BMC Cancer

RESEARCH

Open Access

Genomic investigation of innate sensing pathways in the tumor microenvironment

Gabriella Quinn[1], Gianna Maggiore[2] and Bo Li[1,3,4,5,6*]

Abstract

The innate immune system is the first responder to infectious agents, cellular debris, and cancerous growths. This system plays critical roles in the antitumor immune responses by boosting and priming T cell-mediated cytotoxicity but is understudied due to the complexity and redundancy of its various downstream signaling cascades. We utilized a mathematical tool to holistically quantify innate immune signaling cascades and immunophenotype over 8,000 tumors from The Cancer Genome Atlas (TCGA). We found that innate immune activation was predictive of patient mortality in a subset of cancers. Further analysis identified PHF genes as transcripts that were associated with genomic stability and innate activation. Knockdown of PHF gene transcripts in vitro led to an increase in cell death and *IFNB1* expression in a cGAS-dependent manner, validating PHF genes as potential anti-tumor targets. We also found an association between innate immune activation and both tumor immunogenicity and intratumor microbes, which highlights the versatility of this model. In conclusion, interrogating activation of innate immune signaling cascades demonstrated the importance of studying innate signaling in cancer and broadened the search for new therapeutic adjuvants.

Key points

1. The custom ssGSEA algorithm presented in this article is an effective tool for estimating innate immune activation
2. This algorithm highlighted a new target to increase cGAS signaling in cancer cell lines.
3. In colorectal cancer, innate immunity was associated with tumor immunogenicity.
4. Innate immunity demonstrated weak associations with intratumor microbial abundance.

Keywords TCGA, Innate immunity, CGAS-STING, Tumor microenvironment

*Correspondence:
Bo Li
lib3@chop.edu
[1] Department of Immunology, UT Southwestern Medical Center, Dallas, TX, USA
[2] Children's Research Institute, UT Southwestern Medical Center, Dallas, TX, USA
[3] Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center, Dallas, TX, USA
[4] Harold C. Simmons Comprehensive Cancer Center, UT Southwestern Medical Center, Dallas, TX, USA
[5] Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[6] Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, USA

Introduction

With the increasing interest and utilization of microbes as cancer therapy adjuvants [1, 2], the study of innate immune system activation in the tumor microenvironment is of high interest. One well established microbial adjuvant is the attenuated strain of *Mycobacterium,* bacilli Calmette-Guerin (BCG) in bladder cancer treatment [3]. In more recent years, microbes have been engineered to secrete antigens or specifically enter and accumulate in the tumor microenvironment [1, 2]. The efficacy of this strategy may lie in the subsequent activation of the innate immune system in a previously immunosuppressed environment.

In cancer, the innate immune system is tasked with sensing DAMPs and PAMPs in the tumor microenvironment and subsequently producing cytokines and displaying neoantigens to the adaptive immune system [4]. This preliminary alarm system is essential for mounting a proper immune response, but the study of the innate immune system in cancer has been hindered by its complexity. This system consists of dozens of PRRs that are triggered by an even more diverse group of ligands. Further, the signaling cascades of the innate immune system consist of many intertwining pathways that converge and diverge in their secondary messengers and protein modifications. Due to this entanglement, the PRRs of the innate immune system are underexplored in the field of cancer immunology.

The increased accessibility of mathematical models and eruption of openly available genomic data allow complex topics to be more thoroughly explored. Gene set enrichment analysis (GSEA) is a tool used to calculate the enrichment of a biological pathway into a single score [5], commonly used to summarize RNA sequencing data. This tool can also be used to quantify the enrichment of a pathway within a single sample, also known as single-sample gene set enrichment analysis (ssGSEA) [6]. Using the ssGSEA algorithm with a custom ontology, our study investigated the level of activation of multiple innate immune pathways in the pan-cancer dataset from The Cancer Genome Atlas (TCGA). The innate immune pathways explored included toll-like receptor (TLR) signaling, C-type lectin receptor (CLR) signaling, retinoic acid induced gene (RIG-I) signaling, nucleotide binding oligomerization domain (NOD) signaling, and cyclic GMP-AMP synthase (cGAS) signaling. TLR's are common receptors of bacterial wall and membrane products as well as foreign DNA. CLR's recognize fungal and viral particles. NOD receptors are known to respond to bacterial cell wall components and self-antigens [7, 8]. RIG-I is a sensor for double stranded viral RNA [9]. Finally, cGAS recognizes cytosolic DNA from foreign sources and self [10]. All of these signaling pathways have the goal of triggering an immune response to resolve a threat, either in the form of a foreign invasion or cancerous growths.

The aim of this study is to highlight a new way of quantifying and interrogating the innate immune system. It offers a broad introduction to the various utilities of this score system and a foundation for future work in therapeutic targets and adjuvants. In this study, the five main pathways of the innate immune system, mentioned above, were quantified in over 8,000 tumors of 29 different cancer types. Results from this work feature the importance of the innate immune system in patient outcomes, its association with tumor immunogenicity, and its association with intratumor microbes. The insights from this model also identified a potential therapeutic target: PHD finger (PHF) proteins that can boost activity of the cGAS pathway and possibly the overall inflammatory state of a tumor.

Methods

The cancer genome atlas data acquisition

Normalized RNA sequencing data was acquired from Wang et al., *Scientific Data*, 2018. In brief, reads from TCGA tumors were aligned to the UCSC hg19 human reference, quantified and normalized via RSEM to allow for comparisons across multiple cancer types [11]. Copy number of each protein coding gene was acquired by R packages "TCGAbiolinks" and "RTCGAToolbox" (https://github.com/BioinformaticsFMRP/TCGAbiolinks; https://github.com/mksamur/RTCGAToolbox) using the "getFirehoseData()" function. Intratumor microbe abundance was acquired from a SHOGUN analysis of the unaligned RNA sequencing reads from TCGA (ftp://ftp.microbio.me/pub/cancer_microbiome_analysis/) [12]. All data normalization and handling were performed in R for appropriate downstream analysis.

Defining a custom ontology

A comprehensive gene list for each of the five innate immune pathways; cGAS, CLR, TLR, NOD, RIG-I; was manually curated using a variety of publicly available databases and literature reviews [7, 13–17]. Gene lists were first established based off large and commonly used ontologies including KEGG, Gene Ontology, and STRING [13–15]. Gene sets were summarized from all three lists to ensure there were no duplicate gene names. Gene names were also all individually verified with the UCSC hg19 annotation files to ensure proper gene name usage. Various review articles were used to supplement each summarized gene list [7, 16, 17]. Final gene lists for each of the five innate pathways can be found in Supplemental Table 1. The gene list was compiled into a ".gmt" file for use in ssGSEA score calculations.

ssGSEA score calculation

An activation score for each of the five pathways was calculated for each RSEM normalized RNA sequencing result following the "single sample extension" of GSEA as described in Barbie et al. [18]. The R script for these calculations was obtained via GitHub repository (https://github.com/broadinstitute/ssGSEA2.0). Each score was calculated using a raw rank metric, with a weight parameter of 75%. A Kolmogorov–Smirnov statistic was calculated with 1000 permutations. The normalized enrichment scores (NES) were compiled and used for downstream analyses.

siRNA treatments and interferon expression measurements

HCT116, Wildtype BJ, and cGAS knock out BJ cell lines (kindly provided by Dr. James Chen [19]) were cultured in Dulbecco's Modified Eagle Medium (DMEM) and 10% fetal bovine serum (FBS). All cell lines were grown in 37°C and 5% CO_2. ON-TARGETplus siRNA molecules were obtained from Horizon Discovery for PHF2 (L-012912–00-0005), PHF8 (L-004291–01-0005), and a non-targeting control (D-001810–10-05). Cells were treated with a final siRNA concentration of 25 nM using RNAiMAX (Invitrogen) as a transfection reagent. After treatment with siRNA for 48 h, cells were washed and harvested for subsequent RNA isolation and analysis. Cells were lysed and RNA was extracted with Trizol. cDNA was synthesized from RNA via the iScript cDNA synthesis kit (Bio-Rad 1,708,891). Relative expression was measured via RT-PCR with SYBR Green Reagents using a Bio-Rad CFX384 Real-Time System. Primers used to quantify expression of *PHF* and *IFN* transcripts are outlined in Supplemental Table 2.

Statistical analyses

All statistical analyses were performed in R or GraphPad Prism 9. Partial correlations were performed with the R package, "ppcor". Cox regression analyses were performed with the R package, "survival". ANOVA, Mann–Whitney, and Student t-tests were performed in GraphPad Prism 9 where appropriate. Multiple hypothesis testing was implemented using the Benjamini–Hochberg method in R.

Data availability

Previously published RNA-seq experiments used in this study can be found on Gene Expression Omnibus under the following accessions: GSE174141, GSE99298, GSE146009, and GSE17538. TCGA level-2 sequencing data were obtained via controlled access. TCGA level-3 and patient metadata were directly downloaded from the NCI GDC website (https://portal.gdc.cancer.gov). Copy numbers of genes in TCGA tumors were obtained using the "TCGAbiolinks" and "RTCGAToolbox" R packages. ssGSEA algorithm was obtained from its original authors' submission to GitHub (https://github.com/broadinstitute/ssGSEA2.0). Finally, intratumor microbe abundance values were obtained from Poore et al. [10] via ftp://ftp.microbio.me/pub/cancer_microbiome_analysis/.

Results

Quantification of innate immune activation through ssGSEA

We utilized of a mathematical tool called single sample gene set enrichment analysis (ssGSEA) to quantify 5 individual branches of intratumor innate immune activation using a custom ontology [5]. The quantification scores for each branch were then used to further elucidate various tumor phenotypes. In brief, ssGSEA generates a ranked gene list from which a single score is generated that summarizes expression of a signaling cascade. A gene set was established for the signaling cascades of toll-like receptors (TLR), C-type lectin receptors (CLR), nucleotide-binding and oligomerization domain (NOD)-like receptors (NLR), cyclic GMP-AMP synthase (cGAS), and retinoic acid-inducible gene I (RIG-I) based on curated databases and previous literature (See Methods, Supplemental Table 1) [7, 13–17]. After an ontology was established for each pattern recognition receptor, ssGSEA scores were calculated to summarize the degree of activation of each pathway from RNA expression levels (Fig. 1A).

To first validate that this strategy can quantitatively model innate immune activation, ssGSEA scores were calculated from a previously published sequencing experiment [20]. In Stothers et al., bone marrow derived macrophages (BMDM) were treated with lipopolysaccharide (LPS) for 4 or 72 h. LPS is a strong, natural ligand of TLR4, which triggers downstream signaling in the TLR pathway. Whole transcriptome RNA-sequencing was then performed for each condition. Proxies for TLR activation, such as *Il6* (Gene ID:16,193) and *Tnf* (Gene ID:7124) expression, were significantly elevated after LPS stimulation (Fig. 1B). We calculated the ssGSEA score for intracellular TLR pathway activation after LPS stimulation and we observed similar changes after 4 or 72 h (Fig. 1C). Next, we investigated another published dataset of samples exposed to viral stimulation [21]. A549 cells were infected with respiratory syncytial virus (RSV) for 24 or 48 h and RNA-sequencing was performed. Genes involved in the RIG-I and NOD pathways generally increased with longer viral exposure (Fig. 1D). Similarly, expression of interferon genes also increased

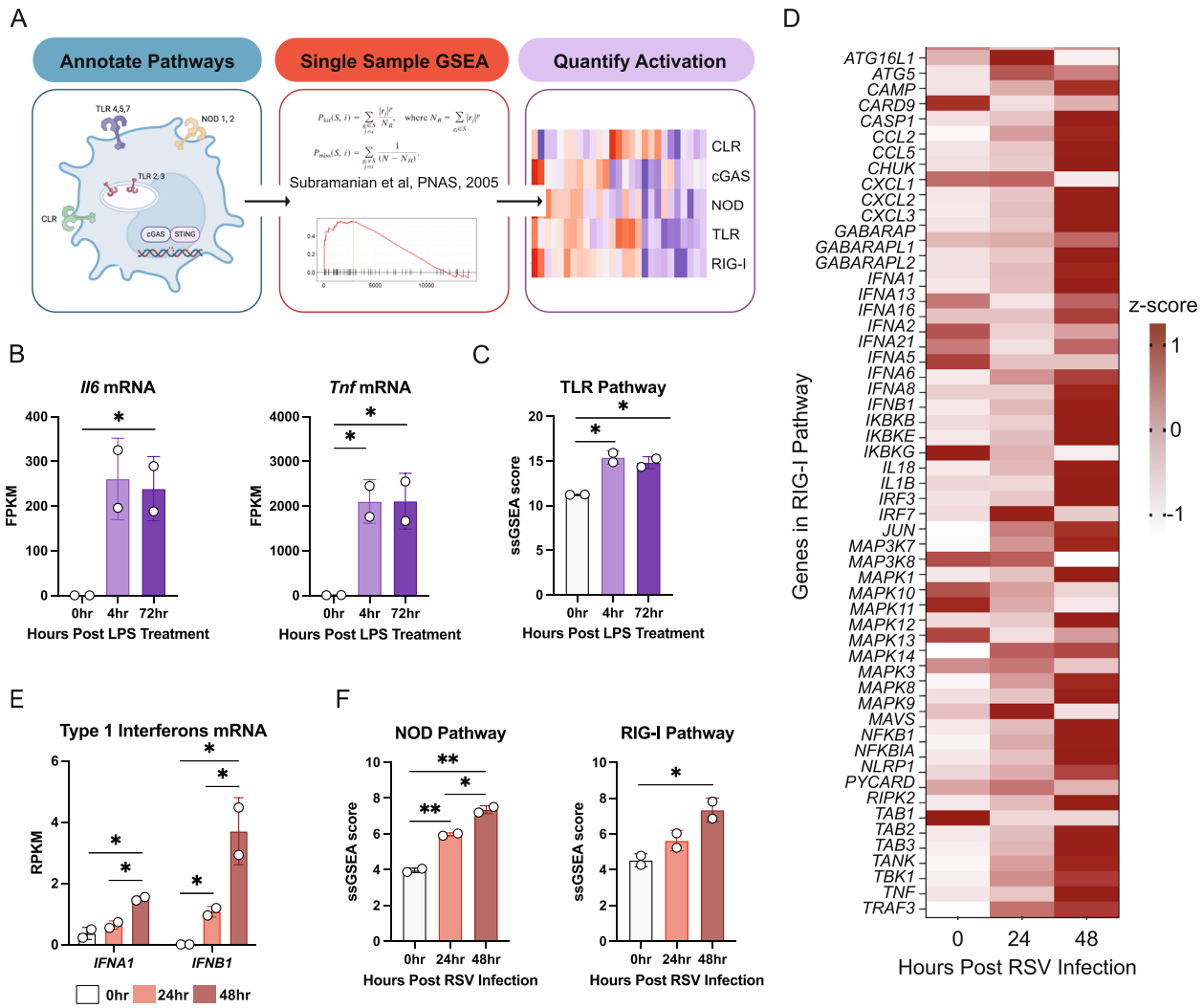

Fig. 1 Quantification of Innate Immune Pathways Using Single Sample Gene Set Enrichment Analysis (ssGSEA). **A** The workflow of generating ssGSEA scores for each of the 5 innate immune pathways. **B** mRNA expression of *Il6* and *Tnf* from mouse BMDMs at baseline, 4 h, and 72 h after LPS stimulation. **C** ssGSEA scores for TLR pathway calculated from mouse BMDMs at baseline, 4 h, and 72 h after LPS stimulation. **D** Heatmap of mRNA expression of genes in the RIG-I pathway, colored by z-score, from A549 cells treated with viral stimulus for 0, 24, and 48 h. All gene are annotated with official NCBI gene symbols. **E** RNA expression of type 1 interferons from A549 cells treated with a viral stimulus for 0, 24, and 48 h. **F** ssGSEA scores for NOD and RIG-I pathways calculated from A549 cells treated with viral stimulus for 0, 24, and 48 h. Comparisons that reached statistical significance were denoted by asterisks. Symbols: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$ from a Student's t-test. Abbreviations: TLR (Toll Like Receptor), CLR (C-type Lectin Receptor), RIG-I (Retinoic acid Inducible Gene I), NOD (Nucleotide Binding Oligomerization Domain), cGAS (Cyclic GMP-AMP Synthase), LPS (Lipopolysaccharide), FPKM (Fragments Per Kilobase of transcript per Million mapped reads), RPKM(Reads Per Kilobase of transcript, per Million mapped reads)

with viral exposure (Fig. 1E). ssGSEA scores of NOD and RIG-I activation unambiguously recapitulated this trend (Fig. 1F). Therefore, by summarizing the key genes involved in the PRR pathways, we demonstrated that ssGSEA is an effective tool to generate scores that reflect innate immune activation by both viral and bacterial stimuli.

Pan-cancer analysis of innate immune activation

To explore the functional impact of innate immune pathways in cancers, we calculated the custom ssGSEA scores for 8,554 human tumor samples, distributed among 29 cancer types in the Cancer Genomic Atlas (TCGA) cohort. Within each innate immune pathway, cancer types exhibited varying degrees of activation.

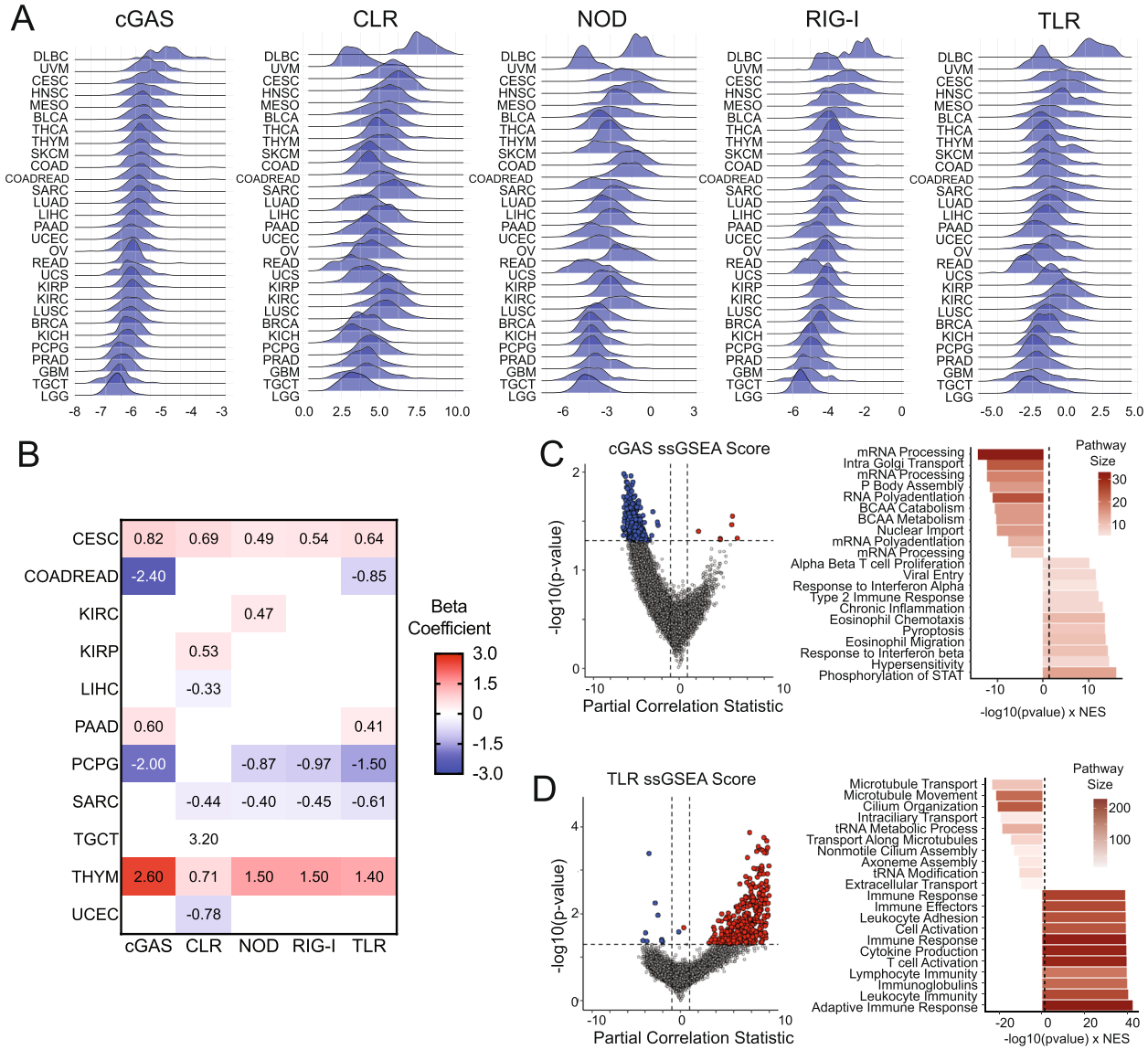

Fig. 2 Pan-Cancer Analysis of ssGSEA scores to Quantify Innate Immune System Activation. **A** Ridgeline plots of ssGSEA scores for five innate immune pathways compared across different cancer types. **B** A partial cox regression of ssGSEA scores and patient survival was calculated for each cancer type, which controlled for age, gender, pathological stage where applicable, tumor purity, and quantity of immune cell infiltrates. Beta values are shown only for regressions which met a statistical cut off *p* value < 0.05. **C** cGAS ssGSEA scores were associated with RNA sequencing data from all 29 cancer types and the statistics were averaged and plotted. GSEA was performed on a list of ranked r statistics from the pan-cancer associations and the top 10 downregulated and upregulated pathways are shown. **D** TLR ssGSEA scores were associated with RNA sequencing data from all 29 cancer types and the statistics were averaged and plotted. GSEA was performed on a list of ranked r statistics from the pan-cancer associations and the top 10 downregulated and upregulated pathways are shown. Abbreviations: TLR (Toll Like Receptor), CLR (C-type Lectin Receptor), RIG-I (Retinoic acid Inducible Gene I), NOD (Nucleotide Binding Oligomerization Domain), cGAS (Cyclic GMP-AMP Synthase)

For example, brain low grade glioma tumors (LGG) had demonstrably less PRR stimulation than cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) (Fig. 2A). We interrogated a possible association with innate immune scores and high-level patient demographics such as age and pathologic staging of disease. There were some weak correlations between innate immune scores and patients' age at diagnosis (Supplemental Fig. 1). We also investigated if there was a pattern of innate immune activation in various pathologic stages (Stage I-IV). There was no clear pattern between disease staging and our activation scores (Supplemental Fig. 2).

The association of innate immune activation scores and patient survival was investigated via partial cox

regression analyses, controlling for immune cell infiltration, tumor purity, patient age, and gender (Fig. 2B). In some cancer types, high innate immune activation was hazardous and in others advantageous. These results demonstrate the importance of the innate immune system in clinical outcomes and highlight the diversity of their influence.

ssGSEA scores were also associated with transcript levels of all protein coding genes, excluding genes already used to calculate the scores. For this analysis, tumor purity was controlled for as increasing immune cell infiltration will influence expression profiles. Associations were done across all cancer types and results were summarized to show pan-cancer trends. cGAS activation scores were negatively associated with 834 genes and positively associated with only 6 genes (Fig. 2C). The observation of decreased transcription when cGAS is highly activated is notable. In previous reports, DNA damage has been shown to induce global transcriptional stress by hindering RNA polymerase II [22]. Pathway enrichment analysis revealed genes involved in RNA processing were enriched in downregulated genes and genes involved in type 2 immunity were enriched in upregulated genes (Fig. 2C). Genes involved in mismatch repair and nucleotide excision repair also were negatively associated with cGAS activation (Supplemental Fig. 3).

TLR activation scores were associated with all transcripts in each tumor except for genes annotated to the TLR pathway in the same pan-cancer fashion. Most transcripts were positively associated with innate immune activation. Pathway enrichment analysis suggests microtubule assembly was negatively associated with TLR activation and cell-mediated immunity was positively associated as expected [23] (Fig. 2D).

Histone demethylases are negatively associated with genome instability and cGAS activation

Genes negatively associated with cGAS signaling in a pan-cancer fashion were of particular interest. cGAS signaling is commonly suppressed in tumors to evade the anti-cancer immune response [17]. Focusing on genes that are upregulated when cGAS activation is possibly suppressed, we hoped to uncover gene targets that, when interfered with, led to greater cGAS activation. Eleven candidate genes were found due to their strong association with cGAS activation. Among them were zinc fingers and chromatin modifiers (Fig. 3A). Two genes of interest include *PHF2* and *PHF8* which are Jumonji-C histone demethylases. *PHF2* has been shown to be crucial for genomic stability and participate in regulating expression of DNA damage repair and cell-cycle proteins [24, 25]. *PHF8* has been noted to have possible oncogenic properties [26, 27], interacts with DNA damage response (DDR) proteins, and encourages genomic stability similar to *PHF2* [28].

These previous observations regarding PHF proteins are also applicable in vivo. In tumors with high microsatellite instability, there was significantly less *PHF2* expression (Fig. 3B). When tumors were identified by either the *PHF2* or *PHF8* copy number individually there were inconsistent patterns of cGAS activation (Fig. 3C). Interestingly, when tumors were grouped based on both *PHF2* and *PHF8* copy number together, tumors that lost both copies of *PHF2* and one copy of *PHF8* had significantly elevated cGAS activation (Fig. 3C). Of note, there were no tumors in the database that had both copies of *PHF2* and *PHF8* simultaneously deleted. This further supports the hypothesis that *PHF2* and *PHF8* proteins impact genomic stability as seen in previous reports [24, 25, 28] and are potential suppressors of DNA damage, a ligand for cGAS activation.

These results are particularly exciting as there is conflicting evidence regarding *PHF2*'s influence in tumorigenesis. *PHF2* has been noted as a tumor suppressor or an oncogene depending on the examined cancer type [29–33]. When evaluated in a holistic manner and across multiple cancer models, PHF transcripts were identified as crucial genomic stabilizers that are associated with the cGAS-STING pathway activation. This strongly suggests that PHF proteins may be targeted to increase DNA damage and enhance cGAS activation to combat the immunosuppressive environment of malignancies [34].

To determine if reducing PHF expression alone is enough to increase cGAS activation, *PHF2* and *PHF8* were knocked down via siRNA's in HCT116 and BJ cell lines. Reduction of target PHF transcripts was confirmed via qPCR (Fig. 3D). When *PHF2* and *PHF8* were both downregulated, there was an increase in *IFNB1* gene expression, a gene highly induced by the cGAS-STING activation (Fig. 3E). In ad dition to an induction of *IFNB1* expression there is also more cell death as measured by lactate dehydrogenase (LDH) cytosolic leakage (Supplemental Fig. 4). To evaluate if this effect was indeed dependent on cGAS, a cGAS knockout fibroblast BJ cell line was used. As expected, wild type fibroblasts had a decrease in PHF expression with respective siRNA treatment and an increase in *IFNB1* expression (Fig. 3F). cGAS knockout cell lines did not increase *IFNB1* expression with reduction of PHF transcripts suggesting that this increase in *IFNB1* expression may be cGAS-dependent (Fig. 3G).

Innate immune activation is heterogeneous in colorectal cancer and is associated with tumor immunogenicity and immune cell exhaustion markers

Innate immune activation of colorectal cancer is of particular interest as this cancer type grows at the interface between tissue and trillions of bacteria in the microbiota.

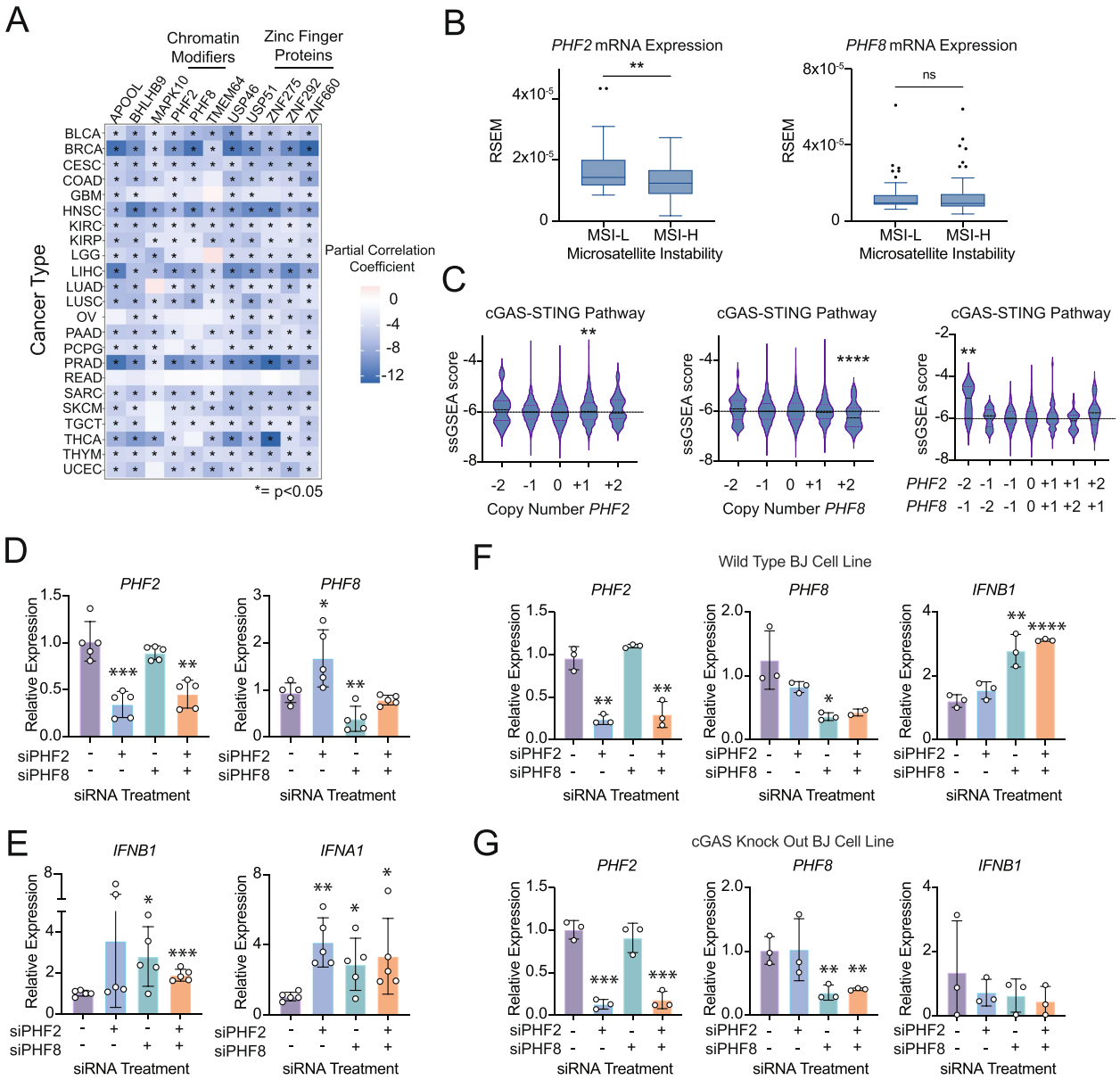

Fig. 3 cGAS ssGSEA scores Highlight Immunomodulatory Effects of Targeting PHF Proteins. **A** Heatmap of the partial correlation coefficients between the top 11 genes with cGAS ssGSEA scores while controlling for tumor purity and immune cell infiltration. Asterisks indicate a *p* value < 0.05. **B** PHF2 and PHF8 expression from tumors classified by microsatellite instability groupings. MSI-L refers to tumors with low microsatellite instability and MSI-H refers to tumors with high microsatellite instability. **C** cGAS activation scores of tumors stratified based off copy number of PHF2 and PHF8. The dotted line represents the mean cGAS activation score of tumors with the two copies of both PHF2 and PHF8. Statistical tests performed in comparison to tumors with 0 copy number of genes of interest. **D** PHF2 and PHF8 expression measured via qPCR in HCT116 cells after siRNA treatment ($n = 5$ replicates per condition). Statistical tests performed for each treatment group in comparison with Control Non-Targeting siRNA treatment. **E** IFNB1 and IFNA2 expression measured via qPCR in HCT116 cells after siRNA treatment ($n = 5$ replicates per condition). Statistical tests performed for each treatment group in comparison with Control Non-Targeting siRNA treatment. **F** PHF2, PHF8, and IFNB1 expression was measured via qPCR in wild type BJ fibroblasts after siRNA treatment ($n = 3$ replicates per condition). Statistical tests performed for each treatment group in comparison with Control Non-Targeting siRNA treatment. **G** PHF2, PHF8, and IFNB1 expression was measured via qPCR in cGAS knock out BJ fibroblasts after siRNA treatment ($n = 3$ replicates per condition). Statistical tests performed for each treatment group in comparison with Control Non-Targeting siRNA treatment. Symbols in panels B-C: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$ from a Mann–Whitney U test. Symbols in panels D-G: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$ from a Student's t test. Abbreviations: TLR (Toll Like Receptor), CLR (C-type Lectin Receptor), RIG-I (Retinoic acid Inducible Gene I), NOD (Nucleotide Binding Oligomerization Domain), cGAS (Cyclic GMP-AMP Synthase), RSEM (RNA-Seq by Expectation–Maximization)

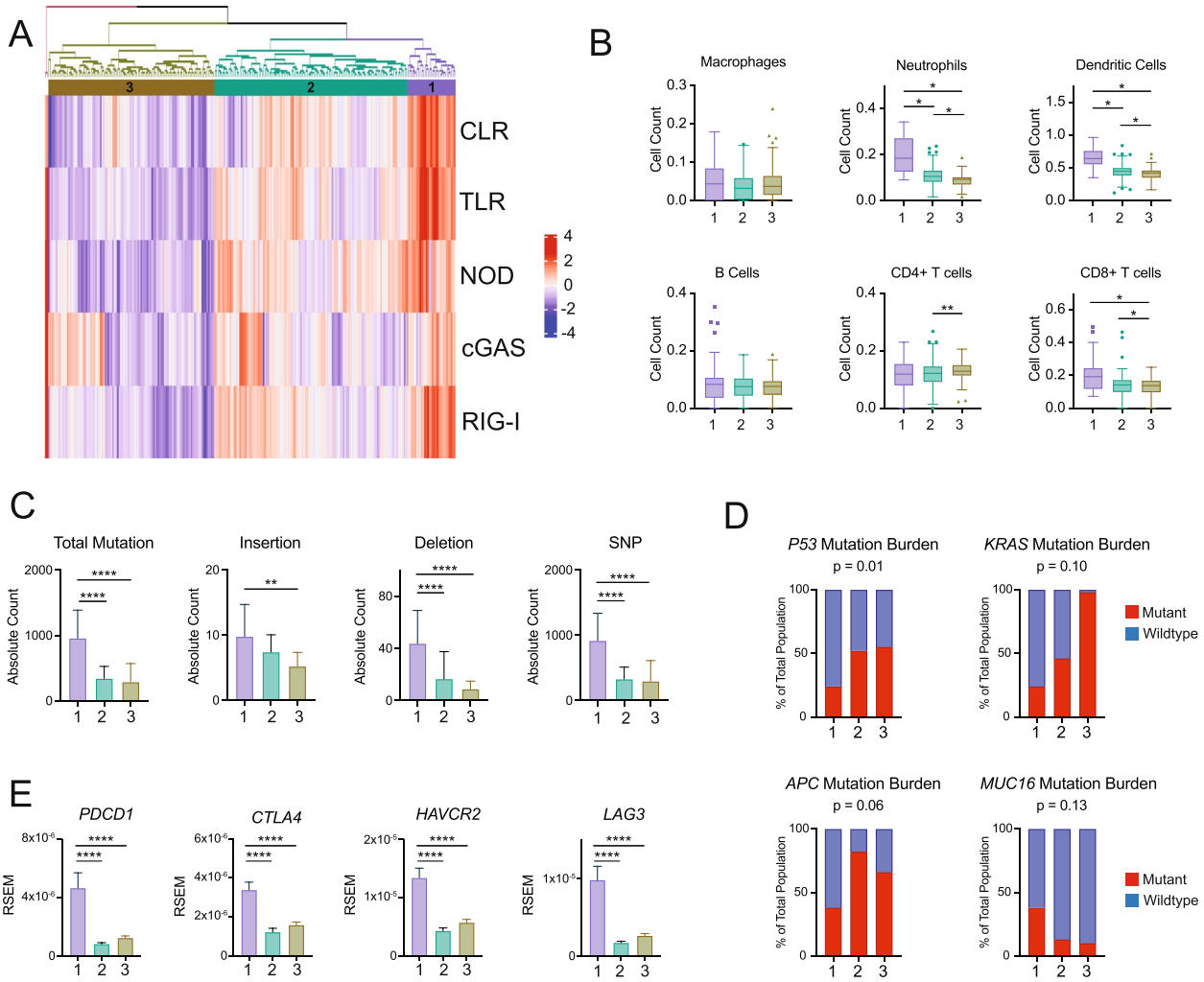

Fig. 4 Colon Adenocarcinomas Demonstrated Heterogenous Innate Immune Activation Associated with Tumor Immunogenicity and Immune Cell Exhaustion. **A** Clustered heatmap of *z*-scores of the ssGSEA scores of all 5 innate immune pathways in colon adenocarcinomas. **B** TIMER results of the quantification of immune cell populations within tumors in each of the 3 groups of innate immune activation. **C** Absolute count and class of each mutation event for the three innate group identified. **D** Mutations within the protein-coding region of P53, KRAS, APC, and MUC16 were quantified for each innate group and then translated into a ratio of mutated versus non-mutated tumors. Chi-Square test *p*-values shown above. **E** RSEM normalized expression values of immune cell exhaustion markers *PCDC1, CTLA4, HAVCR2,* and *LAG3* in each of the three innate immune groups. Symbols: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$ from a Mann–Whitney U test. Abbreviations: TLR (Toll Like Receptor), CLR (C-type Lectin Receptor), RIG-I (Retinoic acid Inducible Gene I), NOD (Nucleotide Binding Oligomerization Domain), cGAS (Cyclic GMP-AMP Synthase), HR (Hazard Ratio), FDR (False Discovery Rate), RSEM (RNA-Seq by Expectation–Maximization)

As expected, innate immune activation in colorectal cancer is largely heterogeneous (Fig. 4A). This result was investigated further by clustering tumors into 3 separate "innate groups" based on all 5 innate activation scores. Innate group 1 demonstrated the highest innate immune activity and group 3 demonstrated the lowest innate immune activation. To determine if higher immune activation corresponded with more immune cell infiltration, an algorithm to quantify immune cell populations in tumor tissue, TIMER was used [35]. Some, but not all, immune cell populations were elevated with innate immune signaling. Specifically, neutrophils, dendritic cells, and CD8 + T cells were significantly elevated in innate group 1 compared to innate group 3 (Fig. 4B).

The heterogeneity of innate activation in colon cancer tumors provided the opportunity to investigate the association between recognition by the innate immune system and tumor immunogenicity. Classically, a higher mutation burden in tumors is associated with more neoantigens [36]. This allows for greater recognition of

tumor cells by antigen presenting cells and CD8 + T cells. It has been hypothesized that pattern recognition receptors may also contribute to the immune recognition of self-antigens. TLR7 is a key contributor to development of anti-nuclear autoantibodies in systemic lupus erythematosus [37]. Additionally, NOD1 has been implicated as a key receptor in triggering cell death in malignancies. Finally, cGAS is known to recognize self-DNA [8, 38].

To further understand the association between tumor immunogenicity and the innate immune response, mutation burden was measured in the three innate groups. Innate group 1, which had the highest innate activation, demonstrated the largest mutation burden (Fig. 4C). This observation cannot be confounded by tumor purity as less pure tumors would have reduced power to detect mutations. In addition to global mutation burden, frequency of mutations in specific protein-coding regions commonly mutated in colorectal cancers was also significantly different between groups. *P53*, *KRAS*, and *APC* mutations were less abundant in group 1, while mutations in a putative tumor antigen *MUC16* (which encodes for CA125) were significantly higher in this group (Fig. 4D). This result is of particular interest as CA125 is a known prognostic marker for colon cancer and can be used in conjunction with CEA levels to screen for cancer progression and recurrence [39]. Although in our dataset, mutation burden of the four protein-coding genes was not associated with outcomes (Supplemental Fig. 5), *MUC16* is commonly mutated in colon cancer tumors [40]. *MUC16* is highly expressed in early-stage tumors [39], and associated with high innate immune activation. All three of these findings suggest the protein produced from the mutated gene of *MUC16* may be an interesting and valuable target for cancer vaccine development.

The "hot" tumors of innate group 1 also exhibited increased expression of immune cell exhaustion markers such as *PDCD1* (PD1), *CTLA4, HAVCR2* (TIM3), and *LAG3* (Fig. 4E). The association between high innate immune activity and exhaustion markers was recapitulated in a separate cohort of colorectal cancer as well (Supplemental Fig. 6). These data support the hypothesis that innate immune activation is associated with mutation frequency and therefore potentially tumor immunogenicity and immune cell exhaustion.

Associations with Innate Immune activation highlights a role of intratumor microbes in the tumor immunological microenvironment

Co-opting engineered microbes for cancer vaccines is an exciting new avenue for cancer therapy. Engineering microbes such as *E. coli Nissle* has the potential to reawaken the innate immune system and elicit a systemic anti-tumor response [1]. Having an accessible tool to summarize the innate immune system activation in the tumor microenvironment can be valuable to understanding the response upon vaccination. In this paper, we took advantage of the innate immune activation scores to interrogate the relationship between tumor associated microbes and the induction or lack thereof of pattern recognition receptors and their signaling cascades. Abundance measurements of intratumor microbes were sourced from Poore et al. [12]. Principal component analysis indicated that the composition of the intratumor microbes was distinct among different tumor sites as has been previously reported (Fig. 5A) [41]. For example, esophageal cancer had a particularly distinct taxonomy, with over half of all tumor-associated microbes belonging to Proteobacteria (Fig. 5B). To evaluate the relationship between the tumor microbiome and the innate immune activation, scores were associated with tumor microbe relative abundance. Colon adenocarcinoma and lung squamous cell carcinoma samples demonstrated multiple associations with multiple taxa, but the direction and degree of the associations was drastically different. In colon adenocarcinoma samples, there are distinct associations of innate activation scores with taxa commonly found in the gut such as Enterobacteriaceae and more specifically Escherichia (Fig. 5C) [42]. In lung squamous cell carcinoma on the other hand, there were few associations with these taxa and instead associations with Cloacibacterium, Alcanivorax, and Bacillus species (Fig. 5D). In addition, when a species is more abundant in the tumor, it is more likely to show associations with innate activation scores. For example, Escherichia is associated with cGAS activation in COAD tumors but not LUSC tumors and this is most likely because Escherichia species was more abundant in COAD tumors (Fig. 5E). Alternatively, LUSC tumors harbored more sequences belonging to Alcanivorax and this taxon showed a unique association with multiple innate activation scores (Fig. 5F). Of note, these associations are not as strong as we expected in respect to their R values. There may be other contributors that trigger strong innate immune signals in the tumor microenvironment, such as self-antigens mentioned previously. These results might provide a way to explain the conflicting observations from literature showing opposite effects of intratumor microbes in various cancer types [41].

Discussion

This study applies a mathematical tool to measure innate immune activation and thus phenotype 8,554 tumors across 29 different cancer types in the TCGA database. Different cancer types demonstrated various degrees of innate immune activation and in some cases, innate immune activation was predictive of patient outcomes.

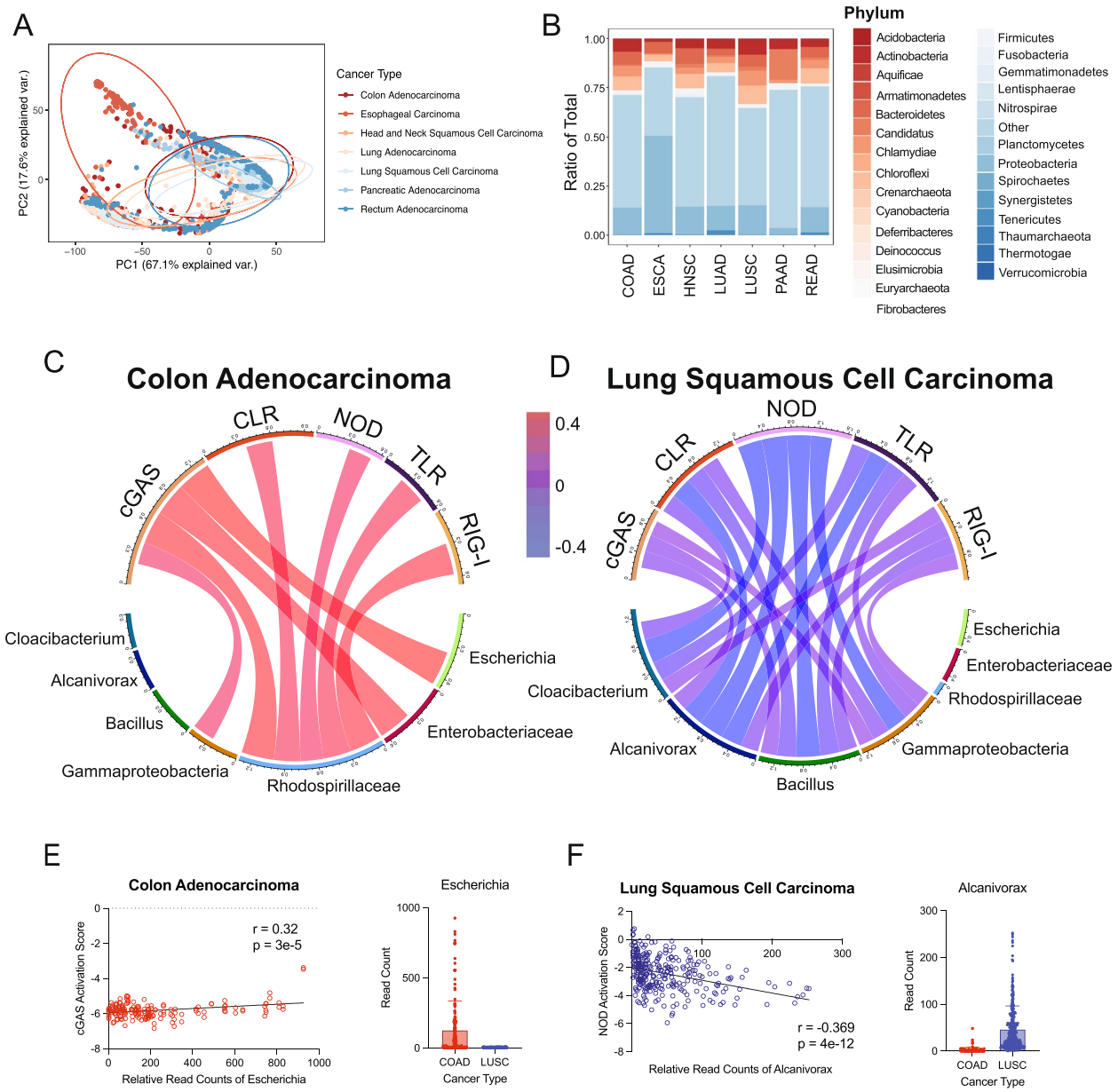

Fig. 5 Intratumor Microbes were Associated with Innate Immune Activation in Mucosal Associated Cancers. **A** Taxonomic information from tumors within 7 different cancer types was used to make a Bray–Curtis dissimilarity matrix and plotted via a principal component analysis (PCoA). **B** Stacked barplots of the relative abundance of each phylum of bacteria compared across 7 different cancer types. **C** Chord plot of the associations between intratumor microbe abundance and ssGSEA scores in colon adenocarcinoma. Associations are colored by correlation coefficients. **D** Chord plot of the associations between intratumor microbe abundance and ssGSEA scores in lung squamous cell carcinoma. Associations are colored by correlation coefficients. **E** Scatter plot of the relative abundance of the common resident gut genus, Escherichia, and cGAS activation in colon adenocarcinomas. Pearson R values and *P* values displayed on graph. (left) Barplot of read counts assigned to Escherichia from COAD and LUSC tumors (right). **F** Scatter plot of the relative abundance of genus Alcanivorax and NOD activation in lung squamous cell carcinoma. Pearson R values and *P* values displayed on graph. (left) Barplot of read counts assigned to Alcanivorax from COAD and LUSC tumors (right). Abbreviations: PC (Principal Component), TLR (Toll Like Receptor), CLR (C-type Lectin Receptor), RIG-I (Retinoic acid Inducible Gene I), NOD (Nucleotide Binding Oligomerization Domain), cGAS (Cyclic GMP-AMP Synthase)

This result emphasized the importance of the innate immune system in the antitumor response.

Immunophenotyping tumors based on specifically cGAS activation highlighted potential therapeutic targets. cGAS is a DNA sensing molecule that synthesizes the secondary messenger, cGAMP, which activates STING and facilitates its movement from the ER to the Golgi [38]. Multiple preclinical trials have attempted to target this pathway to boost antitumor immunity with varying degrees of success [43]. In this study, interrogating the cGAS signaling cascade as a complete pathway instead of as a singular protein led to new insights. This demonstrated an inverse relationship with global transcription in tumors and with chromosome stabilizing proteins, *PHF2* and *PHF8*. Carcinogenic cells classically have unstable genomes yet cGAS and STING are downregulated [44]. This study used PHF proteins as possible targets to boost cGAS activation in tumor cells. When *PHF2* and *PHF8* were transiently knocked down, this increased the downstream cytokine released upon cGAS activation, *IFNB1,* and increased cell death [17].

In addition to highlighting new therapeutic targets, this method of quantifying innate immune activation showed associations with tumor immunogenicity. When stratified based on activation scores, colorectal tumors with greater activation harbored significantly more mutations. Tumor mutation burden is largely thought to increase the display of neoantigens on MHC class I molecules alarming the adaptive immune system [36]. These results suggest that there is also a concurrent increase in the innate immune system activation with greater mutation burden. How and if pattern recognition receptors can help to identify self-antigens is an interesting question that should be further explored.

Finally, this custom ontology interrogated the effect of intratumor microbes on the immunologic tumor microenvironment. Intratumor microbes have been of particular interest in recent years. In pancreatic cancer, it was estimated that there was an average of 1 microbe for every 146 cancer cells [45]. In some cancers, it is hypothesized that intratumor microbes help stimulate the immune system and aid in immune cell infiltration. A great example of this phenomenon is the effective use of the attenuated strain of *Mycobacterium,* bacilli Calmette-Guerin (BCG), as an adjuvant in bladder cancer treatment [3]. Conversely, intratumor microbes may aid in resistance to chemotherapy. In pancreatic cancer, intratumor Gammaproteobacteria were found to metabolize the chemotherapeutic agent, gemcitabine [45]. Due to the unique effects of different tumor microbes on treatment outcomes, interrogating how the tumor microbiome interacts with the innate immune system is of great interest. Tumor localization in the body affected the diversity of the tumor microbiome and the relationships with the innate immune system. Colon adenocarcinoma had some weak associations with intratumor microbes. On the other hand, lung squamous cell carcinoma not only had a distinct microbiome population but also an inverse relationship with its intratumor microbial abundance. Overall, associations with intratumor microbes remain weak, although statistically significant. This may suggest that there are additional triggers of PRR activation outside of intratumor microbes. One likely candidate may be neoantigens as mentioned previously. These results will hopefully propel further research into therapy adjuvants to trigger a greater anti-tumor immune response.

This study is limited by its exploratory nature. Its purpose is to highlight a new way of quantifying and interrogating the innate immune system. It offers a broad introduction to the various utilities of this score system and a foundation of future work into therapeutic targets and therapeutic adjuvants. In conclusion, evaluating the innate immune system using the scoring system developed in this study demonstrated the immune system's complex relationship with patient outcomes, intratumor microbes, and demonstrated potential new therapeutic targets to further boost anti-tumor immunity.

Abbreviations

DAMP	Damage Associated Molecular Pattern
PAMP	Pattern Associated Molecular Pattern
PRR	Pattern Recognition Receptor
GSEA	Gene Set Enrichment Analysis
ssGSEA	Single Sample Gene Set Enrichment Analysis
TCGA	The Cancer Genome Atlas
TLR	Toll Like Receptor
CLR	C-type Lectin Receptor
RIG-I	Retinoic acid Inducible Gene I
NOD	Nucleotide Binding Oligomerization Domain
cGAS	Cyclic GMP-AMP Synthase
PHF	Plant Homeodomain Finger
DMEM	Dulbecco's Modified Eagle Medium
FBS	Fetal Bovine Serum
BMDM	Bone Marrow Derived Macrophages
LPS	Lipopolysaccharide
RSV	Respiratory Syncytial Virus
LGG	Brain Lower Grade Glioma
HNSC	Head and Neck Squamous Cell Carcinoma
THYM	Thymoma
CESC	Cervical Squamous Cell Carcinoma
COAD	Colon Adenocarcinoma
DDR	DNA Damage Response
LDH	Lactate Dehydrogenase
BCG	Bacilli Calmette-Guerin

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-024-12944-w.

Supplementary Material 1.

Acknowledgements

We would like to thank the lab of Dr. James Chen for providing necessary cell lines for these experiments. We would also like to thank the consortium

of contributors of The Cancer Genome Atlas for providing such a valuable tool for cancer research. Finally, we would like to thank the UT Southwestern scientific community for their support.

Authors' contributions
B.L. wrote manuscript text, prepared Figs. 1, 2, 3, 4, and 5, directed experiments. G.Q. wrote manuscript text, prepared Figs. 1, 2, 3, 4, and 5, performed experiments, ran analyses. G.M. prepared Fig. 3, performed experiments. All authors reviewed the manuscript.

Funding
This work is funded by NCI grants 1R01CA245318 (B.L.), and 1R01CA258524 (B.L.).

Availability of data and materials
Also stated in Methods Section in Manuscript Body:
Previously published RNA-seq experiments used in this study can be found on Gene Expression Omnibus under the following accessions: GSE174141, GSE99298, GSE146009, and GSE17538. TCGA level-2 sequencing data were obtained via controlled access. TCGA level-3 and patient metadata were directly downloaded from the NCI GDC website (https://portal.gdc.cancer.gov). Copy numbers of genes in TCGA tumors were obtained using the "TCGAbiolinks" and "RTCGAToolbox" R packages. ssGSEA algorithm was obtained from its original authors' submission to GitHub (https://github.com/broadinstitute/ssGSEA2.0). Finally, intratumor microbe abundance values were obtained from Poore et al. (10) via ftp://ftp.microbio.me/pub/cancer_microbiome_analysis/.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 25 January 2024 Accepted: 12 September 2024
Published online: 18 September 2024

References

1. Vincent RL, Gurbatri CR, Li F, Vardoshvili A, Coker C, Im J, et al. Probiotic-guided CAR-T cells for solid tumor targeting. Science. 2023;382(6667):211–8.
2. Chen YE, Bousbaine D, Veinbachs A, Atabakhsh K, Dimas A, Yu VK, et al. Engineered skin bacteria induce antitumor T cell responses against melanoma. Science. 2023;380(6641):203–10.
3. Herr HW, Morales A. History of bacillus Calmette-Guerin and bladder cancer: an immunotherapy success story. J Urol. 2008;179(1):53–6.
4. Mitchell D, Chintala S, Dey M. Plasmacytoid dendritic cell in immunity and cancer. J Neuroimmunol. 2018;322:63–73.
5. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–50.
6. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
7. Kawai T, Akira S. The roles of TLRs, RLRs and NLRs in pathogen recognition. Int Immunol. 2009;21(4):317–37.
8. da Silva CJ, Miranda Y, Austin-Brown N, Hsu J, Mathison J, Xiang R, et al. Nod1-dependent control of tumor growth. Proc Natl Acad Sci U S A. 2006;103(6):1840–5.
9. Kasumba DM, Grandvaux N. Therapeutic Targeting of RIG-I and MDA5 Might Not Lead to the Same Rome. Trends Pharmacol Sci. 2019;40(2):116–27.
10. Hopfner KP, Hornung V. Molecular mechanisms and cellular functions of cGAS-STING signalling. Nat Rev Mol Cell Biol. 2020;21(9):501–21.
11. Wang Q, Armenia J, Zhang C, Penson AV, Reznik E, Zhang L, et al. Unifying cancer and normal RNA sequencing data from different sources. Sci Data. 2018;5: 180061.
12. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature. 2020;579(7800):567–74.
13. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
14. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
15. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447–52.
16. Kingeter LM, Lin X. C-type lectin receptor-induced NF-kappaB activation in innate immune and inflammatory responses. Cell Mol Immunol. 2012;9(2):105–12.
17. Chen Q, Sun L, Chen ZJ. Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing. Nat Immunol. 2016;17(10):1142–9.
18. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462(7269):108–12.
19. Li T, Huang T, Du M, Chen X, Du F, Ren J, et al. Phosphorylation and chromatin tethering prevent cGAS activation during mitosis. Science. 2021;371(6535):eabc5386.
20. Stothers CL, Burelbach KR, Owen AM, Patil NK, McBride MA, Bohannon JK, et al. β-glucan induces distinct and protective innate immune memory in differentiated macrophages. J Immunol. 2021;207(11):2785–98.
21. Pei J, Beri NR, Zou AJ, Hubel P, Dorando HK, Bergant V, et al. Nuclear-localized human respiratory syncytial virus NS1 protein modulates host gene transcription. Cell Rep. 2021;37(2): 109803.
22. Lans H, Hoeijmakers JHJ, Vermeulen W, Marteijn JA. The DNA damage response to transcription stress. Nat Rev Mol Cell Biol. 2019;20(12):766–84.
23. Duan T, Du Y, Xing C, Wang HY, Wang RF. Toll-like receptor signaling and its role in cell-mediated immunity. Front Immunol. 2022;13: 812774.
24. Alonso-de Vega I, Paz-Cabrera MC, Rother MB, Wiegant WW, Checa-Rodriguez C, Hernandez-Fernaud JR, et al. PHF2 regulates homology-directed DNA repair by controlling the resection of DNA double strand breaks. Nucleic Acids Res. 2020;48(9):4915–27.
25. Pappa S, Padilla N, Iacobucci S, Vicioso M, Alvarez de la Campa E, Navarro C, et al. PHF2 histone demethylase prevents DNA damage and genome instability by controlling cell cycle progression of neural progenitors. Proc Natl Acad Sci U S A. 2019;116(39):19464–73.
26. Ablain J, de The H. Retinoic acid signaling in cancer: The parable of acute promyelocytic leukemia. Int J Cancer. 2014;135(10):2262–72.
27. Wang YH, Israelsen WJ, Lee D, Yu VWC, Jeanson NT, Clish CB, et al. Cell-state-specific metabolic dependency in hematopoiesis and leukemogenesis. Cell. 2014;158(6):1309–23.
28. Ma S, Cao C, Che S, Wang Y, Su D, Liu S, et al. PHF8-promoted TOPBP1 demethylation drives ATR activation and preserves genome stability. Sci Adv. 2021;7(19):eabf7684.
29. Sinha S, Singh RK, Alam N, Roy A, Roychoudhury S, Panda CK. Alterations in candidate genes PHF2, FANCC, PTCH1 and XPA at chromosomal 9q22.3 region: pathological significance in early- and late-onset breast carcinoma. Mol Cancer. 2008;7: 84.
30. Ghosh A, Ghosh S, Maiti GP, Mukherjee S, Mukherjee N, Chakraborty J, et al. Association of FANCC and PTCH1 with the development of early dysplastic lesions of the head and neck. Ann Surg Oncol. 2012;19(Suppl 3):S528–38.
31. Baba A, Ohtake F, Okuno Y, Yokota K, Okada M, Imai Y, et al. PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B. Nat Cell Biol. 2011;13(6):668–75.
32. Lee KH, Park JW, Sung HS, Choi YJ, Kim WH, Lee HS, et al. PHF2 histone demethylase acts as a tumor suppressor in association with p53 in cancer. Oncogene. 2015;34(22):2897–909.
33. Ge Z, Gu Y, Han Q, Sloane J, Ge Q, Gao G, et al. Plant homeodomain finger protein 2 as a novel IKAROS target in acute lymphoblastic leukemia. Epigenomics. 2018;10(1):59–69.

34. Garner H, de Visser KE. Immune crosstalk in cancer progression and metastatic spread: a complex conversation. Nat Rev Immunol. 2020;20(8):483–97.
35. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res. 2020;48(W1):W509–14.
36. Lakatos E, Williams MJ, Schenck RO, Cross WCH, Househam J, Zapata L, et al. Evolutionary dynamics of neoantigens in growing tumors. Nat Genet. 2020;52(10):1057–66.
37. Brown GJ, Canete PF, Wang H, Medhavy A, Bones J, Roco JA, et al. TLR7 gain-of-function genetic variation causes human lupus. Nature. 2022;605(7909):349–56.
38. Decout A, Katz JD, Venkatraman S, Ablasser A. The cGAS-STING pathway as a therapeutic target in inflammatory diseases. Nat Rev Immunol. 2021;21(9):548–69.
39. Li C, Zhang D, Pang X, Pu H, Lei M, Fan B, et al. Trajectories of perioperative serum tumor markers and colorectal cancer outcomes: A retrospective, multicenter longitudinal cohort study. EBioMedicine. 2021;74:103706.
40. Chen J, Apizi A, Wang L, Wu G, Zhu Z, Yao H, et al. TCGA database analysis of the tumor mutation burden and its clinical significance in colon cancer. J Gastrointest Oncol. 2021;12(5):2244–59.
41. Xavier JB, Young VB, Skufca J, Ginty F, Testerman T, Pearson AT, et al. The Cancer Microbiome: Distinguishing Direct and Indirect Effects Requires a Systemic View. Trends Cancer. 2020;6(3):192–204.
42. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016;8(1):51.
43. Yum S, Li M, Frankel AE, Chen ZJ. Roles of the cGAS-STING pathway in cancer immunosurveillance and immunotherapy. Ann Rev Cancer Biol. 2019;3:323–44.
44. Xia T, Konno H, Ahn J, Barber GN. Deregulation of STING signaling in colorectal carcinoma constrains DNA damage responses and correlates with tumorigenesis. Cell Rep. 2016;14(2):282–97.
45. Geller LT, Barzily-Rokni M, Danino T, Jonas OH, Shental N, Nejman D, et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. Science. 2017;357(6356):1156–60.

Publisher's Note