



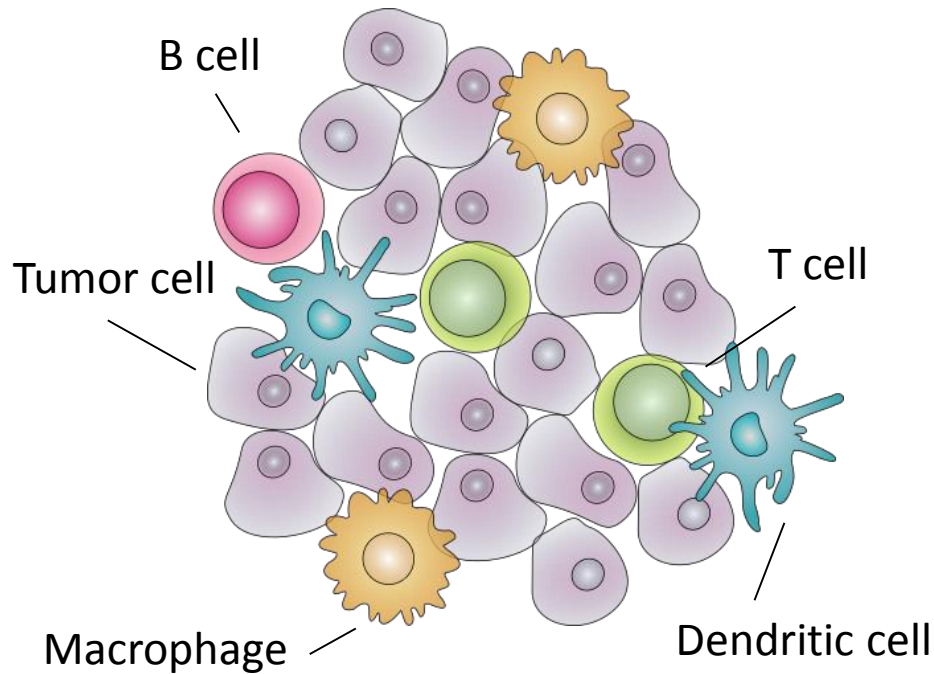
# IO17 | Large Scale Bioinformatics for Immuno-Oncology

Computational quantification of tumor-infiltrating immune cells

Francesca Finotello, Federica Eduati, and Pedro L. Fernandes

GTPB | The Gulbenkian Training Programme in Bioinformatics  
Instituto Gulbenkian de Ciência, Oeiras, Portugal | Sept 19th-22nd, 2017

# Computational quantification of tumor-infiltrating immune cells



Bulk tumors are mixtures of tumor and normal cells (also immune cells)

RNA-seq/microarrays → cumulative expression of the mixture of cells

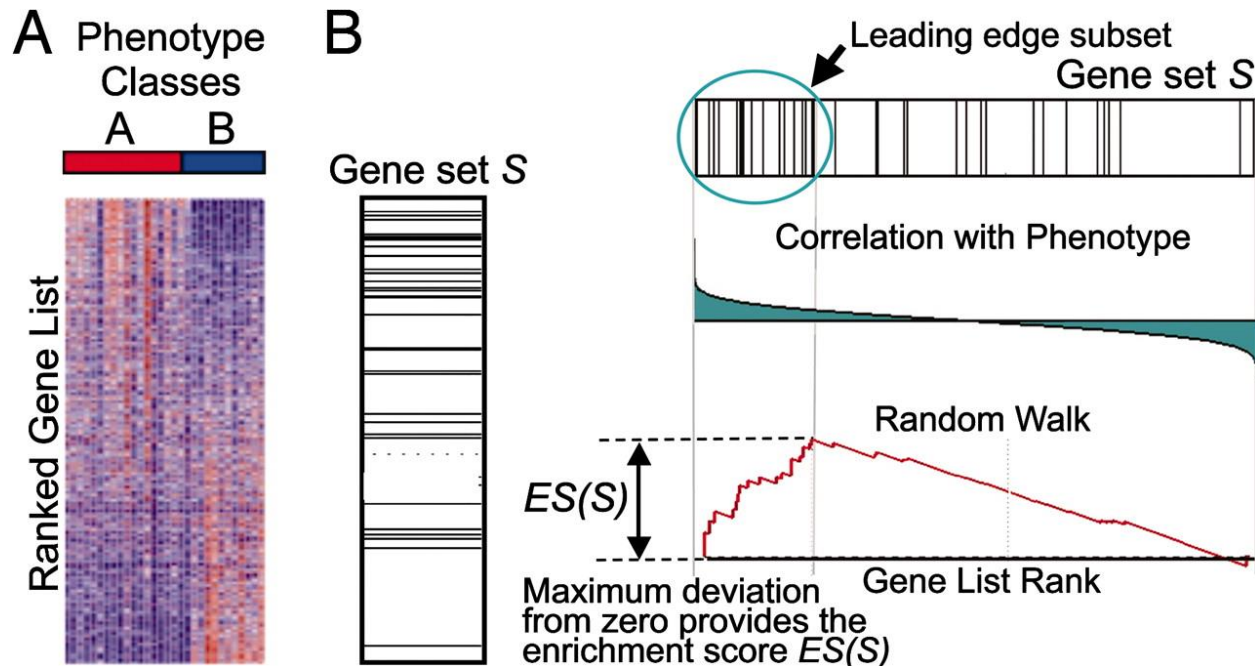
Expression data from microarrays or RNA-seq of bulk tumors or cell mixtures can be analyzed with computational approaches based on:

- Gene set enrichment analysis (GSEA) → semi-quantitative enrichment score (ES)
- Deconvolution → relative cell fractions

# Gene Set Enrichment Analysis

**GSEA:** ranks genes by correlation with a phenotype and determines whether an *a priori* defined gene-set  $S$  shows statistically significant enrichment at either end of the ranked list  $L$

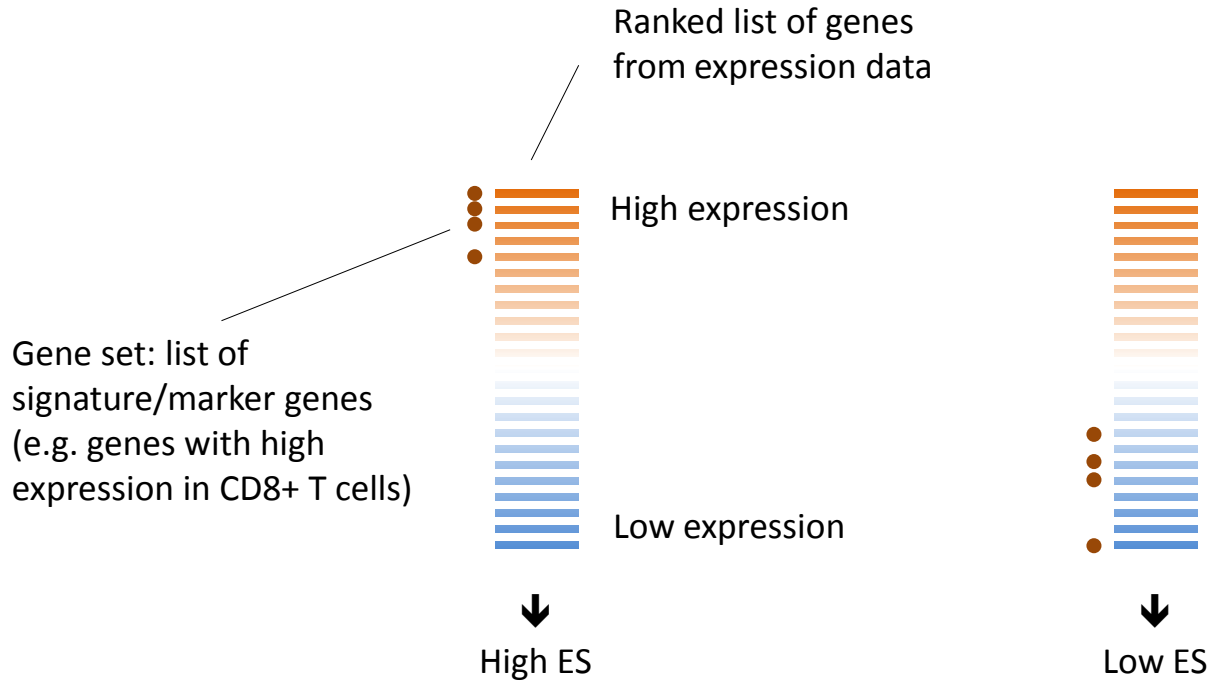
**Enrichment score (ES):** reflects the degree to which a gene set  $S$  is overrepresented at the extremes (top or bottom) of the ranked list  $L$



Walk down the list  $L$  and increase a running-sum statistic when a gene that is present in  $S$  (decrease if not)

→ ES computed as the maximum deviation from zero

# Enrichment of immune cells



## Gene sets from immune cells:

- G Bindea et al., Immunity, 2013: 28 immune cells types infiltrating CRC
- M Angelova et al., Genome biology, 2015: 28 immune cells types infiltrating CRC
- P Charoentong et al., Cell reports, 2017: 28 immune cell types infiltrating solid cancers
- D Aran et al., bioRxiv, 2017: 64 cell types, 489 gene sets, xCell method
- P Danaher et al., Journal for immunotherapy of cancer, 2017: 15 cell types/states
- (J Godec et al., Immunity, 2016: ~5000 gene sets from immune cell state/type changes)



## CIBERSORT: Cell-type identification by estimating relative subsets of RNA transcripts

**Signature matrix:** LM22 defined from microarray data from 22 human immune cell types/phenotypes (547 signature genes)

**Algorithm:** based on linear support vector regression (SVR)

Robust to:

- Closely related cell types (multicollinearity)
- Noise or unknown mixture content (in solid tumors, <50% immune cells)

