Date: 01 March 2022
SSF Project Report

*Performance and Benchmarking*

# Comparisons of two tools for identifying mutational signatures in cancer

Katarina Truvé[1,*]

[1]Bioinformatics Core Facility, Sahlgrenska Academy, University of Gothenburg, Box 115, 405 30, Sweden.

## Summary

This project compares two tools: DeconstructSigs and SigPprofiler, to identify most suitable tool in the identification of mutational signatures in cancer.

**Contact:** katarina.truve@gu.se

## 1 Background

Somatic mutations can occur during DNA replication and repair. It can be caused by contacts with mutagens e.g. UV-rays, X-rays or tobacco smoke. Different mutational processes are active during the development of different types of cancer. These mutational processes give rise to an unique combination of mutation types that has been termed "Mutational Signatures". Some mutational signatures have been shown to be associated with clinical outcome, which offers the potential to use them as biomarkers for novel therapies (Maura, 2019).

COSMIC (Catalogue of Somatic Mutations in Cancer) has in collaboration with Wellcome Sanger Institute, Cambridge, UK and Alexandrov lab at the University of California revealed many mutational signatures in many human cancer types, analyzing more than 23.000 cancer patients.

Considering single nucleotide variants there are six possible substitutions: C>A, C>G, C>T, T>A, T>C, and T>G. Considering also the bases immediately 5' and 3' there are 96 different contexts. At the COSMIC website (https://cancer.sanger.ac.uk/signatures/) by March 2021 there are 60 different reference signatures based on single base substitutions (SBS) that has been extracted using the tool Sigprofiler2. For each signature there is a description what is known about the aetiology. In the website there are also information about doublet base substitutions (DBS) and small insertions and deletions Signatures (ID).

## 2 Methodology

In this study we tried two different software tools for analysis of mutational signatures: DeconstructSigs (Rosenthal, 2016) and Sigprofiler (Alexandrov, 2020). The tools were run on a Linux server (7.5 x86_64) on sequencing data from patients with colon cancer. Both programs were run with default settings. DeconststructSigs uses a multiple linear regression model to determine the linear combination of predefined signatures while Sigprofiler uses a method based on nonnegative matrix factorization (NMF). Before identifying signatures, somatic mutations were extracted and quality filtered using Mutect2 (Cybulskis, 2013).

## 3 Results

We first ran DeconstructSigs. For each patient the different kind of mutations in the genome can be counted and illustrated as in Figure 1, with the fraction (y-axis) of the 96 different mutation types (x-axis). For an individual tumor sample DeconstructsSigs aims to determine the contribution of known mutational processes. The results can be visualized with a circle diagram as exemplified in Figure 2. The output is easy to understand and the program is fast to run. On the downside we ran into problems when trying to run with a newer version of the human genome. The program seems to run best using hg19. The default input signature set has not been updated to the latest version of Cosmic signatures, but this could probably be solved by a user defined parameter.

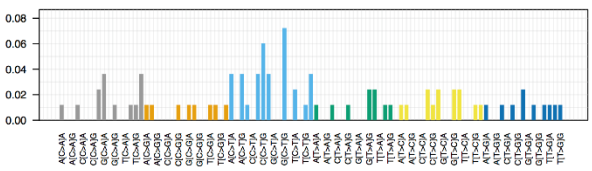


**Figure 1**. Representation of different types of somatic mutations found in one cancer patient using DeconstructSigs.

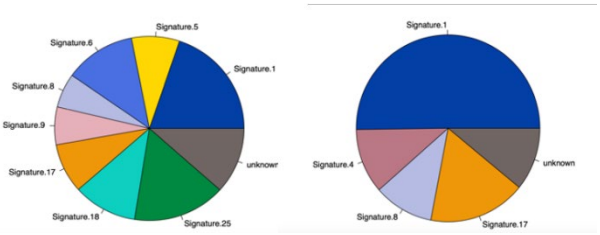


**Figure 2**. Output from DeconstructSigs, showing contribution of known mutational signatures for two individual tumor samples.

Next, we ran SigProfiler, the same tool used to extract the signatures at the COSMIC website. The tool takes longer time to run, but the tool works well on latest version of the human genome (hg38) and known signatures are up to date. Sigprofiler can be run with a batch of samples which unlike Deconstructsigs give different results if samples are instead run individually. Sigprofiler produce a lot of different output files. The ones that would be comparable to the Deconstructsigs pie plots are called COSMIC_SBS96_Activity_Plots_refit.pdf. The output is illustrated in Figure 3. As can be seen no SBS was found for the first individual using Sigprofiler only one ID compared to DeconstructSigs that proposed 9 different signatures. The second individual has two signatures that overlap between the two tools, signature 1 and 17. When running the two samples together Sigprofiler seem to pick up signature 1 and 5 also in the second sample. When running with more samples signature 1 and 17 disappear and are replaced with signature 40.

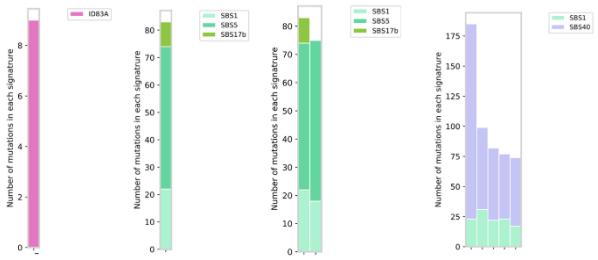


**Figure 3**. Output from Sigprofiler fitting plots. The two first pictures from left correspond to the individuals shown in figure 2, while the third picture from left shows result when the two samples were run together and finally when these two were run together with three more samples.

## 4 Conclusion

We show just a few samples and it is hard to draw any strong conclusions. It has been proposed before that DeconstructSigs might give false positives (Maura, 2019). We can see here that DeconstructSigs fits more Cosmic signatures than Sigprofiler to the individually run samples. On the other hand, it is hard to understand how the algorithm for Sigprofiler works giving very different results for individual samples depending on the context the samples are run. It could be proposed that signatures that overlap between the tools would be more likely to be correct, but more analysis are needed to understand how to best extract accurate information and how to interpret the results.

## Abbreviations

**COSMIC** Catalogue of Somatic Mutations in Cancer
**SBS** single base substitutions
**DBS** doublet base substitutions
**ID** small insertions and deletions Signatures
**NMS** nonnegative matrix factorization

## Availability of data and materials

No datasets generated during the current study. However, use case data may be available from the authors upon reasonable request and with permission of the Anna Rohlin, at the University of Gothenburg.

## Acknowledgements

The author thanks the Bioinformatics Core Facility at the University of Gothenburg for providing computational resources for data analysis and storage. Special thanks to Anna Rohlin (University of Gothenburg) for providing sample.

## Funding

This study was supported by the Swedish Foundation for Strategic Research (RIF14–0081). The funding body did not play any role in the design of the study, or collection, analysis, or interpretation of data, or in writing the manuscript.

*Conflict of Interest:* none declared.

## References

Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101. doi:10.1038/s41586-020-1943-3

Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnology (2013).doi:10.1038/nbt.2514

Maura, F., Degasperi, A., Nadeu, F. et al. A practical guide for mutational signature analysis in hematological malignancies. Nat Commun 10, 2969 (2019). https://doi.org/10.1038/s41467-019-11037-8

Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016 Feb 22;17:31. doi: 10.1186/s13059-016-0893-4. PMID: 26899170; PMCID: PMC4762164.