Laboratory Investigation (2019) 99:1287–1295
https://doi.org/10.1038/s41374-019-0248-3

ARTICLE

Potential MMP2-mediated availability of HLA binding, mutant ECM peptides reflects better melanoma survival rates and greater T-cell infiltrates

Saif Zaman[1] · Boris I. Chobrutskiy[1] · Jay S. Patel[1] · Blake M. Callahan[1] · Moody M. Mihyu[1] · Andrea Diviney[1] · George Blanck[1,2]

Received: 15 October 2018 / Revised: 19 February 2019 / Accepted: 8 March 2019 / Published online: 24 April 2019

Abstract

Proteases in the cancer microenvironment have been studied for some time, with a general conclusion that such proteases facilitate the spread of cancer, although there is some controversy regarding that conclusion in later-stage cancer development. More recently, a very large collection of data regarding mutant amino acids in the potential substrates of cancer microenvironment proteases have become available. To better understand the potential impact of these mutant amino acids on protease function and cancer progression, we established a bioinformatics approach to assessing the impact of melanoma mutants, among a previously defined set of extracellular matrix (ECM) structural proteins, on the sensitivity of matrix metalloproteinase-2 (MMP2), extensively associated with melanoma. The results indicated that tumor samples with mutant amino acids adjacent to the ECM structural protein, MMP2 sites also represented a better survival rate and a larger proportion of mutant peptides with high HLA class I-binding affinities, particularly in comparison with melanoma samples with a reduced or absent T-cell infiltrate. Furthermore, even better HLA class I binders were identified among the samples representing the ECM structural protein mutants resistant to MMP2. Samples representing only MMP2-resistant mutants also represented a worse overall survival. Overall, this analysis suggested that MMP2 has the capacity of freeing mutant peptides that could facilitate an anti-tumor response and a better survival rate, and this analysis has the potential of resolving some of the controversy surrounding the role of cancer proteases in cancer progression.

Introduction

Protease aspects of cancer biology have been investigated for decades [1], and proteases represent a common target for the development of new anti-cancer therapeutics [2]. Proteases in the cancer microenvironment are considered to be expressed and secreted by the cancer cell as an aspect of aberrant regulation of protease expression [3]. For example, MMP2 and MMP14 are upregulated in a variety of cancer types by TGF-β [3]. Also, proteases can be present in the cancer microenvironment due to secretion by immune function cells, e.g., mast cells in breast cancer [4]. Cancer microenvironment proteases are thought to facilitate a loss of contact inhibition, with subsequent commencement of cell proliferation [1], and to facilitate metastasis, although more recent data and conclusions indicate that this facilitation of metastasis is an early-stage event [2]. In other words, there is some doubt about whether proteases facilitate late-stage cancer progression.

Over the last decade, there has been an explosion of bioinformatics approaches to cancer, particularly with the advent of large cancer sample databases and bioinformatics tools [5, 6]. Interestingly, protease biochemistry could be considered one of the earliest developments of bioinformatics in that the collection of protease-binding sites has been accumulating for almost a century, beginning with very early work with common enzymes, such as pepsin. These results are tabulated in, and have been available from,

Supplementary information The online version of this article (https://doi.org/10.1038/s41374-019-0248-3) contains supplementary material, which is available to authorized users.

✉ George Blanck
gblanck@health.usf.edu

[1] Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL, USA

[2] Immunology Program, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA

the MEROPS database for over 20 years [7, 8]. Nevertheless, the merging of protease bioinformatics and big data from cancer has been slow in coming, with but a few recent examples [9, 10].

In this report, we take advantage of automated application of the MEROPS data to melanoma mutant datasets to further understand how the melanoma mutants can impact protease sensitivity, particularly with regard to the extracellular matrix (ECM) and MMP2 and MMP14, two of the most well-studied proteases in the case of melanoma metastasis [11–13] and two proteases which are particularly a focus for drug development [14, 15]. The bioinformatics approaches representing automated application of MEROPS, MMP data were integrated with approaches for understanding cancer microenvironment immune activity and effects. Overall, the results indicated that the tools developed and employed in this report have the potential of identifying relationships between protease-mediated release of mutant peptides, the anti-cancer immune response, and associated survival rates among patients in large datasets. The results may assist in the explanation of a more recently appreciated conundrum in the cancer protease biology, namely that proteases appear to have their negative effect only early in cancer development.

Methods

Acquisition of the ECM structural protein amino acid (AA) substitution data from TCGA database

The entire collection of proteins in the ECM, i.e., the matrisome, has been previously, bioinformatically determined by Naba et al. and grouped into several distinct categories [16, 17]. In this study, we focused on the structural proteins of the ECM and thus removed regulatory factors and small secreted proteins, leaving a total of 445 ECM structural proteins for our analysis (Table S1). The ECM structural protein AA substitution data from the TCGA provisional SKCM and additional cancer datasets were downloaded from cbioportal.org [5, 6] (Table S2).

Identification of AA substitutions that affect protease binding in silico

The protease cleavage matrices for matrix metalloproteinase-2 (MMP2) and matrix metalloproteinase-14 (MMP14) were obtained from the MEROPS Sanger Database [7, 8] and saved in a comma-seperated values file format (Table S3). The previously acquired ECM structural protein mutant AA as well as the MMP2-binding matrix were used as input files for the script, *WholeProteanomeMutationAssesorV5.pl* (Table S4). A calculation was performed with the mutant amino acid (AA) in all positions of the matrix, respectively, to obtain a maximal score with the mutant AA, based on the MEROPS protease cleavage matrix. That mutant AA-based score is compared with the maximum score producible by the matrix. In our analysis, we used a threshold of 0.7 for all calculations, which refers to the mutant AA score representing at least 70% of the maximum possible score obtainable from the protease cleavage matrix. Example calculations are available in the SOM (Table S5). In essence, the software elucidates which mutant AA in the ECM structural proteins occur within four amino acids on either side of a protease cleavage site (with the protease-binding site defined as at least 70% of the maximum possible score obtainable from the MEROPS protease-binding matrix). Moreover, specifically for the Results section we also determined which AA represented mutants at sites with resistance to MMP2, with resistance being defined as ECM mutant peptide binding at 30%, or less, of the maximum possible MEROPS MMP2 protease-binding matrix score. Our analysis only investigated the impact of missense mutations. Furthermore, due to limitations of the software, only missense mutation data that did not represent AA substitutions within eight peptides of the terminal ends of a protein were used.

Kaplan–Meier outcomes analyses

Overall survival data, for the TCGA provisional SKCM and other TCGA datasets, were downloaded from www.cbioportal.org. Next, either TCGA barcodes representing mutations at MMP2-sensitive sites, or V(D)J recombination reads from exome files, as detailed below (in the Methods section), were matched to overall survival data associated with specific TCGA barcodes. Next, Kaplan–Meier survival curves were generated through Graphpad prism software, as further detailed in Results.

Identification of barcodes that represent pre-existing versus new MMP2-sensitive sites

To determine whether AA substitutions occurred at pre-existing MMP2-sensitive sites, or whether a new site was generated, the *WholeProteanomeMutationAssesorV5.pl* was applied using wild-type AA at the same site that AA substitutions occurred in the SKCM ECM structural proteins. Thus, barcodes were identified when using wild-type AA (in the mutant positions) as input that represent peptides that reached the 70% threshold of MMP2 sensitivity. Such identified barcodes have at least one "pre-existing" MMP2-sensitive site. Barcodes that were identified as having MMP2-sensitive sites using mutant AA, but that were not outputted with the wild-type analyses, were considered

barcodes with AA substitutions in ECM structural proteins that only created "new" MMP2-sensitive sites, i.e., barcodes that did not represent any pre-existing MMP2 sites. To be clear, these two sets of barcodes were thus separated into categories of "pre-existing" MMP2-sensitive sites (that may also include new MMP2-sensitive sites) and barcodes with only "new" MMP2-sensitive sites (i.e., barcodes that have no mutant AA substitutions at pre-existing MMP2-sensitive sites). For further details, see Results.

Obtaining TRA and TRB V(D)J recombination reads from TCGA provisional SKCM exome files

Whole-exome sequence (WXS) files of both primary and metastatic tumor samples were downloaded, in a binary alignment map (BAM) format, from the genomic data commons (GDC) to University of South Florida Research Computing using the GDC data transfer tool (https://gdc.cancer.gov/access-data/gdc-data-transfer-tool) with authorization via dbGaP approved project #6300 (Table S6, GDC download manifest). Recovery of immune receptor recombination reads was performed, as described previously [10] using original software available by e-mail to the corresponding author. The final list of productive V(D)J recombination reads for both TRA and TRB was further filtered, to ensure read fidelity, to include only reads with at least 90% nucleotide match fidelity and at minimum 19 nucleotide match length in both V and J regions (Tables S7, S8).

Determination of HLA class I alleles for the TCGA SKCM dataset and assessment of the HLA class I/ECM structural protein, mutant peptide-binding affinities

The determination of the HLA class I alleles was performed as described previously [10, 18]. The results of initial HLA typing from SKCM exome files were further verified by reassessing the HLA types with (a subset of) blinded files and by comparing the tumor- and blood-matched results. The SKCM HLA typing results are available by e-mail request to the corresponding author. HLA allele-matched binding affinities of mutant ECM structural peptides were determined, as described previously [9, 10]. The script used for the HLA-binding assessments is available in the SOM (Table S9). Furthermore, detailed output representing the binding peptides with an IC50 < 5000 nM, along with annotation is provided in the SOM (Table S10). HLA-binding predictions were verified using the IEDB analysis resource ANN tool [19, 20]. Manual verification of the HLA class I-binding results using the IEDB analysis resource ANN web tool are available in the SOM (Table S11). Barcodes with binding peptides were matched to TRA recombination reads; or with the presence of pre-existing or new cut sites, as detailed in the Results section. Statistical tests were conducted using the web tool, https://www.medcalc.org/calc/comparison_of_proportions.php.

Results

Sensitivity to MMP2 correlates with overall survival in SKCM

MMP2-sensitive barcodes (patient melanoma samples), defined as barcodes that have mutant ECM structural peptides (as defined in Methods) with sensitivity scores at least 70% of the highest possible MMP2-binding score, were identified via the *WholeProteanomeMutationAssesorV5.pl* (comprehensive output, Table S12). This processing revealed that there is an increased proportion of barcodes that represent ECM structural mutants that are present within four AAs of, i.e, adjacent to sensitive MMP2 sites, in comparison with barcodes where there were no mutant AA adjacent to the MMP2 sites (Table 1). Of 342 barcodes representing at least one mutant AA substitution in the ECM structural proteins, 197 barcodes represented at least one mutant AA substitution (in the ECM structural proteins) adjacent to an MMP2-sensitive site. One hundred and forty-five barcodes represented AA substitutions, whereby none of those substitutions was adjacent to an MMP2-sensitive site (Table 1). To be clear, the first group of 197 barcodes represented mutant AA adjacent to an MMP2-sensitive site as well as additional mutant AA among the ECM structural proteins, whereas the second group of 145 barcodes had no AA substitutions adjacent to an MMP2-sensitive site (see also summary table, for use of other MMP2 sensitivity thresholds, at the end of Table S12).

Table 1 SKCM barcodes representing ECM structural proteins with mutant AA adjacent or NOT adjacent to MMP2-sensitive sites

Number of SKCM barcodes with a mutant AA adjacent to an MMP2-sensitive site, defined as > 70% of the maximum score (Methods)	Number of SKCM barcodes with NO mutant AA adjacent to an MMP2-sensitive site (i.e., all barcodes with an ECM structural protein mutant AA not in left-side column)	Total number of SKCM barcodes with mutant AA in the ECM structural protein set	p-value representing the distinction between columns 1 and 2
197	145	342	0.0001

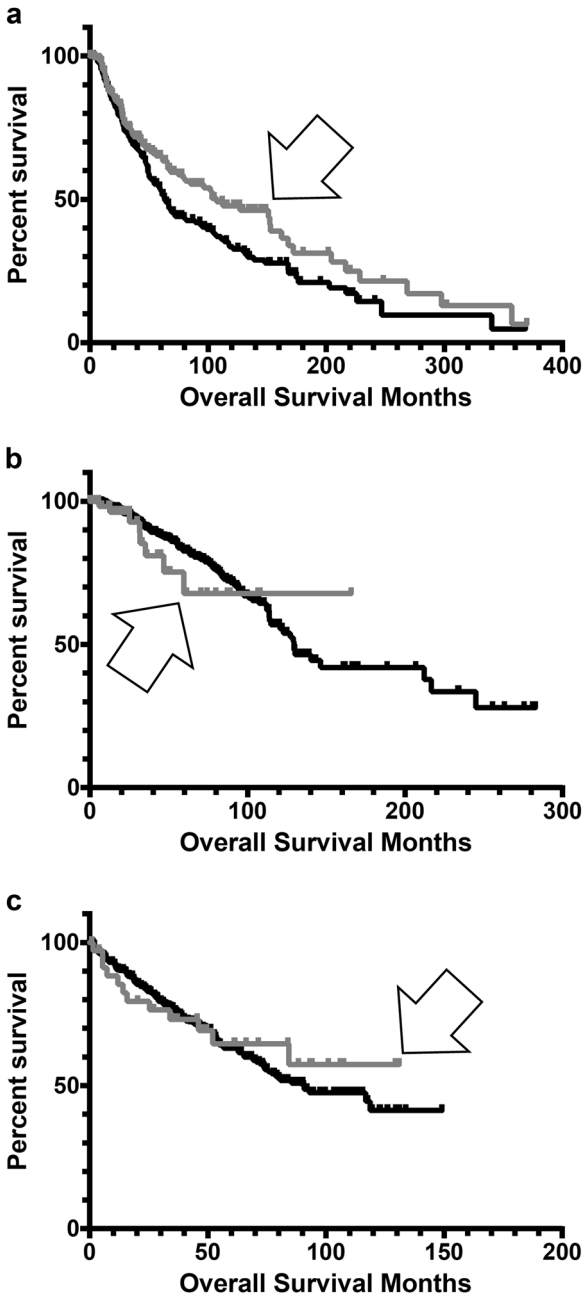

Fig. 1 Kaplan–Meier (KM) curves representing barcodes with mutant AA adjacent to MMP2 sites, for SKCM, BRCA, and KIRC. **a** KM overall survival (OS) curve for Skin Cutaneous Melanoma (SKCM) barcodes representing ECM structural mutant AA adjacent to MMP2-sensitive sites ($n = 197$, arrow), compared with the OS for all remaining barcodes ($n = 268$). Mean OS for the barcodes representing ECM structural mutant AA adjacent to MMP2-sensitive sites, 107.29 months; mean OS for all remaining barcodes, 63.30 months. Log rank comparison p-value, $p = 0.028$. **b** KM OS curve for Breast Invasive Carcinoma (BRCA) barcodes representing ECM structural mutant AA adjacent to MMP2-sensitive sites ($n = 64$, arrow), compared with the OS for all remaining barcodes ($n = 1032$). Mean OS for the barcodes representing ECM structural mutant AA adjacent to MMP2-sensitive sites, undefined months; mean OS for all remaining barcodes, 129.47 months. Log rank comparison p-value, $p = 0.478$. **c** KM OS curve for Kidney Renal Clear Cell Carcinoma (KIRC) barcodes representing ECM structural mutant AA adjacent to MMP2-sensitive sites ($n = 36$, arrow), compared with the OS for all remaining barcodes ($n = 500$). Mean OS for the barcodes representing ECM structural mutant AA adjacent to MMP2-sensitive sites, undefined months; mean OS for all remaining barcodes, 90.41 months. Log-rank comparison p-value, $p = 0.681$

To determine whether AA mutants adjacent to an MMP2-sensitive site were associated with improved survival outcomes, the barcodes that represented MMP2 sites with adjacent mutant AA were compared with all remaining barcodes in the dataset. The results indicated that the barcode group representing MMP2 site-adjacent AA substitutions also represented an increased overall survival (OS) rate (Fig. 1a, $p = 0.028$; Table S13).

To determine whether the above results were specific to SKCM, we performed the same analysis with BRCA, KIRC, COADREAD, and LUAD ECM structural protein mutants. In the case of each dataset, the barcodes that represented mutant AA adjacent to MMP2-sensitive sites were compared with all remaining barcodes in the dataset (Fig. 1b, c; Figs. S1, S2). A significant increase in OS was only observed in the case of SKCM. (See also Fig. S3, indicating a trend for increased survival of barcodes representing the 0.6 sensitivity threshold, for the MMP2 MEROPS matrix, i.e., representing 60% of the maximal MMP2 MEROPS score, whereby such barcodes have mutant AA adjacent to the MMP2 sites as defined by the 0.6 threshold.)

The ECM structural protein, AA mutants adjacent to MMP2-sensitive sites, AA mutants creating new MMP2-sensitive sites, and AA mutants not adjacent to any MMP2-sensitive sites

To determine whether barcodes representing AA substitutions adjacent to MMP2-sensitive sites correlated with a different survival pattern than barcodes representing AA substitutions that were not adjacent to an MMP2-sensitive site (Table 1), we compared the OS of these two barcode groups, with results indicating that barcodes representing ECM structural protein mutant AA adjacent to an MMP2-sensitive site also represented a higher OS (Fig. 2a, $p = 0.0016$).

Next, barcodes were identified as having mutant AA adjacent to pre-existing MMP2-sensitive sites or, exclusively, mutant AA adjacent to "new" MMP2-sensitive sites, i.e., generated by the mutant AA, as indicated in Methods (Table S14). To determine whether barcodes representing

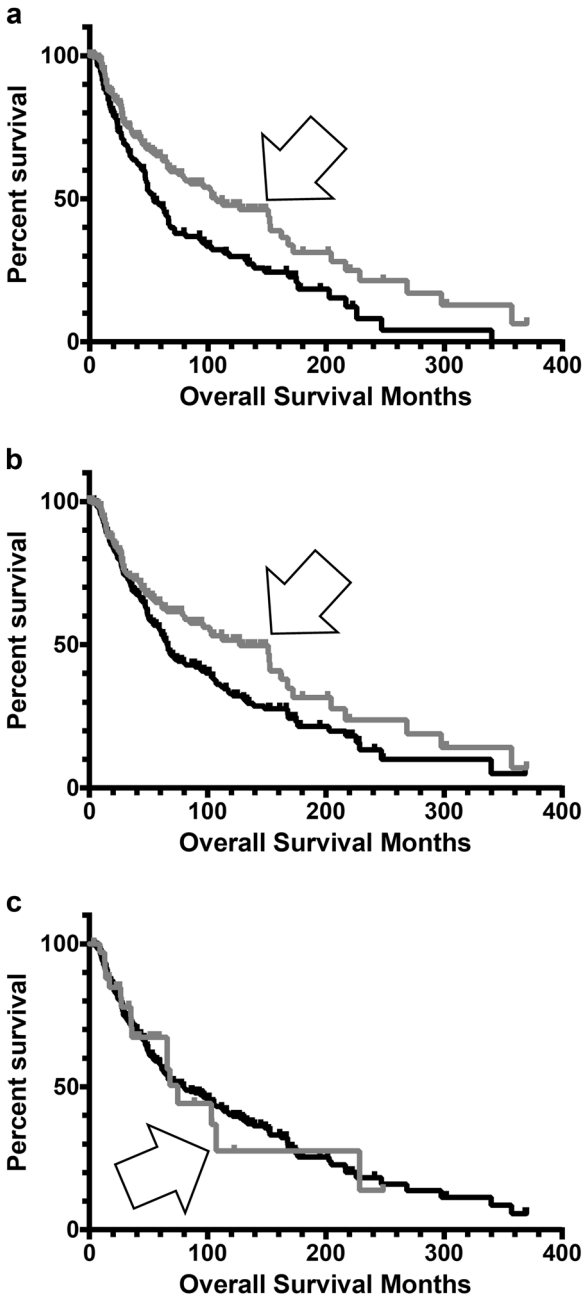

Fig. 2 Kaplan–Meier analyses represented by barcodes representing mutant AA adjacent to MMP2 sites versus mutant AA not adjacent to MMP2 sites. **a** KM OS curve for Skin Cutaneous Melanoma (SKCM) barcodes representing ECM structural mutant AA adjacent to MMP2-sensitive sites ($n = 197$, arrow), compared to the OS for barcodes with ECM structural mutant AA not adjacent to MMP2-sensitive sites ($n = 143$, value differs from Table 1 due to lack of survival data for two barcodes). Mean OS for the barcodes representing ECM structural mutant AA adjacent to MMP2-sensitive sites, 107.29 months; mean OS for the barcodes representing ECM structural mutant AA not adjacent to MMP2-sensitive sites, 53.88 months. Log rank comparison p-value, $p = 0.0016$. **b** KM OS curve for Skin Cutaneous Melanoma (SKCM) barcodes representing ECM structural mutant AA adjacent to pre-existing MMP2-sensitive sites ($n = 163$, arrow), compared to all remaining barcodes ($n = 302$). Mean OS for the barcodes representing ECM structural mutant AA adjacent to pre-existing MMP2-sensitive sites, 127.10 months; mean OS for all remaining barcodes, 65.87 months. Log-rank comparison p-value, $p = 0.0228$. **c** KM OS curve for Skin Cutaneous Melanoma (SKCM) barcodes representing ECM structural mutant AA adjacent to new MMP2-sensitive sites ($n = 34$, arrow), compared with all remaining barcodes ($n = 428$). Mean OS for the barcodes representing ECM structural mutant AA adjacent to new MMP2-sensitive sites, 74.67 months; mean OS for all remaining barcodes, 79.53 months. Log-rank comparison p-value, $p = 0.9697$

mutant AA adjacent to pre-existing MMP2-sensitive sites have a different OS than barcodes exclusively representing mutant AA adjacent to new MMP2-sensitive sites, the two barcode groups were independently compared with all remaining barcodes. The comparison of barcodes that represent mutant AA adjacent to pre-existing MMP2-sensitive sites to all remaining barcodes revealed that barcodes representing mutant AA adjacent to pre-existing MMP2-sensitive sites had a better OS (Fig. 2b, $p = 0.0228$). Comparing barcodes that represented only MMP2-sensitive sites generated from mutant AA substitutions, or new MMP2-sensitive sites, to all remaining barcodes did not reveal a statistically significant difference in overall survival (Fig. 2c, $p = 0.9697$).

The overlap of V(D)J recombination reads and ECM structural protein mutants

To determine whether any of the previously indicated survival distinctions based on mutant AA adjacent to MMP2-sensitive sites could be immunological in nature, V(D)J recombination reads for TRA and TRB were extracted from SKCM WXS files as indicated in Methods (Tables S7, S8). The results indicated that there were no significant differences in TRA or TRB recovery when comparing barcodes representing ECM structural protein AA mutants adjacent to MMP2-sensitive sites and ECM structural protein AA mutants not adjacent to MMP2 cut sites. Moreover, there were no significant differences in TRA or TRB recovery when comparing barcodes representing ECM structural proteins with mutant AA adjacent to pre-existing cut sites versus barcodes representing ECM structural mutant AA only at new cut sites (Table 2).

Next, we determined that TRA recombination was independently associated with improved survival outcomes among the SKCM barcodes ($p < 0.0001$, Fig. 3a). Given the association between TRA read recombination and improved survival outcomes in SKCM, we determined whether barcodes representing ECM structural protein mutant AA adjacent to an MMP2-sensitive site could be further subdivided based on their TRA recombination read recovery status. Thus, the OS represented by barcodes representing ECM structural mutant AA

Table 2 Recovery of TCR V(D)J recombination reads from SKCM WXS files: proportions for barcodes representing ECM structural protein mutant AA adjacent to an MMP2-sensitive site

	Fraction of total SKCM WXS with the indicated recombination read recovery	Fraction of WXS files, with the indicated recombination read recovery, representing barcodes with mutant AA adjacent to an MMP2-sensitive site	Fraction of WXS files, with the indicated recombination read recovery, representing barcodes with mutant AA NOT adjacent to an MMP2-sensitive site	Fraction of WXS files, with the indicated recombination read recovery, representing barcodes with mutant AA adjacent to a new MMP2-sensitive site (i.e., a site dependent on an adjacent mutant AA)	Fraction of WXS files, with the indicated recombination read recovery, representing barcodes with mutant AA adjacent to an MMP2-sensitive site but not including barcodes that have ONLY new MMP2-sensitive sites (which are sites dependent on an adjacent mutant AA)
TRA	0.605	0.644	0.572	0.705	0.632
TRB	0.554	0.588	0.510	0.588	0.589
Maximum possible WXS files	471	197	145	34	163

adjacent to an MMP2-senstive site, and with TRA recombination read recovery, was compared with the OS of barcodes representing ECM structural protein mutant AA adjacent to an MMP2-sensitive site, but with no TRA recombination read recovery. The results indicated that barcodes representing the mutant AA adjacent to an MMP2-sensitive site and TRA recombination read recovery had improved OS rates ($p = 0.0266$, Fig. 3b). The full Graphpad Prism output for Fig. 3b is available in the SOM (Table S15).

Sample-specific HLA class I alleles and peptide/HLA class I-binding analyses for ECM structural mutant AA adjacent to MMP2-sensitive sites

We have previously determined that HLA class I-binding affinities of the mutant cytoskeleton and ECM peptides that had increased sensitivity to certain proteases in melanoma and glioblastoma multiforme were different from the binding affinities of the mutant peptides that represented decreased protease sensitivity [9, 10]. Thus, in this analysis, we determined whether mutant AA of ECM structural proteins adjacent to MMP2-sensitive sites with higher HLA class I affinities were more common among the TRA-positive barcodes. The HLA class I alleles were determined for each SKCM barcode (Methods), and binding affinities of the mutant AA and the barcode-specific HLA-A and HLA-B allelic variants were elucidated. Thus, we determined the HLA class I-binding affinities of the samples used for Fig. 3b, namely barcodes representing ECM structural protein mutant AA adjacent to an MMP2-sensitive site and with TRA recovery; and of the barcodes representing ECM structural mutant AA adjacent to an MMP2-senstive site, but with no TRA recovery from the corresponding WXS files (Table S16). The results indicated that there were more HLA allele-specific class I-binding epitopes among the ECM structural mutant peptides with TRA recombination read recovery than without TRA recombination read recovery ($p = 0.0251$, Table 3). Furthermore, the barcodes representing the ECM structural mutant peptides with TRA recombination read recovery also represented a much higher level of CD4 and CD8 RNAseq values, in comparison with all remaining SKCM barcodes, validating, as in numerous other past studies, the linkage of TRA recombination read recoveries and T-cell expression markers.

Next, we determined the HLA class I-binding affinities of the barcodes that represented ECM structural protein mutant AA adjacent to pre-existing MMP2-sensitive sites, as well as for barcodes that have only new MMP2-sensitive sites (i.e., sites dependent on ECM structural protein mutant AA, discussed in detail above). The results indicated that there were more HLA class I-binding epitopes among the barcodes representing ECM structural protein mutant peptides adjacent to pre-existing MMP2-sensitive sites ($p = 0.06$, Table 4), although the statistical analysis only established a trend, rather than statistical significance.

The ECM structural protein mutant AA and sensitivity of MMP14

The *WholeProteanomeMutationAssesorV5.pl* software was used to determine the presence of AA substitutions adjacent to MMP14-sensitive sites, as defined by sites that were above 70% of the maximal score indicated by

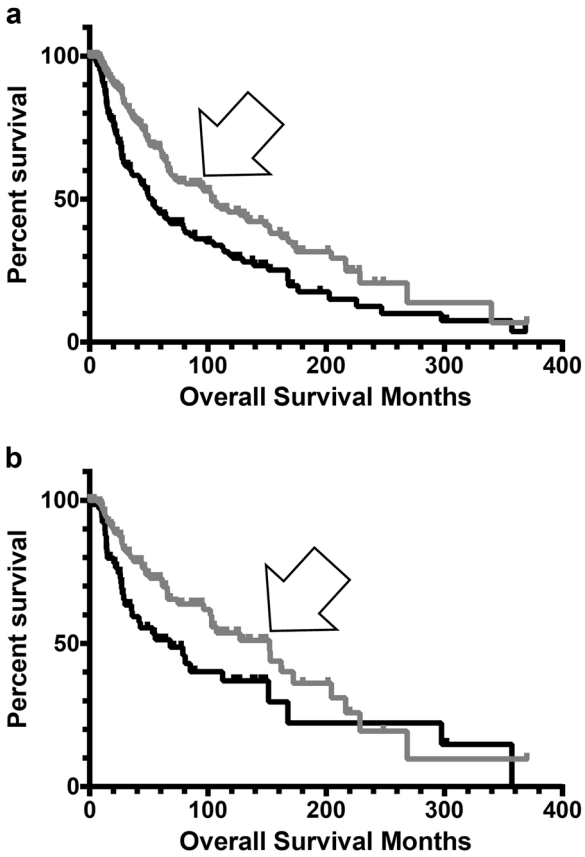

Fig. 3 Kaplan–Meier (KM) analyses representing barcodes with or without TRA recombination read recovery and ECM structural mutant AA adjacent to MMP2-sensitive sites. **a** KM OS curve for Skin Cutaneous Melanoma (SKCM) barcodes representing TRA recombination read recovery ($n = 284$, arrow), compared to the OS for barcodes without TRA recombination read recovery ($n = 184$). Mean OS for the barcodes representing TRA recombination read recovery, 103.19 months; mean OS for barcodes without TRA recombination read recovery, 53.15 months. Log-rank comparison p-value, $p < 0.0001$. **b** KM OS curve for Skin Cutaneous Melanoma (SKCM) barcodes representing ECM structural mutant AA adjacent to an MMP2-senstive site with TRA recombination read recovery ($n = 127$, arrow), compared with the OS for barcodes representing ECM structural mutant AA adjacent to an MMP2-senstive site without TRA recombination read recovery ($n = 69$). Mean OS for the barcodes representing ECM structural mutant AA adjacent to an MMP2-senstive site with TRA recombination read recovery, 152.23 months; mean OS for barcodes representing ECM structural mutant AA adjacent to an MMP2-senstive site without TRA recombination read recovery, 68.1 months. Log-rank comparison p-value, $p = 0.0266$

the MEROPS MMP14-binding matrix. The full protease digestion output for MMP14 is available in the SOM (Table S17). To determine whether barcodes representing AA substitutions adjacent to MMP14-sensitive sites represented a different survival pattern than barcodes representing AA substitutions that were not adjacent to an MMP14-sensitive site, we compared the OS of these two barcode groups, with results indicating that barcodes representing ECM structural protein mutant AA adjacent to an MMP14-sensitive site also represented a higher OS (Fig. 4; $p = 0.001$). The full Graphpad Prism output for this result is available in the SOM (Table S18).

Sample-specific HLA class I-binding comparison for ECM structural protein, mutant AA adjacent to MMP2-sensitive sites versus ECM structural protein, mutant AA adjacent to MMP2-resistant sites

To determine whether the ECM structural mutants adjacent to relatively resistant MMP2 sites may be good binders to HLA class I, HLA class I-binding affinities for all peptides adjacent to resistant MMP2 sites were determined, as described in Methods (Table S19). The results indicated that there is an increased proportion of HLA-binding peptides with mutant AA adjacent to MMP2-resistant sites, in comparison with peptides represented by mutant AA adjacent to MMP2-sensitive sites (Table 5; $p = 0.0077$). When barcodes with only MMP2-resistant sites are compared with all remaining SKCM barcodes in a KM analysis, the barcodes representing the MMP2-resistent sites also represent a significantly worse overall survival (log rank, $p = 0.002$).

Discussion

The above results indicated that patients with ECM structural protein mutants adjacent to MMP2- and MMP14-sensitive sites had a better overall survival than patients not having mutant AA adjacent to those protease sites. This result is consistent with a higher mutation burden in the barcode set representing mutant AA adjacent to MMP2

Table 3 HLA class I allele-matched epitope totals for barcodes representing mutant AA adjacent to MMP2-sensitive sites with or without TRA recovery (using < 5000 nM as a cut off value for HLA-binding affinity; Table S16)

	HLA class I epitopes, among peptides with mutant AA adjacent to MMP2-sensitive sites and with TRA recombination read recovery	HLA class I epitopes, among peptides with mutant AA adjacent to MMP2-sensitive sites but without TRA recombination read recovery	p-value
Proportion	0.401	0.242	$P = 0.0251$
Number of HLA class I-binding epitopes	51	17	
Number of total barcodes	127	70	

Table 4 HLA class I allele-matched epitope totals for barcodes representing mutant AA adjacent to pre-exiting MMP2-sensitive sites versus new MMP2-sensitive sites (using < 5000 nM as a cutoff value for HLA-binding affinity)

	HLA class I epitopes, among peptides with mutant AA adjacent to pre-existing MMP2-sensitive sites	HLA class I epitopes, among peptides with mutant AA adjacent to new MMP2-sensitive sites	*p*-value
Proportion	0.374	0.205	$P = 0.0600$
Number of HLA class I-binding epitopes	61	7	
Number of total barcodes	163	34	

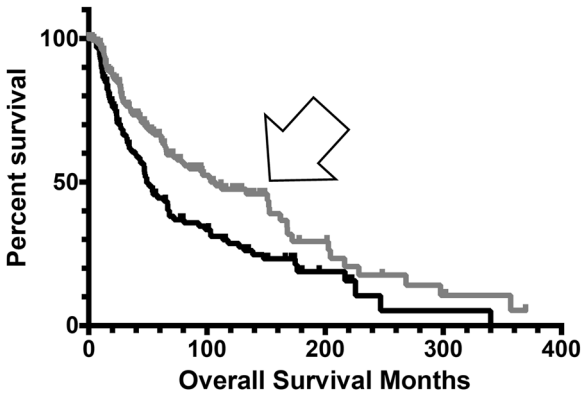

Fig. 4 Kaplan–Meier (KM) curves representing barcodes with mutant AA adjacent to MMP14 sites. KM overall survival (OS) curve for Skin Cutaneous Melanoma (SKCM) barcodes representing ECM structural mutant AA adjacent to MMP14-sensitive sites ($n = 210$, arrow), compared with the OS for barcodes with ECM structural mutant AA not adjacent to MMP14-sensitive sites ($n = 130$). Mean OS for the barcodes representing ECM structural mutant AA adjacent to MMP14-sensitive sites, 103.12 months; mean OS for barcodes with ECM structural mutant AA not adjacent to MMP14-sensitive sites, 48.85 months. Log-rank comparison *p*-value, $p = 0.0010$

(and MMP14)-sensitive sites (Table S20), in that melanoma patients with a higher mutation burden have a better prognosis, particularly with regard to immune checkpoint blockade responses. Thus, at a minimum, the above results indicate a biomarker surrogate for mutation burdens, and the above results may point to a biochemically practical approach to rapidly and very inexpensively identifying, with protease technologies, patients with relatively high mutation burdens.

However, the above results also offer starting points, i.e., the development of mechanistic hypotheses for why a high mutation burden, and the adjacency of mutant AA to protease sensitive sites, may lead to a better outcome. In particular, the above results were linked to the observation that samples representing better HLA class I-binding mutants, that were adjacent to MMP2 sites, had a greater level of T-cell infiltrate than did samples with statistically, significantly fewer good HLA-binding mutant adjacent to the MMP2 sites. Thus, these data are consistent with the hypothesis that MMP2 has the capacity to make available mutant epitopes that facilitate a T-cell response to the tumor.

However, it is important to point out that the measure of the T-cell infiltrate, the recovery of TRA and TRB recombination reads from the melanoma exome files, did not differ in the cases of barcodes with mutants adjacent to MMP2 sites versus barcodes representing no mutants adjacent to MMP2 sites. The lack of differences in T-cell infiltrates between the barcodes with mutant AA adjacent to an MMP2 site and barcodes without such mutant AA, among the ECM structural proteins, would raise the question of what is the source of good HLA class I binders in non-MMP2 cases, or in any event, what is facilitating the T-cell infiltrate? The answer to such a question is beyond the scope of this work, but either intracellular or extracellular sources may be possible, including sources facilitated by other proteases. Or, the infiltrate may not be related to epitope availability or to an ongoing anti-cancer, cytotoxic T-cell immune response, a possibility consistent with the reduced OS represented by barcodes with mutant AA not adjacent to an MMP2 site.

This study also investigated barcodes representing only new MMP2-binding sites, generated by mutant AA. These sites represented a trend toward a reduce level of good HLA class I binders ($p = 0.06$), in comparison with mutant AA adjacent to wild-type MMP2 sites. This results may be consistent with the constraints on being both a good HLA class I-binding peptide and a good MMP2 site. Thus, for both the cases of the new MMP2 sites, and the mutant AA not adjacent to MMP2 sites at all, there is an indication that the T-cell infiltrate is not facilitated by good HLA class I binders among the ECM structural proteins mutants.

Finally, the apparent, good, HLA binding peptides, that were represented by mutant AA and were associated with MMP2 resistance (Table 5), further supports the importance of further inquiry into the possibility that lack of protease mediated availability of good HLA class I binders impedes the anti-tumor immune response.

Table 5 HLA class I allele-matched epitope totals for barcodes representing ECM structural mutant AA adjacent to MMP2-sensitive sites versus ECM structural mutant AA adjacent to MMP2-resistant sites (using < 5000 nM as a cut off value for HLA-binding affinity)

	HLA class I epitopes, among peptides with mutant AA adjacent to MMP2-sensitive sites	HLA class I epitopes, among peptides with mutant AA adjacent to MMP2-resistant sites	*p*-value
Proportion	0.139	0.187	$P = 0.0077$
Number of HLA class I-binding epitopes	68	1726	
Number of total mutants	489	9206	

Acknowledgements The authors wish to thank USF research computing and the taxpayers of the State of Florida. SZ, BIC, MMM, and AD were recipients of USF Morsani College of Medicine RISE fellowships.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Lipkin G, Rosenberg M, Knecht ME. Factors affecting growth of normal and malignant cells in vitro. Biochem Pharmacol. 1976;25:1333–7.
2. Winer A, Adams S, Mignatti P. Matrix metalloproteinase inhibitors in cancer therapy: turning past failures into future successes. Mol Cancer Ther. 2018;17:1147–55.
3. Santibanez JF, Obradovic H, Kukolj T, Krstic J. Transforming growth factor-beta, matrix metalloproteinases, and urokinase-type plasminogen activator interaction in the cancer epithelial to mesenchymal transition. Dev Dyn. 2018;247:382–95.
4. Aponte-Lopez A, Fuentes-Panana EM, Cortes-Munoz D, Munoz-Cruz S. Mast cell, the neglected member of the tumor microenvironment: role in breast cancer. J Immunol Res. 2018;2018:2584243.
5. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:pl1.
6. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–4.
7. Rawlings ND. Peptidase specificity from the substrate cleavage collection in the MEROPS database and a tool to measure cleavage site conservation. Biochimie. 2016;122:5–30.
8. Rawlings ND, Barrett AJ. MEROPS: the peptidase database. Nucleic Acids Res. 1999;27:325–31.
9. Callahan BM, Patel JS, Fawcett TJ, Blanck G. Cytoskeleton and ECM tumor mutant peptides: Increased protease sensitivities and potential consequences for the HLA class I mutant epitope reservoir. Int J Cancer. 2018;142:988–98.
10. Zaman S, Chobrutskiy BI, Patel JS, Callahan BM, Tong WL, Blanck G. Mutant cytoskeletal and ECM peptides sensitive to the ST14 protease are associated with a worse outcome for glioblastoma multiforme. Biochem Biophys Res Commun. 2018;503:2218–25.
11. Dahal U, Kang L, Gupta M. RNA m6A methyltransferase METTL3 regulates invasiveness of melanoma cells by matrix metallopeptidase 2. Melanoma Res. 2019.
12. Redmer T. Deciphering mechanisms of brain metastasis in melanoma—the gist of the matter. Mol Cancer. 2018;17:106.
13. Pekkonen P, Alve S, Balistreri G, Gramolelli S, Tatti-Bugaeva O, Paatero I, et al. Lymphatic endothelium stimulates melanoma metastasis and invasion via MMP14-dependent Notch3 and beta1-integrin activation. Elife 2018;7:e32490.
14. Marusak C, Bayles I, Ma J, Gooyit M, Gao M, Chang M, et al. The thiirane-based selective MT1-MMP/MMP2 inhibitor ND-322 reduces melanoma tumor growth and delays metastatic dissemination. Pharmacol Res. 2016;113:515–20.
15. Pietraszek K, Chatron-Colliet A, Brezillon S, Perreau C, Jakubiak-Augustyn A, Krotkiewski H, et al. Lumican: a new inhibitor of matrix metalloproteinase-14 activity. FEBS Lett. 2014;588:4319–24.
16. Naba A, Clauser KR, Whittaker CA, Carr SA, Tanabe KK, Hynes RO. Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. BMC Cancer. 2014;14:518.
17. Naba A, Clauser KR, Lamar JM, Carr SA, Hynes RO. Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. Elife. 2014;3:e01308.
18. Callahan BM, Yavorski JM, Tu YN, Tong WL, Kinskey JC, Clark KR, et al. T-cell receptor-beta V and J usage, in combination with particular HLA class I and class II alleles, correlates with cancer survival patterns. Cancer Immunol Immunother. 2018;67:885–92.
19. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res. 2008;36:W509–12.
20. Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 2003;12:1007–17.