Minireview

A Comprehensive Overview of RNA Deconvolution Methods and Their Application

Yebin Im[1] and Yongsoo Kim[2,*]

[1]School of Biological Sciences, Seoul National University, Seoul 08826, Korea, [2]Department of Pathology, Cancer Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, 1081 HZ Amsterdam, The Netherlands
*Correspondence: yo.kim@amsterdamumc.nl
https://doi.org/10.14348/molcells.2023.2178
www.molcells.org

Tumors are surrounded by a variety of tumor microenvironmental cells. Profiling individual cells within the tumor tissues is crucial to characterize the tumor microenvironment and its therapeutic implications. Since single-cell technologies are still not cost-effective, scientists have developed many statistical deconvolution methods to delineate cellular characteristics from bulk transcriptome data. Here, we present an overview of 20 deconvolution techniques, including cutting-edge techniques recently established. We categorized deconvolution techniques by three primary criteria: characteristics of methodology, use of prior knowledge of cell types and outcome of the methods. We highlighted the advantage of the recent deconvolution tools that are based on probabilistic models. Moreover, we illustrated two scenarios of the common application of deconvolution methods to study tumor microenvironments. This comprehensive review will serve as a guideline for the researchers to select the appropriate method for their application of deconvolution.

Keywords: statistical deconvolution, tumor microenvironment

INTRODUCTION

Conventional bulk transcriptome analysis has made a significant contribution to our understanding of the molecular mechanisms behind complex biological phenomena, yet it has been unable to fully uncover the intrinsic heterogeneity of samples. Previous studies revealed that tumors are surrounded by collections of various microenvironmental cells, including endothelial, stromal and infiltrating immune cells, which mutually interact with malignant cells to regulate tumor progression and its therapeutic resistance (Baghban et al., 2020; Jin and Jin, 2020). Moreover, tumors consist of multiple subpopulations of malignant cells with different genotypic and phenotypic features (Dagogo-Jack and Shaw, 2018; Nguyen et al., 2016). However, bulk RNA-seq data measures accumulated gene expression levels of all cells in each sample, which makes it limited to studying cellular heterogeneity. For these reasons, technologies like LCM (laser capture microdissection) or FACS (fluorescence-activated cell sorting) have been developed to isolate and identify each single cell, further extending to single-cell RNA sequencing (scRNA-seq) (Lee et al., 2020). Although promising, single-cell technologies still face obstacles to retaining enough samples and determining proper markers for cell labeling due to their labor-intensiveness and high cost (Lähnemann et al., 2020). Furthermore, the tissue dissociation step in scRNA-seq enriches cells that are easily detachable, essentially introducing a bias in the composition of cells to be profiled (Denisenko et al., 2020).

Scientists thus developed various computational deconvolution methods to infer the abundance of different cell types from bulk RNA-seq data to make the most of pre-existing large cohort-based studies (e.g., The Cancer Genome At-

Table 1. Overview of the 20 deconvolution methods covered in this review

Method	Characteristics of methodology			Use of prior knowledge			Outcome		
	Algorithm	Supervised/ unsupervised	Linear/ log-linear	Use of marker gene expression profile (signature)	Use of single-cell RNA-seq data for signature	Account for gene expression variability	Cell type fractions	In silico purification	No. of cell types can be handled*
ABIS	Robust linear regression	Supervised	Linear	Yes	No	No	Yes	No	Flexible (29)
DSA	Regularized linear regression	Supervised	Linear	Yes	No	No	Yes	Yes (group-mode)	Flexible (6)
TIMER	Regularized linear regression (multivariate normal)	Supervised	Linear	Yes	No	No	Yes	Yes	Flexible (6)
csSAM	Linear regression	Supervised	Log-linear	No	No	No	No	Yes (group-mode)	Flexible (5)
MuSiC	Weighted non-negative least squares	Supervised	Linear	Yes	Yes	Cross-subject variance	Yes	No	Flexible (13)
DECODER	NMF + Regularized linear regression	Unsupervised	Log-linear	No	No	No	Yes	Yes (group-mode)	Flexible (8)
CIBERSORT	nu-SVR (linear)	Supervised	Linear	Yes	No	No	Yes	No	Flexible (22)
CIBERSORTx	nu-SVR (linear)	Supervised	Linear	Yes	Yes	No	Yes	Yes (high resolution)	Flexible (10)
Bseq-SC	CIBERSORT + csSAM	Supervised	Linear	Yes	Yes	No	Yes	No	Flexible (6)
quantTIseq	constrained least squares	Supervised	Linear	Yes	No	No	Yes	Yes (group-mode)	Flexible (10)
MCP-counter	Relative gene expression levels	Supervised	Log-linear	Yes	No	No	No; score	No	Flexible (10)
ESTIMATE	Gene set enrichment analysis (GSEA)	Supervised	Log-linear	Yes	No	No	No; score	No	2
xCell	GSEA	Supervised	Log-linear	Yes	No	No	No; score	No	Flexible (64)
CDSeq	Probabilistic model (multinomial)	Unsupervised	Linear	No	No	No	Yes	Yes (group-mode)	Flexible (22)
Demix	Probabilistic model (log-normal)	Semi-supervised	Log-linear	Yes	No	Yes	Yes	Yes	2
DemixT	Probabilistic model (log-normal)	Semi-supervised	Linear	Yes	No	Yes	Yes	Yes	3
EPIC	Least-square	Supervised	Linear	Yes	No	Yes	Yes	No	Flexible (8)
BLADE	Probabilistic model (log-normal)	Supervised	Linear	Yes	Yes	Yes	Yes	Yes (high resolution)	Flexible (20)
ISOpure	Probabilistic model (multinomial)	Semi-supervised	Linear	Yes	No	No	Yes	Yes (high resolution)	2
BayesPrism	Probabilistic model (multinomial)	Supervised	Linear	Yes	Yes	No	Yes	Yes (high resolution)	Flexible (10)

The deconvolution methods are categorized by characteristics of methodology, use of prior knowledge and outcome.

NMF, non-negative matrix factorization; nu-SVR, nu-support vector regression.

*The maximum number of cell types used in the original study. The deconvolution technique may be able to handle more cell types when they are classified as flexible.

las [TCGA] and International Cancer Genome Consortium) (Campbell et al., 2020). Beyond the cell type compositions, advanced techniques can infer cell-type-specific gene expression levels—often referred to as purification. Based on this advancement, recent studies revealed differential cellular states among the same cell type defined by their specific transcriptome profiles (Andrade Barbosa et al., 2021; Chu et al., 2022; Luca et al., 2021). Since so many deconvolution methods have been published, it is challenging for the users to select the most suitable method. Previous review articles have evaluated some deconvolution tools but only focused on benchmarking their performance using well-controlled gold standard data (Avila Cobos et al., 2020; Sturm et al., 2019). However, there have not been many reviews that focus on the theoretical background of the methods that can help users to understand the strengths and the weaknesses of each deconvolution method.

Here, we provide a comprehensive overview of 20 statistical deconvolution tools, including some recently established techniques. We organized the tools by their algorithms and delineated practical limitations due to their technical groundings. We focused on the recent deconvolution techniques that can predict not only the cellular composition but also the cell-type-specific gene expression profiles. The majority of these methods are based on probabilistic models, which benefit from their flexible nature. Furthermore, we will highlight the recent application to show how these techniques can de-lineate tumor ecosystems in large numbers of samples using commonly available bulk RNA-seq data. Finally, we will share some practical recommendations on how to apply deconvolution tools and perspectives on how deconvolution tools can contribute to tumor microenvironment studies.

OVERVIEW OF THE PUBLISHED DECONVOLUTION METHODS

We constructed an overview of 20 different deconvolution methods, including recent approaches categorized by the characteristics of methodology, use of prior knowledge for deconvolution and the outcome of the methods (Table 1). Among the 20 methods, nine deconvolution methods are based on linear approaches, which include robust linear regression (ABIS [Moncao et al., 2019]), regularized linear regression (DSA [Zhong et al., 2013], TIMER [Li et al., 2020], csSAM [Shen-Orr et al., 2010], MuSiC [Wang et al., 2019], quanTIseq [Fintello et al., 2019]), and non-negative matrix factorization (DECODER [Peng et al., 2019]). Three other methods, CIBERSORT (Newman et al., 2015), CIBERSORTx (Newman et al., 2019), and Bseq-SC (Baron et al., 2016), are based on support vector regression using a linear kernel, which makes them similar to linear regression methods. Other methods applied gene set enrichment approaches (e.g., MCP-counter [Becht et al., 2016]) or, more recently, probabilistic models (Fig. 1). Eighteen of the methods take
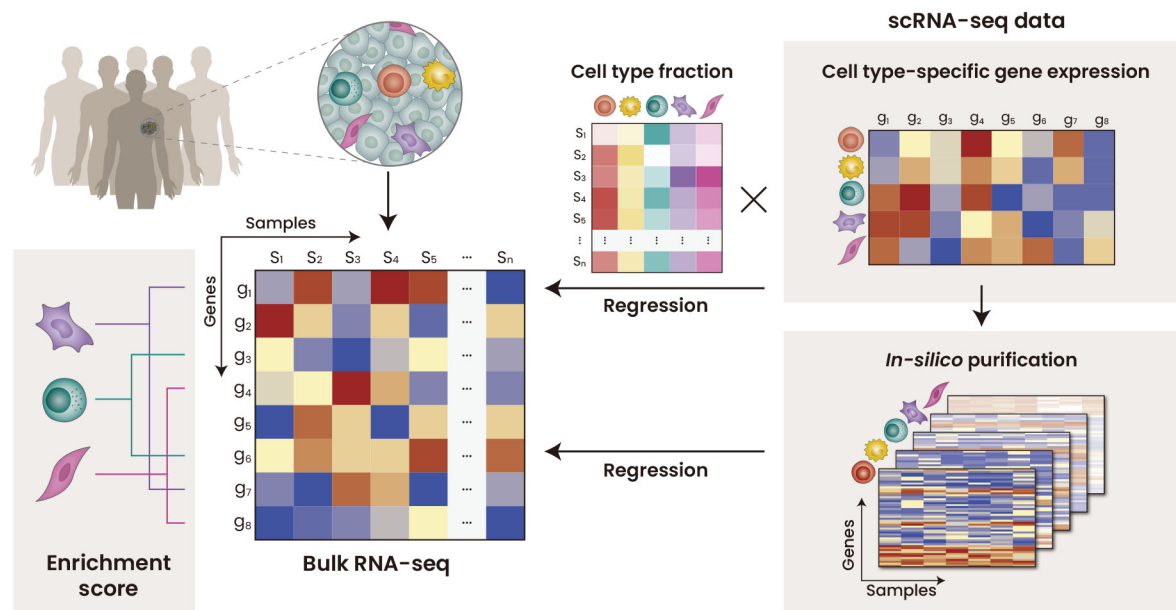

Fig. 1. Overview of common strategies for deconvolution. A diagram illustrates the basic concepts of the two main categories of deconvolution methodologies: Enrichment Analysis (left) and Regression Analysis (right). Enrichment methods calculate the enrichment scores of each cell type by combining the expression profiles of cell type markers from bulk RNA-seq data (left). However, due to variations in the set of marker genes, these scores can vary greatly in scale, making it impossible to convert them to cellular fractions. In contrast, regression-based models estimate cell type fractions by combining cell-type-specific gene expression profiles to reconstruct bulk RNA-seq data. These cell-type-specific gene expression profiles are often obtained from scRNA-seq data (top right). Some advanced techniques, such as probabilistic models, can also perform *in-silico* purification simultaneously to estimate cell-type-specific gene expression profiles (bottom right).

prior knowledge of cell types for deconvolution (supervised/semi-supervised approach), among which csSAM (Shen-Orr et al., 2010) assumes cell type fractions are known. On the contrary, few methods do not require such an input, such as DECODER (Peng et al., 2019) and CDSeq (Kang et al., 2019) (unsupervised approach). Though it is known that deconvolution after logarithmic transformation leads to a downwards bias (Zhong and Liu, 2012), some of the old and recent approaches perform deconvolution in log-linear space.

Among the supervised/semi-supervised methods that take prior knowledge of cell types for deconvolution, the vast majority employ a predetermined reference matrix of cell-type-specific gene expression data, often referred to as a signature. To construct a signature of preset cell types for deconvolution, cell-type-specific marker genes can be selected from databases or by performing differentially expressed gene analysis among each of the cell types. Early approaches constructed such a signature from gene expression data of purified cell populations, while recent approaches take that from scRNA-seq data. CIBERSORTx (Newman et al., 2019) offers an internal tool to provide signatures that represent each cell type by selecting marker gene reference profiles from scRNA-seq data. Although the vast majority of methods take only the expected gene expression profiles as the signature, MuSiC (Wang et al., 2019), Demix/DemixT (Ahn et al., 2013; Wang et al., 2018), EPIC (Racle et al., 2017), and BLADE (Andrade Barbosa et al., 2021) also take into account the variability of gene expression in each cell type for an enhanced robustness. On the other hand, CDSeq (Kang et al., 2019) and DECODER (Peng et al., 2019) estimate the number of constituent cell types as well as their populations from the bulk data without any signature (unsupervised approach). However, CDSeq offers quasi-unsupervised learning strategy, which augments the input bulk gene expression data with additional gene expression profiles of pure cell lines to get some guidance on cell type selection.

The 20 methods in Table 1 can also be categorized by the type of outcomes. Practically, a majority of supervised deconvolution methods practically can handle as many cell types as in the signature. However, enrichment-based approaches and probabilistic approaches often come with a limit in the number of cell types that can be used (e.g., both 2 cell types for ESTIMATE [Yoshihara et al., 2013] and ISOpure [Quon et al., 2013]), with exception of xCell (Aran et al., 2017). Furthermore, enrichment analysis offers a less precise estimate of cell type abundance, which is often called an enrichment score. The score cannot be compared between cell types, unlike fractions. Probabilistic methods require a sophisticated and often complex optimization strategy which may limit the number of cell types, as in Demix/DemixT (Ahn et al., 2013; Wang et al., 2018). Although the vast majority of methods predict only the cell type fractions, recent approaches can also estimate the gene expression profiles of each cell type, often referred to as in silico purification. The purification can be done either per group of samples (group-mode purification) or per one sample (high-resolution–mode purification). Those that offer purification often belong to probabilistic models, the technique that can model complex relationships between many variables, except for CIBERSORTx, which takes

a two-step approach for fraction estimation and purification (Fig. 1).

FLEXIBLE NATURE OF PROBALISTIC MODEL-BASED DECONVOLUTION METHODS

Probabilistic model-based deconvolution methods stand out for their exceptional flexibility at the expense of complex mathematical formulation and high computational costs. First of all, unlike linear regression methods, including support vector regression in CIBERSORT/CIBERSORTx, that are bound to the normal variability assumptions for gene expression data, probabilistic models can take other variability assumptions like log-normal (e.g., BLADE and Demix/DemixT) and multinomial variability (e.g., BayesPrism) (Chu et al., 2022). In particular, log-normal distribution is more suitable than normal distribution in modeling gene expression data, which cannot be log-transformed for deconvolution due to the known risk of bias (Zhong and Liu, 2012). Furthermore, probabilistic models enable the integration of multiple variables, both observed and hidden, to perform a more sophisticated deconvolution. BLADE and Demix/DemixT account for the gene expression variability observed from other data, such as scRNA-seq data. BayesPrism, CDseq, BLADE and Demix/DemixT can perform in silico purification at the same time as cell type fraction estimation; unlike CIBERSORTx, the only non-probabilistic method that offers the same results by taking a two-step approach. Probabilistic deconvolution methods enable such a combined prediction thanks to the joint probability model that includes both cell-type-specific gene expression values and cell type fractions as hidden variables. However, inference of the probabilistic models, the process of identifying hidden variables and parameters that optimize the joint probability, is a significant computational problem. That practically limits the number of cell types that can be handled in some methods (e.g., 2, 3 cell types in Demix, DemixT) and takes a long time to compute.

APPLICATION OF DECONVOLUTION METHODS TO UNRAVEL TUMOR MICROENVIRONMENT

By harnessing the established deconvolution techniques and commonly available RNA-seq data from previous cancer studies, there have been many applications of deconvolution techniques to study the tumor microenvironment (Fig. 2). Thorsson et al. (2018) identified six immune subtypes from 33 cancer types from TCGA using CIBERSORT deconvolution method in combination with several immunogenic scoring methods, such as gene set analysis. In the downstream characterization of these six subtypes, they observed that a subtype with an elevated expression level of T helper cell markers, TH17 and TH1, was associated with the best prognosis. In contrast, subtypes with a mixed signature were associated with poor overall survival. Moreover, they suggested a global regulatory network model for all tumor types and immune subtypes that consists of regulatory relationships between subtype-specific transcription factors and cancer-type-specific somatic mutations. This network model illustrates how cancer-type-specific somatic mutations lead to a specific tumor
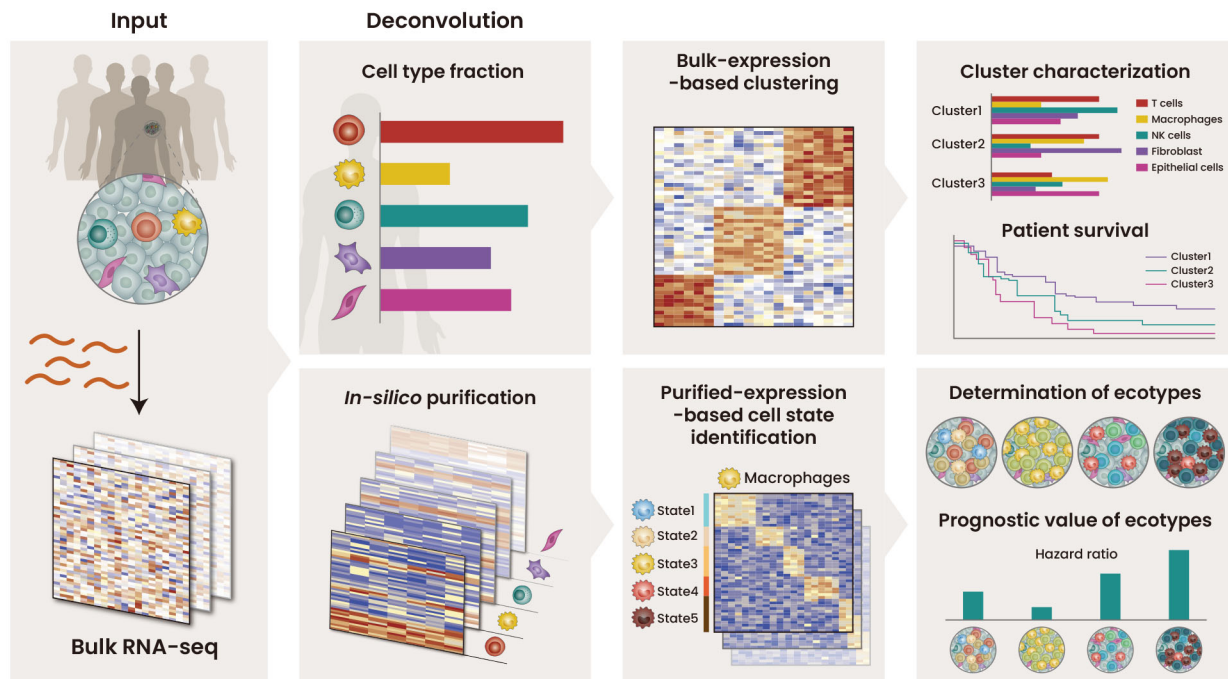
Fig. 2. Application of deconvolution to study tumor microenvironment. A diagram illustrates two scenarios of application and interpretation of deconvolution. Bulk tissue RNA-seq data from multiple cancer patients (left) are subject to deconvolution to estimate cell type fraction and gene expression profiles of each cell type (*in-silico* purification; second column). Cell type fraction results are further used to characterize each subtype of cancer, which is defined by bulk gene expression data in this example (third column/top). Cell type fractions determined by the deconvolution method may delineate the survival difference between the subtypes. Along with the cell type fraction, *in-silico* purification possible with a subset of deconvolution methods can determine transcriptional states for each cell type (third column/bottom). The co-occurrence of a group of cell states is identified to define tumor ecotypes. Further downstream analysis, such as the prognostic value of each ecotype, can be done depending on the available sample information.

immune microenvironment and an underlying key transcription factor. Recent studies have utilized the estimated gene expression profiles of each cell type (i.e., the outcome of the *in silico* purification) to further identify multiple cellular states through clustering analysis. For example, Luca et al. (2021) characterized cell states and multicellular community structure, referred to as an ecotype, across 16 types of solid carcinoma from bulk RNA-seq with the EcoTyper framework, in which CIBERSORTx deconvolution was applied (Luca et al., 2021). They defined each ecotype as the frequent co-occurrence of a group of cell types with a specific combination of cellular states. Furthermore, distinct spatial organization between the ecotypes from the same sample was confirmed by the spatial transcriptomics data, indicating molecular regulation underlying the cellular spatial organization, such as tumor infiltration.

DISCUSSION AND PERSPECTIVE

Although deconvolution is an attractive and powerful tool to obtain an extra resolution of information from standard bulk RNA-seq data, it comes with a risk when it is applied arbitrarily. Given the complexity of the problem, it is best to make use of prior knowledge of cell types, as in the supervised

deconvolution technique, comprehending that cell quality is crucial. It is ideal to extract prior knowledge from the relevant scRNA-seq data that best reflects the biological context. This process will include the critical selection of cell types for deconvolution while keeping in mind that as the more cell types with a detailed classification increase, the number of parameters to be optimized will increase and make deconvolution more challenging. On the contrary, missing an abundant cell type in bulk gene expression data will violate the common assumption in most deconvolution methods that they can reconstruct bulk gene expression profiles by combining cell-type-specific gene expression profiles. Finally, the accuracy of the predicted results can be different between cell types depending on many factors, including their abundance and unique gene expression pattern. Therefore, we recommend benchmarking the deconvolution performance for the specific prior knowledge extracted before its application. As in many benchmark experiments in deconvolution tools, an *in silico* mixture of scRNA-seq data can serve as a suitable gold standard data, although it can be rather optimistic due to the lack of technical difference between simulated bulk RNA-seq data and the prior knowledge. Furthermore, for real applications, performance per sample should be assessed using the reconstructed bulk gene expression profiles that resulted

from the deconvolution method.

There is still much more room for development in deconvolution techniques. So far, there has been only limited application of deconvolution techniques to non-RNA molecule types, except for methylation data (Chakravarthy et al., 2018). Since the estimation of cell-type-specific molecular profiling is possible with recent deconvolution techniques, enabling deconvolution with non-RNA molecules can help us delineate cellular states with multiomics rather than RNA alone. The challenge of deconvolution with other molecular types is less available cell-type-specific information. Borrowing information from other data types can be an option, for instance, RNA-based deconvolution or tumor fractions estimated by allele fraction of mutational data (Poell et al., 2019). Although it is not straightforward for most deconvolution methods to account for the extra information, probabilistic models bear the possibility of integrating extra variables into the model. Multiomics profiling of individual cell types may enable us to delineate cellular molecular mechanisms that determine specific cellular behavior. For instance, integrating spatially resolved data, such as spatial transcriptome profiling (Moffitt et al., 2022) and multiplex immunofluorescence technique (Gorris et al., 2018), may enable us to study active immune cell infiltration and associated cellular signaling pathways. Using spatial transcriptome techniques, such as Nanostring's GeoMX and 10× Visium, although providing a high-resolution spatial information, the techniques fall short on single cell resolution, leading to the use of a deconvolution technique to gain a more accurate understanding of the cellular makeup (e.g., cell2location, a Bayesian deconvolution method for spatial transcription data) (Kleshchevnikov et al., 2022). With the abundance of large-scale multiomics studies and spatially resolved data already available, particularly in the field of oncology, advanced deconvolution techniques can be applied to gain an in-depth characterization of the tumor microenvironment.

AUTHOR CONTRIBUTIONS

Y.K. conceived and provided expertise. Y.K. and Y.I. wrote the manuscript.

CONFLICT OF INTEREST

ORCID

Yebin Im   https://orcid.org/0000-0003-0966-6733
Yongsoo Kim   https://orcid.org/0000-0002-2995-2131

REFERENCES

Ahn, J., Yuan, Y., Parmigiani, G., Suraokar, M.B., Diao, L., Wistuba, I.I., and Wang, W. (2013). DeMix: deconvolution for mixed cancer transcriptomes using raw measured data. Bioinformatics 29, 1865-1871.

Andrade Barbosa, B., van Asten, S.D., Oh, J.W., Farina-Sarasqueta, A., Verheij, J., Dijk, F., van Laarhoven, H.W.M., Ylstra, B., Garcia Vallejo, J.J., van de Wiel, M.A., et al. (2021). Bayesian log-normal deconvolution for enhanced in silico microdissection of bulk gene expression data. Nat. Commun. 12, 6106.

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 18, 220.

Avila Cobos, F., Alquicira-Hernandez, J., Powell, J.E., Mestdagh, P., and De Preter, K. (2020). Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat. Commun. 11, 5650.

Baghban, R., Roshangar, L., Jahanban-Esfahlan, R., Seidi, K., Ebrahimi-Kalan, A., Jaymand, M., Kolahian, S., Javaheri, T., and Zare, P. (2020). Tumor microenvironment complexity and therapeutic implications at a glance. Cell Commun. Signal. 18, 59.

Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. Cell Syst. 3, 346-360.e4.

Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 17, 218.

Campbell, P.J., Getz, G., Korbel, J.O., Stuart, J.M., Jennings, J.L., Stein, L.D., Perry, M.D., Nahal-Bose, H.K., Ouellette, B.F.F., Li, C.H., et al. (2020). Pan-cancer analysis of whole genomes. Nature 578, 82-93.

Chakravarthy, A., Furness, A., Joshi, K., Ghorani, E., Ford, K., Ward, M.J., King, E.V., Lechner, M., Marafioti, T., Quezada, S.A., et al. (2018). Pan-cancer deconvolution of tumour composition using DNA methylation. Nat. Commun. 9, 3220.

Chu, T., Wang, Z., Pe'er, D., and Danko, C.G. (2022). Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat. Cancer 3, 505-517.

Dagogo-Jack, I. and Shaw, A.T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nat. Rev. Clin. Oncol. 15, 81-94.

Denisenko, E., Guo, B.B., Jones, M., Hou, R., de Kock, L., Lassmann, T., Poppe, D., Clément, O., Simmons, R.K., Lister, R., et al. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol. 21, 130.

Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., Krogsdam, A., Loncova, Z., Posch, W., Wilflingseder, D., et al. (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. Genome Med. 11, 34.

Gorris, M.A.J., Halilovic, A., Rabold, K., van Duffelen, A., Wickramasinghe, I.N., Verweij, D., Wortel, I.M.N., Textor, J.C., de Vries, I.J.M., and Figdor, C.G. (2018). Eight-color multiplex immunohistochemistry for simultaneous detection of multiple immune checkpoint molecules within the tumor microenvironment. J. Immunol. 200, 347-354.

Jin, M.Z. and Jin, W.L. (2020). The updated landscape of tumor microenvironment and drug repurposing. Signal Transduct. Target. Ther. 5, 166.

Kang, K., Meng, Q., Shats, I., Umbach, D.M., Li, M., Li, Y., Li, X., and Li, L. (2019). CDSeq: a novel complete deconvolution method for dissecting heterogeneous samples using gene expression data. PLoS Comput. Biol. 15, e1007510.

Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H.W., Li, T., Elmentaite, R., Lomakin, A., Kedlian, V., Gayoso, A., et al. (2022). Cell2location maps fine-grained cell types in spatial transcriptomics. Nat. Biotechnol. 40, 661-671.

Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerenwinkel, N., Mahfouz, A., et al. (2020). Eleven grand challenges in single-cell data science. Genome Biol. 21, 31.

Lee, J., Hyeon, D.Y., and Hwang, D. (2020). Single-cell multiomics: technologies and data analysis methods. Exp. Mol. Med. 52, 1428-1442.

Li, B., Li, T., Liu, J.S., and Liu, X.S. (2020). Computational deconvolution of tumor-infiltrating immune components with bulk tumor gene expression data. Methods Mol. Biol. 2120, 249-262.

Luca, B.A., Steen, C.B., Matusiak, M., Azizi, A., Varma, S., Zhu, C., Przybyl, J., Espín-Pérez, A., Diehn, M., Alizadeh, A.A., et al. (2021). Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell 184, 5482-5496.e28.

Moffitt, J.R., Lundberg, E., and Heyn, H. (2022). The emerging landscape of spatial profiling technologies. Nat. Rev. Genet. 23, 741-759.

Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y.Y., Carré, C., Burdin, N., Visan, L., Ceccarelli, M., Poidinger, M., et al. (2019). RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. Cell Rep. 26, 1627-1640.e7.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods 12, 453-457.

Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. 37, 773-782.

Nguyen, A., Yoshida, M., Goodarzi, H., and Tavazoie, S.F. (2016). Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. Nat. Commun. 7, 11246.

Peng, X.L., Moffitt, R.A., Torphy, R.J., Volmar, K.E., and Yeh, J.J. (2019). De novo compartment deconvolution and weight estimation of tumor samples using DECODER. Nat. Commun. 10, 4729.

Poell, J.B., Mendeville, M., Sie, D., Brink, A., Brakenhoff, R.H., and Ylstra, B. (2019). ACE: absolute copy number estimation from low-coverage whole-genome sequencing data. Bioinformatics 35, 2847-2849.

Quon, G., Haider, S., Deshwar, A.G., Cui, A., Boutros, P.C., and Morris, Q. (2013). Computational purification of individual tumor gene expression profiles leads to significant improvements in prognostic prediction. Genome Med. 5, 29.

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D.E., and Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. Elife 6, e26476.

Shen-Orr, S.S., Tibshirani, R., Khatri, P., Bodian, D.L., Staedtler, F., Perry, N.M., Hastie, T., Sarwal, M.M., Davis, M.M., and Butte, A.J. (2010). Cell type–specific gene expression differences in complex tissues. Nat. Methods 7, 287-289.

Sturm, G., Finotello, F., Petitprez, F., Zhang, J.D., Baumbach, J., Fridman, W.H., List, M., and Aneichyk, T. (2019). Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. Bioinformatics 35, i436-i445.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The immune landscape of cancer. Immunity 48, 812-830.e14.

Wang, X., Park, J., Susztak, K., Zhang, N.R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat. Commun. 10, 380.

Wang, Z., Cao, S., Morris, J.S., Ahn, J., Liu, R., Tyekucheva, S., Gao, F., Li, B., Lu, W., Tang, X., et al. (2018). Transcriptome deconvolution of heterogeneous tumor samples with immune infiltration. iScience 9, 451-460.

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. 4, 2612.

Zhong, Y. and Liu, Z. (2012). Gene expression deconvolution in linear space. Nat. Methods 9, 8-9.

Zhong, Y., Wan, Y.W., Pang, K., Chow, L.M., and Liu, Z. (2013). Digital sorting of complex tissues for cell type-specific gene expression profiles. BMC Bioinformatics 14, 89.