glycosylation, acetylation, ubiquitylation, sumoylation, and methylation PTM (PTM^{mut}). As already discussed in the text, the matrisome seems less prone to accumulate PTM^{mut} than the rest of the genome according to both the frequency of PTM^{mut} per gene length and the $r^{dN/dS}$, suggesting a higher selective pressure on these mutations and hinting at a poor representation of the matrisome—and, in particular, of its PTM^{mut}—among the germline mutations that precede and predispose to cancer and rather suggesting their appearance along somatic cancer evolution [67,68], though this aspect will require further investigations into different data sets than TCGA.

This point is of particular interest, since we recently demonstrated that the matrisome accumulates more point mutations and copy number alterations (CNAs) than the rest of the genome in general [40] and thus the limited number of PTM-affecting mutations and their poor preservation is surprising, hinting at negative selection processes against this type of mutations in the matrisome. Unfortunately, a recent retrospective analysis on the timing of mutation insurgence in TCGA [69] does not help in clarifying this point, as it covers chromosome-level events rather than the intragenic mutations we are after. On the other hand, approximate time-series of colon and lung cancer show that mutational events that affect structural matrisome components such as fibronectin 1 and collagen 1 (*FN1* and *COL1A1*) as well as proteinases (*MMP2* and *ADAM10*) and other functional moieties (*NTN4*, *PCSK6*, and *SVEP1*) are likely tumor driver [70]—though, again, these data are on a different scale than that of this manuscript, so comparison is only approximate and conceptual; it is worth noticing, however, that we found PTM^{mut} for all these genes but *NTN4*, that *FN1*, *COL1A1*, *MMP2*, and *ADAM10* all harbor PTM^{mut} in functional domains and interact reciprocally and that, in the network of interactions between PTM^{mut}-affected matrisome elements, these genes (especially *FN1* and *MMP2*) are major hubs, all these evidence suggesting that matrisome PTM^{mut} might significantly contribute to altered TME dynamics. Further along this line, we notice that 143 PTM^{mut} are not found at all in healthy samples from TCGA irrespective of tumor type (Table S10) and that 9 PTM^{mut} affect "landmark" matrisome genes that deeply characterize the given tumors [71], possibly candidating the carrier genes to a more prominent role in cancer.

In this context, based on the high variability of PTMmut (which often occur only once or twice across the whole Pan-Cancer cohort) and by similarity with other matrisome mutations in general [40,63,72], we deem unlikely that any of these mutations might have a role as driver and they seem more probably the result of passenger mutations with lower historical selection [73]. More likely, the lower amount of PTM^{mut} in the matrisome is the result of the convergent structural and functional damages wrought to the proteins by both

the altered PTM profile *and* the changes at the amino acid sequence level where the PTM should be [21,74–76]. It must be taken into consideration, in fact, that a significant portion of all PTM^{mut} we assessed (20% to 30% of total, according to either SIFT or PolyPhen) can be classified as deleterious to the protein function, involving various types of out-of-frame mutations. Additionally, the majority of PTM^{mut} affect lysine, asparagine, proline, serine, and threonine (see Table S11), all of which are under different selective pressures [77–79] as targets of mutations because of their physico-chemical properties.

Still, the majority of PTM^{mut} that we have investigated has no clear effect on the protein structure, seemingly being tolerable or outside of the threshold for potentially damaging effect. We speculate that the effect of these mutations may be at the system-level scale, where the matrisome proteins harboring the PTM^{mut} might interact incorrectly (*how incorrectly* remains to be determined at the single protein level) with its partners, impairing the network structure and stability of the matrisome in the TME and potentially affecting the signaling pathways that depend or impinge on it. In this context, we have identified at least 286 PTM^{mut} falling in a functional protein region, where they might alter the activity of the protein by proximity with catalytic residues and binding sites [80]. In example, 124 PTM^{mut} in collagens (124 out of 156, 79.5%) occur within the triple helical region, where alterations in the PTM profile can easily damage the conformation and thermal stability of the proteins [81]. Similarly, the PTM^{mut} found in versican (*VCAN*) span both the alpha and beta glucosaminoglycan (GAG) attachment domains and thus impact on the addition of GAGs to the core versican protein, potentially disrupting a high number of local interactions [82], the PTM^{mut} in fibroblast growth factor 2 and fibronectin (*FGF2* and *FN1*, respectively) affect the heparin-binding domains of these proteins, necessary for cell–matrix and matrix–matrix interactions [83,84], and those in matrix metalloproteinase 2 (*MMP2*, type IV collagenase) localize in the collagen-binding domain, where they likely play a role in the regulation of enzyme activity [85]. Even more interestingly, 6 patients (Table S12) also showed concomitant PTM^{mut} in proteins involved in heterodimeric interactions, representing a particular case of study to identify network-scale disarrays to the matrisome which, once more, are significantly less frequent in the Pan-Cancer cohort vs. co-occurring non-matrisome PTM^{mut} (X-square test < 2.2×10^{-16}).

In conclusion, we believe that our results demonstrate, at the genomic level, the potential impact of PTM^{mut} on the tumor matrisome and mark a starting point to their functional characterization, enabling a more comprehensive and integrated view of this critical piece of the TME puzzle whose fine understanding may lead to significant translational applications to improve cancer patient treatment and eventually the outcome.

ratios tumor-wise, Figure S3. Transitions and transversions among PTMmut in the tumor matrisome. The % amount of transitions and transversions in PTMmut (A) and (B) in non-PTM mutations was calculated and compared for matrisome and rest of the genome. The same values were further tabulated across different types of mutation (C) and, (D), transitions and transversions within PTMmut of the matrisome were compared across matrisome families. Abbreviations: n.s., not significant. *p*-values are from Chi-square tests, Figure S4. Development and calculation of the PTMmut "burden". The amount of mutations across loci with different types of PTM in the matrisome and rest of the genome (A) and the different baseline quantities of such PTMs in the two groups (B) were used to calculate a PTM-specific ratio (the "burden") across the different tumor types ("local burdens", only matrisome shown in C). *p*-values are from Chi-square tests, Figure S5. Variety and abundance of PTMmut in the tumor matrisome. The amount of PTMmut across the different matrisome categories and tumor types (A) varies considerably (*p*-value < 2e-16 for all, Chi-square test) though similar patterns of PTMmut acquisition in genes can be found across multiple cancers, depending on the type of PTM (B), Figure S6. PTMmut "hotspots" in the tumor matrisome. The occurrence of each matrisome PTMmut for the total Pan-Cancer cohort (number of occurrences) and the number of cohorts in which it appeared (number of cohorts) were calculated, and the likely "hotspots" (those occurring at least once in three cohorts or more) were identified (in red), Figure S7. Mutational effects and conservation of PTMmut in the tumor matrisome. The effects of mutations across loci with different types of PTM were calculated for the tumor matrisome and rest of the genome using SIFT and POLYPHEN algorithms (A–D). Note the lower abundance of deleterious or damaging PTMmut in the tumor matrisome. In (E), the ratio of PTMmut to silent mutations for each gene (rdN/dS) was calculated and then averaged for the Pan-Cancer cohort across the different types of mutations for the matrisome and rest of the genome, and (F) compared (matrisome vs. rest of the genome) to evaluate differences. Abbreviations: n.s., not significant. *p*-values are from Chi-square tests, Figure S8. Mapping and abundance of PTMmut at the protein domain level. Each PTMmut and non-PTM mutation in the tumor matrisome or in the rest of the genome was mapped to protein domains according to Pfam coordinates and the % of mapping mutations (PTMmut vs. all mutations) was calculated to compare the relative tolerance of any protein domain to PTMmut across different types of PTM in the two groups, Figure S9. Ontological enrichment of PTMmut-abundant protein domains. Protein domains (from Pfam) whose PTMmut vs. all mutations ratio was at least 2 times higher in (left) the tumor matrisome than in the rest of the genome or, (right) vice versa, were mapped back into the genes of origin and these subjected to ontological enrichment using the Kyoto En-cyclopedia of Genes and Genomes (KEGG) annotations. Only terms with false discovery rate (FDR) < 0.01 were main-tained, Figure S10. Sequence features of PTM-mutated loci. Amino acid (AA) positions affected by PTMmut in the matrisome and rest of the genome were expanded by 10 AA in both directions and consensus sequences (logos) were generated according to the relative frequency of AA in each position and their chemistry. Note that, due to the different length of the protein isoforms the mutations map to, the AA affected by PTMmut is always in position 11 or 12.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The R code sustaining this submission is available at https://github.com/Izzilab and https://rpubs.com/Izzilab/. The necessary data for the code are stored in a freely-accessible Zenodo repository (https://doi.org/10.5281/zenodo.4490484). All data were deposited on 2 February 2021.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tomczak, K.; Czerwinska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An Immeasurable Source of Knowledge. *Contemp. Oncol.* **2015**, *19*, 68–77. [CrossRef] [PubMed]
2. Quail, D.F.; Bowman, R.L.; Akkari, L.; Quick, M.L.; Schuhmacher, A.J.; Huse, J.T.; Holland, E.C.; Sutton, J.C.; Joyce, J.A. The Tumor Microenvironment Underlies Acquired Resistance to CSF-Inhibition in Gliomas. *Science* **2016**, *352*, aad3018. [CrossRef] [PubMed]
3. Balkwill, F.R.; Capasso, M.; Hagemann, T. The Tumor Microenvironment at a Glance. *J. Cell. Sci.* **2012**, *125*, 5591–5596. [CrossRef]
4. Rianna, C.; Kumar, P.; Radmacher, M. The Role of the Microenvironment in the Biophysics of Cancer. *Semin. Cell Dev. Biol.* **2018**, *73*, 107–114. [CrossRef] [PubMed]
5. Thorsson, V.; Gibbs, D.L.; Brown, S.D.; Wolf, D.; Bortone, D.S.; Ou Yang, T.H.; Porta-Pardo, E.; Gao, G.F.; Plaisier, C.L.; Eddy, J.A.; et al. The Immune Landscape of Cancer. *Immunity* **2018**, *48*, 812–830.e14. [CrossRef]
6. Semenza, G.L. The Hypoxic Tumor Microenvironment: A Driving Force for Breast Cancer Progression. *Biochim. Biophys. Acta* **2016**, *1863*, 382–391. [CrossRef] [PubMed]
7. DeClerck, Y.A.; Pienta, K.J.; Woodhouse, E.C.; Singer, D.S.; Mohla, S. The Tumor Microenvironment at a Turning Point Knowledge Gained over the Last Decade, and Challenges and Opportunities Ahead: A White Paper from the NCI TME Network. *Cancer Res.* **2017**, *77*, 1051–1059. [CrossRef] [PubMed]
8. Brassart-Pasco, S.; Brézillon, S.; Brassart, B.; Ramont, L.; Oudart, J.; Monboisse, J.C. Tumor Microenvironment: Extracellular Matrix Alterations Influence Tumor Progression. *Front. Oncol.* **2020**, *10*, 397. [CrossRef]
9. Naba, A.; Clauser, K.R.; Ding, H.; Whittaker, C.A.; Carr, S.A.; Hynes, R.O. The Extracellular Matrix: Tools and Insights for the "Omics" Era. *Matrix Biol.* **2016**, *49*, 10–24. [CrossRef]
10. Naba, A.; Clauser, K.R.; Hoersch, S.; Liu, H.; Carr, S.A.; Hynes, R.O. The Matrisome: In Silico Definition and in Vivo Characterization by Proteomics of Normal and Tumor Extracellular Matrices. *Mol. Cell Proteom.* **2012**, *11*, M111.014647. [CrossRef] [PubMed]
11. Yuzhalin, A.E.; Urbonas, T.; Silva, M.A.; Muschel, R.J.; Gordon-Weeks, A.N. A Core Matrisome Gene Signature Predicts Cancer Outcome. *Br. J. Cancer* **2018**, *118*, 435–440. [CrossRef]
12. Jia, D.; Liu, Z.; Deng, N.; Tan, T.Z.; Huang, R.Y.; Taylor-Harding, B.; Cheon, D.J.; Lawrenson, K.; Wiedemeyer, W.R.; Walts, A.E.; et al. A COL11A1-Correlated Pan-Cancer Gene Signature of Activated Fibroblasts for the Prioritization of Therapeutic Targets. *Cancer Lett.* **2016**, *382*, 203–214. [CrossRef] [PubMed]
13. Izzi, V.; Lakkala, J.; Devarajan, R.; Kääriäinen, A.; Koivunen, J.; Heljasvaara, R.; Pihlajaniemi, T. Pan-Cancer Analysis of the Expression and Regulation of Matrisome Genes across 32 Tumor Types. *Matrix Biol. Plus* **2019**, *1*, 100004. [CrossRef]
14. Hanahan, D.; Weinberg, R.A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144*, 646–674. [CrossRef]
15. Pickup, M.W.; Mouw, J.K.; Weaver, V.M. The Extracellular Matrix Modulates the Hallmarks of Cancer. *EMBO Rep.* **2014**, *15*, 1243–1253. [CrossRef] [PubMed]
16. Socovich, A.M.; Naba, A. The Cancer Matrisome: From Comprehensive Characterization to Biomarker Discovery. *Semin. Cell Dev. Biol.* **2019**, *89*, 157–166. [CrossRef] [PubMed]
17. Jeschke, J.; Collignon, E.; Fuks, F. DNA Methylome Profiling Beyond Promoters—Taking an Epigenetic Snapshot of the Breast Tumor Microenvironment. *FEBS J.* **2015**, *282*, 1801–1814. [CrossRef]
18. Takeshima, H.; Ushijima, T. Accumulation of Genetic and Epigenetic Alterations in Normal Cells and Cancer Risk. *NPJ Precis. Oncol.* **2019**, *3*, 1–8.
19. Jin, M.; Jin, W. The Updated Landscape of Tumor Microenvironment and Drug Repurposing. *Signal Transduct. Target. Ther.* **2020**, *5*, 1–16. [CrossRef] [PubMed]
20. Müller, M.M. Post-Translational Modifications of Protein Backbones: Unique Functions, Mechanisms, and Challenges. *Biochemistry* **2018**, *57*, 177–185. [CrossRef] [PubMed]
21. Karve, T.M.; Cheema, A.K. Small Changes Huge Impact: The Role of Protein Posttranslational Modifications in Cellular homeostasis and disease. *J. Amino. Acids* **2011**, *2011*, 207691. [CrossRef] [PubMed]
22. Virág, D.; Dalmadi-Kiss, B.; Vékey, K.; Drahos, L.; Klebovich, I.; Antal, I.; Ludányi, K. Current Trends in the Analysis of Post-Translational Modifications. *Chromatographia* **2020**, *83*, 1–10. [CrossRef]
23. Barber, K.W.; Rinehart, J. The ABCs of PTMs. *Nat. Chem. Biol.* **2018**, *14*, 188–192. [CrossRef]
24. Kam, R.K.T.; Poon, T.C.W. The Potentials of Glycomics in Biomarker Discovery. *Clin. Proteom.* **2008**, *4*, 67–79. [CrossRef]
25. Shriver, Z.; Raguram, S.; Sasisekharan, R. Glycomics: A Pathway to a Class of New and Improved Therapeutics. *Nat. Rev. Drug Discov.* **2004**, *3*, 863–873. [CrossRef] [PubMed]
26. Varki, A. Biological Roles of Glycans. *Glycobiology* **2017**, *27*, 3–49. [CrossRef]
27. Santos, A.L.; Lindner, A.B. Protein Posttranslational Modifications: Roles in Aging and Age-Related Disease. *Oxid. Med. Cell Longev.* **2017**, *2017*, 5716409. [CrossRef] [PubMed]
28. Reimand, J.; Wagih, O.; Bader, G.D. Evolutionary Constraint and Disease Associations of Post-Translational Modification Sites in Human Genomes. *PLoS Genet.* **2015**, *11*, e1004919. [CrossRef]
29. Yang, Y.; Peng, X.; Ying, P.; Tian, J.; Li, J.; Ke, J.; Zhu, Y.; Gong, Y.; Zou, D.; Yang, N.; et al. AWESOME: A Database of SNPs that Affect Protein Post-Translational Modifications. *Nucleic Acids Res.* **2019**, *47*, D874–D880. [CrossRef] [PubMed]

30. Kim, Y.; Kang, C.; Min, B.; Yi, G. Detection and Analysis of Disease-Associated Single Nucleotide Polymorphism Influencing Post-Translational Modification. *BMC Med. Genom.* **2015**, *8* (Suppl. 2), S7. [CrossRef] [PubMed]

31. Leeming, D.J.; Bay-Jensen, A.C.; Vassiliadis, E.; Larsen, M.R.; Henriksen, K.; Karsdal, M.A. Post-Translational Modifications of the Extracellular Matrix are Key Events in Cancer Progression: Opportunities for Biochemical Marker Development. *Biomarkers* **2011**, *16*, 193–205. [CrossRef] [PubMed]

32. Karsdal, M.A.; Nielsen, M.J.; Sand, J.M.; Henriksen, K.; Genovese, F.; Bay-Jensen, A.; Smith, V.; Adamkewicz, J.I.; Christiansen, C.; Leeming, D.J. Extracellular Matrix Remodeling: The Common Denominator in Connective Tissue Diseases. Possibilities for Evaluation and Current Understanding of the Matrix as More than a Passive Architecture, but a Key Player in Tissue Failure. *Assay Drug Dev. Technol.* **2013**, *11*, 70–92. [CrossRef] [PubMed]

33. Rappu, P.; Salo, A.M.; Myllyharju, J.; Heino, J. Role of Prolyl Hydroxylation in the Molecular Interactions of Collagens. *Essays Biochem.* **2019**, *63*, 325–335. [PubMed]

34. Reily, C.; Stewart, T.J.; Renfrow, M.B.; Novak, J. Glycosylation in Health and Disease. Nature reviews. *Nephrology* **2019**, *15*, 346–366.

35. Hsiao, C.; Cheng, H.; Huang, C.; Li, H.; Ou, M.; Huang, J.; Khoo, K.; Yu, H.W.; Chen, Y.; Wang, Y.; et al. Fibronectin in Cell Adhesion and Migration Via N-Glycosylation. *Oncotarget* **2017**, *8*, 70653–70668. [CrossRef] [PubMed]

36. Chen, L.; Miao, Y.; Liu, M.; Zeng, Y.; Gao, Z.; Peng, D.; Hu, B.; Li, X.; Zheng, Y.; Xue, Y.; et al. Pan-Cancer Analysis Reveals the Functional Importance of Protein Lysine Modification in Cancer Development. *Front. Genet.* **2018**, *9*, 254. [CrossRef] [PubMed]

37. Chen, L.; Liu, S.; Tao, Y. Regulating Tumor Suppressor Genes: Post-Translational Modifications. *Signal Transduct. Target. Ther.* **2020**, *5*, 1–25. [CrossRef]

38. Jin, H.; Zangar, R.C. Protein Modifications as Potential Biomarkers in Breast Cancer. *Biomark Insights* **2009**, *4*, 191–200. [CrossRef] [PubMed]

39. Winkler, J.; Abisoye-Ogunniyan, A.; Metcalf, K.J.; Werb, Z. Concepts of Extracellular Matrix Remodelling in Tumour Progression and Metastasis. *Nat. Commun.* **2020**, *11*, 1–19. [CrossRef] [PubMed]

40. Izzi, V.; Davis, M.N.; Naba, A. Pan-Cancer Analysis of the Genomic Alterations and Mutations of the Matrisome. *Cancers* **2020**, *12*, 2046. [CrossRef]

41. Trinh, A.; Del Alcazar, C.R.G.; Shukla, S.A.; Chin, K.; Chang, Y.H.; Thibault, G.; Eng, J.; Jovanović, B.; Aldaz, C.M.; Park, S.Y.; et al. Genomic Alterations during the in Situ to Invasive Ductal Breast Carcinoma Transition Shaped by the Immune System. *Mol. Cancer Res.* **2020**. [CrossRef] [PubMed]

42. Campbell, P.J.; Getz, G.; Korbel, J.O.; Stuart, J.M.; Jennings, J.L.; Stein, L.D.; Perry, M.D.; Nahal-Bose, H.; Ouellette, B.F.F.; Li, C.H.; et al. Pan-Cancer Analysis of Whole Genomes. *Nature* **2020**, *578*, 82–93.

43. Hutter, C.; Zenklusen, J.C. The Cancer Genome Atlas: Creating Lasting Value Beyond its Data. *Cell* **2018**, *173*, 283–285. [CrossRef]

44. Zapata, L.; Pich, O.; Serrano, L.; Kondrashov, F.A.; Ossowski, S.; Schaefer, M.H. Negative Selection in Tumor Genome Evolution Acts on Essential Cellular Functions and the Immunopeptidome. *Genome Biol.* **2018**, *19*, 67. [CrossRef] [PubMed]

45. Hohenester, E.; Engel, J. Domain structure and organisation in extracellular matrix proteins. *Matrix Biol.* **2002**, *21*, 115–128. [CrossRef]

46. Xu, J.; Liao, K.; Wang, X.; He, J.; Wang, X. Combining Bioinformatics Techniques to Explore the Molecular Mechanisms Involved in Pancreatic Cancer Metastasis and Prognosis. *J. Cell Mol. Med.* **2020**, *24*, 14128–14138. [CrossRef] [PubMed]

47. Zheng, X.; Liu, X.; Zheng, H.; Wang, H.; Hong, D. Integrated Bioinformatics Analysis Identified COL11A1 as an Immune Infiltrates Correlated Prognosticator in Pancreatic Adenocarcinoma. *Int. Immunopharmacol.* **2020**, *90*, 106982. [CrossRef] [PubMed]

48. Winkler, I.; Bitter, C.; Winkler, S.; Weichenhan, D.; Thavamani, A.; Hengstler, J.G.; Borkham-Kamphorst, E.; Kohlbacher, O.; Plass, C.; Geffers, R.; et al. Identification of Pparγ-Modulated miRNA Hubs that Target the Fibrotic Tumor Microenvironment. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 454–463. [CrossRef]

49. Bai, J.; Zhang, X.; Kang, X.; Jin, L.; Wang, P.; Wang, Z. Screening of Core Genes and Pathways in Breast Cancer Development via Comprehensive Analysis of Multi Gene Expression Datasets. *Oncol. Lett.* **2019**, *18*, 5821–5830. [CrossRef] [PubMed]

50. Prieto-Vila, M.; Takahashi, R.; Usuba, W.; Kohama, I.; Ochiya, T. Drug Resistance Driven by Cancer Stem Cells and their Niche. *Int. J. Mol. Sci.* **2017**, *18*, 2574. [CrossRef]

51. Jabłońska-Trypuć, A.; Matejczyk, M.; Rosochacki, S. Matrix Metalloproteinases (MMPs), the Main Extracellular Matrix (ECM) Enzymes in Collagen Degradation, as a Target for Anticancer Drugs. *J. Enzyme. Inhib. Med. Chem.* **2016**, *31*, 177–183. [CrossRef] [PubMed]

52. Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. Crystal Structure of the Collagen Triple Helix Model [(Pro-Pro-Gly)$_{10}$]$_3$. *Protein Sci.* **2002**, *11*, 262–270. [CrossRef] [PubMed]

53. Caswell, C.C.; Barczyk, M.; Keene, D.R.; Lukomska, E.; Gullberg, D.E.; Lukomski, S. Identification of the First Prokaryotic Collagen Sequence Motif that Mediates Binding to Human Collagen Receptors, Integrins alpha2beta1 and alpha11beta1. *J. Biol. Chem.* **2008**, *283*, 36168–36175. [CrossRef] [PubMed]

54. Manwar Hussain, M.R.; Iqbal, Z.; Qazi, W.M.; Hoessli, D.C. Charge and Polarity Preferences for N-Glycosylation: A Genome-Wide in Silico Study and its Implications regarding Constitutive Proliferation and Adhesion of Carcinoma Cells. *Front. Oncol.* **2018**, *8*, 29. [CrossRef] [PubMed]

55. Vallabhajosyula, R.R.; Chakravarti, D.; Lutfeali, S.; Ray, A.; Raval, A. Identifying Hubs in Protein Interaction Networks. *PLoS ONE* **2009**, *4*, e5344. [CrossRef]

56. Lu, P.; Weaver, V.M.; Werb, Z. The Extracellular Matrix: A Dynamic Niche in Cancer Progression. *J. Cell Biol.* **2012**, *196*, 395–406. [CrossRef]

57. Venning, F.A.; Wullkopf, L.; Erler, J.T. Targeting ECM Disrupts Cancer Progression. *Front. Oncol.* **2015**, *5*, 224. [CrossRef] [PubMed]

58. Tomko, L.A.; Hill, R.C.; Barrett, A.; Szulczewski, J.M.; Conklin, M.W.; Eliceiri, K.W.; Keely, P.J.; Hansen, K.C.; Ponik, S.M. Targeted Matrisome Analysis Identifies Thrombospondin-2 and Tenascin-C in Aligned Collagen Stroma from Invasive Breast Carcinoma. *Sci. Rep.* **2018**, *8*, 12941. [CrossRef] [PubMed]

59. Todd, J.R.; Ryall, K.A.; Vyse, S.; Wong, J.P.; Natrajan, R.C.; Yuan, Y.; Tan, A.C.; Huang, P.H. Systematic Analysis of Tumour Cell-Extracellular Matrix Adhesion Identifies Independent Prognostic Factors in Breast Cancer. *Oncotarget* **2016**, *7*, 62939–62953. [CrossRef] [PubMed]

60. Hoye, A.M.; Erler, J.T. Structural ECM Components in the Premetastatic and Metastatic Niche. *Am. J. Physiol. Cell Physiol.* **2016**, *310*, 955–967. [CrossRef] [PubMed]

61. Izzi, V.; Lakkala, J.; Devarajan, R.; Ruotsalainen, H.; Savolainen, E.R.; Koistinen, P.; Heljasvaara, R.; Pihlajaniemi, T. An Extracellular Matrix Signature in Leukemia Precursor Cells and Acute Myeloid Leukemia. *Haematologica* **2017**, *102*, e245–e248. [CrossRef] [PubMed]

62. Naba, A.; Ricard-Blum, S. *The Extracellular Matrix Goes-Omics: Resources and Tools*, 7th ed.; Springer: Cham, Switzerland, 2020; pp. 1–16.

63. King, R.J.; Yu, F.; Singh, P.K. Genomic Alterations in Mucins Across Cancers. *Oncotarget* **2017**, *8*, 67152–67168. [CrossRef]

64. Sharma, A.; Jiang, C.; De, S. Dissecting the Sources of Gene Expression Variation in a Pan-Cancer Analysis Identifies Novel Regulatory Mutations. *Nucleic Acids Res.* **2018**, *46*, 4370–4381. [CrossRef]

65. Kanwal, M.; Ding, X.J.; Song, X.; Zhou, G.B.; Cao, Y. MUC16 Overexpression Induced by Gene Mutations Promotes Lung Cancer Cell Growth and Invasion. *Oncotarget* **2018**, *9*, 12226–12239. [CrossRef]

66. Aran, D.; Sirota, M.; Butte, A.J. Systematic Pan-Cancer Analysis of Tumour Purity. *Nat. Commun.* **2015**, *6*, 8971. [CrossRef] [PubMed]

67. Vicens, A.; Posada, D. Selective Pressures on Human Cancer Genes Along the Evolution of Mammals. *Genes* **2018**, *9*, 582. [CrossRef] [PubMed]

68. Lipinski, K.A.; Barber, L.J.; Davies, M.N.; Ashenden, M.; Sottoriva, A.; Gerlinger, M. Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends Cancer* **2016**, *2*, 49–63. [CrossRef] [PubMed]

69. Gerstung, M.; Jolly, C.; Leshchiner, I.; Dentro, S.C.; Gonzalez, S.; Rosebrock, D.; Mitchell, T.J.; Rubanova, Y.; Anur, P.; Yu, K.; et al. The Evolutionary History of 2,658 Cancers. *Nature* **2020**, *578*, 122–128. [CrossRef]

70. Auslander, N.; Wolf, Y.I.; Koonin, E.V. In Silico Learning of Tumor Evolution through Mutational Time Series. *Proc. Natl. Acad Sci. USA* **2019**, *116*, 9501–9510. [CrossRef] [PubMed]

71. Kääriäinen, A.; Pesola, V.; Dittmann, A.; Kontio, J.; Koivunen, J.; Pihlajaniemi, T.; Izzi, V. Machine Learning Identifies Robust Matrisome Markers and Regulatory Mechanisms in Cancer. *Int. J. Mol. Sci.* **2020**, *21*, 8837. [CrossRef]

72. Henke, E.; Nandigama, R.; Ergün, S. Extracellular Matrix in the Tumor Microenvironment and its Impact on Cancer Therapy. *Front. Mol. Biosci.* **2020**, *6*, 160. [CrossRef]

73. Salvadores, M.; Mas-Ponte, D.; Supek, F. Passenger Mutations Accurately Classify Human Tumors. *PLoS Comput. Biol.* **2019**, *15*, e1006953. [CrossRef] [PubMed]

74. Salo, A.M.; Myllyharju, J. Prolyl and Lysyl Hydroxylases in Collagen Synthesis. *Exp. Dermatol.* **2021**, *30*, 38–49. [CrossRef]

75. Myllyharju, J.; Kivirikko, K.I. Collagens, Modifying Enzymes and their Mutations in Humans, Flies and Worms. *Trends Genet.* **2004**, *20*, 33–43. [CrossRef] [PubMed]

76. Gjaltema, R.A.F.; Bank, R.A. Molecular Insights into Prolyl and Lysyl Hydroxylation of Fibrillar Collagens in Health and Disease. *Crit. Rev. Biochem. Mol. Biol.* **2017**, *52*, 74–95. [CrossRef]

77. Vitkup, D.; Sander, C.; Church, G.M. The Amino-Acid Mutational Spectrum of Human Genetic Disease. *Genome Biol.* **2003**, *4*, R72. [CrossRef]

78. Creixell, P.; Schoof, E.M.; Tan, C.H.S.; Linding, R. Mutational Properties of Amino Acid Residues: Implications for Evolvability of Phosphorylatable Residues. Philosophical transactions. *Biol. Sci.* **2012**, *367*, 2584–2593. [CrossRef] [PubMed]

79. Schaefer, C.; Rost, B. Predict Impact of Single Amino Acid Change upon Protein Structure. *BMC Genom.* **2012**, *13* (Suppl. 4), S4. [CrossRef]

80. Holehouse, A.S.; Naegle, K.M. Reproducible Analysis of Post-Translational Modifications in Proteomes—Application to Human Mutations. *PLoS ONE* **2015**, *10*, e0144692. [CrossRef] [PubMed]

81. Yamauchi, M.; Sricholpech, M. Lysine Post-Translational Modifications of Collagen. *Essays Biochem.* **2012**, *52*, 113–133. [PubMed]

82. Wu, Y.J.; La Pierre, D.P.; Jin, W.U.; Albert, J.Y.; Burton, B.Y. The Interaction of Versican with its Binding Partners. *Cell Res.* **2005**, *15*, 483–494. [CrossRef]

83. Li, Y.; Sun, C.; Yates, E.A.; Jiang, C.; Wilkinson, M.C.; Fernig, D.G. Heparin Binding Preference and Structures in the Fibroblast Growth Factor Family Parallel their Evolutionary Diversification. *Open Biol.* **2016**, *6*, 150275. [CrossRef]

84. Dalton, B.A.; McFarland, C.D.; Underwood, P.A.; Steele, J.G. Role of the Heparin Binding Domain of Fibronectin in Attachment and Spreading of Human Bone-Derived Cells. *J. Cell Sci.* **1995**, *108*, 2083–2092. [PubMed]
85. Madzharova, E.; Kastl, P.; Sabino, F.; auf dem Keller, U. Post-Translational Modification-Dependent Activity of Matrix Metalloproteinases. *Int. J. Mol. Sci.* **2019**, *20*, 3077. [CrossRef]