


REVIEW

Open Access



# Unraveling the complexity: understanding the deconvolutions of RNA-seq data

Kavoos Momeni<sup>1\*</sup> , Saeid Ghorbian<sup>1</sup>, Ehsan Ahmadpour<sup>2</sup> and Rasoul Sharifi<sup>3</sup>

## Abstract

Deconvolution of RNA sequencing data is a computational method used to estimate the relative proportions of different cell types or subpopulations within a heterogeneous sample based on gene expression profiles. This technique is particularly useful in studies where the goal is to identify changes in gene expression that are specific to a particular cell type or subpopulation.

The deconvolution process involves using reference gene expression profiles from known cell types or subpopulations to infer the relative abundance of these cells within a mixed sample. This is typically done using linear regression or other statistical methods to model the observed gene expression data as a linear combination of the reference profiles.

Once the relative proportions of each cell type or subpopulation have been estimated, downstream analyses can be performed on each component separately, allowing for more precise identification of cell-type-specific changes in gene expression.

Overall, deconvolution of RNA sequencing data is a powerful tool for dissecting complex biological systems and identifying cell-type-specific molecular signatures that may be relevant for disease diagnosis and treatment.

**Keywords** Deconvolution techniques, RNA-seq data analysis, Differential gene expression analysis, Transcriptome profiling, CIBERSORT, xCell, MCP-counter

## Introduction

RNA sequencing (RNA-seq) has revolutionized the field of transcriptomics by providing a comprehensive view of gene expression at the transcript level. However, analyzing RNA-seq data can be challenging due to its high dimensionality and complexity.

One common approach is to perform differential gene expression analysis, which identifies genes that are

differentially expressed between two or more conditions. However, this approach does not take into account the relationships between genes and their functions.

Deconvolution is a recently developed method that addresses this issue by deconvolving RNA-seq data into cell-type-specific expression profiles.

In this article, we will explore the concept of deconvolution and its applications in understanding complex biological systems.

## Deconvolution of RNA-seq data

RNA sequencing (RNA-seq) is a powerful tool for studying gene expression and transcriptome profiling. However, the analysis of RNA-seq data can be challenging due to the complexity of the data and the presence of various sources of noise. Deconvolution is a computational

\*Correspondence:

Kavoos Momeni  
kavoosmomeni@gmail.com

<sup>1</sup>Department of Molecular Genetics - Ahar Branch, Islamic Azad University, Ahar, Iran

<sup>2</sup>Tabriz University of Medical Sciences, Tabriz, Iran

<sup>3</sup>Department of Biology, Faculty of Basic Science - Ahar Branch - Islamic, Azad University, Ahar, Iran



method that can be used to separate different sources of variation in RNA-seq data, such as cell type-specific gene expression or batch effects.

Deconvolution is a mathematical technique that aims to estimate the underlying components of a mixture based on their observed signals. In RNA-seq data, deconvolution can be used to estimate the relative contribution of different cell types or biological processes to the observed gene expression profiles. The basic idea behind deconvolution is to use a reference dataset that contains known expression profiles for each component of interest, and then estimate the contribution of each component to the observed data using linear regression or other statistical methods.

One common application of deconvolution in RNA-seq data analysis is cell type-specific gene expression analysis. In many tissues and organs, different cell types have distinct gene expression profiles that reflect their specialized functions. However, bulk RNA-seq experiments often measure gene expression from a mixture of multiple cell types, which can obscure cell type-specific signals. Deconvolution can be used to estimate the relative abundance of each cell type in a mixed sample based on their known gene expression profiles. This approach

has been applied to various tissues and diseases, such as brain tumors [1], immune cells [2], and lung cancer [3].

Another application of deconvolution in RNA-seq data analysis is batch effect correction. Batch effects are systematic variations in gene expression that arise from technical factors such as sample preparation or sequencing runs. Batch effects can confound the analysis of RNA-seq data and lead to false positive or negative results. Deconvolution can be used to estimate the batch effect from a reference dataset that contains samples with known batch labels, and then adjust the observed gene expression profiles accordingly. This approach has been shown to improve the accuracy and reproducibility of RNA-seq data analysis [4, 5].

As RNA-seq data continue to grow in size and complexity, deconvolution will become an increasingly important tool for understanding gene expression regulation in health and disease (Fig. 1).

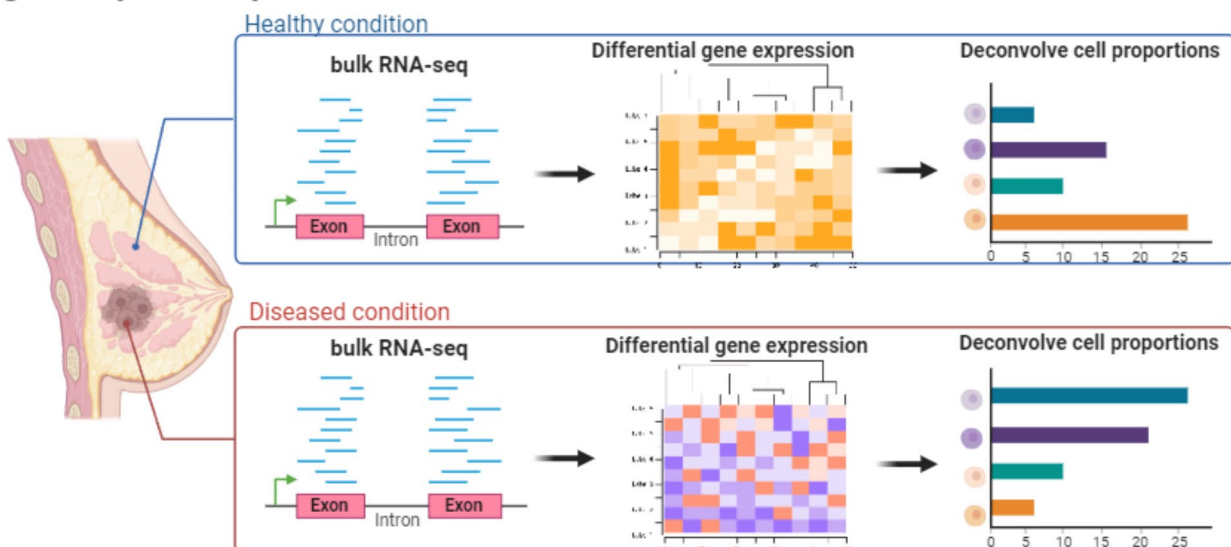
#### Advantages of deconvolution over traditional gene expression analysis methods

Deconvolution is a computational method that has gained popularity in recent years for analyzing gene expression data. It is a powerful tool that allows

## What is RNA-seq Deconvolution?

Our goal is to identify the cell types in a given spot.

This method predicts the proportion of different cell types in a spot based on the gene expression profile.



Created in BioRender.com

**Fig. 1** In RNA-seq deconvolution, a biopsy is obtained and subjected to RNA sequencing and Differential Gene Expression analysis. The resulting data is then processed using deconvolution algorithms and combined with prior knowledge from cell genomes. Through this analysis, the number and characteristics of cells within the tissue can be calculated

researchers to estimate the cell-type-specific gene expression profiles from bulk tissue samples. Traditional gene expression analysis methods, on the other hand, rely on the assumption that all cells in a sample have similar gene expression profiles.

One of the main advantages of deconvolution is its ability to identify cell-type-specific changes in gene expression. In traditional gene expression analysis methods, it is difficult to distinguish between changes in gene expression that are due to changes in the proportion of different cell types and changes in gene expression within individual cells. Deconvolution allows researchers to separate these two sources of variation and identify genes that are specifically upregulated or downregulated in certain cell types. This can provide valuable insights into how different treatments or interventions affect specific cell types and help researchers develop more targeted therapies [6, 7].

Another advantage of deconvolution is its ability to identify rare cell populations. In many tissues, there are small populations of cells that play important roles in disease progression or tissue regeneration. These rare cell populations can be difficult to detect using traditional gene expression analysis methods because their signal may be drowned out by the more abundant cell types. Deconvolution can help researchers identify these rare cell populations and study their role in disease or tissue regeneration [8].

Deconvolution also allows for more accurate interpretation of results from bulk tissue samples. Traditional gene expression analysis methods assume that all cells within a sample have similar gene expression profiles, which may not be true for complex tissues such as tumors or immune tissues. Deconvolution can help researchers identify which genes are expressed by which cell types within a sample, allowing for more accurate interpretation of results [9].

Deconvolution can be used to identify new biomarkers for disease diagnosis and prognosis. Traditional gene expression analysis methods are limited in their ability to identify biomarkers that are specific to individual cell types. Deconvolution overcomes this limitation by allowing researchers to identify biomarkers that are specific to individual cell types and can be used for disease diagnosis and prognosis [10].

Finally, deconvolution can be used to study complex biological processes such as immune responses or tissue regeneration. These processes involve multiple cell types with distinct functions and gene expression profiles. Deconvolution can help researchers understand how different cell types interact and contribute to these processes [11]. Cancer is a complex disease that involves multiple different cell types. Deconvolution can be used to identify the specific cell types that are involved in

cancer development and progression and to identify the specific genes that are expressed in these cell types. This can provide valuable insights into cancer development and progression mechanisms and help researchers develop more effective treatments [12].

Deconvolution allows researchers to identify cell-type-specific changes in gene expression, identify rare cell populations, interpret results from bulk tissue samples more accurately, and study complex biological processes. As such, it has become an essential tool for many researchers studying gene expression in complex tissues.

#### **Limitations of deconvolution of RNA-seq data**

Deconvolution of RNA-seq data is a computational method that aims to estimate the cell type-specific gene expression profiles from bulk RNA-seq data. However, there are several limitations to this approach, including technical and biological factors that can affect the accuracy and reliability of the results.

One major limitation of deconvolution is the lack of reliable cell type-specific markers. The identification of cell type-specific markers is crucial for accurate estimation of gene expression profiles in different cell types. However, many cell types share common markers, and some markers may be expressed in multiple cell types, leading to inaccurate estimates. Moreover, some cell types may have low expression levels or be rare in the sample, making it difficult to accurately estimate their gene expression profiles [8].

Another limitation is the heterogeneity within cell types. Even within a single cell type, there can be significant heterogeneity due to differences in developmental stage, activation state, or environmental cues. Failure to properly address heterogeneity can lead to several issues. For instance, it can result in overestimation or underestimation of gene expression levels for specific cell types or subpopulations. This can have significant implications for downstream analyses, such as understanding disease mechanisms, identifying biomarkers, or developing targeted therapies [13].

Technical factors such as batch effects and sequencing depth can also affect the accuracy of deconvolution results. Batch effects arise when samples are processed at different times or by different technicians, leading to systematic differences in gene expression levels that are not related to biological variation. Sequencing depth can also affect the accuracy of deconvolution results since low coverage may result in inaccurate estimates of gene expression levels [14].

Finally, deconvolution assumes that all genes are expressed independently across different cell types. However, this assumption may not hold true for all genes since some genes may be co-regulated or co-expressed across multiple cell types [15].

In conclusion, while deconvolution is a useful tool for estimating cell type-specific gene expression profiles from bulk RNA-seq data, it has several limitations that must be considered. These limitations include the lack of reliable cell type-specific markers, heterogeneity within cell types, technical factors such as batch effects and sequencing depth, and the assumption of independent gene expression across different cell types. Therefore, careful consideration of these factors is necessary when interpreting deconvolution results.

#### **Using deconvolution of RNA-seq data to identify new biomarkers for disease diagnosis**

Deconvolution of RNA-seq data has emerged as a promising approach to address this challenge and identify new biomarkers for disease diagnosis.

Deconvolution is a computational method that separates mixed signals into their individual components. In the context of RNA-seq data, deconvolution can be used to separate the expression profiles of different cell types or tissues within a sample. This approach has been applied to identify biomarkers for various diseases, including cancer, autoimmune disorders, and neurological disorders [16].

One example of using deconvolution to identify biomarkers is in the study of breast cancer. Breast cancer is a heterogeneous disease with different subtypes that have distinct molecular characteristics and clinical outcomes. Deconvolution of RNA-seq data from breast cancer samples can be used to identify the expression profiles of different cell types within the tumor microenvironment, such as immune cells and stromal cells. By comparing these profiles between different subtypes of breast cancer, researchers can identify genes that are specifically expressed in certain cell types and may serve as biomarkers for diagnosis or prognosis [8].

Another example is in the study of autoimmune disorders such as rheumatoid arthritis (RA). RA is characterized by chronic inflammation in the joints, which leads to joint damage and disability if left untreated. Deconvolution of RNA-seq data from RA patients can be used to identify genes that are specifically expressed in immune cells that are involved in the pathogenesis of RA, such as T cells and B cells. These genes may serve as biomarkers for early diagnosis or monitoring of disease activity [17, 18].

In addition to identifying new biomarkers, deconvolution can also improve our understanding of disease mechanisms by revealing changes in gene expression patterns within specific cell types or tissues. For example, deconvolution of RNA-seq data from Alzheimer's disease patients has revealed changes in gene expression patterns in microglia, the immune cells of the brain, which may

contribute to the neuro inflammation and neuronal damage observed in this disease [19].

In one of the recent articles, the effect of cutaneous leishmaniasis infection on skin tissue has been investigated using the deconvolution method on RNA-seq data. Remarkably, despite the absence of any microscopic observations, they discovered a significant increase in the population of immune cells in the damaged tissue. This innovative application of the deconvolution technique elucidates the immunological dynamics associated with cutaneous leishmaniasis [20].

By separating mixed signals into their individual components, deconvolution can reveal changes in gene expression patterns within specific cell types or tissues that may be missed by traditional analysis methods. This approach has the potential to improve our understanding of disease mechanisms and facilitate the development of more effective diagnostic and therapeutic strategies.

#### **Using deconvolution of RNA-seq data for understanding cancer**

Deconvolution of RNA-seq data has emerged for understanding the complex biology of cancer. RNA sequencing (RNA-seq) is a widely used technique for measuring gene expression levels in cells and tissues. However, the heterogeneity of cancer samples, which often contain multiple cell types, can confound the interpretation of RNA-seq data. Deconvolution methods aim to estimate the relative abundance of different cell types in a mixed sample based on their gene expression profiles.

One application of deconvolution in cancer research is to identify the cell types that contribute to tumor progression and metastasis. For example, immune cells such as T cells and macrophages have been shown to play important roles in shaping the tumor microenvironment and influencing cancer progression [21]. By deconvolving RNA-seq data from tumor samples, researchers can identify changes in immune cell populations that are associated with different stages of cancer development.

Another use of deconvolution is to identify molecular pathways that are dysregulated in specific cell types within tumors. For example, a recent study used deconvolution to identify genes that are specifically upregulated in cancer-associated fibroblasts (CAFs), a type of stromal cell that promotes tumor growth and invasion [22]. By targeting these CAF-specific genes with small molecules or other therapies, it may be possible to disrupt the supportive environment that allows tumors to thrive.

Deconvolution can also be used to study the effects of cancer treatments on different cell types within tumors. For example, chemotherapy drugs often target rapidly dividing cells such as tumor cells but can also affect normal cells such as immune cells and stromal cells. By

deconvolving RNA-seq data from pre- and post-treatment samples, researchers can identify changes in the relative abundance of different cell types and assess how they respond to treatment [23].

By identifying the cell types that contribute to tumor progression, dysregulated molecular pathways, and treatment effects on different cell types, deconvolution can provide insights into the mechanisms underlying cancer development and inform the development of new therapies.

#### **Deconvolution in single-cell RNA sequencing**

Deconvolution has become a valuable tool in the analysis of single-cell RNA sequencing (scRNA-seq) data. The application of deconvolution in scRNA-seq involves estimating the cell type composition and abundance within a heterogeneous cell population based on the gene expression profiles obtained from individual cells.

One common approach is to use reference gene expression profiles from known cell types as a basis for deconvolution. These reference profiles can be obtained from bulk RNA-seq data or from existing databases of cell type-specific gene expression signatures. By comparing the gene expression patterns of individual cells to the reference profiles, deconvolution algorithms can infer the relative proportions of different cell types present in the population.

Deconvolution in scRNA-seq can provide valuable insights into cellular heterogeneity and the composition of complex tissues or disease states. It allows researchers to identify and quantify specific cell types, characterize cell type-specific gene expression patterns, and investigate changes in cell type proportions under different conditions or disease states.

Furthermore, deconvolution can be used to infer signaling interactions and cellular communication within a tissue or between different cell types. By estimating the abundance of specific cell types and their interactions, deconvolution methods can provide a more comprehensive understanding of cellular dynamics and functional relationships.

The application of deconvolution in scRNA-seq data analysis enhances our ability to unravel the cellular complexity of tissues and diseases, enabling more accurate characterization and interpretation of single-cell gene expression profiles.

#### **Deconvolution in rare cell population**

Deconvolution methods can be particularly useful in studying rare cell populations within scRNA-seq data. Rare cell populations often present challenges in their identification and characterization due to their low abundance and potential overlap with more abundant cell

types. However, deconvolution can aid in unraveling the presence and properties of these rare cells.

By leveraging reference gene expression profiles from known cell types, deconvolution algorithms can estimate the proportions of different cell types, including those that are rare, within a heterogeneous population. This allows researchers to identify and quantify the specific rare cell populations present in the data.

Deconvolution can also assist in distinguishing rare cell types from closely related or overlapping cell populations. By comparing the gene expression profiles of individual cells to the reference profiles, deconvolution methods can help discriminate between similar cell types that may share some gene expression patterns. This can provide insights into the specific gene expression signatures or markers that distinguish the rare cell population of interest.

Furthermore, deconvolution can facilitate downstream analyses of rare cells by enabling their isolation for further experimental validation or functional characterization. Once the rare cell population is identified and quantified, researchers can target and isolate these cells for additional experiments such as flow cytometry, single-cell sequencing, or functional assays. This targeted isolation can greatly enhance our understanding of the biological properties and functions of these rare cells [24, 25].

Deconvolution methods play a vital role in aiding the study of rare cell populations in scRNA-seq data by accurately estimating their proportions, distinguishing them from similar cell types, and enabling their targeted isolation for further analysis. This contributes to a more comprehensive understanding of rare cell populations and their significance in various biological processes and disease contexts.

#### **Different methods of RNA-seq data deconvolution**

Here are some commonly used methods for deconvolution of RNA-seq:

1. CIBERSORT
  - CIBERSORT (Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts) is a widely used deconvolution method for RNA-seq data. It employs a support vector regression algorithm to estimate the proportions of cell types in a mixed sample. CIBERSORT relies on a predefined signature matrix consisting of gene expression profiles from pure cell types. It quantifies the relative contributions of these signatures to the gene expression patterns observed in the mixed sample [8, 26–29].
2. quanTIseq:
  - quanTIseq is a deconvolution method specifically designed for tumor-infiltrating immune cells in

cancer samples. It uses RNA-seq data to estimate the proportions of different immune cell types within the tumor microenvironment. quanTIseq integrates a machine learning approach with a gene signature matrix to infer immune cell proportions. It accounts for the heterogeneity of cell types and the potential presence of unknown cell subtypes [12].

### 3. xCell:

- xCell is a deconvolution method that estimates the proportions of various cell types in a mixed sample using RNA-seq data. It employs a gene set enrichment analysis approach, comparing the expression profiles of genes in the mixed sample to a reference gene set database representing different cell types. xCell takes advantage of the distinct gene expression patterns associated with specific cell types to estimate their relative abundances [30–32].

### 4. DeconRNASeq:

- DeconRNASeq is a deconvolution method that utilizes RNA-seq data to estimate cell type proportions in a mixture. It employs a constrained non-negative matrix factorization approach, which decomposes the gene expression matrix of the mixed sample into two matrices representing the cell type proportions and gene expression signatures of individual cell types. By iteratively optimizing the factorization, DeconRNASeq estimates the relative abundances of cell types [33].

### 5. MuSiC:

- MuSiC (Multi-subject Single-cell deconvolution) is a deconvolution method specifically designed for single-cell RNA-seq data. It estimates the proportions of cell types within a sample by leveraging reference single-cell RNA-seq datasets. MuSiC employs a probabilistic model to infer the cell type proportions based on the similarity between the gene expression profiles of the sample and reference datasets [34].

### 6. MCP-counter:

- MCP-counter (Microenvironment Cell Populations-counter) is a deconvolution method that can be used for RNA-seq data. It estimates the proportions of different cell types within a sample based on RNA-seq gene expression profiles. MCP-counter specifically focuses on characterizing the tumor microenvironment by quantifying the abundance of tumor-infiltrating immune cells and stromal cells [35].

MCP-counter employs a reference gene expression signature matrix that represents different cell types within the tumor microenvironment.

It uses a single-sample gene set enrichment analysis (ssGSEA) approach to assess the enrichment of these cell type signatures in the gene expression data of the sample. By comparing the sample's gene expression profiles to the reference signatures, MCP-counter estimates the relative proportions of immune and stromal cell populations within the tumor [8].

MCP-counter has been widely used in cancer research to explore the composition of the tumor microenvironment and its association with clinical outcomes. It provides valuable insights into the immune and stromal cell components of tumors, allowing researchers to study their roles in tumor progression, immune response, and therapy response [36].

### Comparison of the deconvolution methods.

#### 1. CIBERSORT

- Widely used gene expression-based deconvolution method.
- Relies on a predefined signature matrix of gene expression profiles from pure cell types.
- Utilizes support vector regression algorithm for estimation.
- Effective for estimating cell type proportions in heterogeneous samples.
- Particularly useful for studying immune cell composition in various diseases, including cancer.

#### 2. quanTIseq:

- Specialized for deconvolving tumor-infiltrating immune cells in cancer samples.
- Employs a machine learning approach with a gene signature matrix.
- Accounts for cell type heterogeneity and possible unknown subtypes.
- Enables analysis of immune cell composition and its association with tumor biology and clinical outcomes.

#### 3. xCell:

- Utilizes gene set enrichment analysis to estimate cell type proportions.
- Relies on a reference gene set database representing different cell types.
- Takes advantage of distinct gene expression patterns associated with specific cell types.
- Provides relative abundance estimates of various cell types in mixed samples.

#### 4. DeconRNASeq:

- Uses a constrained non-negative matrix factorization approach.
- Decomposes the gene expression matrix into cell type proportions and gene expression signatures.
- Suitable for estimating cell type proportions in mixed RNA-seq samples.

- Facilitates characterization of cell type-specific gene expression profiles.
5. MuSiC:
- Developed for deconvolving single-cell RNA-seq data.
  - Utilizes a probabilistic model with reference single-cell datasets.
  - Estimates cell type proportions based on similarity between gene expression profiles.
  - Enables identification of cell type composition in single-cell transcriptomic data.
6. MCP-counter:
- Focuses on characterizing the tumor microenvironment.
  - Estimates the abundance of tumor-infiltrating immune cells and stromal cells.
  - Relies on a reference gene expression signature matrix.
  - Utilizes single-sample gene set enrichment analysis (ssGSEA).
  - Provides insights into immune and stromal cell components of tumors.

These deconvolution methods offer various approaches and algorithms for estimating cell type proportions in RNA-seq data. They differ in their methodology, target cell types, and specific applications. Researchers should consider the biological context, research question, and available reference data when selecting the most appropriate method for their specific study. Additionally, it is important to validate and interpret the results carefully, considering the limitations and assumptions of each method.

## Conclusion

Deconvolution of RNA-seq data has become an increasingly popular method for analyzing complex gene expression profiles. It offers several advantages, including the ability to identify cell-type-specific gene expression patterns and to infer changes in cell composition within a tissue or sample. However, there are also limitations to this approach, such as the need for accurate reference datasets and the potential for bias in the deconvolution process. Despite these challenges, researchers continue to refine and improve deconvolution methods, making it a valuable tool for understanding gene expression in complex biological systems. As our understanding of this technique continues to evolve, we can expect it to play an increasingly important role in advancing our knowledge of cellular biology and disease pathology.

## Acknowledgements

I would like to express my sincere gratitude to all those who have contributed to the completion of this review article on gene ontology.

## Authors contribution

The authors confirm their contribution to the paper as follows: Study conception and design: Kavoos Momeni, Saeid Ghorbian; Data collection: Kavoos Momeni; Ehsan Ahmadpour; Analysis and interpretation of results: Kavoos Momeni, Saeid Ghorbian, Ehsan Ahmadpour, Rasoul Sharifi; Draft manuscript preparation: Kavoos Momeni, Saeid Ghorbian, Ehsan Ahmadpour. All authors reviewed the results and approved the final version of the manuscript.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Data availability

The data and materials used in this review article on Deconvolution of RNA-seq data are readily available in public repositories such as the Gene Expression Omnibus (GEO) and the Sequence Read Archive (SRA). The software tools and algorithms discussed in this article are also freely accessible online, enabling researchers to replicate the analyses presented here.

## Declarations

### Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any authors. In this paper, I have presented a part of the results of my doctoral dissertation, titled "Bioinformatics Analysis of differential gene expression in cutaneous leishmaniasis Lesions" with Research Ethical Committee Certificate: IR.IAU.TABRIZ.REC.1401.179.

### Consent for publication

We, the authors, give our consent for the publication of identifiable details, which can include photograph(s) and/or details within the text "Unraveling the complexity: Understanding the Deconvolutions of RNA-seq Data" to be published in the "Journal of Translational Medicine".

### Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 2 July 2023 / Accepted: 11 September 2023

Published online: 26 September 2023

## References

1. Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, et al. Single-cell RNA-Seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep*. 2017;21(5):1399–410.
2. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37(7):773–82.
3. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016;17(1):174.
4. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3.
5. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol*. 2014;32(9):896–902.
6. Chu T, Wang Z, Pe'er D, Danko CG. Cell type and gene expression deconvolution with BayesPrism enables bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat Cancer*. 2022;3(4):505–17.
7. Marquez-Galera A, de la Prida LM, Lopez-Atalaya JP. A protocol to extract cell-type-specific signatures from differentially expressed genes in bulk-tissue RNA-seq. *STAR Protocols*. 2022;3(1):101121.

8. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
9. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2018.
10. Maity AK, Stone TC, Ward V, Webster AP, Yang Z, Hogan A, et al. Novel epigenetic network biomarkers for early detection of esophageal cancer. *Clin Epigenetics*. 2022;14(1):23.
11. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3, et al. Comprehensive integration of single-cell data. *Cell*. 2019;177(7):1888–902e21.
12. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med*. 2019;11(1):34.
13. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics*. 2019;35(14):i436–i45.
14. Tirosch I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016;352(6282):189–96.
15. Yuan GC, Cai L, Elowitz M, Enver T, Fan G, Guo G, et al. Challenges and emerging directions in single-cell analysis. *Genome Biol*. 2017;18(1):84.
16. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
17. Shen-Orr SS, Gaujoux R. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. *Curr Opin Immunol*. 2013;25(5):571–8.
18. Zhang F, Wei K, Slowikowski K, Fonseka CY, Rao DA, Kelly S, et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol*. 2019;20(7):928–42.
19. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O’Keefe S, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci*. 2014;34(36):11929–47.
20. Momeni K, Ghorbian S, Ahmadpour E, Sharifi R. Identification of molecular mechanisms causing skin lesions of cutaneous leishmaniasis using weighted gene coexpression network analysis (WGCNA). *Sci Rep*. 2023;13(1):9836.
21. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med*. 2018;24(5):541–50.
22. Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvisse M, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med*. 2017;214(3):579–96.
23. Zhao W, Dovas A, Spinazzi EF, Levitin HM, Banu MA, Upadhyayula P, et al. Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell RNA-seq. *Genome Med*. 2021;13(1):82.
24. Cobos FA, Panah MJN, Epps J, Long X, Man TK, Chiu HS, et al. Effective methods for bulk RNA-seq deconvolution using scRNA-seq transcriptomes. *Genome Biol*. 2023;24(1):177.
25. Tsoucas D, Dong R, Chen H, Zhu Q, Guo G, Yuan GC. Accurate estimation of cell-type composition from gene expression data. *Nat Commun*. 2019;10(1):2975.
26. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor infiltrating Immune cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243–59.
27. Chiu Y-J, Hsieh Y-H, Huang Y-H. Improved cell composition deconvolution method of bulk gene expression profiles to quantify subsets of immune cells. *BMC Med Genom*. 2019;12(8):169.
28. Ali HR, Chlon L, Pharoah PD, Markowitz F, Caldas C. Patterns of Immune infiltration in breast Cancer and their clinical implications: a gene-expression-based retrospective study. *PLoS Med*. 2016;13(12):e1002194.
29. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature*. 2017;541(7637):321–30.
30. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017;18(1):220.
31. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*. 2015;6:8971.
32. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype Relationships and Predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248–62.
33. Gong T, Szustakowski JD. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics*. 2013;29(8):1083–5.
34. Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun*. 2019;10(1):380.
35. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17(1):218.
36. 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*. 2017;6.

## Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.