



OPEN ACCESS

EDITED BY

Elisa Frullanti,
University of Siena, Italy

REVIEWED BY

Xiaotong Yao,
Foundation Medicine Inc., United States
Petros Christopoulos,
Heidelberg University Hospital, Germany

*CORRESPONDENCE

Alfredo Pulvirenti,
✉ alfredo.pulvirenti@unict.it

†These authors have contributed equally to
this work

RECEIVED 04 September 2023

ACCEPTED 11 March 2024

PUBLISHED 05 April 2024

CITATION

Privitera GF, Alaimo S, Caruso A, Ferro A, Forte S
and Pulvirenti A (2024), TMBcalc: a
computational pipeline for identifying pan-
cancer Tumor Mutational Burden
gene signatures.
Front. Genet. 15:1285305.
doi: 10.3389/fgene.2024.1285305

COPYRIGHT

© 2024 Privitera, Alaimo, Caruso, Ferro, Forte
and Pulvirenti. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

TMBcalc: a computational pipeline for identifying pan-cancer Tumor Mutational Burden gene signatures

Grete Francesca Privitera¹, Salvatore Alaimo¹, Anna Caruso²,
Alfredo Ferro¹, Stefano Forte^{3†} and Alfredo Pulvirenti^{1*†}

¹Department of Clinical and Experimental Medicine, Bioinformatics Unit, University of Catania, Catania, Italy, ²Department of Physics and Astronomy, University of Catania, Catania, Italy, ³Istituto Oncologico del Mediterraneo (IOM) Ricerca, Viagrande, Italy

Background: In the precision medicine era, identifying predictive factors to select patients most likely to benefit from treatment with immunological agents is a crucial and open challenge in oncology.

Methods: This paper presents a pan-cancer analysis of Tumor Mutational Burden (TMB). We developed a novel computational pipeline, TMBcalc, to calculate the TMB. Our methodology can identify small and reliable gene signatures to estimate TMB from custom targeted-sequencing panels. For this purpose, our pipeline has been trained on top of 17 cancer types data obtained from TCGA.

Results: Our results show that TMB, computed through the identified signature, strongly correlates with TMB obtained from whole-exome sequencing (WES).

Conclusion: We have rigorously analyzed the effectiveness of our methodology on top of several independent datasets. In particular we conducted a comprehensive testing on: (i) 126 samples sourced from the TCGA database; few independent whole-exome sequencing (WES) datasets linked to colon, breast, and liver cancers, all acquired from the EGA and the ICGC Data Portal. This rigorous evaluation clearly highlights the robustness and practicality of our approach, positioning it as a promising avenue for driving substantial progress within the realm of clinical practice.

KEYWORDS

personalized medicine, Tumor Mutational Burden, DNA-seq, analysis pipeline, pan-cancer

1 Introduction

Cancer Immunotherapy aims to activate or boost patients' adaptive or innate immune systems to attack tumor cells. Indeed, tumor cells carry out several mechanisms to evade their recognition and elimination by T cells. Firstly, they express the programmed cell death ligand 1 (PD-L1), which binds the programmed death receptor 1 (PD1) on T cells, making them inactive. Secondly, when cytotoxic T lymphocyte antigen-4 (CTLA-4), a co-inhibitory molecule that regulates the T cell activation, interacts with its ligands (CD80 and CD86), it inhibits T cell activity promoting immunological escape. Lastly, tumor cells can lose the expression of both Major Histocompatibility Complex (MHC) classes, thus becoming invisible to the host immune system. In the last few years, it has been observed that T cell

therapies and monoclonal antibodies blocking the CTLA-4 and PD1 immune checkpoints can induce durable responses across tumors. In particular, the PD-L1 ligand has been studied since it is commonly upregulated on several human solid tumors, including Melanoma, Lung, and Ovarian cancers, leading to peripheral T cell exhaustion and inhibition of apoptosis of malignant cells (Harview et al., 2014). Immune checkpoint inhibitors enhance anti-tumor T-cell activity by inhibiting immune checkpoint molecules. While the immune system plays a pivotal role in neoplastic controls, its tolerance to physiological elements must be granted by specific signals distinguishing cancer cells from normal cells. Indeed, cancer cells are recognized thanks to particular antigens such as tumor-associated antigens (TAA), over-expressed in tumor cells, and neoantigens, also called tumor-specific antigens (TSA). Usually, these antigens are not expressed by normal cells, but they are produced in tumor cells due to mutations in coding sequences. For this reason, these are ideal targets for T cell-based cancer immunotherapy. Point-mutated genes encode the majority of these neoantigens. Identifying reliable predictive factors allowing the selection of patients most likely to benefit from treatment with immunological agents is still an open challenge in oncology (Hegde and Chen, 2020; Kossai et al., 2021). Unfortunately, some patients do not respond to immunotherapy and, in addition, immunotherapy could lead to unpleasant side effects such as skin rash, colitis, hepatotoxicity, pneumonitis, endocrinopathies, and autoimmune diseases (Golshani and Zhang, 2020). The biomarkers used in cancer immunotherapy include PD-L1 and PD-L2 expression levels, microsatellite instability (MSI) status, neoantigen presence, and Tumor Mutational Load or Tumor Mutational Burden (TMB) (Shu et al., 2020).

In the past years, the most used biomarker to decide immunotherapy was immunohistochemistry's evaluation of PD-1/PD-L1 protein expression. However, this biomarker is challenging to interpret. Hence, TMB has been considerably studied as a biomarker in recent years. TMB is defined by counting all the mutations found in a tumor sample divided by the total length of the sequenced sample in DNA Megabase (Mb). Many mutations in a tumor harbor more neoantigens, making them targets of activated immune cells. In addition, the mutation number and DNA structural alterations lead to the production of foreign proteins recognizable by the immune system. Therefore, establishing the mutational load in cancer cells would allow identifying those patients who can benefit significantly from this type of therapy compared to conventional chemotherapy treatment. However, there is no standardized specific value to decide which is a high TMB (H-TMB) or a low TMB (L-TMB) level (Lawrence et al., 2013; Stenzinger et al., 2019). Few commercial and free pipelines have been implemented to analyze the TMB. In 2020 Yao et al. developed ecTMB to predict TMB using a statistical model to correct the panel design biases (Yao et al., 2020). Commercial solutions for the TMB calculation include Illumina (Illumina, 2021), ThermoFisher (ThermoFisher, 2021), Qiagen (Qiagen, 2021), and Q2 Solutions (Q2solution, 2021). These solutions have been created specifically to calculate TMB in the clinical setting. However, their actual workflow is not known. The commercial tools (approved by FDA) recommended to establish the TMB are The FoundationOne CDx assay and the MSK-IMPACT (Memorial Sloan Kettering Cancer Center) panel, which have been authorized by the 510k

pathway (Food U and Administration D, 2017b; Food U and Administration D, 2017a; Klempner et al., 2020).

Concerning cancer types, Melanoma has the highest mutation rate and the highest number of neoantigens (Snyder et al., 2014). Therefore, immunotherapy represents a common choice for its treatment. However, high-impact cancer disease on the lungs (Rizvi et al., 2015), prostate, colorectal (Stein et al., 2018), and breast could also benefit from this therapy. Monoclonal antibodies have shown promising efficacy against programmed death 1 (PD-1), such as Pembrolizumab for gastrointestinal (GI)-related cancer (Doi et al., 2018), or Nivolumab a PD-1 inhibitor used for patients with hepatocellular carcinoma (El-Khoueiry et al., 2017). Some monoclonal antibodies have already been used in angiogenesis for CRC, and there are two approved immune checkpoint inhibitors targeting PD-1 (pembrolizumab and nivolumab) used after progressed chemotherapy. In Cercsek et al. (2022), authors show that 12 patients with advanced rectal cancer treated with single-agent PD-1 blockade had a complete clinical response, with no evidence of tumor on magnetic resonance imaging.

The gold standard to measure the TMB is the Whole Exome Sequencing (WES) analysis with tumor and normal samples. WES analysis, however, has a high cost and requires extensive data management. Indeed, two samples are needed to discard germline mutations. Unfortunately, the availability of these matched samples in clinical practice varies across organizations. Germline variants in tumor-only sequencing can be filtered out using available databases. Still, this procedure needs a high level of standardization for each type of tumor and each population (Meléndez et al., 2018). Therefore, to go beyond such a limitation, targeted panels are under investigation (Campesato et al., 2015; Johnson et al., 2016) to speed up the analysis and keep high precision and sensitivity. Some authors recommend using targeted gene panels, ideally with 1 Megabase as the lower bound limit to yield reliable TMB estimation (Allgäuer et al., 2018; Buchhalter et al., 2018; Endris et al., 2018).

Furthermore, differential expression genes (DEGs) analysis has clarified genes' role in cancer patients between high and low TMB patients. Comparing tumor and normal colon samples, Gao et al. (2018) found that DEGs were mainly involved in protein transport, apoptotic, and neurotrophin signaling pathways. Wang et al. (2020) screened the TCGA-BRCA dataset splitting patients in TMB high and TMB low and analyzed them with the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) databases. They found that DEGs were primarily enriched in epidermis development, extracellular matrix, and receptor-ligand activity among Biological processes, Cellular Components, and Molecular Functions, respectively. Zhang et al. (2020) found that differential genes were involved in catalytic activity in bladder urothelial carcinoma, acting on DNA, single-stranded DNA-dependent ATPase activity. Moreover, TMB enrichment of related signatures correlated with multiple cancer-related crosstalks, including cell cycle, DNA replication, cellular senescence, and p53 signaling pathway.

This study presents a Docker-based pipeline designed for pan-cancer applications, specifically aimed at computing the Tumor Mutational Burden (TMB) in DNA-Seq samples. Leveraging data from The Cancer Genome Atlas (TCGA), our research delves into the prospect of stratifying patients' TMB using different approaches with the aim of identifying a small pan-cancer mutational signature

allowing to effectively predict the actual TMB. Through our computational model, we identified a small signature composed of 389 genes showing a strong differential mutation rate between high and low TMB patients. Our findings reveal that such a signature effectively stratifies patients without necessitating Whole Exome Sequencing (WES), thus establishing its suitability for clinical contexts. We have rigorously validated its reliability, employing multiple independent datasets from TCGA and the European Genome Archive (EGA).

To further analyze the implications of this gene signature, we conducted a functional analysis for both the High TMB (H-TMB) and Low TMB (L-TMB) groups. Remarkably, our results highlight either the perturbation of several immunity-related pathways and the enrichment of many immune-involved Transcription Factors, shedding light on the potential biological significance of our findings.

2 Materials and methods

The forthcoming sections present an in-depth description of our methodology, delineating the comprehensive pipeline that integrates both upstream and downstream analyses. Our entire approach was constructed upon the foundation of the TCGA Harmonized Cancer Dataset, and its efficacy was rigorously examined across several independent datasets sourced from the EGA.

2.1 Datasets

We downloaded raw Colon adenocarcinoma samples used as our study's primary tumor. BAM cancer and normal tissue biopsy files were analyzed for each sample, extracting the somatic mutations. All analyses have been repeated for five thresholds. Other analyzed solid cancer types were: Ovarian serous cystadenocarcinoma (OV, $n = 441$); Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC, $n = 305$); Thyroid carcinoma (THCA, $n = 496$); Bladder Urothelial Carcinoma (BLCA, $n = 412$); Uterine Corpus Endometrial Carcinoma and Uterine Carcinosarcoma (UCEC and UCS, $n = 628$); Esophageal carcinoma (ESCA, $n = 181$); Kidney renal papillary cell carcinoma (KIRP, $n = 288$); Kidney renal clear cell carcinoma (KIRC, $n = 339$); Liver hepatocellular carcinoma (LIHC, $n = 415$); Stomach adenocarcinoma (STAD, $n = 450$); Pancreatic adenocarcinoma (PAAD, $n = 183$); Prostate adenocarcinoma (PRAD, $n = 497$); Adrenocortical carcinoma (ACC, $n = 240$); Skin Cutaneous Melanoma (SKCM, $n = 466$); Lung Squamous Cell Carcinoma (LUSC, $n = 494$); Lung Adenocarcinoma (LUAD, $n = 512$). For these cancers, we directly downloaded the VCF files supplied by TCGA.

To test our signature we employed six independent datasets. These comprise 126 Colon Cancer WES obtained from TCGA; 72 colon WES samples from Genentech (Seshagiri et al., 2012); 42 samples taken from Colonomics (Díez-Villanueva et al., 2022); 101 Breast Cancer WES samples obtained from dbGaP (NCBI, 2021); 291 colon samples from COCA-CN study taken from ICGC (Zhang et al., 2019) and 338 samples from LINC-JP taken from ICGC. The first three datasets were analyzed with our pipeline starting from the BAM files. While, the somatic mutation files of the latter two, from ICGC, have been directly annotated using Annovar.

2.2 The upstream pipeline

The TMBCalc pipeline (see Figure 1) has been implemented in Bash and R. It computes the TMB, yielding as output (i) the TMB value and (ii) a list of the mutations implicated in its calculation. The source code and user manual are available at <https://github.com/knowmics-lab/TMBCalc>. In addition, the containerized software can be downloaded with Docker through the DockerHub Container Image Library tmbcalc.

Our pipeline comprises five modules:

1. Alignment: WES BAM files are converted in FASTQ with the “bam2fq” command [Samtools (Li et al., 2009)]. Next, FASTQ files are aligned with Bowtie 2 (Langmead and Salzberg, 2012) using the hg38 Genome assembly gathered from NCBI.
2. BAM Processing: First, each BAM file is modified by adding a read group to each sequencing read using the “AddOrReplaceReadGroups” command in Picard tool (Picard toolkit, 2018). Then, all reads are sorted by genomic coordinate [“SortSam” command (Picard toolkit, 2018)] and contig ordering [“ReorderSam” command (Picard toolkit, 2018)]. Finally, duplicates are identified and removed [“MarkDuplicates” command (Picard toolkit, 2018)].
3. Variant calling: Gatk (v4.2.5.0) Mutect2 (McKenna et al., 2010) is employed to perform variant calling. In this step, all mutations labeled as “PASS” were retained. Furthermore, we run a second variant calling using the “somatic” and “somaticFilter” commands of VarScan (v2.4.4) (Koboldt et al., 2012) (min-var-freq = 10). This dual-phase procedure allows us to get a more precise variant calling by intersecting the two VCF output files.
4. Annotation: In this step, we remove all variants known to be unusable for computing TMB (Chalmers et al., 2017). Therefore, first, we annotate all inferred mutations with Annovar (Wang et al., 2010; Yang and Wang, 2015). The following databases are used: 1000genome (2015_08) (Consortium et al., 2015), snp146 (Sherry et al., 2001), cosmic88 (Tate et al., 2018), NHLBI Exome Sequencing Project (ESP6500) (Nhlbi go exome sequencing project, 2021). Then, variants that present annotations in these databases are filtered out, yielding a file containing only mutations valid for the TMB computation.
5. TMB calculation: This step produces two outputs: (i) the TMB value and (ii) the list of contributing variants. Although the TMB is commonly defined as “the number of the counted non-synonymous mutations that alter the amino acid sequence of a protein,” we also decided to include synonymous ones, following (Chalmers et al., 2017; Klempner et al., 2020), since this inclusion might improve sensitivity. Furthermore, such mutations are indicative of a mutational process. Therefore, the TMB has been computed as

$$TMB = \frac{\#mutations}{GenomeSize} \quad (1)$$

where $\#mutations$ is the number of mutations identified in the sample, and the genome size is 38 MB (Chalmers et al., 2017).

Given the absence of well-defined TMB thresholds, our analyses incorporated numerous values documented in existing literature

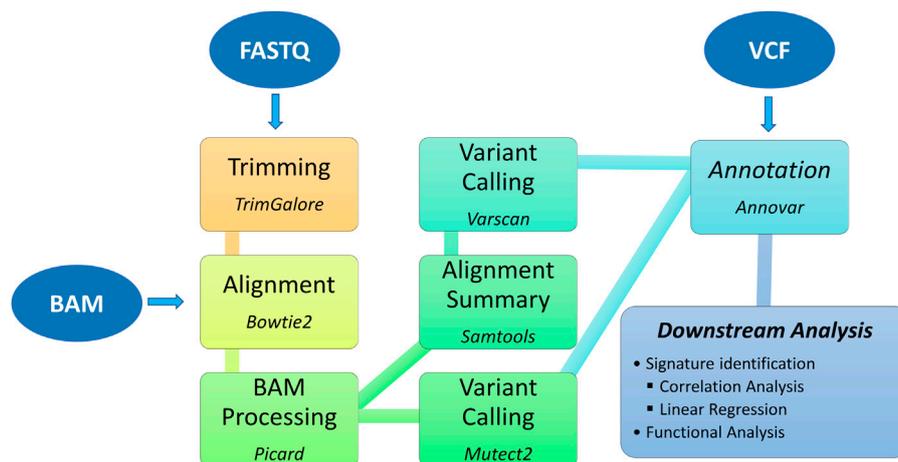


FIGURE 1
Bioinformatics pipeline: upstream and downstream analysis.

pertaining to colon cancer. The selected thresholds were as follows: 5 mutations per megabase (5mut/Mb), 10 mutations per megabase (10mut/Mb), 20 mutations per megabase (20mut/Mb) as referenced in (Chalmers et al., 2017) mutations per megabase (25.29mut/Mb) calculated using the formula $TMB + 1.25 \times IQR(TMB)$ as outlined in (Fernandez et al., 2019), where IQR is the Interquartile range, and 34.66 mutations per megabase (34.66mut/Mb) as cited in (Wu et al., 2019a).

2.3 The Downstream pipeline

To create a gene signature panel suitable for the TMB analysis, our study started from the “AmpliSeq for Illumina Comprehensive Cancer Panel” - a pre-built Illumina panel of length 1.7 Mb, that covers all exons of 409 cancer-associated genes. To derive the size of each gene we make use of the R package EDASeq (Risso et al., 2011) with the “getGeneLengthAndGCCContent” function. We measured its discriminant power in identifying patients with high and low TMB.

Next, we investigated the predictive efficacy of custom panels by conducting two parallel analyses. Initially, we curated sets of randomly selected genes, varying in size (50, 100, 200, and 300 genes), featuring mutations from WES data obtained from colon cancer patients. This process was iterated 1,000 times for each panel size to ensure robustness. Subsequently, we replicated the analysis using the most commonly mutated genes observed in colon cancer WES data, thus constructing panels comprising 50, 100, 200, 300, 409, and 500 genes. The primary objective was to juxtapose panels derived from randomly selected genes against those enriched with frequently mutated ones, thereby facilitating a comparative assessment of their predictive capabilities.

Subsequently, we constructed a concise signature by carefully selecting the top 10 most frequently mutated genes across 17 distinct cancer types. This curation resulted in a signature comprising 44 genes (refer to Supplementary Table S1), encompassing a genomic length of 1.05 Mb.

Finally, we devised a panel comprising genes exhibiting a noteworthy discrepancy in mutational signatures between high and low TMB (Tumor

Mutational Burden) patients. To achieve this objective, we employed the following methodology. Initially, for each cancer type, we calculated the mutation rate per kilobase (kb) of each gene g_i in each patient P_j , denoted as $MutationRate(g_i, P_j)$, using the formula:

$$MutationRate(g_i, P_j) = \frac{M(g_i, P_j)}{length(g_i)} \times 1000 \quad (2)$$

where $M(g_i, P_j)$ is the number of mutations for gene g_i in patient P_j and $length(g_i)$ is the length of gene g_i in bases.

Subsequently, we compared the gene mutation rates between H-TMB and L-TMB patients against the expected rates. A significant deviation from the expected value suggested that such differences could not be solely attributed to gene length. To ascertain the statistical significance of such differences, we conducted Mann-Whitney tests and corrected the p -values for multiple hypotheses using the Benjamini-Hochberg False Discovery Rate (FDR) method. Then, for each statistically significant gene g_i , we computed its rank as the difference between the expected gene mutation rates in H-TMB and L-TMB:

$$rank(g_i) = E_{H-TMB}[MutationRate(g_i, P_j)] - E_{L-TMB}[MutationRate(g_i, P_j)] \quad (3)$$

Finally, we aggregated the rankings of all genes (one from each cancer type) and selected the top k genes to approximate a total size of 1 Mb (1.08 Mb). This methodology enabled us to construct a signature composed of 389 genes, facilitating robust analysis and inference.

3 Results

For each panel, we conducted a comprehensive data analysis involving the computation of correlation and logistic regression. In particular, we started annotating the mutated genes in all samples using the “RefSeq gene” Annovar database (Pruitt et al., 2012); after that, we correlated, using Pearson, the TMB calculated using WES with the one calculated using our custom gene signatures and the “AmpliSeq for Illumina Comprehensive Cancer Panel”; finally, we

TABLE 1 For each random group of panels, we have calculated the average values of the following quantities: panel length, correlation, precision, recall, specificity and negative predicted value. For other gene panels, we have calculated their correlation, precision, recall, sensitivity and negative predicted value. The threshold used was 20mut/Mb. The table also reports the confusion matrix of each experiment.

Number of genes and panel length		Pearson correlation	H-TMB correlation	L-TMB correlation	PPV	TPR	TNR	NPV	TP	FP	TN	FN
50	0.27 R	0.94	0.90	0.31	0.95	0.79	0.96	0.95	9.4	0.5	46.5	2.5
	1.08 H	0.97	0.95	0.81	1	1	1	1	12	0	47	0
100	0.55 R	0.97	0.94	0.42	0.98	0.88	0.99	0.97	10.63	0.23	46.77	1.37
	1.80 H	0.98	0.97	0.86	1	1	1	1	12	0	47	0
200	1.09 R	0.98	0.97	0.54	0.99	0.94	1	0.98	11.27	0.07	46.93	0.73
	3.05 H	0.98	0.97	0.88	1	1	1	1	12	0	47	0
300	1.64 R	0.99	0.98	0.62	1	0.95	1	0.99	11.44	0.02	46.98	0.56
	4.28 H	0.99	0.98	0.90	0.91	1	1	1	12	0	47	0
409	5.49 H	0.99	0.98	0.92	1	1	1	1	12	0	47	0
	1.7 A	0.93	0.88	0.56	1	1	1	1	12	0	47	0
500	6.44 H	0.99	0.98	0.93	1	1	1	1	12	0	47	0
44	1.05 C	0.97	0.96	0.74	1	0.83	1	0.96	10	0	47	2
389	1.08 C	0.97	0.96	0.82	1	1	1	1	12	0	47	0

The bold values in Table 1 refer to the highest values.

classified the patients using the R package caret (caret, 2016) employing the 10-fold cross-validation procedure to measure gene signatures reliability and calculating measures such as sensitivity (TPR), specificity (TNR), Positive Predictive Value (PPV), Negative Predictive Value (NPV).

3.1 Mutational Burden on custom gene panels

We conducted a comprehensive analysis of custom gene panels to compute the Tumor Mutational Burden (TMB), comparing their performance with whole-exome sequencing (WES) data. As anticipated, increasing the number of genes in the signature naturally boosts the correlation. Notably, even a modest set of 50 randomly selected genes exhibits a strong correlation with WES-derived TMB (refer to the Correlation column of Table 1). Upon stratifying patients into High-TMB (H-TMB) and Low-TMB (L-TMB) groups using a threshold of 20 mutations per megabase (mut/Mb), the correlation notably diminishes within the L-TMB class (see Table 1). The 500-gene panel maintains a robust correlation within the L-TMB class, albeit its larger size of 6.44 Mb justifies this correlation. Remarkably, our results reveal that a 389-gene signature maintains a high correlation within the L-TMB class despite its smaller size of 1.08 Mb (refer to Supplementary Tables S1, S2 for additional threshold analyses).

Furthermore, we investigated the predictive capability of these panels. We categorized the actual TMB derived from WES into two

classes, H-TMB and L-TMB, based on a 20 mut/Mb threshold (see Supplementary Material for alternative thresholds). This categorization was then utilized for logistic regression analysis. Table 1 presents the classification outcomes, demonstrating that TMB calculated with the gene panels effectively aligns with WES data. Particularly noteworthy is the performance of the 389-gene signature, which outperforms others when normalized by signature length.

3.2 In-silico 389 genes signature validation

We extended our analysis to the entire TCGA cancer dataset, focusing on the 389 genes signature panel. As depicted in Table 2, the Pearson correlation across all tumors exceeds 85%. Notably, the correlation within the High-TMB (H-TMB) and Low-TMB (L-TMB) groups surpasses that of the 44-gene panel (see Supplementary Material). Additionally, employing logistic regression allowed us to establish precision and recall metrics. Across all tumors, we observed minimal occurrences of False Positives and False Negatives, with a consistently high Specificity exceeding 0.90 Figure 2.

We also validated the panel's performance across six independent datasets (see Table 3):

- 126 TCGA COAD samples that were not utilized in creating the signature. Their correlation stands at 0.95 for H-TMB and 0.85 for L-TMB.

TABLE 2 Correlation, Precision, and Recall of each TCGA tumor between TMB analyzed with the panel built with the 389 and WES TMB using threshold 20 mut/Mb.

	Pearson	Pearson H-TMB	Pearson L-TMB	PPV	TPR	TNR	NPV	TP	FP	TN	FN
UCEC-UCS	0.98	0.98	0.80	1.00	0.93	1	0.95	42	0	62	3
STAD	0.97	0.96	0.84	0.82	0.87	0.95	0.97	14	3	63	2
SKCM	0.98	0.98	0.86	0.91	0.86	0.90	0.95	44	4	37	7
BLCA	0.95	0.96	0.85	1.00	0.78	1	0.97	7	0	72	2
CESC	0.99	0.99	0.80	1	1	1	1	8	0	51	0
LUSC	0.93	0.91	0.82	0.78	0.5	0.96	0.86	11	3	73	11
LUAD	0.97	0.91	0.82	0.77	0.96	0.91	0.98	23	7	70	1
LIHC	0.93	0.86	0.83	0.5	0.50	0.99	0.99	1	1	76	1
THCA	0.99	1	0.86	0.67	1	0.99	1	2	1	95	0
OV	0.91	0.86	0.83	0.75	0.43	0.97	0.90	3	1	35	4
ESCA	0.90	0.92	0.82	-	0.75	1	0.88	0	0	32	4
PRAD	0.99	1	0.67	-	0	1	0.99	0	0	97	1
PAAD	0.99	1	0.76	1	1	1	1	1	0	34	0
KIRC	0.99	1	0.70	-	0	1	0.98	0	0	67	1
KIRP	0.87	0.97	0.81	-	0	1	0.98	0	0	56	1
ACC	0.98	1	0.88	1	1	1	1	1	0	43	0

The table also reports the True Negative Rate, Negative Predicted Value, and the confusion matrix for each cancer type.

- 72 samples from Genentech exhibit a high level of correlation between TMB calculated with WES and the signature gene panels. This correlation is 0.95 for the H-TMB patients and 0.87 for the L-TMB ones.
- 291 COCA-CN colon cancer samples from EGA demonstrate an excellent correlation between TMB calculated on the WES and the gene panels (0.99). This correlation persists within the H-TMB group, while the L-TMB group has a correlation coefficient equal to 0.8.
- 42 colon cancer samples from Colonomics show a Pearson correlation of 0.69. However, since all patients have Low TMB, we could not determine a correlation for the H-TMB group.
- dbGaP, consisting of 101 breast cancer samples, shows a Pearson correlation of 0.89. Upon splitting the samples into H-TMB and L-TMB with a threshold of 10 mut/Mb, the panel exhibits a correlation of 1 for H-TMB and 0.86 for L-TMB.
- 338 Liver-JP samples from ICGC show that although the H-TMB correlation coefficient is as high as in colon cancer, the L-TMB decreases to 0.73. Despite this decrease, the correlation level can still be considered moderate.

Therefore, we can conclude that the 389 gene panel is a promising solution to assess the TMB with reasonable accuracy for many cancer types. Furthermore, due to the relatively small size, it could speed up the TMB analysis, cutting the costs while reaching in principle the same results as WES (see [Supplementary Table S5](#) for the same analysis on the 44 genes panel).

3.3 Signature functional analysis

We conducted a comprehensive functional analysis using the 389-gene signature. Differential Expression Analysis (DEGs) was performed on TCGA RNA-seq data by using the raw counts (HTSeq-Counts) for the 389 genes of the panel. We collected the expression data with TCGAAbiolinks (Colaprico et al., 2015; Silva et al., 2016; Mounir et al., 2019) and performed the analysis with Limma (Ritchie et al., 2015). Genes were considered differentially expressed if $|\log_2FC| \leq \log_2(1.5)$ and adjusted p -value < 0.05 . Pathway perturbation and enrichment analysis were conducted using MITHrIL (Alaimo et al., 2016; Alaimo et al., 2017). Additionally, gProfiler2 (Kolberg et al., 2020) was employed for gene set enrichment analysis.

The comparative analysis between H-TMB and L-TMB patients across 17 tumors helped to elucidate differences in enriched pathways. Our investigation yielded significant outcomes emphasizing pathways intricately linked with immune and inflammatory responses. As depicted in [Figure 3](#), we show substantial immune pathways, including (i) the Cytosolic DNA-sensing pathway, which exhibited upregulation in H-TMB patients, indicating innate immunity activation and serving as a potential adjuvant for anticancer immune therapy (Yu and Liu, 2021); and (ii) the defensins pathway, also upregulated, known for its role as immunomodulators and attractors of immune cells (Contreras et al., 2020).

Furthermore, enrichment analysis conducted using gProfiler2 3 highlighted a significant presence of Transcription Factors (TFs) within our 389 gene signature. This enrichment suggests that these TFs likely play pivotal roles in regulating the expression of these genes. Understanding the impact of these TFs on the immune system holds crucial implications. For instance, NF- κ B, AP-1, IRF, and GATA3,

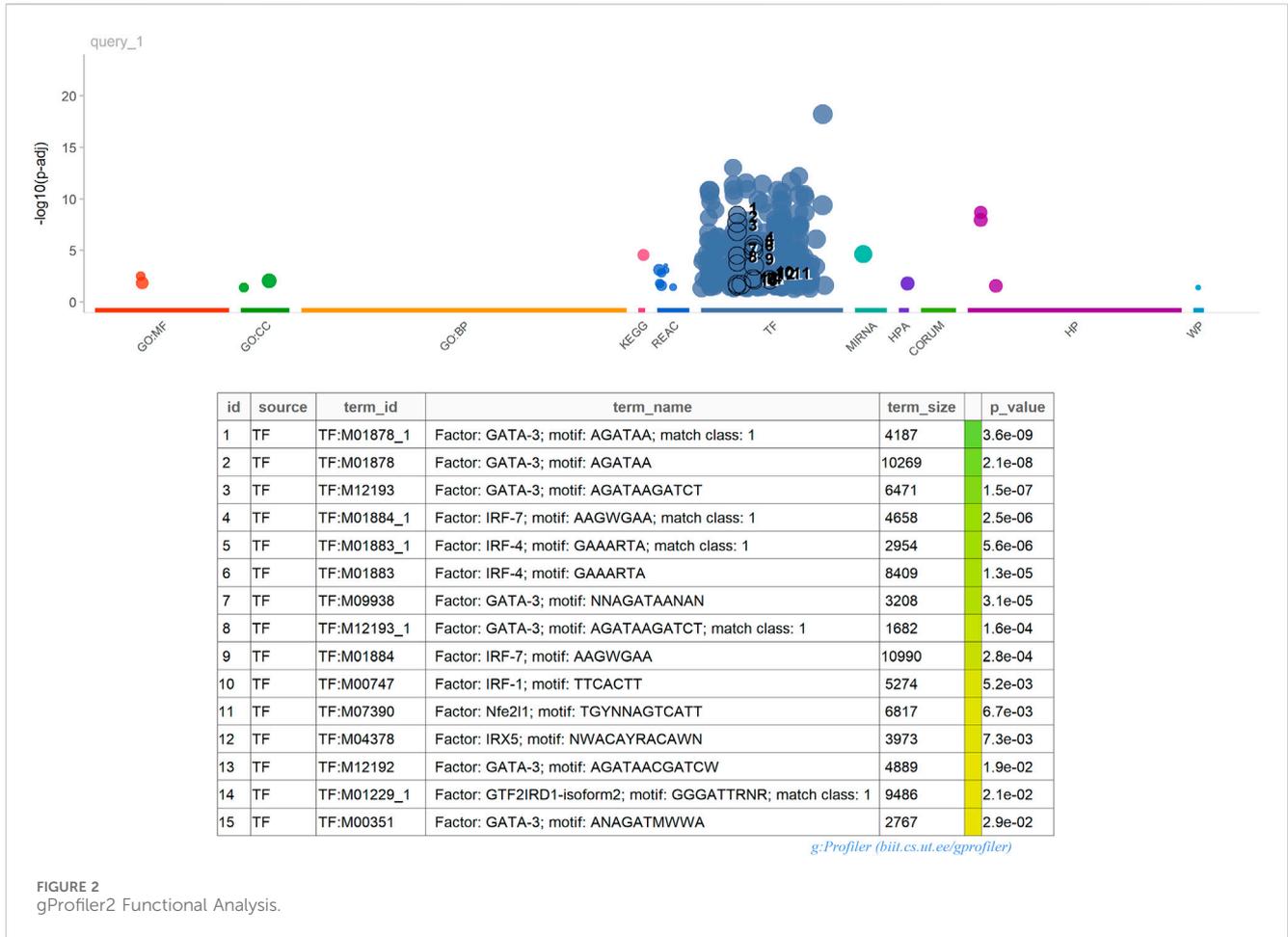


TABLE 3 Correlation, precision, and recall between TMB computed with WES and 389 genes panel using thresholds 20mut/Mb for each cancer except dbGap where the threshold was 10 mut/Mb.

	Pearson correlation	Pearson H-TMB	Pearson L-TMB	PPV	TPR	TNR	NPV	TP	FP	TN	FN
Genentech	0.94	0.95	0.87	1	1	1	1	2	0	11	0
Colonomics	0.69	-	0.69	-	-	-	-	-	-	8	-
Li-JP	0.99	0.99	0.73	0.87	0.87	0.68	0.68	42	6	13	6
TCGA 126 COAD	0.97	0.95	0.85	0.75	1	0.95	1	3	1	20	0
dbGap*	0.89	1	0.86	-	0	1	0.95	0	0	18	1
COCA	0.99	0.99	0.80	0.89	0.62	1	0.94	5	0	50	3

The table also reports the True Negative Rate, Negative Predicted Value, and the confusion matrix for each cancer type.

prominent regulators of immune responses, govern the expression of genes involved in inflammation, cytokine production, cell differentiation, and immune cell activation. By comprehending the regulatory mechanisms of these TFs on immune-related genes, valuable insights can be gained into immune responses against pathogens, antigens, and environmental stimuli. For example, GATA3 is critical in differentiating T helper 2 (Th2) cells, which produce cytokines crucial for allergic responses and immune regulation. Moreover, the interplay between Transcription Factors, including NF-κB, AP-1, IRF, and GATA3, often involves intricate interactions with each other and with other TFs to orchestrate coordinated immune responses.

4 Discussion

The immune response holds considerable promise in combating tumors. However, as tumors progress, they evolve mechanisms to evade detection by the immune system, often by producing a range of inhibitory molecules. Immunotherapy, leveraging immune checkpoint inhibitors, works to remove these brakes on the immune response. Tumors harboring a higher load of neoantigens and mutations tend to derive greater benefit from this approach, as they can provoke a more robust immune response. Identifying patients who are suitable candidates for immunotherapy relies heavily on Tumor

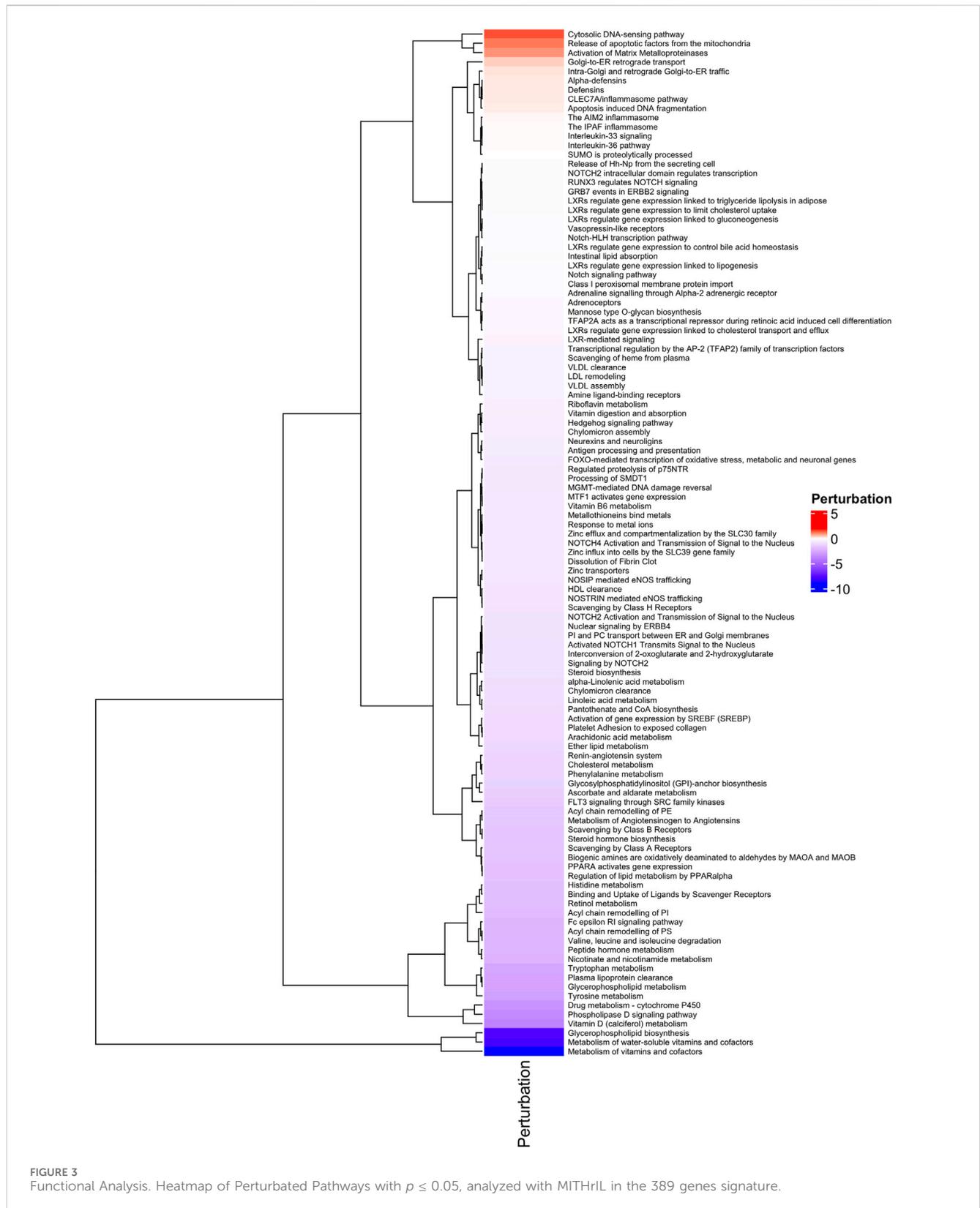


FIGURE 3
Functional Analysis. Heatmap of Perturbed Pathways with $p \leq 0.05$, analyzed with MITHrIL in the 389 genes signature.

Mutational Burden (TMB) assessment. The computation and analysis of TMB necessitate sophisticated Next-Generation Sequencing (NGS) techniques and advanced bioinformatics expertise. Efforts within the scientific community are directed towards streamlining this process for

clinical application. By obtaining a molecular portrait of the tumor, precision medicine approaches in immunotherapy can be realized.

Currently, Whole Exome Sequencing (WES) remains the preferred method for TMB computation, although there are

several panel-based alternatives available to expedite the analysis. Notably, two such panels have already received FDA approval. We have highlighted the utility of the AmpliSeq for Illumina Comprehensive Cancer Panel as a novel tool for TMB analysis. This panel, commonly utilized in laboratories for identifying variants across various cancer types, offers the advantage of incorporating additional markers for cost-effective TMB evaluation. This parallels the trend seen in some NGS panels, which now include assessment markers for Microsatellite Instability (MSI). While cancer-specific panels may offer improved accuracy for TMB estimation, a unified panel for multiple analyses presents a more cost-effective and flexible solution for laboratories dealing with diverse tumor types. Wu et al. (2019b) note that currently available Next-Generation Sequencing (NGS) panels can accurately assess Tumor Mutational Burden (TMB) only within specific cancer types. They emphasize the importance of relying on accuracy rather than correlation when evaluating panel performance. Notably, existing TMB commercial panels have been evaluated solely based on correlation, achieving coefficients of 0.74 for the FDA-approved FoundationOne CDx Panel and 0.76 for the MSK-IMPACT panel (Chalmers et al., 2017; Zehir et al., 2017; Wu et al., 2019b).

Our analysis emphasizes an alternative method for stratifying patients for immunotherapy by utilizing our 389 gene panel. Our panel, assessed across 17 tumor types, exhibits higher correlation coefficients (refer to Table 2 and Supplementary Table S4) and demonstrates favorable outcomes in regression analysis measured by Precision/Recall. The genesis of our 44 gene panel stemmed from identifying the top-10 most mutated genes in colon cancer. These genes, consistently highlighted in various studies as highly mutated (Kang et al., 2020a), have been subject to extensive functional investigations. For instance, Jia et al. (2019) established that mutations in *TTN* predict elevated TMB and increased response rates to Immune checkpoint blockade immunotherapy. Additionally, they observed favorable clinical outcomes correlated with *TTN* mutations. Furthermore, within the 389 gene signature panel, we identified genes such as (i) *ACE2*, recognized for its involvement in innate immunity; (ii) *TRIM51*, which exhibits increased expression in patients with high immune cell infiltration (Chen et al., 2022); (iii) *SERPINB4*, whose high mutation frequency correlates with improved survival post-immunotherapy in melanoma patients (Riaz et al., 2016); and (iv) *ADAM2*, whose expression enhances the cytotoxicity of CD8 T-cells (Dervovic et al., 2023). These findings underscore the potential of our panel not only for TMB assessment but also for uncovering crucial immunoregulatory mechanisms underlying tumor response to therapy.

Building upon the insights of Wang et al. (2020) and Kang et al. (2020b), our analysis delved into the identification of Differentially Expressed Genes (DEGs) between the two Tumor Mutational Burden (TMB) groups, elucidating their impact on signaling pathways. Our findings revealed perturbations in numerous immunity pathways, highlighting the significance of our signature in modulating immune and inflammatory responses. Consequently, mutations in these genes or alterations in their activity are likely to

exert a substantial influence on immune responses, thereby affecting the efficacy of immunotherapy interventions.

5 Limitations of the study

The main limitation of our study is the lack of patients who have undergone immunotherapy to understand better our panel role in predicting its clinical outcome, in a future work we aim to investigate those data. Moreover, the lack of standardization for the TMB thresholds makes determining the best one for each cancer type impossible. Indeed, clinical trials are needed to confirm our data.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.cancer.gov/tcga>, <https://www.ncbi.nlm.nih.gov/gap/>, <https://dcc.icgc.org/>.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

GP: Formal Analysis, Investigation, Software, Validation, Visualization, Writing—original draft, Writing—review and editing. SA: Conceptualization, Funding acquisition, Methodology, Supervision, Writing—review and editing. AC: Formal Analysis, Investigation, Software, Writing—review and editing. AF: Funding acquisition, Writing—review and editing. SF: Project administration, Supervision, Writing—review and editing. AP: Conceptualization, Funding acquisition, Project administration, Supervision, Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. AP, SA, and AF have been partially supported by the following research projects: PO-FESR Sicilia 2014–2020 “DiOncoGen: Innovative diagnostics” (CUP G89J18000700007). AP, has been also partially supported by the following research project: “PROMOTE: Identificazione di nuovi biomarcatori per la diagnosi precoce di mesotelioma maligno pleurico in soggetti ex esposti a fibre asbestiformi,” University of Catania—Piano di incentivi per la ricerca 2020–2022.

Acknowledgments

The results presented in the paper are based on data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

References

- Alaimo, S., Giugno, R., Acunzo, M., Veneziano, D., Ferro, A., and Pulvirenti, A. (2016). Post-transcriptional knowledge in pathway analysis increases the accuracy of phenotypes classification. *Oncotarget* 7, 54572–54582. doi:10.18632/oncotarget.9788
- Alaimo, S., Marceca, G., Ferro, A., and Pulvirenti, A. (2017). Detecting disease specific pathway substructures through an integrated systems biology approach. *Noncoding RNA* 3, 20. doi:10.3390/ncrna3020020
- Allgäuer, M., Budczies, J., Christopoulos, P., Endris, V., Lier, A., Rempel, E., et al. (2018). Implementing tumor mutational burden (TMB) analysis in routine diagnostics—a primer for molecular pathologists and clinicians. *Transl. Lung Cancer Res.* 7, 703–715. doi:10.21037/tlcr.2018.08.14
- Buchhalter, I., Rempel, E., Endris, V., Allgäuer, M., Neumann, O., Volckmar, A. L., et al. (2018). Size matters: dissecting key parameters for panel-based tumor mutational burden analysis. *Int. J. Cancer* 144, 848–858. doi:10.1002/ijc.31878
- Campestrato, L., Barroso-Sousa, R., Jimenez, L., Correa, B. R., Sabbaga, J., Hoff, P. M., et al. (2015). Comprehensive cancer-gene panels can be used to estimate mutational load and predict clinical benefit to PD-1 blockade in clinical practice. *Oncotarget* 6, 34221–34227. doi:10.18632/oncotarget.5950
- caret (2016). *Classification and regression training*.
- Cercek, A., Lumish, M., Sinopoli, J., Weiss, J., Shia, J., Lamendola-Essel, M., et al. (2022). Pd-1 blockade in mismatch repair-deficient, locally advanced rectal cancer. *N. Engl. J. Med.* 386, 2363–2376. doi:10.1056/NEJMoa2201445
- Chalmers, Z., Connelly, C., Fabrizio, D., Gay, L., Ali, S., Ennis, R., et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* 9, 34. doi:10.1186/s13073-017-0424-2
- Chen, C., Tang, D., Gu, C., Wang, B., Yao, Y., Wang, R., et al. (2022). Characterization of the immune microenvironmental landscape of lung squamous cell carcinoma with immune cell infiltration. *Dis. Markers* 2022, 2361507. doi:10.1155/2022/2361507
- Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2015). TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 44, e71. doi:10.1093/nar/gkv1507
- Consortium, G. P., Auton, A., Brooks, L., Durbin, R., Garrison, E., Kang, H., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi:10.1038/nature15393
- Contreras, G., Shirdel, I., Braun, M. S., and Wink, M. (2020). Defensins: transcriptional regulation and function beyond antimicrobial activity. *Dev. Comp. Immunol.* 104, 103556. doi:10.1016/j.dci.2019.103556
- Dervovic, D., Malik, A., Chen, E., Narimatsu, M., Adler, N., Afuni-Zadeh, S., et al. (2023). *In vivo* CRISPR screens reveal Serpinb9 and Adam2 as regulators of immune therapy response in lung cancer. *Nat. Commun.* 14, 3150. doi:10.1038/s41467-023-38841-7
- Diez-Villanueva, A., Sanz-Pamplona, R., Solé, X., Cordero, D., Crous-Bou, M., Guinó, E., et al. (2022). Colonomics - integrative omics data of one hundred paired normal-tumoral samples from colon cancer patients. *Sci. data* 9, 595. doi:10.1038/s41597-022-01697-5
- Doi, T., Piha-Paul, S., Jalal, S., Saraf, S., Lunceford, J., Koshiji, M., et al. (2018). Safety and antitumor activity of the anti-programmed death-1 antibody pembrolizumab in patients with advanced esophageal carcinoma. *J. Clin. Oncol.* 36, 61–67. doi:10.1200/JCO.2017.74.9846
- El-Khoueiry, A., Sangro, B., Yau, T., Crocenzi, T., Kudo, M., Hsu, C., et al. (2017). Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* 389, 2492–2502. doi:10.1016/S0140-6736(17)31046-2

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2024.1285305/full#supplementary-material>

- Endris, V., Buchhalter, I., Allgauer, M., Rempel, E., Lier, A., Volckmar, A. L., et al. (2018). Measurement of tumor mutational burden (TMB) in routine molecular diagnostics: in-silico and real-life analysis of three larger gene panels. *Int. J. Cancer* 144, 2303–2312. doi:10.1002/ijc.32002
- Fernandez, E., Eng, K., Beg, S., Beltran, H., Faltas, B., Mosquera, J., et al. (2019). Cancer-specific thresholds adjust for whole exome sequencing-based tumor mutational burden distribution. *JCO Precis. Oncol.* 3, 1–12. doi:10.1200/PO.18.00400
- Food U, Administration D (2017a). *FDA announces approval, CMS proposes coverage of first breakthrough-designated test to detect extensive number of cancer biomarkers*.
- Food U, Administration D (2017b). *FDA unveils a streamlined path for the authorization of tumor profiling tests alongside its latest product action*.
- Gao, P., He, M., Zhang, C., and Geng, C. (2018). Integrated analysis of gene expression signatures associated with colon cancer from three datasets. *Gene* 654, 95–102. doi:10.1016/j.gene.2018.02.007
- Golshani, G., and Zhang, Y. (2020). Advances in immunotherapy for colorectal cancer: a review. *Ther. Adv. Gastroenterol.* 13, 1756284820917527. doi:10.1177/1756284820917527
- Harview, T. C. P. C., Yearley, J., Shintaku, I., Ej, T., Taylor, E. J. M., Robert, L., et al. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* 515, 568–571. doi:10.1038/nature13954
- Hegde, P. S., and Chen, D. S. (2020). Top 10 challenges in cancer immunotherapy. *Immunity* 52, 17–35. doi:10.1016/j.immuni.2019.12.011
- Illumina 2021. *illumina* <https://emea.illumina.com/areas-of-interest/cancer/research/cancer-immunotherapy-research/neoantigens-mutational-load.html> (Accessed: 2021-June-09).
- Jia, Q., Wang, J., He, N., He, J., and Zhu, B. (2019). Titin mutation associated with responsiveness to checkpoint blockades in solid tumors. *JCI Insight* 4, e127901. doi:10.1172/jci.insight.127901
- Johnson, D., Frampton, G., Rioth, M., Yusko, E., Xu, Y., Guo, X., et al. (2016). Targeted next generation sequencing identifies markers of response to PD-1 blockade. *Cancer Immunol. Res.* 4, 959–967. doi:10.1158/2326-6066.CIR-16-0143
- Kang, K., Xie, F., Mao, J., Bai, Y., and Wang, X. (2020a). Significance of tumor mutation burden in immune infiltration and prognosis in cutaneous melanoma. *Front. Oncol.* 10, 573141. doi:10.3389/fonc.2020.573141
- Kang, K., Xie, F., Mao, J., Bai, Y., and Wang, X. (2020b). Significance of tumor mutation burden in immune infiltration and prognosis in cutaneous melanoma. *Front. Oncol.* 10, 573141. doi:10.3389/fonc.2020.573141
- Klempner, S., Fabrizio, D., Bane, S., Peoples, R. T. M., Ali, S., Sokol, E., et al. (2020). Tumor mutational burden as a predictive biomarker for response to immune checkpoint inhibitors: a review of current evidence. *Oncologist* 25, e147–e159. doi:10.1634/theoncologist.2019-0244
- Koboldt, D., Zhang, Q., Larson, D., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576. doi:10.1101/gr.129684.111
- Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset. *bioRxiv*. *F1000Research* 9 (ELIXIR). R package version 0.2.3. doi:10.12688/f1000research.24956.2
- Kossai, M., Radošević-Robin, N., and Penault-Llorca, F. (2021). Refining patient selection for breast cancer immunotherapy: beyond pd-1. *ESMO Open* 6, 100257. doi:10.1016/j.esmoop.2021.100257

- Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923
- Lawrence, M., Stojanov, P., Polak, P., Kryukov, G., Cibulskis, K., Sivachenko, A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218. doi:10.1038/nature12213
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110
- Meléndez, B., Van Campenhout, C., Rorive, S., Rummelink, M., Salmon, I., and D'Haene, N. (2018). Methods of measurement for tumor mutational burden in tumor tissue. *Transl. Lung Cancer Res.* 7, 661–667. doi:10.21037/tlcr.2018.08.02
- Mounir, M., Lucchetta, M., Silva, T. C., Olsen, C., Bontempi, G., Chen, X., et al. (2019). New functionalities in the tcgabiobioinformatics package for the study and integration of cancer data from gdc and gtx. *PLOS Comput. Biol.* 15, e1006701–e1006718. doi:10.1371/journal.pcbi.1006701
- NCBI (2021). *dbgap/database of genotypes and phenotypes/national center for biotechnology information, national library of medicine (ncbi/nlm)*. <https://www.ncbi.nlm.nih.gov/gap> (Accessed October 13, 2021).
- Nhlbi go exome sequencing project (esp)*. <http://evs.gs.washington.edu/EVS/> (2021).
- Picard toolkit (2018). *Picard toolkit*.
- Pruitt, K., Tatusova, T., Brown, G., and Maglott, D. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 40, D130–D135. doi:10.1093/nar/gkr1079
- Q2solution. 2021. *Q2solution* <https://www.q2labsolutions.com/en/genomics-laboratories/tmb> (Accessed: 2021-June-09).
- Qiagen. 2021. *Qiagen* <https://digitalinsights.qiagen.com/news/blog/discovery/the-qiaseq-tumor-mutational-burden-tmb-bioinformatics-workflow/> (Accessed: 2021-June-09).
- Riaz, N., Havel, J., Kendall, S., Makarov, V., Walsh, L. A., Desrichard, A., et al. (2016). Recurrent SERPINB3 and SERPINB4 mutations in patients who respond to anti-CTLA4 immunotherapy. *Nat. Genet.* 48, 1327–1329. doi:10.1038/ng.3677
- Risso, D., Schwartz, K., Sherlock, G., and Sandrine, D. (2011). GC-content normalization for RNA-seq data. *BMC Bioinforma.* 12, 480. doi:10.1186/1471-2105-12-480
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007
- Rizvi, N., Hellmann, M., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., et al. (2015). Cancer immunology. mutational landscape determines sensitivity to pd-1 blockade in non-small cell lung cancer. *Science* 348, 124–128. doi:10.1126/science.1257578
- Seshagiri, S., Stawiski, E. W., Durinck, S., Modrusan, Z., Storm, E. E., Conboy, C. B., et al. (2012). Recurrent r-spondin fusions in colon cancer. *Nature* 488, 660–664. doi:10.1038/nature11282
- Sherry, S., Ward, M., Kholodov, M., Baker, J., Phan, L., Smigielski, E., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311. doi:10.1093/nar/29.1.308
- Shu, Y., Wu, X., Shen, J., Luo, D., Li, X., Wang, H., et al. (2020). Tumor mutation burden computation in two pan-cancer precision medicine next-generation sequencing panels. *J. Comput. Biol.* 27, 1553–1560. doi:10.1089/cmb.2019.0055
- Silva, T., Colaprico, A., Olsen, C., D'Angelo, F., Bontempi, G., Ceccarelli, M., et al. (2016). TCGA Workflow: analyze cancer genomics and epigenomics data using Bioconductor packages. *F1000Research* 5, 1542. doi:10.12688/f1000research.8923.2
- Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J., Desrichard, A., et al. (2014). Genetic basis for clinical response to ctla-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189–2199. doi:10.1056/NEJMoa1406498
- Stein, A., Moehler, M., Trojan, J., Goekkurt, E., and Vogel, A. (2018). Immunoncology in gi tumours: clinical evidence and emerging trials of pd-1/pd-l1 antagonists. *Crit. Rev. Oncology/Hematology* 130, 13–26. doi:10.1016/j.critrevonc.2018.07.001
- Stenzinger, A., Allen, J., Maas, J., Stewart, M., Merino, D., Wempe, M., et al. (2019). Tumor mutational burden standardization initiatives: recommendations for consistent tumor mutational burden assessment in clinical samples to guide immunotherapy treatment decisions. *Genes Chromosom. Cancer* 58, 578–588. doi:10.1002/gcc.22733
- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2018). COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47, D941–D947–D947. doi:10.1093/nar/gky1015
- ThermoFisher. 2021. *ThermoFisher* <https://www.thermofisher.com/order/catalog/product/A37909> (Accessed: 2021-June-09).
- Wang, F., Tang, C., Gao, X., and Xu, J. (2020). Identification of a six-gene signature associated with tumor mutation burden for predicting prognosis in patients with invasive breast carcinoma. *Ann. Transl. Med.* 8, 453. doi:10.21037/atm.2020.04.02
- Wang, K. H. H., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603
- Wu, H., Wang, Z., Zhao, Q., Chen, D., He, M., Yang, L., et al. (2019a). Tumor mutational and indel burden: a systematic pan-cancer evaluation as prognostic biomarkers. *Ann. Transl. Med.* 7, 640. doi:10.21037/atm.2019.10.116
- Wu, H., Wang, Z., Zhao, Q., Wang, F., and Xu, R. (2019b). Designing gene panels for tumor mutational burden estimation: the need to shift from 'correlation' to 'accuracy'. *J. Immunother. Cancer* 7, 206. doi:10.1186/s40425-019-0681-2
- Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with annovar and wannovar. *Nat. Protoc.* 10, 1556–1566. doi:10.1038/nprot.2015.105
- Yao, L., Fu, Y., Mohiyuddin, M., and Lam, H. Y. K. (2020). eCTMB: a robust method to estimate and classify tumor mutational burden. *Sci. Rep.* 10, 4983. doi:10.1038/s41598-020-61575-1
- Yu, L., and Liu, P. (2021). Cytosolic DNA sensing by cGAS: regulation, function, and human diseases. *Sig Transduct. Target Ther.* 6, 170. doi:10.1038/s41392-021-00554-y
- Zehir, A., Benayed, R., Shah, R., Syed, A., Middha, S., Kim, H., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* 23, 703–713. doi:10.1038/nm.4333
- Zhang, C., Shen, L., Qi, F., Wang, J., and Luo, J. (2020). Multi-omics analysis of tumor mutation burden combined with immune infiltrates in bladder urothelial carcinoma. *J. Cell. Physiol.* 235, 3849–3863. doi:10.1002/jcp.29279
- Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., et al. (2019). The international cancer genome consortium data portal. *Nat. Biotechnol.* 37, 367–369. doi:10.1038/s41587-019-0055-9