 doi:10.20944/preprints202011.0385.v1

Type of the paper: Article

mySORT: A web framework by using Deconvolution Approach to Estimating Immune Cell Composition from Complex Tissues

Shu-Hwa Chen[1]*, Bo-Yi Yu[2]*, Wen-Yu Kuo[3], Ya-Bo Lin[4], Sheng-Yao Su[5], I-Hsuan Lu[6], Chung-Yen Lin[7]§

1. Research Center of Cancer Translational Medicine, Taipei Medical University, 250 Wu-Xing Street, Taipei, TAIWAN; sophia0715@tmu.edu.tw
2. Research Center for Advanced Science and Technology, the University of Tokyo, 4-6-1 Komaba, Meguro-Ku, Tokyo 153-8904, JAPAN; boy65061120@iis.sinica.edu.tw
3. Institute of Information Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan; wenyu125@gmail.com
4. Institute of Information Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan; lightcoker@gmail.com
5. Institute of Information Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan; daniel0523@gmail.com
6. Institute of Information Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan; lindalu@iis.sinica.edu.tw
7. Institute of Information Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan; cylin@iis.sinica.edu.tw

*These authors contributed equally to this work

§Corresponding author

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

Abstract

Cancer immunotherapy reaches a remarkable achievement in various cancer types and brings new possibilities to improve cancer patients' long-term survival. However, outcomes vary from case to case, and the present protocol benefits a small fraction of patients. One notable factor is the tumor microenvironment, especially the immune cell components, that may reflect the immune response's status quo on site. Thus, understanding the content of infiltrating immune cells in tumors is not only for research interesting but also a crucial subject toward precision medicine.

We implement an algorithm for resolving relative proportions of twenty-one immune cell subclasses from a human tissue profiled transcriptome by microarray technology to reach the goal above. By selecting gene features and then adopting **ν**-Support Vector Regression, we can construct a deconvolution model and resolve the immune cell context. The excellent consistency between the estimated values and the correct immune-cell composition further demonstrates this approach provides a more natural alternative to revealing samples' immune cell content and reliable results like recent single-cell technologies.

Based on this algorithm, the web-based deconvolution tool implemented named mySORT provides a user-friendly interface for estimating the immune cell content by uploading gene expression profiling. We also present comprehensive visualization 2D/3D plots in mySORT so that users can easily make a comparison between different samples. Finally, we synthesized pseudo-bulk expression data from single-cell transcriptomic datasets of 17 melanoma and 16 head and neck cancer patients. The

 doi:10.20944/preprints202011.0385.v1

deconvolution results of microarray-based data in the previous study and synthetic pseudo-bulk data all proved the excellent performance of mySORT. We believe that mySORT can help researchers in all fields easily understand complex immune microenvironment. The website of mySORT is freely accessible on https://symbiosis.iis.sinica.edu.tw/mySORT/.

Keywords

Cancer; Immunotherapy; Deconvolution; Alpha diversity; Beta diversity; Precision medicine; Microenvironment; Single-cell RNA sequencing

 doi:10.20944/preprints202011.0385.v1

Introduction

Cancer is a disease caused by a malicious cell population that can divide unlimitedly and further metastasizes to other remote sites, thus occupying the healthy cells' space and other resources. The immune system can detect not only invasive antigens but also abnormal cells in our bodies. Unfortunately, cancer's presence proves it may find ways to escape from the surveillance of the immune system. A recent breakthrough in finding immune blockage/checkpoint molecules PD-1 or CTLA-4 [1,2] leads a new paradigm of cancer therapy targeting checkpoint inhibitors such as PD-1/PD-L1, which successfully re-activate the immune system [3-5]. However, checkpoint inhibitors' objective response rate varies, and adverse side-effects have been observed [6]. Consequently, finding out possible factors causing the variation of therapeutic outcomes among patients is critical in modifying present cancer immunotherapy to better performance.

Immune cell composition in tumors is proposed to explain the patients' and cancer types' diverse responses [7-9]. To reveal the immune escape mechanism driven by the tumor, a robust approach for estimating immune cell content in the tumor microenvironment is crucial. Traditional methods such as flow cytometry and immunocytochemistry can provide a small range of known biomarkers. These methods are also applied to a small fraction of biopsy and are difficult to scale-up to resolve all interested immune cell types.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

Recent high throughput technologies such as microarray and next-generation sequencing (NGS) have revolutionized the way of gene expression profiling, by which methods to estimate immune cell composition were conceived [10-13]. For example, several statistical approaches on the microarray, such as quadratic programming [14], Digital Sorting Algorithm [15], semi-supervised non-negative matrix factorization [16] were proposed to deconvolute the immune cell composition, *viz*, resolving cellular components from a measure of pooling values [12,17]. Most of these methods focused on a small spectrum of cell types. Newman et al. adopted the novel strategy and implemented their method to a web service CIBERSORT [13]. By benchmarking on the cellular composition result to the ground true (cell fractions, typing by flow cytometry), CIBERSORT is recognized as a superior method [18]. The performance detailed into the cell types varies. Using the CIBERSORT defined scenario, we revisited the dataset, exclude potentially contaminant datasets and weak supporting cell subtypes, optimize the signature gene set, and propose a **ν**-support vector regression method deconvolute the immune cell composition [19]. This method, mySORT, outperformed CIBERSORT in the benchmark testing of microarray datasets.

In this study, we implement the web application of mySORT with an interactive user interface to process uploaded transcriptome profiles. MySORT resolves the relative proportion of twenty-one types of immune cells. Results and statistical analyses (clustering, alpha- and beta-diversity) are presented in a graph-rich output. Furthermore, single-cell RNA sequencing is a novel immerging method to analyze the tumor microenvironment's complexity [20-23]. We use single-cell RNA

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

sequencing data to validate the consistency between these two measures.

Results & discussion

Usage of mySORT website

Users can upload a text or CSV file containing a gene expression matrix of single or multiple samples. The matrix should contain gene symbols as rows and sample names as columns. If the submitted expressed data is not normalized previously (user-defined), mySORT will perform log-transformed on it. Those people who wanted to know the algorithm of mySORT with pseudocodes in detail, please find the information in our previous publication [19]. To quickly understand the operation of mySORT, we prepare the demo, which includes several expression data from different kinds of cancer tissues. Users can categorize these expression datasets as several groups for deep analysis. Simultaneously, users can also visualize both alpha and beta diversity in low dimension 2D or 3D plot. The calculation and visualization processed above were conducted by R packages *Vegan* [24], *Phyloseq* [25], and *Plotly* (https://plot.ly).

The output of mySORT includes four parts shown in **Figure 1**:

1. A proportional table and a stacked bar chart of 21 immune cell types,
2. Hierarchical clustering of samples based on the immune cell composition
3. Visualization plots based on alpha and beta diversity
4. Several text files in Download Area for original submission, result, alpha/ beta diversity, signature matrix, and a log file (please see the result page on the website).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

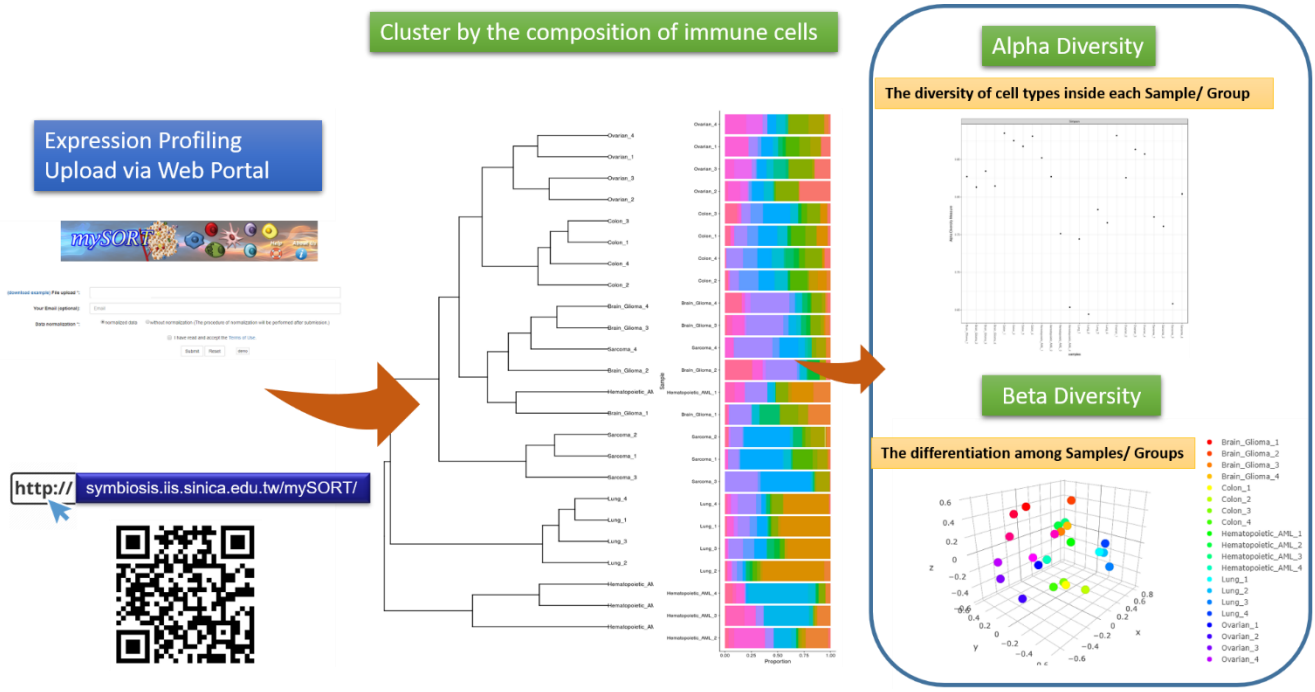


Fig. 1. The workflow of mySORT. mySORT is a web-based tool designed for performing deconvolution analysis of input samples (blue rectangle). After running the analytical process, mySORT will present a relative immune cell composition table and several visualized figures, including hierarchical clustering, alpha diversity, and beta diversity analyses (green rectangle).

Visualization plots of alpha diversity and beta diversity

We further implement alpha and beta diversity plots to visualize the single-cell experiment data's overall distribution and profiling. Here we adopt Simpson's diversity index for alpha diversity to indicate the richness and evenness of immune-cell species and the heterogeneity of a sample [26]. For example, samples with more infiltrating immune cell types or samples with the same cell component but with a more even cell number distribution of each type tend to have higher alpha diversity measurements. The non-metric multidimensional scaling (NMDS) plot is applied to describe the beta diversity; here, the difference of immune cell composition among samples [27]. The shorter distance of two samples on the NMDS plot indicates the overall similarity is relatively higher between this

 doi:10.20944/preprints202011.0385.v1

sample pair. A simple comparison between the web application of mySORT and CIBERSORT is shown in Table 1. The web application of mySORT and CIBERSORT are both in the easy-to-use web interface. However, mySORT provides a more friendly and comprehensive data visualization that allows a glance at complex data and flexible options on deep analyses. The performance of mySORT is compatible or outperforms current state-of-the-art deconvolution methods[19] regarding the performance of RNA-Seq deconvolution. In this new updated version, we expand mySORT deconvolution model to analyze single cell RNA-Seq using blood biopsies as described in the next section.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

Table 1. Comparison of mySORT and CIBERSORT with functions in the web interface.

The table compares the web application of mySORT to CIBERSORT by several functions. mySORT demonstrates the advantage of data visualization functions over CIBERSORT.

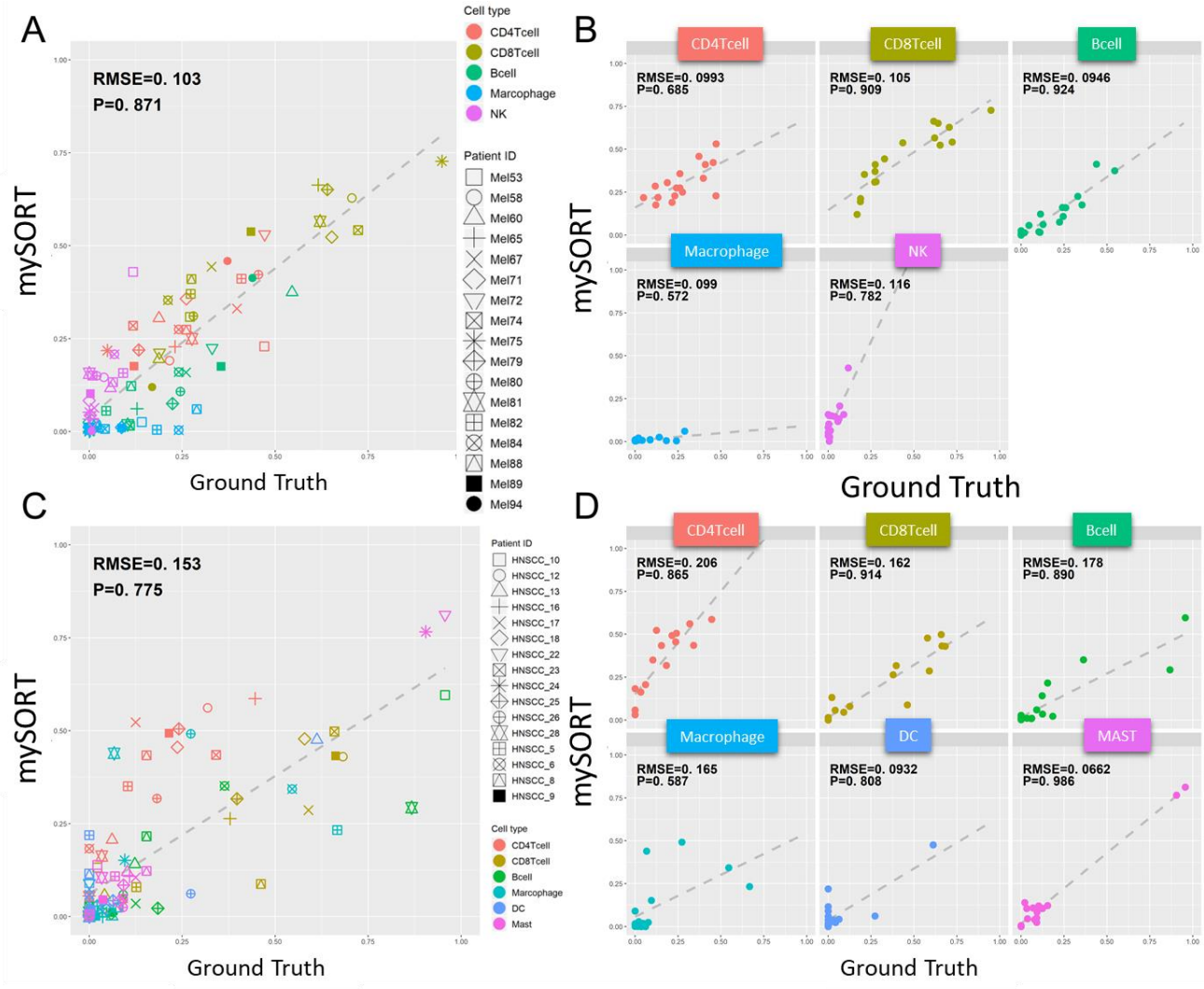
	mySORT	CIBERSORT
Registration	Not required	Required
Custom signature matrix	Not allowed	Allowed
Immune cell composition		
The relative proportion in table	Yes	Yes
Stacked bar chart	Yes	Yes
Multiple data columns	Yes	Yes
Comparison among data		
Hierarchical clustering	Yes	No
Cell diversity within a sample (alpha diversity plot)	Yes	No
Cell diversity among samples (beta diversity plot)	Yes	No
Data export	Yes	Yes

Validation of mySORT performance by real single-cell datasets

Blood biopsies previously benchmarked the performance of mySORT from 20 adults, in which nine immune cell types were identified using flow cytometry analysis, and it demonstrated that the present computations performed better than the current state-of-the-art deconvolution method CIBERSORT [14]. In this study, we further used the cutting-edge technology single-cell RNA sequencing data to validate the performance of mySORT. Here, we collected two public single-cell RNA sequencing datasets of tumor samples from 17 melanoma patients and 16 head and neck cancer patients. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

synthetic pseudo-bulk data of tumor samples were used to estimate the relative proportion of immune cell types. Finally, we found that the prediction of mySORT had a good correlation with the ground truth in both datasets when all immune cell types were considered (Fig. 2A and 2C). If we separated the outcome by cell types, we could also observe good correlation in almost all cell types except for macrophages (Fig. 2B and 2D). The lower accuracy of macrophages possibly results from the relatively lower proportion of macrophage-based datasets in our signature matrix. However, the overall performance of mySORT still presents better consistency with single-cell data and seems to be not affected by this phenomenon seriously.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

Fig. 2. Correlation of the predicted values from mySORT and the ground truth. (A) The scatter plot of the predicted proportion of mySORT and the ground truth from the melanoma dataset. The x-axis indicates the value of the ground truth, and the y-axis indicates the predicted value of mySORT. Shapes represent the different patients, and colors represent the different immune cell types. (B) The similar scatter plot as (A) but plotted based on each cell type only. (C)(D) The scatter plots from the HNSCC dataset.

Materials and Methods

Construction of synthetic pseudo-bulk gene expression data

The single-cell RNA sequencing data of melanoma patients and head and neck cancer patients were downloaded from NCBI GEO Accession GSE72056 and GSE103322, respectively [22,23]. The Smart-seq2 protocol was conducted for both single-cell datasets. Following the instruction of the original publication[22], we kept cells with good quality by the criteria of at least 1,700 expressed genes for melanoma samples and 2,000 expressed genes for head and neck squamous cell carcinoma (HNSCC) samples. An adequate quantity of housekeeping gene expression was also confirmed in every qualified cell. Finally, 4,640 cells of the melanoma dataset and 5,901 cells of the HNSCC dataset passed the quality control. We then constructed the pseudo-bulk gene expression profiles by averaging all the cells' gene expression profiles in each patient.

Calculation of the ground truth of immune cell composition

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

Only four immune cell types, B cells, T cells, NK cells, and macrophages, were initially defined in the melanoma dataset. However, we divided the T cell cluster into CD4 T cells and CD8 T cells, according to *CD4*, *CD8A*, and *CD8B* gene expression of each cell. T cells with *CD4* expression and no *CD8A* and *CD8B* expression were assigned to the CD4-T-cell cluster. On the other hand, T cells with *CD8A* or *CD8B* expression, and no *CD4* expression were classified as CD8 T cells. Besides, we discarded two melanoma samples due to the extremely low quantity of immune cells, so the remaining 17 samples were used for the downstream analysis. Finally, the ground truth of the relative immune cell proportion was calculated based on the authors' cell identity in each sample.

Similarly, the strategy described above was also applied for the HNSCC dataset. Still, six immune cell types, including B cells, CD4 T cells, CD8 T cells, macrophages, dendritic cells, and mast cells, as well as 16 qualified HNSCC samples, were used.

Comparison of estimated and true immune cell composition

The single-cell data and the output of mySORT only share several immune cell types, so we rescale the sum of both the ground-truth value and the predicted value to 1 as the total value for comparison. The Pearson correlation coefficient and root-mean-square were then used to measure the correlation and difference between the estimation of immune cell content and the ground truth.

System Implementation

For intuitive user experience for easy understanding, we build mySORT by composed of LAMP

 doi:10.20944/preprints202011.0385.v1

system architecture (Linux Ubuntu 16.04, Apache 2.04, MySQL 5.7, PHP 5.1) with Bootstrap 3 CSS framework (http://getbootstrap.com/), jQuery1.11.1, and jQuery Validation v1.17. Furthermore, the core of the analysis process is implemented in R (3.4.2). mySORT runs as a virtual machine (CPUs of 2.27 GHz, sixteens' cores, 32 GB RAM, and 1 TB storage) on the Institute of Information Science's cloud infrastructure, Academia Sinica, Taiwan.

Conclusion

The breakthrough in immunotherapy leads to novel anti-cancer drugs/therapies. Undoubtedly, there will be more and newer strategies for overcoming cancer cells' immune suppressive ability in the future. Correlation between the cellular composition of cancer and the drug response suggests the immune cell population's heterogeneity be a critical issue in the clinical practice. Thus, we built mySORT into a user-friendly web framework. We added two cell population diversity measurements to help biomedical researchers understand their samples' tumor microenvironment with comprehensive plots and charts.

The performance of cell deconvolution methods like mySORT largely depends on the quality and coverage of transcriptome data of the cell population, the feature selection strategy, and the implemented model's power. Although the accuracy of mySORT outperformed other concurrent methods on the test datasets from the microarray experiment, the predicting power in some cell types is barely satisfactory. Recent advances in single-cell RNA sequencing methodology provides various cell profiling data in the public depository. It is worthy of revising our methods to adopt the new,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

massive datasets and apply deep learning models to resolve the cell component deconvolution question. We believe that combining these new tools to achieve the concept of precision medicine will serve as a critical point of improving cancer treatment.

Availability and requirements

Project name: mySORT

Project home page: http://symbiosis.iis.sinica.edu.tw/mySORT for academic use

List of abbreviations

CTLA4: cytotoxic T-lymphocyte-associated protein 4

DEG: differentially expressed gene

HNSCC: head and neck squamous cell carcinoma

NMDS: non-metric multidimensional scaling

PD-1: programmed cell death protein 1

PD-L1: programmed death-ligand 1

RMSE: root-mean-square

SVR: Support vector regression

Declarations

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0385.v1

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Availability of data and material

In this study, we used the Melanoma dataset (NCBI GSE72056) and head and neck cancer dataset (NCBI, GSE103322).

Competing interests

The authors declare that they have no competing interests.

Funding

The authors thank the Ministry of Science and Technology (MOST), Taiwan, for financially supporting this research and publication through MOST107-2321-B-002-057 and MOST107-2314-B-001-002 to CYL.

Authors' contributions

SHC, WYK, and CYL planned and implemented the algorithm to decipher infiltrated immune cells' composition. SHC, BYY, WYK, and CYL, composed the whole infrastructure, conducted the experiments, and drafted the manuscript together with HSL. YBL, SYS, and IHL worked on constructing workflow and implementation web platform for data visualization and in-deep analysis. All of the authors had read and approved the final manuscript.

Acknowledgments

We thank Dr. Su-Fang Lin from National Health Research Institutes, Taiwan, for her suggestions on the methodology, and anonymous reviewers for the critical reading and advice throughout the article.

References

1. Buchbinder, E.I.; Desai, A. CTLA-4 and PD-1 Pathways: Similarities, Differences, and Implications of Their Inhibition. *Am J Clin Oncol* **2016**, *39*, 98-106, doi:10.1097/COC.0000000000000239.
2. Seidel, J.A.; Otsuka, A.; Kabashima, K. Anti-PD-1 and Anti-CTLA-4 Therapies in Cancer: Mechanisms of Action, Efficacy, and Limitations. *Front Oncol* **2018**, *8*, 86, doi:10.3389/fonc.2018.00086.
3. Ribas, A.; Wolchok, J.D. Cancer immunotherapy using checkpoint blockade. *Science* **2018**, *359*, 1350-1355, doi:10.1126/science.aar4060.
4. Sanmamed, M.F.; Chen, L. A Paradigm Shift in Cancer Immunotherapy: From Enhancement to Normalization. *Cell* **2018**, *175*, 313-326, doi:10.1016/j.cell.2018.09.035.
5. Sharpe, A.H.; Pauken, K.E. The diverse functions of the PD1 inhibitory pathway. *Nat Rev Immunol* **2018**, *18*, 153-167, doi:10.1038/nri.2017.108.
6. Emens, L.A.; Ascierto, P.A.; Darcy, P.K.; Demaria, S.; Eggermont, A.M.M.; Redmond, W.L.; Seliger, B.; Marincola, F.M. Cancer immunotherapy: Opportunities and challenges in the rapidly evolving clinical landscape. *Eur J Cancer* **2017**, *81*, 116-129, doi:10.1016/j.ejca.2017.01.035.
7. Binnewies, M.; Roberts, E.W.; Kersten, K.; Chan, V.; Fearon, DF; Merad, M.; Coussens, L.M.; Gabrilovich, D.I.; Ostrand-Rosenberg, S.; Hedrick, C.C., et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med* **2018**, *24*, 541-550, doi:10.1038/s41591-018-0014-x.
8. Daud, AI; Loo, K.; Pauli, M.L.; Sanchez-Rodriguez, R.; Sandoval, P.M.; Taravati, K.; Tsai, K.; Nosrati, A.; Nardo, L.; Alvarado, M.D., et al. tumor immune profiling predicts response to anti-PD-1 therapy in human melanoma. *J Clin Invest* **2016**, *126*, 3447-3452, doi:10.1172/JCI87324.
9. Espinosa, E.; Marquez-Rodas, I.; Soria, A.; Berrocal, A.; Manzano, J.L.; Gonzalez-Cao, M.; Martin-Algarra, S.; Spanish Melanoma, G. Predictive factors of response to immunotherapy-a review from the Spanish Melanoma Group (GEM). *Ann Transl Med* **2017**, *5*, 389, doi:10.21037/atm.2017.08.10.
10. Abbas, A.R.; Wolslegel, K.; Seshasayee, D.; Modrusan, Z.; Clark, H.F. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One* **2009**, *4*, e6098, doi:10.1371/journal.pone.0006098.
11. Clarke, J.; Seo, P.; Clarke, B. Statistical expression deconvolution from mixed tissue samples. *Bioinformatics* **2010**, *26*, 1043-1049, doi:10.1093/bioinformatics/btq097.
12. Liebner, D.A.; Huang, K.; Parvin, J.D. MMAD: microarray microdissection with analysis of differences is a computational tool for deconvoluting cell type-specific contributions from tissue samples. *Bioinformatics* **2014**, *30*, 682-689, doi:10.1093/bioinformatics/btt566.
13. Newman, A.M.; Liu, C.L.; Green, MR; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, AA Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **2015**, *12*, 453-457, doi:10.1038/nmeth.3337.
14. Gong, T.; Hartmann, N.; Kohane, I.S.; Brinkmann, V.; Staedtler, F.; Letzkus, M.; Bongiovanni, S.; Szustakowski, J.D. Optimal deconvolution of transcriptional profiling data using quadratic programming with application to complex clinical blood samples. *PLoS One* **2011**, *6*, e27156, doi:10.1371/journal.pone.0027156.
15. Zhong, Y.; Wan, Y.W.; Pang, K.; Chow, L.M.; Liu, Z. Digital sorting of complex tissues for cell type-specific gene expression profiles. *BMC Bioinformatics* **2013**, *14*, 89, doi:10.1186/1471-2105-14-89.
16. Gaujoux, R.; Seoighe, C. Semi-supervised Nonnegative Matrix Factorization for gene expression deconvolution: a case study. *Infect Genet Evol* **2012**, *12*, 913-921, doi:10.1016/j.meegid.2011.08.014.
17. Qiao, W.; Quon, G.; Csaszar, E.; Yu, M.; Morris, Q.; Zandstra, P.W. PERT: a method for expression deconvolution of human blood samples from varied microenvironmental and developmental conditions. *PLoS Comput Biol* **2012**, *8*, e1002838, doi:10.1371/journal.pcbi.1002838.
18. Chen, B.; Khodadoust, M.S.; Liu, C.L.; Newman, A.M.; Alizadeh, AA Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol* **2018**, *1711*, 243-259, doi:10.1007/978-1-4939-7493-1_12.
19. Chen, S.H.; Kuo, W.Y.; Su, S.Y.; Chung, W.C.; Ho, JM; Lu, HH; Lin, C.Y. A gene profiling deconvolution approach to estimating immune cell composition from complex tissues. *BMC Bioinformatics* **2018**, *19*, 154, doi:10.1186/s12859-018-2069-6.
20. Azizi, E.; Carr, A.J.; Plitas, G.; Cornish, A.E.; Konopacki, C.; Prabhakaran, S.; Nainys, J.; Wu, K.; Kiseliovas, V.; Setty, M., et al. Single-Cell Map of Diverse

Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* **2018**, *174*, 1293-1308 e1236, doi:10.1016/j.cell.2018.05.060.

21. Lambrechts, D.; Wauters, E.; Boeckx, B.; Aibar, S.; Nittner, D.; Burton, O.; Bassez, A.; Decaluwe, H.; Pircher, A.; Van den Eynde, K., et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med* **2018**, *24*, 1277-1289, doi:10.1038/s41591-018-0096-5.

22. Puram, SV; Tirosh, I.; Parikh, A.S.; Patel, A.P.; Yizhak, K.; Gillespie, S.; Rodman, C.; Luo, C.L.; Mroz, E.A.; Emerick, K.S., et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **2017**, *171*, 1611-1624 e1624, doi:10.1016/j.cell.2017.10.044.

23. Tirosh, I.; Izar, B.; Prakadan, S.M.; Wadsworth, M.H., 2nd; Treacy, D.; Trombetta, J.J.; Rotem, A.; Rodman, C.; Lian, C.; Murphy, G., et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **2016**, *352*, 189-196, doi:10.1126/science.aad0501.

24. Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **2003**, *14.6*, 927-930.

25. McMurdie, P.J.; Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **2013**, *8*, e61217, doi:10.1371/journal.pone.0061217.

26. Simpson, E.H. Measurement of Diversity. *Nature* **1949**, *163*, 688-688, doi:DOI 10.1038/163688a0.

27. Bray, J.R.; Curtis, J.T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* **1957**, *27*, 326-349.