Journal of Bioinformatics and Computational Biology
(2020) 2050030 (14 pages)

DOI: 10.1142/S0219720020500304

Calculation of immune cell proportion from batch tumor gene expression profile based on support vector regression

Dongmei Ai*, Gang Liu†, Xiaoxin Li‡, Yuduo Wang§ and Man Guo¶

Basic Experimental Center of Natural Science
University of Science and Technology Beijing
Beijing 100083, P. R. China

School of Mathematics and Physics
University of Science and Technology Beijing
Beijing 100083, Beijing, P. R. China
*aidongmei@ustb.edu.cn
†s20180807@xs.ustb.edu.cn
‡s20180729@xs.ustb.edu.cn
§s20190829@xs.ustb.edu.cn
¶s20170735@xs.ustb.edu.cn

Received 21 January 2020
Revised 6 June 2020
Accepted 28 June 2020
Published 21 August 2020

In addition to tumor cells, a large number of immune cells are found in the tumor microenvironment (TME) of cancer patients. Tumor-infiltrating immune cells play an important role in tumor progression and patient outcome. We improved the relative proportion estimation algorithm of immune cells based on RNA-seq gene expression profiling and solved the multiple linear regression model by support vector regression (v-SVR). These steps resulted in increased robustness of the algorithm and more accurate calculation of the relative proportion of different immune cells in cancer tissues. This method was applied to the analysis of infiltrating immune cells based on 41 pairs of colorectal cancer tissues and normal solid tissues. Specifically, we compared the relative fractions of six types of immune cells in colorectal cancer tissues to those found in normal solid tissue samples. We found that tumor tissues contained a higher proportion of CD8 T cells and neutrophils, while B cells and monocytes were relatively low. Our pipeline for calculating immune cell proportion using gene expression profile data can be freely accessed from GitHub at https://github.com/gutmicrobes/EICS.git.

Keywords: Gene expression profile; support vector regression; immune cells; colorectal cancer.

* Corresponding author.

1. Introduction

Histopathological analysis shows that solid tumors are infiltrated by both innate and adaptive immune cells.[1] In the tumor microenvironment (TME) of different solid tumors, and even in different samples of the same type of solid tumor, the density, location, and type of immune cells are heterogeneous and have different effects on the occurrence and development of cancer.[2] Under these circumstances, clinical prognosis depends on accurately characterizing the density of tumor-infiltrating immune cells.[3] Results indicate that the type, density, and location of immune cells in tumor samples are good prognostic biomarkers for colorectal cancer.[4,5] In addition, Chen *et al.*[6,7] have shown that higher numbers of cytotoxic T cell infiltrates in tumor tissues are associated with better prognosis in patients. Thus, characterizing the proportion of immune cells in the TME may be a biomarker for predicting tumorigenesis, which is instructive for the development of new therapeutic targets and strategies.

Till now, many methods of recognizing and counting immune cells have emerged. The traditional methods used to identify immune cells *in vivo*, such as flow cytometry,[8] immunocytochemistry,[9] which rely on a limited library of phenotypic markers, and tissue depolymerization prior to flow cytometry can result in cell loss or damage, which can lead to errors in immune cell counting.[10]

In recent years, algorithms using bioinformatics methods to estimate the relative proportion of tumor-infiltrating immune cells have sprung up. In 2015, Newman *et al.*[10] developed an immune cell estimating algorithm, called CIBERSORT, which estimates the number of each immune cell type (22 in total), using the Support Vector Regression (v-SVR).[11] method based on DNA microarray data. However, this method results in collinearity and nonbiological association owing to the inclusion of too many variables, resulting in incorrect counting of some types of immune cells.[12] In 2016, Li *et al.*[13] developed TIMER. This algorithm uses constrained least squares fitting to infer the abundance of six types of immune cells based on the RNA-seq gene expression profile, solving the multicollinearity problem caused by the high correlation between immune cells subpopulations. In 2017, Racle *et al.*[14] proposed EPIC, an algorithm whereby the proportion of six immune cells was inferred using least squares regression, but this approach was susceptible to noise in the sample, resulting in less robust estimates. However, on the basis of this method, we introduced v-SVR to estimate the proportion of six kinds of tumor-infiltrating immune cells, the result of which increases the robustness of the algorithm and reduces the influence of collinearity on results. In this paper, our new method, termed Estimating Immune Cells by v-SVR (EICS), is presented.

2. Methods and Material

We first standardized the gene expression profile data (RNA-Seq) by Transcripts Per Million (TPM) and then combined the results with a signature matrix. Finally, we

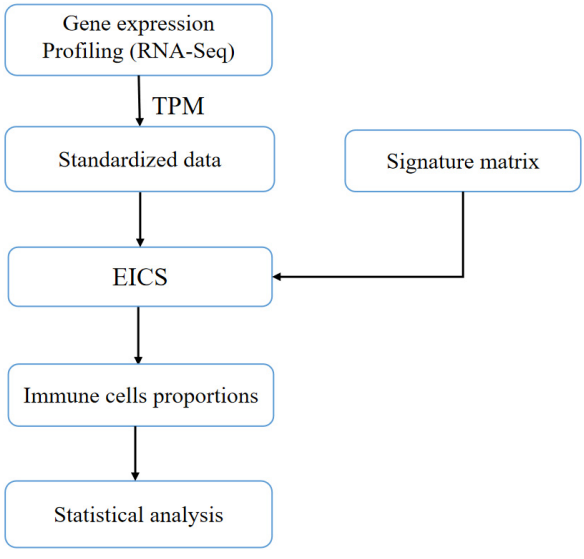


Fig. 1. Flow chart showing the differential analysis of immune cell abundance in normal and infiltrated tumor tissue by EICS.

applied EICS to solve immune cell count. In brief, we downloaded gene expression data from 41 pairs of colorectal cancer patients' tumor samples and normal solid tissue samples. The dataset was from the TCGA database (TCGA-COAD). It showed the flowchart of our approach for statistical analysis of immune cell abundance in normal and infiltrated tumor tissues (see Fig. 1).

2.1. *Building a simulated dataset with noise*

Simulation data were introduced in order to evaluate EICS. Cancer cell heterogeneity is well known, and many other unknown cells can be found in the tumor tissue of cancer patients. This results in the increased expression level of nonimmune cells in the tissue expression profile. Therefore, in order to make the simulation data more practical, noise was added.

Simulation data generation method: Randomly generate a vector p with a sum of 1; that is, $\sum_{i=1}^{6} p_i = 1$, according to the gene expression profile of immune cells (49,898 genes). Next, mix six kinds of immune cells in proportion p to obtain a noise-free simulation sample, as $x_{49898\times1} = S_{49898\times6} \times p_{6\times1}$. $x_{49898\times1}$ represents the gene expression level of 49,898 genes in a simulation sample, and $S_{49898\times6}$ represents the standard gene expression of 49,898 genes in these six immune cells, which is the signature matrix. This method is repeated to obtain 100 noise-free samples. In this paper, the noise simulation method of CIBERSORT[10] is used to add noise to the data, and noise obeys the distribution $2^{N(0,f\times\sigma)}$, where $f \in (0,1)$. The original matrix is X, and σ is assumed to be the standard deviation $\log_2 X$.

Simulation data can be freely accessed from GitHub at https://github.com/gutmicrobes/EICS.git.

2.2. *Two sets of actual data for method validation*

In order to better verify the accuracy and practicality of EICS, we use two sets of actual datasets as a validation tool. The first dataset is the gene expression profile data of blood samples of 20 adult subjects, which comes from the published findings of Newman *et al.*, who developed CIBERSORT. These authors measured the relative proportion of immune cells using flow cytometry. The second dataset was a mixture sample of gene expression profiles for 12 patients with colorectal cancer.[3] The dataset comes from the NCBI database (GSE64385), together with the relative proportion of immune cells measured by immunohistochemistry.

2.3. *A Deconvolution Model Based on v-SVR*

In this paper, the relative proportion of immune cells is estimated based on the gene expression profile of tumor tissue samples of cancer patients; that is, the relative proportion of different immune cell types is inferred from the gene expression profile of the samples. The gene expression profile of a sample can be expressed as a linear combination of standardized gene expression profiles of immune cells.[15]:

$$x_{n\times 1} = S_{n\times m} \times p_{m\times 1}, \tag{1}$$

where x is the expression profile of the individual n genes of the sample, which is used for deconvolution, the matrix S is the data of n gene expression profiles of m immune cells, p is the relative fraction of m immune cells in the specific sample.

The data type used here is RNA-seq, and in order to eliminate the influence of sequencing depth and gene length on the amount of gene expression, the amount of RNA-seq expression was calibrated using TPM.[16] The normalized formula is as follows:

$$\bar{x} = \bar{S} \times \bar{p}, \tag{2}$$

where $\bar{x}$ is the TPM standardized sample, and $\bar{S}$ is the reference gene expression profile.

In this paper, v-SVR is used to solve Eq. (2). v-SVR is a highly robust machine learning method for noise.[17,18] v-SVR uses SVM to fit the curve and perform regression analysis. v-SVR defines a hyperplane that captures as many data points as possible with given defined constraints. To reduce overfitting, we used the loss function that penalizes data points outside a certain error radius, and $\bar{p}$ is the hyperplane we require.

Since we used standardized gene expression data, in which $\bar{p}$ actually corresponds to the ratio of mRNA from each cell, rather than the cell ratio, the relative ratio of

each immune cell can be estimated as follows:

$$p_j = \alpha \cdot \frac{\bar{p}_j}{r_j}, \tag{3}$$

where r_j is the number of RNA nucleotides in type j cell (or equivalent to the total weight of mRNA in each cell type), and α is a normalization constant, which makes $\sum p_j = 1$.

For the standard reference gene expression profile S of formula (2), we used the data which comes from the published findings of Racle J and others, who developed EPIC.[14] Sixty-five marker genes closely related to six kinds of immune cells were selected from 49,898 genes in GSE64655, GSE60424, and GSE51984 in the NCBI database, and the standard reference gene expression profile matrix $S_{65\times6}$ was finally obtained. Six kinds of immune cells were B cells, CD4 T cells, CD8 T cells, monocytes, neutrophils, and NK cells.

In terms of formula (2), the relative proportion p of immune cells is obtained by calculating the coefficients of the multivariate linear regression model using the svm () function of the R language "e1071" package. The parameters in the model were, respectively, set to $v = 0.25, 0.5, 0.75$, and the regression coefficients of three models were obtained. Then the model with the smallest root mean square error was selected as the final model, as shown in Eq. (4):

$$\min \sqrt{[\bar{x}_i|_{i\in T} - (\bar{S} \times \bar{p})_i|_{i\in T}]^2}, \tag{4}$$

where T is the total number of samples.

3. Results

3.1. *Evaluating the correlation between the calculation results of EICS and preset ratios based on simulation data*

First, the EICS algorithm was used to calculate the relative proportions of the six immune cells in simulated data. A scatter plot was drawn based on the calculated result and the actual scale (see Fig. 2). The EICS prediction results and the simulated proportions p were clustered around the straight line $Y = x$, which shows that the algorithm has high accuracy.

3.2. *Comparing the accuracy of different methods on actual datasets by combining the results of flow cytometry and immunohistochemistry*

The relative proportion of immune cells in two groups of actual samples was calculated by EICS, and then the relative proportion of immune cells in two groups of data was compared with the result of flow cytometry and immunohistochemistry, respectively. The root mean square error (RMSE), Spearman and Pearson correlation coefficients are calculated, as shown in Tables 1 and 2.

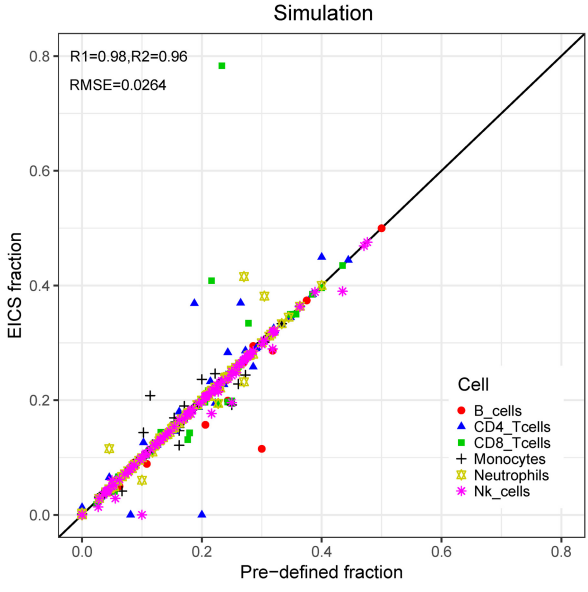


Fig. 2. Scatter plot of the relative proportion of immune cells in simulated data. The x-axis corresponds to the preset simulated proportion; the y-axis corresponds to the calculated result of EICS. The mean Spearman correlation coefficients of the six kinds of immune cells are $R_1 = 0.98$, and P values are $P_1 < 2.2e-16$. Pearson correlation coefficients are $R_1 = 0.96$, and P values are $P_2 < 2.2e-16$. The root mean square error is RMSE $= 0.0264$.

Table 1. EICS performance of actual dataset 1.

Cells	RMSE	Spearman	Pearson
B cells	0.0557	0.4300	0.4705
CD4 T cells	0.1324	0.5845	0.6399
CD8 T cells	0.1619	0.7774	0.7653
Monocytes	0.1903	0.7263	0.8037

The first set of actual data gives flow cytometry results of B cells, CD4 T cells, CD8 T cells, and monocytes. The second set of data gives immunohistochemical results for B cells, monocytes, NK cells, and T cells. In a comparison study, the results of EICS are highly correlated with the results of flow cytometry calculations.

Table 2. EICS performance of actual dataset 2.

Cells	RMSE	Spearman	Pearson
B cells	0.2342	0.9350	0.8791
T cells	0.5158	0.1333	0.7552
Monocytes	0.1060	0.9087	0.8836
NK cells	0.2366	0.9263	0.7639

In the first group of data, the correlation of B cells is less than 0.5, but the RMSE is 0.0557. This could be explained by the small proportion of B cells in the human body. For the second set of data, the Spearman correlation coefficient of B cells and NK cells was 0.9350 and 0.9263, respectively, and the RMSE was 0.2342 and 0.2366, respectively, while the Spearman of T cells was 0.1333, Pearson was 0.7552 and RMSE was 0.5158.

3.3. *Performance comparison with other methods in tumor samples*

Finally, the relative proportion of immune cells in the two groups of real samples were calculated by the algorithms CIBERSORT and EPIC and compared with the algorithm EICS; the results are shown in Table 3, Figs. 3 and 4. The Pearson and Spearman correlation coefficients of EICS in the two groups are higher than those in the algorithms CIBERSORT and EPIC. In that first set of data, the root mean-square error of EICS is minimal, while in the second set it is less than EPIC. Overall, the EICS method is more computationally accurate.

Table 3. Methods comparison of actual data.

	Actual dataset 1			Actual dataset 2		
	RMSE	Spearman	Pearson	RMSE	Spearman	Pearson
CIBERSORT	0.0784	0.5884	0.6302	0.1620	0.6891	0.8204
EPIC	0.0821	0.5864	0.6292	0.3335	0.6891	0.7625
EICS	0.0751	0.6297	0.6698	0.2731	0.7258	0.8204

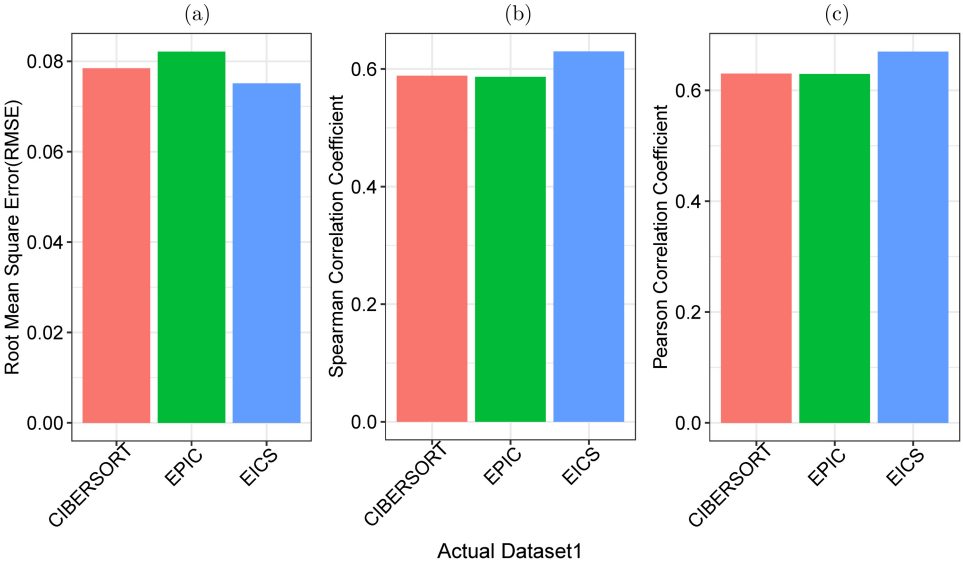


Fig. 3. Bar graph showing algorithms comparison of actual data1.

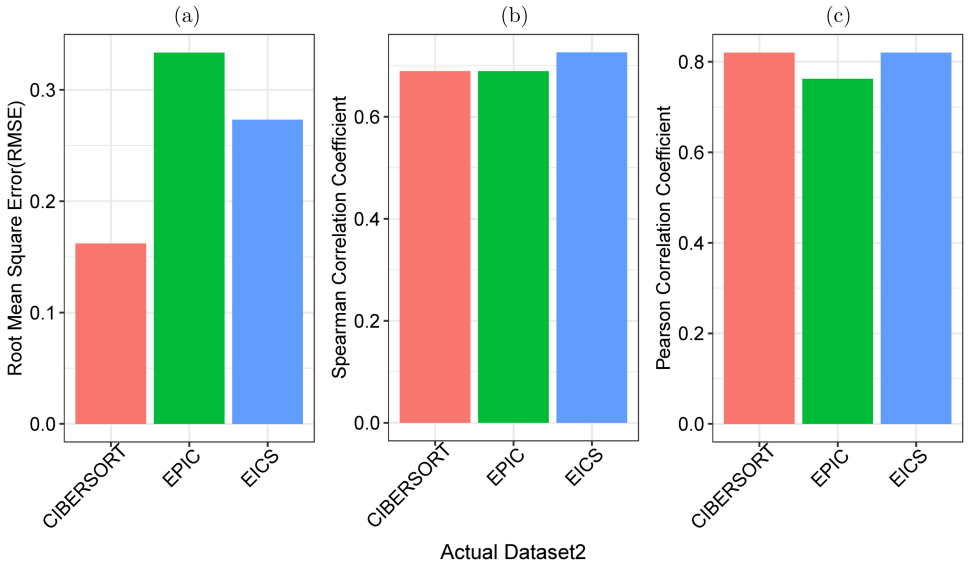


Fig. 4. Bar graph showing algorithms comparison of actual data2.

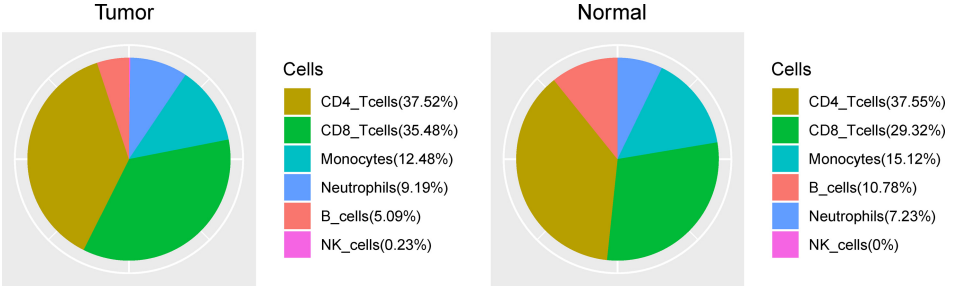
Then we use the Kappa function in the *R* language to calculate the condition number of the variable matrix to determine whether there is multicollinearity. The results are shown in Table 4. The Kappa values of CIBERSORT are all greater than 1000, and the Kappa values of EICS are less than 100 in the two sets of actual data. The degree of EICS multicollinearity is significantly lower than CIBERSORT.

3.4. *Applying EICS to CRC sample*

Colorectal cancer is the third largest cancer and the fourth leading cause of cancer-related death.[19] In order to study the association between immune cells and colorectal cancer, the EICS deconvolution algorithm described above was applied to the gene expression profile data of colorectal cancer patients. By calculating the relative proportion of six immune cells in all samples, we further found infiltrating immune cells with significant differences between tumor and normal solid tissue samples. The dataset comes from the TCGA database (https://portal.gdc.cancer.gov). We downloaded gene expression data for 41 pairs of tumor and normal solid tissue samples of colorectal cancer patients from the TCGA-COAD project.

Table 4. Multicollinearity performance comparison of actual data.

Kappa	Actual dataset 1	Actual dataset 2
CIBERSORT	47849.05	29455.34
EICS	17.5405	35.25597



(a) (b)

Fig. 5. Panel of infiltrating immune cells in tumor and normal solid tissue samples. Graph A shows the relative proportion of infiltrating immune cells in tumor samples, and graph B shows the relative proportion of infiltrating immune cells in normal solid tissue samples.

CRC sample data can be freely accessed from GitHub at https://github.com/gutmicrobes/EICS.git.

First, we analyzed the relative proportion of infiltrating immune cells in the tumor and normal solid tissue samples and plotted a pie chart for the relative proportion of six immune cells (see Fig. 5). The proportion of T cells (CD8 T cells and CD4 T cells) in the solid tissue samples of the two conditions is 73.00% and 66.87%, respectively, both of which occupy a large proportion of all six immune cells calculated. However, B cells are 5.09% and 10.78% compared to the proportion of T cells. It is well known that T cells and B cells are important components of lymphocytes and that T cells are mainly involved in cell-mediated immunity and distributed in the immune organs of the whole body. These are the most abundant and most complex type of lymphocytes, consistent with our results.

We analyzed infiltrating immune cells in tumor samples and normal solid tissue samples in order to ascertain any differences in relative proportion. The nonparametric Mann–Whitney–Wilcoxon test was performed in *R* software. The *P*-values were corrected by Benjamini–Hochberg (FDR). Immune cells with significant differences (Q-value < 0.05) were identified, including B cells, CD8 T cells, monocytes and neutrophils, as shown in Table 5.

Table 5. Immune cells with significant differences between tumor and normal solid tissue samples (Q-value < 0.05).

Cells	*P*-value	*Q*-value (adjusted.fdr)
Bcells	4.52E-10	2.71E-09
CD8 T cells	0.00230	0.0089
Monocytes	0.0302	0.0453
Neutrophils	0.0231	0.0453

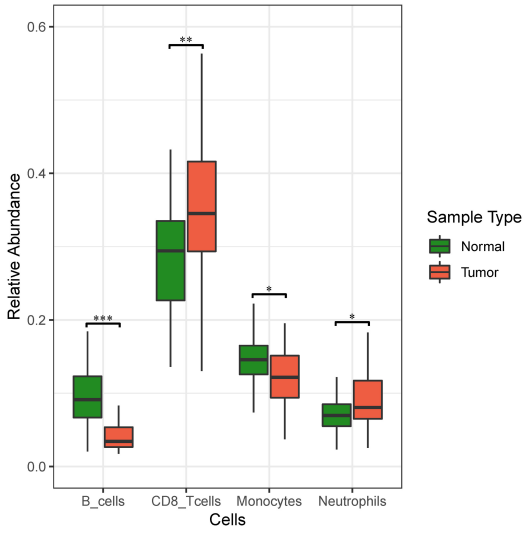


Fig. 6. Immune cells with significant differences between tumor and normal solid tissue samples. "*" means Q-value < 0.05; "**" means Q-value < 0.01 and "***" means Q-value < 0.001.

Based on the above results, we further drew the box diagram from which it can be deduced that the relative proportion of CD8 T cells and neutrophils in tumor tissue is higher, while the relative proportion of B cells and monocytes is lower, compared to normal solid tissue samples (see Fig. 6).

4. Discussion

In this study, we introduced the v-SVR model based on the EPIC method, and we developed a new method, termed EICS, to calculate the relative proportion of six kinds of immune cells in tumor samples.

In order to verify the accuracy of this method, we applied EICS to simulation data and actual data. On the whole, results show that the method can accurately calculate the relative proportion of six kinds of immune cells in cancer tissue.

We further tested the difference of the six kinds of immune cells between colorectal cancer and healthy samples by statistical methods and found four kinds of immune cells with significant difference: B cells, CD8 T cells, monocytes, and neutrophils. Compared with normal tissue samples, colorectal cancer samples contain a higher proportion of CD8 T cells and neutrophils, but normal solid tissue samples have a higher relative proportion of B cells and monocytes. A large body of literature has validated the correlation between immune cells and tumors.[4]

CD8 T cells are not a single cell population; instead, they can be divided into cytotoxic T cells (CTL) and inhibitory T cells[20] according to whether CD 28 is expressed. The former are called effector T cells, and the latter are called regulatory

T (Treg) cells.[21,22] Where Treg cells are involved, not only in the development of autoimmune diseases, but also in the prognosis of cancer patients, studies have shown an elevation of Treg cells in the tumor stroma of a variety of tumors, such as breast cancer,[23] hepatocellular carcinoma,[24] colon cancer.[20,25] and the like.

Many molecular mechanisms are involved in regulating the proliferation of T cells. Studies have shown that tumor antigens promote the expansion of natural Treg cells in tumors and that TGF-β expressed by tumor cells is also responsible for the expansion of Treg cells.[26] The main function of TGF-β is to maintain self-tolerance and inhibit immune response[20] which means that the increase of Treg cells can eliminate antitumor immunity *in vivo* and lead to poor prognosis. Thus, inhibition of Treg cell proliferation may increase survival rate of cancer patients.[27] Studies by Osamu Wakabayashi and others confirm that patients with higher numbers of CD8 T cells within cancer cells showed significantly shorter survival times compared to those with lower numbers of CD8 T cells within cancer cell nests.[28]

Regulatory T cells play a key role in the control of antitumor immune responses. Related studies have demonstrated that Treg cells in tumors of colorectal cancer patients are suppressor cells,[29] the increase of which is associated with immunosuppression and evasion in cancer patients. Treg cells prevent autoimmune diseases by inhibiting immune effector cells. The high frequency of Treg cells in peripheral blood, lymphoid tissue, and TME indicates the key role of CD8 T cells in tumors.[30]

Neutrophils are seen as the first responders of the innate immune system to extracellular pathogens.[31] When an infection or inflammation occurs in the human body, neutrophils will accumulate in the areas of inflammation and start phagocytosis to help the body fight off infection and disease. Therefore, during some infections and fever, as well as some diseases like leukemia, the number of neutrophils increases dramatically.[32] However, neutrophils not only inhibit cancer development, but Barbara and others have shown that tumor cells use neutrophils to enhance the ability of cancer to metastasize. Moreover, circulating tumor cells (CTCs) are precursors to several types of cancer metastasis which, as noted, is closely associated with the increase of neutrophils.[33] Based on this finding, a specific strategy for blocking metastasis of cancer cells can be further developed. In addition, Yanfang *et al.*[34] have found that neutrophil progenitor cells (NeP) are capable of differentiating into neutrophils with PD-L1 ligands on the surface, thereby inhibiting the activation of T cells and promoting tumor growth. The level of NeP cells was significantly elevated in the blood of tumor patients, but scarcely detectable in healthy blood, so NeP could be used as a marker for early cancer diagnosis. Mizuno *et al.*[35] also found a prominent role of neutrophils in infiltrating tumor tissues. Neutrophils, which are involved in the regulation of the innate and adaptive immune system, can respond to environmental signals, induce phagocytosis, release lyase, generate reactive oxygen species (ROS), and promote their growth in various types of cancer.[35]

Monocytes are the direct precursors of macrophages. After their recruitment into the tumor tissue, they can differentiate into tumor-associated macrophages, which

promote tumor genesis, local progression, and distant metastasis.[36] In Sec. 3, we saw that the proportion of monocytes in healthy samples is higher than that in cancer samples. We speculate that more monocytes differentiate into macrophages under the influence of the TME. In addition, studies by Shimabukuro-VA[37] and others confirm that the regulatory B-cell subsets of TME in tumor samples from patients with colorectal cancer is lower than that of peripheral blood with CRC.

In summary, different types of immune cells play important roles in the occurrence and development of tumors. Accordingly, a method for accurately characterizing the heterogeneous composition of immune cells from gene expression profiles of complex tumor tissue would allow, in turn, for more precise prognosis and diagnosis in the context of immunotherapy.

Acknowledgments

This work was supported by a grant from the National Natural Science Foundation of China (Grant Nos. 61873027 and 61370131). We thank Professor Fengzhu Sun at the University of Southern California and Li C. Xia at Stanford University.

Appendix A. Pipeline and Data

Our pipeline and data for calculating immune cell proportion using RNA sequences can be freely accessed from GitHub at https://github.com/gutmicrobes/EICS.git.

References

1. Gajewski TF, Schreiber H, Fu Y-X, Innate and adaptive immune cells in the tumor microenvironment, *Nat Immunol* **14**:1014, 2013.
2. Di Caro G, Castino GF, Bergomas F, Cortese N, Chiriva-Internati M, Grizzi F, Mantovani A, Marchesi F, Tertiary lymphoid tissue in the tumor microenvironment: From its occurrence to immunotherapeutic implications, *Int Rev Immunol* **34**:123–133, 2015.
3. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A, Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, *Genome Biol* **17**:218, 2016.
4. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoué F, Bruneval P, Cugnenc P-H, Trajanoski Z, Fridman W-H, Pagès F, Type, density, and location of immune cells within human colorectal tumors predict clinical outcome, *Science* **313**:1960, 2006.
5. Angell H, Galon J, From the immune contexture to the Immunoscore: The role of prognostic and predictive immune markers in cancer, *Current Opin Immunol* **25**:261–267, 2013.
6. Chen K, Zhou L, Xie H, Ahmed T-E, Feng X-W, Zheng S-S, Intratumoral regulatory T cells alone or in combination with cytotoxic T cells predict prognosis of hepatocellular carcinoma after resection, *Med Oncol* **29**:1817–1826, 2012.
7. Guidoboni M, Gafà R, Viel A, Doglioni C, Russo A, Santini A, Del Tin L, Macrì E, Lanza G, Boiocchi M, Dolcetti R, Microsatellite instability and high content of activated

cytotoxic lymphocytes identify colon cancer patients with a favorable prognosis, *Am J Pathol* **159**:297–304, 2001.
8. Deenitchina SS, Ando T, Okuda S, Kinukawa N, Hirakata H, Nagashima A, Fujishima M, Cellular immunity in hemodialysis patients: A quantitative analysis of immune cell subsets by flow cytometry, *Am J Nephrol* **15**:57–65, 1995.
9. Phillips T, Simmons P, Inzunza HD, Cogswell J, Novotny JJr., Taylor C, Zhang X, Development of an automated PD-L1 immunohistochemistry (IHC) assay for non-small cell lung cancer, *Appl Immunohistochem Mol Morphol.* **23**:541–549, 2015.
10. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA, Robust enumeration of cell subsets from tissue expression profiles, *Nat Meth* **12**:453–457, 2015.
11. Schölkopf B, Williamson RC, Smola AJ, Shawe-Taylor J, Platt JC, Support vector method for novelty detection, *Advances in Neural Information Processing Systems*, pp. 582–588, 2000.
12. Zheng S, Benchmarking: Contexts and details matter, *Genome Biol* **18**:129, 2017.
13. Li B, Severson E, Pignon J-C, Zhao H, Li, T, Novak J, Jiang P, Shen H, Aster JC, Rodig S, Signoretti S, Liu JS, Liu XS, Comprehensive analyses of tumor immunity: Implications for cancer immunotherapy, *Genome Biol* **17**:174, 2016.
14. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D, Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data, *eLife* **6**:e26476, 2017.
15. Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF, Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus, *PloS One* **4**:e6098, 2009.
16. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN, RNA-Seq gene expression estimation with read mapping uncertainty, *Bioinformatics* **26**:493–500, 2009.
17. Chang C-C, Lin C-J, LIBSVM: A library for support vector machines, *ACM Trans Intell Syst Technol* **2**:27, 2011.
18. Mangasarian OL, Musicant DR, Robust linear and support vector regression, *IEEE Trans Pattern Anal Mach Intell* **22**:950–955, 2000.
19. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray, F, Global patterns and trends in colorectal cancer incidence and mortality, *Gut* **66**:683–691, 201.
20. Sinicrope FA, Rego RL, Ansell SM, Knutson KL, Foster NR, Sargent DJ, Intraepithelial effector (CD3+)/regulatory (FoxP3+) T-cell ratio predicts a clinical outcome of human colon carcinoma, *Gastroenterology* **137**:1270–1279, 2009.
21. Filaci G, Suciu-Foca N, CD8+ T suppressor cells are back to the game: Are they players in autoimmunity? *Autoimmunity Rev* **1**:279–283, 2002.
22. Suciu-Foca N, Manavalan JS, Cortesini R, Generation and function of antigen-specific suppressor and regulatory T cells, *Transplant Immunol* **11**:235–244, 2003.
23. Bates GJ, Fox SB, Han C, Leek RD, Garcia JF, Harris AL, Banham AH, Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk of late relapse, *J Clin Oncol* **24**:5373–5380, 2006.
24. Gao Q, Qiu S-J, Fan J, Zhou J, Wang X-Y, Xiao Y-S, Xu Y, Li Y-W, Tang Z-Y, Intratumoral balance of regulatory and cytotoxic T cells is associated with prognosis of hepatocellular carcinoma after resection, *J Clin Oncol* **25**:2586–2593, 2007.
25. Salama P, Phillips M, Grieu F, Morris M, Zeps N, Joseph D, Platell C, Iacopetta, B, Tumor-infiltrating FOXP3+ T regulatory cells show strong prognostic significance in colorectal cancer, *J Clin Oncol* **27**:186–192, 2009.
26. Coe D, Addey C, White M, Simpson E, Dyson J, Chai J-G, The roles of antigen — specificity, responsiveness to transforming growth factor-β and antigen — presenting cell

subsets in tumour — induced expansion of regulatory T cells, *Immunology* **131**:556–569, 2010.
27. Wang Y, Ma Y, Fang Y, Wu S, Liu L, Fu D, Shen X, Regulatory T cell: A protection for tumour cells, *J Cellular Mol Med* **16**:425–436, 2012.
28. Wakabayashi O, Yamazaki K, Oizumi S, Hommura F, Kinoshita I, Ogura S, Dosaka A, Hirotoshi NM, CD4+ T cells in cancer stroma, not CD8+ T cells in cancer cell nests, are associated with favorable prognosis in human non - small cell lung cancers, *Cancer Sci* **94**:1003–1009, 2003.
29. Nishikawa H, Sakaguchi S, Regulatory T cells in tumor immunity, *Int J Cancer* **127**:759–767, 2010.
30. Johdi NA, Ait-Tahar K, Sagap I, Jamal R, Molecular signatures of human regulatory T cells in colorectal cancer and polyps, *Front. Immunol.* **8**:620, 2017.
31. Odobasic D, Kitching AR, Holdsworth SR, Neutrophil-mediated regulation of innate and adaptive immunity: The role of myeloperoxidase, *J Immunol Res* **2016**:234987, 2016.
32. Amulic B, Cazalet C, Hayes GL, Metzler KD, Zychlinsky A, Neutrophil function: From mechanisms to disease, *Annual Rev Immunol* **30**:459–489, 2012.
33. Szczerba BM, Castro-Giner F, Vetter M, Krol I, Gkountela S, Landin J, Scheidmann MC, Donato C, Scherrer R, Singer J, Neutrophils escort circulating tumour cells to enable cell cycle progression, *Nature* **566**:553, 2019.
34. Zhu YP, Padgett L, Dinh HQ, Marcovecchio P, Blatchley A, Wu R, Ehinger E, Kim C, Mikulski Z, Seumois G, Identification of an early unipotent neutrophil progenitor with pro-tumoral activity in mouse and human bone marrow, *Cell Rep* **24**:2329–2341, 2018.
35. Mizuno R, Kawada K, Itatani Y, Ogawa R, Kiyasu Y, Sakai Y, The role of tumor-associated neutrophils in colorectal cancer, *Int J Mol Sci* **20**:529, 2019.
36. Richards DM, Hettinger J, Feuerer M, Monocytes and macrophages in cancer: Development and functions, *Cancer Microenviron* **6**:179–191, 2013.
37. Shimabukuro-Vornhagen A, Schlößer HA, Gryschok L, Malcher J, Wennhold K, Garcia-Marquez M, Herbold T, Neuhaus LS, Becker HJ, Fiedler A, Characterization of tumor-associated B-cell subsets in patients with colorectal cancer, *Oncotarget* **5**:4651, 2014.