

Independent component analysis of single cell RNA-seq data: from batch effect correction to biological processes

Petr V Nazarov
petr.nazarov@lih.lu

Poster 51

Challenge

A

batch 1

○ **Cell line 1**
(entire population)

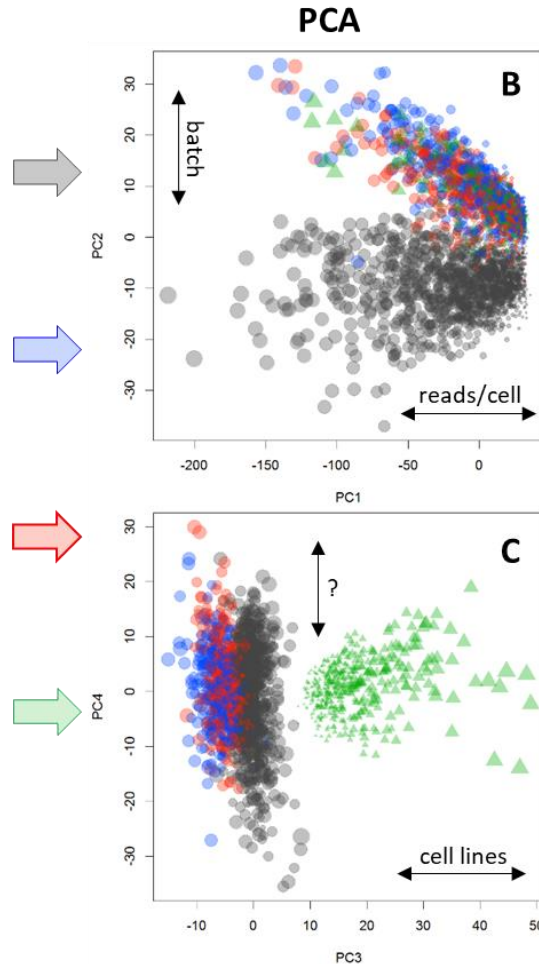
batch 2

○ **Cell line 1**
CD133+, CD44-,
A2B5+, CD15+

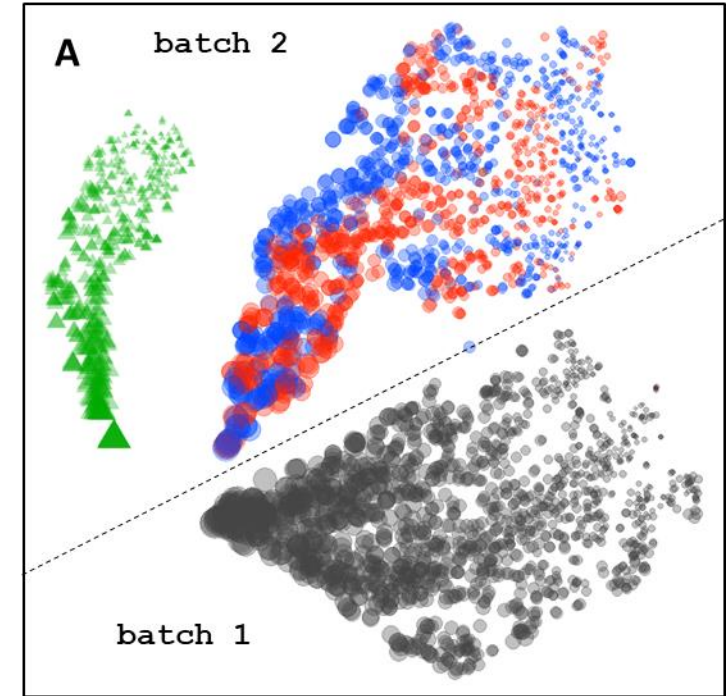
○ **Cell line 1**
CD133+, CD44+,
A2B5+, CD15+

△ **Cell line 2**
(entire population)

“Ill-posed” problem...

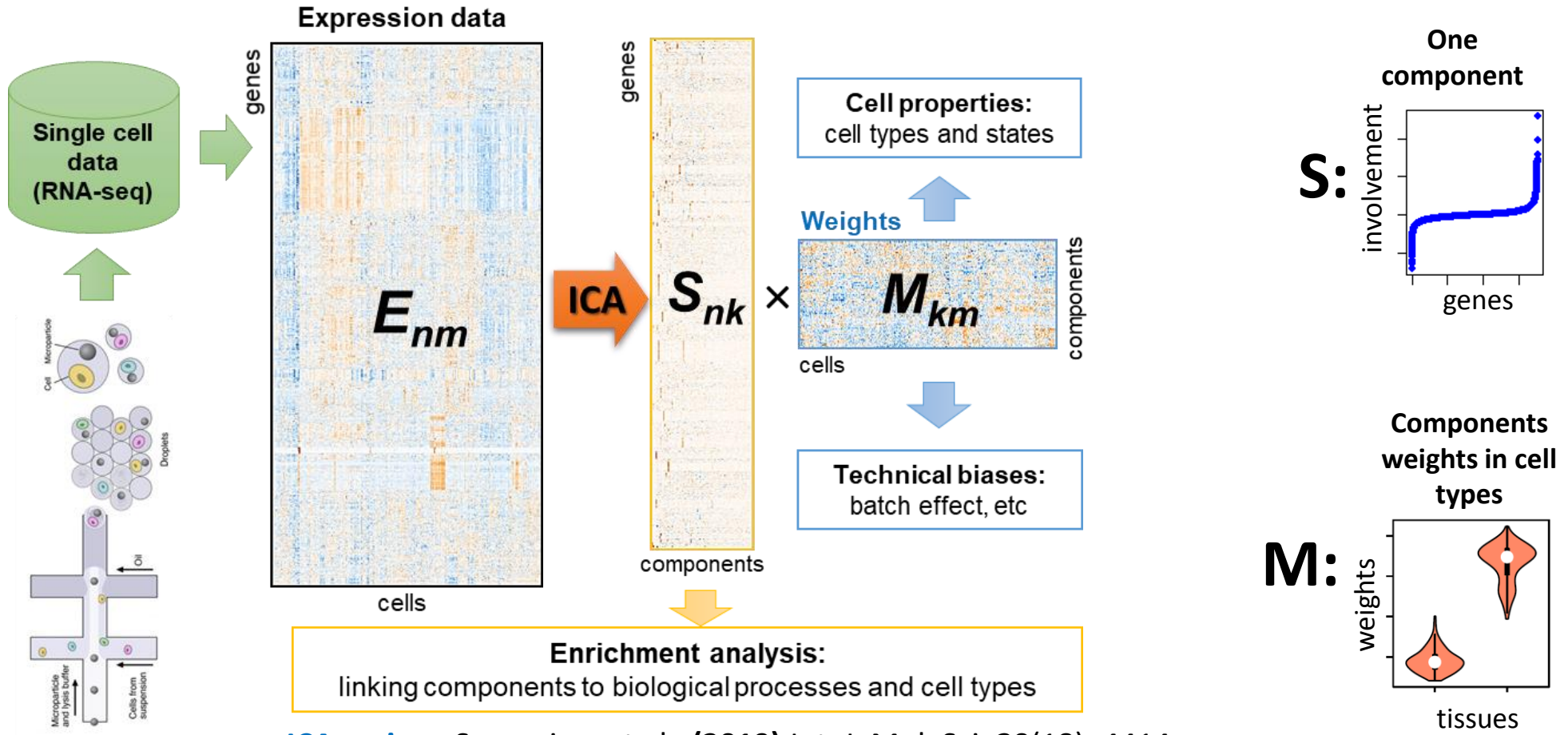


t-SNE



Protocol optimization at the scRNA-seq platform lead to a batch effect (A) between entire cell population (grey) and its subpopulations (red, blue). Difference between batches is larger than difference between two cell lines (B,C). Adapted from [1].

Method: Consensus ICA

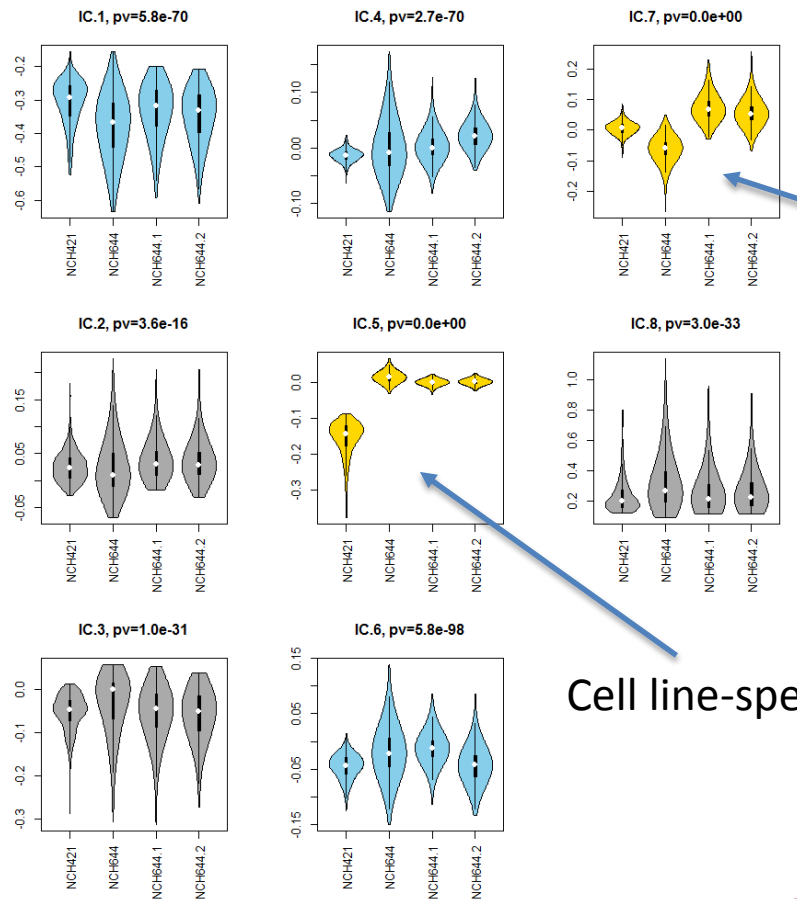
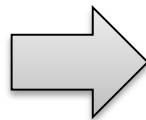
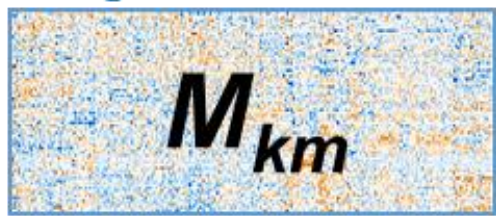


ICA review: Sompairac et al., (2019) Int. J. Mol. Sci, **20**(18): 4414

Tool and application: Nazarov et al., (2019) BMC Medical Genomics, **12**:132

ICA with 8 components: M (weights) over experiments

Weights



Batch effect

Cell line-specific

Analysis of contributing genes (S): biological processes

1 → SRP-dependent cotranslational protein ta...
cotranslational protein targeting to mem...
protein targeting to ER a lot of MTs

★ 2 → cell cycle & DNA replication
CDK1, TOP2A, CDK2, CCNA2

★ 3 → mitotic cell cycle,
CCNA2, CDK1, TOP2A, CDK2

4 → generation of precursor metabolites and ...
ATP metabolic process
SRP-dependent cotranslational protein ta...

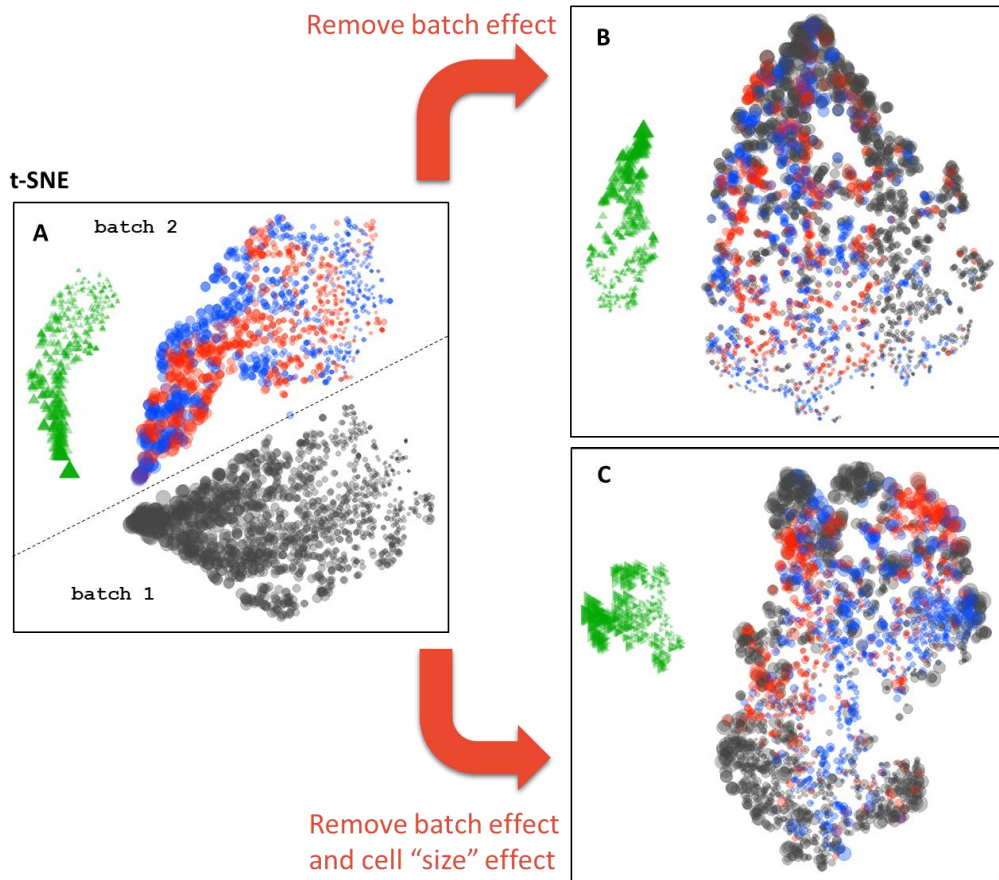
5 → Cell lines, PDGFRA

6 → RNA processing
ribonucleoprotein complex biogenesis
ribosome biogenesis

7 → **Batch effect +**
organonitrogen compound metabolic proces...
mitochondrial translational elongation
ribonucleoprotein complex biogenesis

8 → Technical – library size

Correction of batch effect



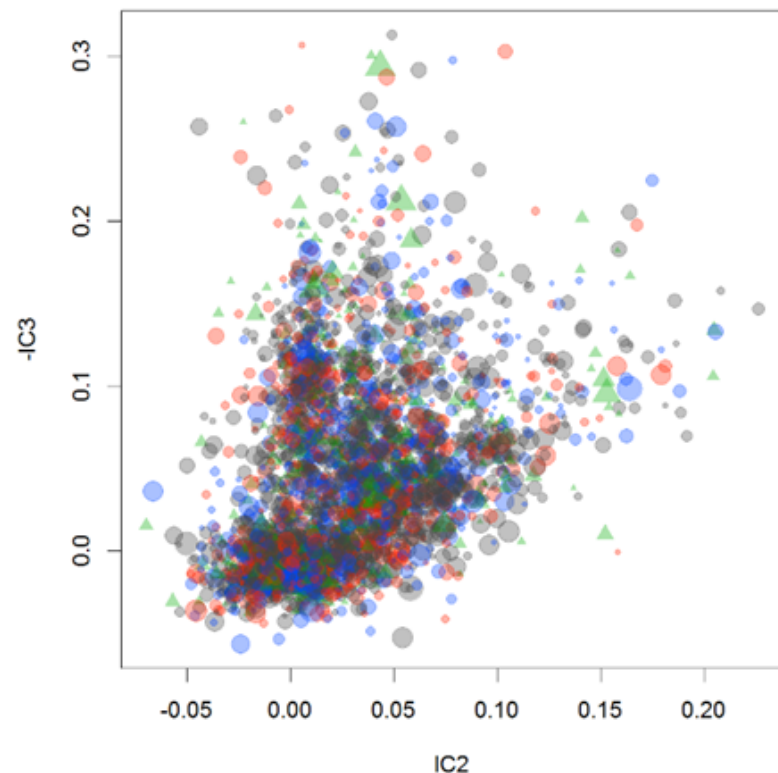
$$E_{nm} \Rightarrow S_{nk} \times M_{km}$$
$$M'_{7,m} \leftarrow 0$$
$$E'_{nm} \leftarrow S_{nk} \times M'_{km}$$

t-SNE representations of original data (A) and ICA-recovered data, after excluding batch effect (B) or several (C) components linked to technical factors.

Dirkse et al. (2019) Nature Communications, 10:1787

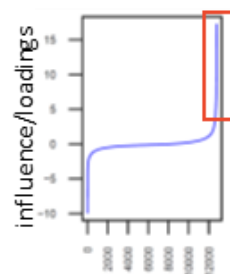
Gene expression: CCNA2

ICA (*M*-matrix)



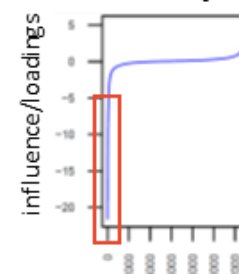
ICA (*S*-matrix)

GO of IC2 (positive side)



GO	FDR
cell cycle process	1.45e-28
DNA replication	1.45e-28
chromosome organization	1.45e-28
DNA metabolic process	1.45e-28
DNA-dependent DNA replication	1.45e-28
DNA conformation change	1.45e-28
DNA repair	2.35e-27
DNA recombination	7.95e-27
chromosome segregation	1.10e-19
DNA packaging	1.76e-19

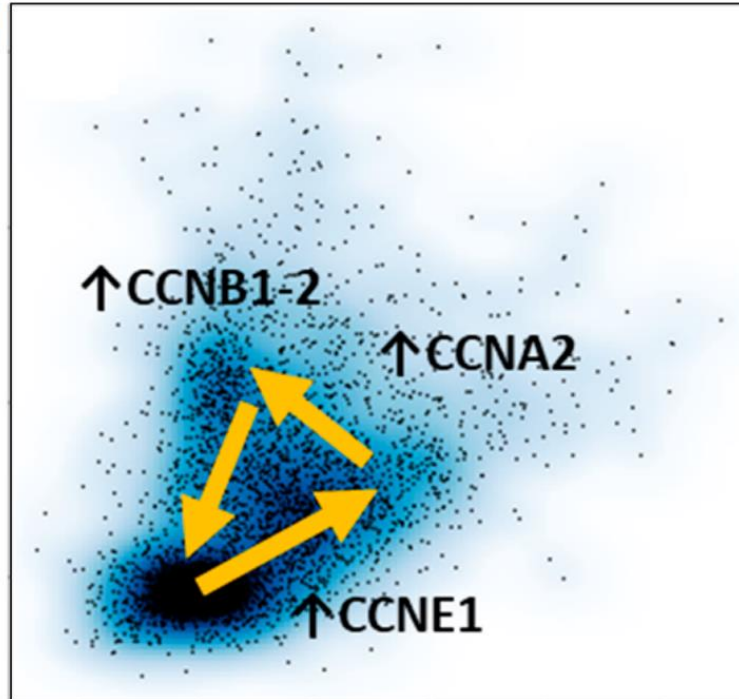
GO of IC3 (negative side)



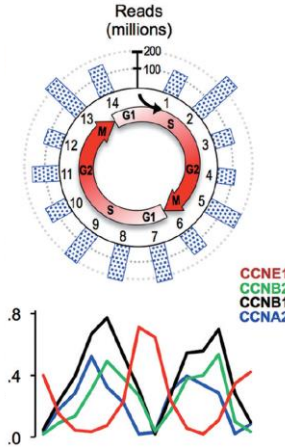
GO	FDR
mitotic cell cycle process	7.68e-29
mitotic cell cycle	7.68e-29
cell cycle process	7.68e-29
cell cycle	7.68e-29
sister chromatid segregation	7.68e-29
microtubule-based process	7.68e-29
sister chromatid cohesion	7.68e-29
regulation of cell cycle process	8.71e-28
single-organism organelle organization	8.25e-26
microtubule cytoskeleton organization	3.00e-25

Same cell cycle in depth

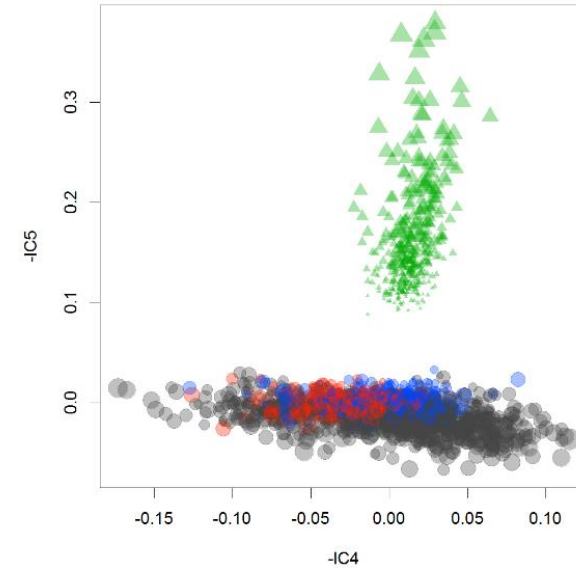
IC: DNA replication



IC: mitotic cell cycle



Dominiguez (2016) Cell Research



- ❑ Consensus ICA can be used to **correct technical biases**.
- ❑ ICA provides information about **biological processes** in scRNA-seq.
- ❑ Weights of the components can be used to **predict cell classes**.

Quantitative Biology Unit, Luxembourg Institute of Health (LIH)

Arnaud MULLER

Dr. Francisco AZUAJE



NORLUX Neuro-Oncology, LIH

Dr. Anna GOLEBIEWSKA

Prof. Simone NICLOU



Institut Curie, France

Dr. Andrei ZINOVYEV



Supported by Luxembourg National Research Fund
C17/BM/11664971/**DEMICS**