Weekly Report

# 25/04/2022

BOUDEMIA ALA EDDINE

TEAM CHEN

# Goal:

* Extract mutational signatures operating in active and inactive genes

# Procedure:

## Select active and Inactive genes:

* Download RNA-Seq data for TCGA-BRCA and TCGA-BLCA from recount2 project.
* Extract the read counts computed for paired end sequencing, then compute CPM.
* Select the threshold of 1 to determine active/inactive genes – If considered as log transformed this will mean that the cut-off is 0.
  + This yielded 14154 active gene and 3955 inactive genes for BRCA and 14140 active genes and 3969 inactive genes for BLCA.
  + 231 genes were not present in the RNA-seq data. Some of them are absent because these genes were only added into GRCh38 annotation, but I am not sure if this applies to all of them. Following is the Panther GO for some of these genes (65) the remaining that were not mapped seem to be the novel genes aforementioned (through a skim, if necessary, I will make thorough analysis):

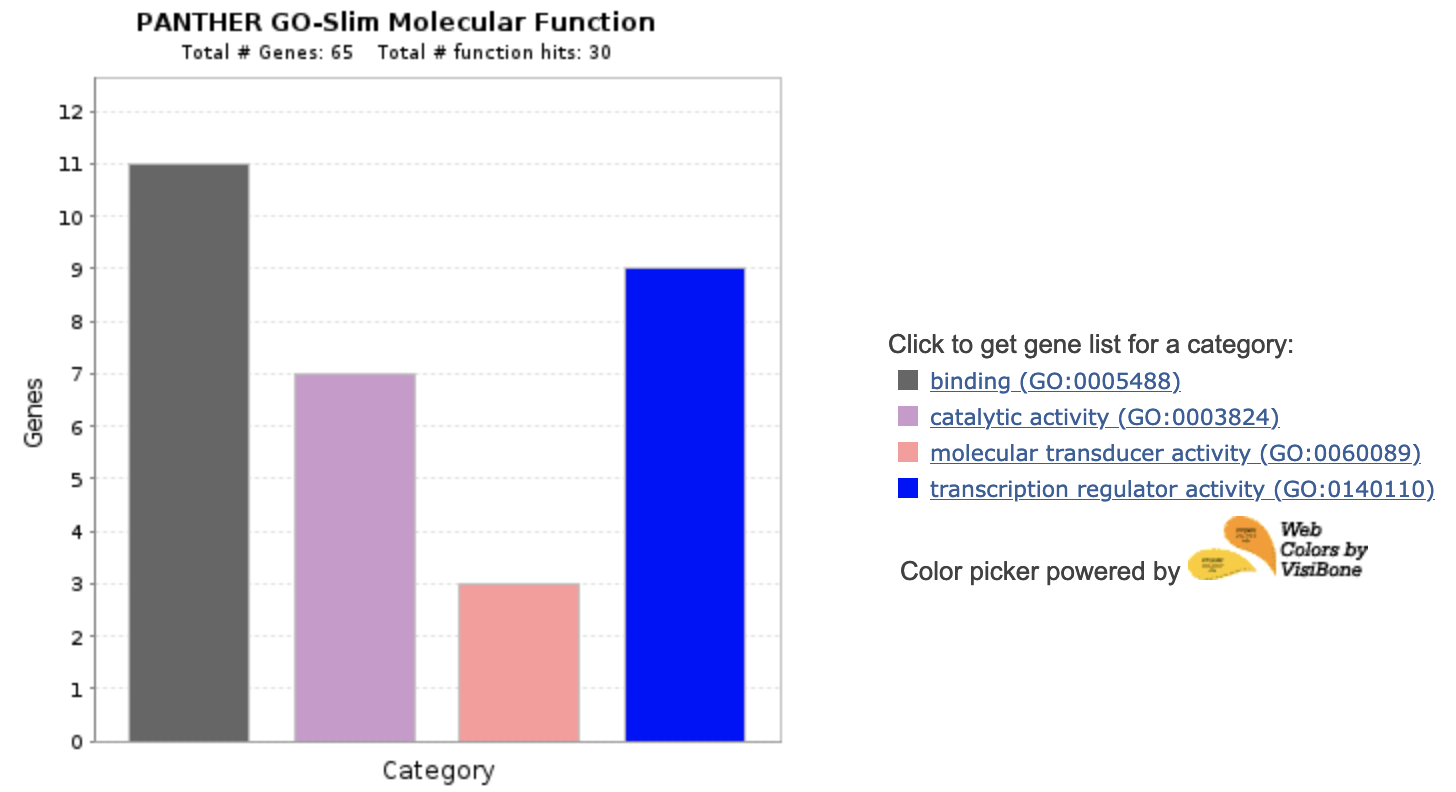


Figure 1 Panther GO-slim Molecular Function for the Missing Genes

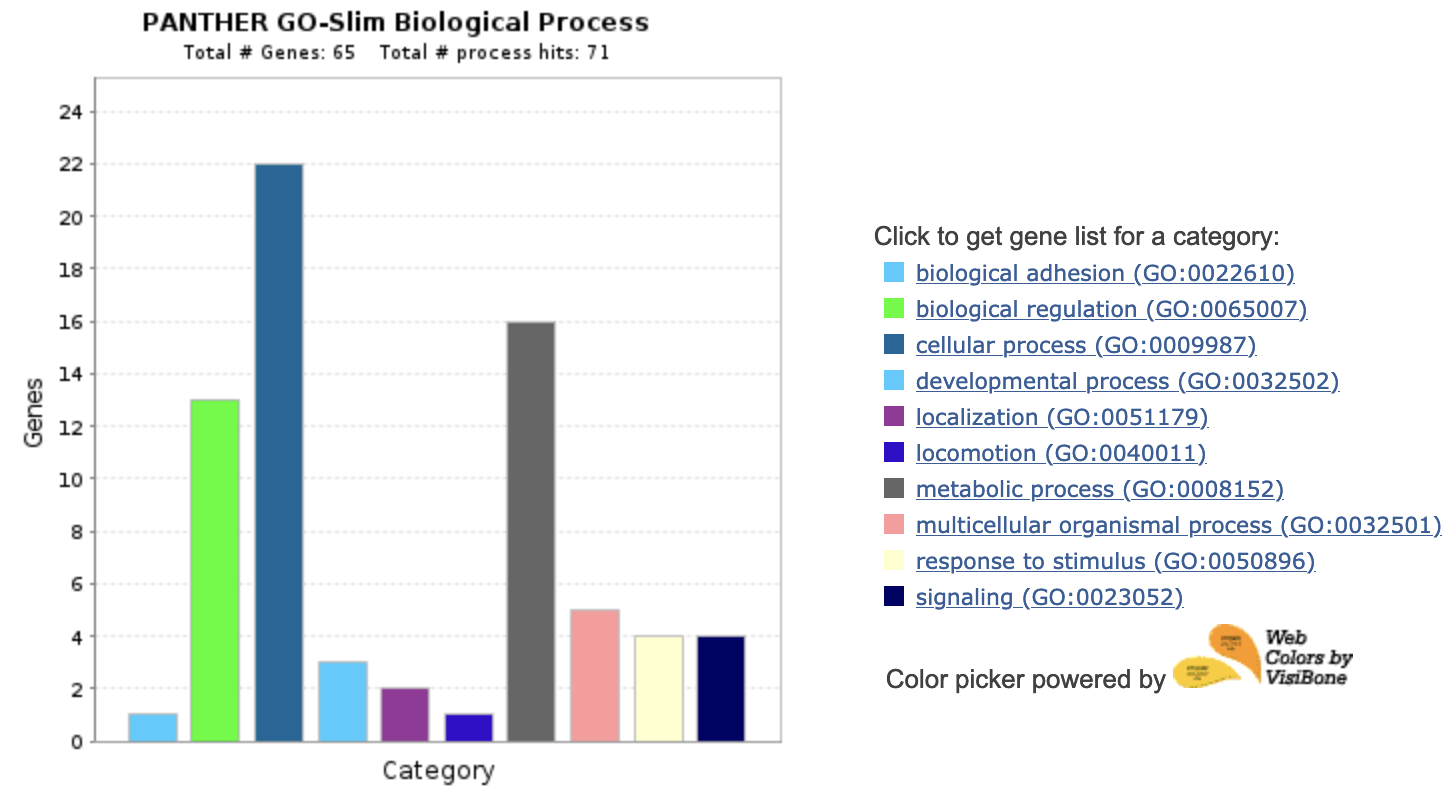


Figure 2 Panther GO-slim Biological Process for the Missing Genes

* After selecting the active-inactive genes from the set of the 60k genes. The TSS and TTS are extracted relative to each cancer (+/- 3kb window).
* Based on the extracted TSS-TTS the mutations will be sorted to each region. This yielded:

Table 1 Mutation distribution between TSS and TTS of active and inactive genes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cancer | BRCA | | BLCA | |
| State | Active | Inactive | Active | Inactive |
| TSS | 9840 | ~~4805~~ 4798 | 14263 | ~~5427~~ 5414 |
| TTS | 16933 | ~~6706~~ 6687 | 19359 | ~~7242~~ 7222 |
| Common | 3378 | 2796 | 4914 | 3102 |

* As previously, there is a bias in the distribution of mutations between the TSS and TTS. That seems independent of the state of the genes but is higher in active genes. This could be due just related to the fact that there are more active genes (to be verified).
* Finally, run SigProfiler to generate the mutational profiles and then the mutational signatures for each case of each cancer.

# Results:

## BRCA:

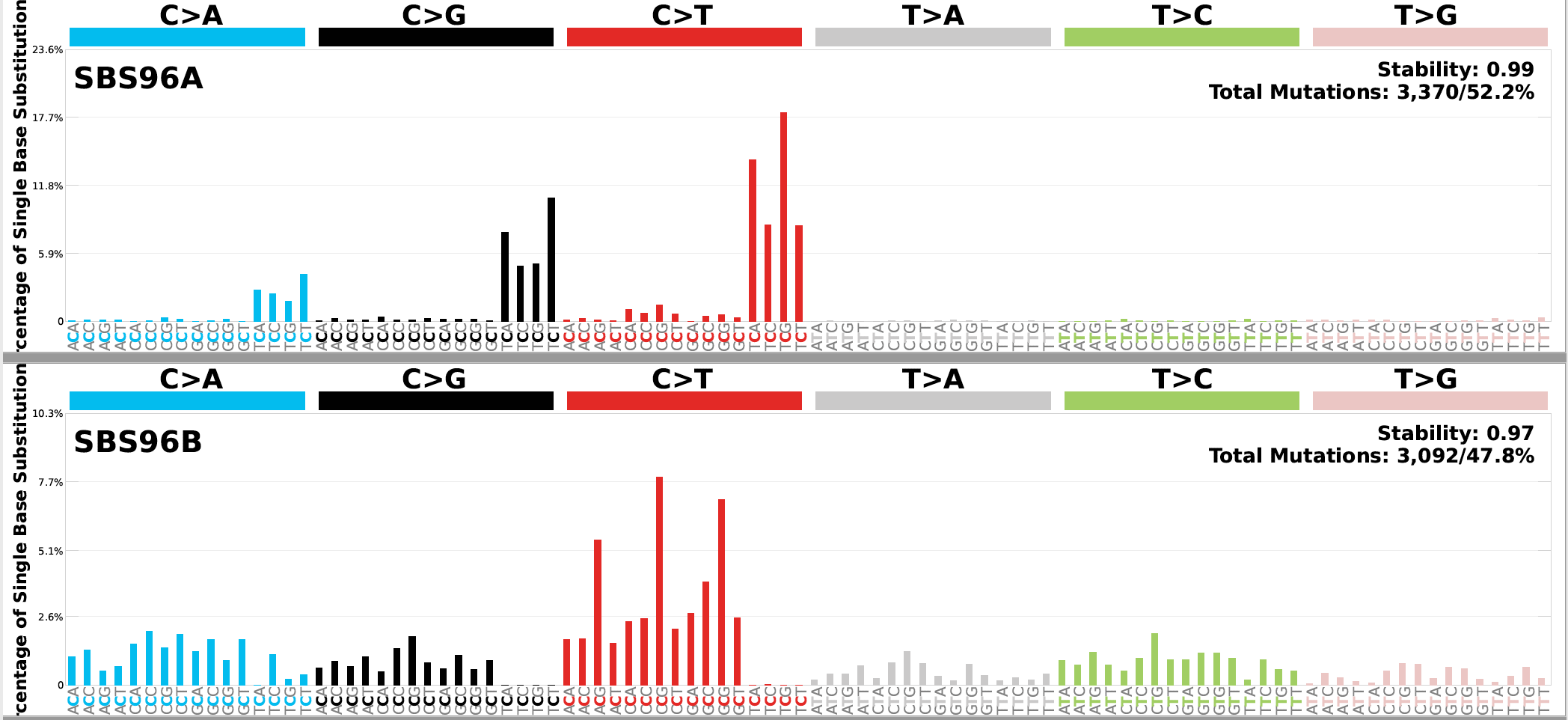


Figure 3 Signatures extracted from the TSS of active genes

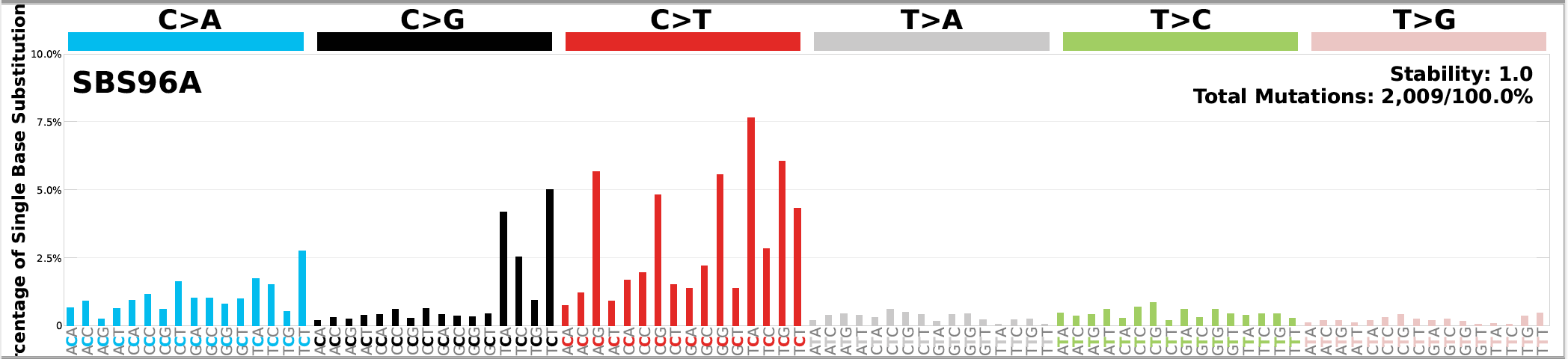


Figure 4 Signatures extracted from the TSS of inactive genes

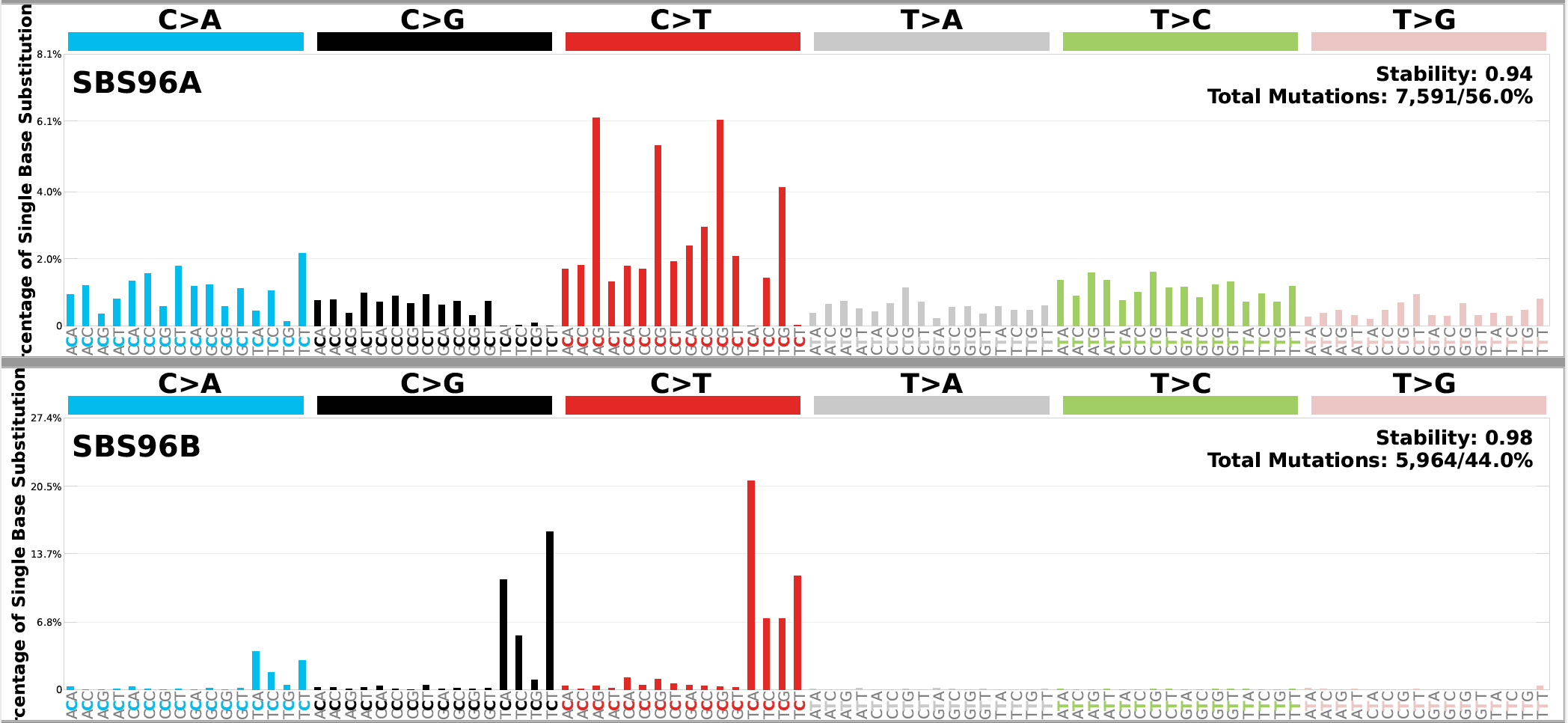


Figure 5 Signatures extracted from the TTS of active genes

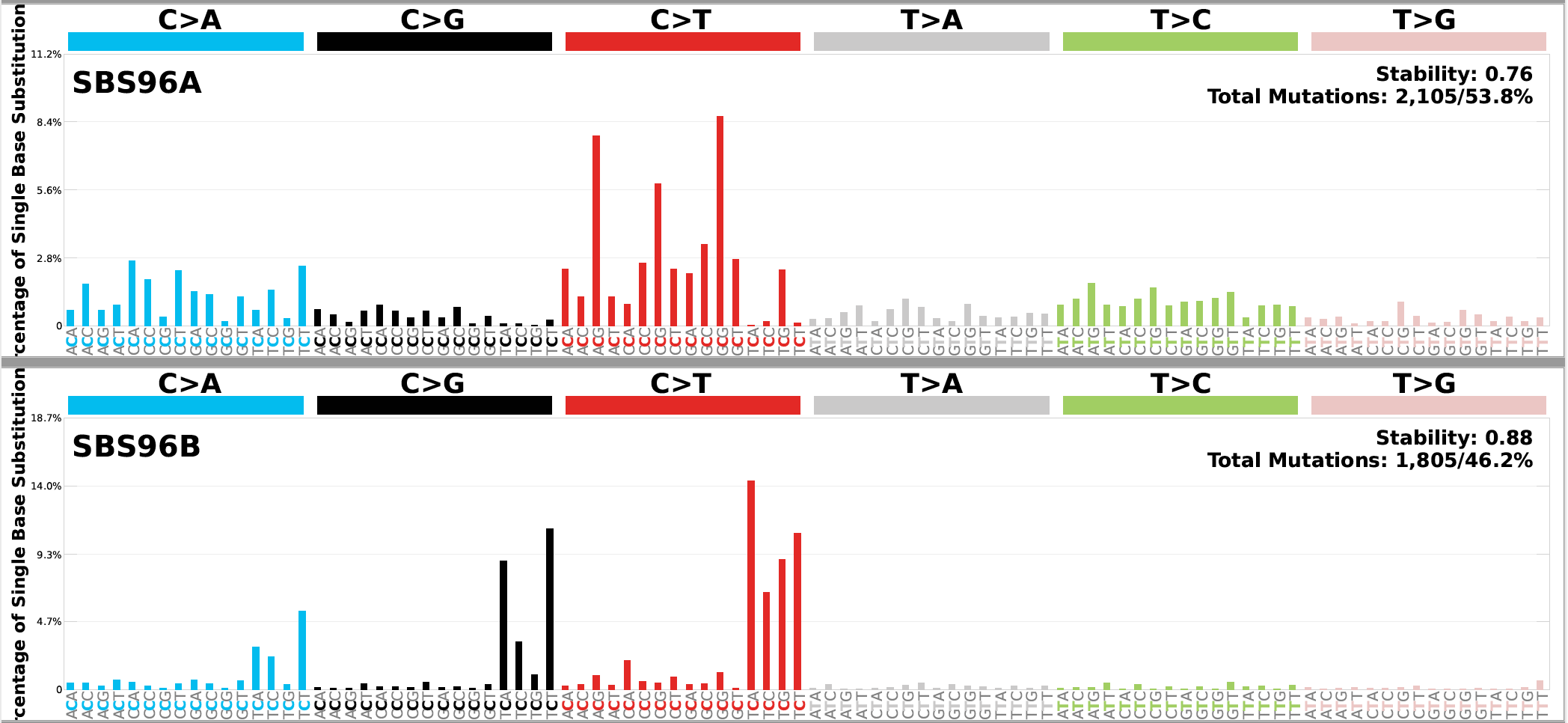


Figure 6 Signatures extracted from the TTS of inactive genes

* These results show a slight difference between the signatures operating on the TSS of active genes and those operating at their TTS. (Better to compare the signatures using cosine similarity; but from the decomposition plot only one of them show SBS6)
* There is no difference between the signatures operating at active or inactive genes. However, at the TSS level SigProfiler choose only one signature as a best result probably because it had difficulties to separate the signature into two due to the lack of mutational density. This further confirmed when checking the results from two signatures I can see that there is a similarity, but the stability of the signature dropped to 0,66.
* Cosine similarity between SBS96A from active TSS and SBS96B from active TTS is 0.885
* Cosine similarity between SBS96B from active TSS and SBS96A from active TTS is 0.92

## BLCA:

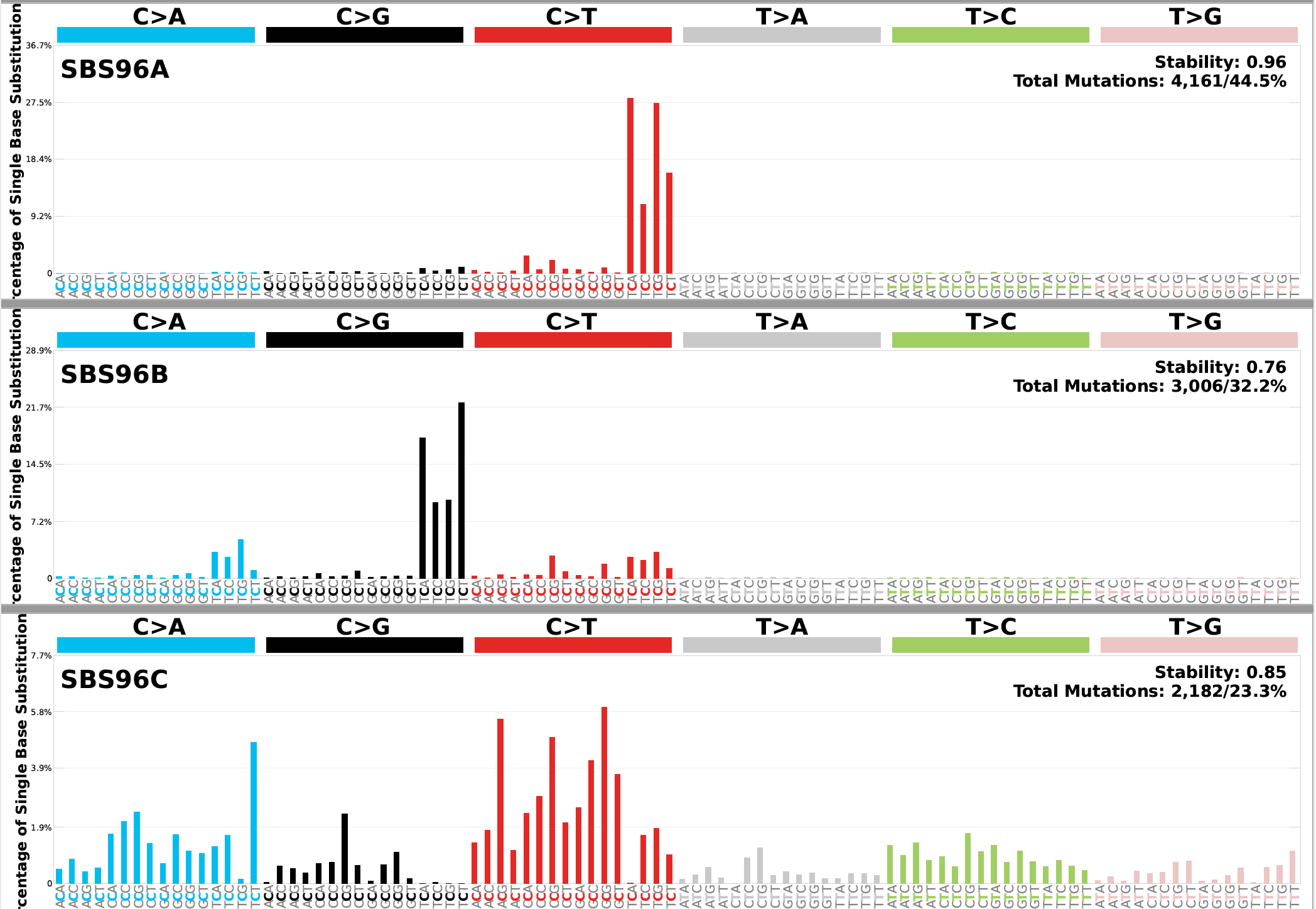


Figure 7 Signatures extracted from the TSS of active genes

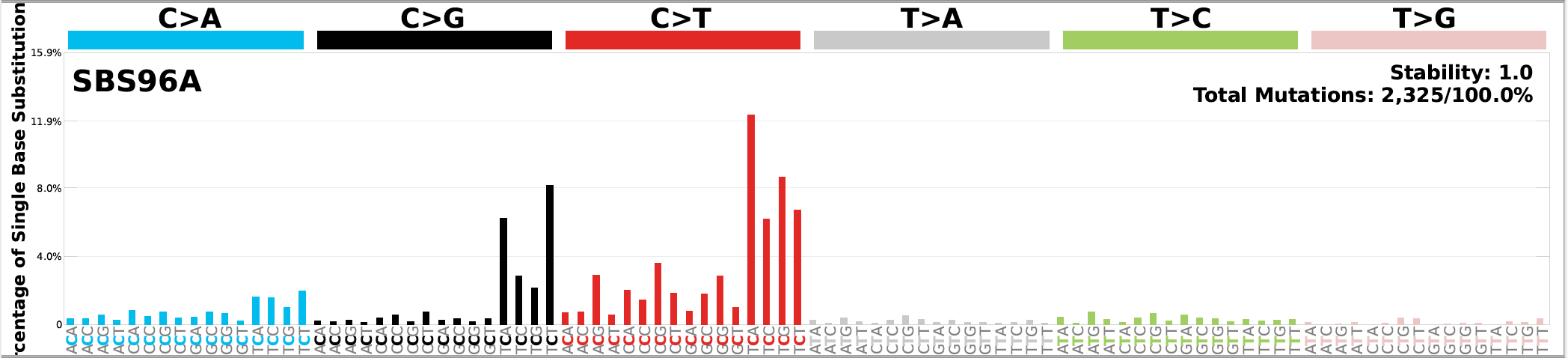


Figure 8 Signatures extracted from the TSS of inactive genes

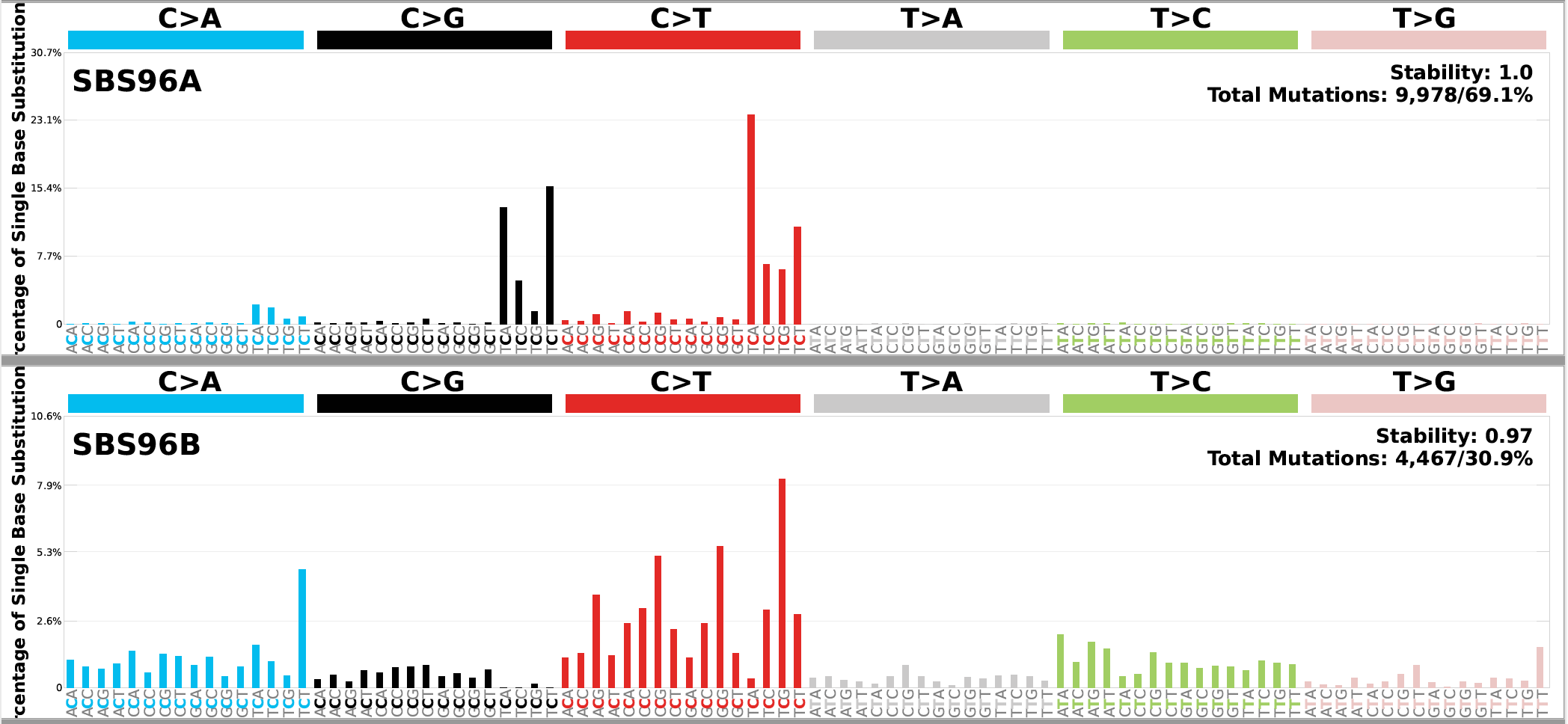


Figure 9 Signatures extracted from the TTS of active genes

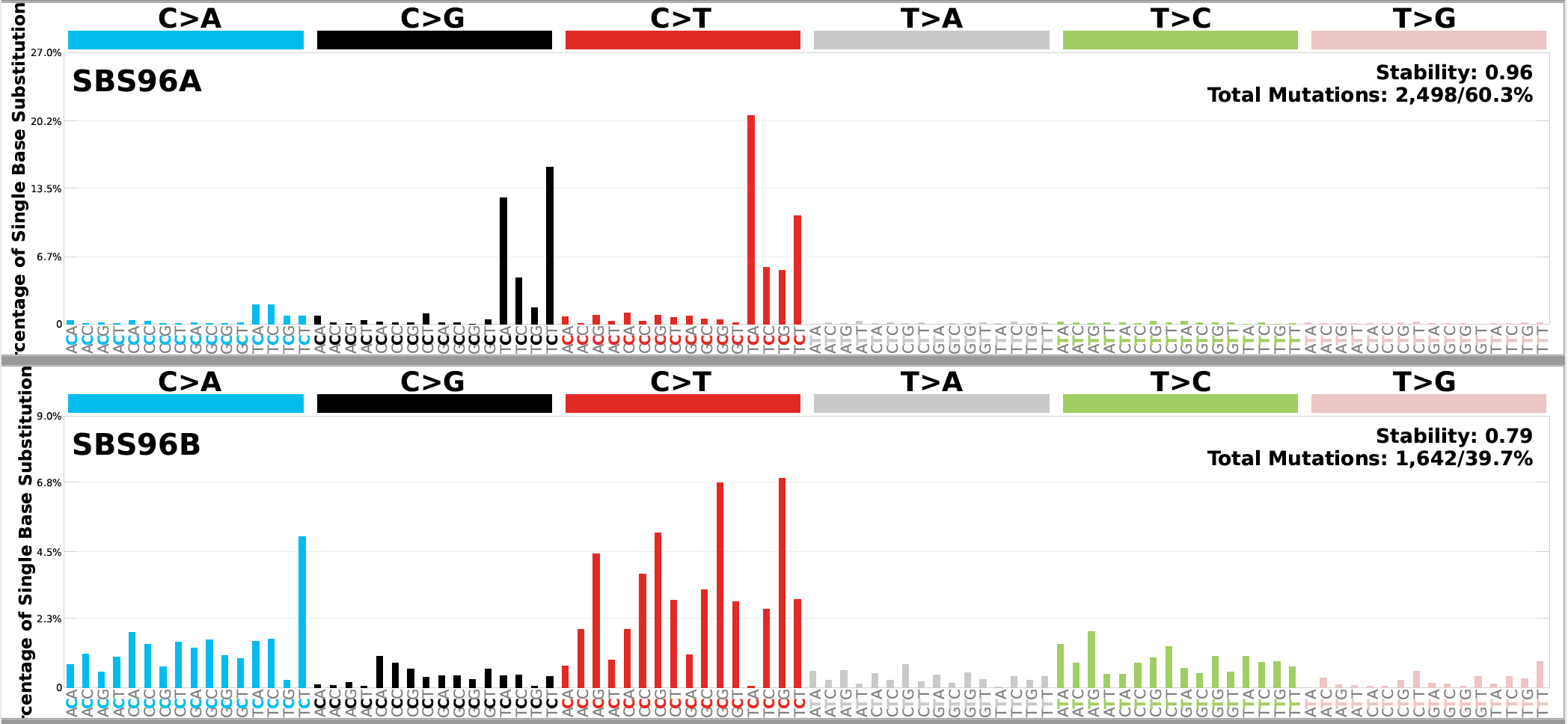


Figure 10 Signatures extracted from the TTS of inactive genes

* Same observation for BRCA apply for BLCA as well