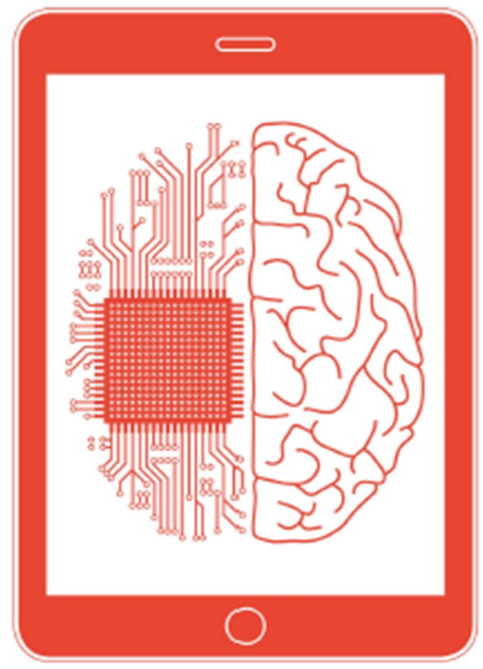
University of Applied Sciences and Arts Northwestern Switzerland
School of Life Sciences

Development of a Bioinformatics Pipeline for the Detection and Visualization of Mutational Signatures in Next-Generation-Sequencing Data

Hodel Selina
Master-Thesis

Supervisor: Prof. Dr. Abdullah Kahraman, FHNW University of Applied Sciences and Arts Northwestern Switzerland
Expert: Prof. Dr. Kjong Lehmann, RWTH Aachen
External partner, Dr. Martin Zoche, USZ University Hospital Zurich



Introduction

Cancer is predominantly caused by genetic mutations resulting from exogenous or endogenous processes. Mutational signatures are characteristic patterns of somatic mutations that can indicate the underlying process and thus the etiology of the tumor, such as ultraviolet light, tobacco smoking or defective DNA repair processes and many others. Single base substitutions (SBS) are the most studied mutational signatures. There are 96 SBS categories resulting from the six substitution subtypes: C>A, C>G, C>T, T>A, T>C, and T>G including the immediately neighboring nucleotides (Alexandrov et al., 2020). Detection of mutational signatures in targeted sequencing is much more challenging than in whole genome sequencing data due many times fewer variants present (Lee et al., 2023). Reference signatures can be used to map mutational signatures of tumors.

Objectives

- Develop pipline for the detection of mutational signatures in FMI panel
- Evaluate feasibility and validate mutational signature mapping in FMI panel

Method / Developed Pipeline

Somatic variant detection

- Variant calling of short nucleotide variants is performed with Mutect2. The variants are filtered using a self-assembled panel of normals and population databases GnomAD and dbSNP to remove common SNPs. Variants are filtered for VAF ≥ 0.05, read depth ≥ 200 and others.

Mapping of mutational signatures

- The remaining variants are put into the trinucleotide context and the 96 SBS mutation type matrix is generated and a corresponding plot created.

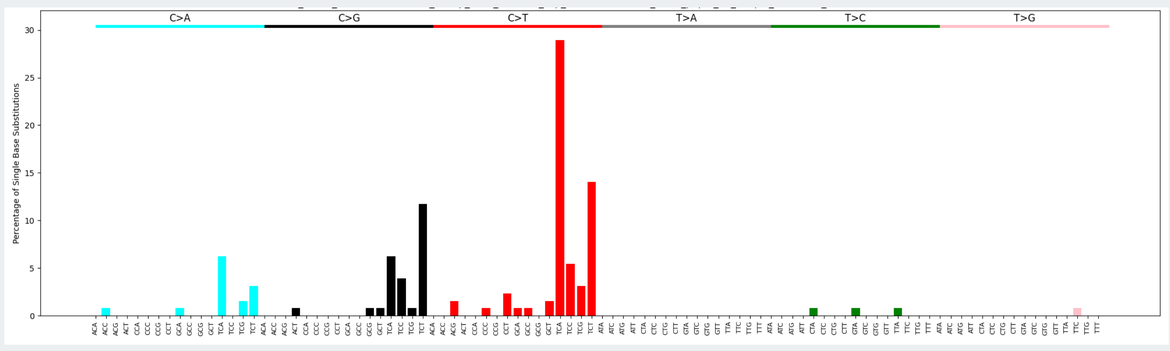


Fig. 1: 96 SBS mutation type matrix of a breast cancer sample

- The mutation type matrix is used to perform signature mapping with SigProfilerAssignment and the reference signatures from COSMIC v3.2 (Fig 2). Activity of signatures of the sample (red) can be compared to cases of the same primary, e.g. breast (blue) (Fig 3).

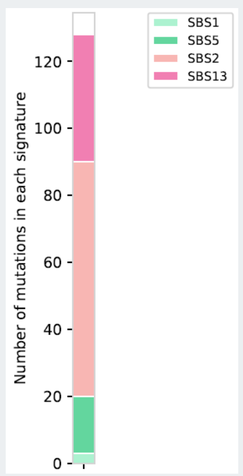


Fig. 2: Signature activity in sample

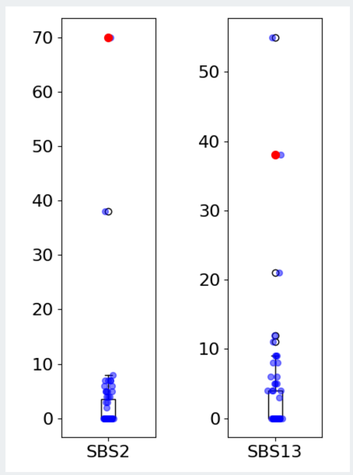


Fig. 3: Signature activity compared to other breast cancer samples

- Samples with close cosine similarity of the mutation type matrix are identified.

Sample Name	Cosine similarity	Disease	TMB [Muts/Mb]	SBS active
Sample	-	Breast invasive lobular carcinoma (ILC)	27	SBS2, SBS13
Hit1	0.941	Anus squamous cell carcinoma	8	SBS2, SBS13
Hit2	0.919	Head and neck squamous cell carcinoma (HNSCC)	22	SBS2, SBS13
Hit3	0.885	Bladder urothelial (transitional cell) carcinoma	18	SBS2, SBS13

Validation / Results

- Signature mapping performance was better using SigProfilerAssignment than SATS.
- Using larger reference signature sets led to higher precision.
- Precision was ≥ 80 % for signatures SBS2, SBS4, SBS7a, SBS7b, SBS13, SBS17a and SBS17b in PCAWG data (J. Zhang et al., 2019) if at least 20 variants were present in the panel target region.

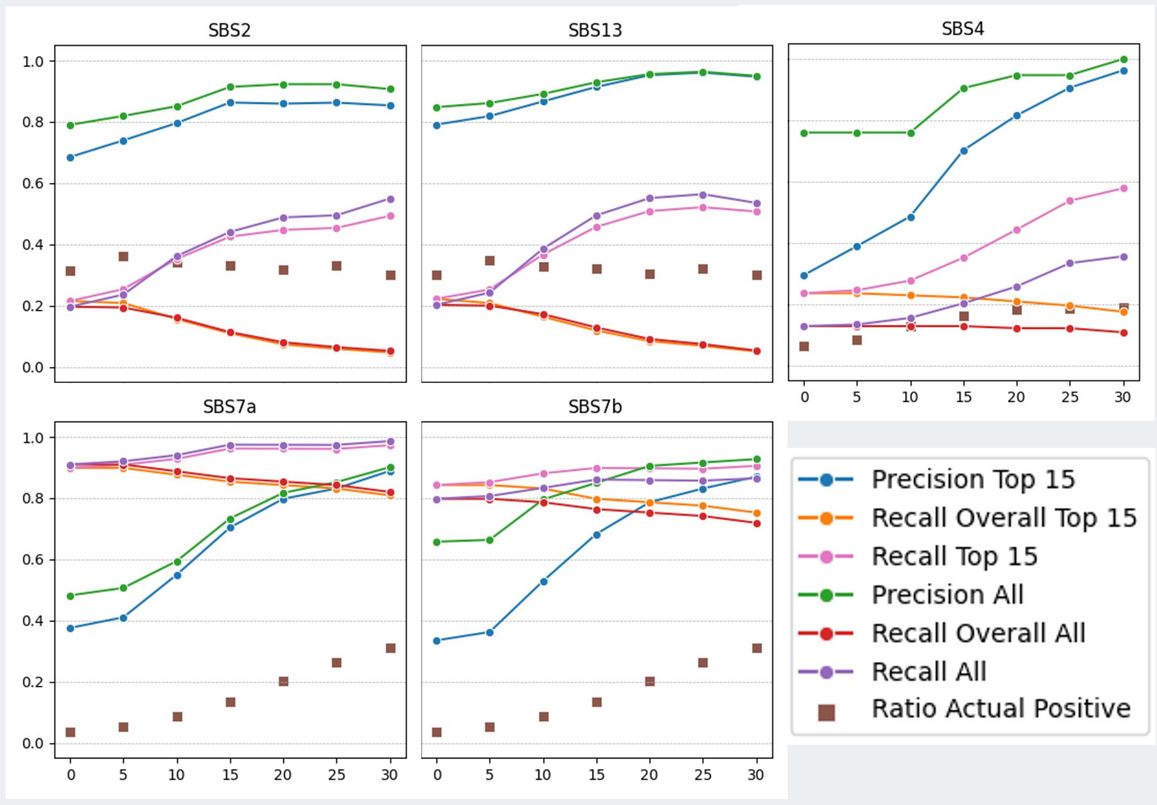


Fig. 4: SigprofilerAssignment performance in 5 of the 7 best performing signatures in PCAWG validation. X-axis of the figures shows the set threshold for number of variants in total in the target region. Y-axis shows fraction of precision or recall. For the all COSMIC v3.2 signatures set and the top15 set the precision and recall are indicated, as well as the overall recall. The brown rectangles show the ratio of actual positive cases of a signature divided by the total number of cases that have at least the indicated number of variants.

- Signature mapping in FMI panel data was validated using corresponding WGS data as ground truth. Detection of signatures mentioned above works for samples starting from a TMB of approximately 5 Muts/Mb.

Conclusion

- Single base substitution (SBS) signatures SBS7a and SBS7b (UV light), SBS2 and SBS13 (APOBEC), SBS4 (tobacco smoking) and SBS17a and SBS17b (unknown etiology) can be detected in the FoundationOne®CDx panel.
- Although this work showed promising results in the detection of mutational signatures in panel data, it is important to recognize that mutational signatures represent only part of the broader picture of tumor characterization and should always be seen in the context of other relevant information.

References
Alexandrov, L. B et al (2020). The repertoire of mutational signatures in human cancer. Nature, 578 (7793), 94–101. https://doi.org/10.1038/s41586-020-1943-3
Zhang et al.(2019). The International Cancer Genome Consortium Data Portal, Nature Biotechnology, 37 (4), 367–369. https://doi.org/10.1038/s41587-019-0055-9
Lee et al. (2023). Pan-cancer mutational signature analysis of 111,711 targeted sequenced tumors using SATS. https://doi.org/10.1101/2023.05.18.23290188