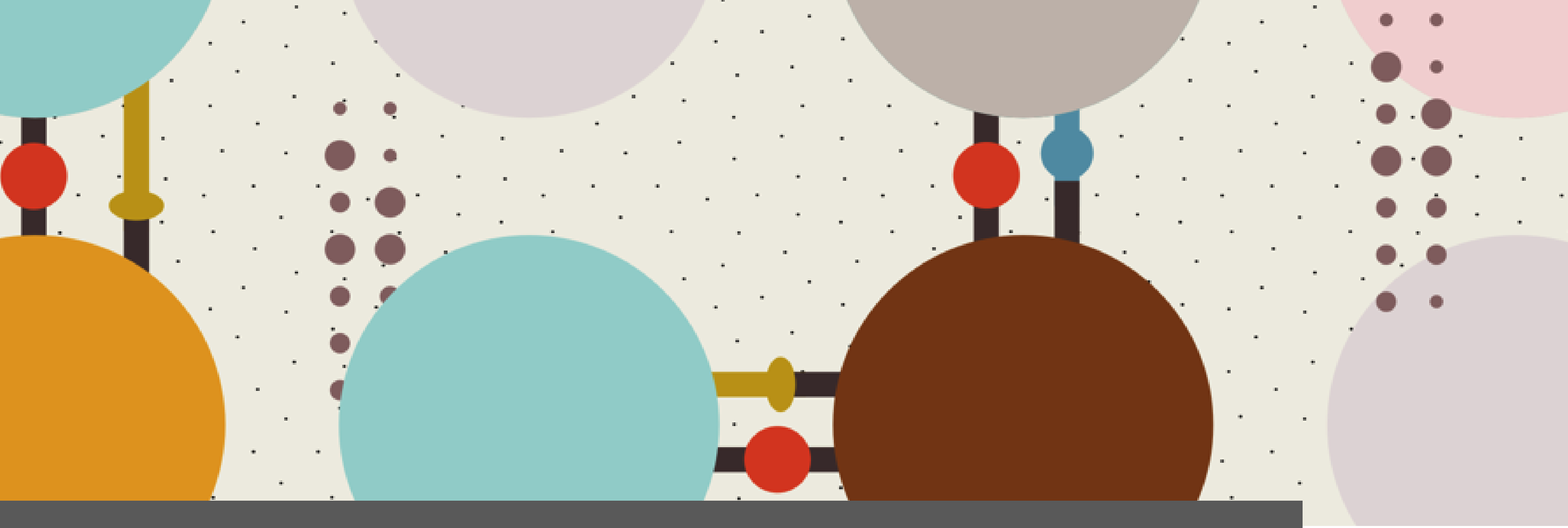


IO17 | Large Scale Bioinformatics for Immuno-Oncology

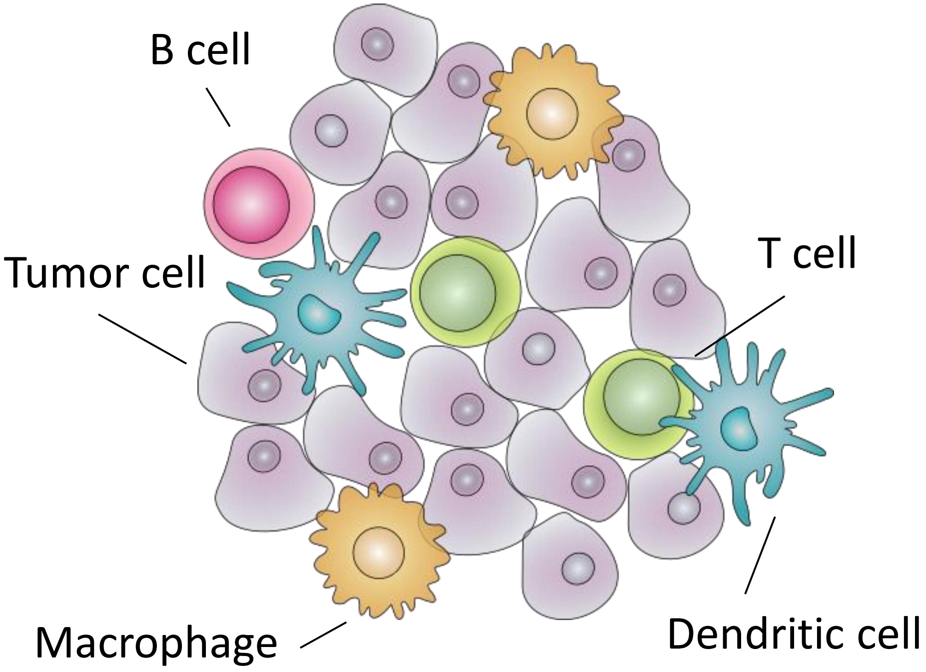
Assessing deconvolution performance

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

Quantification of tumor-infiltrating immune cells: expression da



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

Expression data of the mixture (RNA-seq/microarrays) → estimation of relative cell fractions with **deconvolution** algorithms

How can we assess whether deconvolution results are accurate?

We can measure their **agreement** with **gold-standard** measures

Quantification of tumor-infiltrating immune cells: cell suspensio

Coulter Counter

- Cells suspended in a low-concentration electrolyte pass through an aperture causing a short-term change in the impedance proportional to their volume, which is measured as a voltage/current pulse
- Used in hospital laboratory to compute of complete blood counts (CBC)

Flow cytometry

- Cells are attached to antibodies for specific markers (e.g. CD4 or CD8) attached to different fluorochromes.
- Single-cells are through a nozzle and stimulated with lasers
- Different detectors measures the light scatter (forward=size, side=granularity) and the fluorescence of specific fluorochormes (light intensity ~ maarker expression)

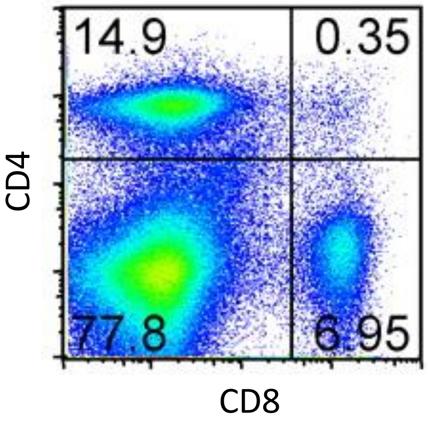
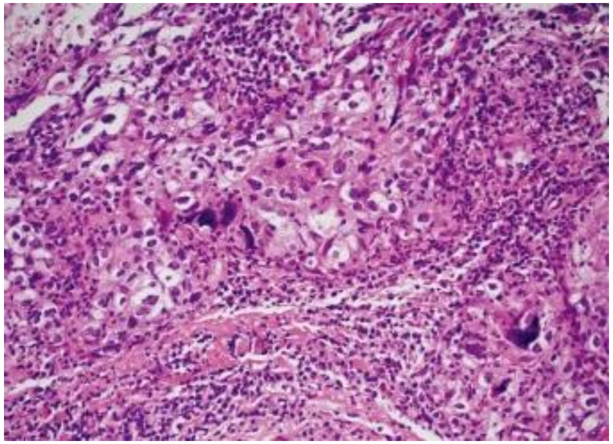


Figure adapted from S Zhang et al., Journal of Experimental Medicine, 2010

Quantification of tumor-infiltrating immune cells: tissue imag

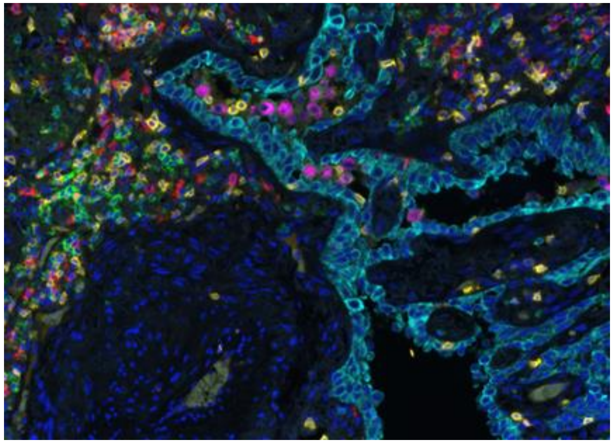


Hematoxylin and eosin stain (H&E)

- Hematoxylin: positive, blue/violet stain that binds to negatively-charged substances (e.g. DNA in the cell nucleus)
- Eosin: negative, red/pink stain that binds to positively-charged substances (e.g. proteins in the cell cytoplasm)
- Widely used tissue pathology

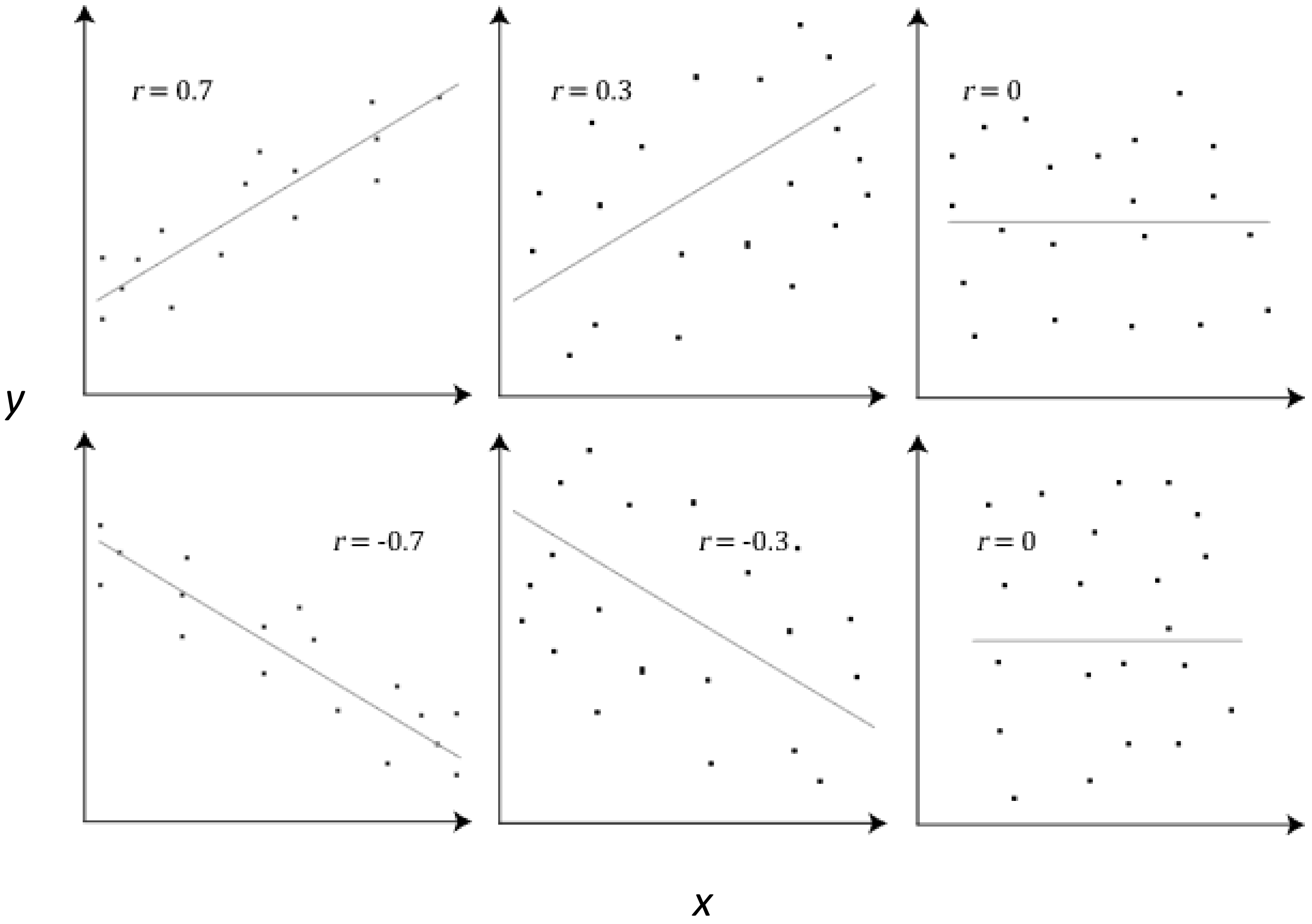
Immunohistochemistry (IHC)

- Imaging antigens from specific proteins expressed in the cells of a tissue ("histo") section using labeled antibodies ("immuno")
- Sample preparation is critical to maintain cell morphology, tissue architecture, and access to target antigens → formalin-fixed and paraffin-embedded (FFPE) samples



Pearson’ s correlation

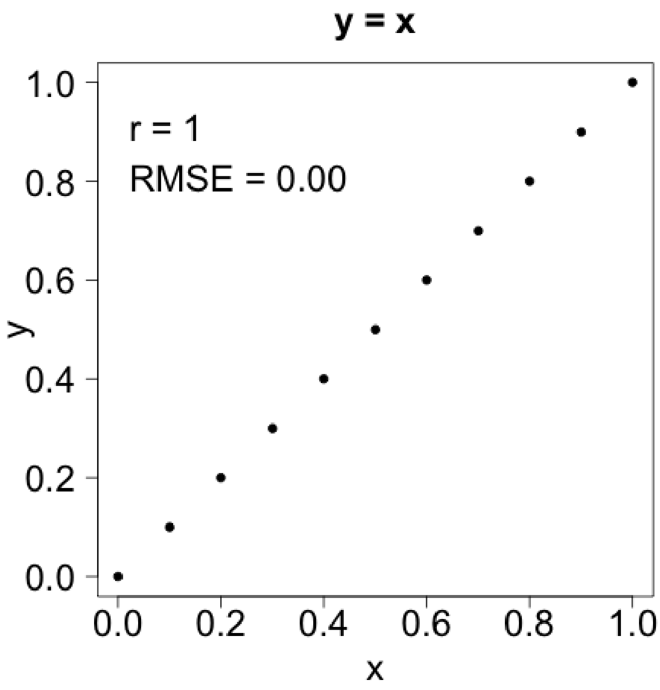
Pearson‘s correlation ***r*** measures of the strength of the **linear** association between two variables ***x*** and ***y***, ranging from -1 to 1

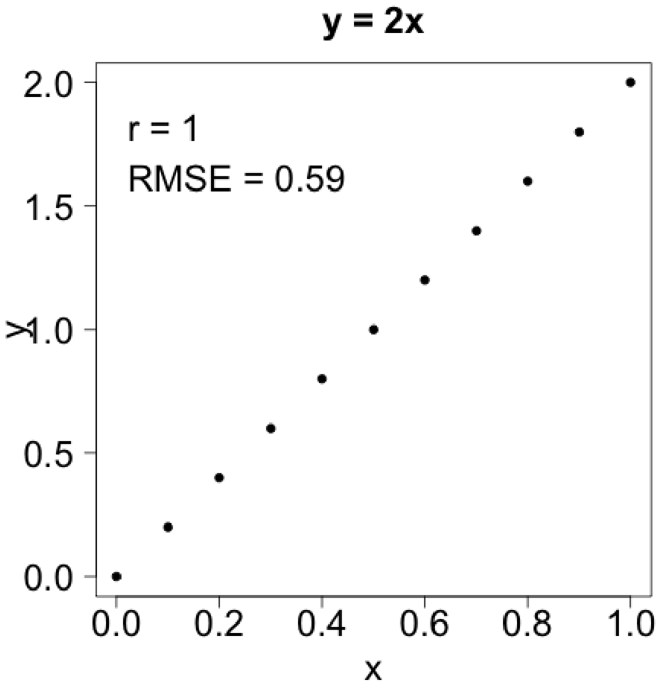


Root-mean-square error

The root-mean-square error (RMSE) or root-mean-square deviation (RMSD) measures the differences between the values predicted by a mode/estimator and the values actually observed:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y - x)^2}{n}} = \sqrt{mean[(y - x)^2]}$$





Descriptive statistics and plots: the Anscombe's quartet (1

x1	y1	x2	y2	x3	y3	x4	y4
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.5
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

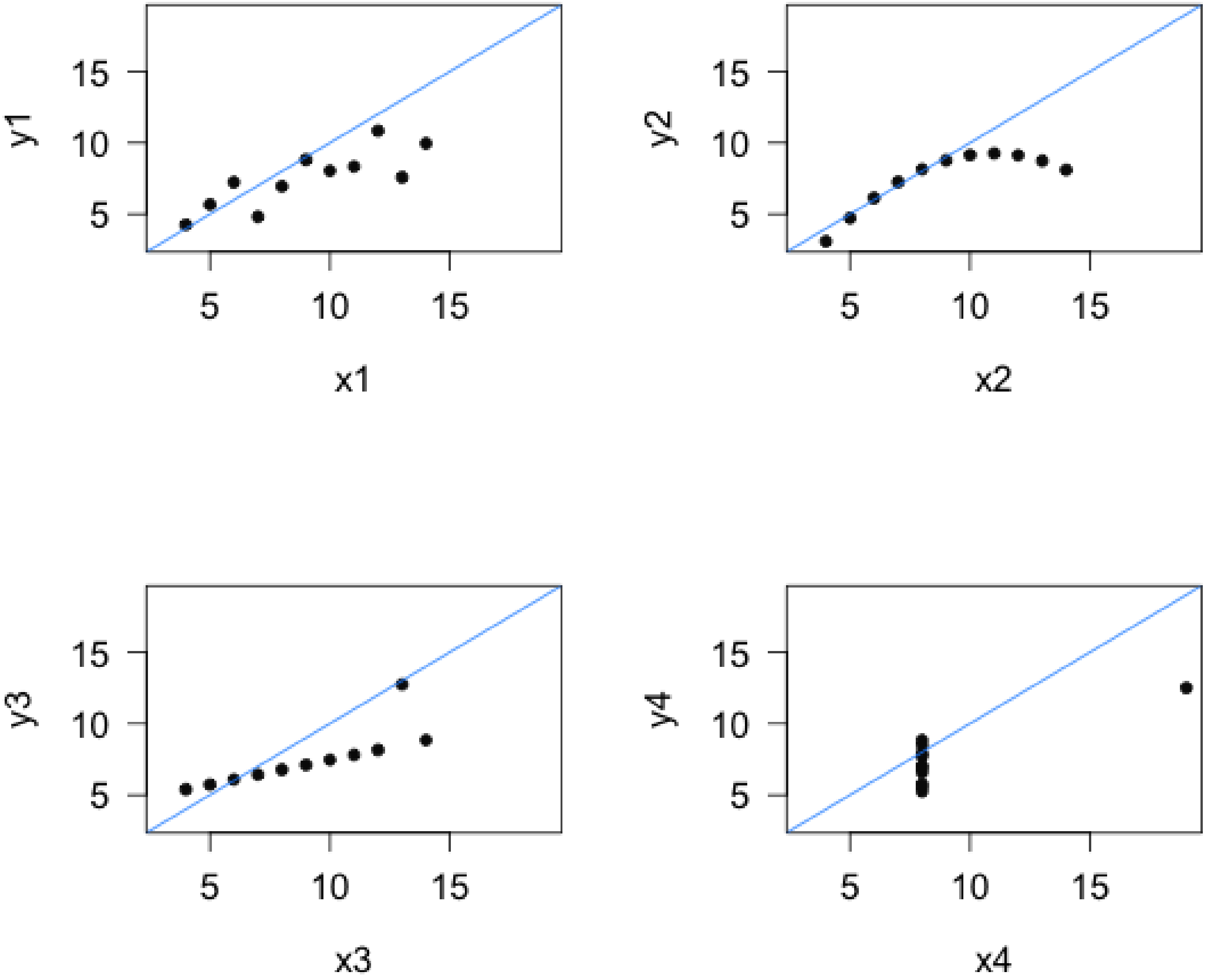
mean(x) = 9.0
var(x) = 11.0

mean(y) = 7.5
var(y) = 4.1

cor(x,y) = 0.82

Lin. Regression y~x:
y = 3x + 0.5

Descriptive statistics and plots: the Anscombe's quartet (2

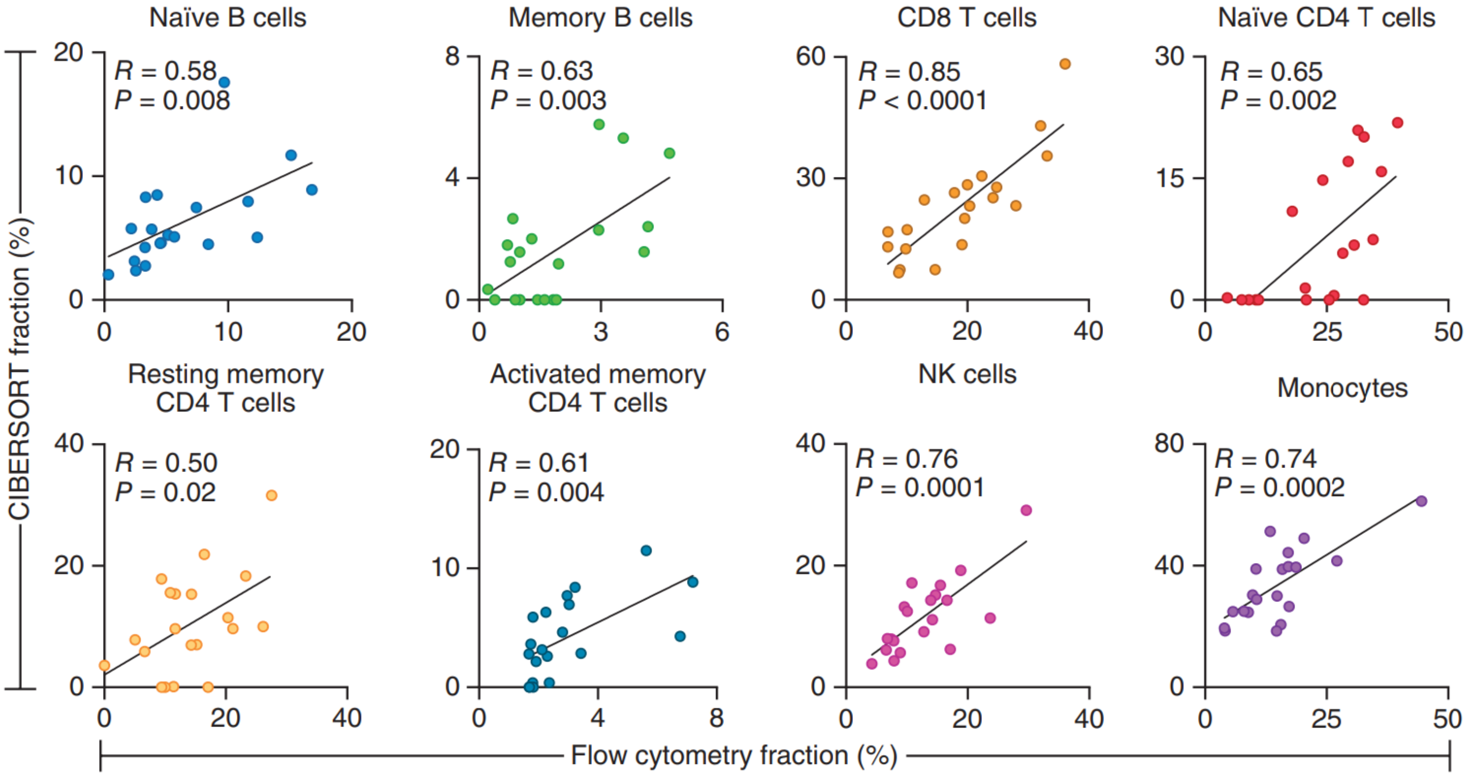


Example: CIBERSORT validation

cell type	corr	RMSE
Naïve B cells	0.58	3.75
Memory B cells	0.63	1.42
CD8 T cells	0.85	8.04
Naïve CD4 T cells	0.65	18.25
Resting memory CD4 T	0.50	8.38
Activated memory CD4 T	0.61	2.88
Gamma delta T cells	0.29	5.13
NK cells	0.76	4.39
Monocytes	0.74	20.29

Peripheral blood mononuclear cells (PBMC) from 20 healthy individuals:

- Mixture expression data (microarrays)
- Relative fractions from flow cytometry



Note: cell fractions in 0-100% (not 0-1)

AM Newman et al, Nature methods, 2015