Analysis of cell deconvolution methods: A comparison of reference-based and reference-free cell deconvolution

Author: Stanisław Howard
S.M.Howard@student.tudelft.nl

Supervisors: Stavros Makrodimitris, Bram Pronk, Daan Hazelaar
Responsible Professor: Marcel Reinders

1. Introduction

- **Fragmentomics**, a field in bioinformatics, can offer a non-invasive solution to cancer detection by analysing circulating DNA fragments in blood samples.
- We focused on comparing two methods of **cell deconvolution.**
- Technique used for estimating cell type proportions in blood samples based on circulating DNA fragments. It can be used for anomaly detection.
- Cell deconvolution is divided into reference-based and reference-free methods.. We chose **UXM**[1] and **cfSort**[2].

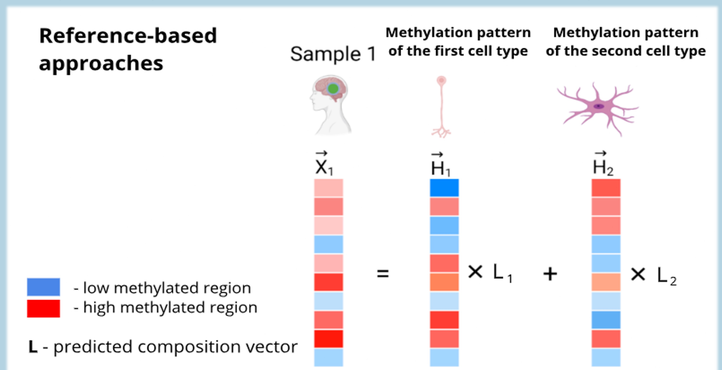

Figure 1. general idea of reference-based deconvolution

- Decompose a mixture of DNA fragments using a reference
- Reference is information about methylation in DNA region
- Proportions of cell types in the sample are estimated

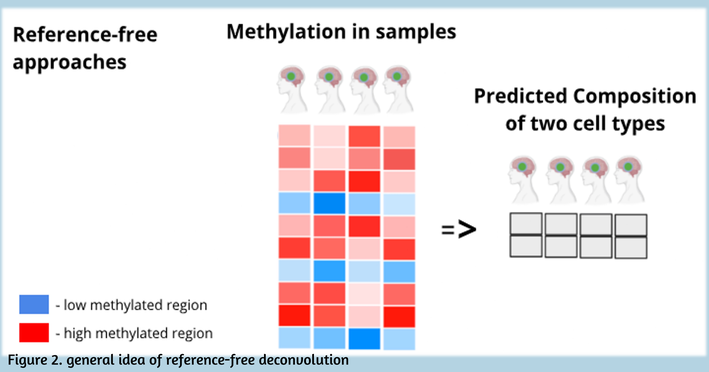

Figure 2. general idea of reference-free deconvolution

- Analyse the gene expression profile of a sample directly
- Using e.g. Machine Learning approach to derive the compositions based on methylation

2. Research questions

- Which method is better in terms of performance using Pearson's correlation coefficient?
- How do the two models compare in a sensitivity detection test in a case with low percentages of a secondary cell type in a mixture with white blood cells?

3. Methodology

- We devised an experiment with 6 cell types: breast, prostate, colon, lung, kidney and white blood cells.
- We then calculated the metrics on the synthesized mixtures

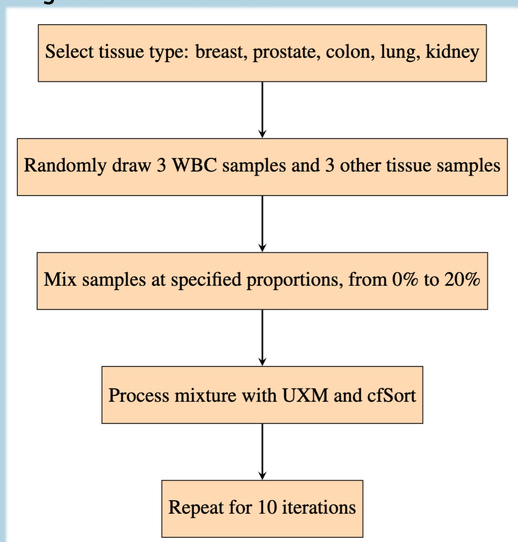

Figure 3. Sample mixing experiment flow

7. Future Work

Investigate Coverage of Files: Conduct an analysis of the coverage of files used in each sample to determine how low coverage correlates with poorer Pearson's correlation coefficients.

4. Results

Pearson Correlation Experiment:

- UXM achieved higher average Pearson correlation coefficients across all tissue types compared to cfSort.
- UXM demonstrated lower standard deviation in correlation coefficients.
- Plots revealed two outliers in the colon tissue for both UXM and cfSort, suggesting a need for further investigation.

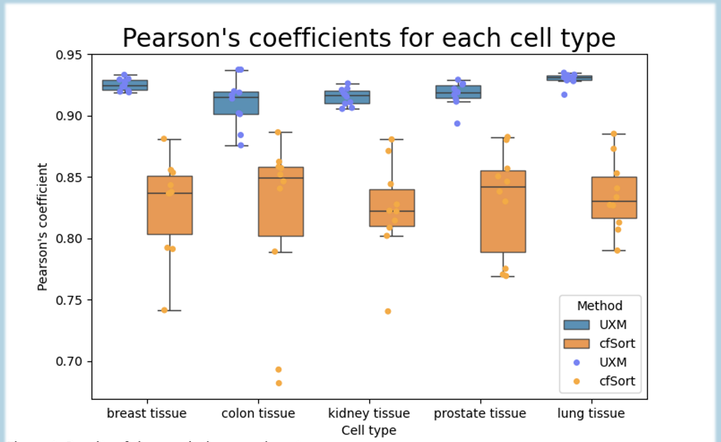

Figure 4. Results of the correlation experiment

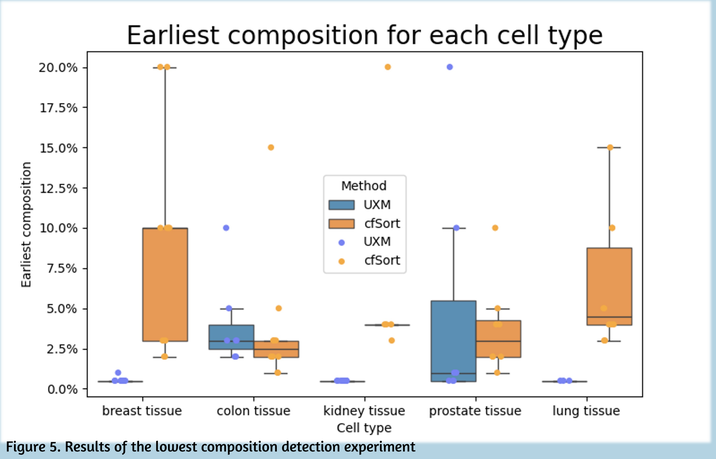

Figure 5. Results of the lowest composition detection experiment

Early Detection Experiment:

- UXM showed lower mean values for earliest detection across most tissue types.
- Breast tissue had the highest detection mean for cfSort, while colon tissue had the lowest. Variability across types.
- The overall average for UXM was 2.07%, while cfSort averaged 5.95%.
- UXM superior sensitivity in early detection.

5. Conclusions

- UXM provides a more transparent and consistent method for cell deconvolution in cancer detection compared to cfSort, making it more suitable for clinical applications.
- Enhancing the data diversity and improving the transparency of cfSort is essential for advancing the accuracy and reliability of cell deconvolution.

6. Limitations

- **Limited Tissue Types**: Only 6 tissue types may provide a limited view of the performance of the deconvolution methods.
- **Suboptimal Architecture**: cfSort was tested using Linear Regression instead of Deep Neural Network, possibly affecting its performance.

[1] - N. Loyfer et al., "A DNA methylation atlas of normal human cell types," *Nature*, vol. 613, no. 7943, pp. 355-364, Jan. 2023, doi: 10.1038/s41586-022-05580-6.
[2] - S. Li et al., "Comprehensive tissue deconvolution of cell-free DNA by deep learning for disease diagnosis and monitoring," *Proceedings of the National Academy of Sciences*, vol. 120, no. 28, p. e2305236120, Jul. 2023, doi: 10.1073/pnas.2305236120.