Weekly Report

# 28/03/2022-08/04/2022

BOUDEMIA ALA EDDINE

TEAM CHEN

# Goal:

* Re-extract the TSS/TTS regions.
* Validate the extracted TSS/TTS regions.
* Re-extract the mutational signatures acting at the TSS and TTS in BRCA.
* Extract mutational signatures acting at the TSS and TTS in BLCA.

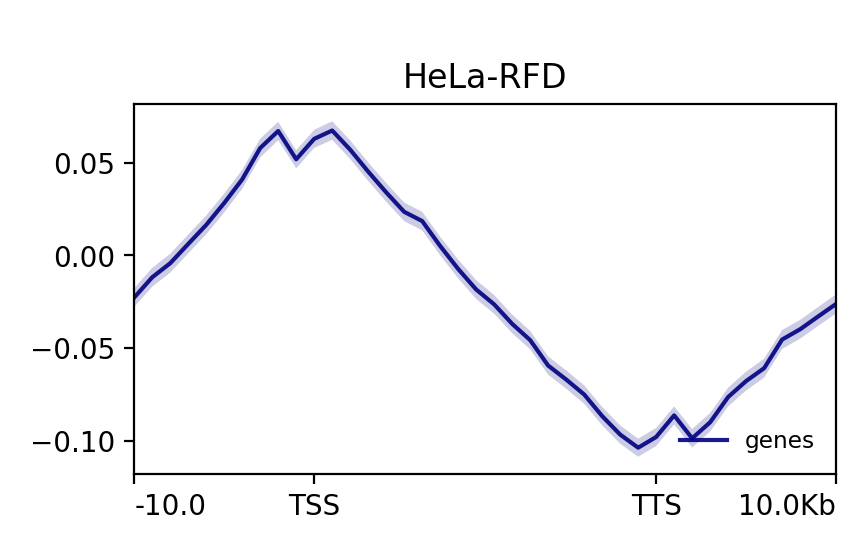
# Procedure:

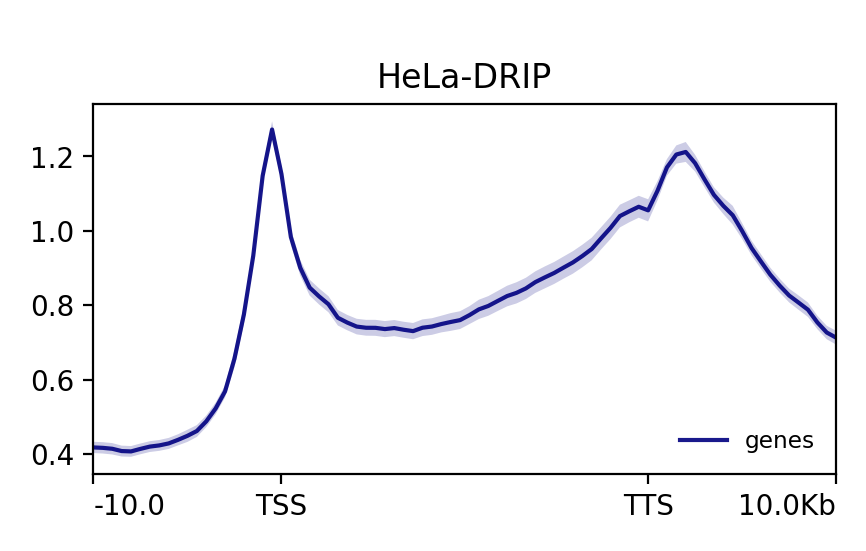
## Extract and validate TSS and TTS regions:

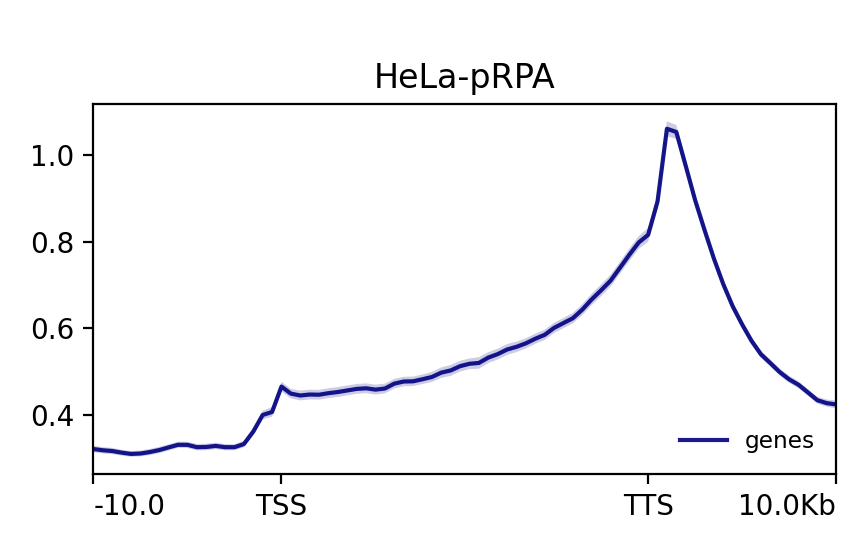
* Using GFF3 annotation from ENSEMBL, I extracted the genes and saved the file in a bed format. I also tried to extract the longest transcript and the results were identical. This yielded approximately 19500 genes.
* Using the extracted genes, I set a window of +/- 3kb to define the TSS and TTS regions.
* To validate these results, I plotted the mean RFD, DRIP and pRPA profiles using the data from the HeLa cells. Because the data were aligned on the GRCh37, I used the liftover tool to convert the coordinates of my data, as I am working with GRCh38.

# Results:

## RFD, DRIP and pRPA profiles:







## Mutational signatures at TSS and TTS in BRCA:

* The results didn’t change much, except for the selection plot in the TSS it has selected only two signatures, but the decomposition gave the same results as previously and the 4 signatures were very similar to previous results as well.

## Mutational signatures at TSS and TTS in BLCA:

## TSS:

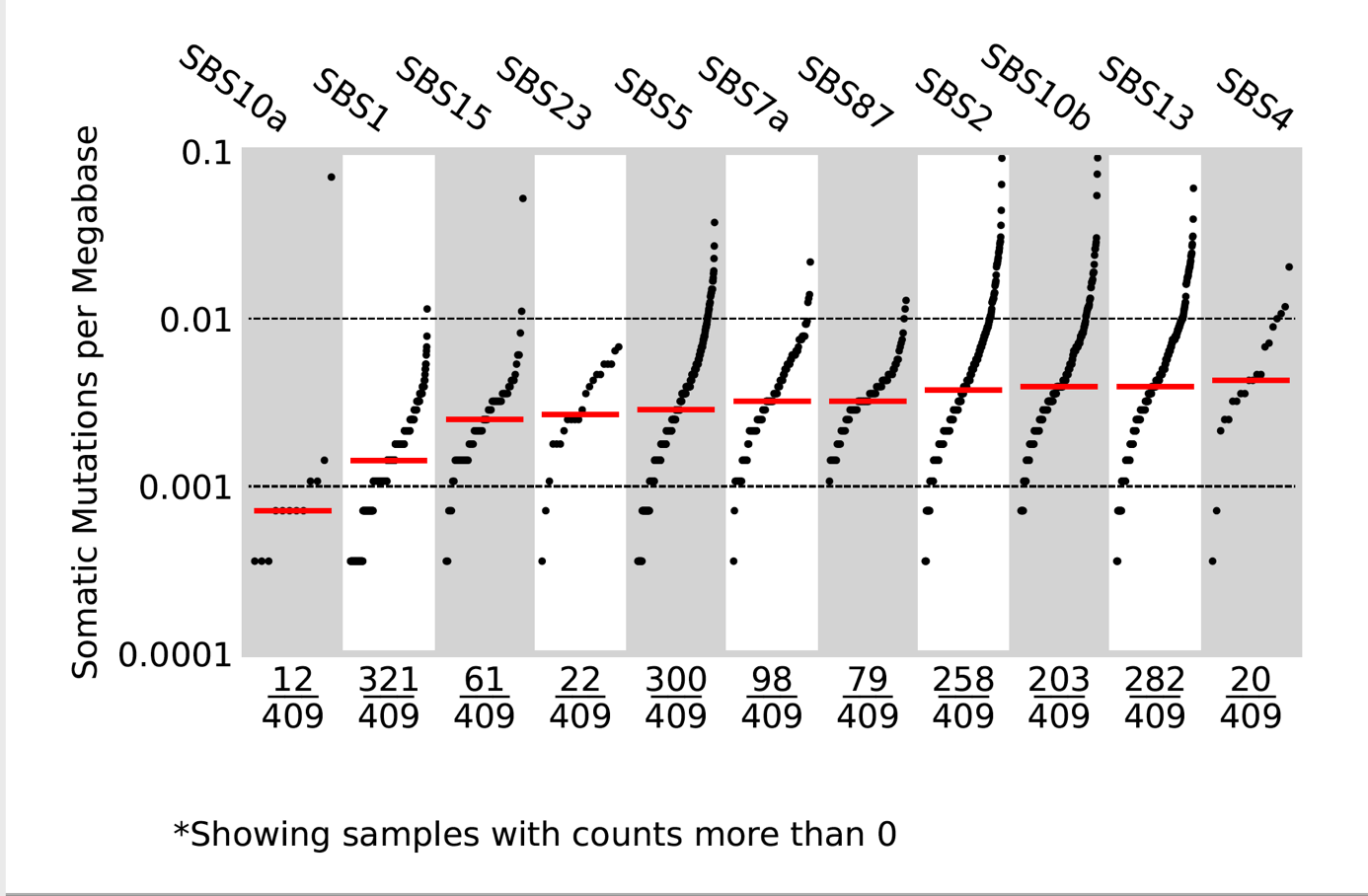


Figure 1 Activity plot of the COSMIC signatures after decomposition

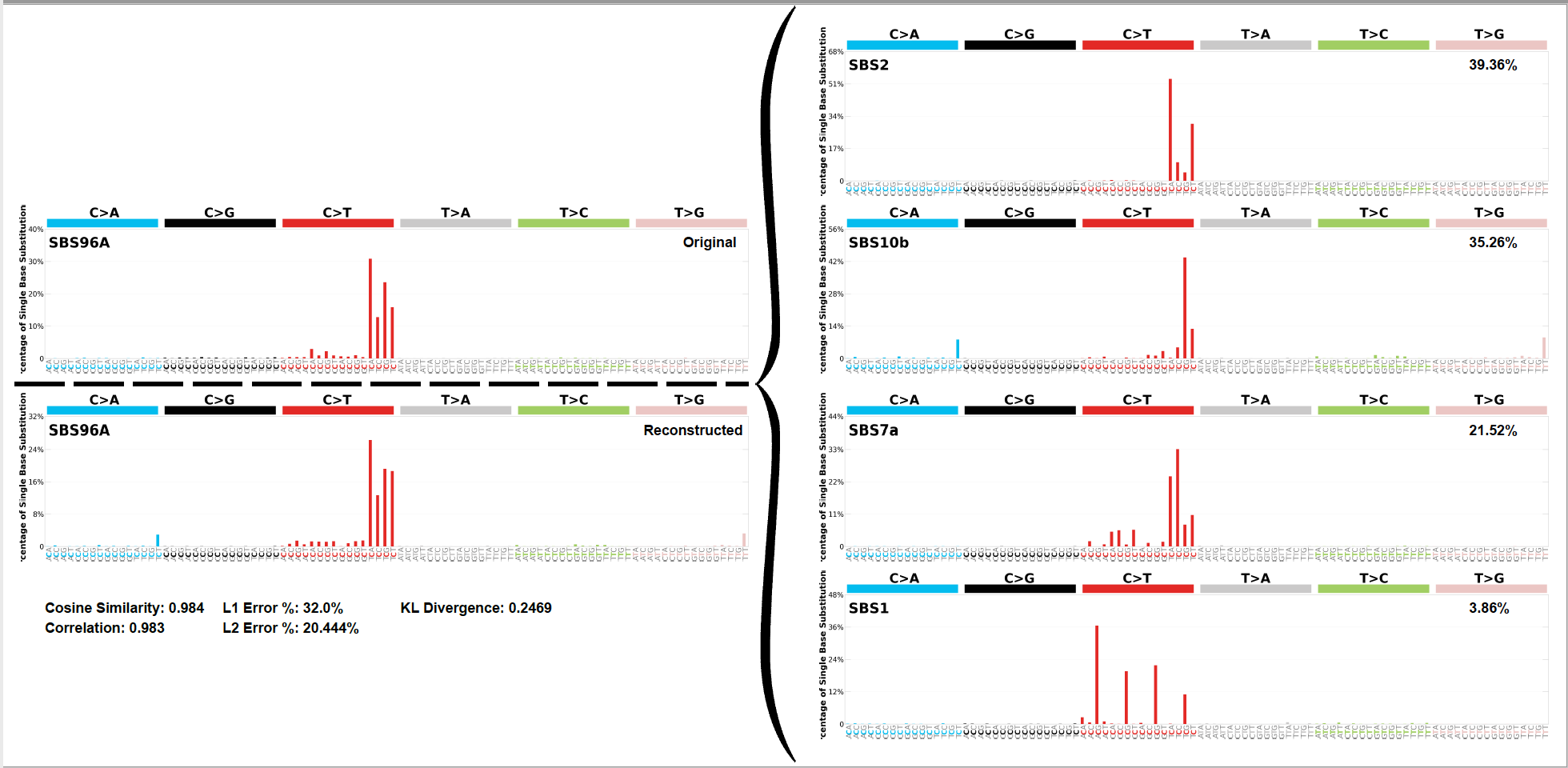


Figure 2 Decomposition plot for SBSA at the TSS

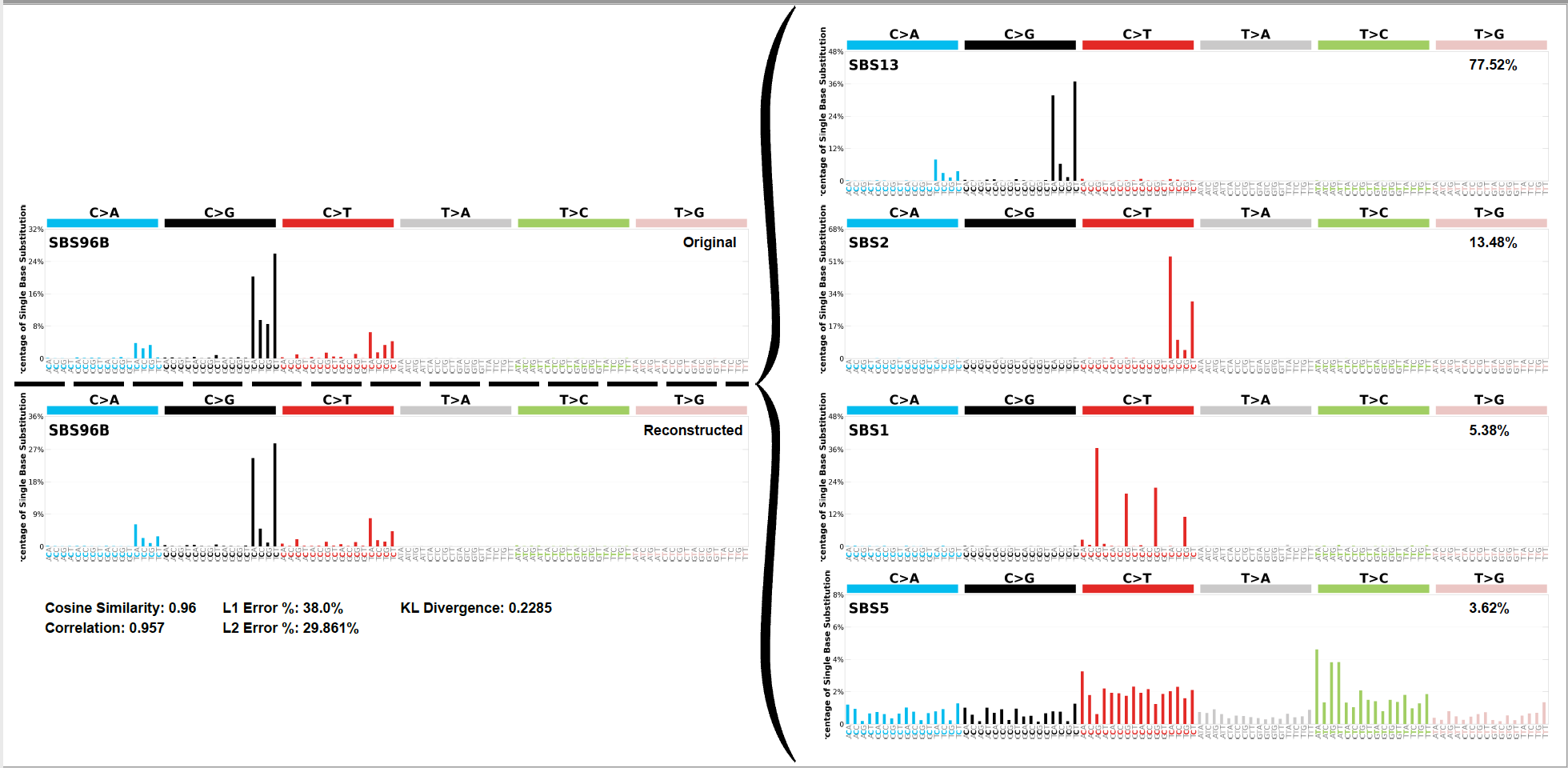


Figure 3 Decomposition plot for SBSB at the TSS

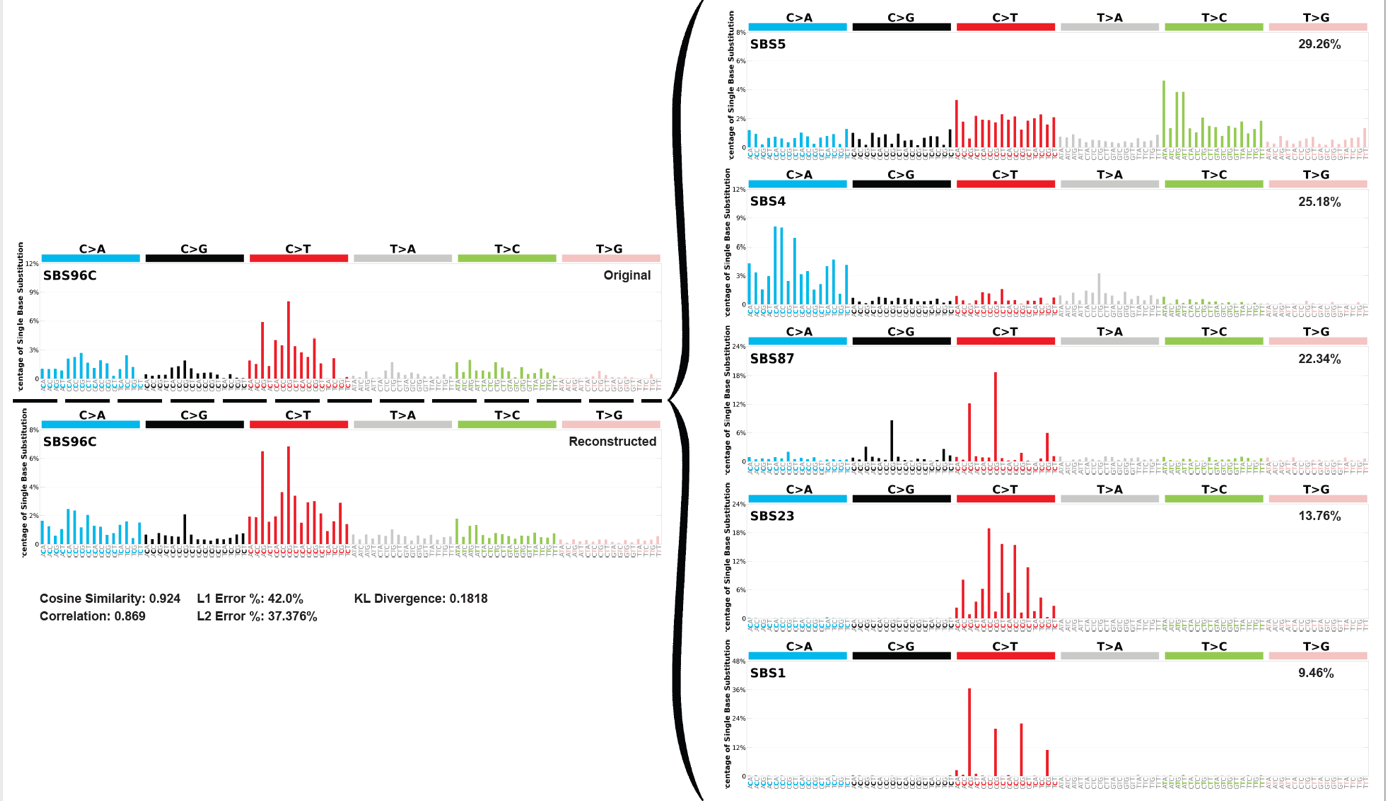


Figure 4 Decomposition plot for SBSC at the TSS

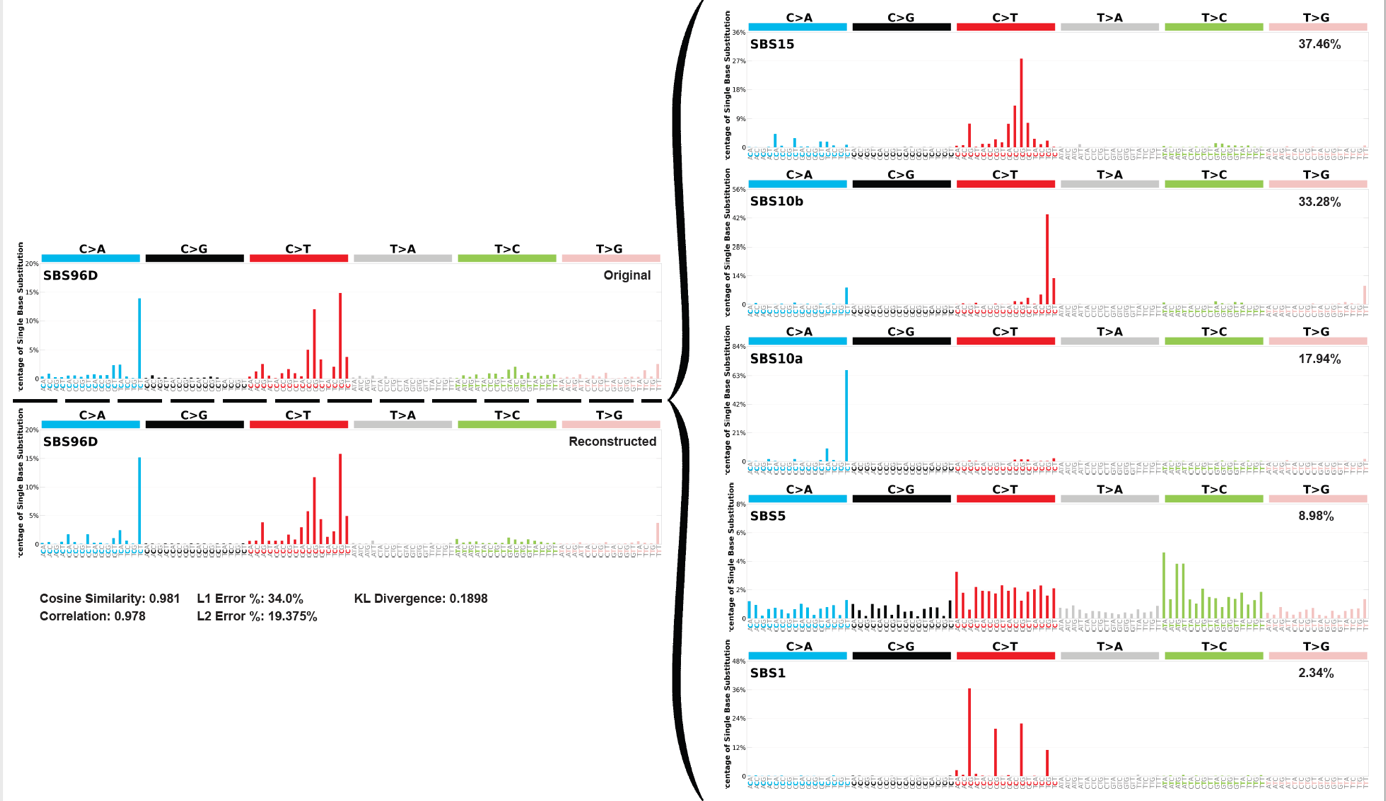


Figure 5 Decomposition plot for SBSD at the TSS

## TTS:

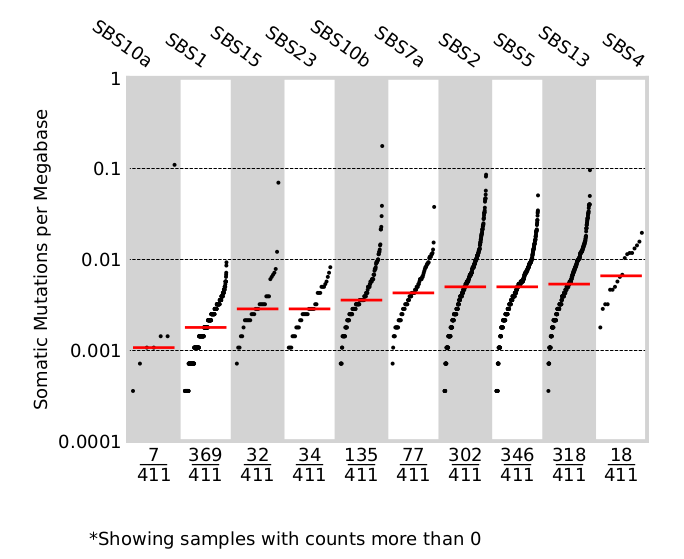


Figure 6 Activity of each COSMIC signature in the samples at TTS

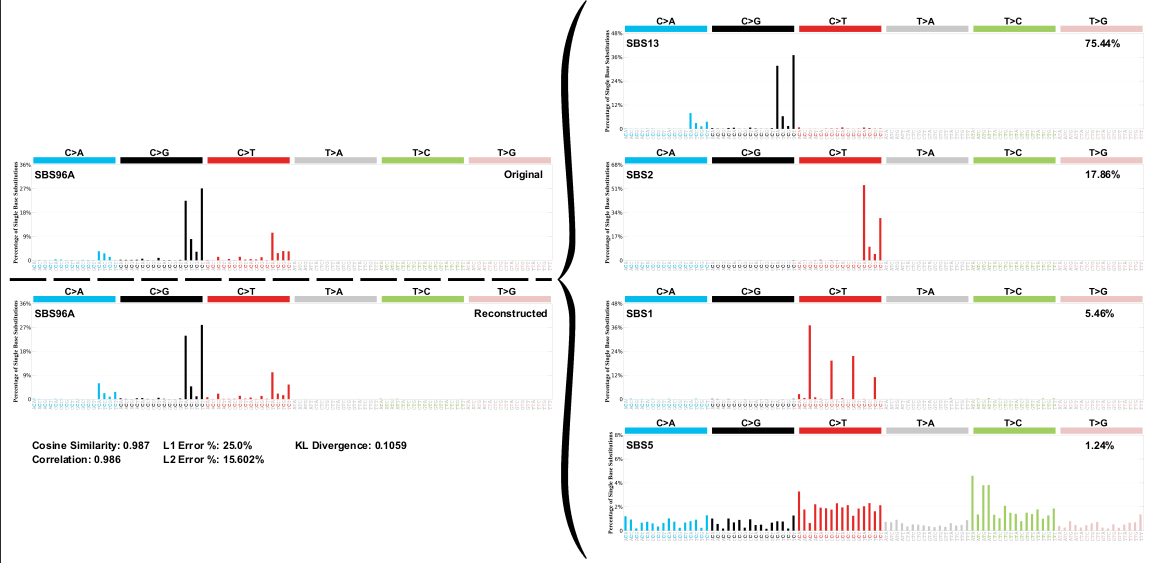


Figure 7 Decomposition plot for SBSA at the TTS

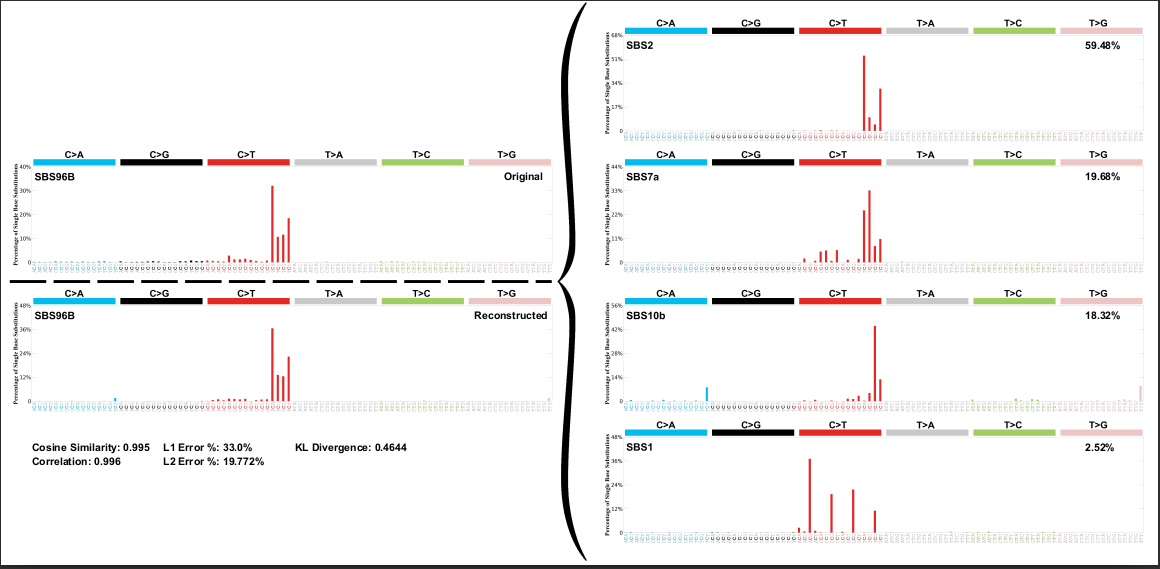


Figure 8 Decomposition plot for SBSB at the TTS

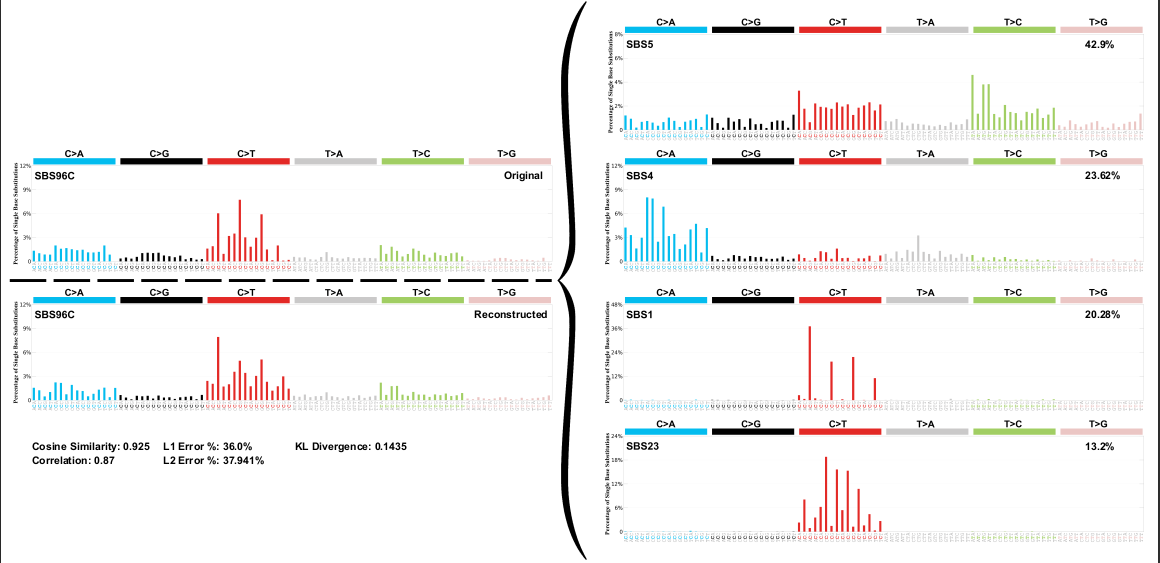


Figure 9 Decomposition plot for SBSC at the TTS

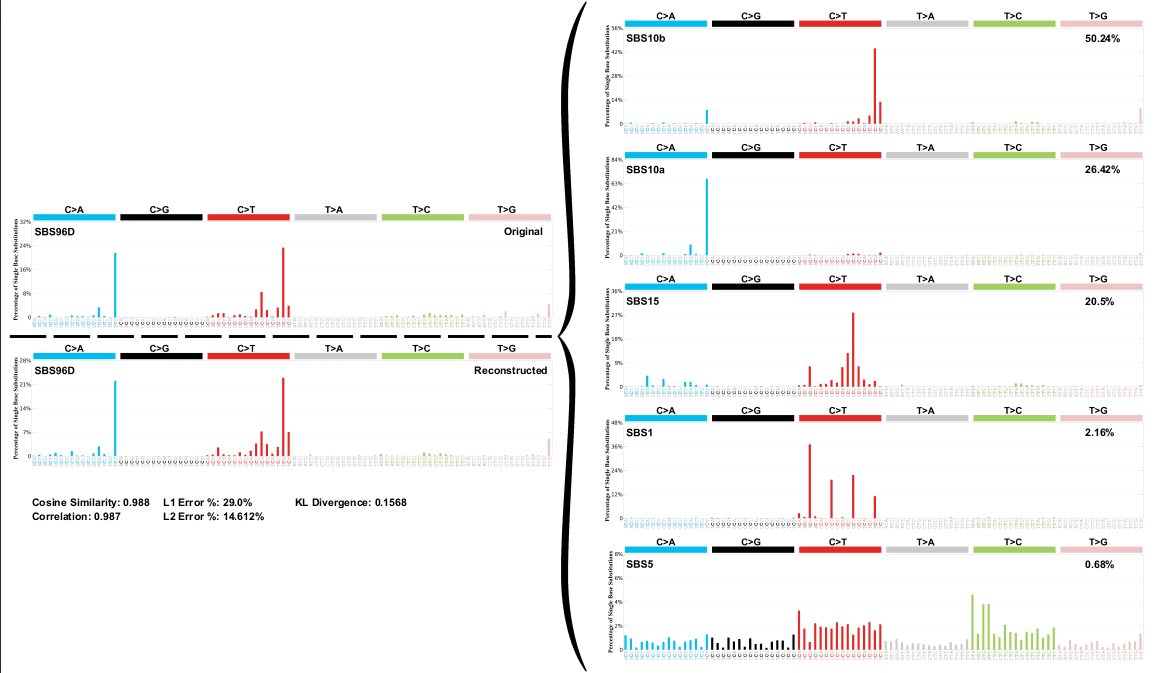


Figure 10 Decomposition plot for SBSD at the TTS

* The signatures extracted at the TSS and TTS regions of BLCA project are similar also but different from the signatures extracted previously on the whole dataset as there are new signatures that were detected.
* Globally, it is necessary to perform this analysis on WGS because the WES data do not provide mutations that are upstream/downstream TSS/TTS especially when considering a +/- 3kb window.
* I think that it is also better to keep the analysis specific to each tissue for the sake of this project because depending on the tissue type the chromatin organization change and thus one specific TSS/TTS could be in open chromatin in one tissue and in closed chromatin in another especially that the mutation rates change between the two types.
* Unfortunately, the MAF files I am using (which groups every MAF file in only one file) are no longer accessible on TCGA (I only have BRCA, BLCA). TCGA WGS is available only as BAM files and require permission. PCAWG at ICGC is available as vcf but require permission as well.