

# Lesson 4: Cancer immunogenomics

Matteo Dugo, PhD

IRCCS Ospedale San Raffaele

Department of Medical Oncology – Breast Cancer Unit

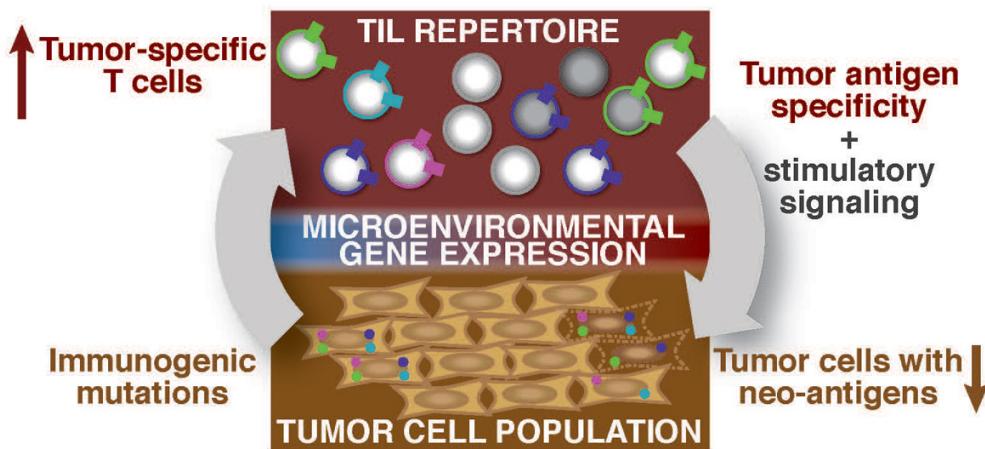
Clinical translational and immunotherapy research

# Cancer immunogenomics

Research field focusing on the study of tumour–immune cell interactions using genomic tools

## Evolving and heterogeneous nature of two multicellular ecosystems

Immune system with numerous innate and adaptive immune cell subpopulations some of which show phenotypic plasticity



Cancer compartment developing following an evolutionary process

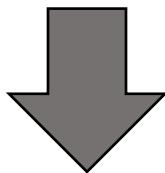
<https://www.mskcc.org/research-programs/immunogenomics-and-precision-oncology-platform/what-immunogenomics>

# Cancer immunogenomics

---

Comprehensive characterization of cancer immunity requires the determination of the following broad characteristics:

- a. cellular composition of immune infiltrates
- b. neoantigens
- c. immune contexture
- d. tumour microenvironment (TME)
- e. host and environmental factors.

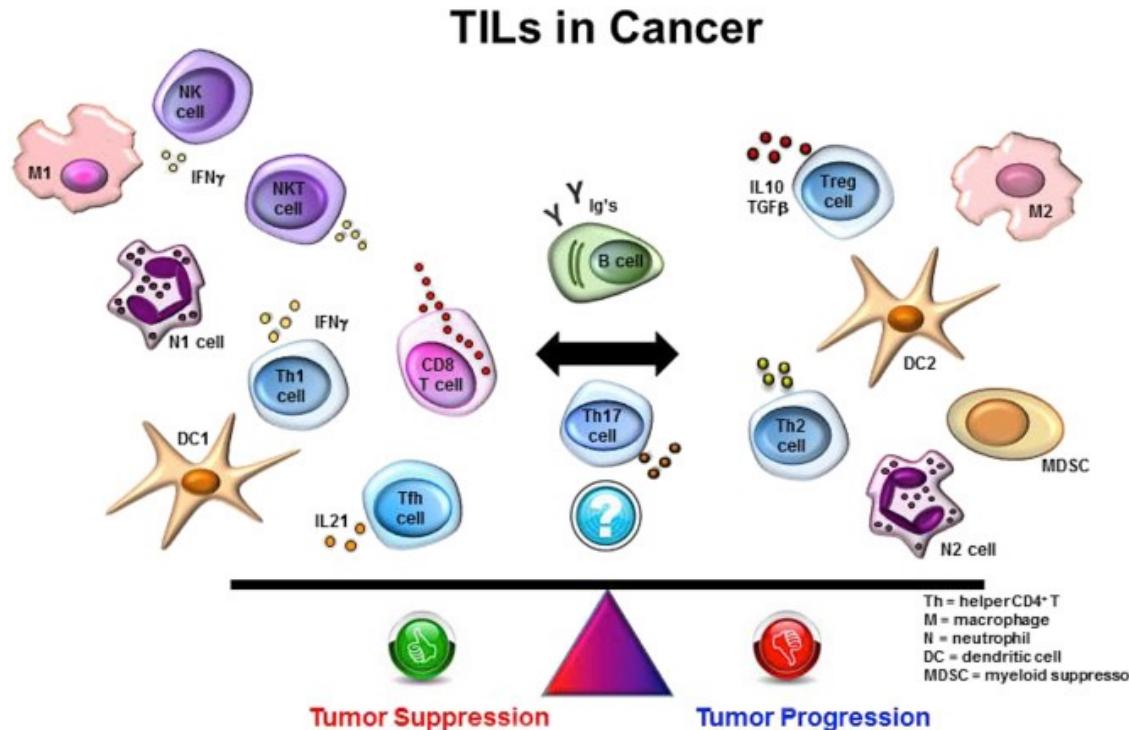


Multiparametric assessment rather than the use of single parameters is necessary to dissect the complex tumor–immune cell interactions and inform cancer immunotherapy.

# Cancer immunogenomics: Compositions

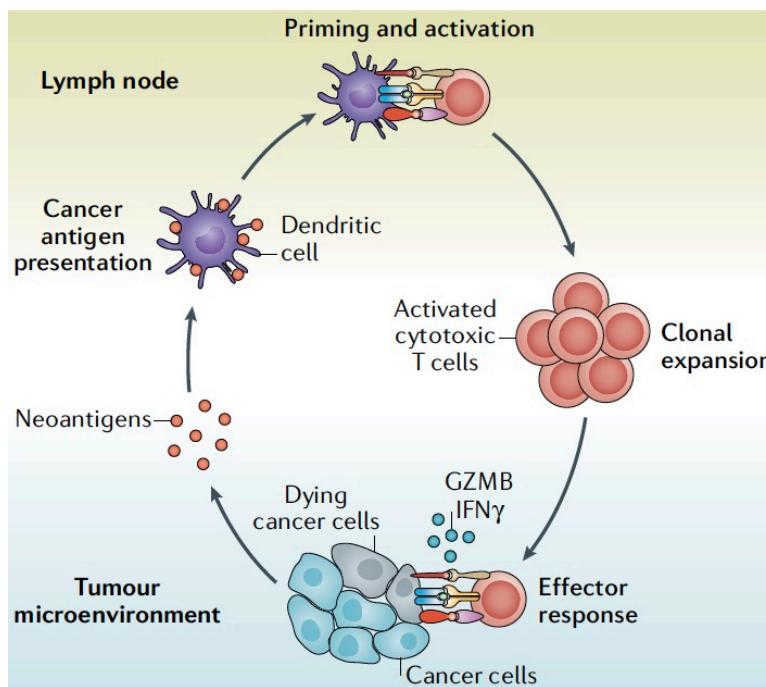
Different types of immune cells have different effects on tumour progression. The determination of the cellular composition of immune infiltrates in tumours provides:

- prognostic information
- predictive markers
- novel therapeutic strategies



# Cancer immunogenomics: Neoantigens

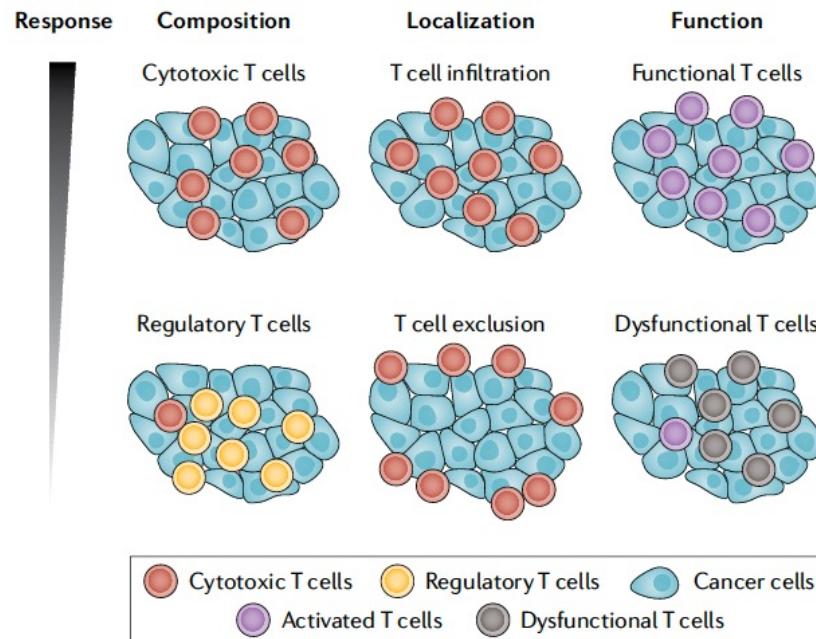
- Short peptides generated from the expression of mutated or rearranged genes in cancer cells, but not in normal cells.
- Recognized by T cells through the interaction of the T cell receptor with the peptide–HLA complex.
- Major determinants of the response to immunotherapy with checkpoint blockers
- Computational or experimental characterization in cancer patients is the basis for personalized cancer vaccines and T cell-based immunotherapies



**The cancer-immunity cycle**

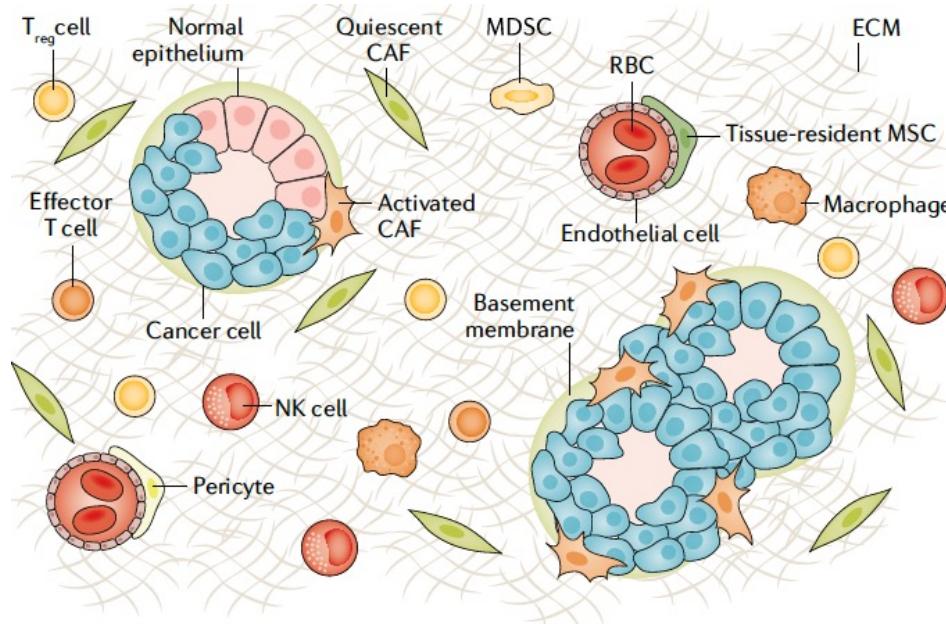
# Cancer immunogenomics: Immune contexture

- Immune contexture: the localization and functional state of tumor-infiltrating immune cells
- Immune contexture determine prognosis and efficacy of anticancer immune responses.
- **Hot tumors** (Immunogenic tumours with high infiltration of T cells) are more amenable to checkpoint-blocker-based monotherapy or combination therapy than **cold tumors** (Poorly immunogenic tumours with low or no infiltration of T cells).



# Cancer immunogenomics: TME

- TME is a complex ecosystem composed of different cell types (cancer cells, epithelial cells, cancer-associated fibroblasts (CAFs) and immune cells)
- Determine the cellular components of the TME to investigate their role and interactions (e.g.: CAFs release various tumour-promoting cytokines and chemokines and contribute to an immunosuppressive TME)
- Measure physical properties to study aberrant cell mechanics (e.g.: ECM stiffness)

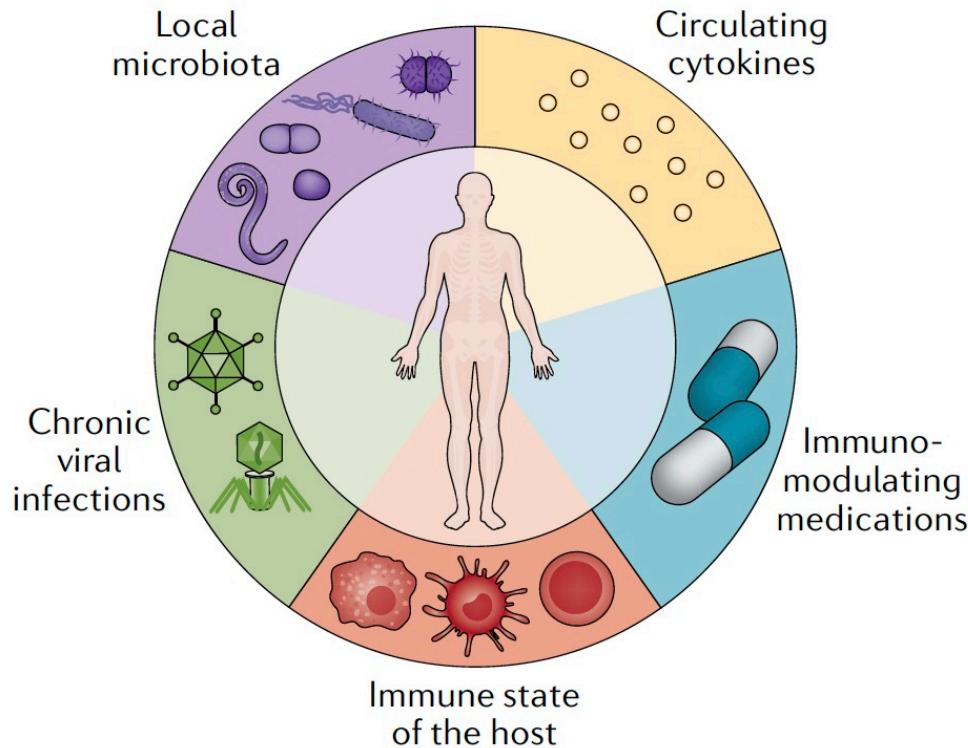


Finotello et al. Nat. Rev. Genet. 2017

# Cancer immunogenomics: host and environmental factors

---

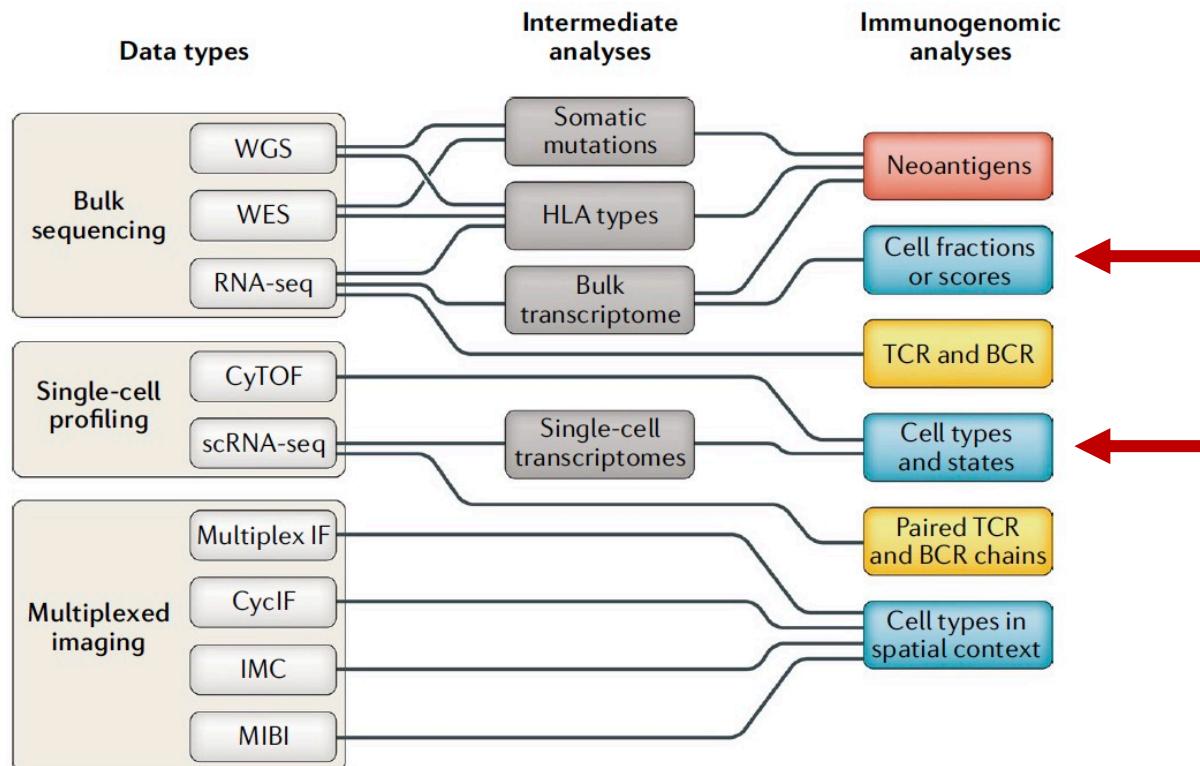
Besides the tumour immune contexture, different systemic parameters influence the patient's outcome and response to therapy



Finotello et al. Nat. Rev. Genet. 2017

# Overview of technologies and analyses

- Technologies for bulk sequencing allow the prediction of candidate neoantigens and deconvolution of cell fractions or computation of abundance scores.
- Single-cell analysis allow the characterization of cell types and states, but also the reconstruction of B cell receptors (BCRs) and T cell receptors (TCRs) of the same cells.
- Recent multiplexed imaging techniques enable the phenotyping of distinct cell types and the reconstruction of the spatial architecture of the tumour microenvironment.



# Quantification of immune cells

The analysis of different types of data can reveal different facets of the tumour immune contexture depending on their pros and cons.

## Bulk RNA-seq

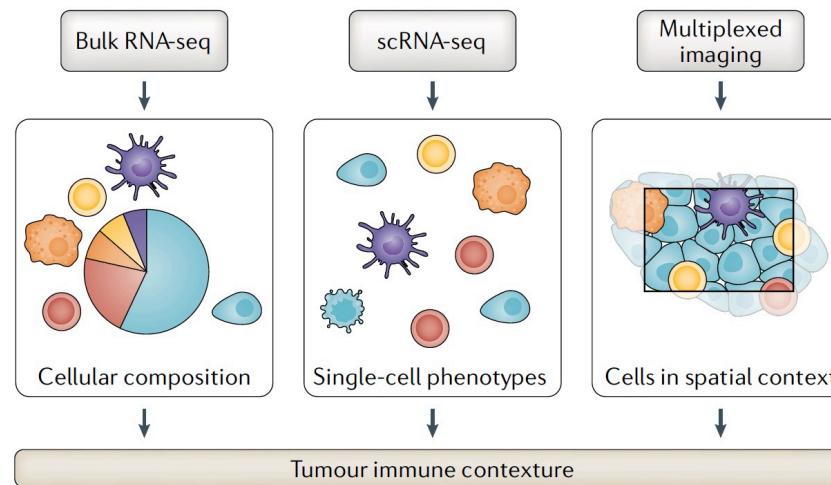
- ▶ useful to quantify the fractions of different cell subpopulations
- ▶ cannot be used to study the phenotypes of single cells.

## scRNA-seq

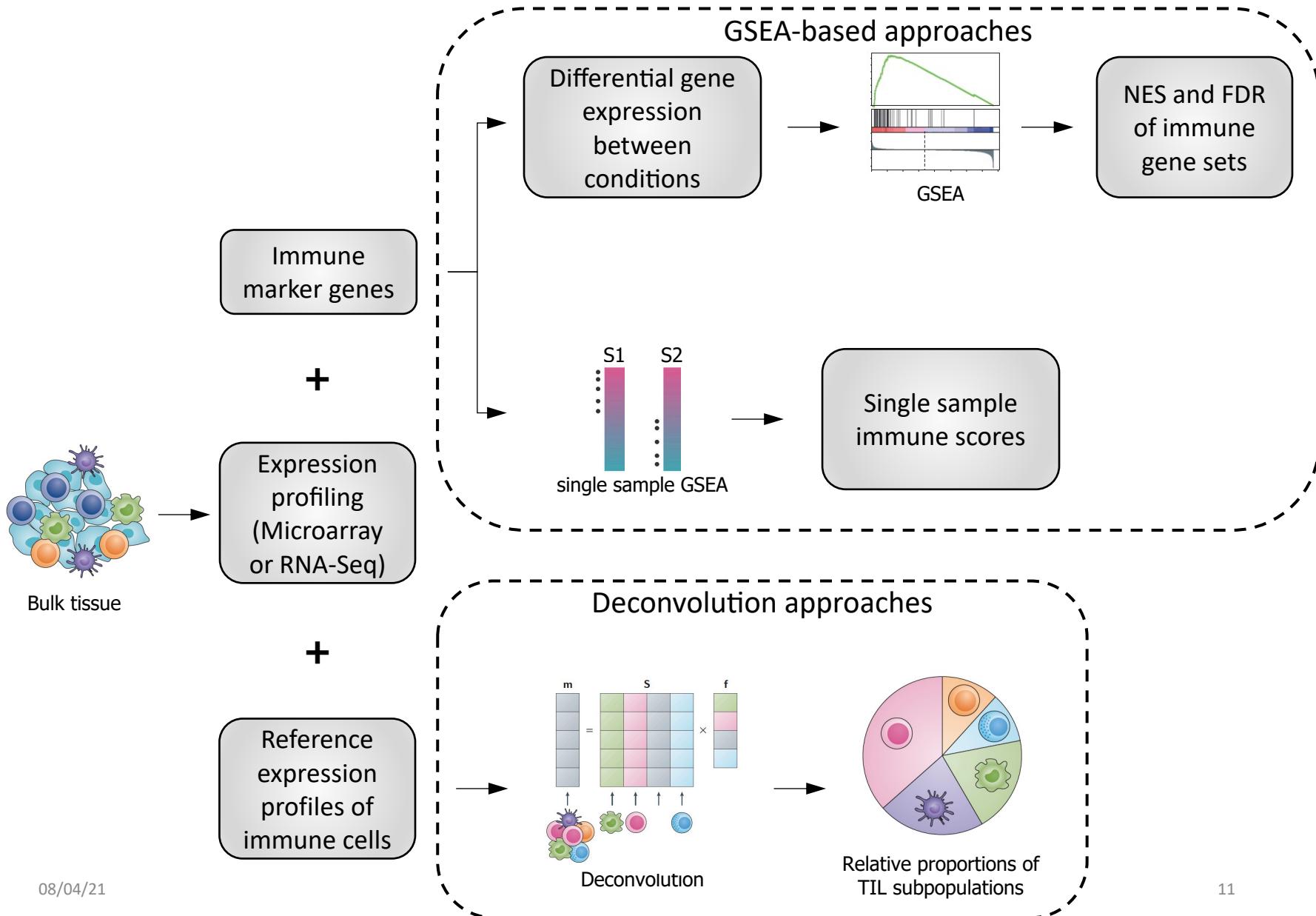
- ▶ not optimal to quantitatively assess the cellular composition of the tumour
- ▶ portray single-cell types and states.

## Multiplexed imaging

- ▶ allows the study of cells in a spatial context
- ▶ only reconstructs a restricted, 2D portion of the tumour microenvironment
- ▶ limitation in the number of markers that can be analyzed



# Quantification of immune cells from bulk RNA-Seq



# GSEA-based approaches

---

GSEA evaluates ranked gene lists for statistical enrichment of genes involved in defined pathways and cellular processes (see lesson 1).

1. In the comparative approach, genes are ranked based on differential expression between two biological states (see lesson 1).
2. An alternative approach is single-sample GSEA (ssGSEA), which computes an ES representing the degree to which genes in a particular gene set are coordinately up- or down-regulated within a single sample.  
ssGSEA ranks the genes by their absolute expression (not differential expression) in a sample and computes ES.

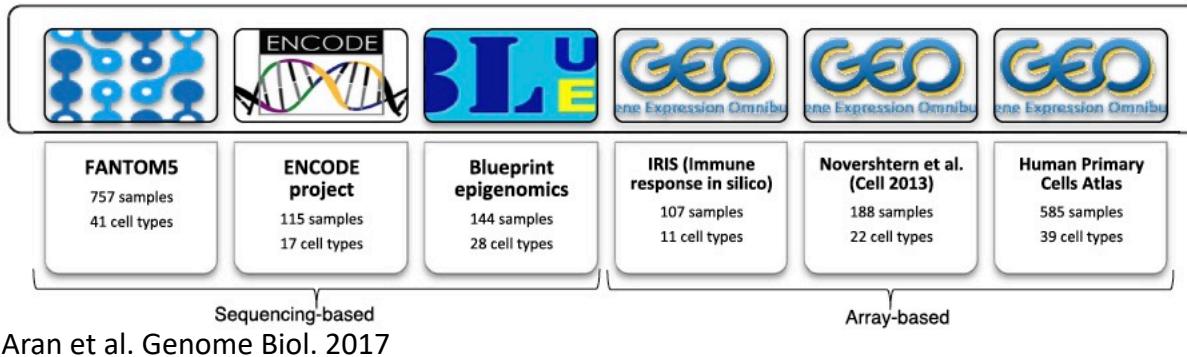
Popular methods:

- xCell (Aran et al. 2017). R package or web-based tool. 64 cell types
- MCP-counter (Becht et al. 2016). R package. 8 immune cell types + fibroblasts + endothelial cells
- ConsensusTME (Jimenez-Sanchez et al. 2019). R package. 8 immune cell types + fibroblasts + endothelial cells. **Cancer specific**.

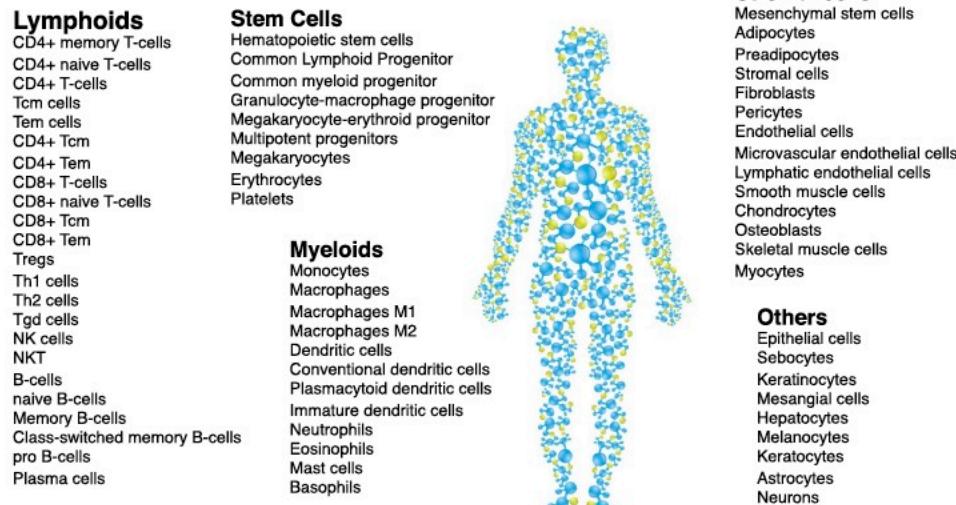
All these methods rely on predefined immune specific marker gene lists.

# GSEA-based approaches: xCell

Data sources for generation of cell-type specific gene signatures



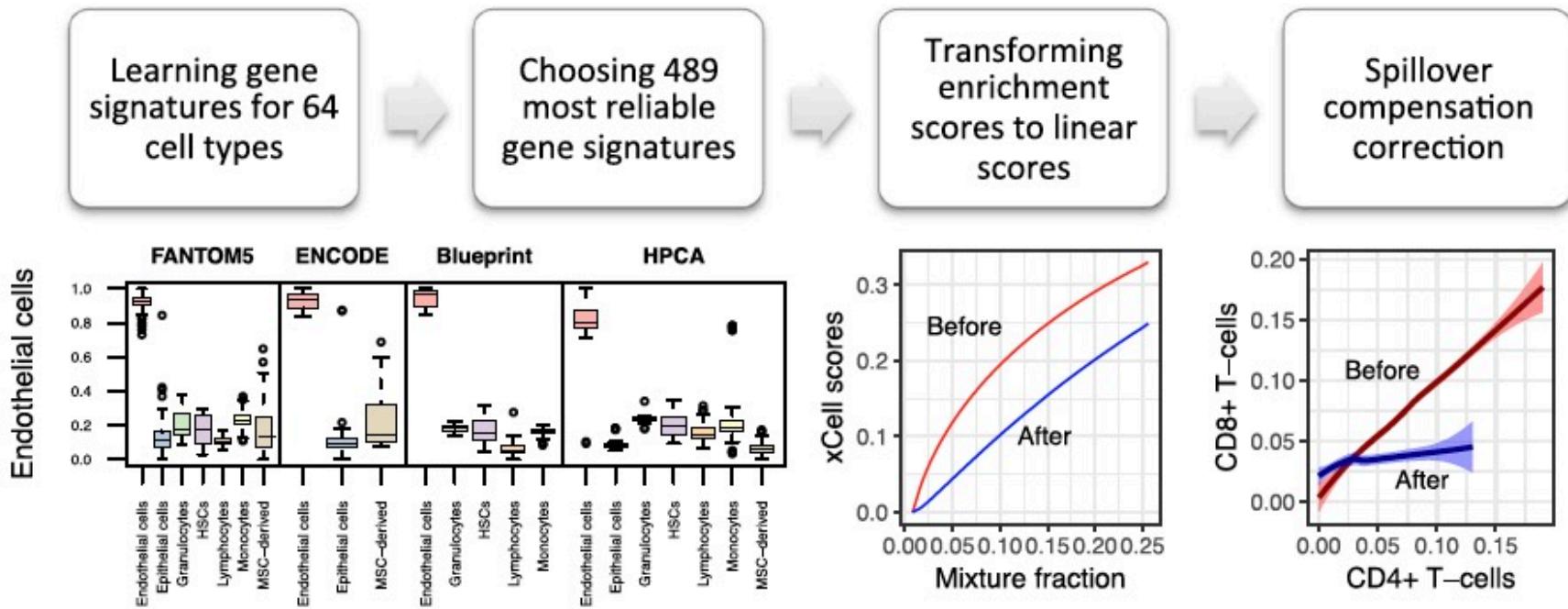
Compendium of gene signatures for 64 cell types



Aran et al. Genome Biol. 2017

# GSEA-based approaches: xCell

## xCell pipeline



Aran et al. Genome Biol. 2017

Spillover: the erroneous prediction of another cell type due to a partial overlap of the signatures

# GSEA-based approaches: xCell

Web interface: <https://xcell.ucsf.edu/>

To use *xCell* simply upload human gene expression data file in tab delimited text format or csv (up to 1Gb). The expression matrix should be a matrix with genes in rows and samples in columns. The rownames should be gene symbols. If the data contains non-unique gene symbols, rows with same gene symbols will be averaged. *xCell* uses the expression levels ranking and not the actual values, thus normalization does not have an effect, however normalizing to gene length (RPKM/FPKM/TPM/RSEM) is required.

## For RNA-Seq data

Importantly, *xCell* performs best with heterogenous dataset. Thus it is recommended to use all data combined in one run, and not break down to pieces (especially not cases and control in different runs).

Mark if the data is RNA-seq gene expression profile, and choose the *xCell* signature set (or other signature sets). The analysis may take a while to complete, therefore please provide an email address. A link to download the results in a tab delimited file format will be sent once the analysis is done.

An example of a gene expression data file can be downloaded [here](#).

Pre-calculated TCGA data by *xCell* is available [here](#). Please do not run TCGA data using the webtool.

\*\*\* We do not save any files. All uploaded data is immediately deleted after the analysis. Analysis results are immediately deleted after first download. \*\*\*

Upload gene expression data:

No file chosen

RNA-seq?



Choose gene signatures:

▾

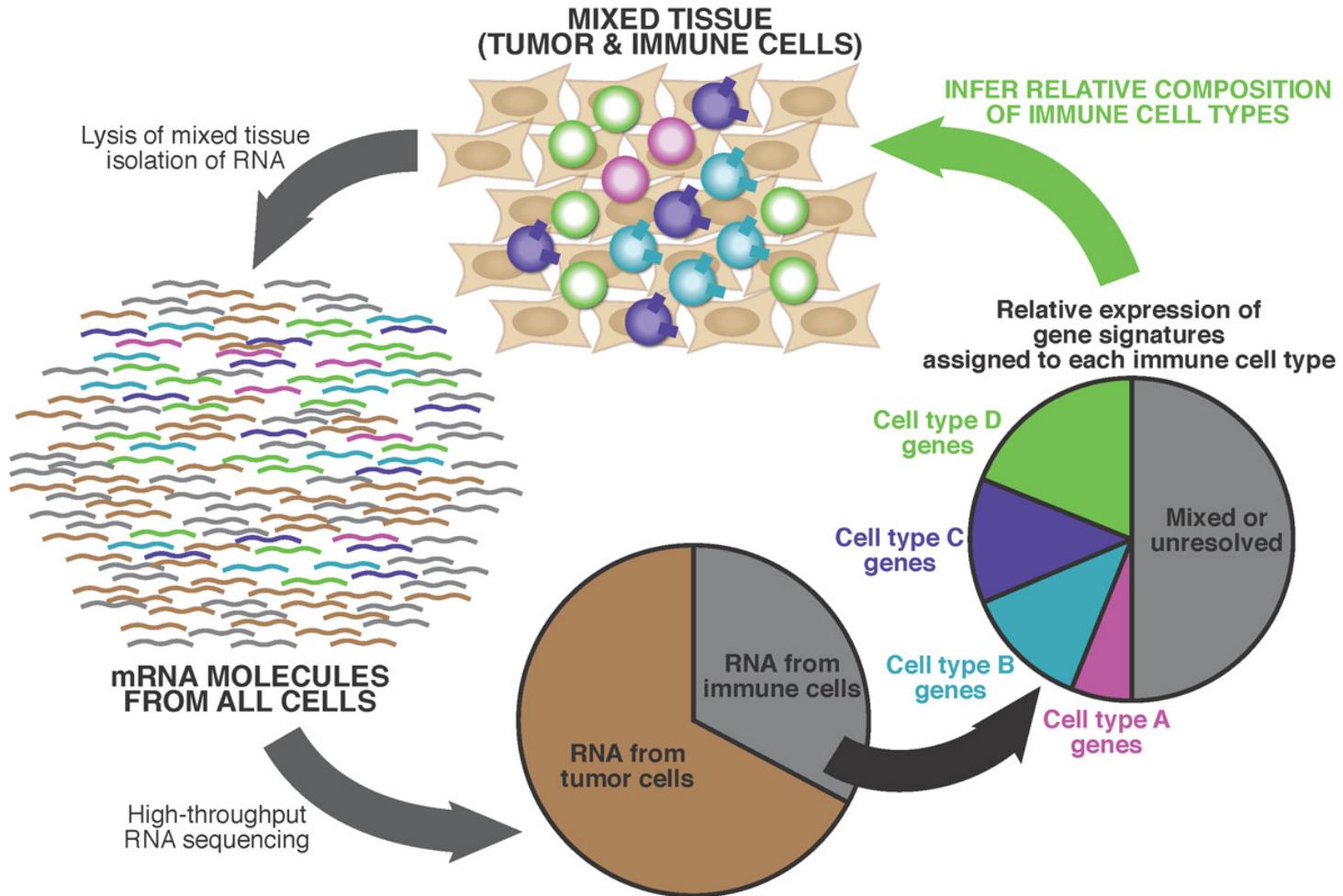
Email address:

An e-mail will be sent with link to download the results once they are ready.

Run!

# Deconvolution approaches

To deconvolve: to decompose a complex signal in its components

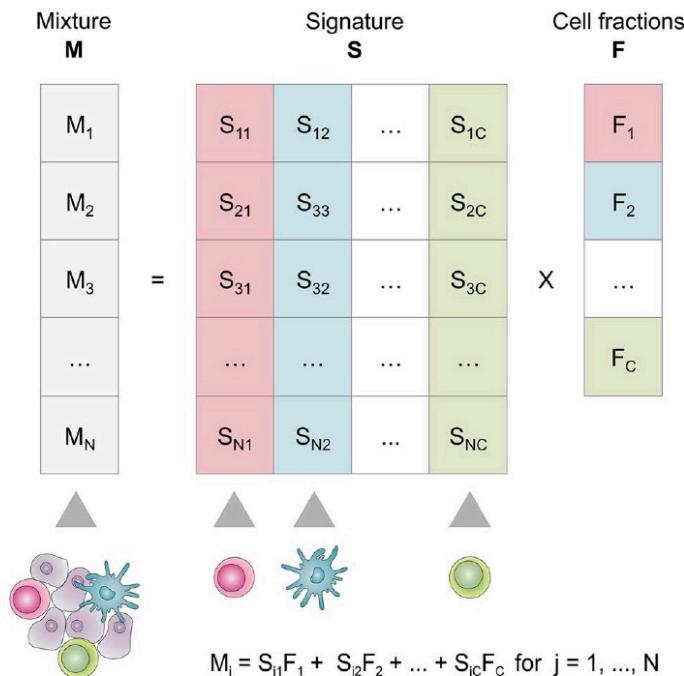


<https://www.mskcc.org/research-programs/immunogenomics-and-precision-oncology-platform/what-immunogenomics>

# Deconvolution approaches

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.

Require cell-type specific gene expression matrices.

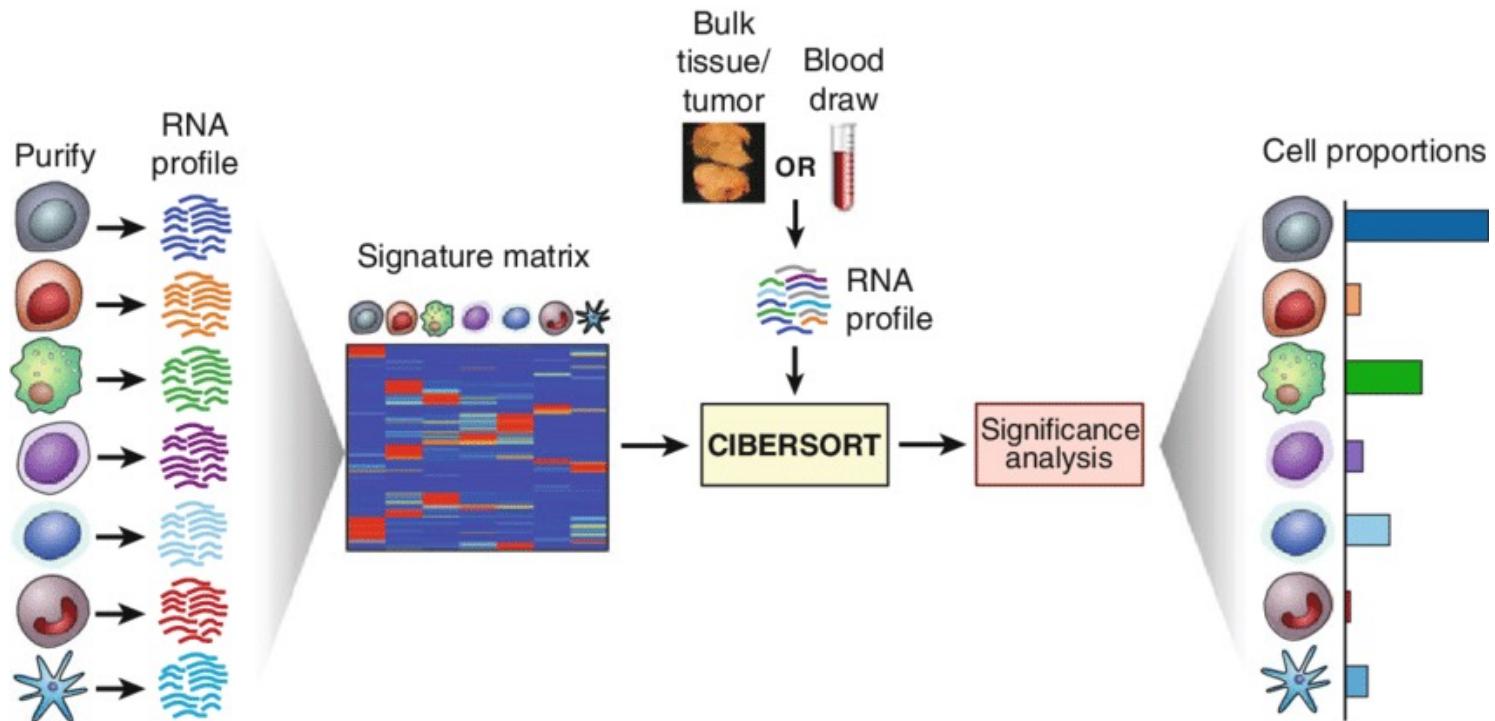


Popular tools:

- CIBERSORT (Newman et al. 2015). 22 immune cell types.
- EPIC (Racle et al. 2017). 6 immune cell types + fibroblasts + endothelial cells
- quanTIseq (Finotello et al. 2017). 10 immune cell types
- TIMER (Li et al. 2016). 6 immune cell types
- ImmuCC for mouse data (Chen et al. 2017). 10 immune cell types

# Deconvolution approaches: CIBERSORT

As input, CIBERSORT requires a “signature matrix” comprised of barcode genes that are enriched in each cell type of interest. Once a suitable knowledgebase is created and validated, CIBERSORT can be applied to characterize cell type proportions in bulk tissue expression profiles.



Chen et al. Cancer Systems Biology: Methods and Protocols. 2018

# Deconvolution approaches: CIBERSORT

Web interface: <https://cibersort.stanford.edu/index.php>

The screenshot shows the 'Configure Basic CIBERSORT Options' page. At the top is the CIBERSORT logo. Below it, there are several input fields:

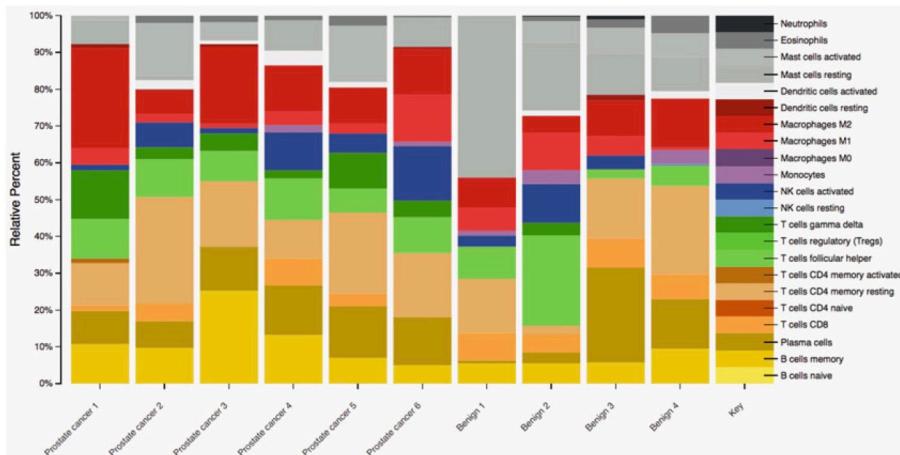
- Signature gene file\***: A dropdown menu with the placeholder "Please select an uploaded file below, or generate a custom signature..." and a "More Information..." link.
- Mixture file\***: A dropdown menu with the placeholder "REQUIRED - please select an uploaded file below:" and a "More Information..." link.
- Permutations**: A dropdown menu set to "100" with a "More Information..." link.
- Disable quantile normalization**
- Custom Signature Genes**: A section with two optional fields:
  - Reference sample file**: A dropdown menu with the placeholder "Optional - please select an uploaded file below:" and a "More Information..." link.
  - Phenotype classes file**: A dropdown menu with the placeholder "Optional - please select an uploaded file below:" and a "More Information..." link.

Before running CIBERSORT, all mixture files need to be uploaded (Menu > “Upload Files”). The user needs to select “Mixture” when uploading mixture files. Select “LM22 (22 immune cell types)” for “Signature gene file.” When using LM22, the user will need to select the uploaded mixture file and specify “LM22 (22 immune cell types)” for the signature gene file. At least 100 permutations are recommended to achieve statistical rigor.

# Deconvolution approaches: CIBERSORT

The website will output a stacked bar plot and a heat map. The output includes a *p*-value for the global deconvolution of each sample. A *p*-value threshold <0.05 is recommended.

By default, deconvolution results are expressed as relative fractions normalized to 1 (e.g., fractions of total leukocyte content). Absolute immune fraction score can be obtained by selecting “absolute” mode.



Input Sample	B cells	CD8 T cells	CD4 T cells	NK cells	Monocytes	Neutrophils	P-value	Pearson Correlation	RMSE
TCGA.EE.A29N.06A.12R.A18S...	0.341	0.108	0.186	0.007	0.358	0	0.000	0.569	0.822
TCGA.GN.A26A.06A.11R.A18T...	0.012	0.07	0.068	0.029	0.821	0	0.000	0.497	1.028
TCGA.EE.A2MR.06A.11R.A18S...	0.386	0.206	0.143	0	0.266	0	0.000	0.432	0.907
TCGA.ER.A2NG.06A.11R.A18T...	0.02	0.545	0	0.051	0.384	0	0.000	0.419	0.914
TCGA.FR.A8YE.06A.11R.A37K...	0.446	0.01	0.362	0.018	0.165	0	0.000	0.397	0.925
TCGA.EE.A3AF.06A.11R.A18S...	0.099	0.191	0	0.019	0.691	0	0.000	0.366	1.069
TCGA.ER.A19A.06A.21R.A18U...	0.051	0.249	0	0.048	0.653	0	0.000	0.365	1.044
TCGA.EE.A29G.06A.12R.A18T...	0.078	0	0.152	0.02	0.75	0	0.000	0.358	1.114
TCGA.GN.A4U.01A.11R.A32P...	0	0.522	0	0.031	0.446	0	0.000	0.354	0.964
TCGA.ER.A193.06A.12R.A18S...	0.056	0.325	0	0.025	0.594	0	0.000	0.353	1.022
TCGA.EE.A2GL.06A.11R.A18S...	0.534	0.144	0.13	0	0.193	0	0.000	0.348	0.969
TCGA.XV.AAZV.01A.11R.A40A...	0	0.068	0.068	0.082	0.781	0	0.000	0.348	1.140
TCGA.EB.A3XC.01A.11R.A239...	0.069	0	0.298	0.063	0.57	0	0.000	0.346	1.017
TCGA.EE.A2MJ.06A.11R.A18S...	0.39	0.264	0	0.032	0.313	0	0.000	0.342	0.968
TCGA.EE.A2GP.06A.11R.A18S...	0.084	0	0.146	0.017	0.752	0	0.000	0.342	1.129

# Interpretation of results

---

- GSEA based approaches: calculation of semi-quantitative abundance scores.
- Deconvolution: quantitatively estimate the fractions of individual cell types in a heterocellular tissue

**Methods that allow between-sample comparisons (*In patient A, there are more CD8+ T cells than in patient B*)**

- MCP-counter
- xCell
- ConsensusTME
- TIMER

**Methods that allow between-cell-type comparisons (*In a certain patient, there are more B cells than T cells*)**

- CIBERSORT

**Methods that allow both**

- EPIC
- quanTIseq
- CIBERSORT absolute mode

# Tools' evolution

---

Most tools rely on cell marker signatures or reference signature matrices obtained from bulk gene expression profiling of purified populations



Single cell  
RNA-SEQ



**New deconvolution tools based on scRNA-Seq data**

- CIBERSORTx (Newman et al. 2019)
- SCDC (Dong et al. 2021)
- MuSiC (Wang et al. 2019)
- deconvSeq (Du et al. 2019)



**New cell marker signatures for GSEA-based approaches**

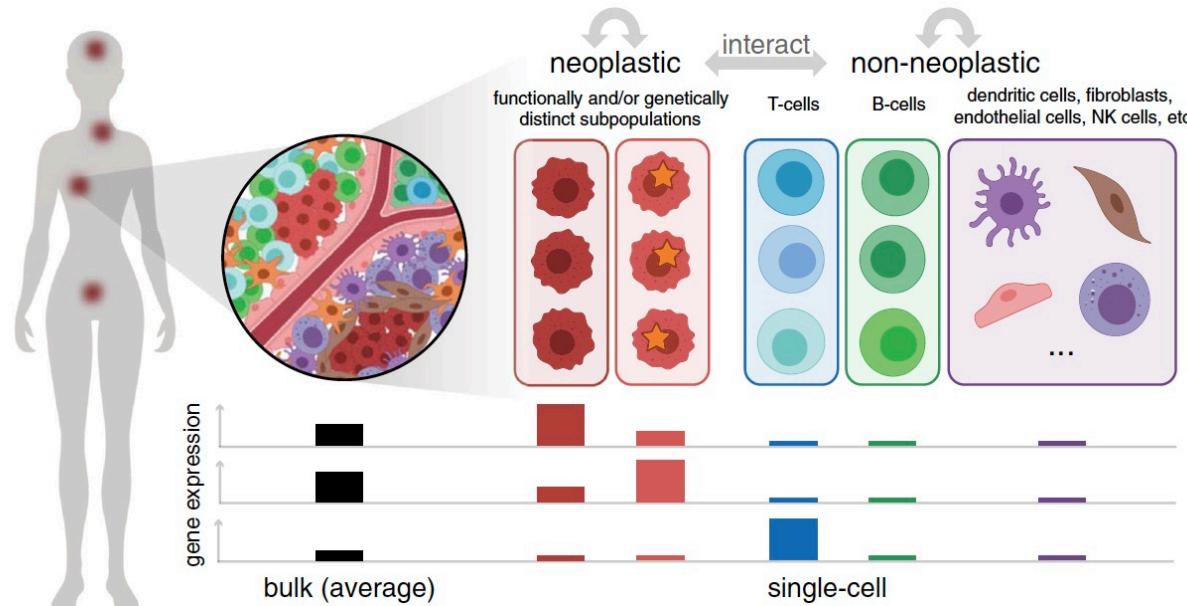
- [CancerSEA](#)
- [MSigDB C8 collection](#)
- [Single-Cell Tumor Immune Atlas](#)

# Single cell RNA-Seq

scRNA-Seq unravels tissue heterogeneity:

- Functional states of cell types (e.g.: functional or exhausted CD8+ T cells)
- Identification of rare or unknown cell types
- Study of tumor cell sub-populations

⚠ Care when using scRNA-seq techniques for quantifying the cellular composition of tumors: differences in single-cell dissociation efficiency relative to immune cells can bias cell-type proportions.



Fan et al. Experimental & Molecular Medicine. 2020

# Single cell RNA-Seq

## Methods for studying single cells

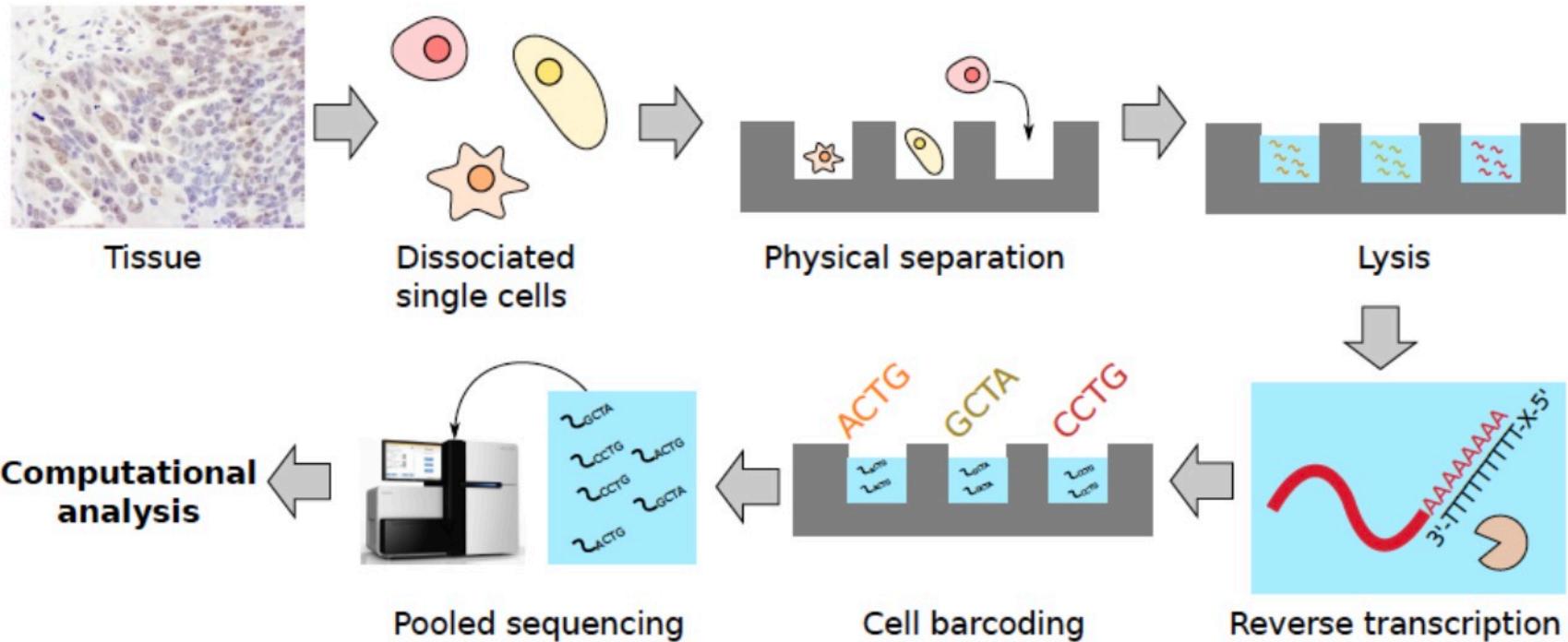
Every method has it's pros and cons. There is no all-encompassing single cell methodology → **It depends on your biological question**

Technology	Measurements (P)	Cells (N)	Throughput	Pro	Con
Flow cytometry	1-15	1k-100k	big N, small P	Technically easy	Limited targets
Mass cytometry	20-50	1k-100k	big N, medium P	>P than flow	Limited targets
RNA FISH	1	~100	small N, small P	Spatial resolution	Technically hard, low throughput
Multiplex FISH	~100	100's	medium N, medium P	Spatial resolution	Technically and analytically hard
SS2 scRNA-seq	~20,000	100-1000	medium N, big P	High throughput	Sparse, low input material
Droplet scRNA-seq	~20,000	100-1M	big N, big P	High throughput	Very sparse, low input material

# Single cell RNA-Seq

## A typical scRNA-Seq experiment

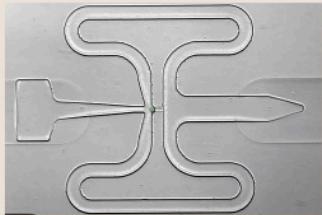
- Dissociation can be easy (blood) or hard (collagenous tissue)
- Separation and reverse transcription differ by protocol



# Single cell RNA-Seq

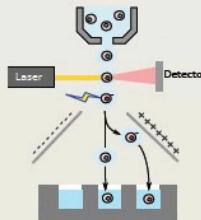
Physical separation defines main scRNA-seq protocols

## Microfluidic device



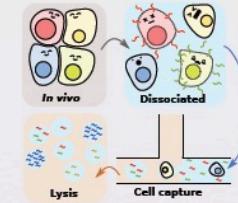
- ❖ 96 or 800 well format
- ❖ Physically check presence of cells
- ❖ High capture efficiency
- ❖ Doublet issues
- ❖ Expensive
- ❖ Full-length cDNA (SMART-seq{2})
- ❖ Spike-in control RNA
- ❖ **High gene coverage**

## Plate-based



- ❖ 96 or 384 well format
- ❖ Sort specific population(s) of cells
- ❖ High capture efficiency
- ❖ Experimental design considerations
- ❖ Full-length cDNA (SMART-seq(2) or end-tagging; UMIs)
- ❖ Spike-in control RNA
- ❖ **High gene coverage**

## Droplet-based



- ❖ 100-1000's of cells
- ❖ Doublet issues
- ❖ Variable capture efficiency
- ❖ Low per-cell cost
- ❖ 3' end tag; UMIs
- ❖ No spike-in control RNA
- ❖ **High cell coverage**

# Single cell RNA-Seq

---

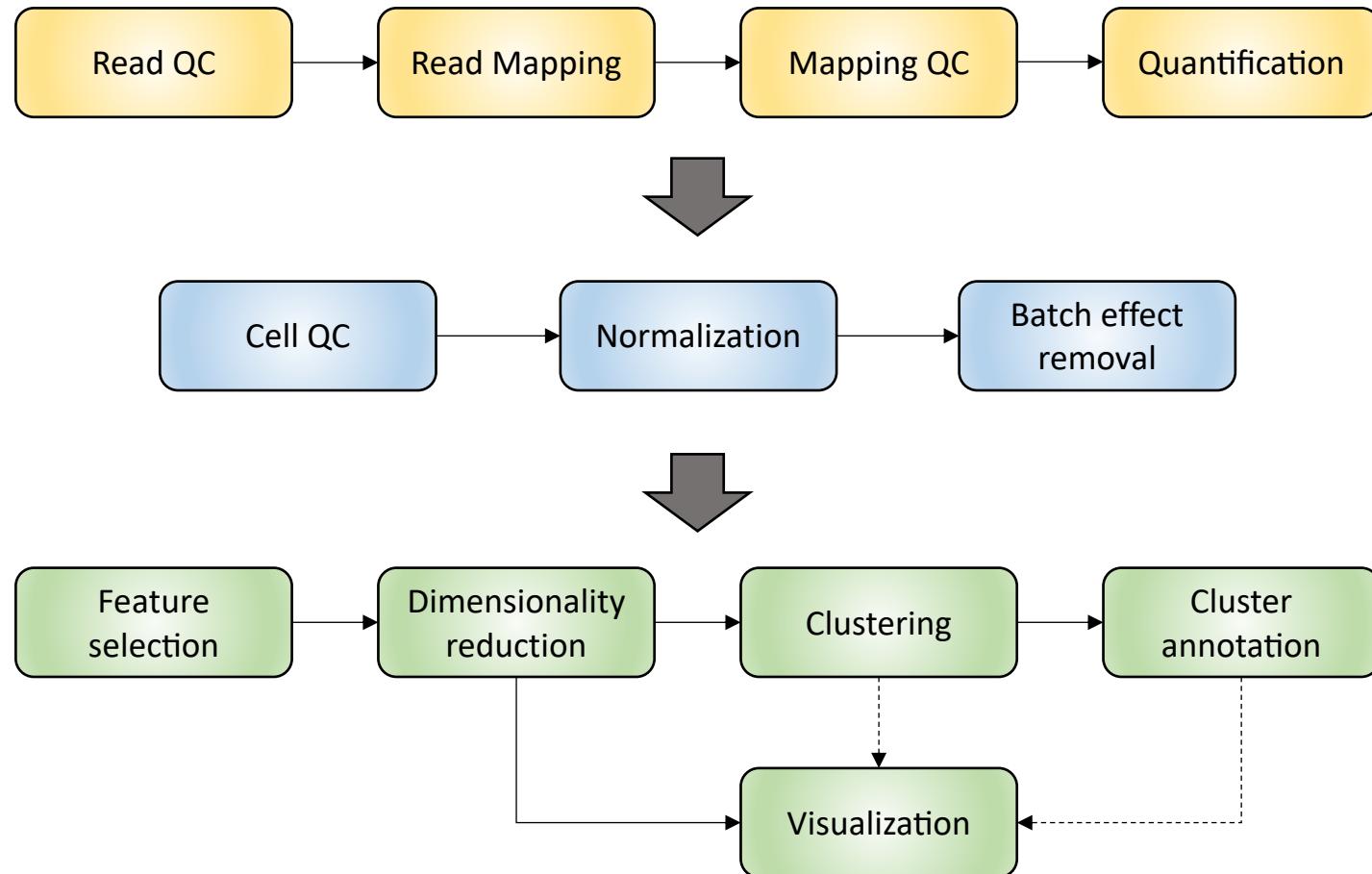
## Summary of scRNA-Seq approaches

Protocol example	C1 (SMARTer)	Smart- seq2	MATQ- seq	MARS-seq	CEL-seq	Drop-seq	InDrop	Chromium	SEQ-well	SPLIT-seq
Transcript data	Full length	Full length	Full length	3'-end counting						
Platform	Microfluidics	Plate-based	Plate-based	Plate-based	Plate-based	Droplet	Droplet	Droplet	Nanowell array	Plate-based
Throughput (number of cells)	$10^2\text{--}10^3$	$10^2\text{--}10^3$	$10^2\text{--}10^3$	$10^2\text{--}10^3$	$10^2\text{--}10^3$	$10^3\text{--}10^4$	$10^3\text{--}10^4$	$10^3\text{--}10^4$	$10^3\text{--}10^4$	$10^3\text{--}10^5$
Typical read depth (per cell)	$10^6$	$10^6$	$10^6$	$10^4\text{--}10^5$	$10^4\text{--}10^5$	$10^4\text{--}10^5$	$10^4\text{--}10^5$	$10^4\text{--}10^5$	$10^4\text{--}10^5$	$10^4$
Reaction volume	Nanoliter	Microliter	Microliter	Microliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Microliter
Reference	[63]	[57]	[39]	[10]	[64]	[45]	[46]	[47]	[101]	[38]

Haque et al. Genome Medicine. 2017

# Single cell RNA-Seq

## Core analytical workflow of scRNA-Seq data

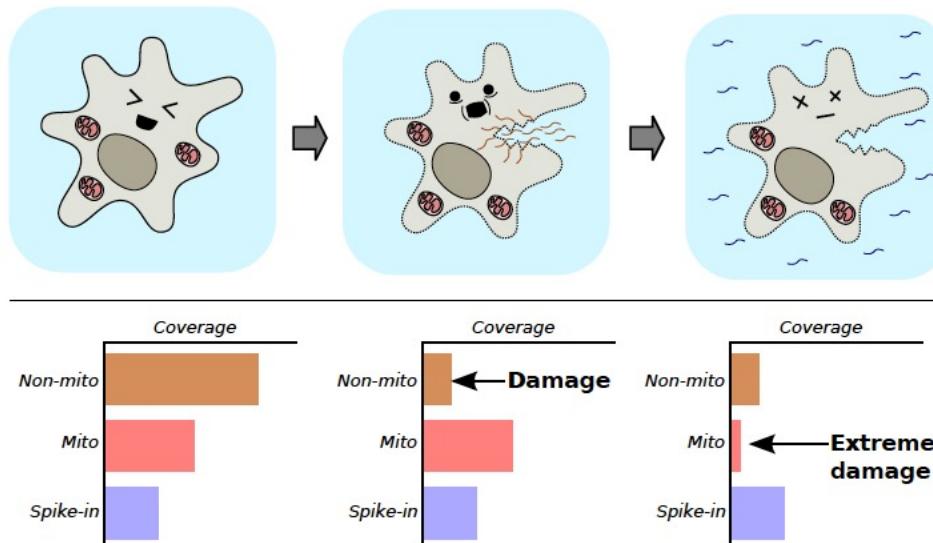


# Single cell RNA-Seq

## A little bit of details about specific scRNA-Seq analysis steps

### ➤ Cell QC

- Removal of low-quality cells
  - Low sequencing depth
  - Low number of detected genes
  - Damaged or dead cells (evaluation of expression of mitochondrial genes)
  - Doublets (pairs of cells captured and sequenced together. Mixed transcriptomes, falsely interpreted as intermediate cell phenotypes)



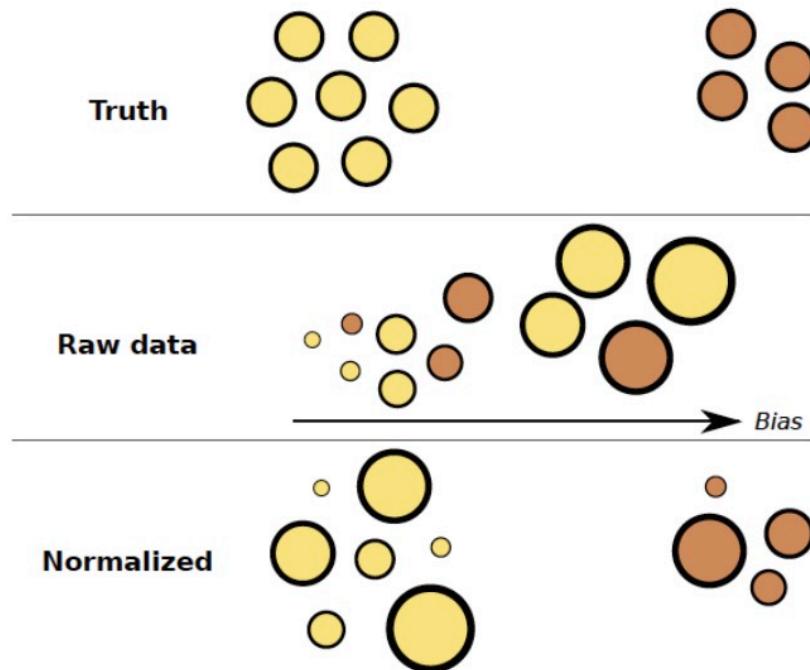
# Single cell RNA-Seq

---

## A little bit of details about specific scRNA-Seq analysis steps

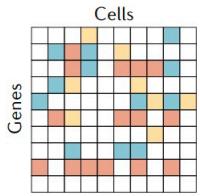
### ➤ Normalization

- The aim is bring all cells onto the same distribution to remove biases between them
  - We want to preserve biological variability, not introduce new technical variation
  - Primary source of bias is sequencing depth: scale down counts accordingly



# Single cell RNA-Seq

## Feature selection and clustering



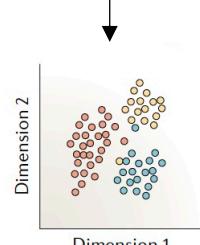
Expression matrices:

- High dimensionality: high number of genes measured in thousands of cells
- Data sparsity: data mainly composed of zeros. Mainly due to dropouts
- Dropouts: expressed genes result in null expression values due to the inefficiency of mRNA capture and/or to the stochasticity of mRNA expression



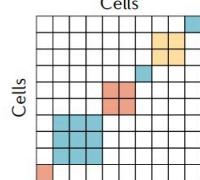
Feature selection

- Identification of most informative genes to reduce noise (e.g.: genes with high variability between cells)



Dimensionality reduction

- Reducing the number of features of a data set to obtain a set of principal features  
Original data are projected onto a lower-dimensional sub-space (PCA, t-SNE, UMAP)
- Much easier to visualize the data in a 2 or 3-dimensional subspace



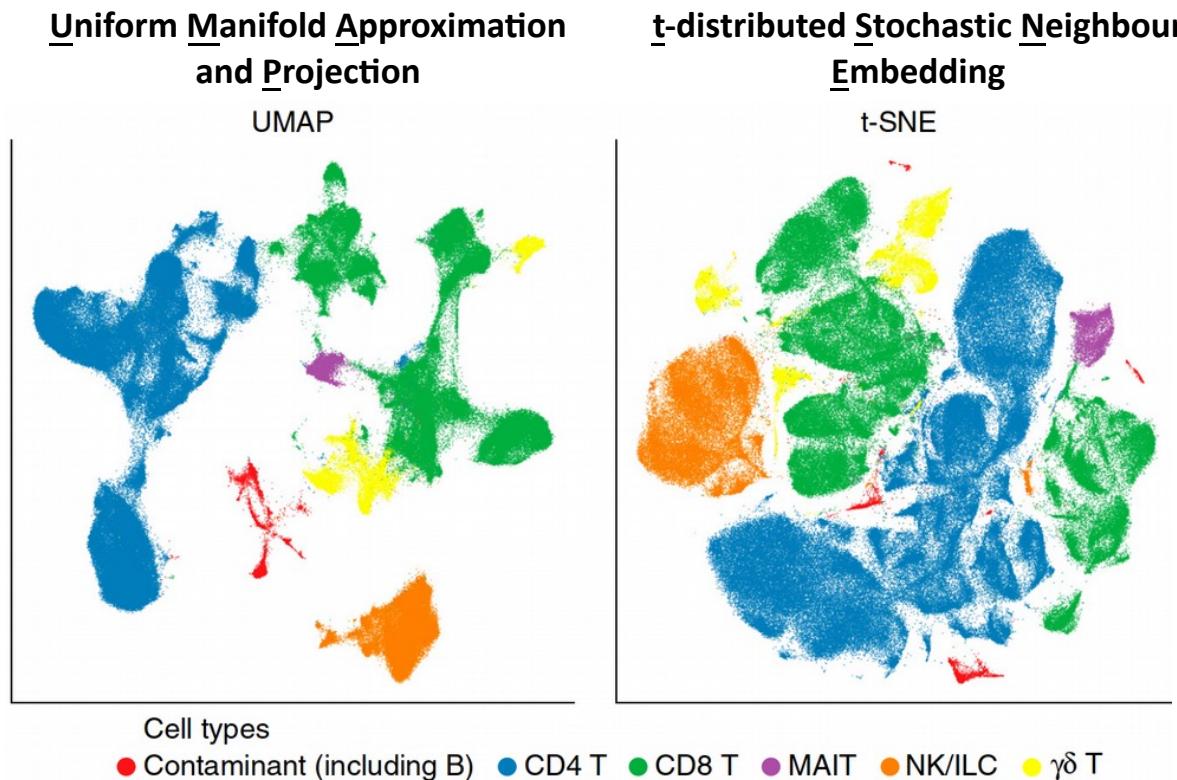
Cell clustering

- Identify groups of cells based on the similarities of the transcriptomes without any prior knowledge of the labels. Moreover, in most situations we do not even know the number of clusters *a priori*. scRNA-Seq specific methods

# Single cell RNA-Seq

## Data visualization

- Graphically represent functionally related groups of cells as clusters with similar gene expression profiles in 2D plots.
- Closer cells have more similar profiles.
- Results depend on input parameters.



# References and additional material

---

## Immune quantification from bulk RNA-Seq data

1. Hackl, H., Charoentong, P., Finotello, F. & Trajanoski, Z. Computational genomics tools for dissecting tumour–immune cell interactions. *Nat. Rev. Genet.* **17**, 441–458 (2016).
2. Finotello, F. & Trajanoski, Z. Quantifying tumor-infiltrating immune cells from transcriptomics data. *Cancer Immunol. Immunother.* **67**, 1031–1040 (2018).
3. Finotello, F., Rieder, D., Hackl, H. & Trajanoski, Z. Next-generation computational tools for interrogating cancer immunity. *Nat. Rev. Genet.* **20**, 724–746 (2019).
4. Liu, C. C., Steen, C. B. & Newman, A. M. Computational approaches for characterizing the tumor immune microenvironment. *Immunology* **158**, 70–84 (2019).
5. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
6. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
7. Aran, D. Cell-Type Enrichment Analysis of Bulk Transcriptomes Using xCell. in *Bioinformatics for Cancer Immunotherapy: Methods and Protocols* (ed. Boegel, S.) 263–276 (Springer US, 2020). doi:10.1007/978-1-0716-0327-7\_19.
8. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. in *Cancer Systems Biology: Methods and Protocols* (ed. von Stechow, L.) 243–259 (Springer, 2018). doi:10.1007/978-1-4939-7493-1\_12.

## scRNA-Seq

1. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 75 (2017).
2. Camara, P. G. Methods and challenges in the analysis of single-cell RNA-sequencing data. *Curr. Opin. Syst. Biol.* **7**, 47–53 (2018).
3. Papalexis, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* **18**, 35–45 (2018).
4. Fan, J., Slowikowski, K. & Zhang, F. Single-cell transcriptomics in cancer: computational challenges and opportunities. *Exp. Mol. Med.* **52**, 1452–1465 (2020).
5. [10x support page with many introductory materials](#)