IO17 | Large Scale Bioinformatics for Immuno-Oncology

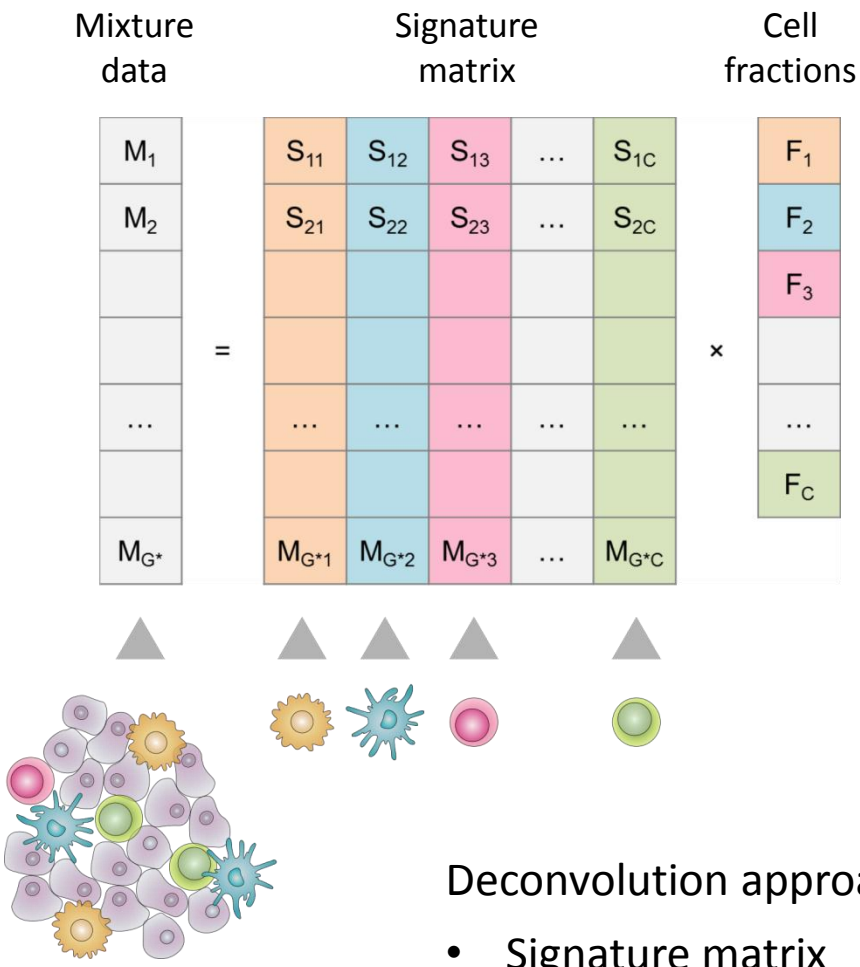
Deconvolution methods

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

INSTITUTO GULBENKIAN DE CIÊNCIA

Deconvolution of cell mixtures



$$\mathbf{M} = \mathbf{S} \times \mathbf{F}$$

$$M_1 = S_{11} F_1 + S_{12} F_2 + S_{12} F_3 + \ldots + S_{1C} F_C$$

$$M_2 = S_{21} F_1 + S_{22} F_2 + S_{13} F_3 + \ldots + S_{2C} F_C$$

…

$$M_{G*} = S_{G*1} F_1 + S_{G*2} F_2 + S_{G*3} F_3 + \ldots + S_{G*C} F_C$$

Deconvolution approach:

- Signature matrix
- Computational method to solve the inverse problem

Computational tools for the deconvolution* of cell fractio

Tool	Deconvolution method	Signature	Reference
EPIC	Constrained least-square regression	5 immune cell types plus uncharacterized cells (RNA-seq)	J Racle et al, bioRxiv, 2017
TIMER	Linear least-square regression	9 immune cell types (no cell fractions but scores)	B Li et al, Genome biology, 2016
CIBERSORT	Support vector regression	22 immune cell phenotypes	AM Newman et al, Nature methods, 2015
DeconRNASeq	Non-negative least-squares solved with quadratic programming	5 tissues from the Human BodyMap 2.0, but no immune cells (RNA-seq data)	T Gong and JD Szustakowski, Bioinformatics, 2013
PERT	Perturbation model (account for variations btw. the mixture and signature	11 immune cell types and progenitors	W Qiao et al, PLoS computational biology, 2012
-	Linear least-square regression	17 immune cell types	AR Abbas et al, PloS one, 2009
xCell	Cell fractions derived with ssGSEA (*no decon.)	489 gene sets for 64 cell types	D Aran et al, bioRxiv, 2017
MCP-counter	Abundance score as geometric mean of markers (*no decon.)	8 immune cell types, endothelial cells, and fibroblasts	B Becht et al, Genome biology, 2016

The CellMix package

The CellMix R package (R Gaujoux and C Seoighe, Bioinformatics, 2013) facilitates the exploration, assessment and deconvolution of gene expression data, by integrating:

- Benchmark data sets (with a gold-standard)
- Signature matrices
- Methods for partial and complete deconvolution

- **Partial deconvolution methods**: assume that either the signature matrix or the cell proportions are known, and estimate the unknown cell fractions of the cell-type-specific expression profiles, respectively.

- **Complete deconvolution methods:** infer both cell-type expression profiles and fractions from the mixture matrix, possibly using a priori information (e.g. marker genes)