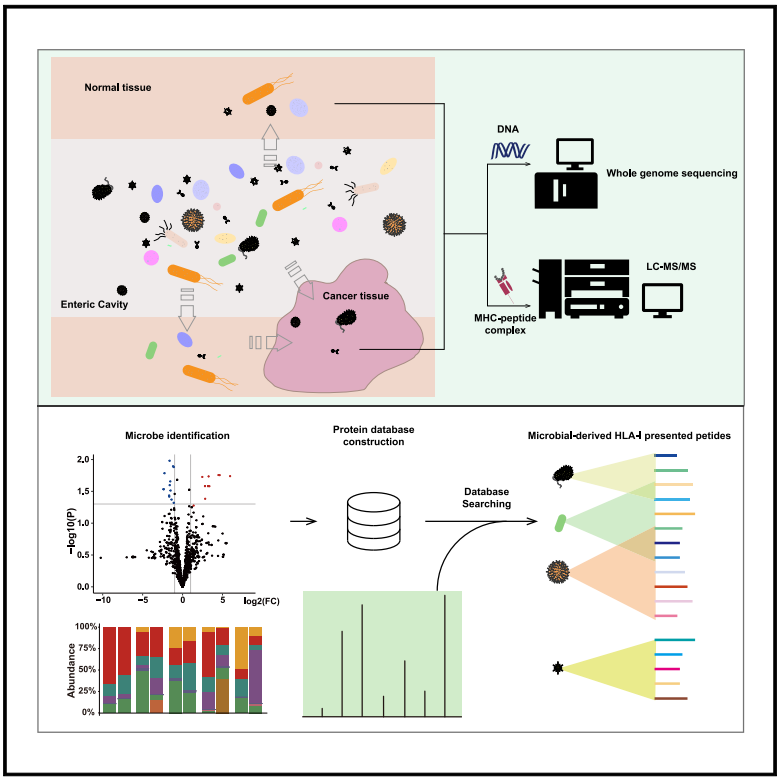
iScience

Article

Immunogenic peptides putatively from intratumor microbes: Opportunities for colorectal cancer treatment

Graphical abstract



Highlights

- Identification of intratumoral microbes through WGS data
- Immunogenic microbe-derived peptides are presented by HLA-I in tumors
- *Fusobacterium nucleatum* and its HLA-I-presented peptides are enriched in tumors

Authors

Xiangyu Guan, Fanyu Bu, Yunyun Fu, ..., Kui Wu, Longqi Liu, Xuan Dong

Correspondence

liulongqi@genomics.cn (L.L.), dongxuan@genomics.cn (X.D.)

In brief

Immune response; Microbiome; Cancer

Guan et al., 2024, iScience *27*, 111338
December 20, 2024
https://doi.org/10.1016/j.isci.2024.111338

CellPress

iScience

CellPress
OPEN ACCESS

Article

Immunogenic peptides putatively from intratumor microbes: Opportunities for colorectal cancer treatment

Xiangyu Guan,[1,2,3] Fanyu Bu,[2] Yunyun Fu,[1,2] Haibo Zhang,[1,2] Haitao Xiang,[2] Xinle Chen,[2,4] Tai Chen,[5] Xiaojian Wu,[6] Kui Wu,[2,3,7,8] Longqi Liu,[2,3,*] and Xuan Dong[2,3,7,8,9,*]

[1]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[2]BGI Research, Hangzhou 310030, China
[3]BGI Research, Shenzhen 518083, China
[4]Center for Mitochondrial Biology and Medicine, The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, China
[5]BGI Research, Changzhou 213299, China
[6]The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong 510655, China
[7]Guangdong Provincial Key Laboratory of Human Disease Genomics, Shenzhen Key Laboratory of Genomics, BGI Research, Shenzhen 518083, China
[8]HIM-BGI Omics Center, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences (CAS), Hangzhou 310022, China
[9]Lead contact
*Correspondence: liulongqi@genomics.cn (L.L.), dongxuan@genomics.cn (X.D.)
https://doi.org/10.1016/j.isci.2024.111338

SUMMARY

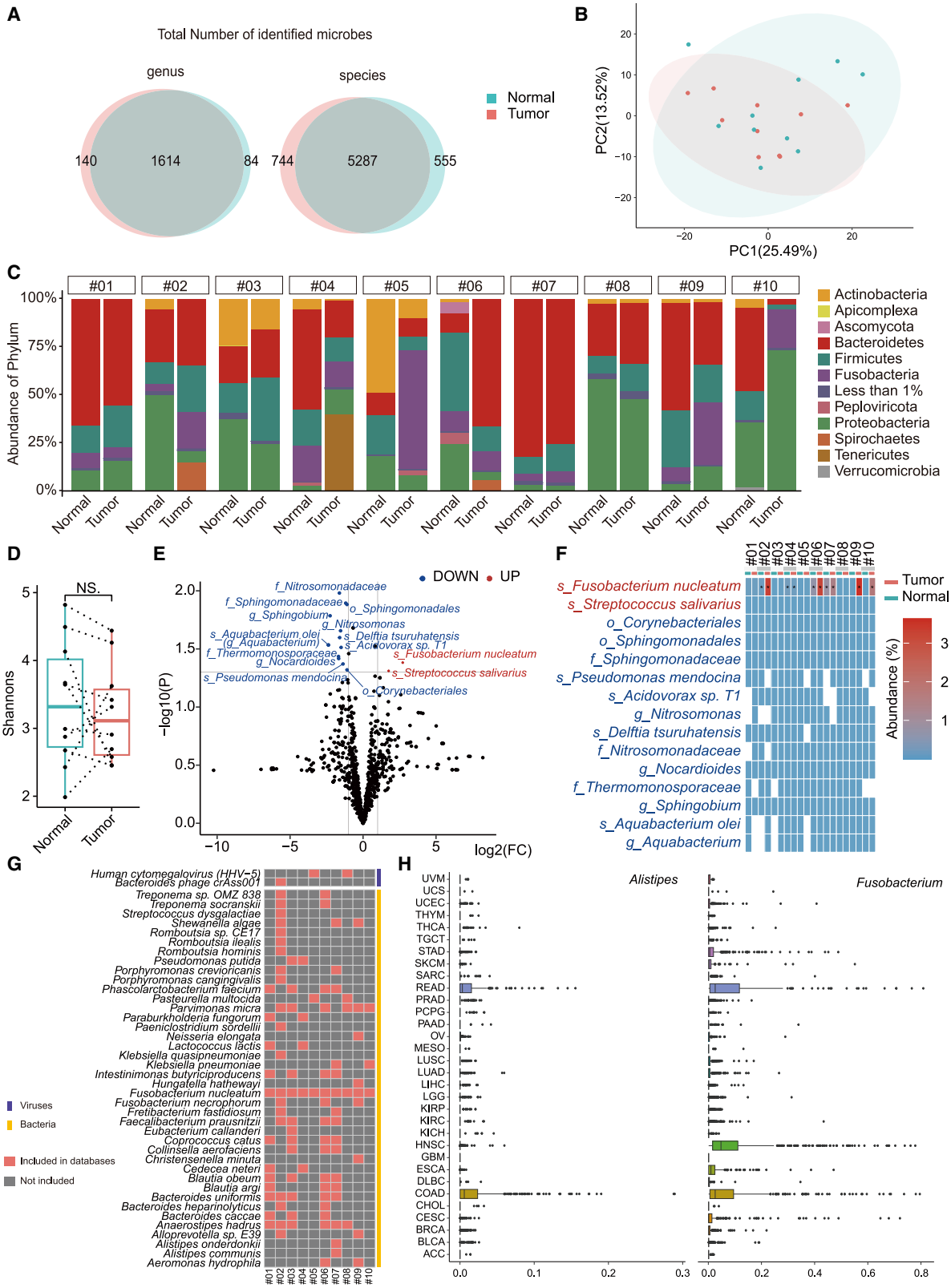
Recent evidence has confirmed the presence of intratumor microbes, yet their impact on the immunopeptidome remains largely unexplored. Here we introduced an integrated strategy to identify the immunopeptidome originated from intratumor microbes. Analyzing 10 colorectal cancer (CRC) patients, we identified 154 putative microbe-derived human leukocyte antigen (HLA)-I ligands. Predominantly bacterial in origin, these peptides were notably abundant in *Fusobacterium nucleatum*, the most prevalent bacterium differentiating between normal and tumor tissues. We discovered 20 peptides originating from *F. nucleatum*, thirteen of which, including two peptides shared across multiple patients, were tumor specific. Validation experiments confirmed that the putative microbe-derived peptide could activate $CD8^+$ T cell responses. Our findings indicate that HLA-I molecules are capable of presenting intratumor microbe-derived peptides in CRC, potentially contributing to $CD8^+$ T cell-mediated immunity and suggesting potential strategies for cancer immunotherapy.

INTRODUCTION

The tumor microenvironment (TME) undergoes alterations during the genesis and development of cancer, which are believed to influence the outcome of cancer therapy, especially with regard to cancer immunotherapy.[1] Immunological therapies, such as immune checkpoint inhibitors (ICIs), immune cell therapy, and tumor therapeutic vaccines have been used in cancer treatment.[2–4] However, the effectiveness of ICI in solid tumors remains unsatisfactory due to their low immunogenicity, lack of tumor-infiltrating lymphocytes, and other immune escape mechanisms or features of tumor cells.[5–9] To improve the effectiveness of immunotherapy, it is crucial to reactivate and enhance the immune system's ability to recognize and kill tumor cells.

The immunopeptidome refer to the complete set of peptide segments presented by antigen-presenting cells in the immune system.[10,11] These peptides are typically derived from the degradation of antigen proteins, cleaved into shorter peptide segments by intracellular proteases, and presented to immune cells via human leukocyte antigen (HLA) molecules, also known as major histocompatibility complex (MHC) molecules.[12,13] Studying the immunopeptidome helps us understand how the immune system recognizes and attacks tumor cells or infected cells.[14–16] By analyzing the immunopeptidome, we can identify peptide segments that bind to HLA molecules and further investigate the interaction between these complexes and specific immune cells.[17] Recent studies have shown that the immunopeptidome can be used to discover tumor-specific antigens (TSAs) expressed in tumor cells, which engage the immune system.[18] By analyzing the immunopeptidome in tumor tissues, it is possible to identify peptide segments associated with tumors, thereby providing potential targets for the development of tumor immunotherapy.[19–21] Overall, research on the immunopeptidome reveals important processes of immune recognition and response, offering possible insights and targets for immune therapy, vaccine development, and personalized medicine.

The polymorphism of microbiomes has been considered one of the hallmarks of cancer.[22] The microbes colonizing in gut,



(legend on next page)

mucosa, and other tissues can either induce or inhibit tumor growth, and inflammation, and influence therapy outcomes by interacting with other tumor characteristics, such as tumor-promoting inflammation, genomic instability, and mutation.[23–27] With the proven existence of intratumor microbe and their relationship with tumorigenesis, invasion, and metastasis, these microbes may perform additional functions in the TME.[28–33] The local interaction of microbes with host cells and immune systems through HLA and microbiome-derived peptides should be explored to further understand TME alterations.

Here, we developed a workflow to identify intratumoral microbes and related HLA-I peptides. Paired samples of tumor and adjacent normal tissues acquired from ten colon rectal cancers (CRCs) were analyzed. *Fusobacterium nucleatum* and *Streptococcus salivarius* were significantly enriched in tumor samples, and 154 putative microbe-derived peptides were identified from a total of 14,260 HLA-I ligands. These peptides may function as HLA-I ligands with high immunogenicity in the local region, interacting with tumor cells and antigen-presenting cells, thereby potentially influencing the immune status in the TME.

RESULTS

Bacteria overwhelmingly dominated the microbe in CRC samples with *Fusobacteria* enriched in tumors

To investigate the contribution of microbes to host immunopeptidome, we first identified the microbial communities in 10 patients diagnosed with CRC (Table S1). Based on the analyzing of the host whole-genome sequencing (WGS) data with an average sequencing depth of 87.36-fold (Table S2), we found 6.74%–44.37% of unaligned reads were of microbe origin, ultimately identifying 1,838 genera and 6,568 species. Of these, 80.50% and 87.81% were shared between tumor and normal samples, respectively (Figure 1A). Principal component analysis (PCA) of microbiome relative abundance and *Bray-Curtis* beta diversity revealed a slight difference between normal and tumor samples (Figures 1B and S1A). At the phylum level, *Bacteroidetes*, *Proteobacteria*, and *Firmicutes* were dominant except in patient #05, while *Spirochaetes* and *Tenericutes* were more abundant in patient #02 and #04 than in paired adjacent samples (Figure 1C).

Despite detecting more microbial taxa in tumors (Figure S1B), tumor samples exhibited lower alpha diversity indicated by Shannon and Simpson indices (not significant, Figures 1D and S1C), suggesting specific microbial-TME interactions that promote colonization and result in higher richness but lower diversity within tumors.[34,35] Bacteria overwhelmingly dominated the microbiome (>90%), even though viruses and fungi were also considered during microbiome identification. *Fusobacteria*, particularly *F. nucleatum*, and *Streptococcus salivarius*, were significantly enriched in tumor samples (Figure 1E and Table S3), with *F. nucleatum* being the only species exceeding 1% abundance among all significantly differential species (Figure 1F).

To comprehensively identify peptides from microbes, dominant microbe species and significant enrichment species identified from tumor sample were included in construction of personalized databases. A totally of forty-two species across all samples were selected based on whether known proteomes has been recorded in UniProt (Figure 1G). These species could be classified into thirty-two genera, twenty-five of which were reported in the BIC database.[36] Compared with thirty other cancer types, ten genera, such as *Alistipes*, *Eubacterium*, and *Fusobacterium*, showed higher abundance in rectum adenocarcinoma (READ) and colon adenocarcinoma (COAD). *Pseudomonas* and *Streptococcus* were commonly found across multiple cancer types (Figures 1I and S1D; Table S4). Additionally, twenty-one species have been reported to be associated with cancer, particularly with CRC (Table 1). The enrichment and dominance of these microbes raise the possibility of their peptides being captured and presented by HLA complex.

HLA-I ligands identification with host microbe information and predominantly originate from human self-protein

We next established a workflow identifying HLA-I ligands isolated by immunoprecipitation with an HLA-I-specific antibody in tumor samples and paired adjacent normal tissues (Figure 2A). Microbes significantly enriched in all tumor samples or with higher relative abundance than paired normal samples were selected. Annotated proteins originating from these microbes and *Homo sapiens*, downloaded from the UniProt database, were merged to create a personalized protein database (PPD) for database searching of LC-MS/MS spectra analyzed using MaxQuant.[83] Peptides meeting stringent criteria, with a false discovery rate[84] (FDR) under 1% and posterior error probability (PEP) below 0.05, were regarded as HLA-I ligands.

In our study, samples ranging from 130.3 to 779.4 mg were used for the immunoprecipitation of HLA-I ligands (Table S5). During the personalized database construction process, 50,572 to 114,310 proteins from 4 to 23 microbe species and human self-proteins were included for each sample (Figure 2B). An

Figure 1. Microbiome profiles in paired tumor and normal tissue samples

(A) Number of microbiomes identified at genus and species levels from different samples.

(B) Principal-component analysis (PCA) based on the relative abundance of all identified microbes.

(C) Relative abundance in phylum.

(D) Shannon's index of alpha diversity within samples. A two-tailed paired Student's t-test were conducted among normal and tumor samples, no significance (NS) was found.

(E) Difference analysis of microbes between tumor and normal samples. Microbes that significantly enrichment in tumor samples highlighted in red and those decreased in blue. *p* values were calculated using a two-tailed Student's t-test.

(F) The abundance of significantly different microbes; relative abundance higher than 1% is marked by an asterisk. Taxonomic levels are represented by prefix: o (order), g (genus), f (family), and s (species).

(G) All 42 microbial species incorporated into the personalized protein database.

(H) The prevalence of Alistipes (left) and Fusobacterium (right) across 32 different cancer types as documented in the BIC database. Abbreviations for all cancer types are listed in Table S4.

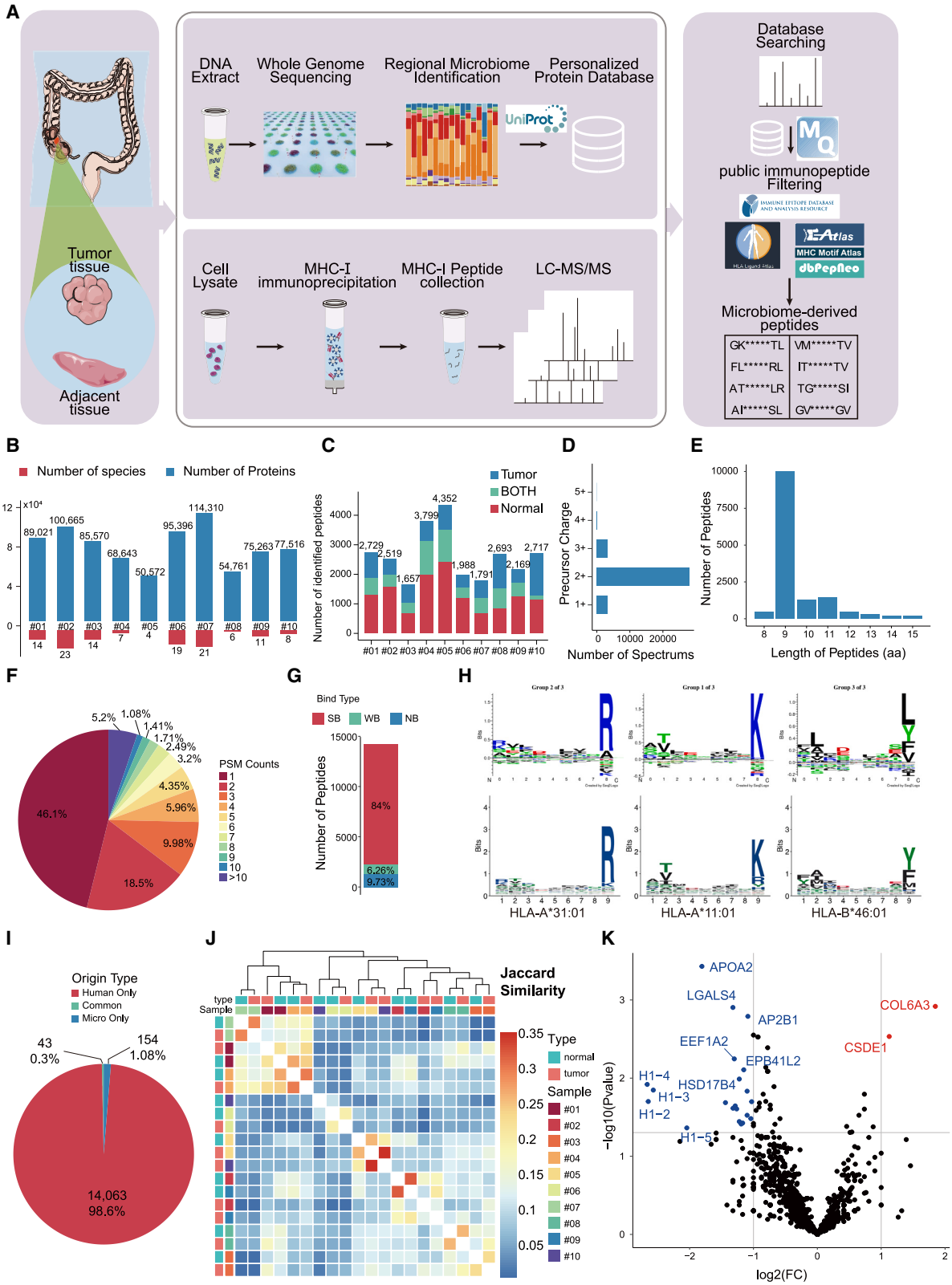
Table 1. Reported relationships between microbes and cancers

Species of microbes	Tumor associated	Tumor-suppressing	Tumor-promoting	Cancer therapy related
Alistipes communis	–	HCC[37]	CRC[38]	PD-1 in NSCLC[39]; durable clinical benefit in lung cancer[27]
Anaerostipes hadrus	PC[40]; CRC[41]	–	–	–
Bacteroides caccae	CRC[42]	–	–	ICI in MM[43]; TNBC[44]
Bacteroides heparinolyticus	CRC[45]; NMIBC[46]; BC; GC; OC; PC; pharyngeal cancer[47]	–	multiple myeloma[48]; hematologic malignancies[49]	–
Collinsella aerofaciens	–	–	–	ICI in MM[27,50]
Coprococcus catus	–	–	CRC[51]; advanced adenoma[51]; precancerous mucosal lesions[52]	ICI in BC[53]
Eubacterium callanderi	–	CRC[54]	–	–
Faecalibacterium prausnitzii	CRC[55]; lung cancer[27]	CRC[56,57]; BC[58]	–	ICI in MM[43,59–61]; radioiodine therapy in thyroid cancer[62]
Fretibacterium fastidiosum	–	lung cancer[63]	–	–
Fusobacterium necrophorum	CRC[64]	–	–	–
Fusobacterium nucleatum	CRC[65]	–	CRC[66,67]	CRC[68]
Hungatella hathewayi	CRC[69]	–	CRC[70]	–
Intestinimonas butyriciproducens	CRC[71,72]	–	–	ICI in melanoma[26]
Lactococcus lactis	–	CRC[73,74]	–	–
Paeniclostridium sordellii	–	–	colonic epithelial lesions[75]	–
Paraburkholderia fungorum	–	ICC[31]	–	–
Parvimonas micra	–	–	CRC[76,77]	Immune response in CRC[78,79]
Phascolarctobacterium faecium	–	–	–	ICI in melanoma[24]
Romboutsia ilealis	pancreatic cancer[40]	–	–	–
Streptococcus dysgalactiae	GC[80]	–	–	–
Treponema sp. *OMZ 838*	–	–	OSCC[81,82]	–

Abbreviations included are listed in Table S4.

average number of 2,641 HLA-I ligands were identified per patient through database searching with MaxQuant (Figure 2C). A Total of 14,260 HLA-I ligands were identified across all Samples. These identified HLA-I ligands corresponded to 46,318 peptide-spectrum matches (PSMs), and 90.13% of precursor charge being +2 (Figure 2D). Of the ligands, 69.8% were nonapeptides (Figure 2E). Although 46.1% of the ligands were reported by individual PSM (Figure 2F), 84% were predicted as strong binders by netMHCPan[85] (Figure 2G). The length distribution and charge states of the peptides were consistent with known HLA-I ligand characteristics.[86,87] Moreover, the amino acid frequencies at specific positions of HLA-I ligands in each patient were mostly consistent with known HLA-I subtype motifs, except for patient #10 (Figures 2H and S2). Furthermore, we observed that the total number and the average number per milligrams of HLA-I ligand identified in tumor samples were fewer than in paired normal samples (Figure S3A), possibly due to the downregulation or loss of HLA-I surface expression in cancer.[88–90] A marginal increase in the total identified HLA-I ligands was observed with increased sample weight (Figure S3B).

Among these HLA-I ligands, we identified that 98.6% (14,063 peptides) were exclusively derived from human self-protein based on 6,359 human proteins and related isoforms (Figure 2I). The Jaccard similarity revealed a higher similarity of HLA-I ligands among samples from the same patient compared to tumor or normal samples from different patients (Figure 2J). Notably, the ligands of patients #01, #04, and #07 showed a higher similarity compared to other samples, indicating that the diversity of immunopeptidome is associated with the subtypes of HLA-I (Figure S3C). We found that protein COL6A3 and CSDE1, which yielded more ligands in tumors than normal tissues, have been reported to be associated with cancer development.[91–93] Conversely, 21 other proteins, including various histones and eukaryotic translation elongation factors (EEFs), generated more HLA-I ligands in normal tissues (Figure 2K, Tables S6 and S7). GO enrichment analysis indicated significant involvement of



(legend on next page)

these 21 Proteins in chromosome condensation and related chromosome dynamics, structural stability, and cell cycle processes. The scarcity of related HLA-I ligands of these proteins in tumors underscores the aberrant proliferation and cell cycle disruption characteristic of cancer cells (Figure S3D).

Immunogenic bacterial peptides are presented by HLA-I in CRC tumor

In addition to peptides exclusively derived from human self-proteins, we identified 197 peptides may originating from microbial proteins or shared with human proteins. Of these, 84 peptides (42.64%) particularly detected in tumor samples (Figure 3A; Table S8). The overlap of peptides between normal and tumor samples was minimal (33 peptides, 16.75%) (Figure 3A). The number of putative microbe-derived peptides identified varied across samples, ranging from 8 to 35, due to the differing PPDs used (Figure 3B).

To characterize putative microbe-derived epitopes, peptides were categorized as 'Common' (43 peptides, 20.83%) if they matched known human proteins, or 'Microbe only' (154 peptides, 78.17%) if they did not (Figure 3C and Table S9). Among the 'Common' peptides, 38 peptides were identified as linear epitopes in public databases. Notably, we observed significant sequence similarity among several peptides (Table 2). For instance, the peptide LLDEPTNHL, reported as a T cell epitope from *Mycobacterium tuberculosis*,[94] was also sourced from human protein ABCF2. Its variant ILDEPTNHL was documented in the HLA Ligand Atlas as a human ABCF3 originated ligand, but our study linked it to proteins from *F. nucleatum, Klebsiella pneumoniae*, and *Parvimonas micra*. The peptide LLDEPTNNL (N/H variant of LLDEPTNHL) was previously reported in both *Bifidobacterium longum* and human. However, our findings indicate it may be derived from multiple proteins of *K. pneumoniae* and *F. nucleatum*.

In contrast, only four 'Microbe only' peptides (2.60%) were retrieved from the public databases. The peptides ITELNSPVL and IIKDNQIIV, consistent with our findings, were confirmed as bacterial HLA-bound peptides in melanoma from *F. nucleatum*.[95] ELLKLLEQL (L/I variant of ELIKIIEQL) was listed as an epitope eluted from human HLA allele in the Immune Epitope Database (IEDB),[96] and INKETLENV (L/I variant of INKETIENV) was reported as an HLA-A*02:01 ligand in the MHC Motif Atlas.[97] Although the source proteins for the last two peptides were not specified in the databases, our study suggested their mirror I/L variants were derived from *Bacteroides caccae* and *F. nucleatum* separately. Additionally, none of our 'Microbe only' peptides were found in the HLA ligand Atlas,[98] Cryptic peptides, IEAtlas[99] or dbPepNeo.[100] IEAtlas and dbPepNeo specifically catalog non-canonical epitopes from non-coding regions and neoantigens derived from somatic mutation-induced protein variations. In contrast, the HLA ligand Atlas and Cryptic peptides databases include ligands originating from known proteins or non-coding regions in benign human tissues. These findings further support our hypothesis that these peptides may originating from microorganisms.

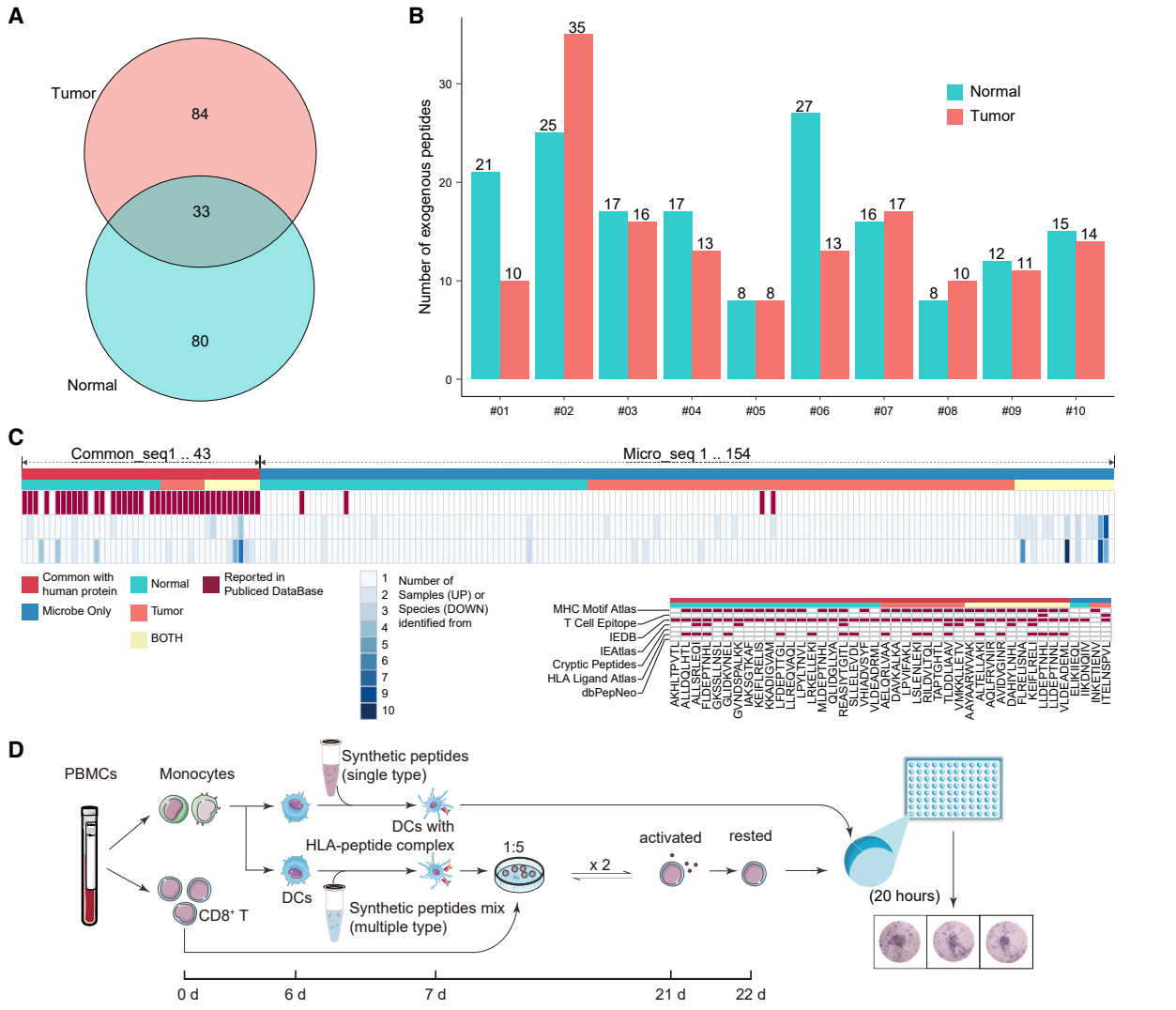
Previous studies have highlighted the presence of tumor-specific intracellular bacteria and their influence on therapy response, cancer metastasis, and patient outcomes in both clinical and animal models.[101–104] However, few studies have examined microbial antigens in tumors and their effect on the TME.[95,105] To investigate whether these microbial antigens could trigger $CD8^+$ T cell responses and inflame the TME, we randomly evaluated eight microbial. After passed verification (7/8) through parallel reaction monitoring (PRM) (Figure S4; Table S11), the evaluation of immunogenicity with synthesized peptides on $CD8^+$ T cells from four healthy donors were performed (Figure 3D). To maximize detect the possibilities of the immunogenicity, donors were selected based on the predicted affinity of theses peptides with every HLA haplotype of patients that peptides were detected from (Table S10). The results showed that ITGNSILTV could stimulate IFN-γ secretion from $CD8^+$ T cells (Figure 3E; Tables S11 and S12). Peptide ITGNSILTV is associated with the transmembrane helices of the multi-antimicrobial extrusion protein (MatE) (Figure 3F). It is prevalent in bacteria and located in cell membranes.[106–108]

Tumor-specific peptides derived from microbes were shared across CRC patients

The application of cancer immunotherapy has clearly demonstrated that, once identified, cancerous cells can be effectively eliminated by immune system. Identifying tumor-specific microbial-originated peptides may offer a possible route to assist the host's immune system, including $CD8^+$ T cells, in recognize and killing cancer cells.[28] In this study, more than 10 peptides were identified as originating from *F. nucleatum*, *Cedecea neteri*, *Bacteroides uniformis* and *Aeromonas hydrophila*, respectively

Figure 2. Overview of HLA-I ligands identified in this study

(A) The identification strategy for microbiome-derived HLA-I ligands utilizing whole-genome sequencing (WGS) and liquid chromatography-tandem mass spectrometry (LC-MS/MS).
(B) Number of species and proteins included in the construction of a personalized database.
(C) Number of HLA-I ligands identified in each patient.
(D) The distribution of precursor ion charges detected.
(E) The length distribution of HLA-I ligands identified.
(F) The peptide-spectrum match (PSM) counts for the identified ligands.
(G) Affinity of peptide presented by HLA-I complex predicted by NetMHCpan. Peptides were classified based on the minimum rank value with specific HLA-I subtypes: SB, strong binder (<0.5%); WB, weak binder (0.5%–2%); NB, no binder (>2%).
(H) Motif analysis of HLA-I ligands identified in patient #02 (upper panel) compared to known motifs of specific HLA alleles (lower panel).
(I) Classification of HLA-I ligands identified across all tumor and normal samples.
(J) Jaccard index assessing the similarity of identified HLA-I ligands between samples.
(K) Selection of twenty-three differentially expressed genes based on the number of matched peptides from the corresponding proteins. *p* values were calculated using a two-tailed Student's t-test.



(legend on next page)

(Figure 4A). Five peptides were identified from multiple tumor samples, including AGDNIGVLLR and ITELNSPVL, which originated from the Elongation factor Tu (EF-Tu) and RND superfamily resistance-nodulation-cell division antiporter of *F. nucleatum*, respectively (Figure 4B).

Additionally, microbial peptides appeared to be more similar among the same type of specimens from different patients (Figure S5A), indicating the potential for microbial peptides to be applied across a wider range of patients, regardless of HLA complex subtyping. Notably, 50% of 'Microbe only' peptides, which were identified exclusively from tumor samples, could serve as potential targets for immunotherapy. Compared to the other half of the peptides, tumor-specific microbial-originated peptides included a higher proportion of ten-amino-acid sequences (Figure 4C) and were stronger binders with HLA-I subtypes in a lower IC_{50} values (Figures 4D and S5B). As a result, these peptides have a higher likelihood of being presented and recognized by the immune system.

When tracing back to source species, the greatest contributor to tumor-specific microbial-originated peptides was *F. nucleatum*. In addition to the two above peptides identified from multiple samples, there were another 18 peptides identified from *F. nucleatum*, 11 of which were tumor-specific peptides (Figure 4E; Table S13). Peptides AIVSALVAI, ITELNSPVL, QHIDSLSSF, SFSASFSIS, and VSKTNSRASSAI originated from proteins located in the membrane. Notably, SFSASFSIS and VSKTNSRASSAI might be generated from the same protein Q8REH6, a hypothetical membrane-spanning protein. Similarly to Immunogenic bacterial peptides proved in this study, these 5 peptides were located within or near a 3D structure or domain of their source proteins, which may be conserved in structure or sequence (Figure 4F).

Peptides AGDNIGVLLR, along with three other peptides (GYRPQFYFR, TTLTAAITTVLAK, and VGEEVEIVGIK) identified in multiple tumor and normal samples, were all derived from Elongation factor Tu (EF-Tu) (Figure 4G). As an abundant and conserved cytoplasmic protein in bacteria, EF-Tu has been reported to stimulate immune responses, and antibodies against EF-Tu can decrease bacterial load and partially protect hosts from infections.[109] To further verify that tumor-specific peptides identified from F. nucleatum could trigger the host's immune response, two peptides identified from multiple samples and 4 other random selected peptides labeled as tumor-specific peptides were synthesized. Following with the same strategy used in former verification, 5 peptides that pass the PRM verification (Figure S6; Table S11) were tested with PBMCs from another 2 donors. Results indicated that all 5 peptides could stimulate IFN-γ secretion from $CD8^+$ T cells (Figure 4H and Table S14). All this evidence supports the assumption that these peptides could be presented and identified by $CD8^+$ T cells, which would be contribute to remolding TME.

DISCUSSION

In this study, we established a strategy by combining the high-depth WGS data with the HLA-I immunopeptidome data to identify the microbe-derived HLA I ligands. We identified 154 peptides derived from microbes, 77 of which were exclusively found in tumor samples. Most of these microbe-derived peptides predominantly originated from bacteria. Peptide ITGNSILTV was proved to be immunogenic, indicating their potential roles in cancer immunotherapy.

During the HLA-I ligand identification process, we first accurately identified the microbial species by utilizing high-depth WGS data from paired normal samples to exclude potential contaminant DNA. The identification of microbes was executed using a robust taxonomic classification system from metagenomic analysis. We found that *F. nucleatum* emerged as highly abundant (>1% relative abundance) and significantly more prevalent within tumor regions. The potential link between *F. nucleatum* and CRC, as well as its utility in cancer diagnosis, classification, and treatment, has been extensively documented.[66–68,71,110] The identification of HLA-I ligands from *F. nucleatum* and other cancer-related microbes may help elucidate their role in carcinogenesis and their potential in cancer therapy, particularly immunotherapy.

To account for the complexities of HLA-I peptide presentation and HLA polymorphisms, we created personalized protein databases for each patient, these databases incorporated proteins from significantly different microbes (SDMs), tumor-specific high-abundance microbes (SHMs), and *H. sapiens*. However, due to gaps in microbial proteomics data, some microbes were excluded during the construction of these protein databases. As a result, only *F. nucleatum* was retained as an SDM. Given that SHMs varied per patient, the number of proteins included in each personalized database differed. For half of the patients (#01, #02, #03, #06, and #07), the protein count exceeded twice the number of human proteins listed in Swiss-Prot (42,367). The limited expansion of the databases was acceptable for target-decoy search strategy, ensuring the accuracy of spectrum analysis.[84,111] However, the absence of proteome data for other microbes also limited the exploring of microbial originated peptides.

Generally, over 98% of HLA-I ligands were host-derived, with paired samples and patients with similar HLA subtypes showing greater ligand similarity.[112] Notably, peptides from proteins COL6A3 and CSDE1 were more abundant in tumor samples.

Figure 3. Presentation of exogenous peptides by HLA-I molecules in the CRC tumor microenvironment

(A) Total number of exogenous peptides identified in normal and tumor samples.

(B) Number of exogenous peptides detected in individual samples.

(C) Features and searching results of identified exogenous peptides. Related databases used in searching process were listed on the key resources table.

(D) Schematic diagram of immunogenicity validation.

(E) Quantification of IFN-γ release from T cell stimulated by exogenous peptides. DMSO and anti-human CD3 monoclonal antibody (CD3-2) served as negative and positive controls, respectively. Data were presented as mean ± SEM. SFU denotes spot forming units. *P*-values were calculated using a two-tailed Student's t-test, wih $p < 0.05$ denoted by *.

(F) Topological features of source proteins yielding immunogenic bacterial peptides.

Table 2. Peptides similar to those reported in different databases and this study[a]

Sequence[b]	Source species	Source protein	Database
LLDEPTN**H**L	*F. nucleatum, H. sapiens, K. pneumoniae, P. micra*	different kinds of ATP-binding proteins	This study
	H. sapiens	ATP-binding cassette sub-family F member 2	IEDB
	M. tuberculosis	ABC transporter ATP-binding protein	IEDB T cell epitope
	-	–	MHC Motif Atlas
ILDEPTN**H**L	*H. sapiens*	ATP-binding cassette sub-family F member 3	IEDB
	H. sapiens	ATP-binding cassette sub-family F member 3	HLA Ligand Atlas
	-	–	MHC Motif Atlas
LLDEPTN**N**L	*F. nucleatum, H. sapiens, K. pneumoniae*	ABC transporter ATP-binding protein, Exonuclease SBCC (EC 3.1.11.-), ATP-binding component of a transport system, Transport system ATP-binding component	This study
	B. longum	ABC-type transporter ATPase component	IEDB
ILDEPTN**N**L	*H. sapiens*	ATP-binding cassette sub-family F member 1	IEDB
	H. sapiens	ATP-binding cassette sub-family F member 1	HLA Ligand Atlas
	-	–	MHC Motif Atlas
EL**I**K**II**EQL	*B. caccae*	histidine kinase (EC 2.7.13.3)	This study
EL**L**K**LL**EQL	*H. sapiens*	*cis*-spliced peptide from unknown protein eluted from human MHC allele	IEDB
INKET**I**ENV	*F. nucleatum*	Methyltransferase	This study
INKET**L**ENV	–	–	MHC Motif Atlas

[a]Information that is not included in databases is marked as "-."
[b]Difference among similar peptides was marked as bold font.

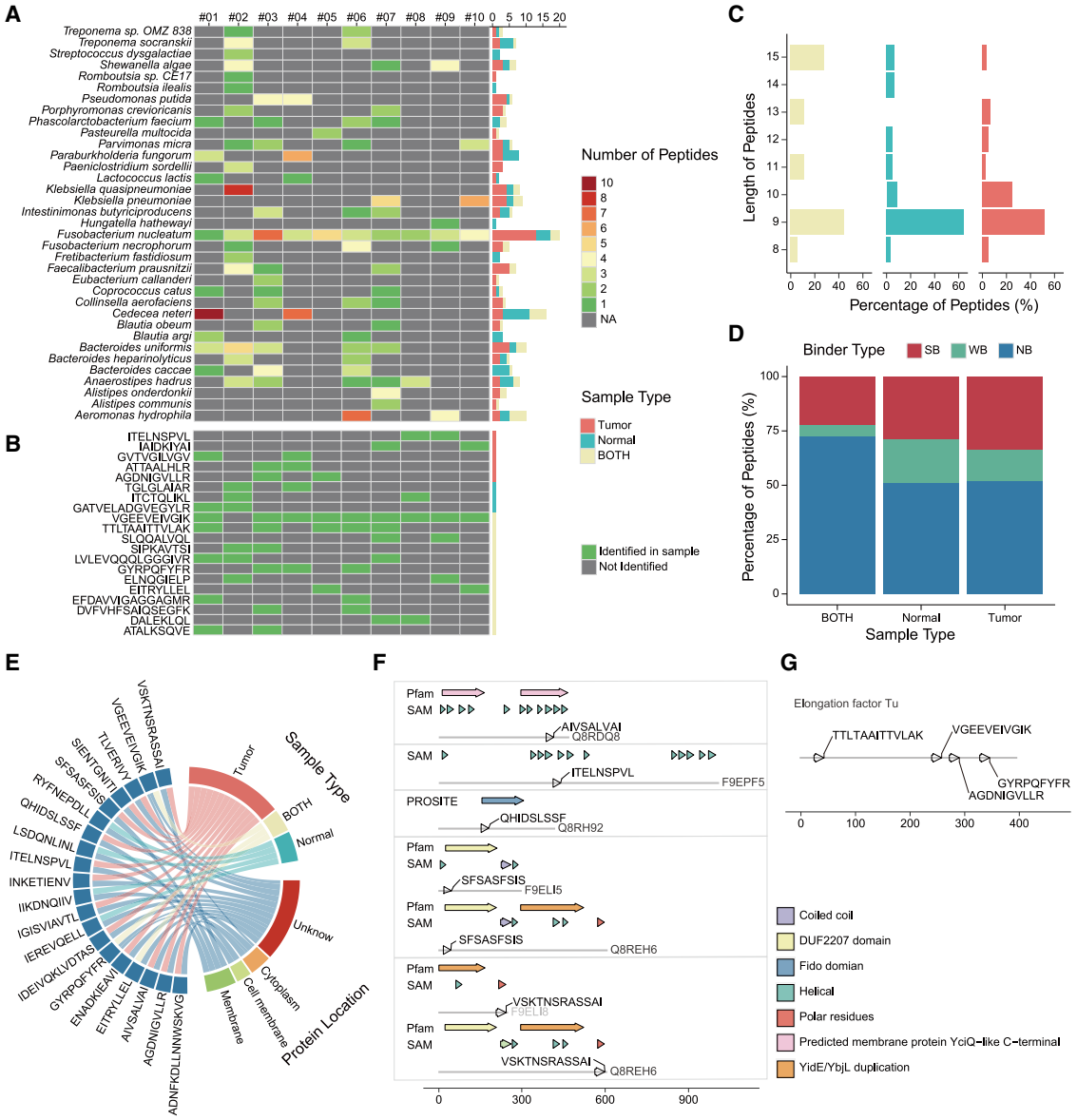
Given their potential as biomarkers and therapeutic targets,[113,114] further research is required to determine whether these ligands can modify the TME in immunotherapy and be utilized in the development of chimeric antigen receptor T (CAR-T) or T cell receptor–engineered T (TCR-T) cells.[5,115]

Although SHMs were included in the databases, we found no significant accumulation or tumor-enrichment of their peptides. The observation applied to *F. nucleatum* in tumor samples potentially due to the limited microbial variety and peptide numbers, as well as MHC downregulation in tumor cells. Nonetheless, we identified 154 microbe-derived peptides as HLA-I ligands across tumor and normal samples, including 18 that are specific to *F. nucleatum*, 15 of which were not recorded in the public databases we utilized. Additionally, peptides from bacterial EF-Tu and EF-4 appeared in multiple samples. Given their abundance and conservation in bacteria,[116] and their known roles in stimulating immune responses,[109,117] these proteins are promising candidates for immunization against infections and potentially cancer. This study also confirmed that peptides ITGNSILTV from MatE could activate $CD8^+$ T cell immunity. These conserved prokaryotic proteins and their peptides warrant further investigation.

In conclusion, our study successfully identified 154 putative microbe-derived HLA-I ligands by utilizing personalized databases that incorporated microbiome-derived proteins to decipher the LC-MS/MS spectra. The majority of these putative microbe-derived ligands were not recorded in related databases. Our findings demonstrated that peptides ITGNSILTV from MatE protein can activate $CD8^+$ T cell immune responses, indicating the potential application of microbe-derived peptides in cancer immunotherapy. Furthermore, another 15 peptides were identified from *F. nucleatum* highlighting the need for further exploration of these peptides in therapeutic contexts.

Limitations of the study

It is important to note that verifying the sources of putative microbe-derived peptides in our approach may be limited. Peptides presented by HLA-I can originate from various sources, such us linear sequences or spliced oligopeptides formed by ligating non-contiguous peptide fragments.[12] These fragments may arise from normal or abnormal translation of genomic regions, including non-coding regions, as well as from extracellular or intracellular microorganisms and toxins.[11,118] In this study, we focused solely on the annotated proteins of humans and microbes during the peptide identification process and traced their origins. While the usage of public epitope databases partially compensates for this limitation, we did not consider somatic mutation or abnormal translation of genomic regions specific to each patient. Future studies could potentially overcome this limitation by incorporation somatic mutation analysis using



(legend on next page)

WGS,[119–121] transcriptome identification via Ribosome Nascent-chain Complex-bound RNA sequencing (RNC-seq)[122,123] and long-read sequencing.[124]

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Xuan Dong (dongxuan@genomics.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Whole genome sequencing data have been deposited at in CNGB Sequence Archive (CNSA)[125] of China National GeneBank (CNGB)[126] and the Genome Sequence Archive for Human[127] in National Genomics Data Center (NGDC)[128] and are publicly available as of the date of publication. Raw mass spectra data have been deposited in CNSA of CNGBdb and are publicly available as of the date of publication. Accession numbers are listed on the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We acknowledge support from the Guangdong Provincial Key Laboratory of Human Disease Genomics (no. 2020B1212070028) and Shenzhen Key Laboratory of Genomics (no. CXB200903110066A). This work was supported by China National GeneBank (CNGB).

AUTHOR CONTRIBUTIONS

Conceptualization, X.D. and X.G.; methodology, X.G., H.X., Y.F., and F.B.; software, X.G. and H.X.; validation, F.B., Y.F., X.C., and T.C.; formal analysis, X.G. and H.Z.; resources, X.W., K.W., and X.D.; writing – original draft, X.G.; writing – review and editing, X.D. and X.G.; visualization, X.G. and H.Z.; project administration, X.D., F.B., and L.L.

DECLARATION OF INTERESTS

BGI Group employees hold company stock.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used ChatGPT to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human participants
- METHOD DETAILS
  - Identification of microbe from samples based on whole genome sequencing (WGS) data
  - Identification of HLA-I subtypes
  - Calculation of reads count and relative abundance in different samples and every level of taxonomy
  - HLA-I ligand identification based on database dependent analysing
  - Collection of HLA-I ligand through immunoprecipitation
  - Acquire mass spectra of HLA-I ligand through LC-MS/MS
  - Construction of personalized database
  - Database searching with MaxQuant
  - Classification of identified peptides
  - Comparison with epitopes in public databases, articles and patents
  - Affinity prediction and motif analysis of immunopeptidome
  - Verification of HLA-I ligands with parallel reaction monitoring with synthesized peptide
  - Immunogenicity validation of bacterial peptides
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111338.

Received: April 18, 2024
Revised: July 23, 2024
Accepted: November 4, 2024
Published: November 9, 2024

REFERENCES

1. Esfahani, K., Roudaia, L., Buhlaiga, N., Del Rincon, S., Papneja, N., and Miller, W. (2020). A Review of Cancer Immunotherapy: From the Past, to the Present, to the Future. Curr. Oncol. *27*, 87–97.
2. Rosenberg, S.A., and Restifo, N.P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. Science *348*, 62–68.
3. Ragoonanan, D., Khazal, S.J., Abdel-Azim, H., McCall, D., Cuglievan, B., Tambaro, F.P., Ahmad, A.H., Rowan, C.M., Gutierrez, C., Schadler, K., et al. (2021). Diagnosis, grading and management of toxicities from immunotherapies in children, adolescents and young adults with cancer. Nat. Rev. Clin. Oncol. *18*, 435–453.

Figure 4. Characterization of tumor-specific microbe-originated peptides

(A) The number of identified microbe-originated peptides in each patient. Total number of peptides sourcing from each microbe was noted in the bar plot on the right.
(B) Shared microbe-originated peptides in different patients.
(C) Length distribution of microbe-originated peptides.
(D) Predicted binding affinity of microbe-originated peptides. Peptides were classified following the same principle described in Figure 2F.
(E) Peptides identified from *F. nucleatum*. Protein location was described based on the annotation messages downloaded from UniProt.
(F) 3D structure and domain characteristics for 7 membrane proteins where 5 tumor-specific peptides were identified from *F. nucleatum*.
(G) Location of peptides identified from EF-Tu.
(H) Quantification of IFN-γ release from T cells stimulated by tumor-specific *F. nucleatum*-derived peptides. Data were represented as mean ± SEM. *P*-values were calculated using a two-tailed Student's t-test, with significance levels denoted as follows: * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). Experiments design and control settings as in Figure 3E.

4. Mahoney, K.M., Rennert, P.D., and Freeman, G.J. (2015). Combination cancer immunotherapy and new immunomodulatory targets. Nat. Rev. Drug Discov. *14*, 561–584.

5. Baulu, E., Gardet, C., Chuvin, N., and Depil, S. (2023). TCR-engineered T cell therapy in solid tumors: State of the art and perspectives. Sci. Adv. *9*, eadf3700.

6. Tay, R.E., Richardson, E.K., and Toh, H.C. (2021). Revisiting the role of CD4+ T cells in cancer immunotherapy—new insights into old paradigms. Cancer Gene Ther. *28*, 5–17.

7. Zlatareva, I., and Wu, Y. (2023). Local γδ T cells: translating promise to practice in cancer immunotherapy. Br. J. Cancer *129*, 393–409. https://doi.org/10.1038/s41416-023-02303-0.

8. Raskov, H., Orhan, A., Christensen, J.P., and Gögenur, I. (2021). Cytotoxic CD8+ T cells in cancer and cancer immunotherapy. Br. J. Cancer *124*, 359–367.

9. Liu, Y., Yan, X., Zhang, F., Zhang, X., Tang, F., Han, Z., and Li, Y. (2021). TCR-T Immunotherapy: The Challenges and Solutions. Front. Oncol. *11*, 794183.

10. Charles, A.J., Jr., Travers, P., Walport, M., and Shlomchik, M.J. (2001). The major histocompatibility complex and its functions. In Immunobiology: The Immune System in Health and Disease, 5th edition (Garland Science).

11. Yewdell, J.W. (2022). MHC Class I Immunopeptidome: Past, Present, and Future. Mol. Cell. Proteomics *21*, 100230.

12. Pishesha, N., Harmand, T.J., and Ploegh, H.L. (2022). A guide to antigen processing and presentation. Nat. Rev. Immunol. *22*, 751–764.

13. Admon, A. (2023). The biogenesis of the immunopeptidome. Semin. Immunol. *67*, 101766.

14. Son, E.T., Faridi, P., Paul-Heng, M., Leong, M.L., English, K., Ramarathinam, S.H., Braun, A., Dudek, N.L., Alexander, I.E., Lisowski, L., et al. (2021). The self-peptide repertoire plays a critical role in transplant tolerance induction. J. Clin. Invest. *131*, e146771.

15. Oberhardt, V., Luxenburger, H., Kemming, J., Schulien, I., Ciminski, K., Giese, S., Csernalabics, B., Lang-Meli, J., Janowska, I., Staniek, J., et al. (2021). Rapid and stable mobilization of CD8+ T cells by SARS-CoV-2 mRNA vaccine. Nature *597*, 268–273.

16. Yamamiya, D., Mizukoshi, E., Kaji, K., Terashima, T., Kitahara, M., Yamashita, T., Arai, K., Fushimi, K., Honda, M., and Kaneko, S. (2018). Immune responses of human T lymphocytes to novel hepatitis B virus-derived peptides. PLoS One *13*, e0198264.

17. Bruno, P.M., Timms, R.T., Abdelfattah, N.S., Leng, Y., Lelis, F.J.N., Wesemann, D.R., Yu, X.G., and Elledge, S.J. (2023). High-throughput, targeted MHC class I immunopeptidomics using a functional genetics screening platform. Nat. Biotechnol. *41*, 980–992.

18. He, J., Xiong, X., Yang, H., Li, D., Liu, X., Li, S., Liao, S., Chen, S., Wen, X., Yu, K., et al. (2022). Defined tumor antigen-specific T cells potentiate personalized TCR-T cell therapy and prediction of immunotherapy response. Cell Res. *32*, 530–542.

19. Chong, C., Coukos, G., and Bassani-Sternberg, M. (2022). Identification of tumor antigens with immunopeptidomics. Nat. Biotechnol. *40*, 175–188.

20. Fujiwara, K., Shao, Y., Niu, N., Zhang, T., Herbst, B., Henderson, M., Muth, S., Zhang, P., and Zheng, L. (2022). Direct identification of HLA class I and class II-restricted T cell epitopes in pancreatic cancer tissues by mass spectrometry. J. Hematol. Oncol. *15*, 154.

21. Foy, S.P., Jacoby, K., Bota, D.A., Hunter, T., Pan, Z., Stawiski, E., Ma, Y., Lu, W., Peng, S., Wang, C.L., et al. (2023). Non-viral precision T cell receptor replacement for personalized cell therapy. Nature *615*, 687–696.

22. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discov. *12*, 31–46.

23. Garrett, W.S. (2015). Cancer and the microbiota. Science *348*, 80–86.

24. Frankel, A.E., Deshmukh, S., Reddy, A., Lightcap, J., Hayes, M., McClellan, S., Singh, S., Rabideau, B., Glover, T.G., Roberts, B., and Koh, A.Y. (2019). Cancer Immune Checkpoint Inhibitor Therapy and the Gut Microbiota. Integr. Cancer Ther. *18*, 1534735419846379.

25. Nunes-Alves, C. (2016). Commensals promote anticancer immunotherapy. Nat. Rev. Microbiol. *14*, 3.

26. Lee, K.A., Thomas, A.M., Bolte, L.A., Björk, J.R., de Ruijter, L.K., Armanini, F., Asnicar, F., Blanco-Miguez, A., Board, R., Calbet-Llopart, N., et al. (2022). Cross-cohort gut microbiome associations with immune checkpoint inhibitor response in advanced melanoma. Nat. Med. *28*, 535–544.

27. Sun, J.Y., Yin, T.L., Zhou, J., Xu, J., and Lu, X.J. (2020). Gut microbiome and cancer immunotherapy. J. Cell. Physiol. *235*, 4082–4088.

28. Zhang, Z., Gao, Q., Ren, X., Luo, M., Liu, Y., Liu, P., Liu, Y., Ye, Y., Chen, X., Liu, H., and Han, L. (2023). Characterization of intratumor microbiome in cancer immunotherapy. Innovation *4*, 100482.

29. Fu, A., Yao, B., Dong, T., and Cai, S. (2023). Emerging roles of intratumor microbiota in cancer metastasis. Trends Cell Biol. *33*, 583–593.

30. Liu, J., and Zhang, Y. (2022). Intratumor microbiome in cancer progression: current developments, challenges and future trends. Biomark. Res. *10*, 37.

31. Chai, X., Wang, J., Li, H., Gao, C., Li, S., Wei, C., Huang, J., Tian, Y., Yuan, J., Lu, J., et al. (2023). Intratumor microbiome features reveal antitumor potentials of intrahepatic cholangiocarcinoma. Gut Microb. *15*, 2156255.

32. Yang, L., Li, A., Wang, Y., and Zhang, Y. (2023). Intratumoral microbiota: roles in cancer initiation, development and therapeutic efficacy. Signal Transduct. Targeted Ther. *8*, 35.

33. Hilmi, M., Kamal, M., Vacher, S., Dupain, C., Ibadioune, S., Halladjian, M., Sablin, M.P., Marret, G., Ajgal, Z.C., Nijnikoff, M., et al. (2023). Intratumoral microbiome is driven by metastatic site and associated with immune histopathological parameters: An ancillary study of the SHIVA clinical trial. Eur. J. Cancer *183*, 152–161.

34. Galeano Niño, J.L., Wu, H., LaCourse, K.D., Kempchinsky, A.G., Baryiames, A., Barber, B., Futran, N., Houlton, J., Sather, C., Sicinska, E., et al. (2022). Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. Nature *611*, 810–817.

35. Sepich-Poore, G.D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J.A., and Knight, R. (2021). The microbiome and human cancer. Science, 371.

36. Chen, K.-P., Hsu, C.-L., Oyang, Y.-J., Huang, H.-C., and Juan, H.F. (2023). a database for the transcriptional landscape of bacteria in cancer. Nucleic Acids Res. *51*, D1205–D1211.

37. Li, J., Sung, C.Y., Lee, N., Ni, Y., Pihlajamäki, J., Panagiotou, G., and El-Nezami, H. (2016). Probiotics modulated gut microbiota suppresses hepatocellular carcinoma growth in mice. Proc. Natl. Acad. Sci. USA *113*, E1306–E1315.

38. Iida, N., Dzutsev, A., Stewart, C.A., Smith, L., Bouladoux, N., Weingarten, R.A., Molina, D.A., Salcedo, R., Back, T., Cramer, S., et al. (2013). Commensal Bacteria Control Cancer Response to Therapy by Modulating the Tumor Microenvironment. Science *342*, 967–970.

39. Jin, Y., Dong, H., Xia, L., Yang, Y., Zhu, Y., Shen, Y., Zheng, H., Yao, C., Wang, Y., and Lu, S. (2019). The Diversity of Gut Microbiome is Associated With Favorable Responses to Anti–Programmed Death 1 Immunotherapy in Chinese Patients With NSCLC. J. Thorac. Oncol. *14*, 1378–1389.

40. Wang, X.-Y., Sun, Z.X., Makale, E.C., Sun, Z.K., Wang, T.C., Hou, Z.Y., Yu, X.Y., Long, W., Du, G.K., and Huang, H. (2022). Gut Microbial Profile in Patients with Pancreatic Cancer. Jundishapur J. Microbiol. *15*.

41. Ai, D., Pan, H., Li, X., Gao, Y., Liu, G., and Xia, L.C. (2019). Identifying Gut Microbiota Associated With Colorectal Cancer Using a Zero-Inflated Lognormal Model. Front. Microbiol. *10*, 826.

42. Png, C.-W., Chua, Y.-K., Law, J.-H., Zhang, Y., and Tan, K.-K. (2022). Alterations in co-abundant bacteriome in colorectal cancer and its persistence after surgery: a pilot study. Sci. Rep. *12*, 9829.

43. Frankel, A.E., Coughlin, L.A., Kim, J., Froehlich, T.W., Xie, Y., Frenkel, E.P., and Koh, A.Y. (2017). Metagenomic Shotgun Sequencing and Unbiased Metabolomic Profiling Identify Specific Human Gut Microbiota and Metabolites Associated with Immune Checkpoint Therapy Efficacy in Melanoma Patients. Neoplasia *19*, 848–855.

44. Swoboda, A.T., Sharma, A., Olopade, O.I., Nanda, R., and Gilbert, J. (2019). Characterizing the gut microbiome of patients with triple-negative breast cancer. J. Clin. Oncol. *37*, e14186.

45. Zhang, M., Fan, X., Fang, B., Zhu, C., Zhu, J., and Ren, F. (2015). Effects of Lactobacillus salivarius Ren on cancer prevention and intestinal microbiota in 1, 2-dimethylhydrazine-induced rat model. J. Microbiol. *53*, 398–405.

46. Chen, C., Huang, Z., Huang, P., Li, K., Zeng, J., Wen, Y., Li, B., Zhao, J., and Wu, P. (2022). Urogenital Microbiota:Potentially Important Determinant of PD-L1 Expression in Male Patients with Non-muscle Invasive Bladder Cancer. BMC Microbiol. *22*, 7.

47. Kabwe, M., Dashper, S., Bachrach, G., and Tucci, J. (2021). Bacteriophage manipulation of the microbiome associated with tumour microenvironments-can this improve cancer therapeutic response? FEMS Microbiol. Rev. *45*, fuab017.

48. Calcinotto, A., Brevi, A., Chesi, M., Ferrarese, R., Garcia Perez, L., Grioni, M., Kumar, S., Garbitt, V.M., Sharik, M.E., Henderson, K.J., et al. (2018). Microbiota-driven interleukin-17-producing cells and eosinophils synergize to accelerate multiple myeloma progression. Nat. Commun. *9*, 4832.

49. D'Angelo, C.R., Sudakaran, S., and Callander, N.S. (2021). Clinical effects and applications of the gut microbiome in hematologic malignancies. Cancer *127*, 679–687.

50. Matson, V., Fessler, J., Bao, R., Chongsuwat, T., Zha, Y., Alegre, M.L., Luke, J.J., and Gajewski, T.F. (2018). The commensal microbiome is associated with anti–PD-1 efficacy in metastatic melanoma patients. Science *359*, 104–108.

51. Xu, J., Zheng, Z., Yang, L., Li, R., Ma, X., Zhang, J., Yin, F., Liu, L., Xu, Q., Shen, Q., et al. (2022). A novel promising diagnosis model for colorectal advanced adenoma and carcinoma based on the progressive gut microbiota gene biomarkers. Cell Biosci. *12*, 208.

52. Wei, H., Dong, L., Wang, T., Zhang, M., Hua, W., Zhang, C., Pang, X., Chen, M., Su, M., Qiu, Y., et al. (2010). Structural shifts of gut microbiota as surrogate endpoints for monitoring host health changes induced by carcinogen exposure. FEMS Microbiol. Ecol. *73*, 577–586.

53. Terrisse, S., Derosa, L., Iebba, V., Ghiringhelli, F., Vaz-Luis, I., Kroemer, G., Fidelle, M., Christodoulidis, S., Segata, N., Thomas, A.M., et al. (2021). Intestinal microbiota influences clinical outcome and side effects of early breast cancer treatment. Cell Death Differ. *28*, 2778–2796.

54. Ryu, S.W., Kim, J.S., Oh, B.S., Choi, W.J., Yu, S.Y., Bak, J.E., Park, S.H., Kang, S.W., Lee, J., Jung, W.Y., et al. (2022). Gut Microbiota Eubacterium callanderi Exerts Anti-Colorectal Cancer Activity. Microbiol. Spectr. *10*, e0253122.

55. Obuya, S., Elkholy, A., Avuthu, N., Behring, M., Bajpai, P., Agarwal, S., Kim, H.G., El-Nikhely, N., Akinyi, P., Orwa, J., et al. (2022). A signature of Prevotella copri and Faecalibacterium prausnitzii depletion, and a link with bacterial glutamate degradation in the Kenyan colorectal cancer patients. J. Gastrointest. Oncol. *13*, 2282–2292.

56. Xia, C., Cai, Y., Ren, S., and Xia, C. (2022). Role of microbes in colorectal cancer therapy: Cross-talk between the microbiome and tumor microenvironment. Front. Pharmacol. *13*, 1051330.

57. Dikeocha, I.J., Al-Kabsi, A.M., Chiu, H.-T., and Alshawsh, M.A. (2022). Faecalibacterium prausnitzii Ameliorates Colorectal Tumorigenesis and Suppresses Proliferation of HCT116 Colorectal Cancer Cells. Biomedicines *10*, 1128.

58. Ma, J., Sun, L., Liu, Y., Ren, H., Shen, Y., Bi, F., Zhang, T., and Wang, X. (2020). Alter between gut bacteria and blood metabolites and the anti-tumor effects of Faecalibacterium prausnitzii in breast cancer. BMC Microbiol. *20*, 82.

59. Chaput, N., Lepage, P., Coutzac, C., Soularue, E., Le Roux, K., Monot, C., Boselli, L., Routier, E., Cassard, L., Collins, M., et al. (2017). Baseline gut microbiota predicts clinical response and colitis in metastatic melanoma patients treated with ipilimumab. Ann. Oncol. *28*, 1368–1379.

60. Brevi, A., Cogrossi, L.L., Lorenzoni, M., Mattorre, B., and Bellone, M. (2022). The Insider: Impact of the Gut Microbiota on Cancer Immunity and Response to Therapies in Multiple Myeloma. Front. Immunol. *13*, 845422.

61. Gopalakrishnan, V., Spencer, C.N., Nezi, L., Reuben, A., Andrews, M.C., Karpinets, T.V., Prieto, P.A., Vicente, D., Hoffman, K., Wei, S.C., et al. (2018). Gut microbiome modulates response to anti–PD-1 immunotherapy in melanoma patients. Science *359*, 97–103.

62. Fernandes, A., Oliveira, A., Carvalho, A.L., Soares, R., and Barata, P. (2023). Faecalibacterium prausnitzii in Differentiated Thyroid Cancer Patients Treated with Radioiodine. Nutrients *15*, 2680.

63. Shi, J., Yang, Y., Xie, H., Wang, X., Wu, J., Long, J., Courtney, R., Shu, X.O., Zheng, W., Blot, W.J., and Cai, Q. (2021). Association of oral microbiota with lung cancer risk in a low-income population in the Southeastern USA. Cancer Causes Control. *32*, 1423–1432.

64. King, M., Hurley, H., Davidson, K.R., Dempsey, E.C., Barron, M.A., Chan, E.D., and Frey, A. (2020). The Link between *Fusobacteria* and Colon Cancer: a Fulminant Example and Review of the Evidence. Immune Netw. *20*, e30.

65. Castellarin, M., Warren, R.L., Freeman, J.D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R.A., and Holt, R.A. (2012). Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. *22*, 299–306.

66. Ou, S., Wang, H., Tao, Y., Luo, K., Ye, J., Ran, S., Guan, Z., Wang, Y., Hu, H., and Huang, R. (2022). Fusobacterium nucleatum and colorectal cancer: From phenomenon to mechanism. Front. Cell. Infect. Microbiol. *12*, 1020583.

67. Wang, S., Liu, Y., Li, J., Zhao, L., Yan, W., Lin, B., Guo, X., and Wei, Y. (2021). Fusobacterium nucleatum Acts as a Pro-carcinogenic Bacterium in Colorectal Cancer: From Association to Causality. Front. Cell Dev. Biol. *9*, 710165.

68. Wang, N., and Fang, J.Y. (2023). Fusobacterium nucleatum, a key pathogenic factor and microbial biomarker for colorectal cancer. Trends Microbiol. *31*, 159–172.

69. Zhang, J., He, Y., Xia, L., Yi, J., Wang, Z., Zhao, Y., Song, X., Li, J., Liu, H., Liang, X., et al. (2022). Expansion of Colorectal Cancer Biomarkers Based on Gut Bacteria and Viruses. Cancers *14*, 4662.

70. Xia, X., Wu, W.K.K., Wong, S.H., Liu, D., Kwong, T.N.Y., Nakatsu, G., Yan, P.S., Chuang, Y.M., Chan, M.W.Y., Coker, O.O., et al. (2020). Bacteria pathogens drive host colonic epithelial cell promoter hypermethylation of tumor suppressor genes in colorectal cancer. Microbiome *8*, 108.

71. Osman, M.A., Neoh, H.M., Ab Mutalib, N.S., Chin, S.F., Mazlan, L., Raja Ali, R.A., Zakaria, A.D., Ngiu, C.S., Ang, M.Y., and Jamal, R. (2021). Parvimonas micra, Peptostreptococcus stomatis, Fusobacterium nucleatum and Akkermansia muciniphila as a four-bacteria biomarker panel of colorectal cancer. Sci. Rep. *11*, 2925.

72. Kharofa, J., Apewokin, S., Alenghat, T., and Ollberding, N.J. (2023). Metagenomic analysis of the fecal microbiome in colorectal cancer patients compared to healthy controls as a function of age. Cancer Med. *12*, 2945–2957.

73. Hosseini, S.S., Goudarzi, H., Ghalavand, Z., Hajikhani, B., Rafeieiatani, Z., and Hakemi-Vala, M. (2020). Anti-proliferative effects of cell wall, cytoplasmic extract of Lactococcus lactis and nisin through down-regulation of cyclin D1 on SW480 colorectal cancer cell line. Iran. J. Microbiol. *12*, 424–430. https://doi.org/10.18502/ijm.v12i5.4603.

74. Kaczmarek, K., Więckiewicz, J., Węglarczyk, K., Siedlar, M., and Baran, J. (2021). The Anti-Tumor Effect of Lactococcus lactis Bacteria-Secreting Human Soluble TRAIL Can Be Enhanced by Metformin Both In Vitro and In Vivo in a Mouse Model of Human Colorectal Cancer. Cancers *13*, 3004.

75. Li, X., He, L., Luo, J., Zheng, Y., Zhou, Y., Li, D., Zhang, Y., Pan, Z., Li, Y., and Tao, L. (2022). Paeniclostridium sordellii hemorrhagic toxin targets TMPRSS2 to induce colonic epithelial lesions. Nat. Commun. *13*, 4331.

76. Chang, Y., Huang, Z., Hou, F., Liu, Y., Wang, L., Wang, Z., Sun, Y., Pan, Z., Tan, Y., Ding, L., et al. (2023). Parvimonas micra activates the Ras/ERK/c-Fos pathway by upregulating miR-218-5p to promote colorectal cancer progression. J. Exp. Clin. Cancer Res. *42*, 13.

77. Zhao, L., Zhang, X., Zhou, Y., Fu, K., Lau, H.C.H., Chun, T.W.Y., Cheung, A.H.K., Coker, O.O., Wei, H., Wu, W.K.K., et al. (2022). Parvimonas micra promotes colorectal tumorigenesis and is associated with prognosis of colorectal cancer patients. Oncogene *41*, 4200–4210.

78. Hatta, M.N.A., Mohamad Hanif, E.A., Chin, S.-F., Low, T.Y., and Neoh, H.M. (2023). Parvimonas micra infection enhances proliferation, wound healing, and inflammation of a colorectal cancer cell line. Biosci. Rep. *43*, BSR20230609.

79. Löwenmark, T., Li, X., Löfgren-Burström, A., Zingmark, C., Ling, A., Kellgren, T.G., Larsson, P., Ljuslinder, I., Wai, S.N., Edin, S., and Palmqvist, R. (2022). Parvimonas micra is associated with tumour immune profiles in molecular subtypes of colorectal cancer. Cancer Immunol. Immunother. *71*, 2565–2575.

80. Zi, M., Zhang, Y., Hu, C., Zhang, S., Chen, J., Yuan, L., and Cheng, X. (2022). A literature review on the potential clinical implications of streptococci in gastric cancer. Front. Microbiol. *13*, 1010465.

81. Nieminen, M.T., Listyarifah, D., Hagström, J., Haglund, C., Grenier, D., Nordström, D., Uitto, V.J., Hernandez, M., Yucel-Lindberg, T., Tervahartiala, T., et al. (2018). Treponema denticola chymotrypsin-like proteinase may contribute to orodigestive carcinogenesis through immunomodulation. Br. J. Cancer *118*, 428–434.

82. Kylmä, A.K., Jouhi, L., Listyarifah, D., Mohamed, H., Mäkitie, A., Remes, S.M., Haglund, C., Atula, T., Nieminen, M.T., Sorsa, T., and Hagström, J. (2018). Treponema denticola chymotrypsin-like protease as associated with HPV-negative oropharyngeal squamous cell carcinoma. Br. J. Cancer *119*, 89–95.

83. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319.

84. Gupta, N., Bandeira, N., Keich, U., and Pevzner, P.A. (2011). Target-decoy approach and false discovery rate: when things may go wrong. J. Am. Soc. Mass Spectrom. *22*, 1111–1120.

85. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. *48*, W449–W454.

86. Rist, M.J., Theodossis, A., Croft, N.P., Neller, M.A., Welland, A., Chen, Z., Sullivan, L.C., Burrows, J.M., Miles, J.J., Brennan, R.M., et al. (2013). HLA Peptide Length Preferences Control CD8+ T Cell Responses. J. Immunol. *191*, 561–571.

87. Johnson, D., Boyes, B., Fields, T., Kopkin, R., and Orlando, R. (2013). Optimization of data-dependent acquisition parameters for coupling high- speed separations with LC-MS/MS for protein identifications. J. Biomol. Tech. *24*, 62–72.

88. Cornel, A.M., Mimpen, I.L., and Nierkens, S. (2020). MHC Class I Downregulation in Cancer: Underlying Mechanisms and Potential Targets for Cancer Immunotherapy. Cancers *12*, 1760.

89. Anderson, P., Aptsiauri, N., Ruiz-Cabello, F., and Garrido, F. (2021). HLA class I loss in colorectal cancer: implications for immune escape and immunotherapy. Cell. Mol. Immunol. *18*, 556–565.

90. Hazini, A., Fisher, K., and Seymour, L. (2021). Deregulation of HLA-I in cancer and its central importance for immunotherapy. J. Immunother. Cancer *9*, e002899.

91. Guo, H.-L., Chen, G., Song, Z.L., Sun, J., Gao, X.H., and Han, Y.X. (2020). COL6A3 promotes cellular malignancy of osteosarcoma by activating the PI3K/AKT pathway. Rev. Assoc. Med. Bras. *66*, 740–745.

92. Wang, J., and Pan, W. (2020). The Biological Role of the Collagen Alpha-3 (VI) Chain and Its Cleaved C5 Domain Fragment Endotrophin in Cancer. OncoTargets Ther. *13*, 5779–5793.

93. Martinez-Useros, J., Garcia-Carbonero, N., Li, W., Fernandez-Aceñero, M.J., Cristobal, I., Rincon, R., Rodriguez-Remirez, M., Borrero-Palacios, A., and Garcia-Foncillas, J. (2019). UNR/CSDE1 Expression Is Critical to Maintain Invasive Phenotype of Colorectal Cancer through Regulation of c-MYC and Epithelial-to-Mesenchymal Transition. J. Clin. Med. *8*, 560.

94. Lin, Y. (2021). Identification of CTL Epitopes on Efflux Pumps of the ATP-Binding Cassette and the Major Facilitator Superfamily of Mycobacterium tuberculosis. J. Immunol. Res. *1–13*.

95. Kalaora, S., Nagler, A., Nejman, D., Alon, M., Barbolin, C., Barnea, E., Ketelaars, S.L.C., Cheng, K., Vervier, K., Shental, N., et al. (2021). Identification of bacteria-derived HLA-bound peptides in melanoma. Nature *592*, 138–143.

96. Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D.K., Sette, A., and Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. *47*, D339–D343.

97. Tadros, D.M., Eggenschwiler, S., Racle, J., and Gfeller, D. (2023). The MHC Motif Atlas: a database of MHC binding specificities and ligands. Nucleic Acids Res. *51*, D428–D437.

98. Marcu, A., Bichmann, L., Kuchenbecker, L., Kowalewski, D.J., Freudenmann, L.K., Backert, L., Mühlenbruch, L., Szolek, A., Lübke, M., Wagner, P., et al. (2021). HLA Ligand Atlas: A benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. J. Immunother. Cancer *9*, e002071.

99. Cai, Y., Lv, D., Li, D., Yin, J., Ma, Y., Luo, Y., Fu, L., Ding, N., Li, Y., Pan, Z., et al. (2023). IEAtlas: an atlas of HLA-presented immune epitopes derived from non-coding regions. Nucleic Acids Res. *51*, D409–D417.

100. Tan, X., Li, D., Huang, P., Jian, X., Wan, H., Wang, G., Li, Y., Ouyang, J., Lin, Y., and Xie, L. (2020). dbPepNeo: a manually curated database for human tumor neoantigen peptides. Database *2020*, baaa004.

101. Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L.T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., et al. (2020). The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science *368*, 973–980.

102. Fu, A., Yao, B., Dong, T., Chen, Y., Yao, J., Liu, Y., Li, H., Bai, H., Liu, X., Zhang, Y., et al. (2022). Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. Cell *185*, 1356–1372.

103. Riquelme, E., Zhang, Y., Zhang, L., Montiel, M., Zoltan, M., Dong, W., Quesada, P., Sahin, I., Chandra, V., San Lucas, A., et al. (2019). Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes. Cell *178*, 795–806.

104. Roelands, J., Kuppen, P.J.K., Ahmed, E.I., Mall, R., Masoodi, T., Singh, P., Monaco, G., Raynaud, C., de Miranda, N.F.C.C., Ferraro, L., et al. (2023). An integrated tumor, immune and microbiome atlas of colon cancer. Nat. Med. *29*, 1273–1286.

105. Naghavian, R., Faigle, W., Oldrati, P., Wang, J., Toussaint, N.C., Qiu, Y., Medici, G., Wacker, M., Freudenmann, L.K., Bonté, P.E., et al. (2023). Microbial peptides activate tumour-infiltrating lymphocytes in glioblastoma. Nature *617*, 807–817.

106. Chen, H., Yu, C., Wu, H., Li, G., Li, C., Hong, W., Yang, X., Wang, H., and You, X. (2022). Recent Advances in Histidine Kinase-Targeted Antimicrobial Agents. Front. Chem. *10*, 866392.

107. Ficici, E., Zhou, W., Castellano, S., and Faraldo-Gómez, J.D. (2018). Broadly conserved Na^+-binding site in the N-lobe of prokaryotic multidrug MATE transporters. Proc. Natl. Acad. Sci. USA *115*, E6172–E6181.

108. Radchenko, M., Symersky, J., Nie, R., and Lu, M. (2015). Structural basis for the blockade of MATE multidrug efflux pumps. Nat. Commun. *6*, 7995.

109. Harvey, K.L., Jarocki, V.M., Charles, I.G., and Djordjevic, S.P. (2019). The Diverse Functional Roles of Elongation Factor Tu (EF-Tu) in Microbial Pathogenesis. Front. Microbiol. *10*, 2351.

110. Brennan, C.A., and Garrett, W.S. (2019). Fusobacterium nucleatum — symbiont, opportunist and oncobacterium. Nat. Rev. Microbiol. *17*, 156–166.

111. Elias, J.E., and Gygi, S.P. (2010). Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol. Biol. *604*, 55–71.

112. (2014). The Major Histocompatibility Complex. In Primer to the Immune ResponsePrimer to the Immune Response (Elsevier), pp. 143–159. https://doi.org/10.1016/B978-0-12-385245-8.00006-6.

113. Thorlacius-Ussing, J., Jensen, C., Nissen, N.I., Cox, T.R., Kalluri, R., Karsdal, M., and Willumsen, N. (2024). The collagen landscape in cancer: profiling collagens in tumors and in circulation reveals novel markers of cancer-associated fibroblast subtypes. J. Pathol. *262*, 22–36.

114. Lv, J., Zhou, Y., Zhou, N., Wang, Z., Chen, J., Chen, H., Wang, D., Zhou, L., Wei, K., Zhang, H., et al. (2023). Epigenetic modification of *CSDE1* locus dictates immune recognition of nascent tumorigenic cells. Sci. Transl. Med. *15*, eabq6024.

115. Baker, D.J., Arany, Z., Baur, J.A., Epstein, J.A., and June, C.H. (2023). CAR T therapy beyond cancer: the evolution of a living drug. Nature *619*, 707–715.

116. Pech, M., Karim, Z., Yamamoto, H., Kitakawa, M., Qin, Y., and Nierhaus, K.H. (2011). Elongation factor 4 (EF4/LepA) accelerates protein synthesis at increased Mg2+ concentrations. Proc. Natl. Acad. Sci. USA *108*, 3199–3203.

117. Widjaja, M., Harvey, K.L., Hagemann, L., Berry, I.J., Jarocki, V.M., Raymond, B.B.A., Tacchi, J.L., Gründel, A., Steele, J.R., Padula, M.P., et al. (2017). Elongation factor Tu is a multifunctional and processed moonlighting protein. Sci. Rep. *7*, 11227.

118. (2014). Antigen Processing and Presentation. In Primer to the Immune ResponsePrimer to the Immune Response (Elsevier), pp. 161–179. https://doi.org/10.1016/B978-0-12-385245-8.00007-8.

119. Lau, T.T.Y., Sefid Dashti, Z.J., Titmuss, E., Pender, A., Topham, J.T., Bridgers, J., Loree, J.M., Feng, X., Pleasance, E.D., Renouf, D.J., et al. (2022). The Neoantigen Landscape of the Coding and Noncoding Cancer Genome Space. J. Mol. Diagn. *24*, 609–618.

120. Cleyle, J., Hardy, M.P., Minati, R., Courcelles, M., Durette, C., Lanoix, J., Laverdure, J.P., Vincent, K., Perreault, C., and Thibault, P. (2022). Immunopeptidomic Analyses of Colorectal Cancers With and Without Microsatellite Instability. Mol. Cell. Proteomics *21*, 100228.

121. Laumont, C.M., Vincent, K., Hesnard, L., Audemard, É., Bonneil, É., Laverdure, J.P., Gendron, P., Courcelles, M., Hardy, M.P., Côté, C., et al. (2018). Noncoding regions are the main source of targetable tumor-specific antigens. Sci. Transl. Med. *10*, eaau5516.

122. Luo, Z., Hu, H., Liu, S., Zhang, Z., Li, Y., and Zhou, L. (2021). Comprehensive analysis of the translatome reveals the relationship between the translational and transcriptional control in high fat diet-induced liver steatosis. RNA Biol. *18*, 863–874.

123. Cai, C., Wang, P., Zhao, C., Lei, W., Chu, Z., Cai, Y., and An, G. (2021). The use of ribosome-nascent chain complex-seq to reveal the translated mRNA profile and the role of ASN1 in resistance to Verticillium wilt in cotton. Genomics *113*, 3872–3880.

124. Pardo-Palacios, F.J., Wang, D., Reese, F., Diekhans, M., Carbonell-Sala, S., Williams, B., Loveland, J.E., De María, M., Adams, M.S., Balderrama-Gutierrez, G., et al. (2024). Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. Nat. Methods *21*, 1349–1363. https://doi.org/10.1038/s41592-024-02298-3.

125. Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., Hua, C., Yang, F., Liu, W., Peng, C., et al. (2020). CNSA: a data repository for archiving omics data. Database *2020*, baaa055.

126. Chen, F.Z., You, L.J., Yang, F., Wang, L.N., Guo, X.Q., Gao, F., Hua, C., Tan, C., Fang, L., Shan, R.Q., et al. (2020). CNGBdb: China National GeneBank DataBase. Yi Chuan *42*, 799–809.

127. Chen, T., Chen, X., Zhang, S., Zhu, J., Tang, B., Wang, A., Dong, L., Zhang, Z., Yu, C., Sun, Y., et al. (2021). The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. Dev. Reprod. Biol. *19*, 578–583.

128. CNCB-NGDC Members and Partners (2023). Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2023. Nucleic Acids Res. *51*, D18–D28.

129. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531.

130. Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta *2*, e107.

131. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arxiv.1303.3997.

132. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 1–13.

133. Lu, J., Rincon, N., Wood, D.E., Breitwieser, F.P., Pockrandt, C., Langmead, B., Salzberg, S.L., and Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. Nat. Protoc. *17*, 2815–2839.

134. Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. *33*, 1152–1158.

135. Marsh, A.N., Sharma, V., Mani, S.K., Vitek, O., MacCoss, M.J., and MacLean, B.X. (2022). Skyline Batch: An Intuitive User Interface for Batch Processing with Skyline. J. Proteome Res. *21*, 289–294.

136. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968.

137. Li, K., Vaudel, M., Zhang, B., Ren, Y., and Wen, B. (2019). an integrative proteomics data viewer. Bioinformatics *35*, 1249–1251.

138. Huber, F., Verhoeven, S., Meijer, C., Spreeuw, H., Castilla, E., Geng, C., van der Hooft, J., Rogers, S., Belloum, A., Diblen, F., and Spaaks, J. (2020). matchms - processing and similarity evaluation of mass spectrometry data. J. Open Source Softw. *5*, 2411.

139. Andreatta, M., Alvarez, B., and Nielsen, M. (2017). GibbsCluster: Unsupervised clustering and alignment of peptide sequences. Nucleic Acids Res. *45*, W458–W463.

140. Thomsen, M.C.F., and Nielsen, M. (2012). Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. Nucleic Acids Res. *40*, W281–W287.

141. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.

142. Xiang, H., Zhang, L., Bu, F., Guan, X., Chen, L., Zhang, H., Zhao, Y., Chen, H., Zhang, W., Li, Y., et al. (2022). A Novel Proteogenomic Integration Strategy Expands the Breadth of Neo-Epitope Sources. Cancers *14*, 3016.

143. Bittremieux, W., Schmid, R., Huber, F., van der Hooft, J.J.J., Wang, M., and Dorrestein, P.C. (2022). Comparison of Cosine, Modified Cosine, and Neutral Loss Based Spectrum Alignment For Discovery of Structurally Related Molecules. J. Am. Soc. Mass Spectrom. *33*, 1733–1744.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-human-CD80-FITC	BioLegend	Cat#305205; RRID: AB_314501
Anti-human-CD83-FITC	BioLegend	Cat#305305; RRID: AB_314513
Anti-human-CD86-PE	BioLegend	Cat#305405; RRID: AB_314525
Anti-human-HLA-A/B/C-PE	BioLegend	Cat#311405; RRID: AB_314874
Anti-Human HLA-A/B/C Monomorphic Antibody (W6/32)	AtaGenix	Cat#FHM00110
Anti-human CD3 monoclonal antibody (CD3-2)	Thermo Fisher Scientific	Cat#14-0037-82; RRID: AB_467057
Biological samples		
tumor biopsies and paired adjacent normal tissues of colorectal cancer from patients, listed in Table S9	The Sixth Affiliated Hospital of Sun Yat-sen University	N/A
Peripheral Blood Mononuclear Cell from Donors, listed in Table S10	Beijing Yuhe Traditional Chinese and Western Medicine Comprehensive Rehabilitation Hospital	N/A
Chemicals, peptides, and recombinant proteins		
Phenylmethanesulfonyl fluoride (PMSF)	Beyotime	Cat#ST505
protein A Sepharose CL-4B	Cytiva	Cat#17096303
HIPP-T009 Lymphocyte Serum-free Medium	Bioengine	Cat#EXP0101302
HyClone™ Characterized Fetal Bovine Serum (FBS)	Cytiva	Cat#SH30071.02
Human GM-CSF Recombinant Protein	Thermo Fisher Scientific	Cat#300-03-1MG
Human IL-6 Recombinant Protein	Thermo Fisher Scientific	Cat#200-06-1MG
Human IL-1 beta Recombinant Protein	Thermo Fisher Scientific	Cat#200-01B-1MG
Human TNF-alpha Recombinant Protein	Thermo Fisher Scientific	Cat#300-01A-1MG
Human IL-4 Recombinant Protein	Thermo Fisher Scientific	Cat#200-04-1MG
Prostaglandin E2 (PGE-2)	PeproTech	Cat#3632464
Synthetic peptides for PRM identification and immunogenicity validation, listed in Table S11	GCAT bio	N/A
Critical commercial assays		
IFN-γ enzyme-linked immunospot (ELISpot) assay	Mabtech	
AllPrep DNA/RNA Mini Kit	Qiagen	Cat#80204
MGIEasy FS DNA Library Prep Set	MGI	Cat#1000006987
Deposited data		
Whole genome sequencing data	This study	CNGB: CNP0004656; GSA for human: HRA005229
Raw mass spectra data	This study	CNGB: CNP0005402
Original in-house codes	This study	https://doi.org/10.5281/zenodo.11312743
Reference proteomes from UniProt	The UniProt Consortium[129]	https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/
IEDB database	Randi Vita et al.[96]	https://www.iedb.org/database_export_v3.php
T Cell Epitope from IEDB	Randi Vita et al.[96]	https://www.iedb.org/database_export_v3.php
MHC Motif Atlas database	Daniel M. Tadros et al.[97]	http://mhcmotifatlas.org/data/classI/all_peptides.txt

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IEAtlas database	Yangyang Cai et al.[99]	http://bio-bigdata.hrbmu.edu.cn/IEAtlas/download.jsp
HLA Ligand Atlas database	Ana Marcu et al.[98]	https://hla-ligand-atlas.org/downloads
Cryptic Peptides from HLA Ligand Atlas	Ana Marcu et al.[98]	https://jitc.bmj.com/highwire/filestream/22897/field_highwire_adjunct_files/2/jitc-2020-002071supp003_data_supplement.xlsx
dbPepNeo database	Xiaoxiu Tan et al.[100]	http://www.biostatistics.online/dbPepNeo/download.html
BIC database	Kai-Pu Chen et al.[36]	http://bic.jhlab.tw/
Software and algorithms		
Fastp (version 0.20.1)	Shifu Chen et al.[130]	https://github.com/OpenGene/fastp
BWA (version 0.7.17)	Heng Li et al.[131]	https://github.com/lh3/bwa
Kraken2 (version 2.1.2)	Derrick E. Wood et al.[132]	https://github.com/DerrickWood/kraken2/tree/master
KrakenTools (version 1.2)	Jennifer Lu et al.[133]	https://github.com/jenniferlu717/KrakenTools/
POLYSOLVER	Sachet A Shukla et al.[134]	https://github.com/jason-weirather/hla-polysolver
MaxQuant (version 2.2.0)	Tyanova, S et al.[83]	https://www.maxquant.org/download_asset/maxquant/latest
Skyline (version 21.2)	Alexandra N. Marsh et al.[135,136]	https://skyline.ms/project/home/software/Skyline/begin.view
PDV (version 1.7.4)	Li Kai et al.[137]	https://github.com/wenbostar/PDV
Matchms (version 0.17.0)	Huber Florian et al.[138]	https://github.com/matchms/matchms
NetMHCpan (version 4.1)	Reynisson Birkir et al.[85]	https://services.healthtech.dtu.dk/services/NetMHCpan-4.1/
GibbsCluster (version 2.0)	Andreatta Massimo et al.[139]	https://services.healthtech.dtu.dk/services/GibbsCluster-2.0/
Seq2Logo (version 2.1)	Martin Christen Frølund Thomsen et al.[140]	https://services.healthtech.dtu.dk/services/Seq2Logo-2.1/
Other		
96-well filter microplate	Agilent	Cat#204495-100
Sep-Pak tC18 96-well Plate	Waters	Cat#186002321
Kimble™ Kontes™ Dounce Tissue Grinders	Thermo Fisher Scientific	Cat#K8853000015
Poly-Prep Chromatography Columns	Bio-Rad	Cat# 7311550
HyperCool 3055 centrifugal vacuum concentrators	Gyrozen	Cat#Hyper-HC3055
CD14 MicroBeads	Miltenyi Biotec	Cat#130-097-052
CD8 MicroBeads	Miltenyi Biotec	Cat#130-097-057

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human participants

Ten patients diagnosed with colon adenocarcinoma (n=9), or rectal adenocarcinoma(n=1) were selected. All patients signed an informed consent form to collect tumor biopsies and paired adjacent normal tissues. All samples were selected based on the surgical specimens obtained from the Sixth Affiliated Hospital of Sun Yat-sen University in 2021 under the approval of E2018032 by the Medical Ethics Committee. The detailed information of clinical characteristics was listed in Table S1. All related samples were flash-frozen in liquid nitrogen within 30 min after surgical resection and stored in at -80°C freezer until use.[141]

Six donors with specific HLA-I subtypes listed in Table S10 were selected at Beijing Yuhe Traditional Chinese and Western Medicine Comprehensive Rehabilitation Hospital. Human specimens were collected in accordance with an institutional review board-approved protocol, and written informed consent was obtained from the donors for the use of de-identified specimens.

All experiments involving the use of human tissue samples were performed in accordance with the WMA Declaration of Helsinki, CIOMS International Ethical Guidelines for Health-related Research Involving Humans 2016, applicable regulations/guidelines, and were approved by the Institutional Review Board of BGI.

METHOD DETAILS

Identification of microbe from samples based on whole genome sequencing (WGS) data

After rapid freezing in liquid nitrogen and milling, approximately 50 mg of each sample was used for DNA extraction using the AllPrep DNA/RNA Mini Kit (Qiagen). Qualified DNA was constructed using the MGIEasy FS DNA Library Prep Set (MGI). The resulting libraries were sequenced on the BGI DNBSEQ platform. Microbiome identifications were performed as follows: raw read sequences of WGS were processed using Fastp,[141] and all low-quality reads and adapters were removed. Acquired clean reads were mapped to the GRCh38 human genome reference using BWA.[131] Finally, all unaligned reads were used for microbe identification using Kraken2.[132]

Identification of HLA-I subtypes

The HLA-I subtypes of all patients were determined using the POLYSOLVER[134] software. For this analysis, BAM files, which had been previously aligned to the GRCh38 human genome reference, were utilized as input. The parameters for POLYSOLVER were configured to cater to the specific demographic and technical aspects of the study: -race: Asian, -includeFreq: 1, -build: hg38, -format: STDFQ, -insertCacl: 0.

Calculation of reads count and relative abundance in different samples and every level of taxonomy

Read count, relative abundance, and alpha and beta diversity were calculated using in-house scripts. In general, read counts were extracted from Kraken2 report files with modified scripts from KrakenTools,[133] the count of reads mapped to Homo sapiens were wiped off from each sample's node of H. sapiens as well as the total read counts for each sample. The relative abundance of different taxa from Domain to Species was calculated as the percentage of read counts corresponding to the total mapped read count.

HLA-I ligand identification based on database dependent analysing

Immunoprecipitation and liquid chromatography with tandem mass spectrometry (LC-MS/MS) methods to acquire spectra of HLA-I ligands was processed according to our previous protocol.[142]

Collection of HLA-I ligand through immunoprecipitation

Snap-frozen tissue samples were minced in lysis buffer and then homogenized on ice using Kimble Kontes Dounce Tissue Grinders. The lysis buffer consisted of phosphate-buffered saline (Gibco) supplemented with 0.5% sodium deoxycholate (Sigma-Aldrich), 2% octyl-β-D-glucopyranoside (Sigma-Aldrich), a protease inhibitor cocktail (Roche), 2 mM PMSF (Beyotime), 2 mM EDTA (Invitrogen), and 0.4 mM iodoacetamide (Sigma-Aldrich). The homogenates were centrifuged at 17,000 × g at 4°C for 50 min. The supernatant was used to immunoaffinity-purify HLA class I complexes using 0.2 mL of W6/32 antibody cross-linked to protein A Sepharose CL-4B beads (5 mg of antibody per 2 mL of beads) in a 96-well filter microplate (Agilent) pre-conditioned with 100% acetonitrile (ACN, Sigma-Aldrich), 0.1% trifluoroacetic acid (TFA, Sigma-Aldrich), and 100 mM Tris-HCl (Thermo Fisher Scientific), pH 8.0. The beads were washed sequentially at room temperature with 2 mL each of buffer A (150 mM NaCl, 20 mM Tris-HCl, pH 8.0), buffer B (400 mM NaCl, 20 mM Tris-HCl, pH 8.0), and buffer C (20 mM Tris-HCl, pH 8.0), followed by elution of the HLA class I complexes and bound peptides with 1% TFA, and then processed through Sep-Pak tC18 96-well cartridges (Waters), pre-washed with 80% ACN in 0.1% TFA, and equilibrated with 0.1% TFA. After washing the cartridges with 0.1% TFA and then with 2% ACN in 0.1% TFA, the peptides were eluted using 14% ACN in 0.1% TFA followed by 28% ACN in 0.1% TFA. The eluted peptides were dried using HyperCool 3055 centrifugal vacuum concentrators (Gyrozen) and stored at −20°C for further analysis.

Acquire mass spectra of HLA-I ligand through LC-MS/MS

Collected peptides were reconstituted in 5% acetonitrile (ACN) with 0.1% formic acid and analyzed using liquid chromatography with tandem mass spectrometry (LC-MS/MS). The peptides were loaded onto a 300 μm × 5 mm trapping column (Acclaim PepMap 100 C18 HPLC column, 5 μm, 100 Å, Thermo Fisher Scientific) at a flow rate of 0.5 μL/min for 5 min. Subsequent separation occurred on a C18 column (150 μm × 35 cm, 1.8 μm particle size, Welch) connected to an Ultimate 3000 RSLCnano system (Thermo Fisher Scientific). A non-linear gradient of ACN, ranging from 5% to 80% over 90 min at a flow rate of 0.5 μL/min, was used for peptide elution. The eluted peptides were ionized via nanospray ionization and analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) in data-dependent acquisition (DDA) mode. Full-scan mass spectra were acquired in the Orbitrap over a range of 350-1,500 m/z at a resolution of 60,000. The top 30 most intense precursor ions were selected for fragmentation by higher-energy collisional dissociation (HCD) with a normalized collision energy (NCE) of 30%. The resulting fragments were detected in the Orbitrap at a resolution of 15,000. A dynamic exclusion period of 30 s was applied to minimize redundant precursor selection. The

automatic gain control (AGC) targets for MS1 and MS2 scans were set to 1.0×10^5 and 2.0×10^4, respectively, with maximum injection times of 50 ms for both MS1 and MS2 scans.

Construction of personalized database

The results generated by LC-MS/MS were decoded using a database searching method based on a patient-specific personalized protein database (PPD). A specific PPD contains three parts: proteins of *H. sapiens*, microbes with significantly different abundance between tumor and normal samples, and other dominant microbe species (relative abundance in tumor samples exceeded the median of all species), whose relative abundance in specific patients was at least 2-fold higher or only identified in tumor samples. All proteins of specific species or subspecies were downloaded from the UniProt[129] database, and microbial species whose proteins were not published in the database were ignored.

Database searching with MaxQuant

Output raw spectra were searched against the PPD using MaxQuant.[83] The tolerances of the precursor and fragment ions were set at 10 ppm and 0.05 Da, respectively. Oxidation (M) and deamination (NQ) were selected as variable modifications, and Carbamidomethyl (C) was set as a fixed modification. False discovery rate at the peptide level was set at 1%. The digestion mode was set to unspecific. Only peptides with lengths of 8–15 amino acids were retained. peptide-spectrum matches (PSMs) whose false discovery rate (FDR) was lower than 1% and posterior error probability (PEP) did not exceed 0.05 remained as potential HLA-I ligands.

Classification of identified peptides

Peptide sequences identified by filtered PSMs were categorized as human-derived, microbe-derived, or overlapping peptides, according to the origin of the protein and species listed in PPD. Due to the indistinguishable of leucine (L) with isoleucine (I) for the same mass difference from other amino acids, peptide sequences were remapped to protein sequences in each patient's PPD with in-house scripts after replacement of I with L in both peptide sequences and protein sequences. The classification of peptide sequences was then determined based on the protein sources identified through this remapping process.

Comparison with epitopes in public databases, articles and patents

Seven databases containing known epitopes, as detailed in the key resources table, were utilized. Specifically, the 'T cell Epitope' dataset was obtained from IED,[96] and the 'Cryptic Peptides' dataset was sourced from supplemental information of published articles associated with HLA Ligand Atlas.[98] From these datasets, only epitopes classified as 'Linear peptide' were selected for inclusion in our analysis. To ensure accurate peptide matching between the datasets from this study and the downloaded databases, all instances of the I was replaced with L due to their biochemical similarity and potential for interchangeability in mass spectrometry data. We conducted articles and patents searches using PubMed and google patents (https://patents.google.com/), respectively, using the peptide sequence as search term.

Affinity prediction and motif analysis of immunopeptidome

Binding affinity of peptides to HLA-I were predicted using NetMHCpan[85] with all HLA-I subtypes of samples. Peptides were categorized based on their lowest predicted binding affinity percentile rank value across different HLA-I subtypes as strong binders (SB, <0.5%), weak binders (WB, 0.5%–2%) or no binders (NB, >2%). For clustering analysis, GibbsCluster[139] was employed, with the number of clusters set to '1-x'. Here, 'x' represents the smallest integer greater than the number of HLA-I subtypes present in each sample. The final number of clusters was determined based on the Kullbach-Leibler distance. Sequence logos, create using Seq2Logo[140] following clustering, were compared with binding motifs from the MHC motif Atals[97] to assess the similarity.

Verification of HLA-I ligands with parallel reaction monitoring with synthesized peptide

Synthetic peptides (>98% purity) were analyzed using the Orbitrap Eclipse Tribrid mass spectrometer by applying the parallel reaction monitoring (PRM) method under a 65 min gradient. Full scan spectra were measured with a resolution of 500,000 and a normalized AGC target of 200%, with a maximum injection time of 100 ms. The precursor ions were selected with a 0.7 m/z isolation window, fragmented by 30% HCD collision energy, and analyzed with a resolution of 60,000, standard AGC target and auto maximum injection time. The PRM data were processed using Skyline[135] and PDV[137] software to validate and visualize the mutated peptides. Modified cosine similarity (MCS)[143] of spectra from DDA and PRM was calculated with matchms.[138] Ligand verification was considered successful if MCS of paired Spectra exceeded 0.33. All Synthetic peptides were listed in Table S11.

Immunogenicity validation of bacterial peptides

Generation of specific DCs and $CD8^+$ T cells

Healthy donors were selected based on the predicted affinity of chosen peptides with various HLA-I complex types. Specifically, validation was conducted on at least one donor whose HLA type included the same subtype with the lowest predicted binding affinity rank value from the corresponding patients. $CD14^+$ monocytes and $CD8^+$ T cells were isolated from 1×10^8 PBMCs of healthy donors using CD14 and CD8 MicroBeads (Miltenyi Biotec), respectively. The isolated $CD8^+$ T cells were cryopreserved in liquid nitrogen for later use. Dendritic cells (DCs) were generated by culturing $CD14^+$ monocytes (1×10^7 cells) in HIPP-T009 medium (Bioengine)

supplemented with 2% FBS (Cytiva), 800 U/mL GM-CSF (Thermo Fisher Scientific) and 1000 U/mL IL-4 for 4 days. Maturation was induced by replacing the medium with HIPP-T009 containing 2% FBS, 20 ng/mL IL-6, 20 ng/mL IL-1β, 40 ng/mL TNF-α, 1 μg/mL PGE-2, 800 U/mL GM-CSF, and 1000 U/mL IL-4 for 2 additional days, confirmed by flow cytometry with Anti-human-CD80-FITC, Anti-human-CD83-PE, Anti-human-CD86-PE, and Anti-human-HLA-A/B/C-PE (BioLegend).

DCs were loaded with synthetic peptides at 10 μg/mL concentration and incubated overnight at 37°C. Excess peptides were removed by centrifugation and washing twice in HIPP-T009 medium with 2% FBS and 30 ng/mL IL-21. Loaded DCs were then co-cultured with autologous $CD8^+$ T cells at an effector-to-target cell ratio of 5:1 in HIPP-T009 medium supplemented with 2% FBS and 30 ng/mL IL-21 for 48 h. Activated $CD8^+$ T cells were then cultured in fresh HIPP-T009 medium containing 2% FBS, 10 ng/mL IL-2, 10 ng/mL IL-7, and 10 ng/mL IL-15 for one week to promote proliferation. This process of antigenic stimulation with DCs and subsequent *in vitro* expansion was repeated twice to enhance the expansion of $CD8^+$ T cells specific to the peptides.

IFN-γ releasing testing by ELISpot assay

The secretion of IFN-γ by $CD8^+$ T Cells was tested using IFN-γ enzyme-linked immunospot (ELISpot) assay (Mabtech). Briefly, $CD8^+$ T cells that had been stimulated and proliferated were allowed to rest in HIPP-T009 medium containing 2% FBS for 24 h. Approximately 1×10^5 $CD8^+$ T cells were then seeded into each well of an ELISpot plate, together with 2×10^4 DCs pulsed with the peptide of interest. The cells were co-cultured in HIPP-T009 medium with 2% FBS for 20 h. DMSO and Anti-human CD3 monoclonal antibody (CD3-2) were served as the negative and positive control, respectively. After the co-culture period, immunospots were visualized and quantified using the AID iSpot ELISPot reader (AID-Autoimmun Diagnostika GmbH). Three technical replicates were performed for each peptide-donor pair during the co-culturing and imaging stages.

QUANTIFICATION AND STATISTICAL ANALYSIS

All Statistical analysis in the script was performed using 'ggsignif' package (v0.6.4) in R (v 3.6.3) with two-side Student's t-test. The significance levels are represented as follows: * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), and **** ($p < 0.0001$).