Weekly Report

# 14/03/2022

BOUDEMIA ALA EDDINE

TEAM CHEN

# **Goal:**

* Extraction of mutational signatures in breast/bladder cancers using SigProfiler.
* Extract mutations by regions type
* Extract mutational signatures based on their regions

# **Procedure:**

* As previously, the mutation signatures are extracted using SigProfiler using the previously downloaded data from TCGA.
* Maftools was used to analyze the MAF files.
* I tried maftools to extract the signatures (very fast but bad results)
* Extract the TSS and UTRs using gtftools.
* Extract the TTS from UCSC.
* Map the mutations from the maf files to their respective regions.
* Run sigprofiler on the obtained files separately.

# **Results:**

## BRCA best signatures:

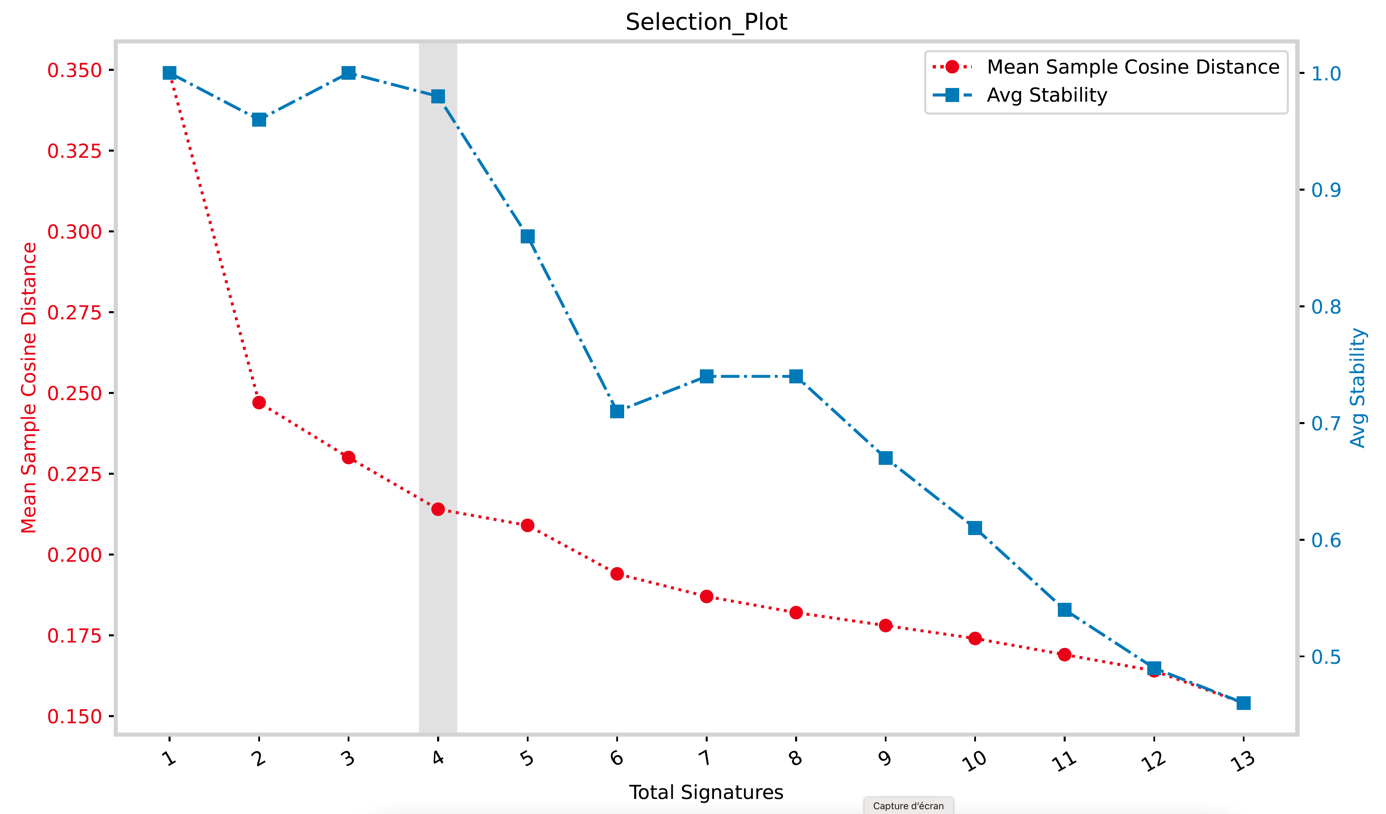


Figure 1 Selection plot of the best number of signatures BRCA

The selection plot from Figure 01 suggests that the best number of signatures to be extracted is 4. Compared to the number of mutational signatures that were reported to be associated with cancer which is 13 sigprofiler failed to deconvolute the signatures from the data at our disposal. Two possible explanations are that WES data do not show some signatures related to other genomic regions and the average mutational load is low for sigprofiler as showed by the authors of the tool. However, when we take a look at the decomposition plot obtained, one can see that most of the significant reference signatures are present, namely: SBS13, SBS2, SBS5, SBS1, and SBS3.

SBS29 as well as SBS6 and SBS10(a&b) are reported in these plots even though they were not reported to contribute to BRCA. SBS6 is associated with defective DNA mismatch repair which could be relevant but SBS29 is related to tobacco, both signatures are not contributing that much to the signatures extracted. However signatures SBS10a and SBS10b which are associated with polymerase exonuclease mutations are present as a distinct signature.

Une image contenant texte, capture d’écran, intérieur

Description générée automatiquement

Figure 2 Decomposition plot for SBSA

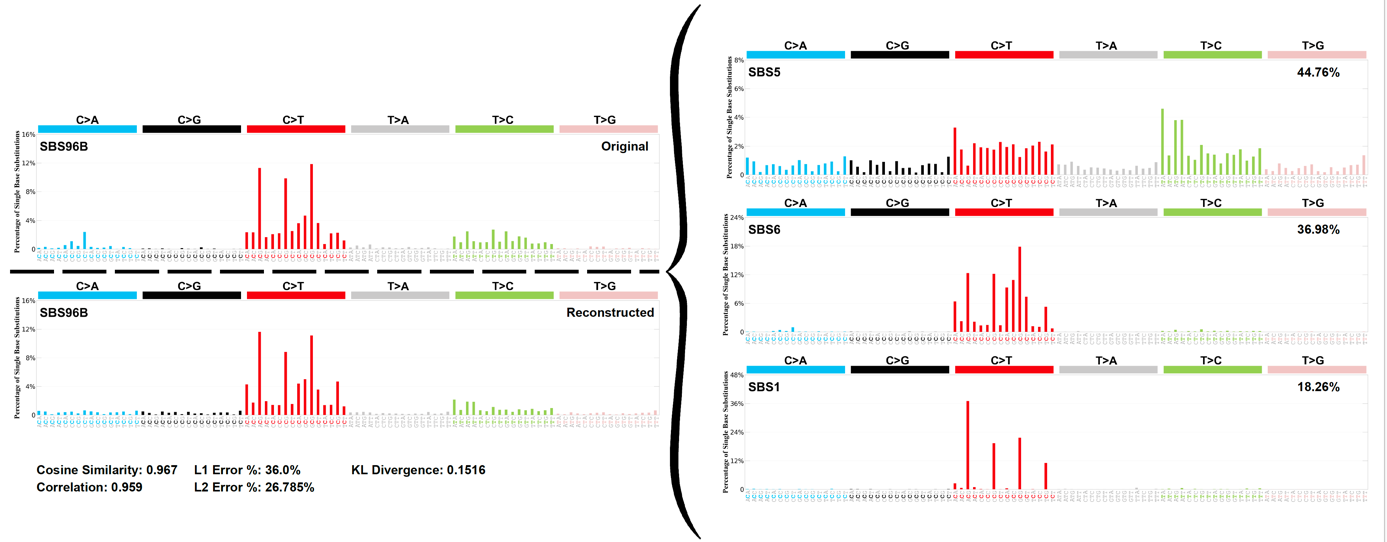


Figure 3 Decomposition plot for SBSB

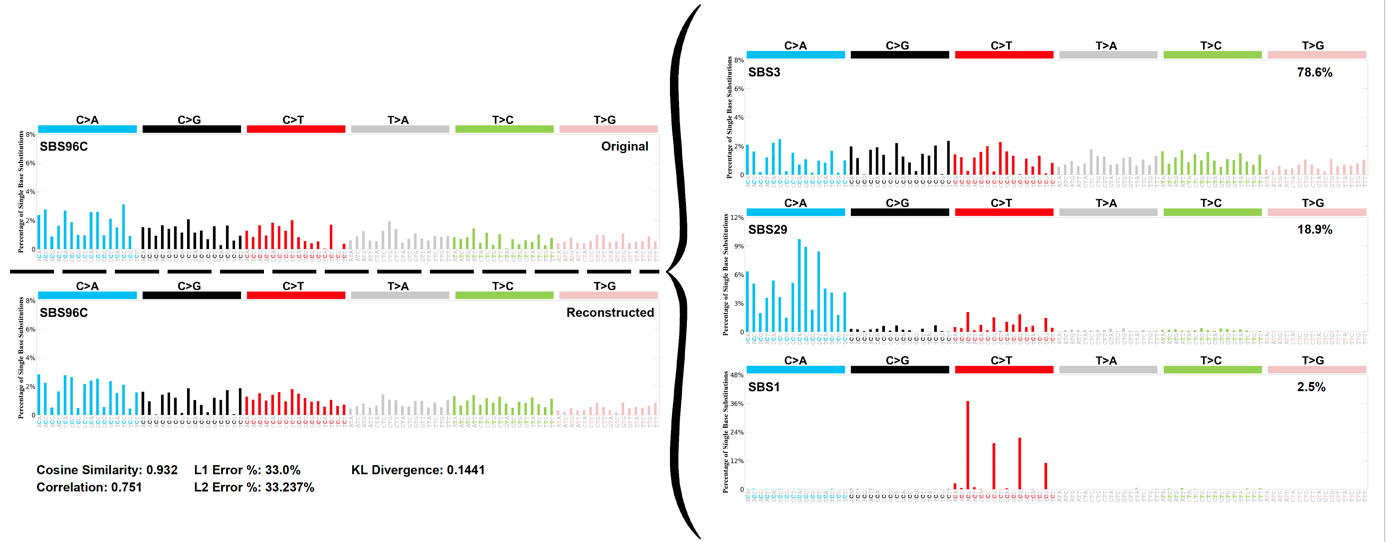


Figure 4 Decomposition plot for SBSC

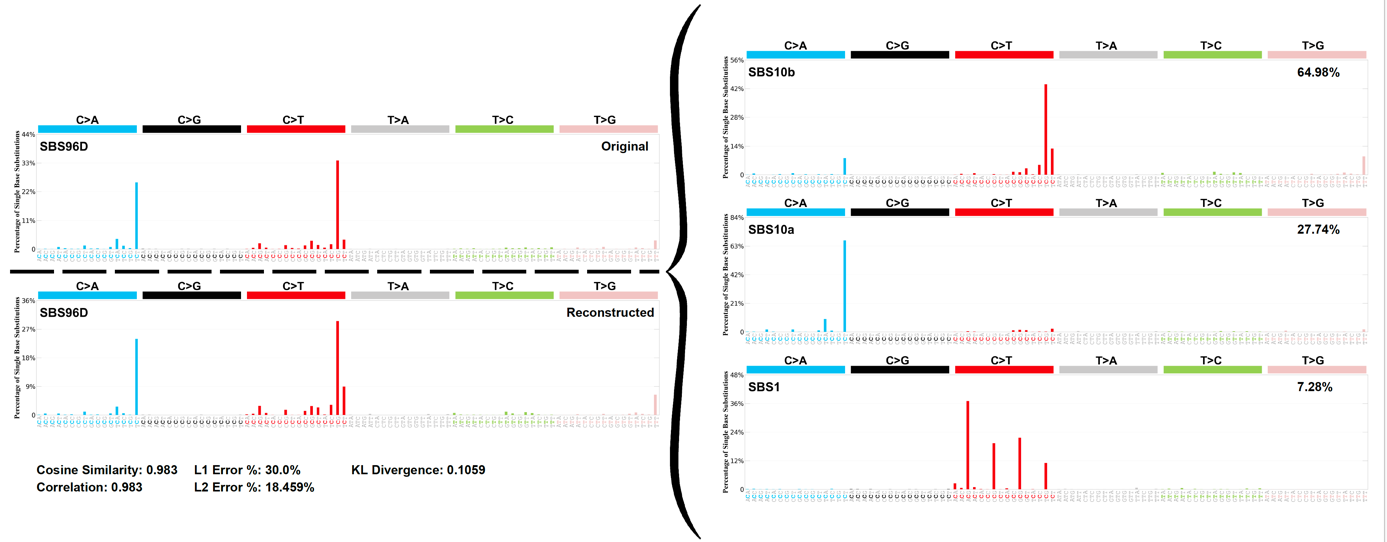


Figure 5 Decomposition plot for SBSD

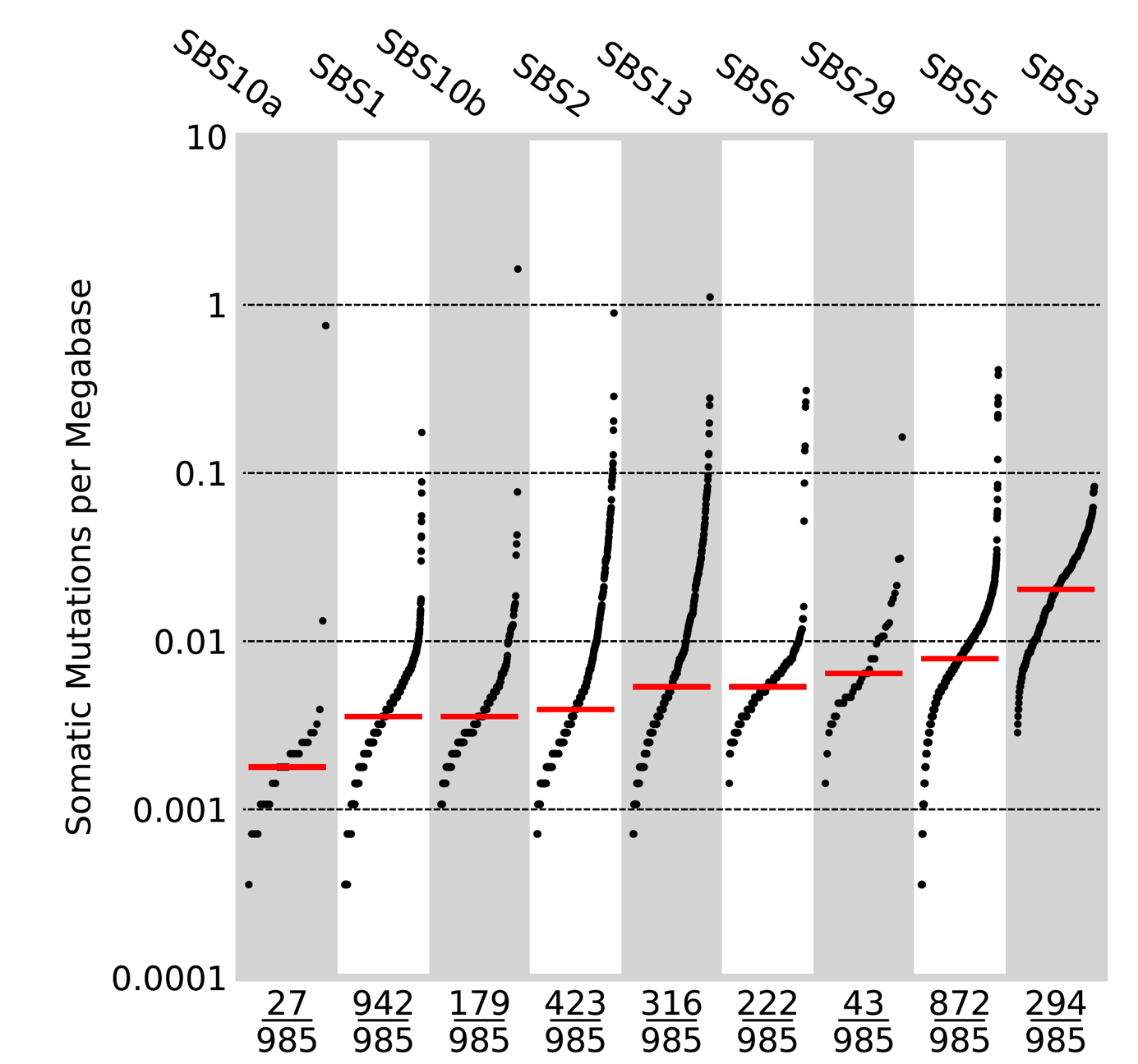
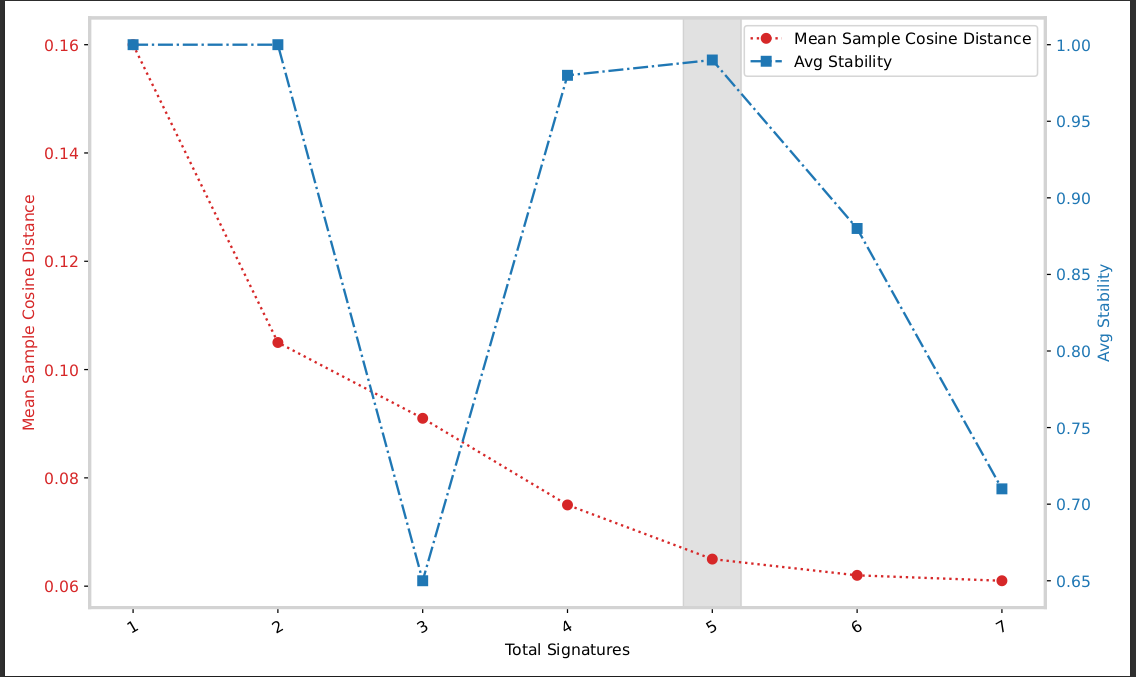


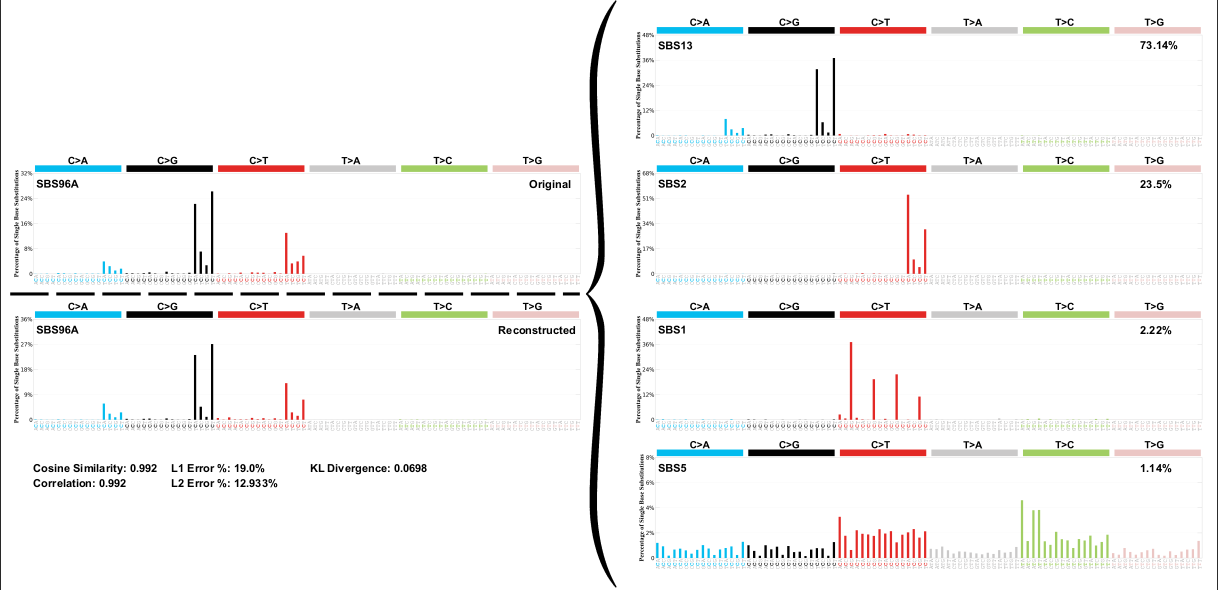
Figure 6 Activity of COSMIC signatures in the BRCA

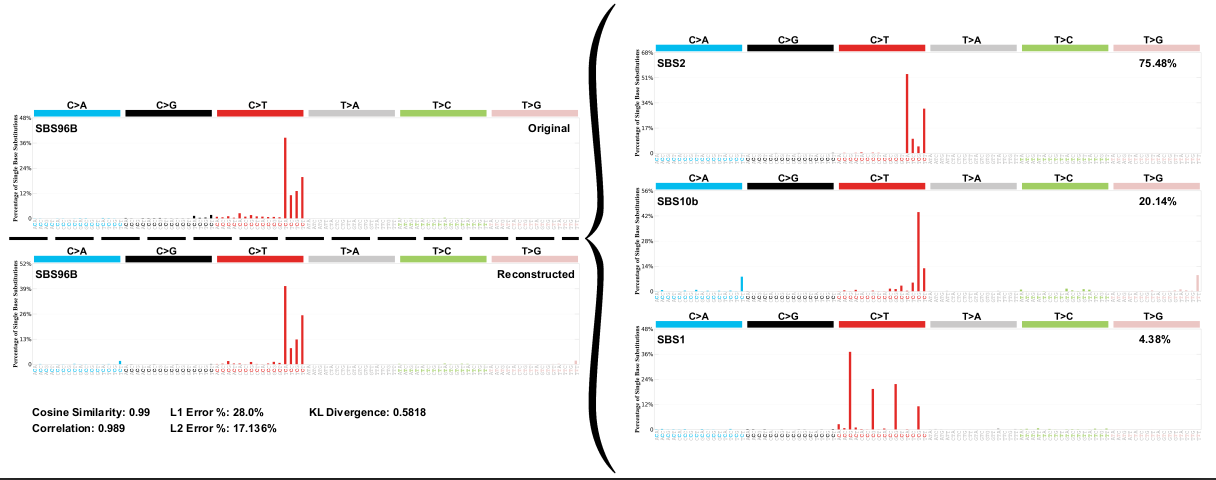
Figure 06 shows that SBS1 is present in all samples but going back to the decomposition plots we observe that it is present in every extracted signatures mainly in moderate percentages, therefore I am not sure how to interpret this. At the same time SBS1 is ubiquitous signature in all cancer types. Similarly, SBS5 is common in all cancer types and is present most samples here.

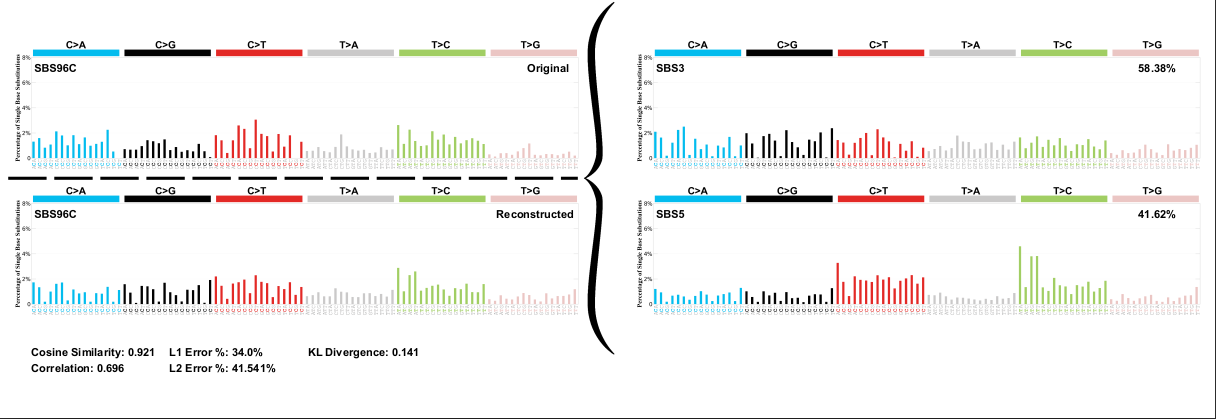
## BLCA best signatures:

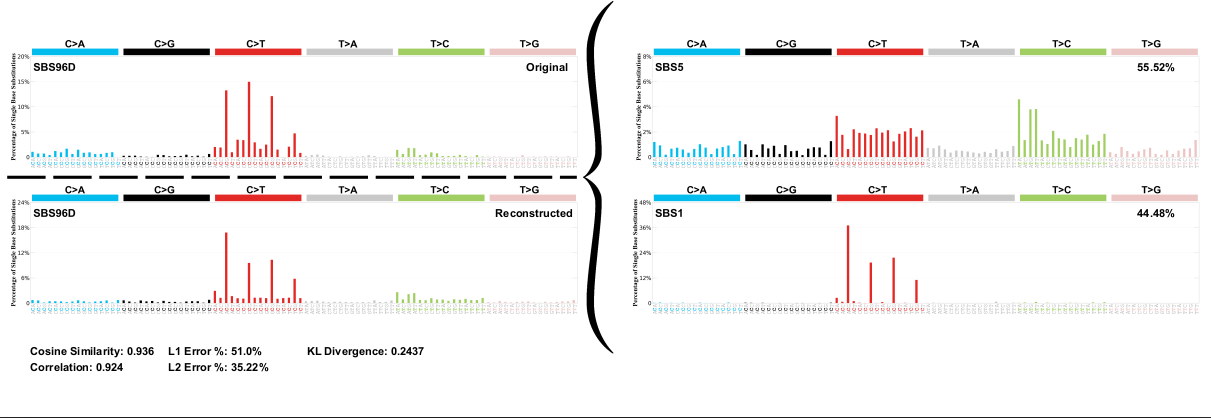
Figure 7: Selection plot for BLCA

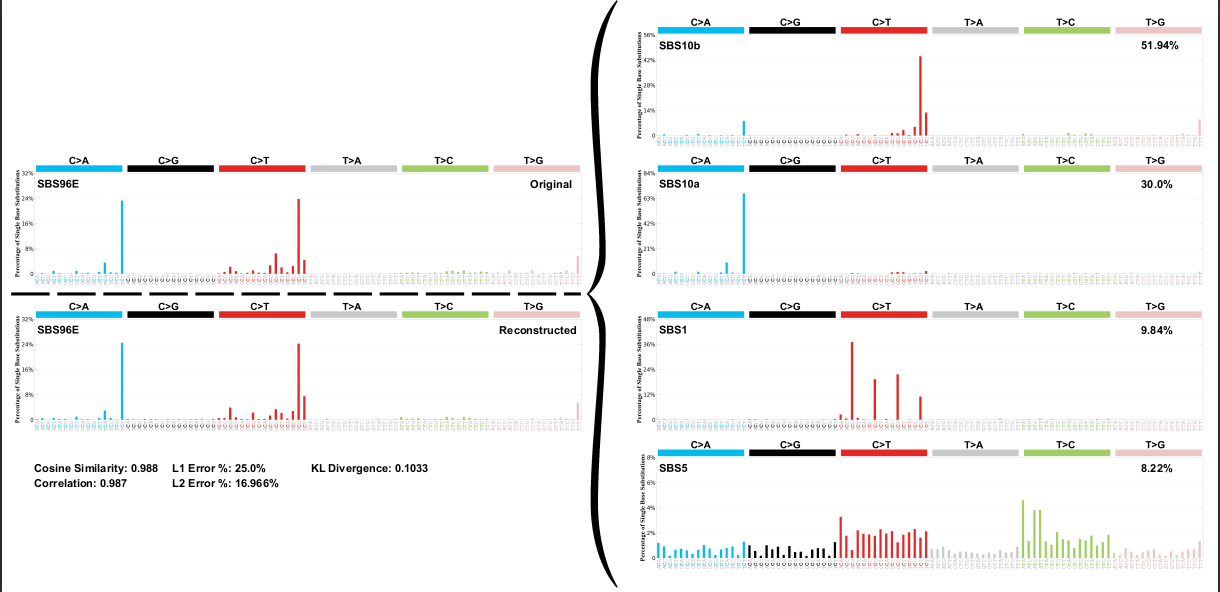
The decomposition plot from figure 07 suggests that the best number of signatures to be extracted from the data we have is 5. The number of mutational signatures that were found to be associated with bladder cancer are 7 from Alexandrov’s paper. The 7 signatures are SBS1, SBS2, SBS5, SBS8, SBS13, SBS29, and SBS40.

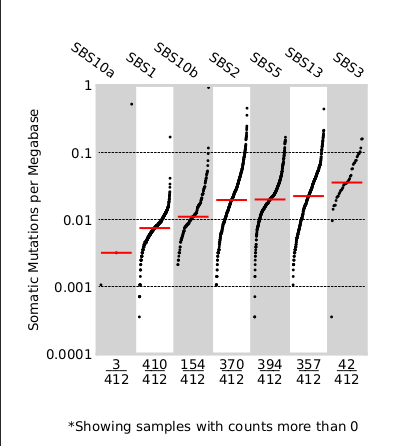
Figure 8: Decomposition plot for SBSA-BLCA

Figure 9: Decomposition plot for SBSB-BLCA

Figure 10: Decomposition plot for SBSC-BLCA

Figure 11: Decomposition plot for SBSD-BLCA

Figure 12: Decomposition plot for SBSE-BLCA

Figure 13: Activity of COSMIC signatures in BLCA

From figure 13 we can see that the most frequent signature is SBS1 as it is present in mostly all the samples. SBS5, SBS2 and SBS13 are also very common which is in accordance with what was shown previously. SBS3 and SBS10a are not that common but SBS10b is relatively well represented despite the fact that it was not expected to show. However, the because it is associated with POLE mutations it makes sense to see it.

SBS8 doesn’t appear but even in Alexandrov’s study it has a weak contribution, therefore I believe that the type and amount of data we have are the reasons for it. Similarly for SBS29 which is associated with tobacco smoking. SBS40 did not appear as well.

So again sigprofiler was unable to deconvolute the signatures totally but the results seems promising as the most significant signatures are present.

## MAFTOOLS to extract the signatures:

I tried maftools to extract mutational signatures and the tools was super-fast. It takes less than a minute to predict the best number of signatures and extract them. However the results did not make sense at all. For example it extracted tobacco signature for breast cancer and SBS46 which is an artefact.

GTFTOOLS to extract UTRs:

Using gtftools I was able to extract the 3’ and 5’ UTRs from the GTF3 annotations. I also extracted the TSS with a window of 1000bp upstream and downstream, making the total size of the window 2000 kb.

This week planning:

* I will start by mapping the mutations to their UTRs and lunch sigprofiler on them for both BRCA and BLCA.
* I could not install sigprofiler on my virtual environment in the cluster due to torch. I don’t know maybe torch is too big (800M) so the system refuses to install it. I sent a ticket but stefano told me that they install only one a year.