Cancer Research

# Cancer Immunity and Gene Expression Data: A Quick Tool for Immunophenotype Evaluation

Masayuki Hirano

The rapid advancement of next generation sequencing technology has resulted in accumulation of many datasets in cancer clinical and research laboratories, many of which do not have bioinformaticians. Xu and colleagues developed a user-friendly web-based tool to define the tumor immunophenotype among patients with cancer. By uploading user-defined datasets on the web, it can systematically track, analyze, and visualize the status of anticancer immune activity and the proportion of tumor-infiltrating immune cells. This tool can help immunologists and clinical researchers to perform quick, efficient, and comprehensive analysis of the tumor immunophenotype. *Cancer Res; 78(23); 6536–8. ©2018 AACR.*

*See related article by Xu et al., p. 6575*

Cancer immunotherapy has been rapidly advancing following the discovery that therapeutic blockade of immune checkpoints unleashes antitumor immunity (1). Tumor cells exploit these checkpoints as a major mechanism for immune escape, especially from tumor-specific T cells. Because many immune checkpoints are initiated by ligand–receptor interactions, they can be simply blocked by antibodies or modulated by recombinant ligands or receptors. Programmed cell death protein 1 (PD-1) and CTL–associated antigen 4 (CTLA4) are major immune checkpoint proteins, and antibodies targeting them induce broad and diverse antitumor immune responses. The process of tumor immune response can be broken down into seven steps called the cancer–immunity cycle (2): release of cancer cell antigen (step 1), antigen capture and presentation by dendritic cells (DC; step 2), effector T-cell recognition against the cancer-specific antigen (step 3), effector T-cell trafficking to tumors (step 4), infiltration of effector T cells into tumors (step 5), T-cell recognition of and binding to cancer cells (step 6), and killing of the target cancer cells (step 7). Each step of the cancer–immunity cycle is coordinated by various stimulatory and inhibitory factors (2). Stimulatory factors promote immune response, whereas inhibitory factors enable cancer progression and downregulate immune activity and/or prevent autoimmunity. Some of the immune checkpoint proteins, including CTLA4, inhibit amplification of an active immune response by acting primarily at the level of T-cell development and proliferation at step 3. PD-1 has an inhibitory function that regulates active immune responses both in the tumor bed (step 7) and in T-cell development and proliferation (step 3). Understanding these steps helps to delineate how cancer interacts with the immune system and enables step-specific targeting in cancer immunotherapy to promote targeted treatment strategies for individual immune responses.

Tumors are mixtures of many cell types, including tumor-infiltrating immune cells as well as malignant cells (3, 4). Furthermore, the immune cells within tumors are composed of various cell types with different functions [e.g., cytotoxic $CD8^+$ T cells that affect antitumor activity and regulatory T (Treg) cells with immunosuppressive function]. Hence, detailed characterization of tumor immunophenotype is critical for understanding tumor status to guide cancer treatment. Imaging and cellular phenotyping assays such as IHC, immunocytochemistry, and flow cytometry analyses are used for immunophenotype characterization in tumors and provide important cellular contextual information. In addition to these techniques, genetic and computational approaches based on immune-specific marker genes or gene expression signatures can be applied to characterize the composition of tumor-infiltrating immune cells. Therefore, gene analysis tools have been developed to provide comprehensive information about tumor-infiltrating immune cells. The computational gene analysis tools can be categorized into gene set enrichment analysis (GSEA) methods and deconvolution methods (3). Both assay types are based on a matrix of expression profiles for individual cell subpopulations. These methods reconstruct tumor-infiltrating immune cell subpopulations, which are defined in the reference matrix of expression profiles. The GSEA method is often used to identify gene markers (3–5). This method ranks genes according to their expression levels in a sample and calculates an enrichment score (ES) based on the position of cell type–specific gene markers in the ranked list. Alternatively, a single-sample GSEA (ssGSEA) ES can be used, which represents the degree to which genes are coordinately up- or downregulated within a single sample (3, 4). Microarray or RNA sequencing (RNA-seq) data from the tumor microenvironment can be used to characterize the immunophenotype and to detect immune cell type–specific markers and gene expression signatures. A merit of the GSEA method is that it can be applied to existing tools and does not require extra sampling compared with conventional gene expression analysis. It only requires the assembly of gene signatures related to each immune subpopulation.

Deconvolution methods make use of expression signature matrices to deduce specific cell proportions from expression data of cell mixtures (3, 6). The recently developed CIBERSORT tool

Department of Pathology and Laboratory Medicine, Emory Vaccine Center, Emory University School of Medicine, Atlanta, Georgia.

**Corresponding Author:** Masayuki Hirano, Department of Pathology and Laboratory Medicine, Emory Vaccine Center, Emory University School of Medicine, 1462 Clifton Rd NE, Atlanta, GA 30322. Phone: 404-727-9588; Fax: 404-727-8795; E-mail: mhiran2@emory.edu

**doi:** 10.1158/0008-5472.CAN-18-3288

enables deconvolution of 22 human immune cell types (547 genes) and converts a microarray dataset into a gene signature matrix (3). It uses an algorithm to assess the cell fractions by nu-support vector regression (ν-SVR) and was validated on microarray data of cell mixtures derived from lymphoid tissue biopsies. It showed a high accuracy in the deconvolution of up to nine immune cell subsets, whereas it demonstrated lower accuracy in the quantification of gamma–delta T cells (3), illustrating that while various successful applications of computational methodologies have been established, several issues still need to be improved (7).

As deconvolution methods are based on gene expression profiles, cell-lineage–specific DNA methylation patterns can be utilized to detect and to estimate immune cell subsets (8). For this purpose, a number of programs and tools (8) were developed using information ranging from a few methylated CpG to genome-wide loci on microarray platforms (i.e., Illumina Infinium 27k and 450k DNA methylation arrays). The epigenome can be highly variable across different cell types as is apparent from epigenome-wide association studies (9). These cell type–specific effects can be useful although the current availability of the reference methylation patterns from purified cell types is still limited.

Generally, an issue with many of these computer analyses is that they often require additional expertise in computer programming and statistical analysis. Some programs can be used to characterize the cancer–immunity cycle, but these commonly require a long execution time. Therefore, there is a need for tools that are more user-friendly to the average cancer biologist or immunologist. In this issue of *Cancer Research*, Xu and colleagues describe the web-based tool, TIP (tracking tumor immunophenotype; ref. 10) that utilizes both GSEA and deconvolution methods to depict the status of anticancer immunity and the proportion of tumor-infiltrating immune cells, respectively. To estimate the status of anticancer immunity, they manually collected 178 signature genes (23 sets) that are involved in the seven steps of the cancer–immunity cycle, with cancer immunity-related keywords (e.g. checkpoints, chemokines and MHC molecules). The activity levels of these signature gene sets were calculated using ssGSEA based on individual gene expression. The stimulatory (positive) and inhibitory (negative) gene sets in each step of the cancer–immunity cycle were calculated separately. To compare ssGSEA scores between different expression platforms and different samples, the activity score for each signature set was generated by estimating the difference between the normalized ssGSEA scores of positive and negative sets.

TIP is also designed to evaluate the proportion of various tumor-infiltrating immune cells, such as T cells, B cells, DCs, natural killer (NK) cells, and macrophages, using the CIBERSORT algorithm. When users upload microarray expression profiling data, TIP will estimate the fractions of infiltrated immune cell types for each sample by using the original leukocyte gene signature matrix (LM22) in CIBERSORT. To infer the cell proportion from bulk RNA-seq expression data, they built a gene signature expression matrix (LM14, involving seven T-cell types, B cells, CD14 and CD16 monocytes, DC and plasmacytoid DC, plasma cells, and NK cells) using single-cell RNA-seq data (scRNA-seq), which is composed of peripheral blood mononuclear cells, $CD4^+$ Th cells, and $CD4^+CD25^+$ Treg cells from the 10× Genomics datasets. For an appropriate signature gene set for LM14, differentially expressed genes among the 14 cell types and signature genes from the LM22 were merged. The LM14 gene signature matrix is a TPM expression matrix of 973 signature genes (row) and 14 immune cell types (columns). Users can just upload bulk RNA-seq expression data (raw count or TPM) or microarray expression data (log or non-log transformed) as input. The results (output) can be viewed from both a global and individual perspective for interactive evaluation. TIP displays the 23 immune activity scores reflecting the activity status of the seven-step cancer–immunity cycle for all samples from a global perspective. The activity scores and the relative proportion of tumor-infiltrating immune cells across all samples are provided. The expression pattern and the principal component analysis of 178 step-specific signature genes are also presented. Conveniently, the overall activity score for each sample from a global perspective is also displayed to investigate the immunophenotype in an individual perspective.

In conclusion, TIP is a very useful web-based tool that will support tumor immunophenotype profiling and diagnosis. Immunologists and clinical researchers can perform comprehensive analysis of tumor-infiltrating immune cells easily and efficiently without additional computer and programming skills. Contamination of normal tissue in tumor samples is always a big issue for tumor diagnosis and evaluation, therefore, consistency of cancer tissue sampling or combined evaluation with IHC and/or flow cytometry analysis may be necessary for precise evaluation and diagnosis. As with CIBERSORT software, cell fraction data from computational TIP assays are not quite consistent with the data from other biological assays for some cells types (e.g., monocytes and naïve $CD4^+$ T cells assayed by flow cytometry, and macrophages by IHC). Modification of signature genes and calculation algorithm, as well as addition of DNA methylation profiling and clinical phenotype data (survival, immunotherapy stage and response, etc.), can improve tumor-infiltrating immune cells evaluation. Computational methods for profiling cellular heterogeneity using various types of large sequencing data derived from patient tissues and cell mixtures will likely increase in popularity; these computer algorithms, including TIP, can also be applied easily to other human diseases.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Acknowledgments

This work was supported by the National Science Foundation under grant nos. 1655163 and 1755418. The author thanks Jonathan P. Rast, Satoru Osuka, and Yoichi Sutoh for their helpful comments.

Received October 17, 2018; accepted October 17, 2018; published first December 1, 2018.

## References

1. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer 2012;12:252–64.
2. Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. Immunity 2013;39:1–10.

3. Hackl H, Charoentong P, Finotello F, Trajanoski Z. Computational genomics tools for dissecting tumour-immune cell interactions. Nat Rev Genet 2016;17:441–58.
4. Finotello F, Trajanoski Z. Quantifying tumor-infiltrating immune cells from transcriptomics data. Cancer Immunol Immunother 2018;67:1031–40.
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;102:15545–50.
6. Gong T, Szustakowski JD. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. Bioinformatics 2013;29:1083–5.
7. Shen-Orr SS, Gaujoux R. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. Curr Opin Immunol 2013;25:571–8.
8. Accomando WP, Wiencke JK, Houseman EA, Nelson HH, Kelsey KT. Quantitative reconstruction of leukocyte subsets using DNA methylation. Genome Biol 2014;15:R50.
9. Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J. Epigenome-wide association studies without the need for cell-type composition. Nat Methods 2014;11:309–11.
10. Xu L, Deng C, Pang B, Zhang X, Liu W, Liao G, et al. TIP: a web server for resolving tumor immunophenotype profiling. Cancer Res 2018;78:6576–80.