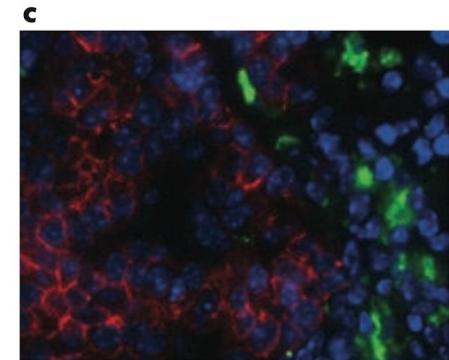
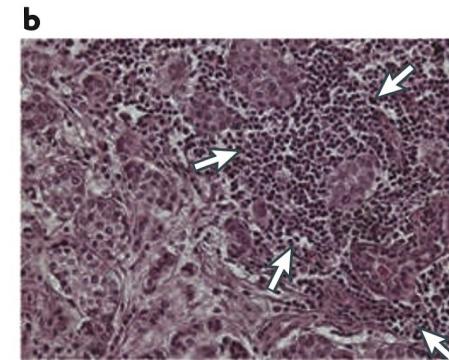
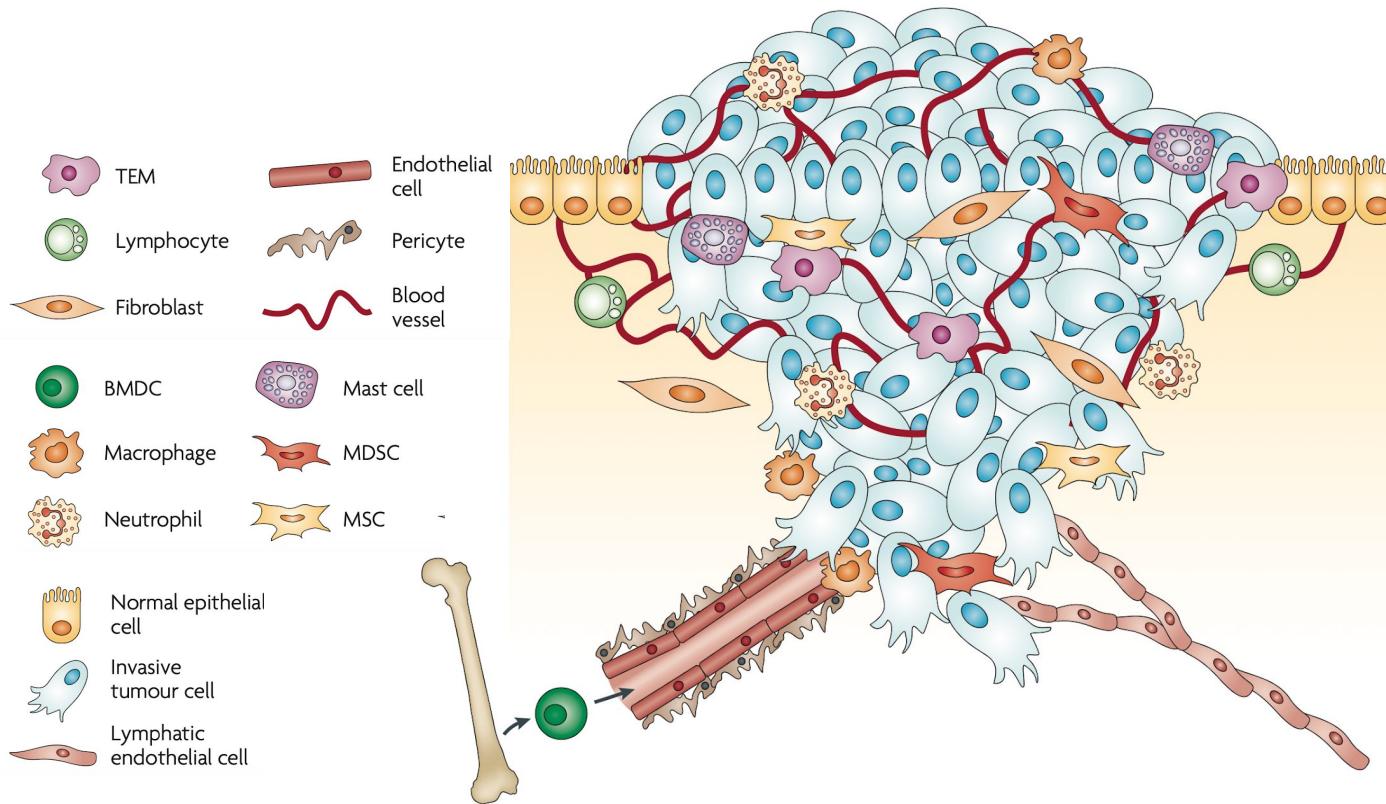


# DECONVOLUTIONAL APPROACHES TO STUDY THE TUMOR MICROENVIRONMENT

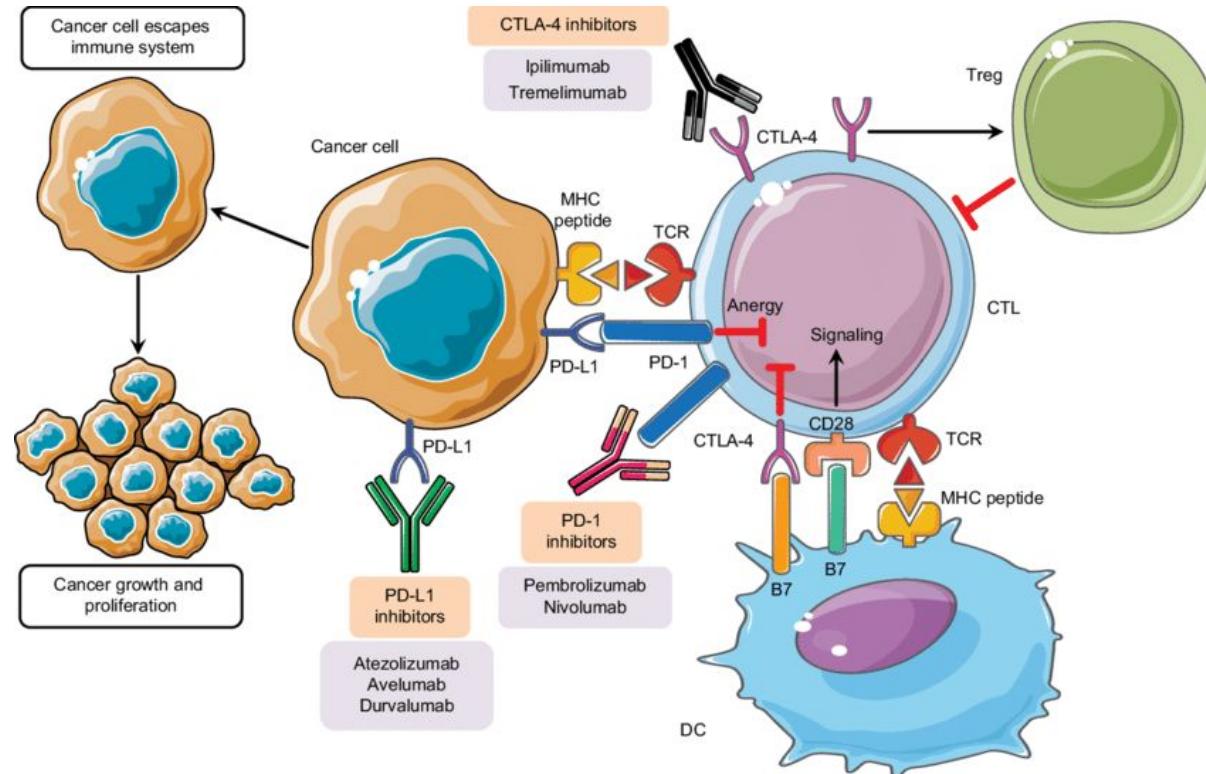
**Mariana Boroni**  
Bioinformatics and  
Computational Biology Lab  
INCA, BR

**Nyasha Chambwe**  
Feinstein Institutes for  
Medical Research, USA

# TUMOR MICROENVIRONMENT (TME)



# TUMOR-INFILTRATING IMMUNE CELLS



# HOW CAN WE STUDY SPECIFIC POPULATIONS IN THE TME?

- **Cell sorting**
  - FACS
  - CyTOF
- **Immunohistochemistry/Immunofluorescence (IHC/IF)**
  - Cell staining
- **Bulk Transcriptomics**
  - Microarrays
  - RNA-seq
- **Single cell RNA-seq**
  - Transcriptomics of single cells
- **Combinations**
  - Spatial transcriptomics

# CELL SORTING

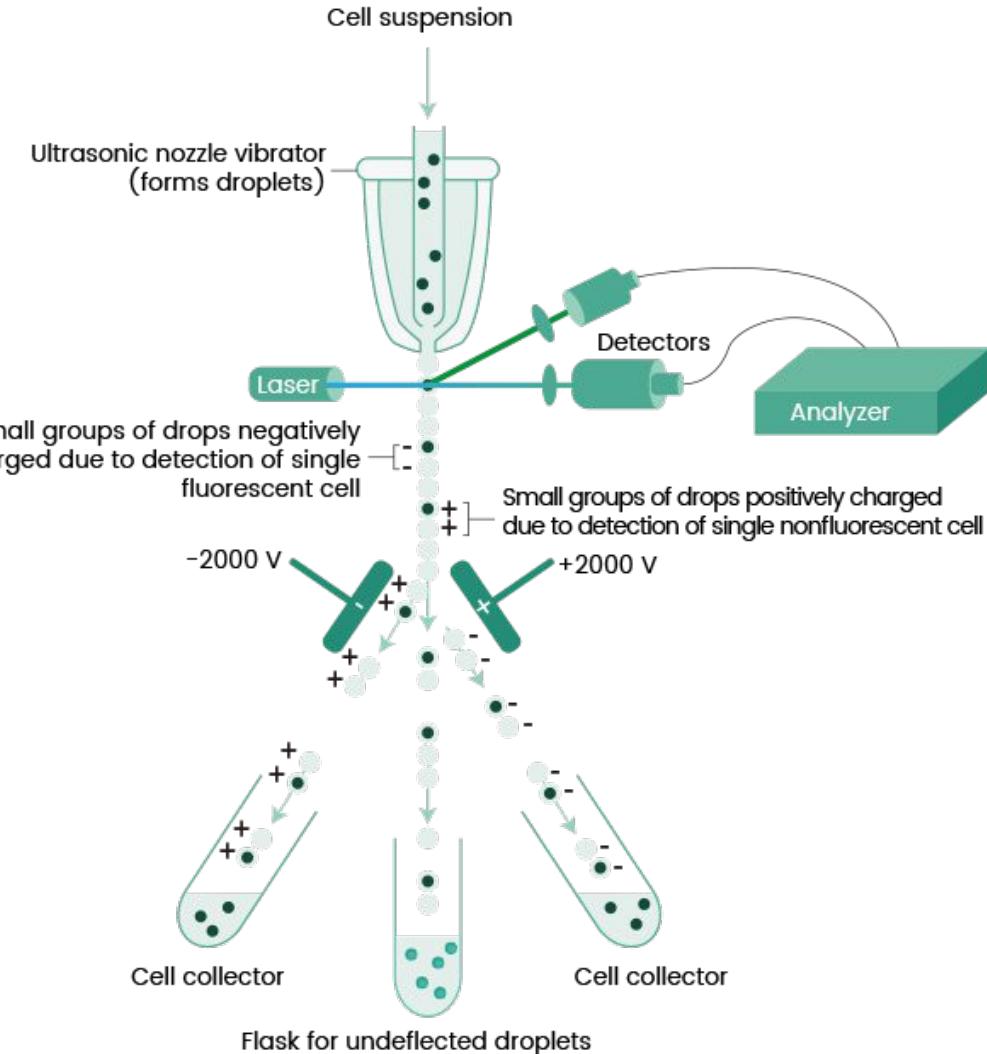
- Sort and label cells using cell type specific antigens
- Detect labels on cells
  - CyTOF – time-of-flight mass spectrometry
  - FACS – fluorescence activated cell sorting

## Pros

- Known technology, established infrastructure
- Comparatively cheap

## Cons

- Limited markers (max 50 for CyTOF)
- CyTOF antibodies are expensive
- Potential disaggregation issues



# IMMUNOHISTOCHEMISTRY/IMMUNOFLUORESCENCE

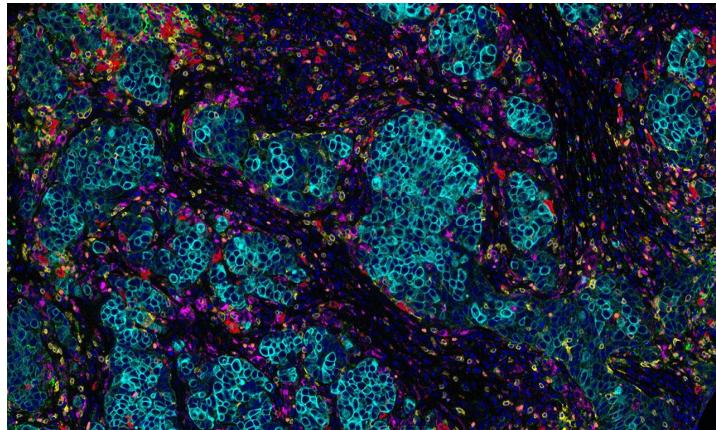
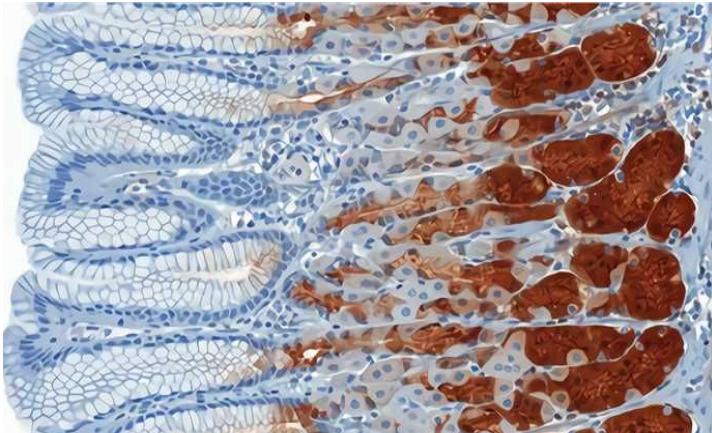
Sectioning and staining for cell type specific markers

## Pros

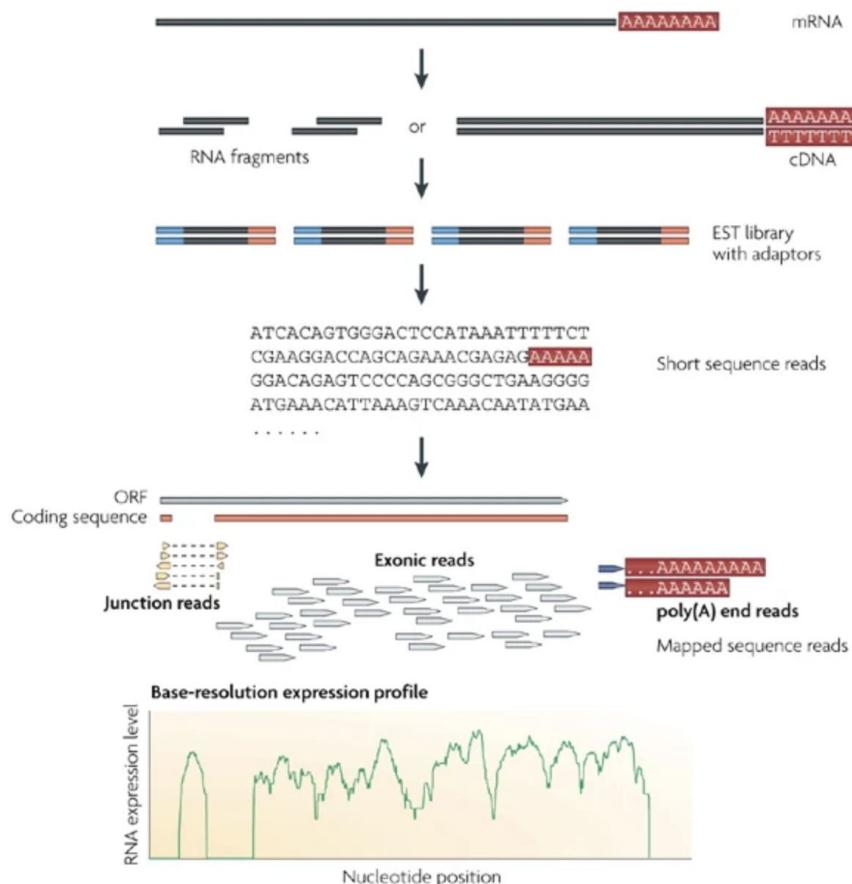
- Known technology, established infrastructure
- Comparatively cheap
- Lots of FFPE and frozen tissue samples available

## Cons

- Sections only, hard/expensive to assay entire tumor
- Limited to a few markers per section



# RNA-SEQ

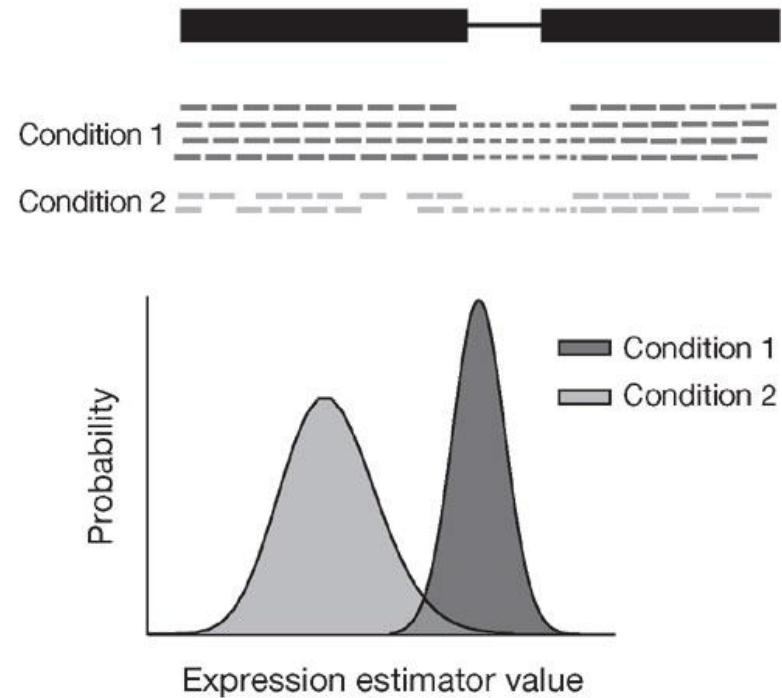


Briefly, long RNAs are first converted into a library of **cDNA fragments** through either RNA fragmentation or DNA fragmentation. **Sequencing adaptors** (blue) are subsequently added to each cDNA fragment and a short sequence is obtained from each cDNA using high-throughput sequencing technology. The resulting sequence **reads** are **aligned with the reference genome** or transcriptome, and classified as three types: exonic reads, junction reads and poly(A) end-reads. These three types are used to generate a **base-resolution expression profile** for each gene, as illustrated at the bottom; a yeast ORF with one intron is shown.

Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57–63 (2009).  
<https://doi.org/10.1038/nrg2484>

# RNA-SEQ

- ❑ Reads are aligned to the reference genome
- ❑ Gene Expression quantification based on number of reads
- ❑ Normalization to compare across conditions and replicates



# BULK TRANSCRIPTOMICS - RNA-SEQ

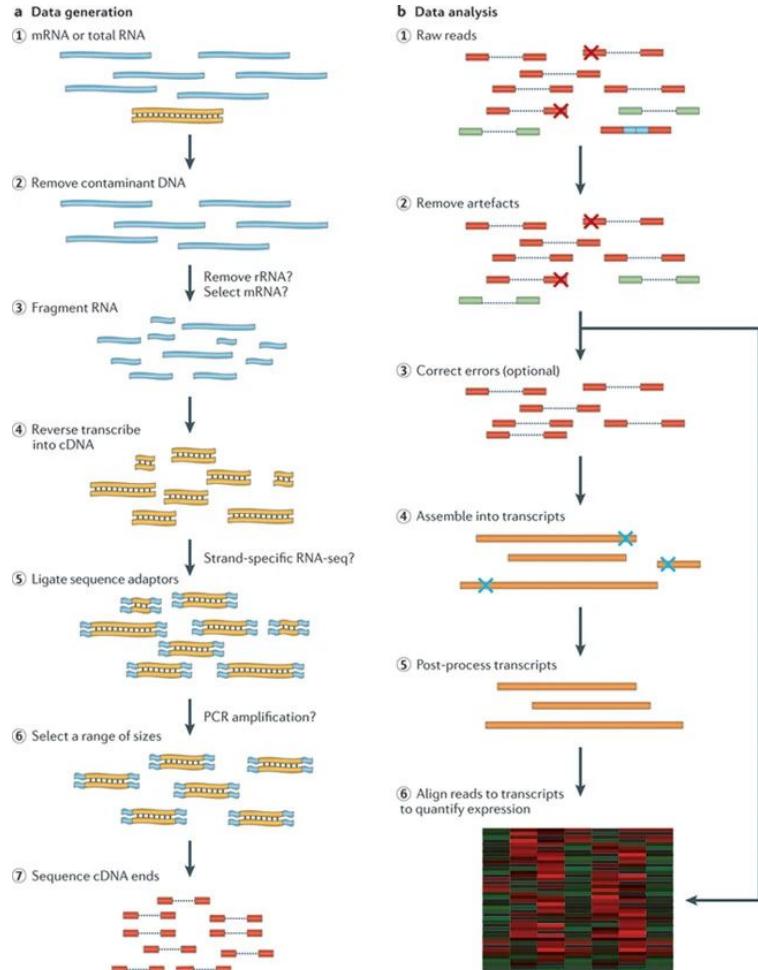
Bulk RNA-seq = all cells within mixture contribute to final expression levels

## Pros

- Can assay entire sample at once
- Can help identify transcription changes in individual cell types
- Huge amount of data out there already
- Cheap(er)

## Cons

- Hard to do well



# SINGLE-CELL TRANSCRIPTOMICS SCRNA-SEQ

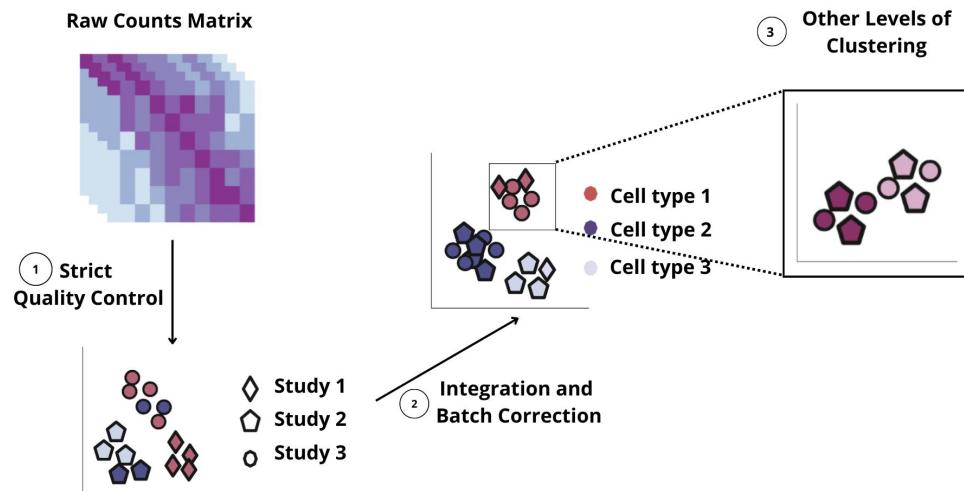
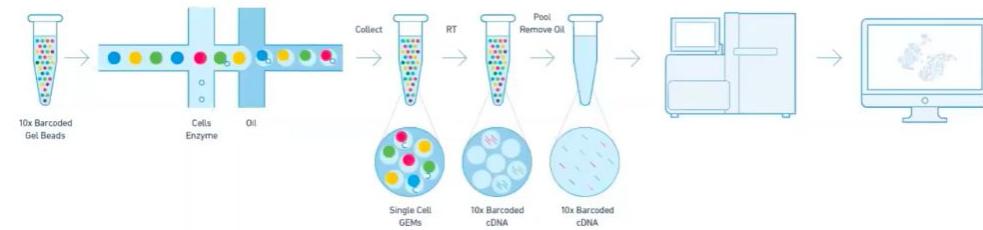
InDrops, DropSeq, SmartSeq, 10x →

## Pros

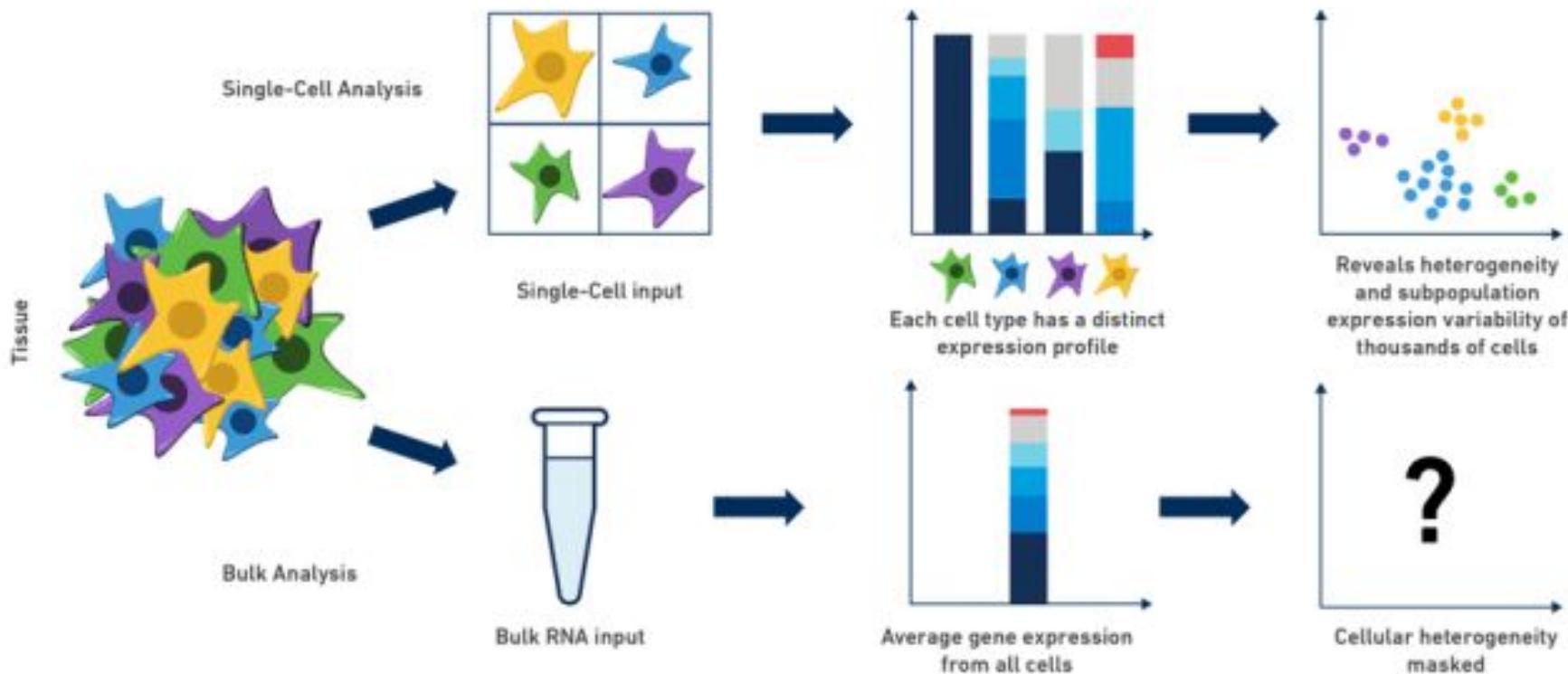
- Powerful
- Effective

## Cons

- Expensive
- Disaggregation bias
- Can't always identify the cells
- Marker issues



# BULK X SINGLE-CELL RNA-SEQ



# BULK X SINGLE-CELL RNA-SEQ

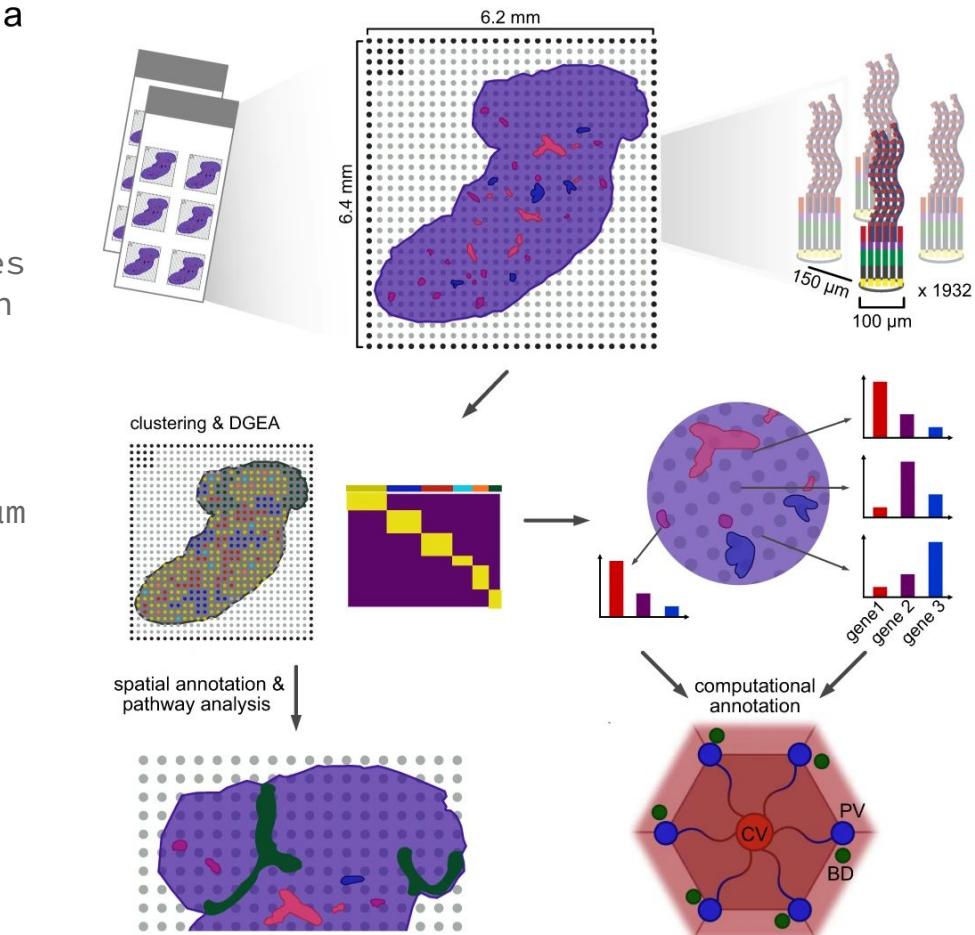
Table 1

Comparison of bulk and single-cell RNA-seq for clinical translation, specifically in onco-immunology. The advantages and limitations of combining both techniques were also explored.

	RNA-sequencing		
	Conventional bulk	Single-cell	Joint use of single-cell and bulk data
<b>Experimental and general aspects</b>			
Cost	+	+++	++++
Size of sample (minimal amount of RNA required)	+++	+	+++
Batch effect	+	++	+++
Dropout amplification/coverage bias	+	+++	++
<b>Computational aspects</b>			
Storage capacity	+	++++	++++
Handling sparsity	+	+++	++
Differentiating complex expression patterns from noise	++	+++	++
Dealing with missing data	++	+++	++
Deconvolution requirement to characterise immune infiltrate	++++	+	+
<b>Acquired information</b>			
Appreciating tissue heterogeneity at the cell level	+	++++	++++
Analysing specific populations (T cells, B cells ...)	+	+++	++++
Describing populations as a whole	++	++	++++
Accuracy of the final analysis	++	+++	++++

# SPATIAL TRANSCRIPTOMICS

- The tissue sections were placed in one of six,  $6.2 \times 6.4$  mm frames on the glass slide ST array. Each frame contains 1932 spots, with  $>200$  M uniquely barcoded mRNA capture probes. The distance between centers of each neighboring spot is  $150 \mu\text{m}$  ( $200 \mu\text{m}$  for spots in the same row).
- Initially, each tissue section was fixed, stained with hematoxylin and eosin (H&E) and followed by imaging.
- Then, tissue sections were permeabilized, followed by mRNA capture, tissue removal and sequencing.



Hildebrandt, F., Andersson, A., Saarenpää, S. et al. Spatial Transcriptomics to define transcriptional patterns of zonation and structural components in the mouse liver. *Nat Commun* 12, 7046 (2021). <https://doi.org/10.1038/s41467-021-27354-w>

# Understanding Complexity

Whole Tissue/Organs  
(Genetic) Disease Model



Complex Tissue



Spatial Transcriptomics



Bulk Genomics



Flow  
Cytometry +  
Bulk Genomics

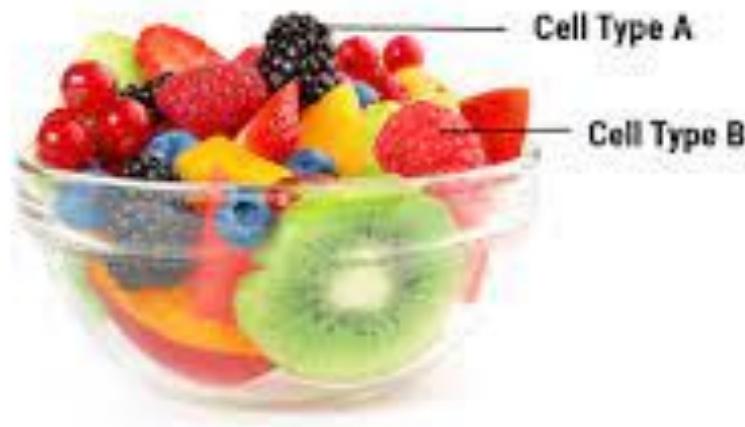
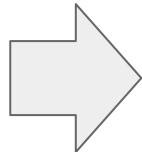


Single Cell Genomics  
(+ Cytometry)

# DECONVOLUTION METHODS – UNMIXING THE SMOOTHIE

How many strawberries, kiwis, pineapples and oranges went into the smoothie?

Can we computationally figure out what went into the mixture?



Bulk RNA Seq

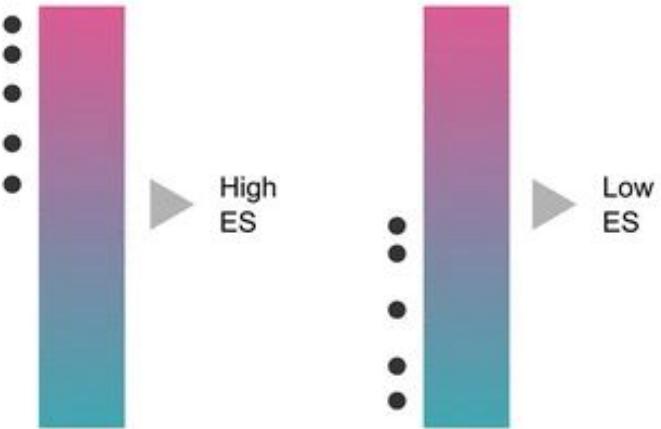
\$200/sample

scRNA Seq

\$4000–10000/sample

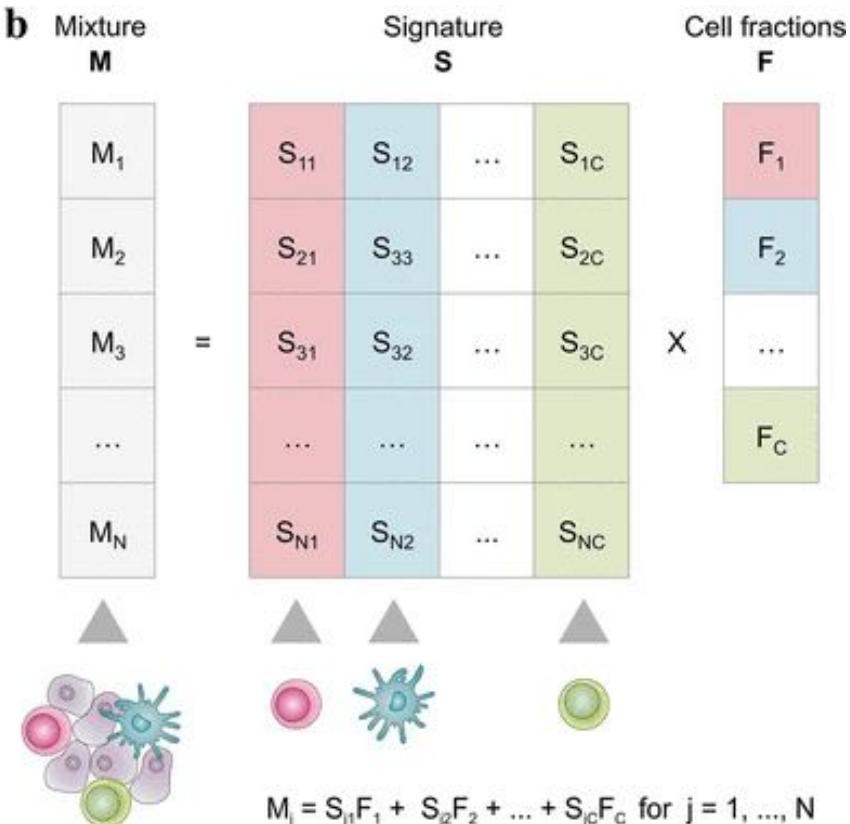
# DECONVOLUTION OF BULK RNA-SEQ

a



Enrichment score (ES) is high when the genes specific for a certain cell type are amongst the top highly expressed in the sample of interest (i.e., the cell type is enriched in the sample) and low otherwise

# DECONVOLUTION OF BULK RNA-SEQ



Deconvolution algorithms model the expression of a gene in a mixture M as a linear combination of the expression of that gene in the different cell types, whose average expression profiles are summarized in a signature matrix S, weighted by the relative fractions F of the cell types in the mixture.

# APPROACH : DECONVOLUTION OF BULK RNA-SEQ

- marker-gene-based approaches (M)
- deconvolution-based approaches (D)

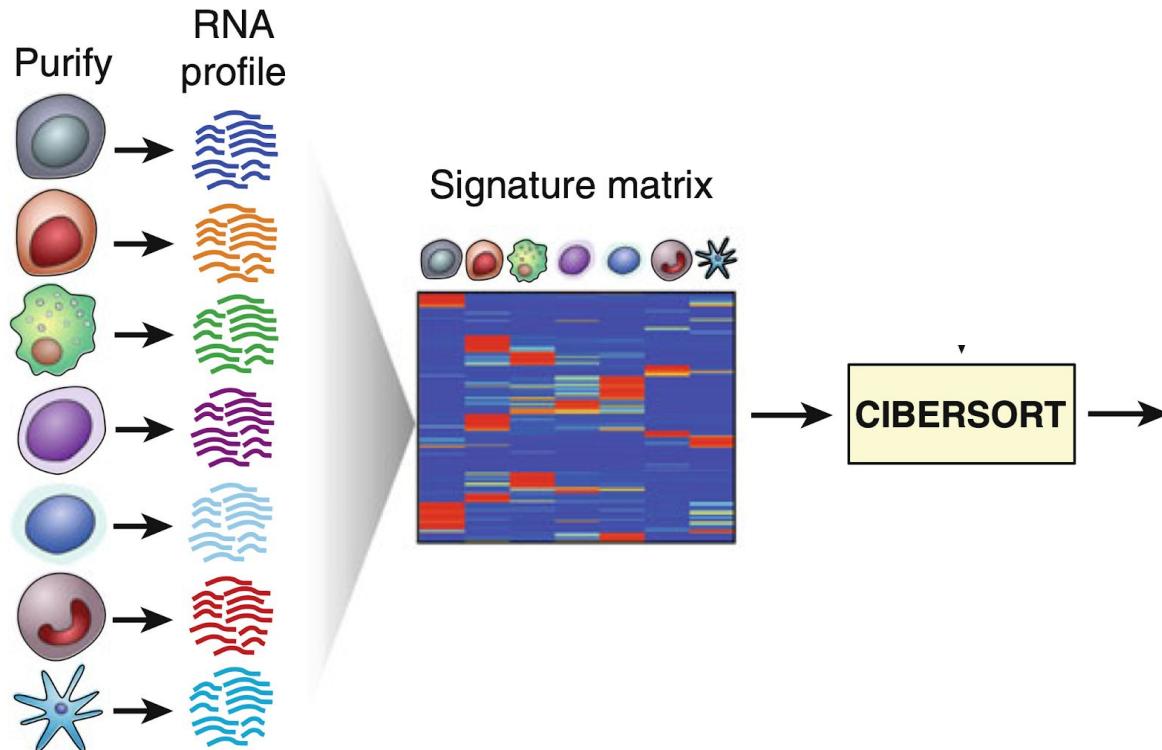
**Table 1.**

Overview of cell type quantification methods providing gene signatures for immuno-oncology

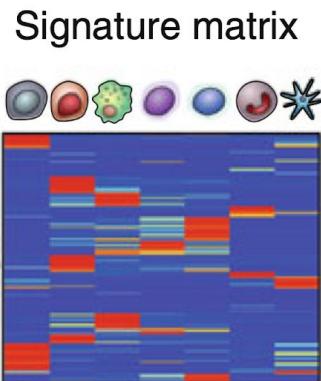
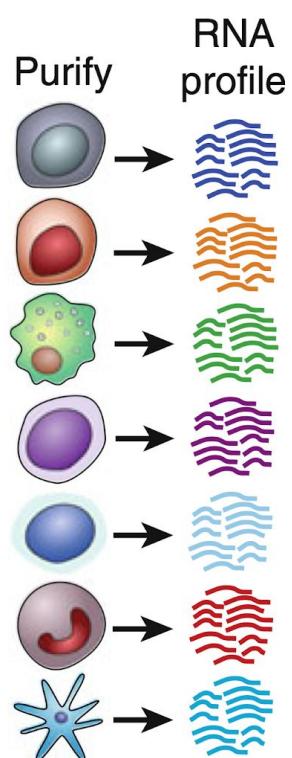
Tool	Abbrev.	Type	Score	Comparisons	Algorithm	Cell types	Reference
CIBERSORT	CBS	D	Immune cell fractions, relative to total immune cell content	Intra	v-support vector regression	22 immune cell types	Newman <i>et al.</i> (2015)
CIBERSORT abs. mode	CBA	D	Score of arbitrary units that reflects the absolute proportion of each cell type	Intra, inter	v-support vector regression	22 immune cell types	Newman <i>et al.</i> (2015, 2018)
EPIC	EPC	D	Cell fractions, relative to all cells in sample	Intra, inter	constrained least square regression	6 immune cell types, fibroblasts, endothelial cells	Racle <i>et al.</i> (2017)
MCP-counter	MCP	M	Arbitrary units, comparable between samples	Inter	mean of marker gene expression	8 immune cell types, fibroblasts, endothelial cells	Becht <i>et al.</i> (2016)
quanTseq	QTS	D	Cell fractions, relative to all cells in sample	Intra, inter	constrained least square regression	10 immune cell types	Finotello <i>et al.</i> (2017)
TIMER	TMR	D	Arbitrary units, comparable between samples (not different cancer types)	Inter	linear least square regression	6 immune cell types	Li <i>et al.</i> (2016)
xCell	XCL	M	Arbitrary units, comparable between samples	Inter	ssGSEA (Hänzelmann <i>et al.</i> , 2013)	64 immune and non-immune cell types	Aran <i>et al.</i> (2017)

Note: Methods can be conceptually distinguished in marker-gene-based approaches (M) and deconvolution-based approaches (D). The output scores of the methods have different properties and allow either intra-sample comparisons between cell types, inter-sample comparisons of the same cell type, or both. All methods come with a set of cell type signatures ranging from six immune cell types to 64 immune and non-immune cell types.

# DECONVOLUTION APPROACH WITH CIBERSORT



# DECONVOLUTION APPROACH WITH CIBERSORT



Bulk  
tissue/  
tumor OR  
Blood  
draw

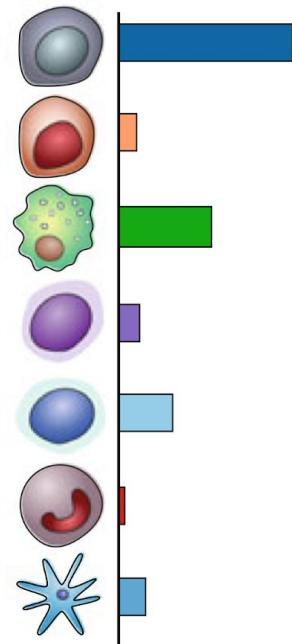


RNA profile

CIBERSORT

Significance analysis

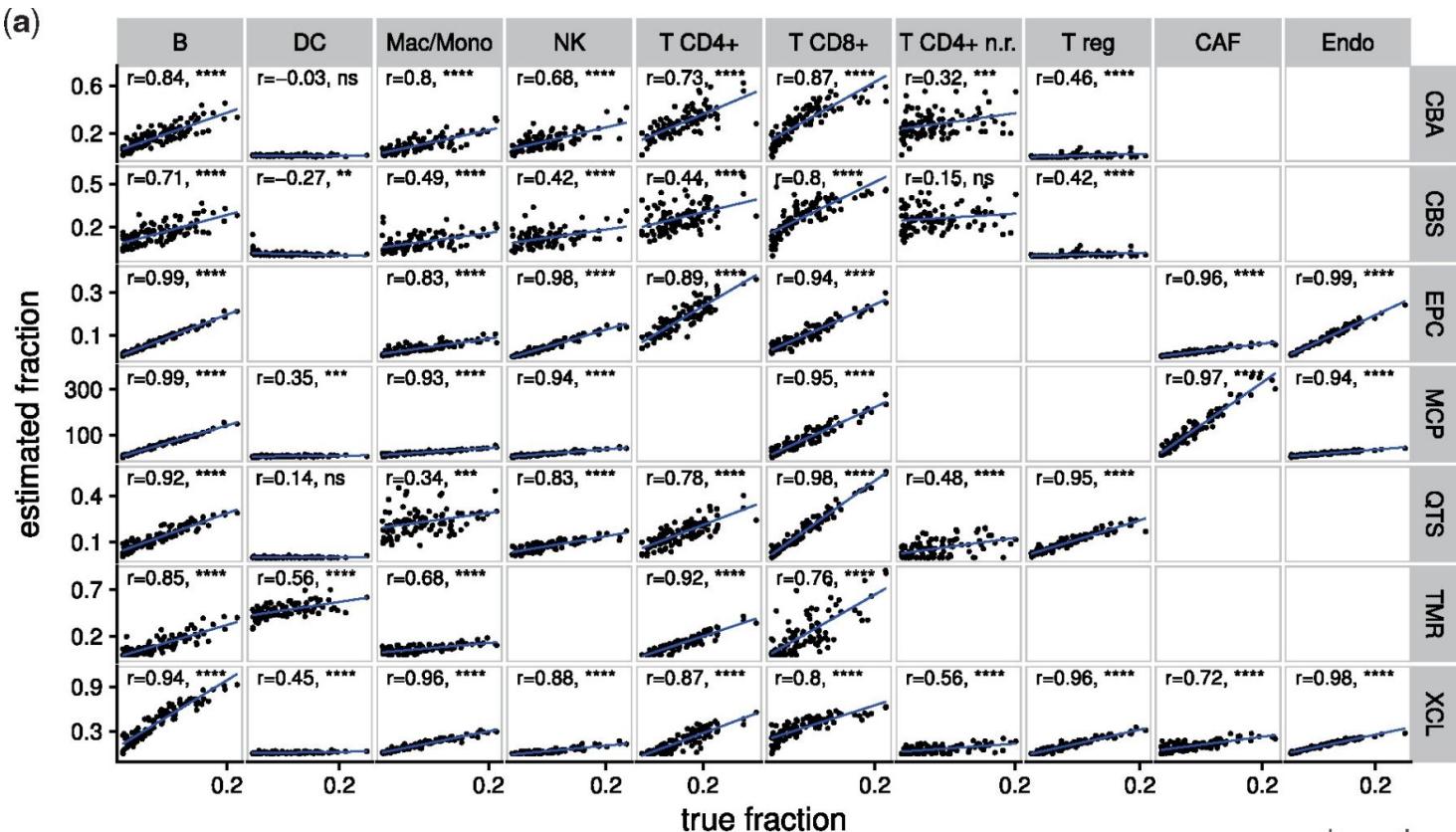
Cell proportions



# DECONVOLUTION DOESN'T ALWAYS WORK WELL

Simulated  
datasets  
drawn from  
scRNAseq  
data

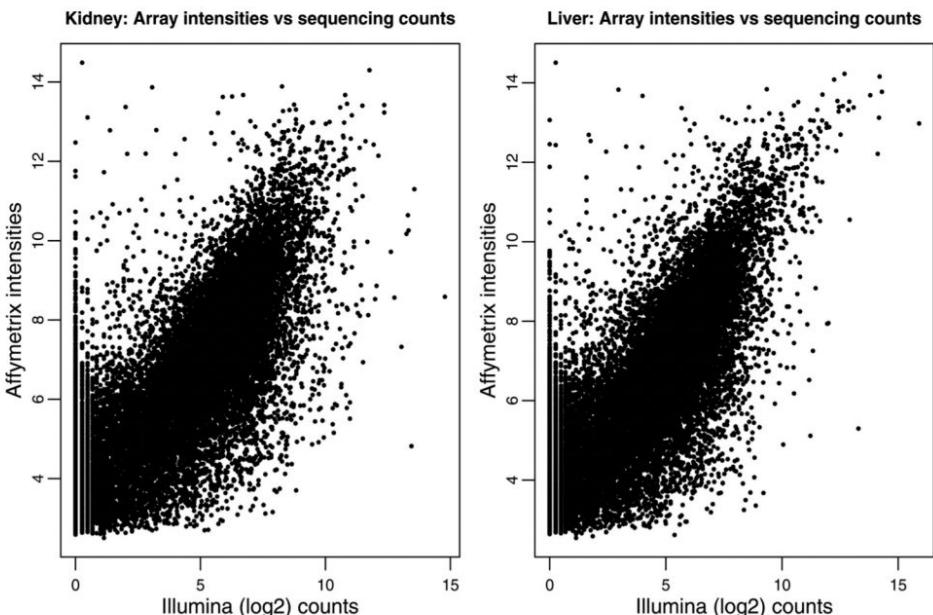
Sturm, G. et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology, Bioinformatics, Volume 35, Issue 14, July 2019, Pages i436–i445, <https://doi.org/10.1093/bioinformatics/btz363>



# ISSUES - TECHNOLOGICAL BIASES

- Some of the methods rely on microarray based **cell type references**
- **Microarrays** = probe intensities
  - continuous measure, best modeled by **normal distribution after log transformation**
- **RNA-seq** – read counts
  - count based measure, best modeled by **negative binomial distribution** of raw counts
- Can transform RNA-seq data to better fit microarray (normal) distributions but count based methods would be better

Sturm, G. et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology, Bioinformatics, Volume 35, Issue 14, July 2019, Pages i436-i445, <https://doi.org/10.1093/bioinformatics/btz363>



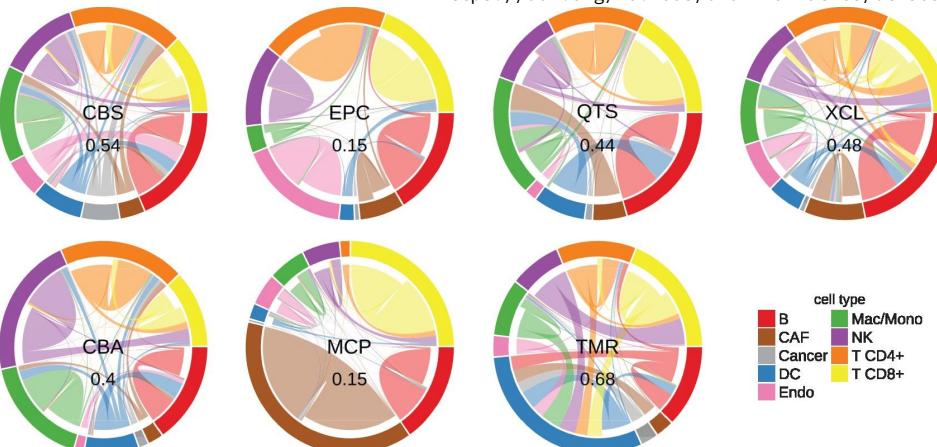
John C. Marioni et al. Genome Res. 2008;18:1509-1517

# ISSUES - "SPILLOVER"

Closely related cell types have similar cell signatures

- scores that predict enrichment of one cell type may also predict enrichment of another cell type
- other cell type might not even be present

Sturm, G. et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology, Bioinformatics, Volume 35, Issue 14, July 2019, Pages i436-i445, <https://doi.org/10.1093/bioinformatics/btz363>



A connection leading to a border segment of the same color indicates a correctly predicted cell type fraction; a **connection leading to a different color indicates spillover, i.e. a prediction of a different cell type than actually present**. Note that not all methods provide signatures for all cell types, in that case the connections are indicative of the cell types wrongly predicted when a method is confronted with cell types it has not been optimized for.

# ISSUES – MICROENVIRONMENT EFFECTS

- **Reference** sets are often derived from **purified non-tumor cells**
- Do pure cell populations accurately reflect the gene expression patterns of cells in a tumor?
- **Cell state versus cell identities** – microenvironment affects cell state

# ISSUES – LIMITED REFERENCE SETS

- Uneven background dataset availability
- Not all cell types available for all methods
- Not all species available

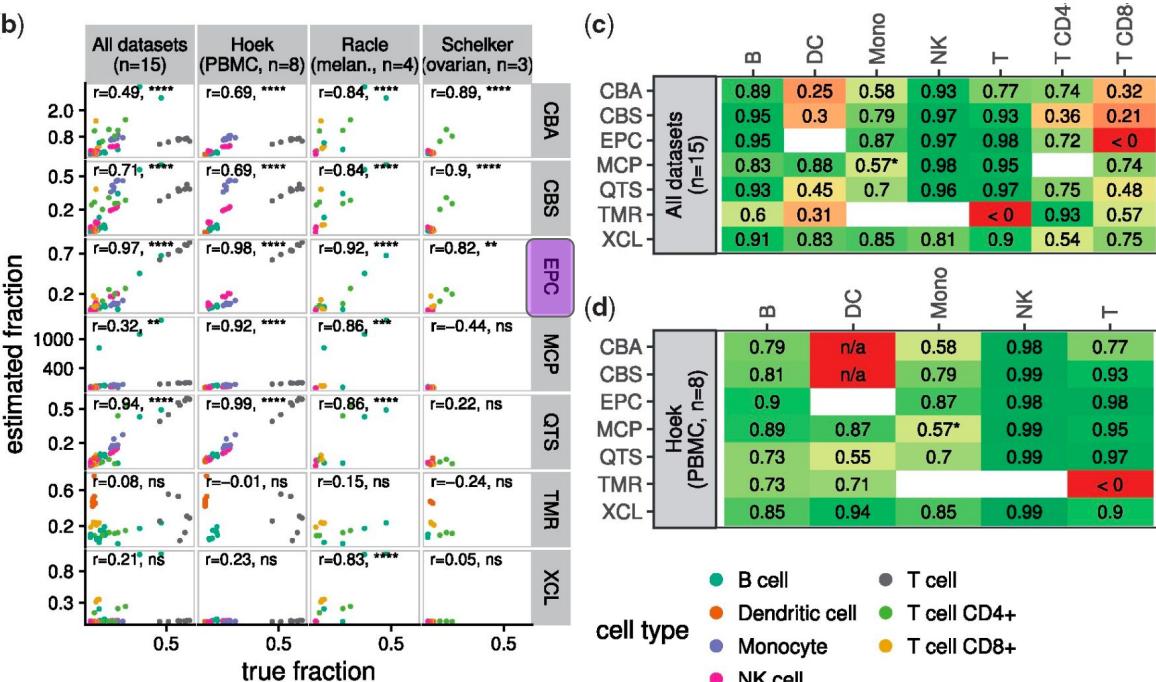
**Table 1.**  
Overview of cell type quantification methods providing gene signatures for immuno-oncology

Tool	Abbrev.	Type	Score	Comparisons	Algorithm	Cell types	Reference
CIBERSORT	CBS	D	Immune cell fractions, relative to total immune cell content	Intra	v-support vector regression	22 immune cell types	Newman <i>et al.</i> (2015)
CIBERSORT abs. mode	CBA	D	Score of arbitrary units that reflects the absolute proportion of each cell type	Intra, inter	v-support vector regression	22 immune cell types	Newman <i>et al.</i> (2015, 2018)
EPIC	EPC	D	Cell fractions, relative to all cells in sample	Intra, inter	constrained least square regression	6 immune cell types, fibroblasts, endothelial cells	Racle <i>et al.</i> (2017)
MCP-counter	MCP	M	Arbitrary units, comparable between samples	Inter	mean of marker gene expression	8 immune cell types, fibroblasts, endothelial cells	Becht <i>et al.</i> (2016)
quanTseq	QTS	D	Cell fractions, relative to all cells in sample	Intra, inter	constrained least square regression	10 immune cell types	Finotello <i>et al.</i> (2017)
TIMER	TMR	D	Arbitrary units, comparable between samples (not different cancer types)	Inter	linear least square regression	6 immune cell types	Li <i>et al.</i> (2016)
xCell	XCL	M	Arbitrary units, comparable between samples	Inter	ssGSEA (Hänzelmann <i>et al.</i> , 2013)	64 immune and non-immune cell types	Aran <i>et al.</i> (2017)

Note: Methods can be conceptually distinguished in marker gene-based approaches (M) and deconvolution-based approaches (D). The output scores of the methods have different properties and allow either intra-sample comparisons between cell types, inter-sample comparisons of the same cell type, or both. All methods come with a set of cell type signatures ranging from six immune cell types to 64 immune and non-immune cell types.

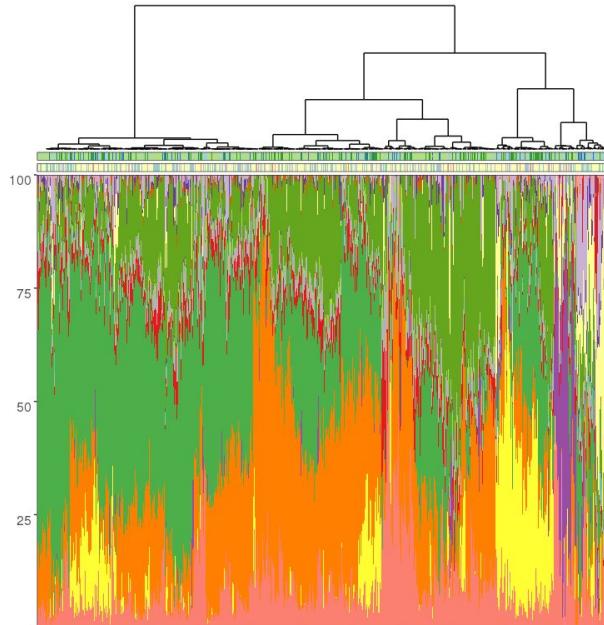
# COMPARING METHODS – CORRELATIONS

- Performance of the methods on three independent datasets that provide immune cell quantification by FACS.
- Different cell types are indicated in different colors.



# PREDICTION OF TME PROFILES IN BREAST TUMORS

MuSiC

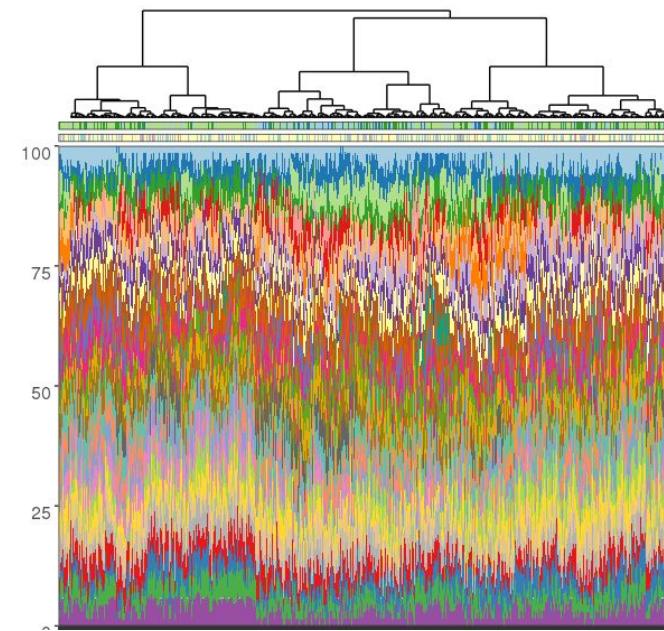


BRCA-TCGA (n=1077)

Stage  
1 2 3 4 NA

Subtype\_mRNA  
Basal Her2 LumA LumB

CibersortX



Subpopulation

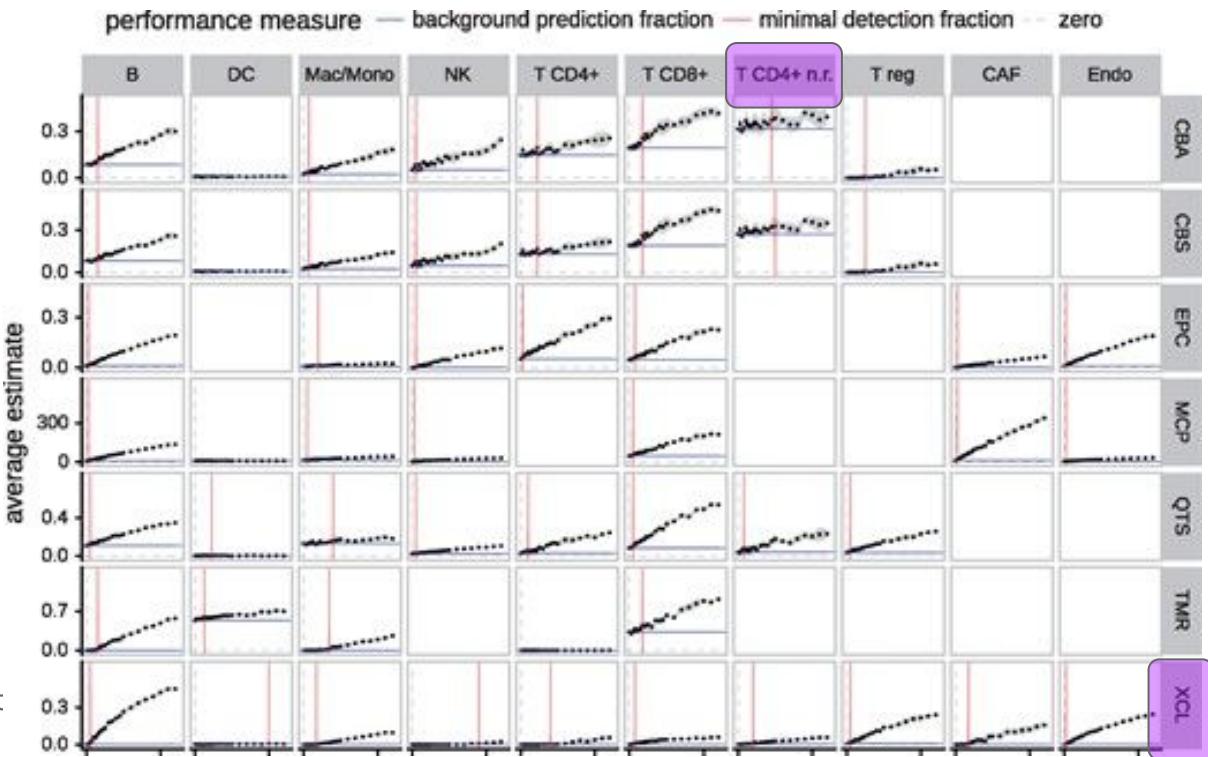
# COMPARING METHODS – DETECTION LIMITS

**Simulated bulk RNA-seq samples with an increasing amount of the cell type of interest (x-axis)**

- background of 1000 cells randomly sampled from the other cell types

Figure explanation

- dots = the mean predicted score across five independently simulated samples for each fraction of spike-in cells
- red line = minimal detection fraction, i.e. the minimal fraction needed for a method to detect its abundance as different from background
- blue line = background prediction level, i.e. average estimate of a method while the cell type is absent



# R PACKAGE, IMMUNEDECONV

Unified interface to the different deconvolution methods. The package is freely available from GitHub: <https://github.com/grst/immunedeconv>.

Sturm, G. et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology, *Bioinformatics*, Volume 35, Issue 14, July 2019, Pages i436-i445, <https://doi.org/10.1093/bioinformatics/btz363>



## Basic usage

Deconvolution of human data:

```
immunedeconv::deconvolute(gene_expression_matrix, "quantiseq")
```

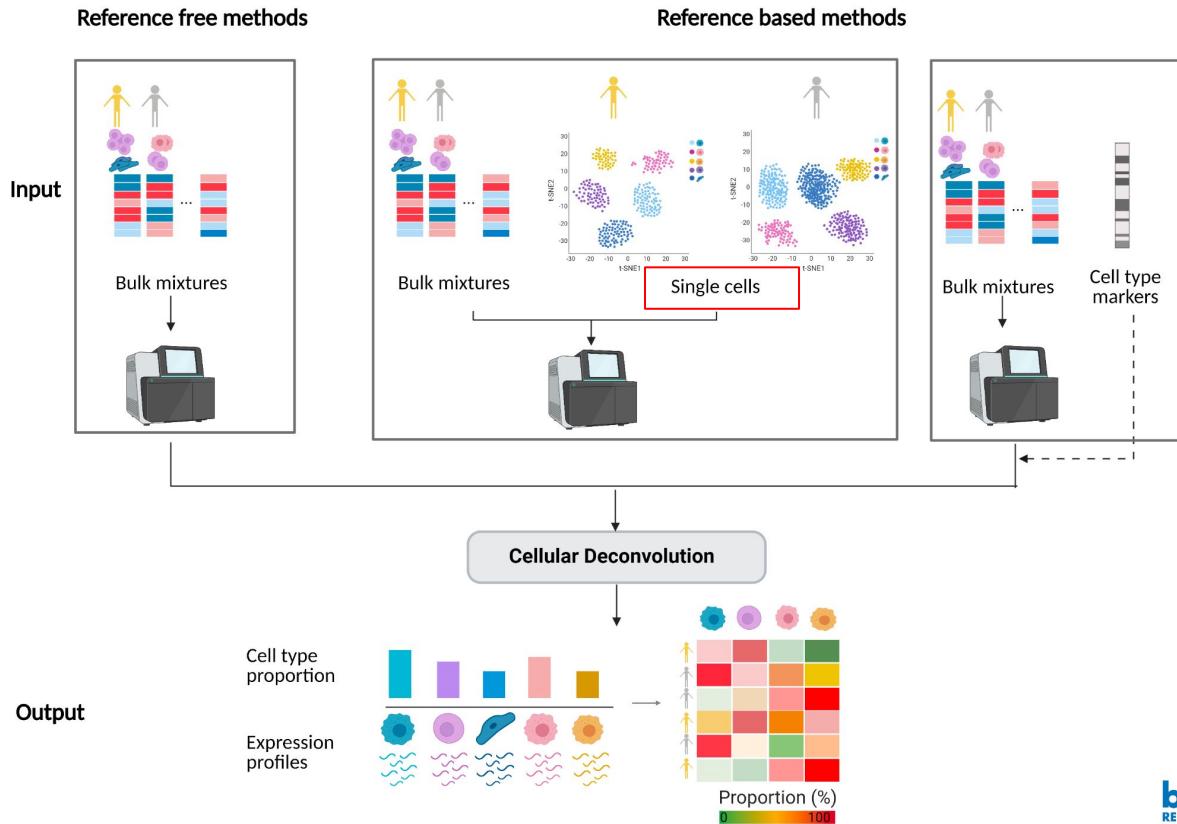
Deconvolution of mouse data:

```
immunedeconv::deconvolute_mouse(gene_expression_matrix, "mmcpc_counter")
```

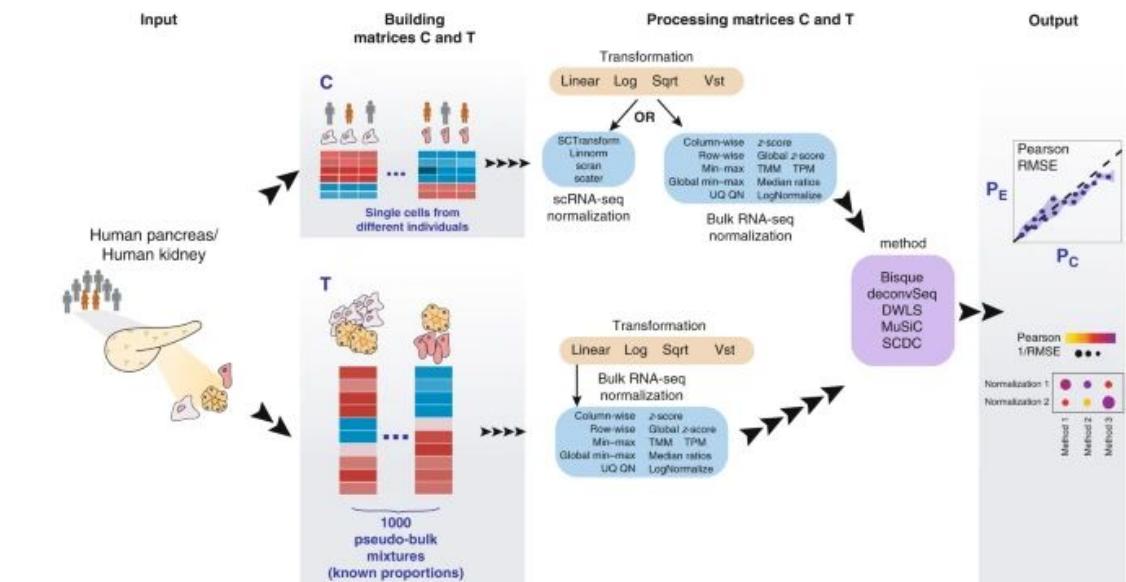
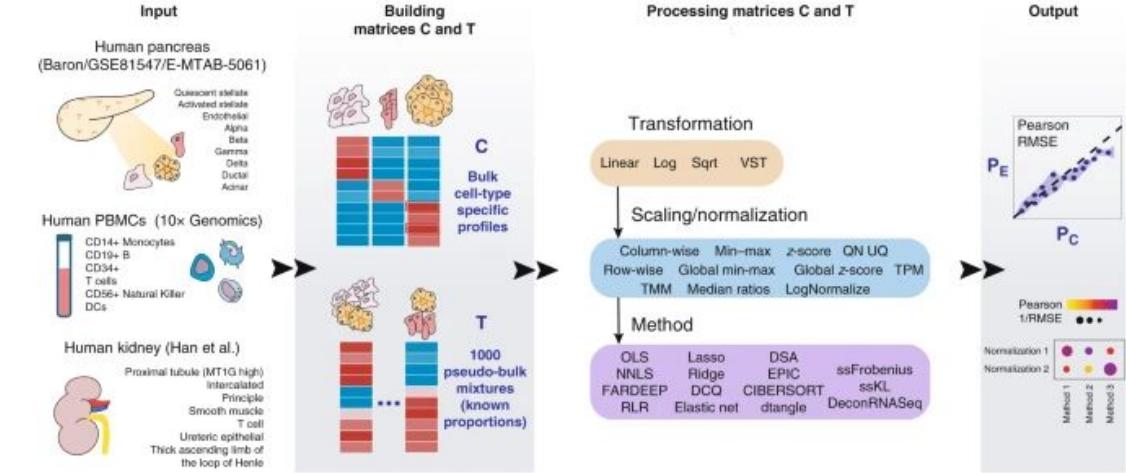
where `gene_expression_matrix` is a matrix with genes in rows and samples in columns. The rownames must be HGNC symbols for human data, or MGI gene symbols for mouse data. The colnames must be sample names. For human data, the method can be one of

```
quantiseq  
timer  
cibersort  
cibersort_abs  
mmcpc_counter  
xcell  
epic  
abis
```

# DECONVOLUTION OF BULK RNA-SEQ

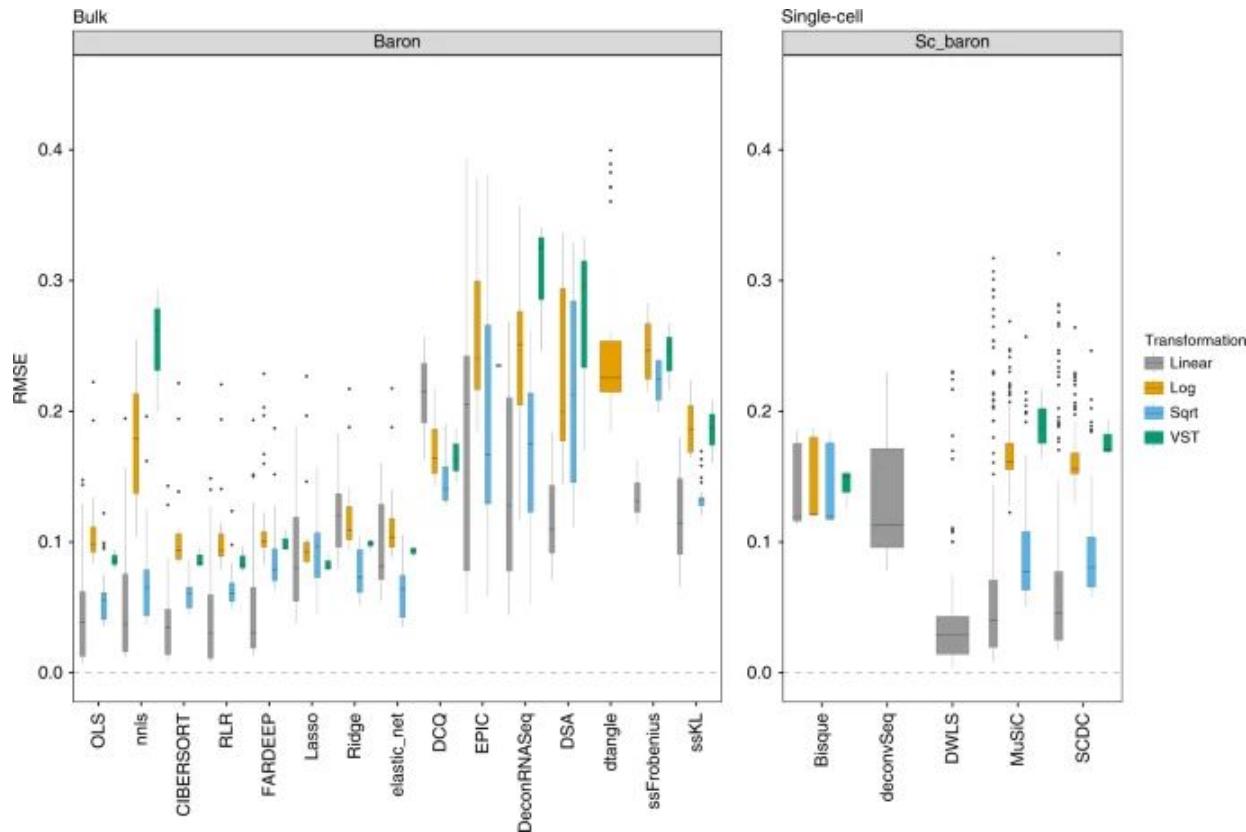


# BENCHMARKING OF CELL TYPE DECONVOLUTION PIPELINES FOR TRANSCRIPTOMICS DATA



# IMPACT OF THE DATA TRANSFORMATION ON THE DECONVOLUTION RESULTS

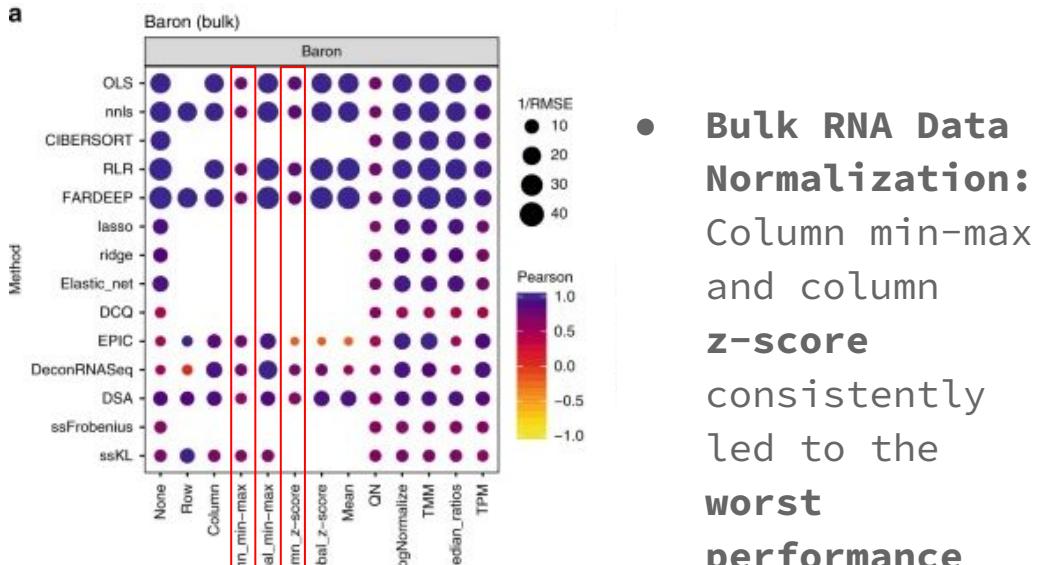
Maintaining the data in **linear scale** (linear transformation, in gray) consistently **showed the best results** (lowest RMSE values) whereas the logarithmic (in orange) and VST (in green; which also performs an internal complex logarithmic transformation) scale led to a poorer performance, with two to four-fold higher median RMSE values.



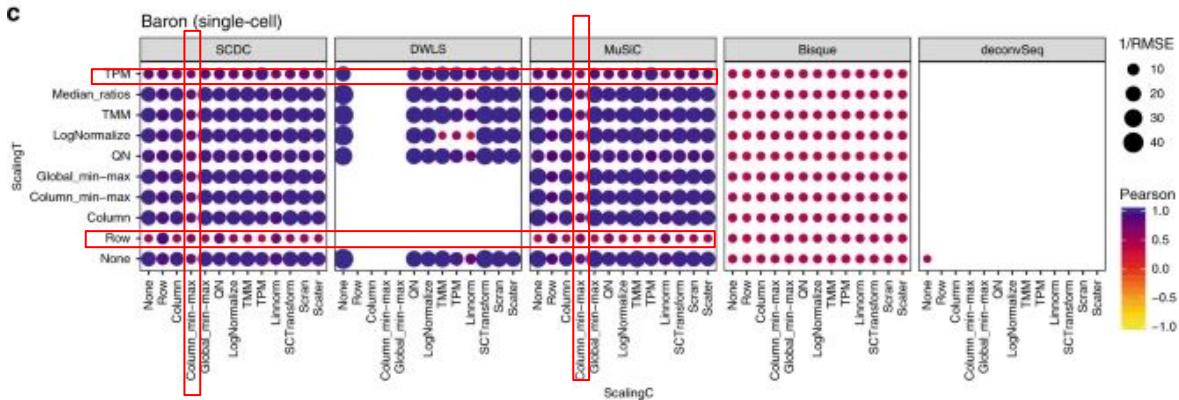
# COMBINED IMPACT OF DATA NORMALIZATION AND METHODOLOGY

- For deconvolution approaches using single cell:  
they evaluated each combination of normalization strategies for both the **pseudo-bulk mixtures** (scalingT, y-axis) and the **single-cell expression matrices** (scalingC, x-axis).

Avila Cobos, F., Alquicira-Hernandez, J., Powell, J.E. et al. Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun 11, 5650 (2020).  
<https://doi.org/10.1038/s41467-020-19015-1>

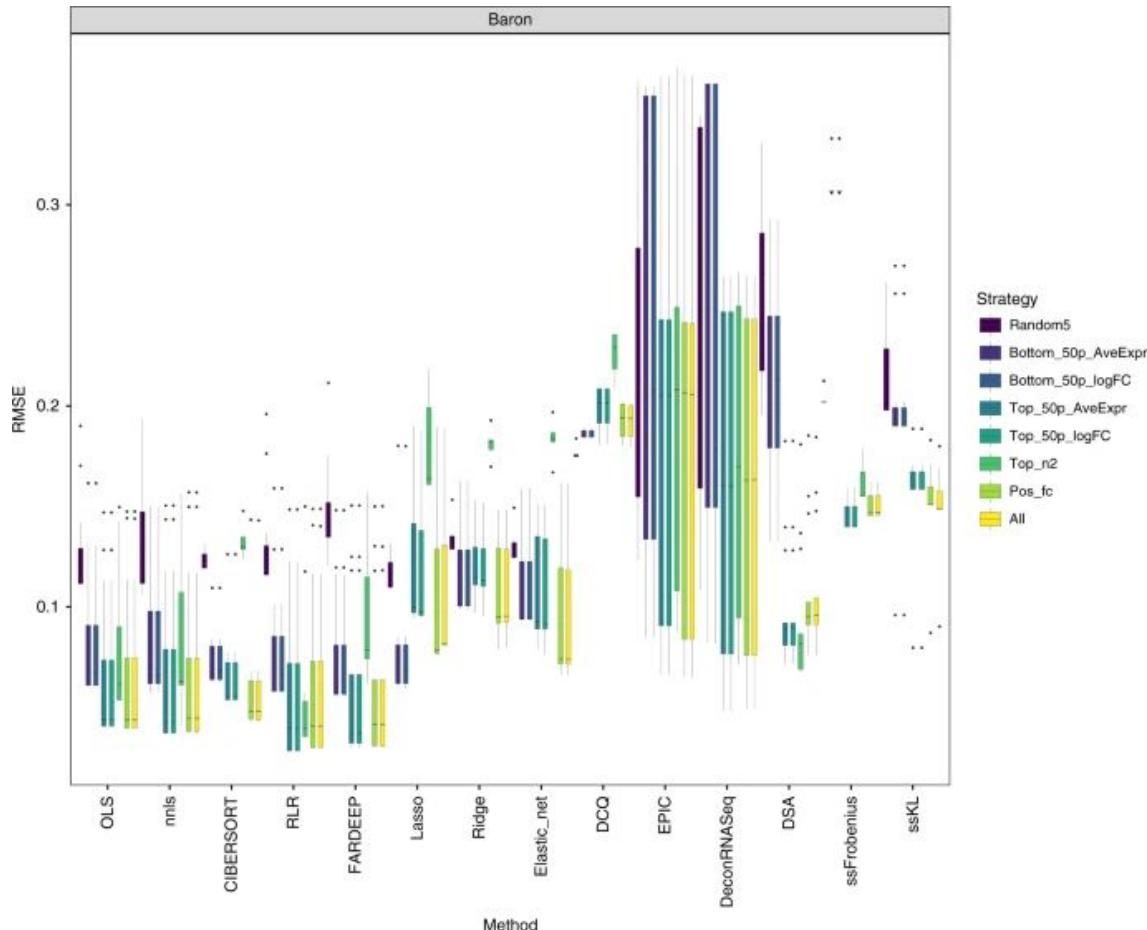


- Bulk RNA Data Normalization: Column min-max and column z-score consistently led to the **worst performance**



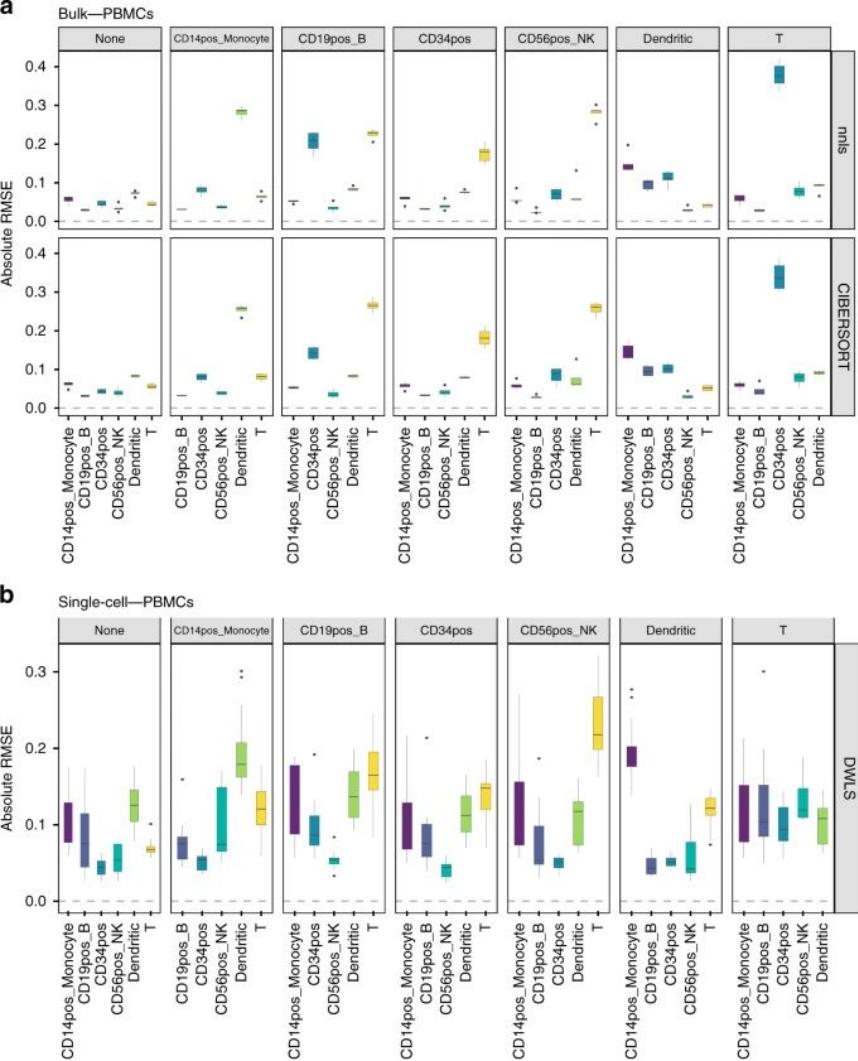
# IMPACT OF THE MARKER SELECTION

The use of all possible markers (all strategy) showed the best performance overall, followed by positive fold-change markers



# EFFECT OF CELL TYPE REMOVAL

Removing any cell type from the reference matrix led to distorted proportions for all other cell types.



# IN CONCLUSION, WHEN PERFORMING A DECONVOLUTION TASK, WE ADVISE USERS TO:

- (a) keep their input data in **linear scale**;
- (b) select any of the scaling/normalization approaches described here with exception of row scaling, column min-max, column z-score or quantile normalization;
- (c) choose a regression-based bulk deconvolution method (e.g., RLR, CIBERSORT or FARDEEP) and also perform the same task in parallel with DWLS, MuSiC or SCDC **if scRNA-seq data is available**;
- (d) use a stringent marker selection strategy that focuses on differences between the first and second cell types with highest expression values;
- (e) use a comprehensive reference matrix that include **all relevant cell types present in the mixtures**.

# OTHER REFERENCES

- Sutton, G.J., Poppe, D., Simmons, R.K. et al. Comprehensive evaluation of deconvolution methods for human brain gene expression. *Nat Commun* 13, 1358 (2022).  
<https://doi.org/10.1038/s41467-022-28655-4>
- Avila Cobos, F., Alquicira-Hernandez, J., Powell, J.E. et al. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun* 11, 5650 (2020).  
<https://doi.org/10.1038/s41467-020-19015-1>
- Jin, H., Liu, Z. A benchmark for RNA-seq deconvolution analysis under dynamic testing environments. *Genome Biol* 22, 102 (2021). <https://doi.org/10.1186/s13059-021-02290-6>
- Monaco, G., Lee, B., Xu, W., et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types, *Cell Reports* 26, 6 (2019).  
<https://doi.org/10.1016/j.celrep.2019.01.041>.
- Chu, T., Wang, Z., Pe'er, D. et al. 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 (2022). <https://doi.org/10.1038/s43018-022-00356-3>
- Jew, B., Alvarez, M., Rahmani, E. et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat Commun* 11, 1971 (2020).  
<https://doi.org/10.1038/s41467-020-15816-6>
- Newman, A.M., Steen, C.B., Liu, C.L. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 37, 773-782 (2019).  
<https://doi.org/10.1038/s41587-019-0114-2>
- <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-3-r25>

TIME FOR THE  
HANDS ON TRAINING

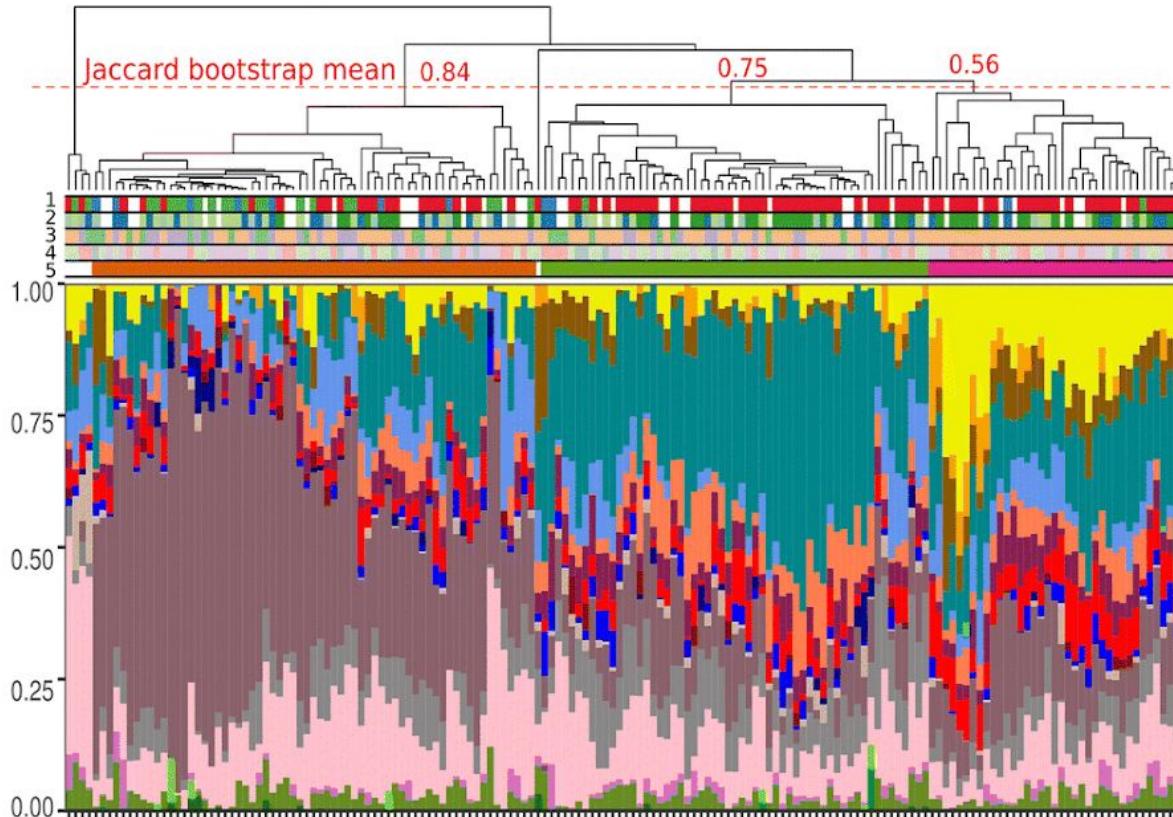
# METASTATIC MELANOMA TME PROFILES

Poor clinical outcome in metastatic melanoma is associated with a microRNA-modulated immunosuppressive tumor microenvironment

Natasha A. N. Jorge, Jéssica G. V. Cruz, Marco Antônio M. Pretti, Martin H. Bonamino, Patricia A. Possik & Mariana Boroni 

*Journal of Translational Medicine* 18, Article number: 56 (2020) | [Cite this article](#)

66 Accesses



1-mRNA Subtype  
■ immune  
■ keratin  
■ MITF-low  
■ Not Identified

2- Methylation Subtype  
■ CpG island-methylated  
■ hyper-methylated  
■ hypo-methylated  
■ normal-like

3-Tumor\_tissue\_site  
■ Distant Metastasis  
■ Regional Cutaneous or Subcutaneous Tissue  
■ Regional Lymph Node

4- Mutation  
■ WT3  
■ NF1  
■ RAS  
■ BRAF

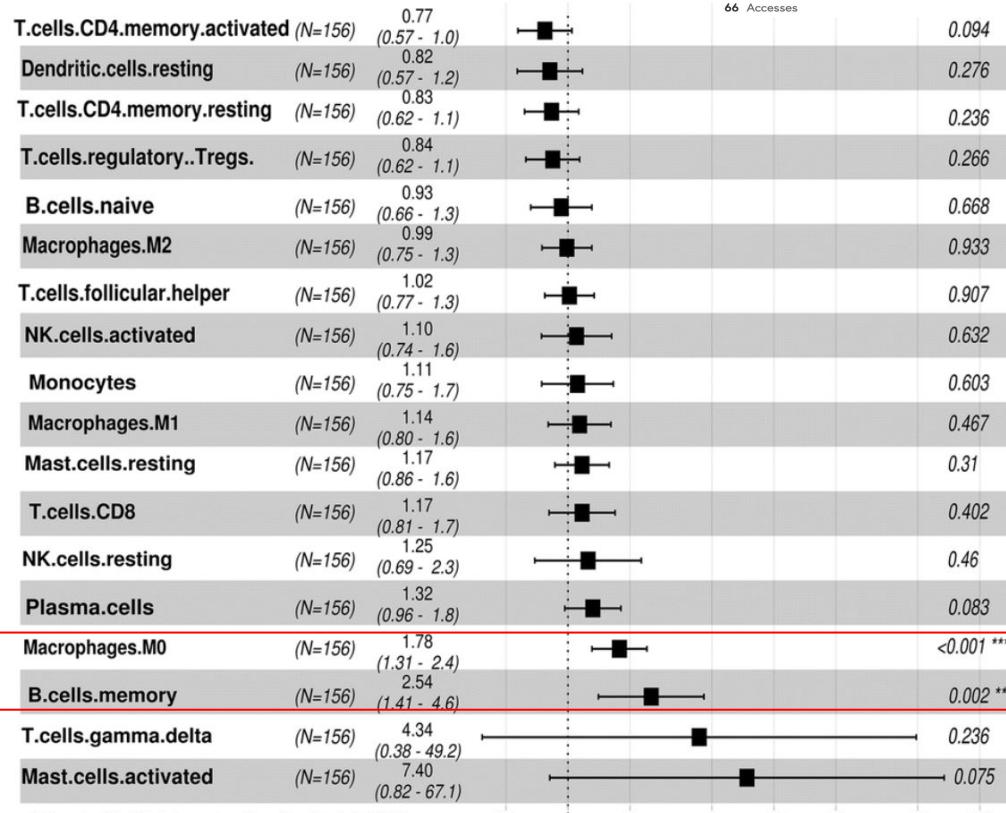
5- Group  
■ 1  
■ 2  
■ 3

## Subpopulation

- B.cells.naive
- B.cells.memory
- Plasma.cells
- T.cells.CD8
- T.cells.CD4.naive
- T.cells.CD4.memory.resting
- T.cells.CD4.memory.activated
- T.cells.follicular.helper
- T.cells.regulatory..Tregs.
- T.cells.gamma.delta
- NK.cells.resting
- NK.cells.activated
- Monocytes
- Macrophages.M0
- Macrophages.M1
- Macrophages.M2
- Dendritic.cells.resting
- Dendritic.cells.activated
- Mast.cells.resting
- Mast.cells.activated
- Neutrophils

# METASTATIC MELANOMA TME PROFILES

## SURVIVAL IMPACT



# METASTATIC MELANOMA TME PROFILES - SURVIVAL IMPACT

