



**IO17 | Large Scale Bioinformatics for Immuno-Oncology**

**Deconvolution: questions - Solution**

**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

## Deconvolution questions

1) What is the number of LM22 signature genes that are present in the PBMC mixture matrix?

**507**

```
sum(rownames(mixture) %in% rownames(signature))
```

## Deconvolution questions

2) What is the percentage of LM22 signature genes that are present in the PBMC mixture matrix? Report the percentage rounded to 2 decimals, with no space before the “%” and with “.” as decimal separator (e.g. 60.2267% should be rounded to 60.23%).

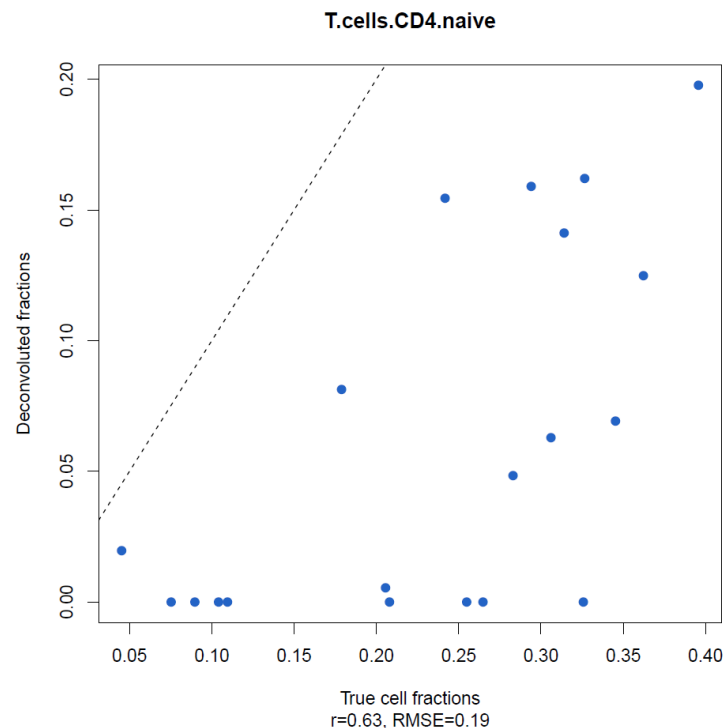
**92.69%**

```
round( sum(rownames(mixture) %in%  
       rownames(signature) ) / nrow(signature)*100, digits=2)
```

3) By comparing CIBERSORT cell fractions with flow cytometry, which cell-type estimates obtained the highest error?

## Naïve CD4+ T cells

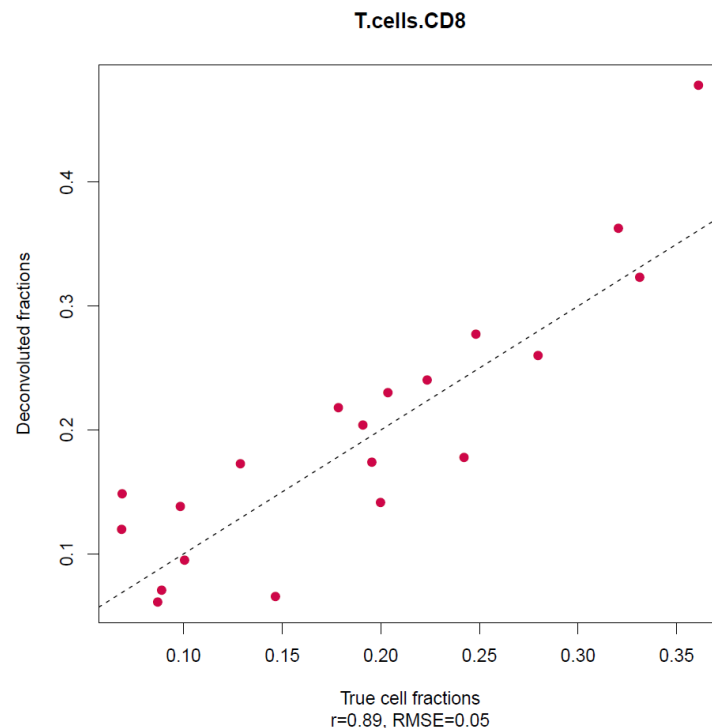
Cell type	r	RMSE
B.cells.naive	0.61	0.04
B.cells.memory	0.68	0.01
T.cells.CD8	0.89	0.05
T.cells.CD4.naive	0.63	0.19
T.cells.CD4.memory.resting	0.50	0.09
T.cells.CD4.memory.activated	0.62	0.02
T.cells.gamma.delta	0.29	0.05
NK.cells	0.79	0.05
Monocytes	0.75	0.13
All cell types	0.62	0.09



4) By comparing CIBERSORT cell fractions with flow cytometry, which cell-type estimates obtained the highest Pearson's correlation?

## CD8+ T cells

Cell type	r	RMSE
B.cells.naive	0.61	0.04
B.cells.memory	0.68	0.01
<b>T.cells.CD8</b>	<b>0.89</b>	0.05
T.cells.CD4.naive	0.63	0.19
T.cells.CD4.memory.resting	0.50	0.09
T.cells.CD4.memory.activated	0.62	0.02
T.cells.gamma.delta	0.29	0.05
NK.cells	0.79	0.05
Monocytes	0.75	0.13
All cell types	0.62	0.09

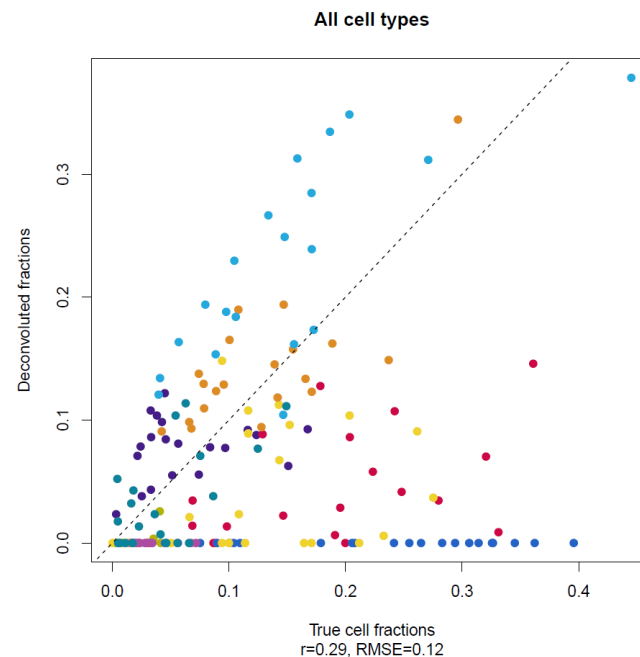
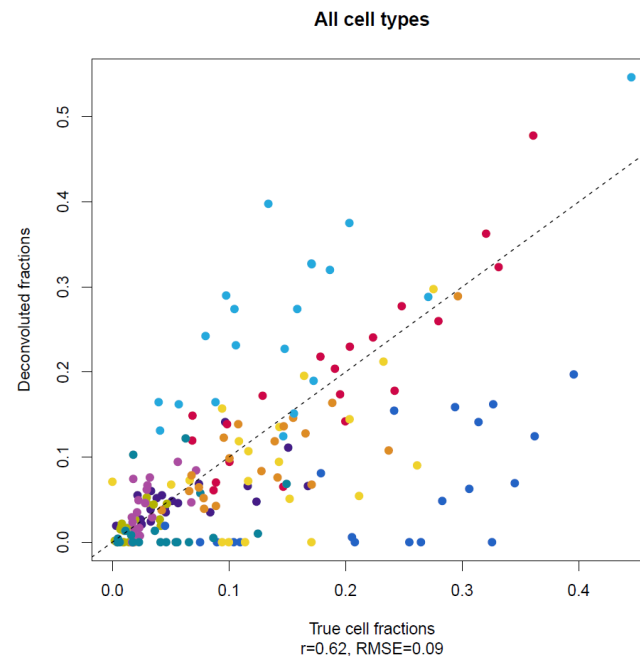


# Deconvolution questions

5) Which method performed better in the deconvolution of the PBMC data?

## CIBERSORT

Cell type	CIBERSORT		DeconRNASeq	
	r	RMSE	r	RMSE
B.cells.naive	0.61	0.04	0.21	0.05
B.cells.memory	0.68	0.01	0.41	0.02
T.cells.CD8	0.89	0.05	0.47	0.16
T.cells.CD4.naive	0.63	0.19	-	0.26
T.cells.CD4.memory.resting	0.5	0.09	0.15	0.12
T.cells.CD4.memory.activated	0.62	0.02	-	0.03
T.cells.gamma.delta	0.29	0.05	0.61	0.04
NK.cells	0.79	0.05	0.72	0.05
Monocytes	0.75	0.13	0.75	0.1
All cell types	0.62	0.09	0.29	0.12

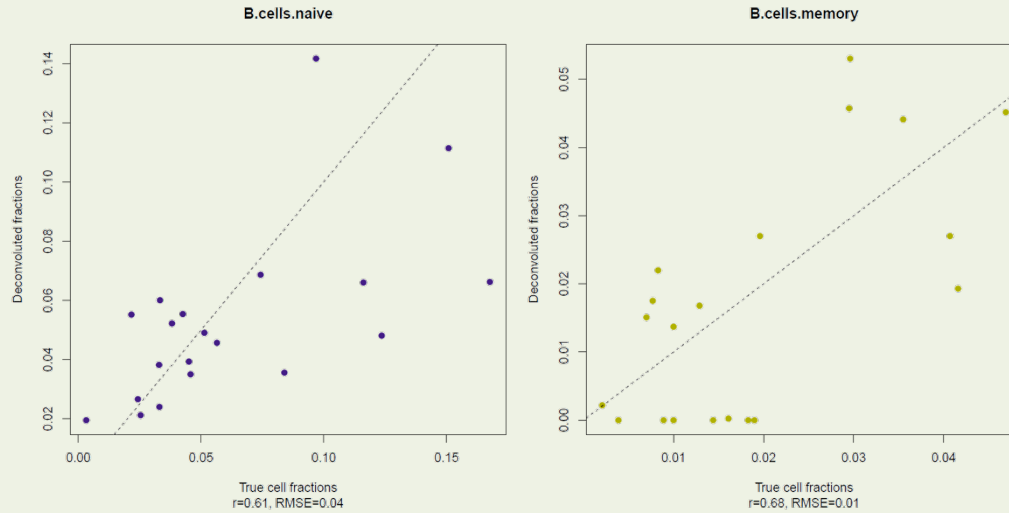


6) What might be the reasons of the underestimation of memory B cell fractions by DeconRNAseq? Select the **wrong** answer.

- Memory B cells have low abundance the mixtures (i.e. <5%)
- The LM22 matrix and the PBMC mixture matrix do not have enough genes in common
- Memory B cells are closely related to other cell types in the LM22 signature matrix

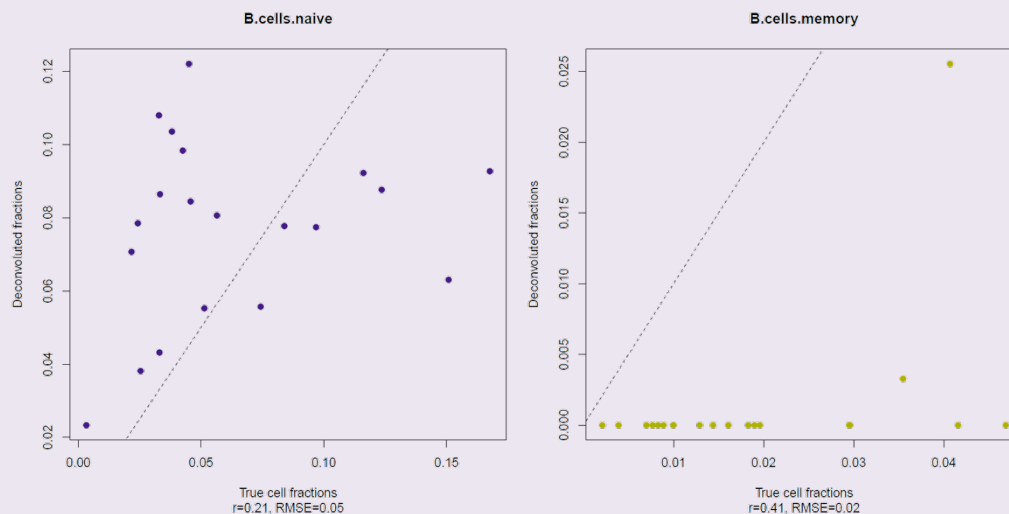
# Deconvolution questions

## CIBERSORT



- Memory B cells have low abundance (<5%)
- Multicollinearity: naïve and memory B cells
- CIBERSORT is robust to multicollinearity
- Memory B cell detected when naïve B cells are removed

## DeconRNASeq



## DeconRNASeq (no naïve B cells)

