Chapter 18

Computational Deconvolution of Tumor-Infiltrating Immune Components with Bulk Tumor Gene Expression Data

Bo Li, Taiwen Li, Jun S. Liu, and X. Shirley Liu

Abstract

Tumor-infiltrating immune cells play critical roles in immune-mediated tumor rejection and/or progression, and are key targets of immunotherapies. Estimation of different immune subsets becomes increasingly important with the decreased cost of high-throughput molecular profiling and the rapidly growing amount of cancer genomics data. Here, we present **T**umor **IM**mune **E**stimation **R**esource (TIMER), an in silico deconvolution method for inference of tumor-infiltrating immune components. TIMER takes bulk tissue gene expression profiles measured with RNA-seq or microarray to evaluate the abundance of six immune cell types in the tumor microenvironment: B cell, CD4+ T cell, CD8+ T cell, neutrophil, macrophage, and dendritic cell. We further introduce its associated webserver for convenient, user-friendly analysis of tumor immune infiltrates across multiple cancer types.

Key words Tumor immune interaction, Infiltrating immune cells, Cancer immunotherapy, Deconvolution, RNA-seq, Interactive website

1 Introduction

Tumor microenvironment usually consists of diverse immune cell types [1–3], including both lymphocytes and myelocytes as a consequence of tumor antigen recognition [Boon, Coul, Eynde] and immune infiltration [4–7]. Recent development in cancer immunotherapies necessitates the understanding of different immune subsets that co-localize with the tumor during cancer progression [8–12]. Clinical efforts to boost the antitumor immune responses by reinvigoration of the infiltrating cytotoxic T cells [13] or elimination of the regulatory T cells [14] and/or myeloid-derived suppressive cells [15] have demonstrated curative potentials for some late-stage cancer patients. Therefore, learning the components of the infiltrating immune cells is critical to understanding tumor immune interaction and designing effective immunotherapies targeting different immune subsets.

Sebastian Boegel (ed.), *Bioinformatics for Cancer Immunotherapy: Methods and Protocols*, Methods in Molecular Biology, vol. 2120, https://doi.org/10.1007/978-1-0716-0327-7_18, © Springer Science+Business Media, LLC, part of Springer Nature 2020

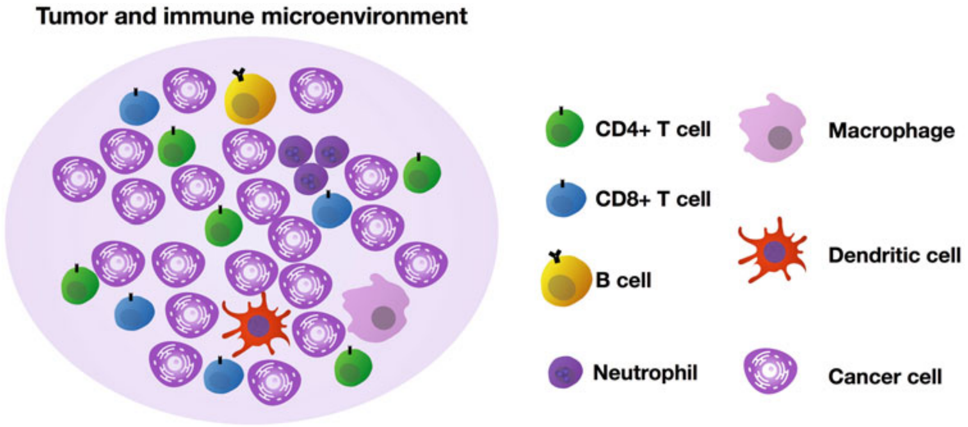

Fig. 1 Overview of the immune components in the tumor microenvironment. Diverse immune subsets have been identified in the tumor microenvironment, and the heterogeneity of immune infiltrates across individuals have been associated with the clinical outcome for diverse cancer types

Bulk tumor tissue is a mixture of diverse cell types, including cancer cells, immune cells, endothelial cells (blood vessel), fibroblasts, and so on (Fig. 1). Experimental procedures to dissociate and individually process each cell type are currently only available for small study cohorts due to cost and logistic considerations. Alternatively, with the rapid accumulation of tumor gene expression data, especially the large cancer consortium studies such as TCGA [16–19], ICGC [8, 20–22], and TARGET [23, 24], it is highly desirable to implement a computational deconvolution algorithm to infer immune cell compositions from bulk tissue gene expression data.

High-throughput molecular profiling takes whole transcriptome as input and generates a measurement of gene expression. Currently, two platforms are commonly used for gene expression profiling: massive parallel RNA sequencing, or RNA-seq, and whole-genome gene expression microarray. In general, RNA-seq produces more reliable and unbiased measurements, where sometimes microarray is less robust for lowly expressed transcripts and saturates for high gene expression [25]. Nonetheless, when taking bulk tissue sample, which is a mixture of different cell types, as input, both platforms produce gene expression measurement for all the cell types. It is expected that the observed expression level for each gene is a weighted linear combination of all different cell types in the sample:

$$\Upsilon_i = \sum_j f_j \times X_i^j + \varepsilon_i \tag{1}$$

where Υ_i is the observed gene expression level of gene i, f_j the relative abundance of cell type j in the tissue, X_i^j the unobserved expression level for gene i in cell type j, and ε_i the measurement error. This equation is a general mathematical representation of deconvolution problems with finite number of hidden components. It is straightforward that when neither f_j nor X_i^j is known, the above inference is computationally unidentifiable: for any hypothesized solution of f_j and X_i^j, $m \times f_j$ and $\frac{1}{m} \times X_i^j$ will be a valid solution as well, where m is any finite positive number.

Depending on the nature of the biological problem, sometimes it is feasible to apply a normalization constraint of f_j to make the above problem identifiable:

$$\sum_j f_j = 1 \tag{2}$$

The constraint dictates that the relative abundances for all the cell types in consideration must sum up to 1. The hidden hypothesis behind it is that even though the tissue is an unknown mixture, the collection of cell types is finite and known. For example, when studying the expression levels of normal brain tissue, it is usually valid to assume that the tissue is a mixture of several well-studied cell types in the central neural system, such as neurons, oligodendrocytes, astrocytes, glia, and microglia cells.

This strategy, however, usually does not apply to tumor tissues, as there remain uncharacterized cell types in the microenvironment. Alternatively, it is sometimes feasible to obtain X_i^j for cell types of interest, to make the equation identifiable. This is true to immune infiltrates, as previous studies have profiled many pure immune subsets through flow cytometry. The major challenge to study tumor tissue gene expression is, due to genome instability, gene expression patterns of tumor cells are usually unknown. Thus, the deconvolution problem for bulk tumor tissue can be generally written as:

$$Y_i = T_i + \sum_{j \in \text{immune cells}} f_j \times X_i^j + \sum_{s \in \text{other cell types}} f_s \times X_i^s + \varepsilon_i \tag{3}$$

In the above equation, T_i is the unknown tumor expression for gene i, with the two summations representing immune and other cell types, and for most cancer types, the latter remains poorly understood. The goal in the TIMER algorithm is to estimate f_j for the immune subsets using bulk tumor gene expression data.

2 Materials

TIMER [26] is written in the R programming language [27]. We applied TIMER to estimate abundances of different immune cell types using the TCGA data and presented the TIMER webserver for users to conveniently explore the associations between immune cells and a wide spectrum of patient clinical features. We will explain the rationale of TIMER methodology in this subheading and introduce the usage of the webserver in the next.

2.1 Tumor Purity Estimation

Tumor purity is defined as the fraction of malignant cells in the tumor tissue in a genomic sequencing experiment. Due to genome instability, it is expected that cancer cells usually carry copy number alterations (CNAs), which distinguish them from normal somatic cells. Purity is estimated using R package CHAT [28], which takes allele-specific SNP array data to infer the fraction of cells with aneuploidy genome (Fig. 2a). Detailed usage of CHAT and its input data format is available at:

https://sourceforge.net/p/clonalhetanalysistool/wiki/CHAT/

For each sample, CHAT outputs an estimation of tumor purity (AGP) and its associated quality score (PoP). PoP is the percentage of genome with copy number changes, which is important in purity estimation. PoP smaller than 0.05 indicates that the inference is unreliable due to limited information and needs to be excluded from downstream analysis. AGP is a value ranging from 0 to 1. Purity equals to 1 means that the sample contains almost none of noncancerous cells. A complete set of purity inference for all the TCGA samples is available at:

http://cistrome.org/TIMER/misc/AGPall.zip

2.2 Selection of Informative Genes and Model Simplification

As it is usually infeasible to learn the true values of T_i for different cancer types, we select genes that are lowly expressed in the cancer cells to exclude this component in the model. Specifically, for each cancer type, we calculated the Pearson's correlation (r) between the expression levels of each gene and tumor purity. Genes with negative r is higher expressed in the tumor microenvironment than in the tumor. We selected genes with $r < -0.2$ as informative markers (Fig. 2b). When the inference is restricted to these genes, Eq. (3) can be rewritten as:

$$\Upsilon_i = \sum_{j \in \text{immune cells}} f_j \times X_i^j + \sum_{s \in \text{other cell types}} f_s \times X_i^s + \varepsilon_i \qquad (4)$$

We further selected markers with expression enriched or restricted within the immune cell lineage [8] from purity-selected genes (Fig. 2c), to exclude the effects of other cell types in the microenvironment. Eq. (4) is then simplified as:

$$\Upsilon_i = \sum_{j \in \text{immune cells}} f_j \times X_i^j + \varepsilon_i \qquad (5)$$

2.3 Selection of Immune Cell Types

Hematopoietic stem cells give rise to all the immune cells, including the lymphocytes and myelocytes [29]. Major immune cell types in the tumor include T/B cell, natural killer (NK) cell, monocyte, macrophage, neutrophil, dendritic cell, and so on. Each of the major cell type may also be further divided into subgroups. For

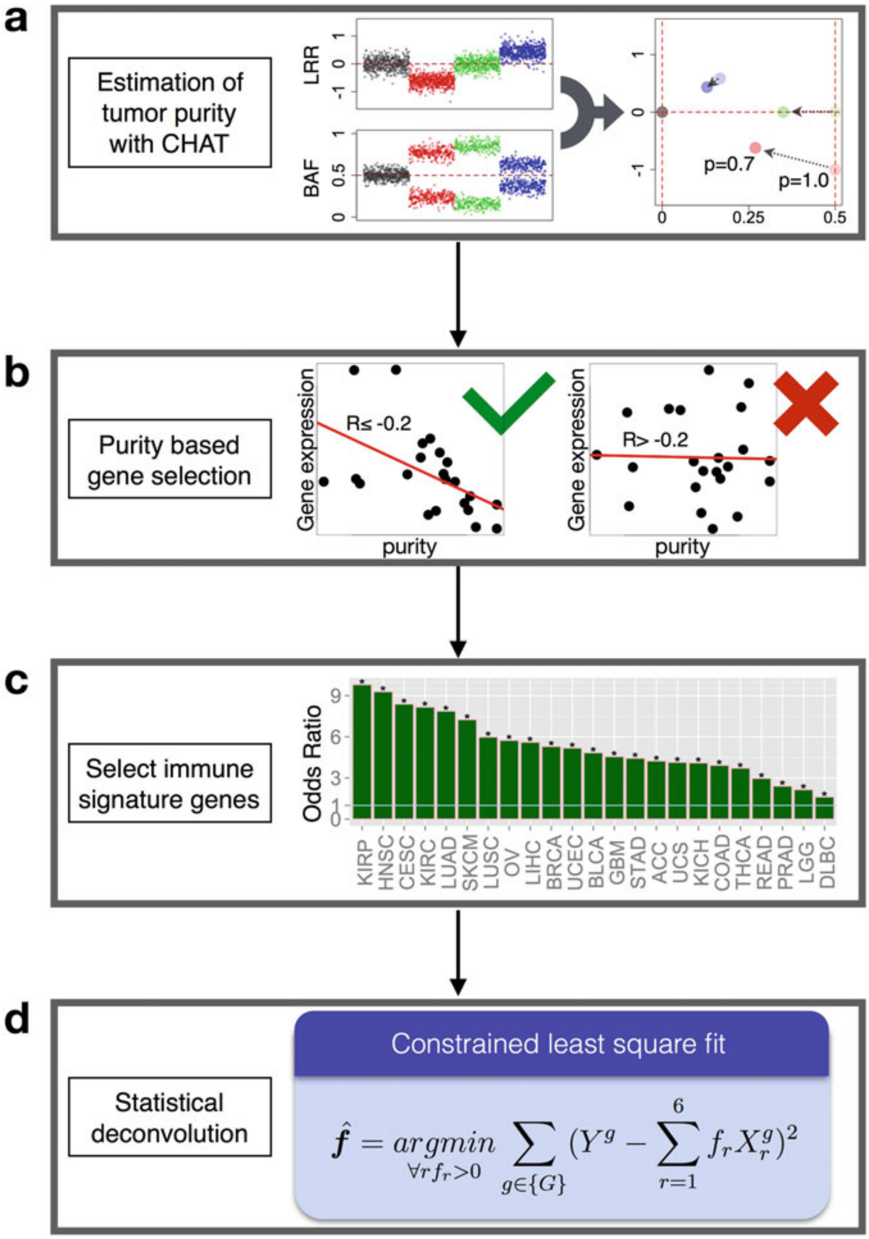
a
Estimation of tumor purity with CHAT
LRR
-1 0 1
BAF
0 0.5 1
1
0
-1
p=0.7
p=1.0
0
0.25
0.5
b
Purity based gene selection
Gene expression
R≤ -0.2
purity
Gene expression
R> -0.2
purity
c
Select immune signature genes
Odds Ratio
9
6
3
1
0
KIRP
HNSC
CESC
KIRC
LUAD
SKCM
LUSC
OV
LIHC
BRCA
UCEC
BLCA
GBM
STAD
ACC
UCS
KICH
COAD
THCA
READ
PRAD
LGG
DLBC
d
Statistical deconvolution
Constrained least square fit
$\hat{f} = \underset{\forall r f_r > 0}{argmin} \sum_{g \in \{G\}} (Y^g - \sum_{r=1}^{6} f_r X_r^g)^2$

Fig. 2 Flowchart of the TIMER methodology. TIMER uses genomic estimation of tumor purity (**a**) to select genes with higher expression in the microenvironment. These genes typically show negative correlations with tumor purity (**b**). A substantial fraction of the purity-selected genes are enriched for immune signature (**c**), which are further selected to estimate the immune components with constraint least square fitting (**d**). (This flowchart is a modified version of Fig. 1 in [26])

example, the T-cell population contains CD4+ and CD8+ T cells and macrophage contains M1 or M2 polarized subtypes. Sometimes these subtypes can be split into even finer subsets. CD4+ T cells consist of a mixture of naïve, regulatory, helper cells, where CD8+ T cells in the tumor usually harbor effector, memory, and exhausted cell types. While it is attractive to deconvolve all the known immune cell subsets, due to computational limitations, it is practical to include only selected cell types in the model.

The first limitation is *identifiability*. Since it is impossible to include all immune cell types, the normalization constraint Eq. (2) usually does not apply. To make Eq. (5) identifiable, one needs to known $X_\bullet^j$, that is, the expression profile for cell type j. Such data can be obtained from previous experiments on sorted immune cells from human blood samples [30] but limited for only a few cell types. The second limitation is *statistical colinearity*. Theoretically, introducing highly correlated terms in regression models will generate unstable estimations of the coefficients. For example, if $X_\bullet^k$ and $X_\bullet^l$ represent two immune cell types with similar gene expression pattern, the estimation for f_k and f_l will be inaccurate [31]. Due to shared lineage or similar functions, different immune cell types can be very similar, for example, NK and CD8+ T cell, regulatory and helper CD4+ T cells, monocyte and macrophage, and so on. Therefore, we selected six linearly-separable cell types: B cell, CD4+ T cell, CD8+ T cell, macrophage, neutrophil, and dendritic cell for deconvolution.

2.4 Constrained Least Square Fitting

Reference expression levels of selected immune cell types, $X_\bullet^j$ were available in the public domain, and downloadable at

`http://cistrome.org/TIMER/misc/HPCTimmune.Rdata`

For tumor samples profiled from either RNA-seq or Affymetrix HGU133plus2 microarray, the immune components can be inferred using Eq. (5) with the reference immune cell expression data. Coefficients f_j for each cell type were estimated with least square fitting with constraint $f_j \geq 0$ (Fig. 2d). Coefficients are comparable across individuals but not between different cell types (*see* **Note 1**). This method can be implemented with the `getFractions.Abbas` function in the following R codes:

`http://cistrome.org/TIMER/codes/CancerImmunePipeline.R`

3 Methods

As discussed above, unbiased estimation of selected immune cell types from bulk tissue data is practical but only accurate when tumor purity can be estimated for informative gene selection. The Cancer Genome Atlas provided both DNA and RNA profiling data for this task. Based on the inference of immune cell abundance, we developed the TIMER website [32] for users to explore different aspects of tumor immune interactions:

`https://cistrome.shinyapps.io/timer/`

3.1 Overview of TIMER Website

TIMER consists of four modules for correlative analysis of immune infiltrates estimated from the TCGA data, including `Gene`, `Survival`, `Mutation`, and `SCNA` modules. There are two additional modules for convenient investigation of differentially expressed genes between tumor and adjacent normal tissues (`DiffExp`), or the correlations between a pair of genes (`Correlation`). In the following are detailed instructions to the website, and there is also a step-by-step guide on YouTube:

https://youtu.be/94v8XboCrXU

3.2 `Gene` Module

This module is intended to study the correlation between gene expression and the abundance of given immune cell type(s). Three input boxes, **Gene Symbol**, **Cancer Types**, and **Immune Infiltrates** are presented on the webpage. Users need to type in the official symbol for one gene. The website will suggest alternative spells if the input was not found in the database. Users may select one or more cancer types by either typing in the cancer disease names or selecting from the drop down list in the **Cancer Types** input box. All TCGA cancer name and abbreviations are available in Table 1. By default, all six immune cell types are included in the **Immune Infiltrates**, and users can delete one or more cell type from the analysis. After the parameters are selected, clicking the "Submit" button will start the analysis, where the partial Spearman's correlation is calculated for each pair of gene/infiltrate. Tumor purity is automatically corrected for each analysis since it is a known confounding factor for both gene expression and immune infiltration levels.

For each cancer type, 1 + **X** scatter plots are returned (Fig. 3), with the first one being gene expression against tumor purity, and the following figures for **X** selected immune cell types ($0 \leq \mathbf{X} \leq 6$). On top of each plot, the Lowess smooth curve with confidence interval estimations is overlaid to visualize the trend of correlation. The figures can be directly downloaded as JPG or PDF format.

3.3 `Survival` Module

This module builds flexible Cox proportional regression models, with diverse variable options. First, users can select one or more diseases in the **Cancer Types** input field. Three categories of covariates are allowed in the model, including clinical factors (**Clinical**), infiltrating immune cell abundance (**Immune Infiltrates**), and the expression levels of given gene(s) (**Gene Symbols**). For each category, one or more covariates can be included. After all the fields have been set, the resulting Cox model will be immediately displayed on the right side of the webpage. Estimations of parameters are displayed as a table, with the following columns: coef (log hazard ratio), HR (hazard ratio), 95%CI_l (lower boundary of 95% confidence interval for HR), 95%CI_u (upper boundary of 95% confidence interval for HR), *p*.value (*p* value for individual

Table 1
Disease full name and abbreviations of 33 cancer types covered in TCGA and TIMER website

Abbreviation	Disease full name
ACC	Adrenocortical carcinoma
BLCA	Bladder urothelial carcinoma
BRCA	Breast invasive carcinoma
CESC	Cervical and endocervical cancers
CHOL	Cholangiocarcinoma
COAD	Colon adenocarcinoma
DLBC	Lymphoid neoplasm diffuse large B-cell lymphoma
ESCA	Esophageal carcinoma
GBM	Glioblastoma multiforme
HNSC	Head and neck squamous cell carcinoma
KICH	Kidney chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LAML	Acute myeloid leukemia
LGG	Brain lower grade glioma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MESO	Mesothelioma
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma and paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin cutaneous melanoma
STAD	Stomach adenocarcinoma
TGCT	Testicular germ cell tumors
THCA	Thyroid carcinoma
THYM	Thymoma
UCEC	Uterine corpus endometrial carcinoma
UCS	Uterine carcinosarcoma
UVM	Uveal melanoma

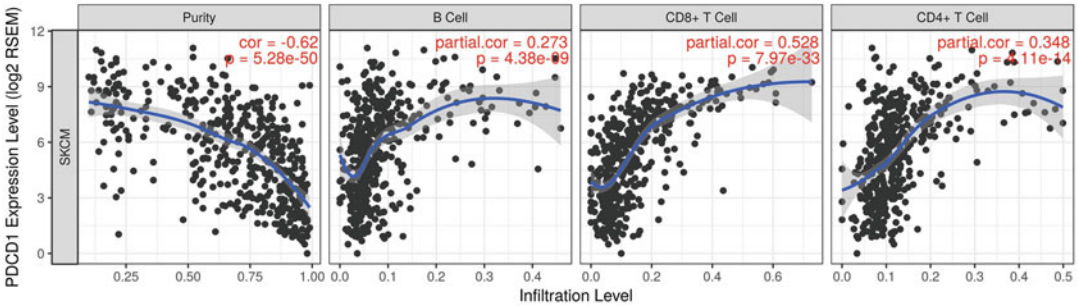

Fig. 3 Example of Gene module in the TIMER website. When choosing PDCD1 as Gene Symbol, metastatic melanoma (SKCM) as Cancer Type, and B cell, CD4+, and CD8+ T cells as immune infiltrates, the website returns 4 scatter plots. The first one is always comparison of gene expression and tumor purity, with the following panels for each of the selected immune cell type

Table 2
Example of Survival module in the TIMER website

Model: Surv(LUAD) ~ Age + Purity + B_cell + CD19						
455 patients with 161 dying						
	coef	**HR**	**95% CI_l**	**95% CI_u**	***p*.value**	**sig**
Age	0.009	1.009	0.992	1.025	0.303	
Purity	−0.163	0.849	0.416	1.736	0.655	
B_cell	−2.638	0.071	0.007	0.701	0.023	*
CD19	−0.069	0.933	0.845	1.03	0.17	
R square = 0.043 (max possible = 9.75e-01)						
Likelihood ratio test: $p = 5.12e04$						
Wald test: $p = 1.31e\text{-}03$						
Score (log rank) test: $p = 1.04e\text{-}03$						

When choosing lung adenocarcinoma (LUAD) as Cancer Type, B cell as immune infiltrates, Age as clinical factor, and CD19 as Gene Symbol, the website returns a summary table of the Cox proportional hazard built with the given variables

covariate), and sig (significance levels). Test results for the complete model are displayed on the bottom. For example, when selecting lung adenocarcinoma as cancer type, age and purity as clinical cofactors, B cell as immune infiltrates, and CD19 as gene symbol, the website returns Table 2.

With the above parameters, users can also generate Kaplan-Meier curves (Fig. 4) for each selected covariate (excluding clinical cofactors) using the "Plot KM Curve" button. Survival of patients with the upper X% and lower X% of the covariate will be compared, with X ranging from 5 to 50 (**Split Percentage of Patients** slider). Users may also choose to view a subset of the data by limiting

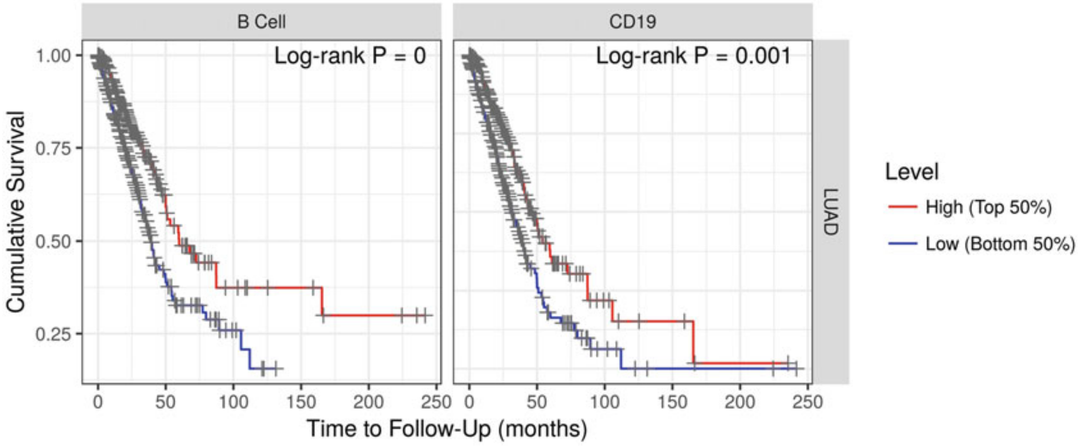

Fig. 4 Example of Survival module in the TIMER website. When choosing lung adenocarcinoma (LUAD) as Cancer Type, B cell as immune infiltrates and CD19 as Gene Symbol, the website returns two Kaplan-Meier curves, each for one covariate. By default, the upper and lower 50% of samples for each variable are displayed in the two groups and compared with Log-rank test for statistical significance

survival time within a certain period (**Survival Time Between** slider). It should be noted that truncating data with an empirical time window is for visualization purposes only and shall not be used for statistical significance estimations.

3.4 `Mutation` and `SCNA` Modules

Cancer somatic single nucleotide mutations (SNVs) or copy number alterations (SCNAs) may induce immune responses either via signaling pathways within the cancer cells or providing novel antigenic targets for the adaptive immune system. The `Mutation` and `SCNA` modules allow users to explore the heterogeneity of immune infiltrates in tumors with different mutational backgrounds. In both modules, the cancer type needs to be selected first. As most somatic SNVs are rare events [33], in the `Mutation` module, we focus on most frequent mutations to ensure statistical power in the analysis. Specifically, if a cancer type has less than 50 mutations with frequency greater than 10%, we select the top 50 mutations. Otherwise mutations with frequency $\geq$ 10% are selected. Users may choose any SNVs in the **Gene with Mutation** drop down list to visualize the results. Clicking on the “Submit” button will return a Boxplot with 6 pairs of distributions, each for one immune cell type. The infiltration levels between mutated or wild type samples are compared with two-sided Wilcoxon rank sum test, with significance level labeled on the top of each pair. For example, when cancer type is selected as ovarian cancer (OV) and mutation as CSMD3, the website returns Fig. 5, where CD4+ T cell and dendritic cell showing significantly higher levels in mutated samples.

In the SCNA module, users can input the official symbol for the gene of interest, and the copy number values estimated from

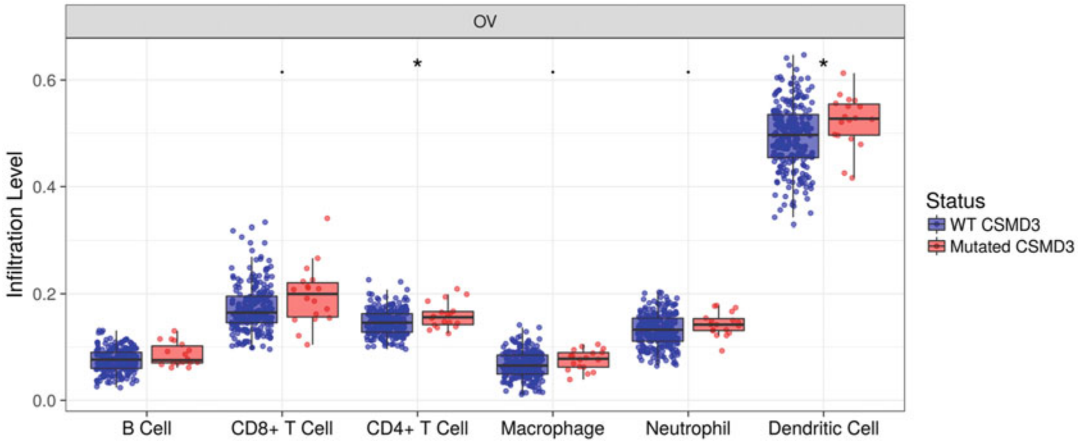

Fig. 5 Example of Mutation module in the TIMER website. When choosing ovarian cancer (OV) as Cancer Type and CSMD3 mutation, the website returns six pairs of boxplots, each for one immune cell type. Infiltration levels in mutated or wild type (WT) individuals are compared with Wilcoxon rank sum test and the statistical significance levels are labeled above the boxes

GISTIC 2.0 [34] will be used for the analysis. There are five types of copy number events defined in the GISTIC method: focal deletion ($n = -2$), arm-level deletion ($n = -1$), normal diploid ($n = 0$), arm-level deletion ($n = 1$), and focal amplification ($n = 2$). Once parameters are selected, clicking the "Submit" button will return a Boxplot with six sets of distributions, each for one cell type. Samples with copy number variations ($n \neq 0$) are compared with diploid samples with Wilcoxon rank sum test, and significance levels are labeled on top of each box.

3.5 DiffExp Module

Perhaps one of the most commonly implemented analyses in cancer research is to compare gene expression levels between tumor and the matched normal samples. In TIMER, we provide a convenient solution for quick visualization of such comparisons for all the genes in the database. In this module, users simply input the gene symbol, and click "Submit." The website will return a comprehensive Boxplot showing the distributions of the gene expression levels (in RSEM [35]). In total, 33 cancer types and 17 normal tissue types are investigated, and Wilcoxon rank sum test is applied to all the cancer/normal pairs to estimate statistical significance. Major subtypes of selected cancers are also displayed as individual columns on the plot for quick comparisons.

3.6 Correlation Module

This module explores the dependency between pairs of genes. Users may type in m genes in the **Gene Symbols: (Y-axis)** input field, and n genes in the **Gene Symbols: (X-axis) field**. One founding factor (either tumor purity or age) can be included in the analysis using the drop down list in the **Correlation Adjusted by** field. If no confounding factor is selected, clicking "Submit"

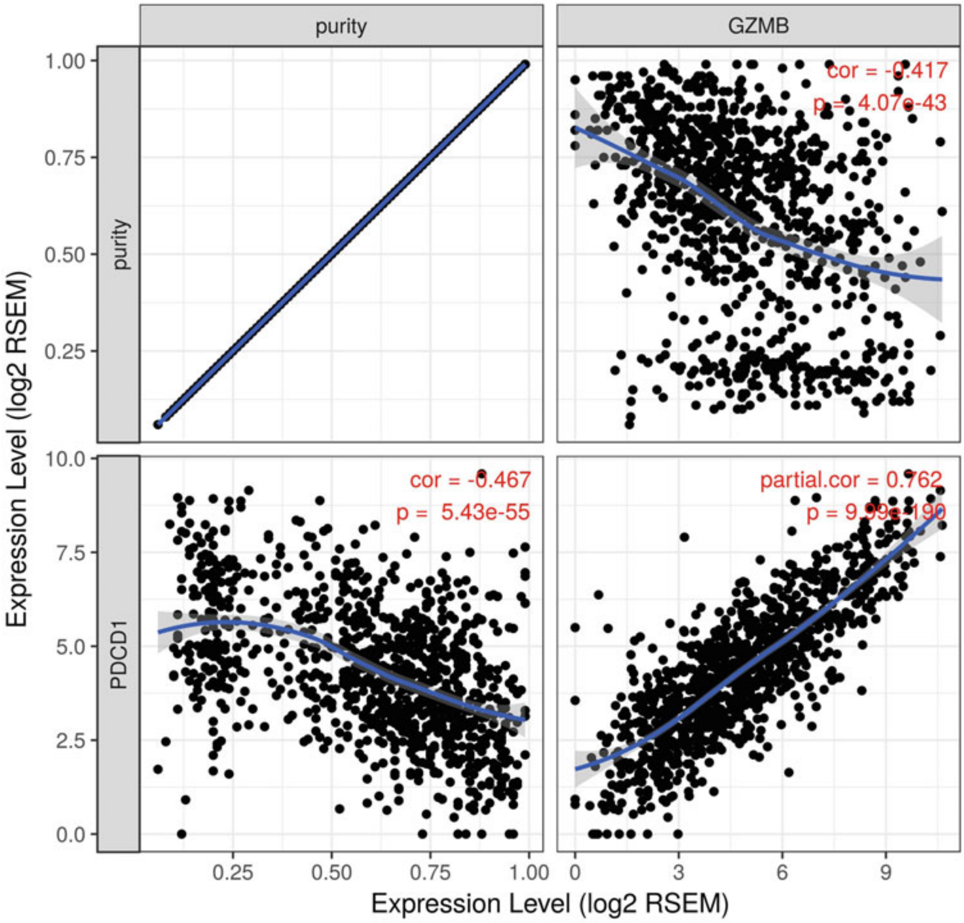

Fig. 6 Example of Correlation module in the TIMER website. With Y-axis gene selected as PDCD1 and X-axis gene as GZMB, when correcting for purity, the website returns a set of four scatter plots. The upper left panel is noninformative, with the rest the correlations between gene expression and either purity or the expression levels of the other gene. Gene–gene correlation (lower-left) is corrected by purity and estimated by partial Spearman's correlation

button will return m-by-n scatter plots, each displaying the correlation of the corresponding pair of genes.

If one of the factor is selected for adjustment, $(m + 1)$-by-$(n + 1)$ plots will be returned, with the first row and column of figures showing the dependencies of the gene expression levels and the confounding factor. Correlations between genes are calculated with partial Spearman's correlation corrected for the selected factor. For example, when choosing metastatic melanoma as cancer type, PDCD1 as Y-axis, and GZMB as X-axis, without purity adjusted, the website returns Fig. 6, and it is clear that the two genes are significantly correlated. This is in line with the observation that effector T cells secreting GZMB in the tumor are likely to become exhausted [36].

4 Notes

1. Since we chose not to apply the normalization constraint in Eq. (2), the estimated fractions usually do not sum up to 1. In addition, different immune cell type may express informative marker genes at different levels, which will impact the scales of f_j. As a result, comparison between two immune cell types of the same sample, that is, f_j and f_k, is meaningless. We have demonstrated that enforcing normalization constraint to make f_j comparable across immune cell types will falsely impose negative correlations between the estimated infiltration levels [31]. Therefore, all the analysis performed in the TIMER website are restricted to comparisons across individuals for the same immune cell type.

References

1. Binnewies M, Roberts EW, Kersten K et al (2018) Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med 24:541–550
2. Junttila MR, de Sauvage FJ (2013) Influence of tumour micro-environment heterogeneity on therapeutic response. Nature 501:346–354
3. Quail DF, Joyce JA (2013) Microenvironmental regulation of tumor progression and metastasis. Nat Med 19:1423–1437
4. Dushyanthen S, Beavis PA, Savas P et al (2015) Relevance of tumor-infiltrating lymphocytes in breast cancer. BMC Med 13:202
5. Eruslanov E, Neuberger M, Daurkin I et al (2012) Circulating and tumor-infiltrating myeloid cell subsets in patients with bladder cancer. Int J Cancer 130:1109–1119
6. Marvel D, Gabrilovich DI (2015) Myeloid-derived suppressor cells in the tumor microenvironment: expect the unexpected. J Clin Invest 125:3356–3364
7. Pages F, Galon J, Dieu-Nosjean MC et al (2010) Immune infiltration in human tumors: a prognostic factor that should not be ignored. Oncogene 29:1093–1102
8. Abbas AR, Baldwin D, Ma Y et al (2005) Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes Immun 6:319–331
9. Gubin MM, Zhang X, Schuster H et al (2014) Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature 515:577–581
10. Hamid O, Robert C, Daud A et al (2013) Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. N Engl J Med 369:134–144
11. Tran E, Turcotte S, Gros A et al (2014) Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. Science 344:641–645
12. Van Allen EM, Miao D, Schilling B et al (2015) Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 350:207–211
13. Rosenberg SA, Restifo NP, Yang JC et al (2008) Adoptive cell transfer: a clinical path to effective cancer immunotherapy. Nat Rev Cancer 8:299–308
14. Simpson TR, Li F, Montalvo-Ortiz W et al (2013) Fc-dependent depletion of tumor-infiltrating regulatory T cells co-defines the efficacy of anti-CTLA-4 therapy against melanoma. J Exp Med 210:1695–1710
15. Ko JS, Zea AH, Rini BI et al (2009) Sunitinib mediates reversal of myeloid-derived suppressor cell accumulation in renal cell carcinoma patients. Clin Cancer Res 15:2148–2157
16. The Cancer Genome Atlas Research Network (2012) Comprehensive molecular portraits of human breast tumours. Nature 490:61–70
17. The Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455:1061–1068
18. The Cancer Genome Atlas Research Network (2012) Comprehensive genomic

characterization of squamous cell lung cancers. Nature 489:519–525

19. The Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. Nature 474:609–615
20. Nik-Zainal S, Davies H, Staaf J et al (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534:47–54
21. Richter J, Schlesner M, Hoffmann S et al (2012) Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. Nat Genet 44:1316–1320
22. Tirode F, Surdez D, Ma X et al (2014) Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov 4:1342–1353
23. Bolouri H, Farrar JE, Triche T Jr et al (2018) The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med 24:103–112
24. Ma X, Liu Y, Liu Y et al (2018) Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature 555:371–376
25. Zhao S, Fung-Leung WP, Bittner A et al (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 9:e78644
26. Li B, Severson E, Pignon JC et al (2016) Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol 17:174
27. R Core Team: R: A language and environment for statistical computing. 2014
28. Li B, Li JZ (2014) A general framework for analyzing tumor subclonality using SNP array and DNA sequencing data. Genome Biol 15:473
29. Orkin SH, Zon LI (2008) Hematopoiesis: an evolving paradigm for stem cell biology. Cell 132:631–644
30. Mabbott NA, Baillie JK, Brown H et al (2013) An expression atlas of human primary cells: inference of gene function from coexpression networks. BMC Genomics 14:632
31. Li B, Liu JS, Liu XS (2017) Revisit linear regression-based deconvolution methods for tumor gene expression data. Genome Biol 18:127
32. Li T, Fan J, Wang B et al (2017) TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res 77: e108–e110
33. Martincorena I, Campbell PJ (2015) Somatic mutation in cancer and normal cells. Science 349:1483–1489
34. Mermel CH, Schumacher SE, Hill B et al (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 12:R41
35. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323
36. Crespo J, Sun H, Welling TH et al (2013) T cell anergy, exhaustion, senescence, and stemness in the tumor microenvironment. Curr Opin Immunol 25:214–221