LUXEMBOURG INSTITUTE OF HEALTH
RESEARCH DEDICATED TO LIFE

Fonds National de la Recherche Luxembourg

consICA: a package for parallel consensus ICA of multi-omics data and interpretation of the results

Petr V. Nazarov, Tony Kaoma, Maryna Chepeleva, Sang-Yoon Kim, Arnaud Muller, Gunnar Dittmar

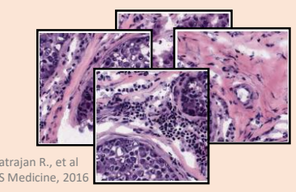
Multiomics Data Science Research Group, Quantitative Biology Unit, Luxembourg Institute of Health, Luxembourg
petr.nazarov@lih.lu

Challenges

Experimental data (RNA-seq)

- Technical heterogeneity
- Sample composition
- Clonal evolution in cancer


http://abrf.masonlab.net/


Natrajan R., et al PLoS Medicine, 2016

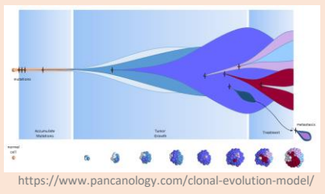
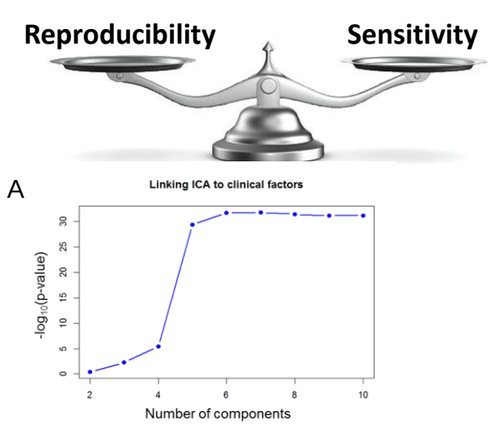

https://www.pancanology.com/clonal-evolution-model/

FIGURE 1. Sources of biases that limit comparison between transcriptomics data of cancer patients.



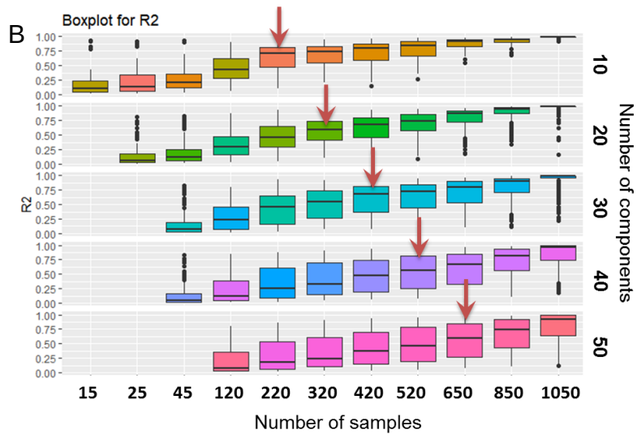


FIGURE 3. (A) Detection of a confounding factor (gender) by ICA. (B) The stability of ICA increases with growth of the dataset and decreases with increase of the number of components.

Method

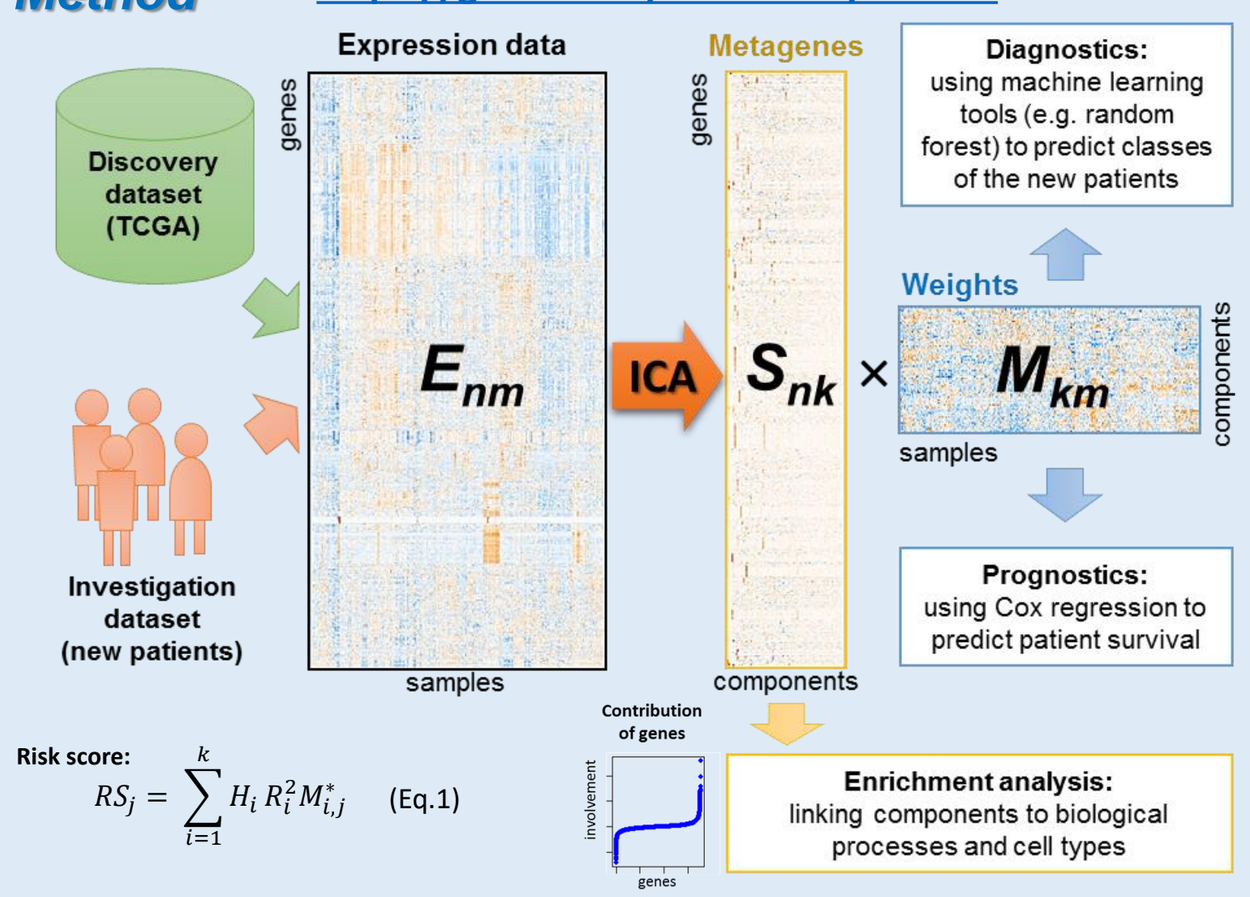
https://gitlab.com/biomodlih/consica

SKCM (melanoma)
Discovery: 472
Validation: 44
Investigation: 5

GBM/LGG (gliomas)
Discovery: 701
Validation: 325
Investigation: 58

LUAD/LUSC (lung cancers)
Discovery: 1128
Validation: 490
Investigation: 36

PAAD (pancreas cancers)
Discovery: 457
Validation: 96
Investigation: 183



Risk score:
$$RS_j = \sum_{i=1}^{k} H_i \, R_i^2 M_{i,j}^* \quad \text{(Eq.1)}$$

FIGURE 2. ICA decomposes combined gene expression matrix from multiple bulk samples into meaningful signals in the space of genes (metagenes, **S**) and weights in the space of samples (weight matrix, **M**). Biological processes and signatures of cell subtypes can be found in **S**, while **M** could be linked to patient groups and patient survival (Cox regression and Eq.1).

Platform bias / biological processes

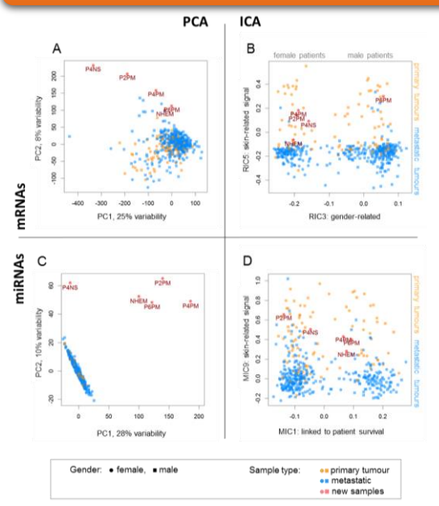


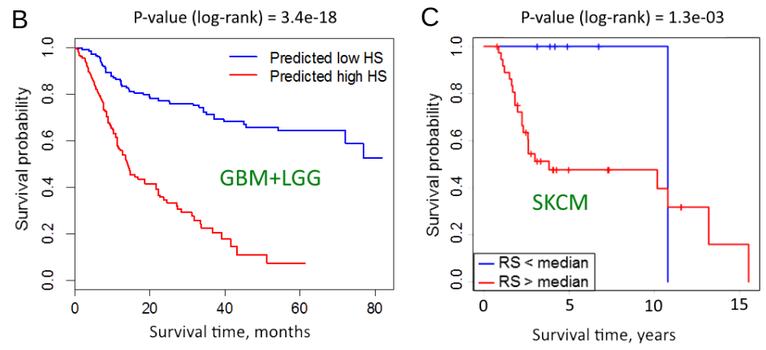
FIGURE 4. ICA-based batch correction in melanoma mRNA (A,B) and miRNA (C,D) datasets. PCA captures mainly technical factors, while ICA correctly projects new samples into the space of biologically meaningful components.

Diagnostics & prognostics

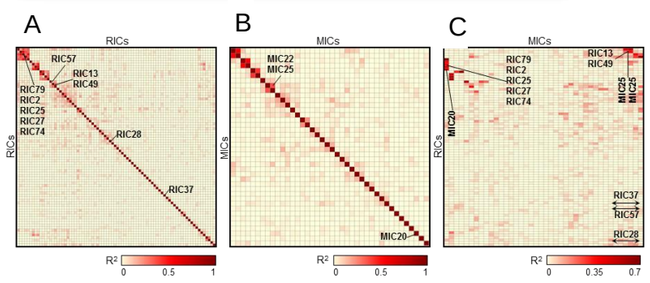
A

Actual → Predicted ↓	immune	keratine	MITF-low
immune	160	9	6
keratine	9	91	6
MITF-low	1	2	47

FIGURE 5. Example of confusion matrix for melanoma patient classification for melanoma (A). Predicted risk correlates with actual survival in independent cohorts of glioma (B) and melanoma (C) patients.



Multi-omics data integration



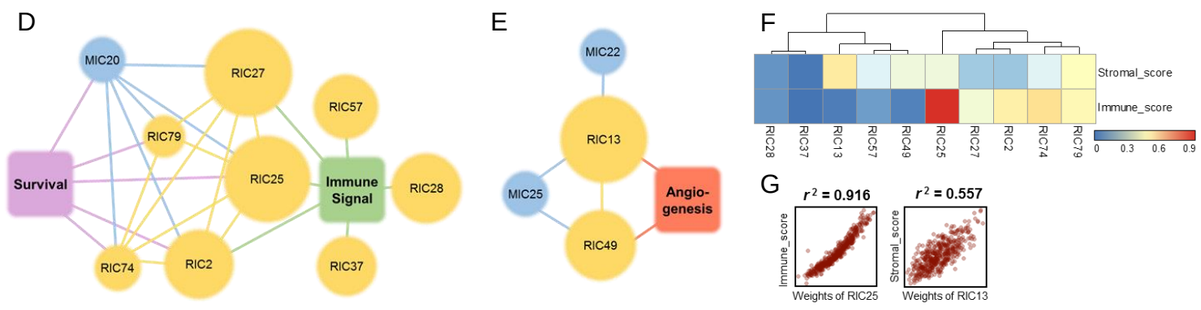


FIGURE 6. Heatmaps of r^2 between weights (**M**-matrix) of RIC-RIC (A), MIC-MIC (B) and RIC-MIC (C). Clusters of components linked to immune response (D) and to angiogenesis and stroma (E). Components were compared with ESTIMATE scores and corresponding r^2 are shown in (F). RIC25 and RIC13 components correlated best to immune and stromal scores (G).

Conclusions

- ❑ Consensus ICA **corrects technical biases** and improves comparability of the new data to the reference dataset.
- ❑ ICA provides information about **biological processes** in the new samples and allows scoring them.
- ❑ Weights of the ICA-derived components can be used to **predict patient survival and classes**, as was demonstrated on an independent datasets.
- ❑ Finally, ICA can be used for **data integration**, when several levels of omics data are available.

Acknowledgements

This work was supported by the Luxembourg National Research Fund C17/BM/11664971 "DEMICS".
We would like to thank all the people participated in the DEMICS project or discussions of the results: Laurène Picandet, Thomas Eveno, Aliaksandra Kakoichankava, Dr. Andrea Bauer, Dr. Stephanie Kreis, Dr. Anna Golebiewska, Dr. Andrei Zinovyev, Prof. Simone Niclou, Prof. Jörg Hoheisel, Prof. Inge Jonassen.

References

1. Sompairac et al., **(2019)** Int. J. Mol. Sci, **20**(18): 4414
2. Nazarov et al., **(2019)** BMC Medical Genomics
3. https://gitlab.com/biomodlih/consica

1A-B rue Thomas Edison
L-1445 Strassen
Luxembourg

T +352 26970-385
F +352 26970-390
www.lih.lu