

Practical: deconvolution of the clonality of a tumour using single-cell transcriptomic data

Martha Estefania Vázquez Cruz

International Laboratory for Human Genome Research

National Autonomous University of Mexico (LIIGH UNAM)

evazquez@liigh.unam.mx

Gabriel Pereira

Federal University of Health Sciences of Porto Alegre (UFCSPA)

Immunotherapy LAB (LAIT)

gabriel.pereira@ufcspa.edu.br



LIIGH UNAM

EMBL-EBI



universidade federal de ciências da saúde
de porto alegre



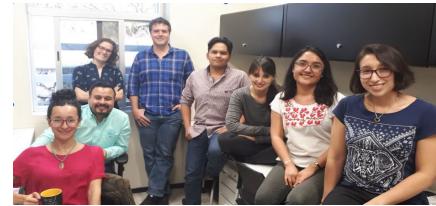
Laboratório de Imunoterapia

Overview

- 1** Introduction
- 2** Approaches and considerations
- 3** Benchmark studies
- 4** Activity 1: Individual exercises (BayesPrism, Bisque, CIBERSORTx)
- 5** Activity 2: Team challenge
- 6** Activity 3: Final quiz

Estef Vazquez

Institution: LIIGH UNAM - National Autonomous University of Mexico
Cancer Genetics and Bioinformatics (CGBioLab)
Led by Professor Daniela Robles-Espinoza, PhD



I am currently investigating the components of the tumor immune microenvironment of acral melanoma. My research interests include gaining a deeper understanding of immune evasion mechanisms in cancer and integration of multi-omics data to uncover biological underlying insights.

Hobbies: Play guitar and sing, play with my cat, running, learn german

X @EstefVazque



Gabriel Pereira



Institution: UFCSPA - Federal University of Health Sciences of Porto Alegre (Brazil)
Immunotherapy Lab (LAIT)
Led by Professor Cristina Bonorino, PhD

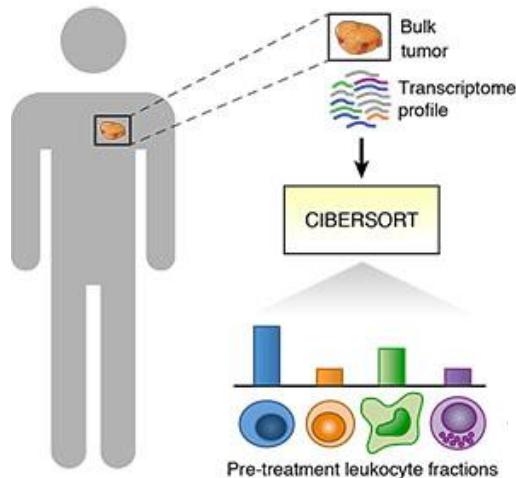
I am working on creating a single-cell and spatially resolved atlas of pancreatic cancer. My goal is to understand the tumor ecosystem and the spatial distribution of immune cells, and how this information can provide critical insights into prognosis and therapy response.

Hobbies: Play football, play with my dog and watching F1



Deconvolution

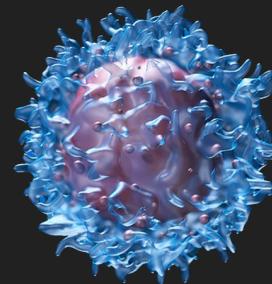
Computational approaches to estimate cell-type composition from bulk RNA-seq.



Newman et al., 2015

Using scRNA-seq profiles to define cell-type-specific signatures

High resolution of single-cell data



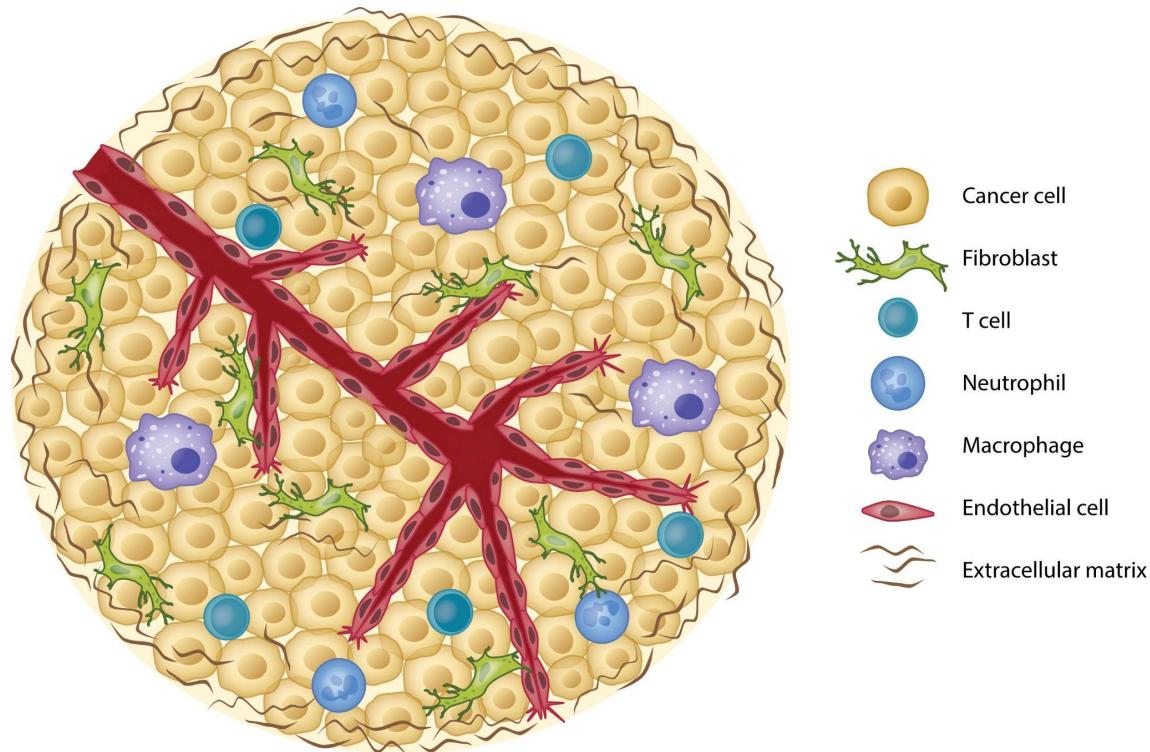
Cost-effectiveness of bulk RNA-seq

The tumour microenvironment impacts therapy response

Constant communication between tumour and immune cells.

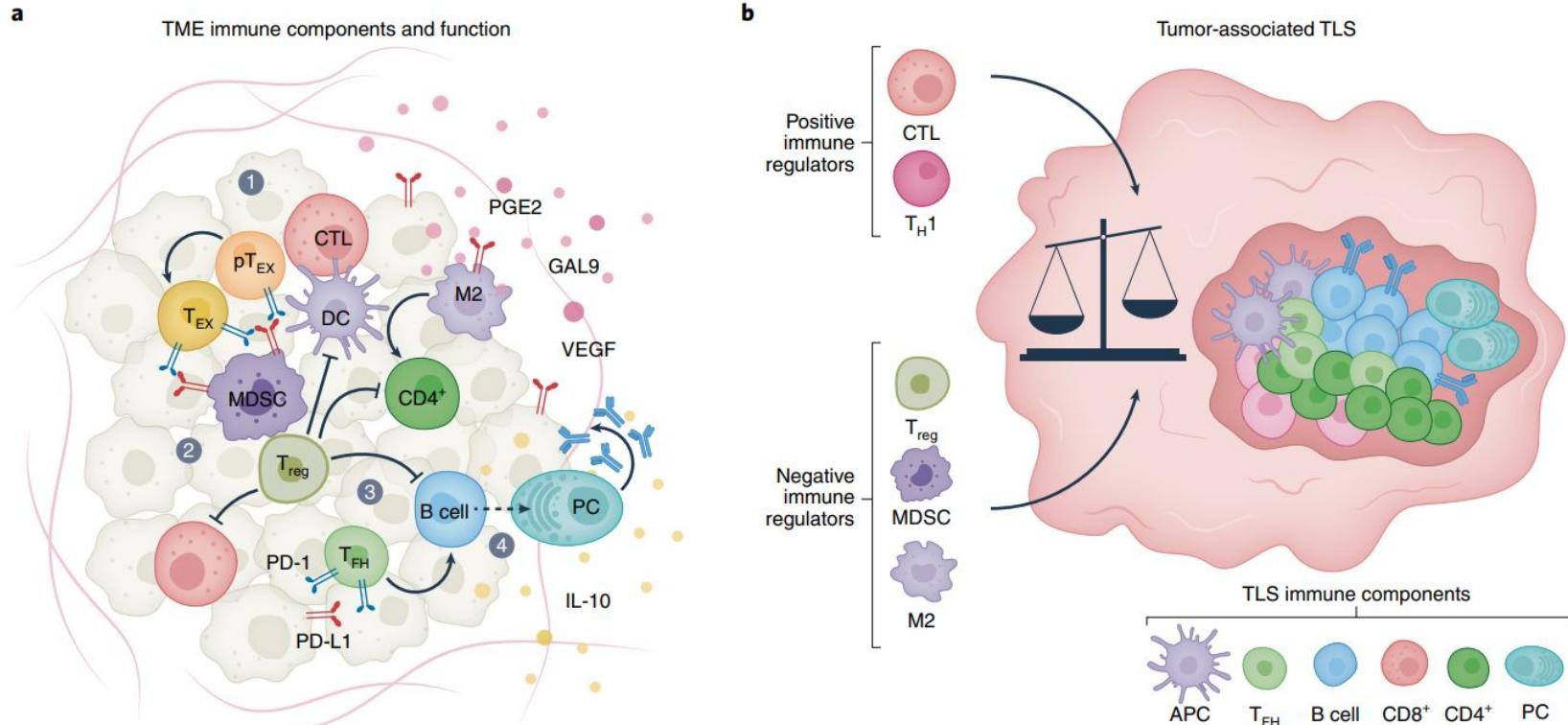
Composition and activity of immune and stromal cell types.

Clinical significance.

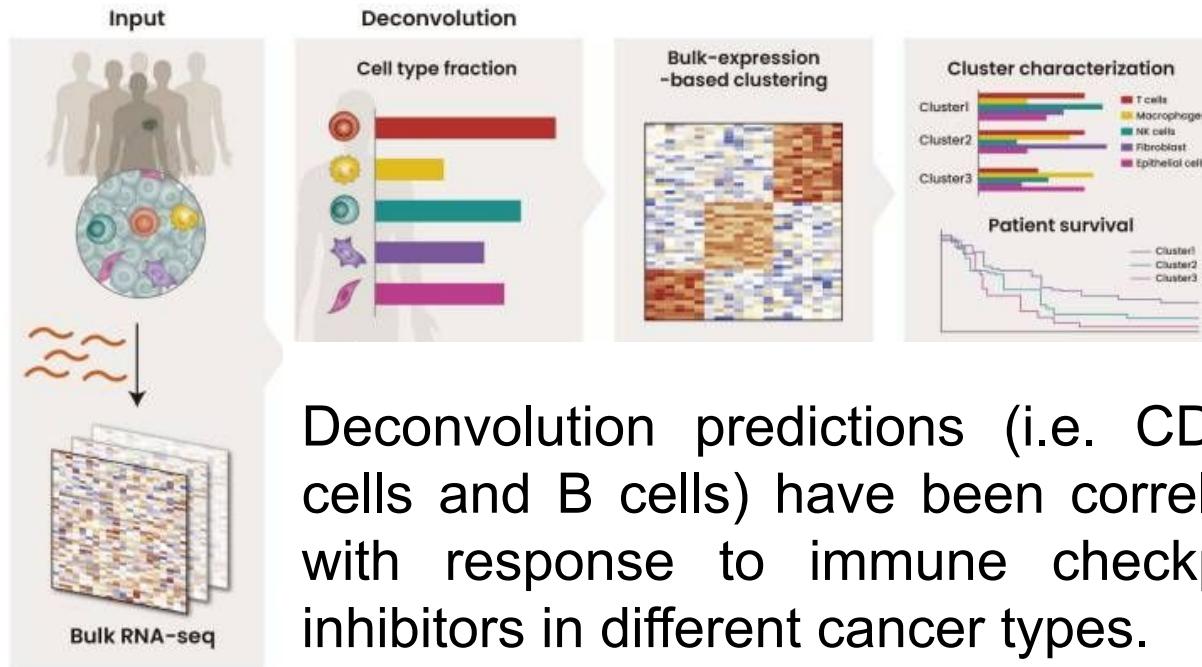


Lau AN et al., 2020

The tumour immune microenvironment - melanoma



Deconvolution to study the TME - prognostic value



Deconvolution predictions (i.e. CD8 T cells and B cells) have been correlated with response to immune checkpoint inhibitors in different cancer types.

Benchmarking studies

Sturm et al., 2019

- No 'one-size-fits-all' method.
- EPIC and quanTseq as general purpose methods.



Cobos et al., 2020

Impact of data normalization strategies on deconvolution performance.

Huuki-Myers et al., 2023

- Mean Ratio in DeconvBuddies package
- Analyzed post-mortem brain tissue
- Impact of library prep (polyA vs RiboZero)
- Bisque and hspe were the top two performers, followed by CIBERSORTx



Finotello et al., 2024

- Impact of sources of variability: single-cell reference size, cellular resolution, cell-type specific mRNA bias, unknown cellular content, single-cell technology and tissue context.
- DWLS and Scaden for single-cell-informed deconvolution.



White et al., 2024

- Levels of most major immune and stromal lineages were well predicted.
- High variability for sub-populations.
- CIBERSORTx exhibited a strong performance.

Mathematical approaches and marker gene selection strategies

Method	Citation	Approach	Marker Gene Selection	Availability	Top Benchmark Performance
DWLS (Dampened weighted least-squares)	Tsoucas et al, Nature Comm, 2019 [5]	weighted least squares	-	R package on CRAN	Cobos et al. [18]
Bisque	Jew et al, Nature Comm, 2020 [6]	Bias correction: Assay	-	R package on GitHub	Dai et al. [17]
MuSiC (Multi-subject Single-cell)	Wang et al, Nature Communications, 2019 [7]	Bias correction: Source	Weights Genes	R package GitHub	Jin et al. [20]
BayesPrism	Chu et al., Nature Cancer, 2022 [8]	Bayesian	Pairwise t-test	Webtool R package on GitHub	Hippen et al. [22]
hspe (dtangle) (hybrid-scale proportion estimation)	Hunt and Gagnon-Bartsch, Ann. Appl. Stat. 2021 [9, 44]	High collinearity adjustment	Multiple options-default "ratio" 1vALL mean expression ratio	R package on GitHub	Dai et al. [17]
CIBERSORTx	Newman et al., Nat Biotech, 2019 [11]	Machine Learning	Differential Gene expression	Webtool, Docker Image	Jin et al. [20]

Huuki-Myers et al., 2023
Pre-print

First and second generation methods

CIBERSORT
EPIC
xCell
quantiSeq
TIMER
MCP-counter

Pre-computed expression
signatures cover limited cell types.

CIBERSORTx
BayesPrism
MuSic
MuSic2
Bisque
Scaden (deep learning)

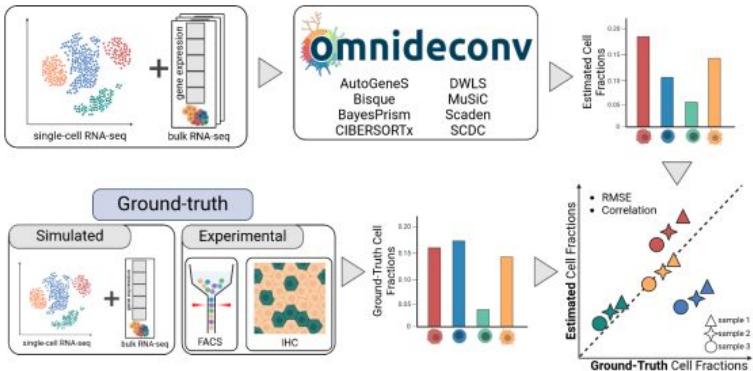
Provide more flexibility.
Use of scRNA-seq to build custom signatures.

Omnideconv

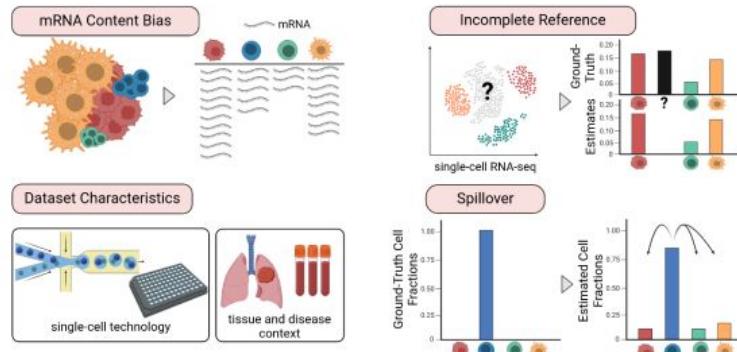
A) Omnidecconv ecosystem



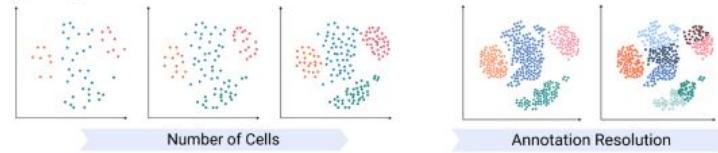
B) Benchmarking framework



C) Challenges addressed



D) Single-cell reference characteristics



Streamline deconvolution analysis and benchmarking

<https://omnideconv.org/>

<https://exbio.wzw.tum.de/deconvexplorer/>

The screenshot shows the homepage of the OMNIDECONV website. At the top, there is a navigation bar with the text "OMNIDECONV" and a "MENU" button. Below the navigation bar, a banner features a dark background with several cells stained in purple and green. The text "ENABLING THE DECONVOLUTION OF ANY CELL TYPE, TISSUE, AND ORGANISM" is displayed over the banner. A large blue section below the banner contains the text "WELCOME TO OMNIDECONV!". Below this, a paragraph describes the ecosystem as "an ecosystem of user-friendly tools and resources for cell-type deconvolution". At the bottom, a note credits the development to two groups: "The omnideconv ecosystem has been developed in a joint effort of the Group of Computational Biomedicine (University of Innsbruck, Austria) and the Group of Data Science in Systems Biology (TUM School of Life Sciences, Germany)". To the right of the text, there is a circular logo featuring a stylized cell with multiple colored protrusions (orange, green, red, blue) surrounding a central white circle with a black outline.

OMNIDECONV

ENABLING THE DECONVOLUTION OF ANY CELL TYPE, TISSUE, AND ORGANISM

MENU

WELCOME TO OMNIDECONV!

omnideconv is an ecosystem of user-friendly tools and resources for cell-type deconvolution.

The omnideconv ecosystem has been developed in a joint effort of the Group of Computational Biomedicine (University of Innsbruck, Austria) and the Group of Data Science in Systems Biology (TUM School of Life Sciences, Germany).

14

Deconvolution DREAM Challenge

- Community assessment of deconvolution methods
- 28 community-contributed methods
- Estimated cell fractions vs ground truth
- Coarse vs fine grained subchallenges
- Challenges across methods
- Specificity, sensitivity, LOD, computational efficiency
- Per cell-type top performers

<https://dreamchallenges.org/>

Mouse transcriptomics deconvolution

Bioinformatics Advances, 2024, 00, vbae032
https://doi.org/10.1093/bioadv/vbae032
Advance Access Publication Date: 28 February 2024
Application Note



Immunoinformatics

Making mouse transcriptomics deconvolution accessible with immunedeconv

Lorenzo Merotto¹, Gregor Sturm^{2,3}, Alexander Dietrich⁴, Markus List¹,
Francesca Finotello¹,*

¹Department of Molecular Biology, Digital Science Center (DiSC), University of Innsbruck, Innsbruck 6020, Austria

²Biocenter, Institute of Bioinformatics, Medical University of Innsbruck, Innsbruck 6020, Austria

³Boehringer Ingelheim International Pharma GmbH & Co KG, Biberach 88400, Germany

⁴Data Science in Systems Biology, TUM School of Life Sciences, Technical University of Munich, Freising 85354, Germany

*Corresponding author: Institute of Molecular Biology, University of Innsbruck, Technikerstraße 25, 6020 Innsbruck, Austria; Digital Science Center (DiSC), University of Innsbruck, Inrain 15, 6020 Innsbruck, Austria. E-mail: francesca.finotello@uibk.ac.at

Associate Editor: Guoqiang Yu

Abstract

Summary: Transcriptome deconvolution has emerged as a reliable technique to estimate cell-type abundances from bulk RNA sequencing data. Unlike their human equivalents, methods to quantify the cellular composition of complex tissues from murine transcriptomics are sparse and sometimes not easy to use. We extended the immunedeconv R package to facilitate the deconvolution of mouse transcriptomics, enabling the quantification of murine immune-cell types using 13 different methods. Through immunedeconv, we further offer the possibility of tweaking cell signatures used by deconvolution methods, providing custom annotations tailored for specific cell types and tissues. These developments strongly facilitate the study of the immune-cell composition of mouse models and further open new avenues in the investigation of the cellular composition of other tissues and organisms.

Availability and implementation: The R package and the documentation are available at <https://github.com/omnideconv/immunedecov>.

1 Introduction

Bulk RNA sequencing (RNA-seq) captures the expression pro-

2 Extension of immunedeconv to the analysis of mouse transcriptomics

	Method	Output	Cell Types	Custom signature
 Mouse-based methods	mMcp counter		12 immune 4 non-immune	
	seqimmucc		10 immune	
	DCQ		17 immune stromal and stem cells	
	BASE		15 immune stromal and stem cells	
 ↓  Human-based methods	quanTseq		10 immune Unknown content	
	EPIC		6 immune 2 non-immune Unknown content	
	CIBERSORT		22 immune	
	TIMER		6 immune	
	xCell		64	
	MCP counter		8 immune 2 non-immune	
	ABIS		17 immune	
	ConsensusTME		15 immune 2 non-immune	
	Estimate		Immune, stromal and tumor score Tumor purity	

 : cell-type scores  : cell-type fractions

Factors that may influence the performance of deconvolution.

Cell-type specific markers

- Genes that are able to distinguish cell types with high specificity.
- There is a need to refine cell type markers.
- Single-cell characterization of rare cell types.

Granularity

Coarse-grained vs fine-grained



Broad cell types



Finer subsets and states
(i.e. T-cell subsets)



Pre-processing of input data

- Normalization corrects for differences in the library size
- Logarithmic transformation of the input matrices can result in worse performance with the exception of BisqueRNA and bseq-sc.
- Keeping input data in linear scale aids in assessing the cell proportions accurately.

Avila Cobos et al. 2020; Jew et al. 2020, Jin & Liu 2021.

Other considerations

- Batch effect (i.e. variation between sc reference and bulk).
- Deconvolution performance differs by cell type
- Specificity (i.e. ‘spillover’)
- Assumptions
- Computational efficiency
- Underrepresentation of a cell type in the sc reference
- Tissue and disease context
- Robust and accurate estimations of cellular composition in a tumor, particularly when a cell type is well characterized.

The three pieces we need to deconvolve RNA-seq

1. Bulk expression matrix
2. Gene expression single-cell reference
3. A deconvolution method

Hands-on

For this practical, we will work with Pancreatic Ductal Adenocarcinoma (PDAC) bulk data from the TCGA project (downsampled) and a PDAC scRNA-seq reference (Steegle et al., 2020).

Connect to your VM

1. Copy the deconvolution_data folder from Penelope Cloud
2. Locate the input .rds and .Rmd files
3. Open Rstudio and the “sc_deconvolution.Rmd” file
4. Set your working directory

Note: All required R libraries and input data are pre-installed on your virtual machine. The input data is also available in the GitHub repository: <https://github.com/estefvazquez/scDeconvolution>

Load libraries

```
Library(BayesPrism)
Library(BisqueRNA)
Library(BioBase)
Library(Immunedeconv)
Library(dplyr)
Library(tidyR)
```

Load input data

```
bulkdata <- tcga_pdac_counts.rds  
bulk_tpms <- tcga_pdac_tpms.rds  
clindata <- tcga_pdac_clindata.rds  
Seurat object sobj <- sc_ref_pdac_seurat.rds
```

Exercise 1

Run BayesPrism

1. Prepare data
2. Run deconvolution
3. Explore output

Execute the code interactively, line by line, to understand what each command does. Once you have completed running all the chunks, you can generate an HTML report by knitting the document. Got questions? We are here to help!

Exercise 2

Run Bisque

1. Convert bulk RNA-seq from a matrix to an ExpressionSet

```
bulk.eset <- Biobase::ExpressionSet(assayData = bulk.matrix)
```

2. Run decomposition
3. Explore output

Execute the code interactively, line by line, to understand what each command does. Once you have completed running all the chunks, you can generate an HTML report by knitting the document. Got questions? We are here to help!

Exercise 3

Run CIBERSORTx

To begin, access your CIBERSORTx account: <https://cibersortx.stanford.edu/>

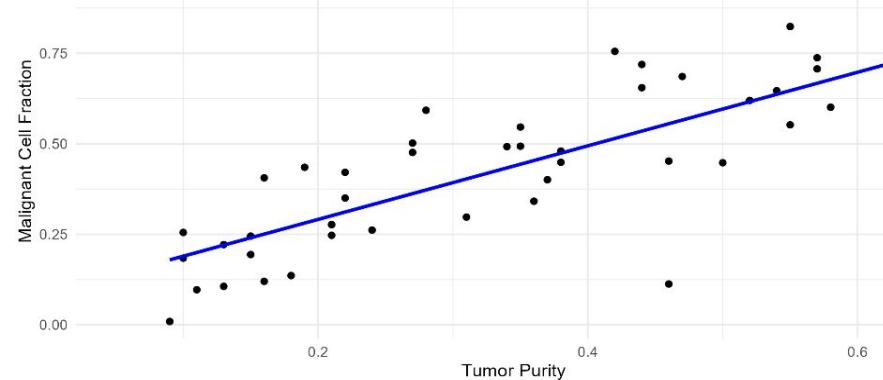
For this demonstration we will use the LM22 signature (Newman et al., 2015).

The screenshot shows the CIBERSORTx website interface. At the top, there is a navigation bar with links for Home, About, Contact, Register, Email (highlighted with a red box), Password, Sign in, and Remember Me. Below the navigation bar, a message in a pink box states "User data generated before 8/15/23 currently not available." The main content area features the CIBERSORTx logo and a brief description of the tool: "CIBERSORTx is an analytical tool from the Alizadeh Lab and Newman Lab to impute gene expression profiles and provide an estimation of the abundances of member cell types in a mixed cell population, using gene expression data." A green callout box highlights the "EcoTyper" feature: "EcoTyper, a machine learning framework for the identification of cell states and ecosystems from bulk, single-cell, and spatially-resolved expression data, is now available. EcoTyper extends CIBERSORTx for large-scale profiling of cellular ecosystems." At the bottom, a yellow box contains the text: "The CIBERSORT legacy website has been incorporated into the CIBERSORTx website under the menu item CS Archives."

Correlating deconvolution results with tumor purity

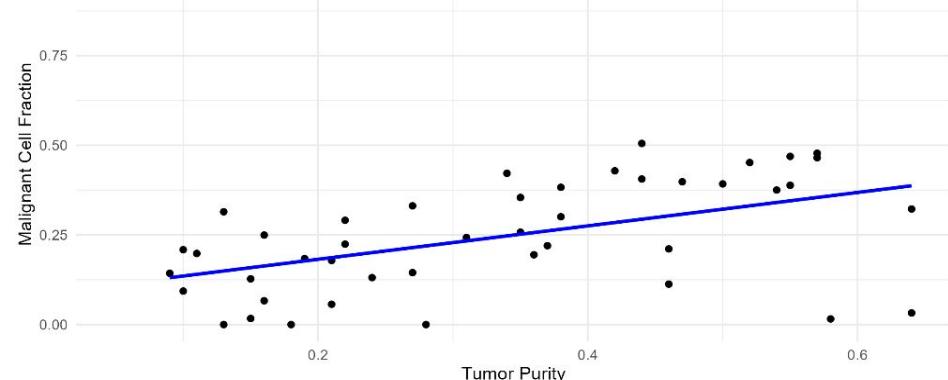
Correlation between Tumor Purity and Malignant Cell Fraction - BayesPrism

$R = 0.8, p = 9e-11$



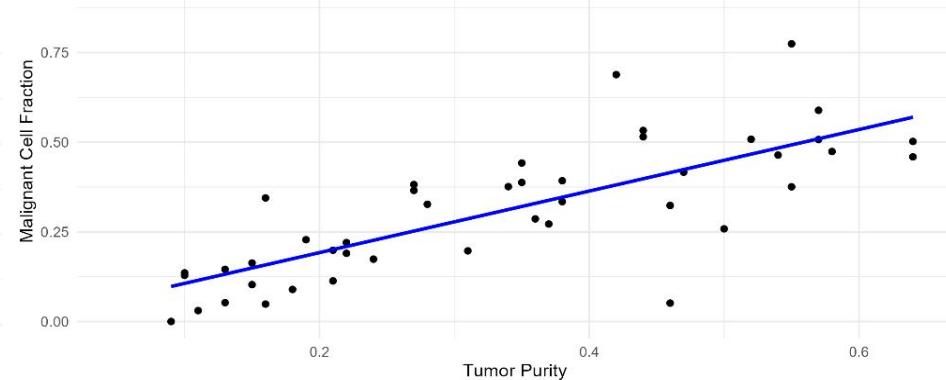
Correlation between Tumor Purity and Malignant Cell Fraction - BisqueRNA

$R = 0.51, p = 0.00039$



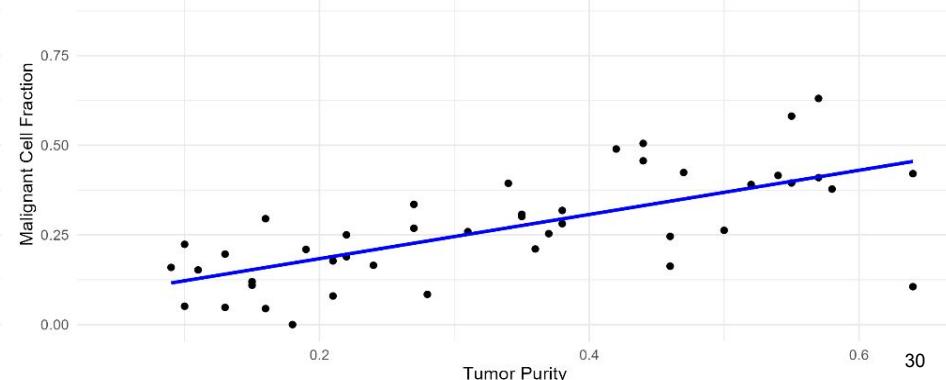
Correlation between Tumor Purity and Malignant Cell Fraction - BayesPrism (Top25 markers)

$R = 0.77, p = 7.6e-10$

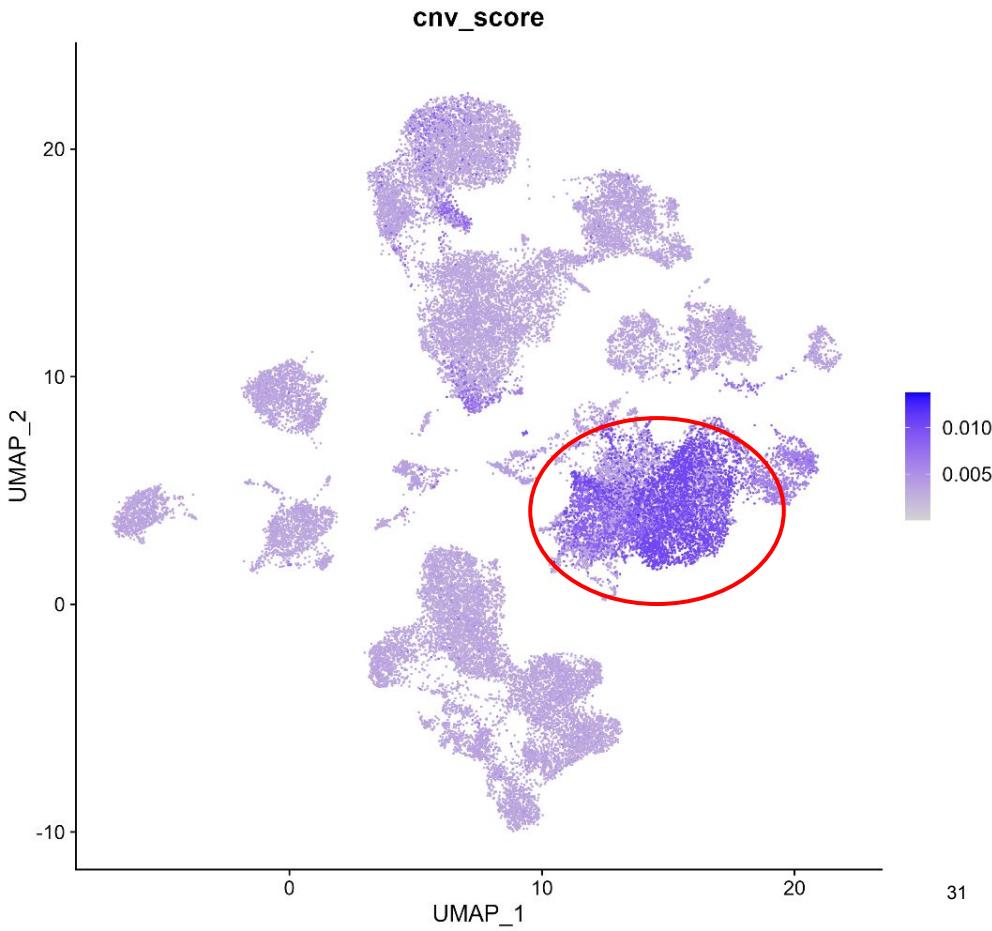
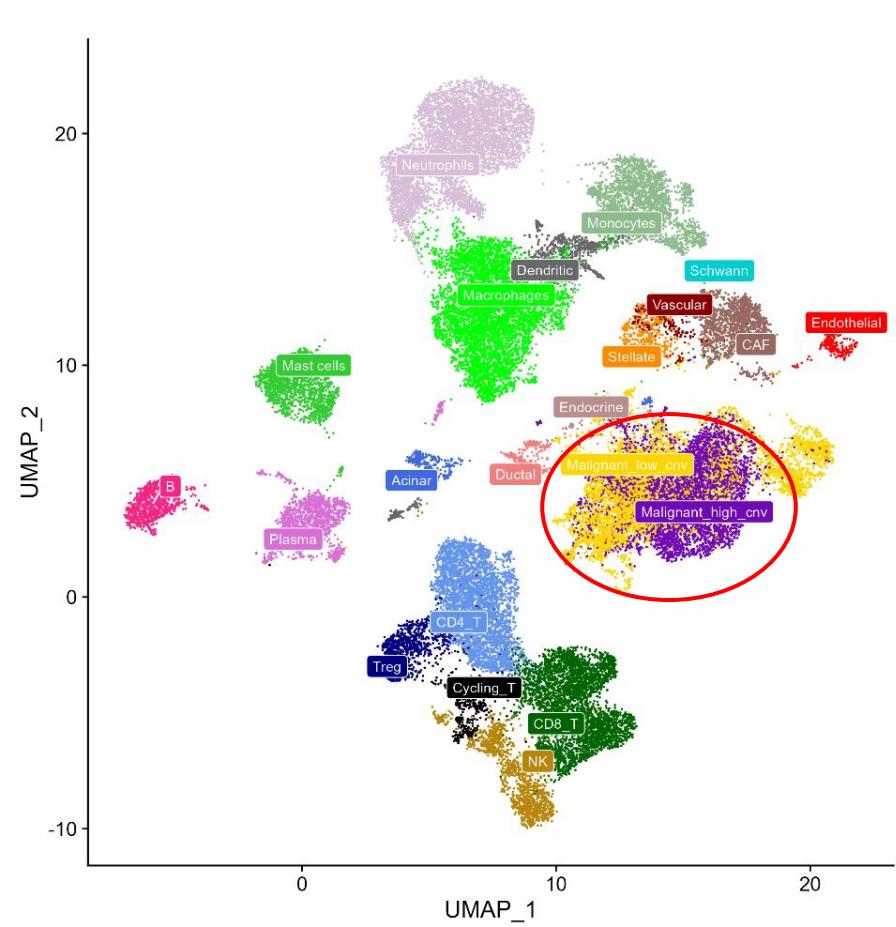


Correlation between Tumor Purity and Malignant Cell Fraction - BisqueRNA (Top25 markers)

$R = 0.69, p = 2.8e-07$

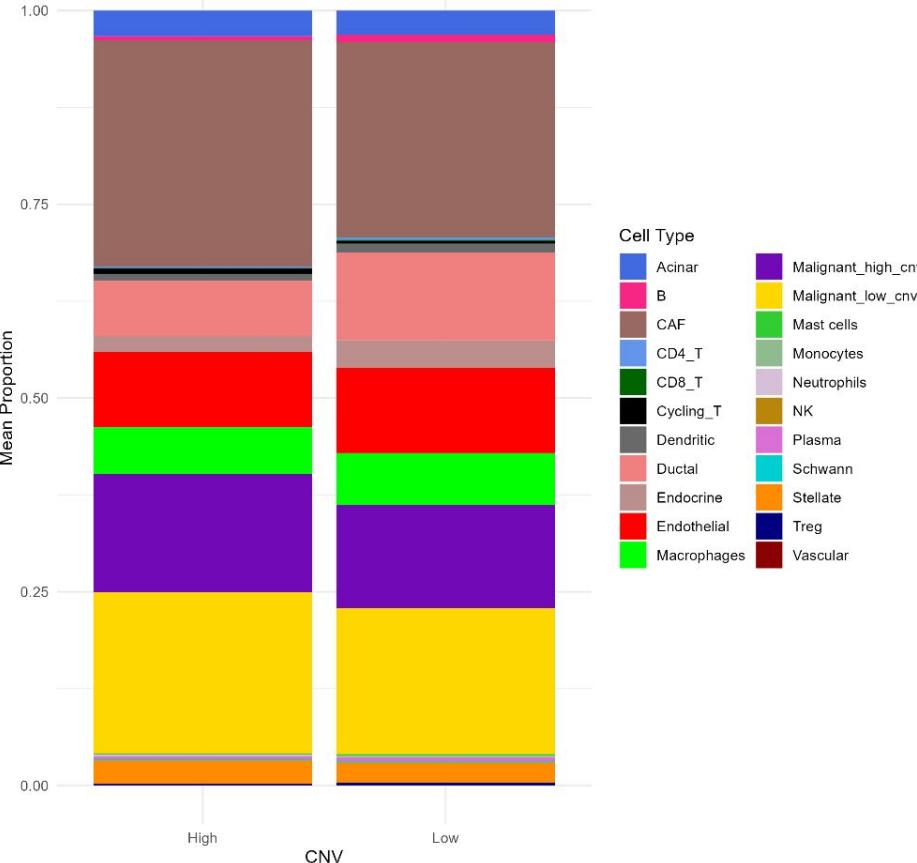


Correlating deconvolution results with CNV

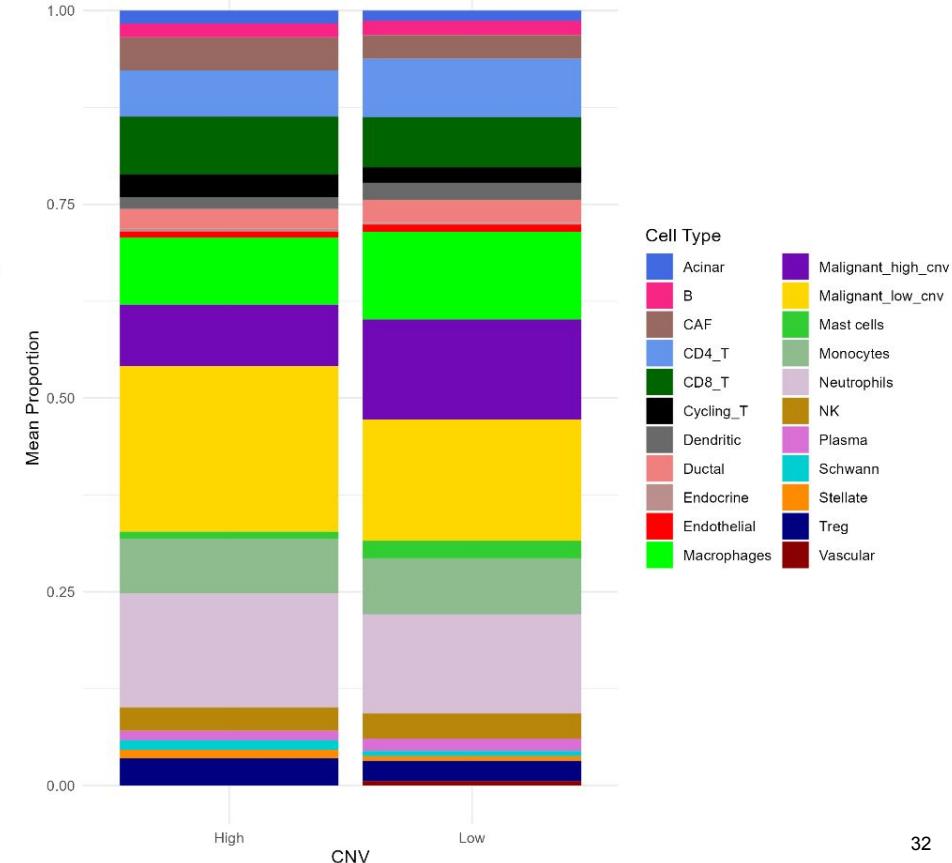


Correlating deconvolution results with CNV

Cell Proportions by CNV class (Top25 markers) - BayesPrism

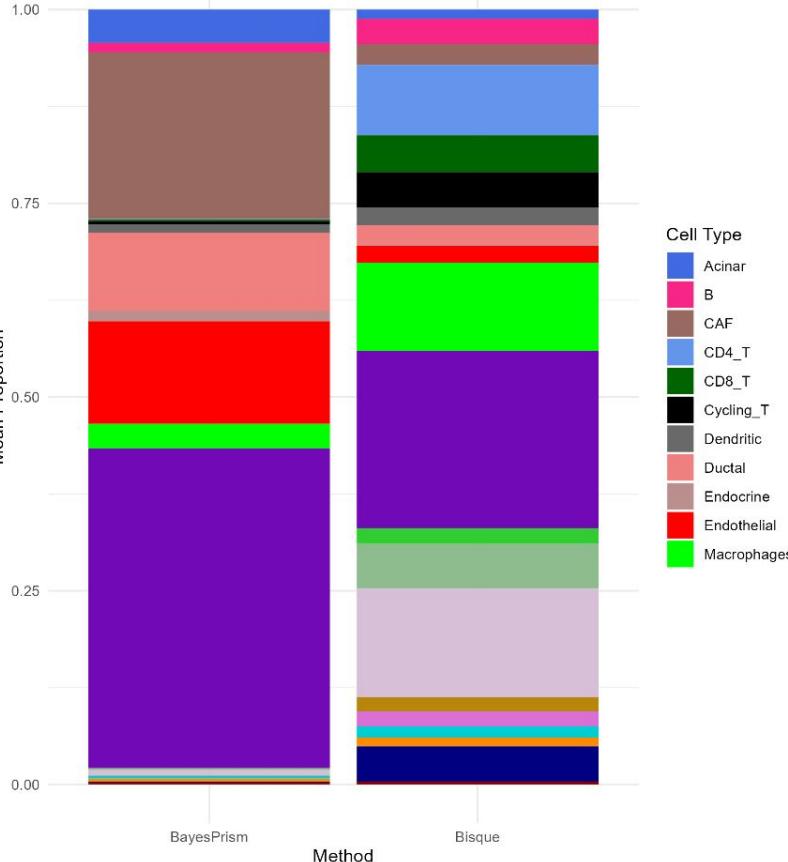


Cell Proportions by CNV class (Top25 markers) - BisqueRNA

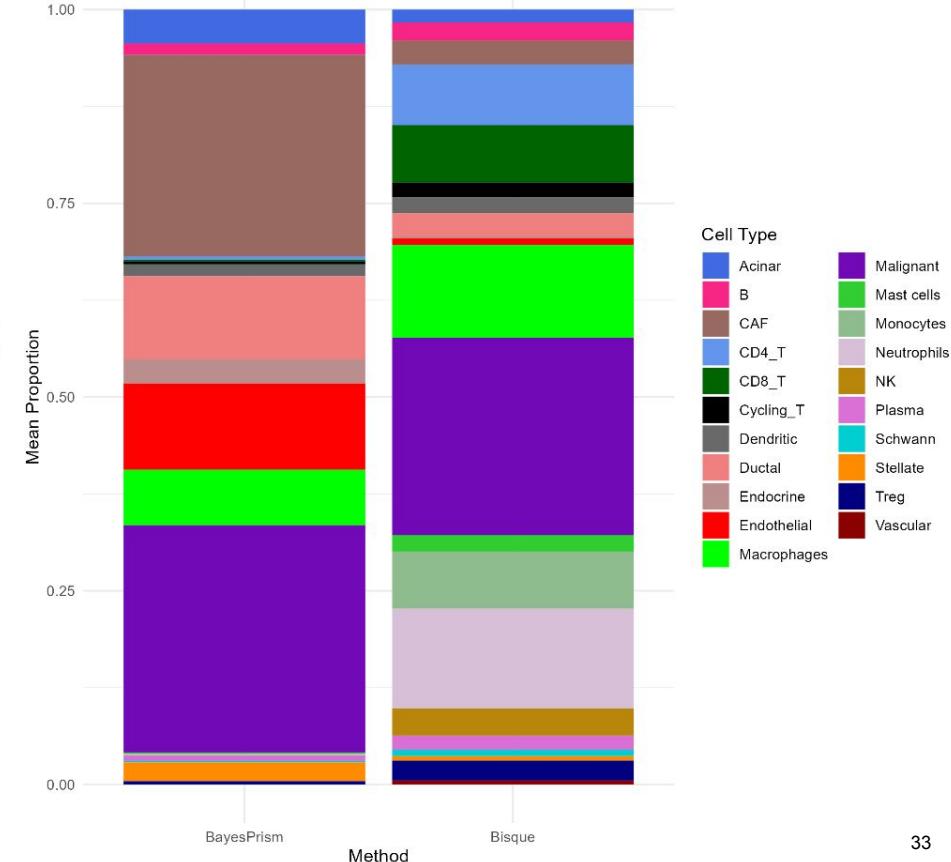


Comparing deconvolution results between methods

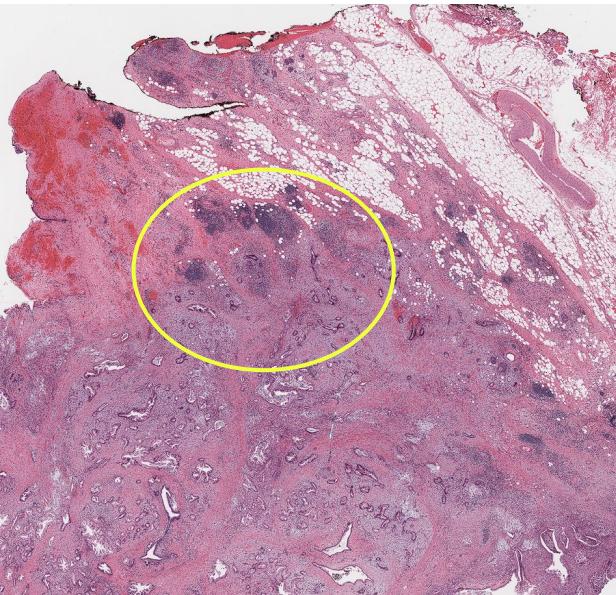
Cell Proportions by Method



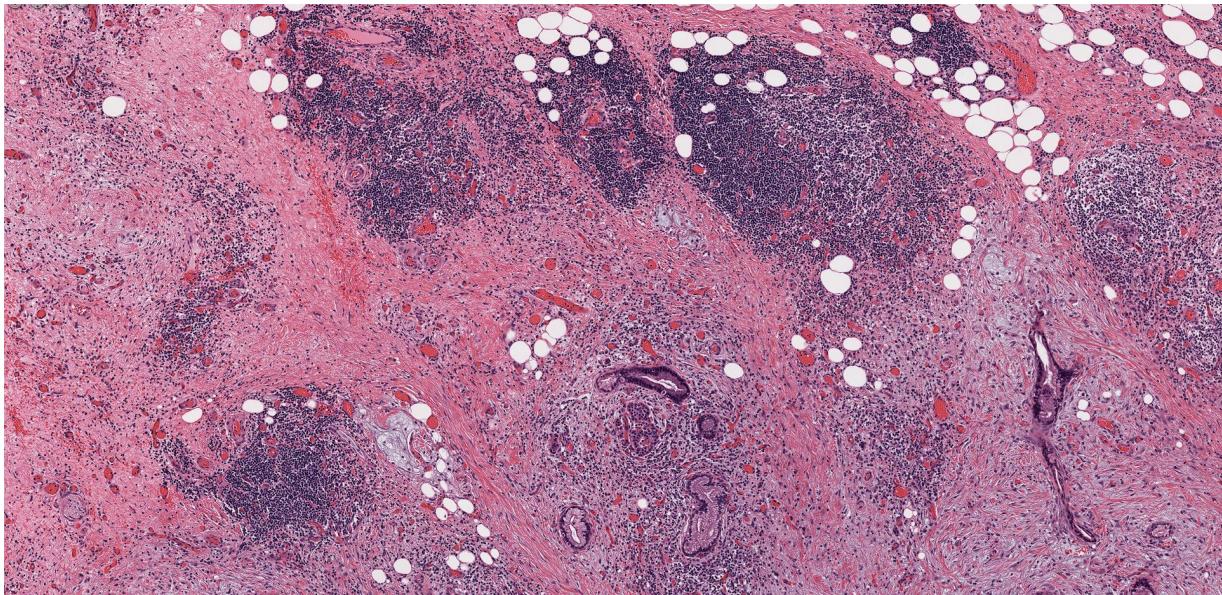
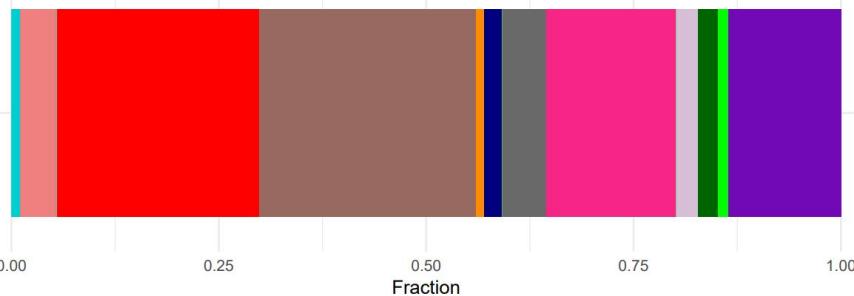
Cell Proportions by Method (Top25 markers)



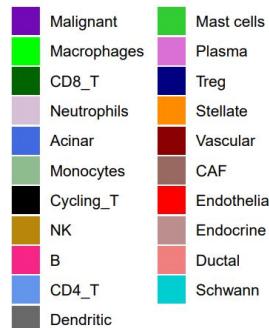
Correlating deconvolution results with H&E slides



Composition of TCGA-IB-AAUR-01A-21R-A38C-07



variable



Team challenge (5 breakout rooms)

Each team will present a maximum of 5 slides, showcasing one research question and conducting a downstream analysis using the deconvolution results.

Examples:

- Cluster bulk samples
- Correlate with clinical covariates
- Correlate deconvolution results with tumor purity
- Z-scores of signature genes of the cell type of interest
- Survival analysis
- How gene expression in malignant cells correlates with the cell type fraction of non-malignant cells

Exercise 4

Final Quiz

- . How can I validate my deconvolution results?
- . What is the spillover effect?
- . How can I deal with rare cell types or cell states?
- . What level of cell annotation (resolution) should I use?

Find more resources at the github repository:

<https://github.com/estefvazquez/scDeconvolution>

Acknowledgments

EMBL EBI Training Team

Cancer Genetics & Bioinformatics Lab - LIIGH UNAM

Immunotherapy Lab UFCSPA

EMBL-EBI

