Class 18 Mini Project

Jessica Diaz-Vigil

2023-06-02

Exploring a Cancer Sequencing Data Portal

Skin Cutaneous Melanoma

Discussion #1

Q1. How many cancer samples are included in the dataset?

448

Q2. Which is the most mutated gene?

TTN

Q3. Which is the most common treatment undergone by patients?

Radiation 1

Downloading Cancer Sequencing Data

```
library(maftools)
skcm = read.maf('skcm_tcga_pan_can_atlas_2018/data_mutations.txt')
```

```
## -Reading
## -Validating
## --Removed 27563 duplicated variants
## -Silent variants: 209854
## -Summarizing
## --Possible FLAGS among top ten genes:
##    TTN
##    MUC16
## -Processing clinical data
## --Missing clinical data
## -Finished in 33.2s elapsed (24.4s cpu)
```

```
mm_skcm = trinucleotideMatrix(maf = skcm, prefix = 'chr', add = TRUE, ref_genome = "BSgenome.Hsapiens.UC
```

```
##
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':
##
```

```
##      anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##      colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##      get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##      match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##      Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##      table, tapply, union, unique, unsplit, which.max, which.min

##
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:base':
##
##      expand.grid, I, unname

##
## Attaching package: 'Biostrings'

## The following object is masked from 'package:base':
##
##      strsplit

## -Extracting 5' and 3' adjacent bases
## -Extracting +/- 20bp around mutated bases for background C>T estimation
## -Estimating APOBEC enrichment scores
## --Performing one-way Fisher's test for APOBEC enrichment
## ---APOBEC related mutations are enriched in  1.818 % of samples (APOBEC enrichment score > 2 ;  8  o
## -Creating mutation matrix
## --matrix of dimension 440x96
```

```r
mm_skcm = t(mm_skcm$nmf_matrix)
```

```r
library(MutationalPatterns)
```

```
## Loading required package: NMF

## Loading required package: registry

## Loading required package: rngtools

## Loading required package: cluster

## NMF - BioConductor layer [OK] | Shared memory capabilities [NO: bigmemory] | Cores 2/2

##   To enable shared memory capabilities, try: install.extras('
## NMF
## ')

##
## Attaching package: 'NMF'

## The following object is masked from 'package:S4Vectors':
##
##      nrun
```
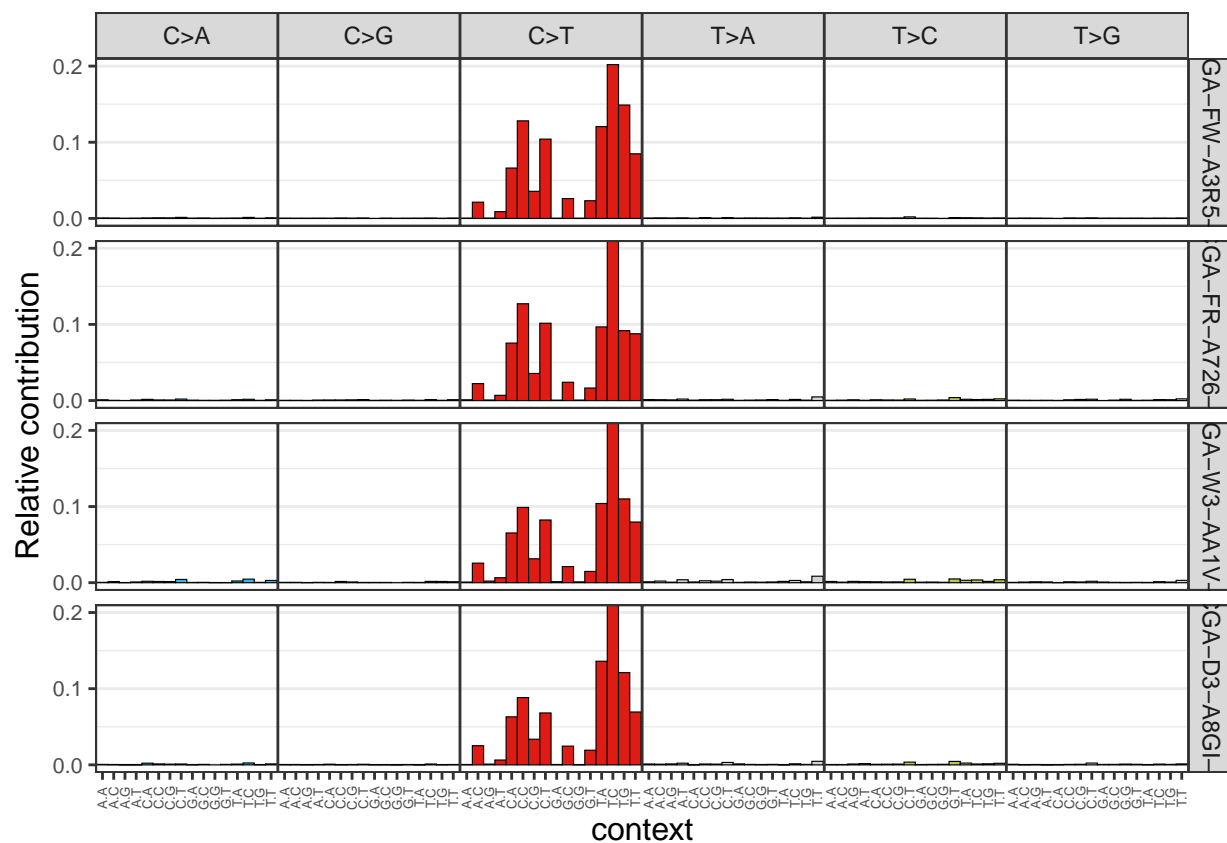
```r
set.seed(11111)
```

```r
samples_to_plot = sample(1:ncol(mm_skcm),4)
plot_96_profile(mm_skcm[,samples_to_plot], condensed = T)
```

```
mutations_in_samples = colSums(mm_skcm)
mutations_in_samples = sort(mutations_in_samples, decreasing = T)
samples_to_plot = names(mutations_in_samples)[1:4]
plot_96_profile(mm_skcm[,samples_to_plot], condensed = T)
```

```
mutations_in_samples = sort(mutations_in_samples, decreasing = F)
samples_to_plot = names(mutations_in_samples)[1:4]
plot_96_profile(mm_skcm[,samples_to_plot], condensed = T)
```
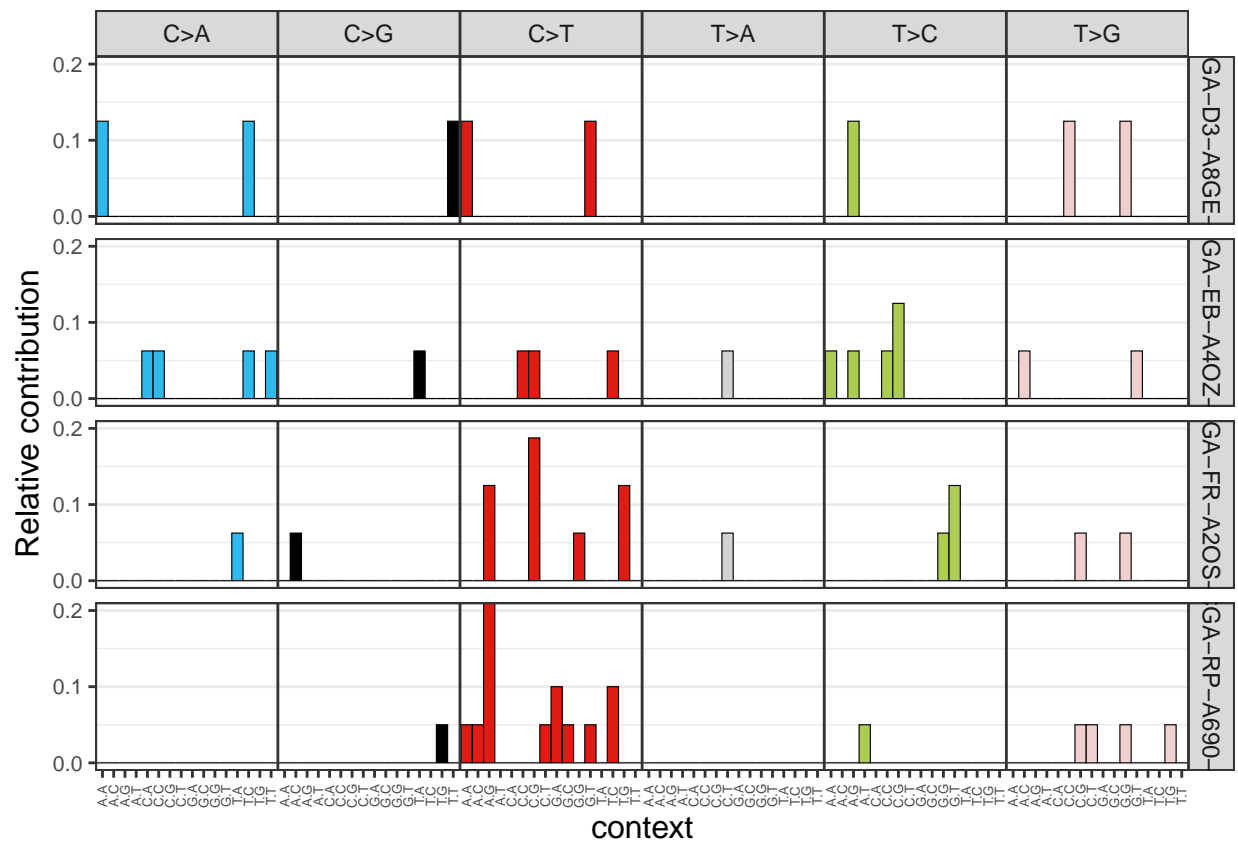
```
relative_mutational_profile = apply(mm_skcm, 2, prop.table)

average_mutational_profile = rowMeans(relative_mutational_profile)
average_mutational_profile = data.frame(average_mutational_profile)
plot_96_profile(average_mutational_profile, condensed = T)
```
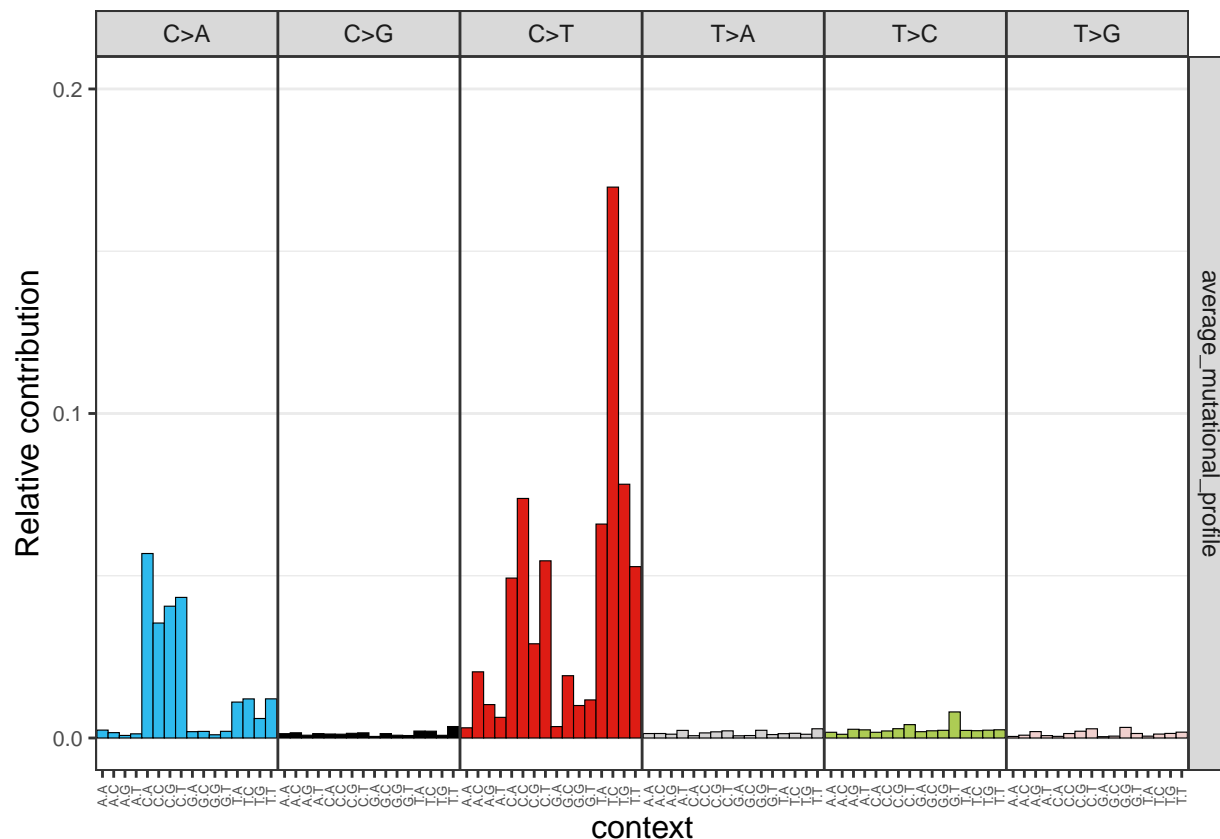
COSMIC Reference Mutational Signatures

Assigning Reference Mutational Signatures

```
cosmic_signatures = get_known_signatures(source = 'COSMIC_v3.2')

fit_res = fit_to_signatures(mm_skcm, cosmic_signatures)

contributions = fit_res$contribution
top_contributing_signatures_abs = rowMeans(contributions)

top_contributing_signatures_abs = sort(top_contributing_signatures_abs, decreasing = T)[1:4]

top_contributing_signatures_abs
```

```
##      SBS7a      SBS7b      SBS38       SBS4
## 366.97614 340.91011 204.44450   99.49106
```

```
relative_contributions = apply(contributions,2,prop.table)

top_contributing_signatures_rel = rowMeans(relative_contributions)

top_contributing_signatures_rel = sort(top_contributing_signatures_rel, decreasing = T)[1:4]

top_contributing_signatures_rel
```

```
##      SBS7b      SBS7a      SBS38      SBS4
## 0.26336351 0.26019455 0.10885595 0.07240978
```

```
fit_res_strict = fit_to_signatures_strict(mm_skcm, cosmic_signatures)

fit_res_strict = fit_res_strict$fit_res

contributions_strict = fit_res_strict$contribution
```
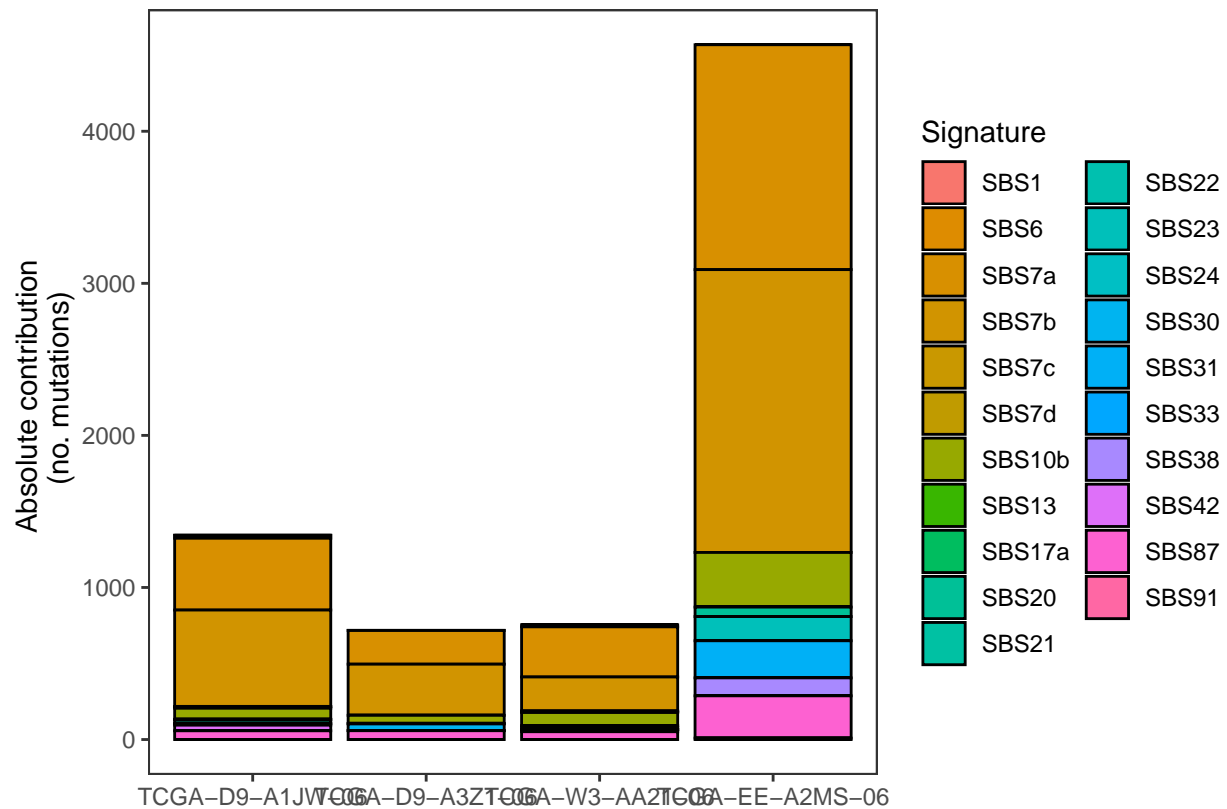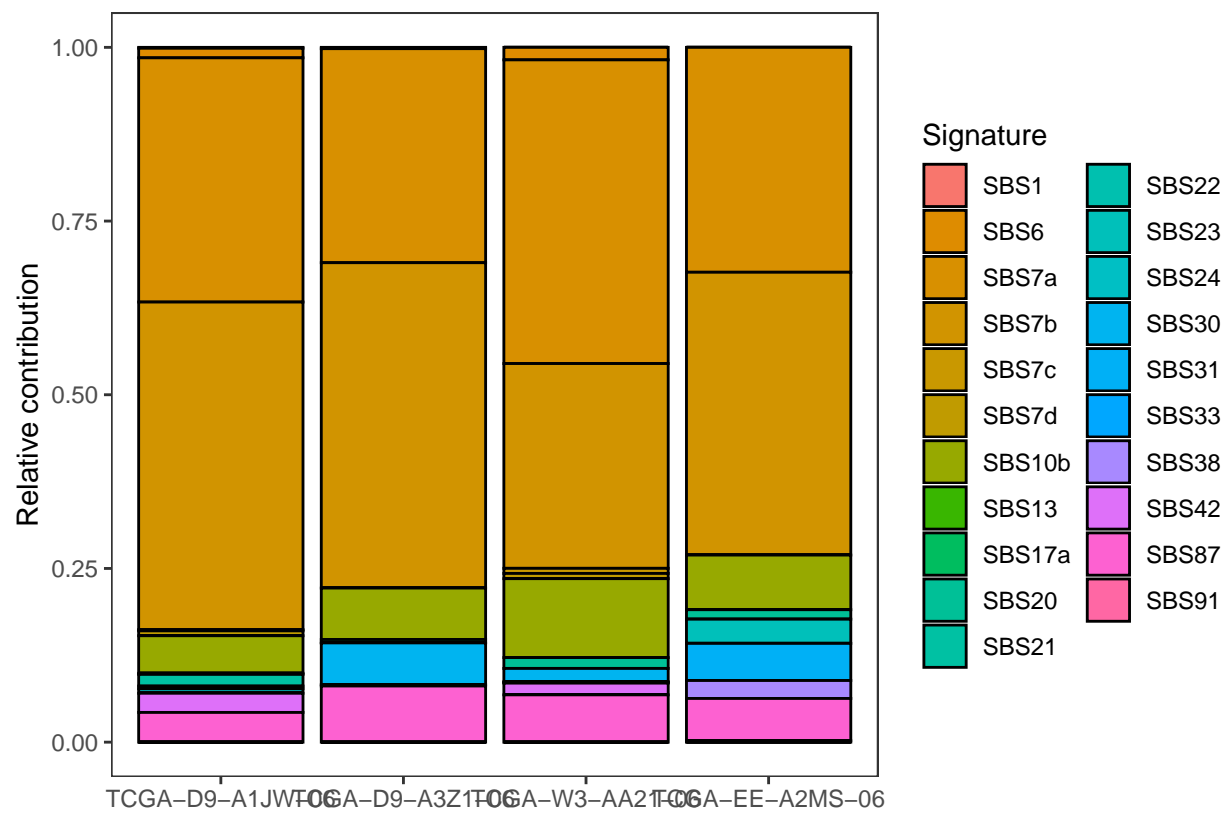
Visualizing Mutational Signature Assignment Results

```
set.seed(11111)
samples_to_plot = sample(1:ncol(mm_skcm),4)

plot_contribution(contributions[,samples_to_plot], mode = "absolute")
```
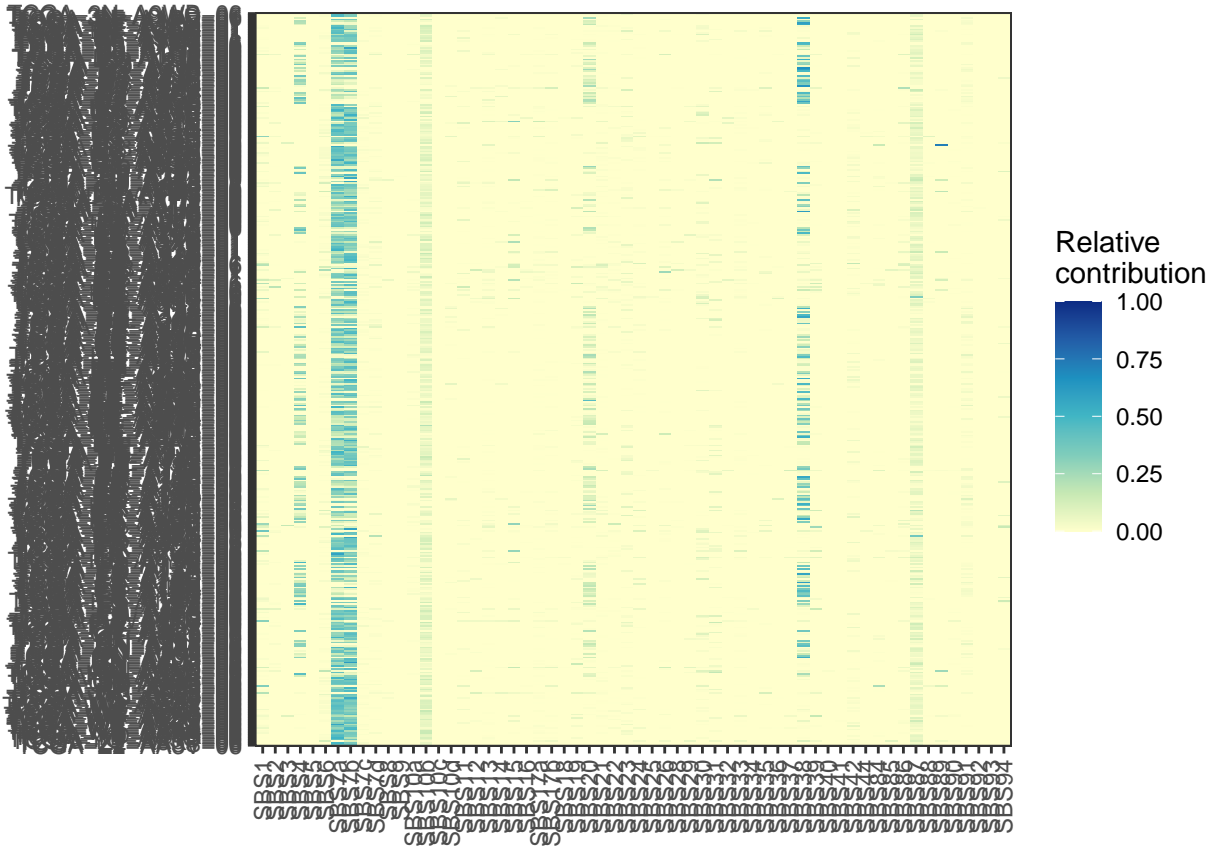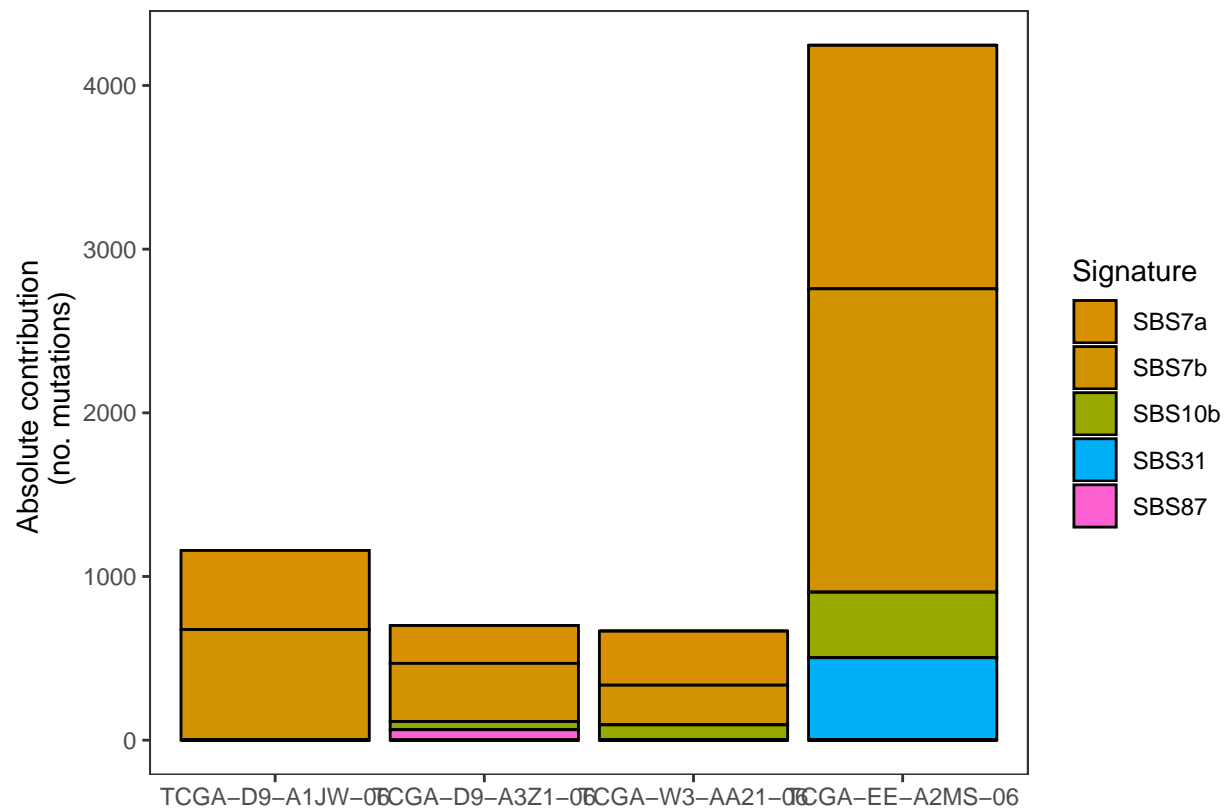


```
plot_contribution(contributions[,samples_to_plot], mode = "relative")
```
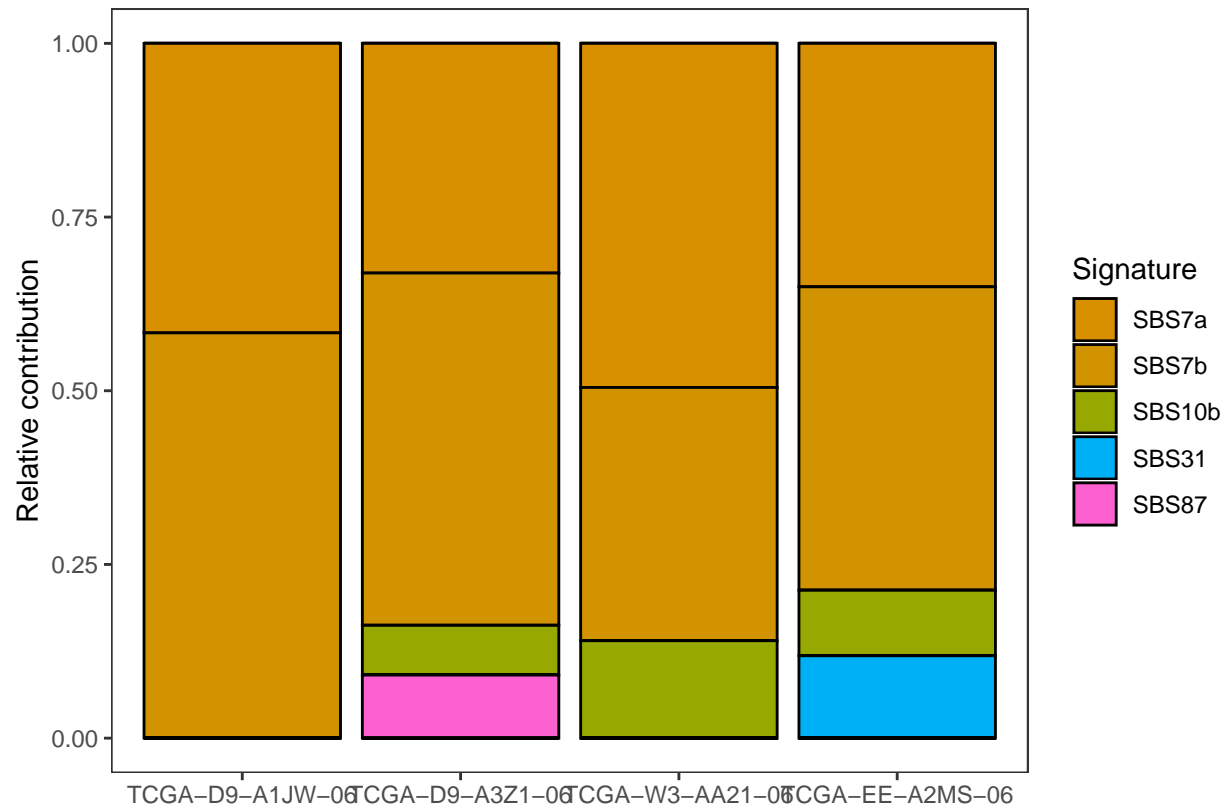
```
plot_contribution_heatmap(contributions, cluster_samples = F)
```

```
plot_contribution(contributions_strict[,samples_to_plot], mode = "absolute")
```
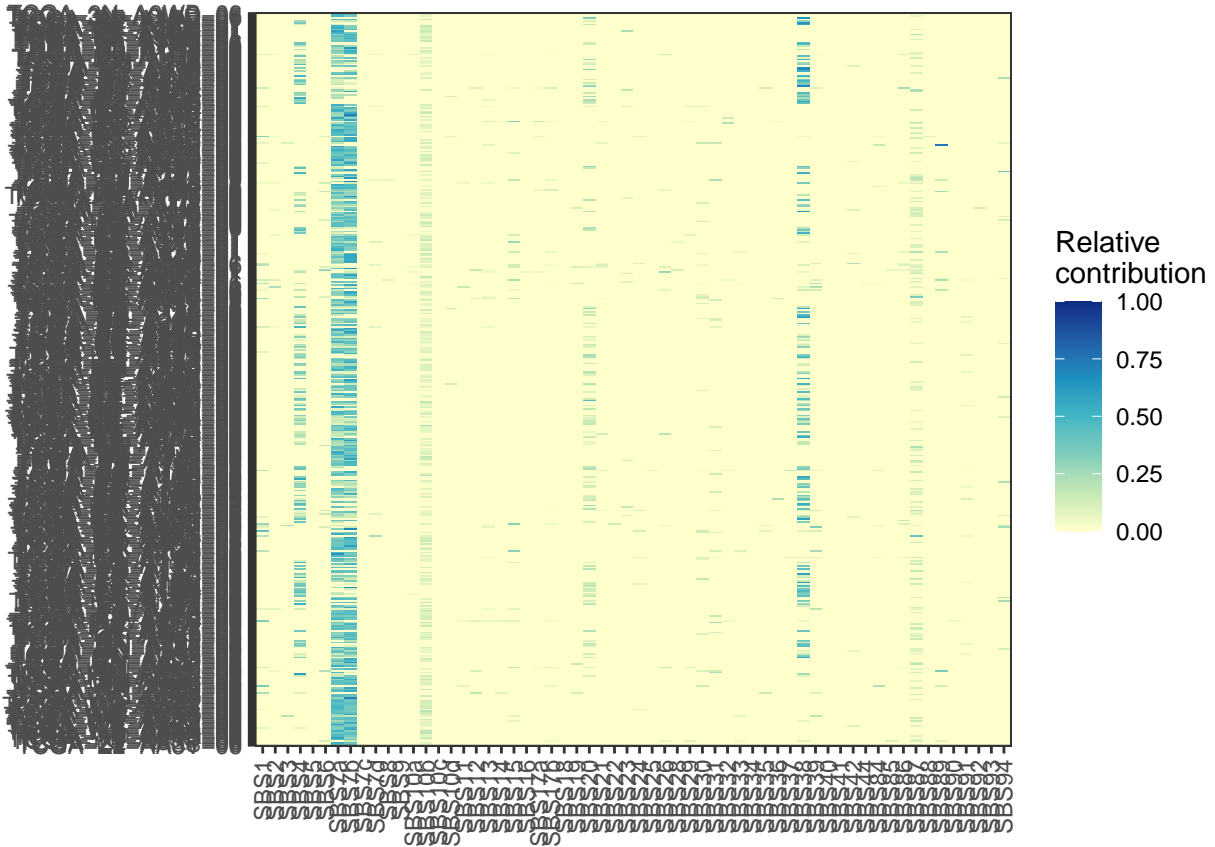
```
plot_contribution(contributions_strict[,samples_to_plot], mode = "relative")
```
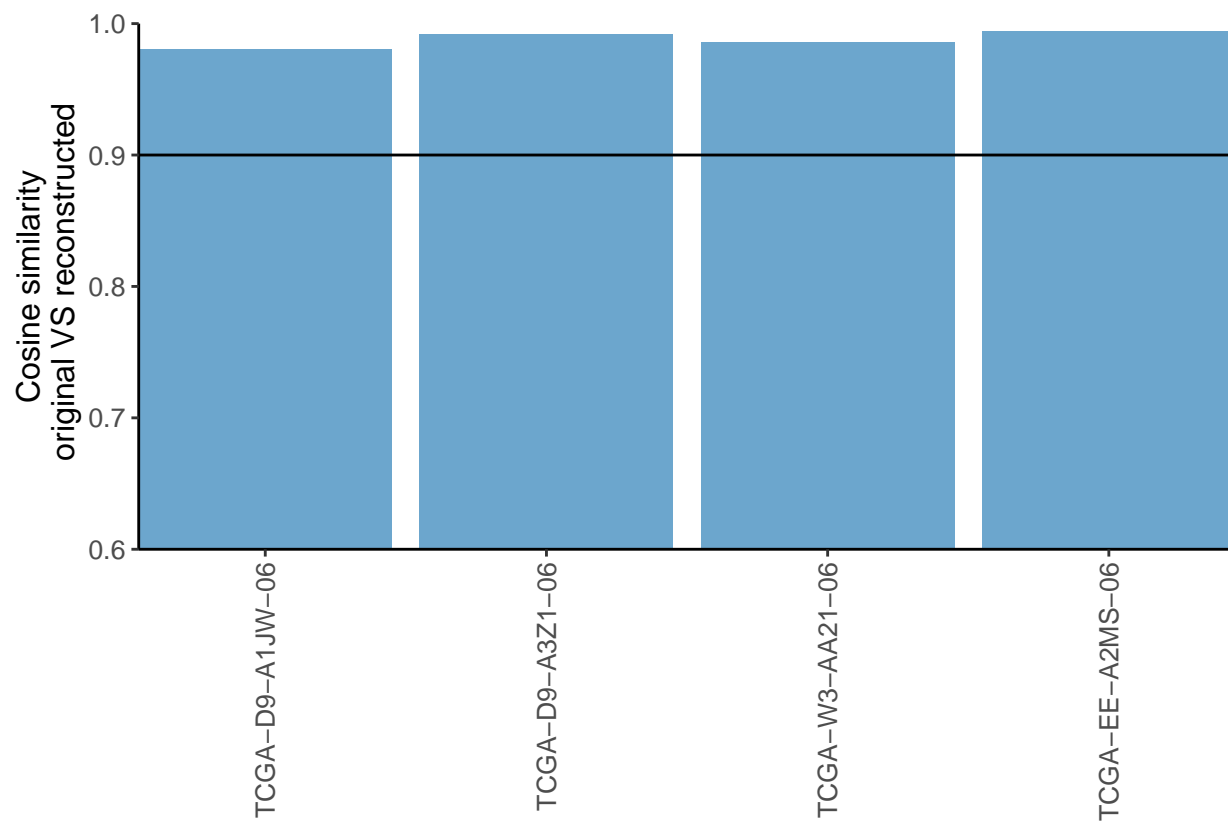
```
plot_contribution_heatmap(contributions_strict, cluster_samples = F)
```
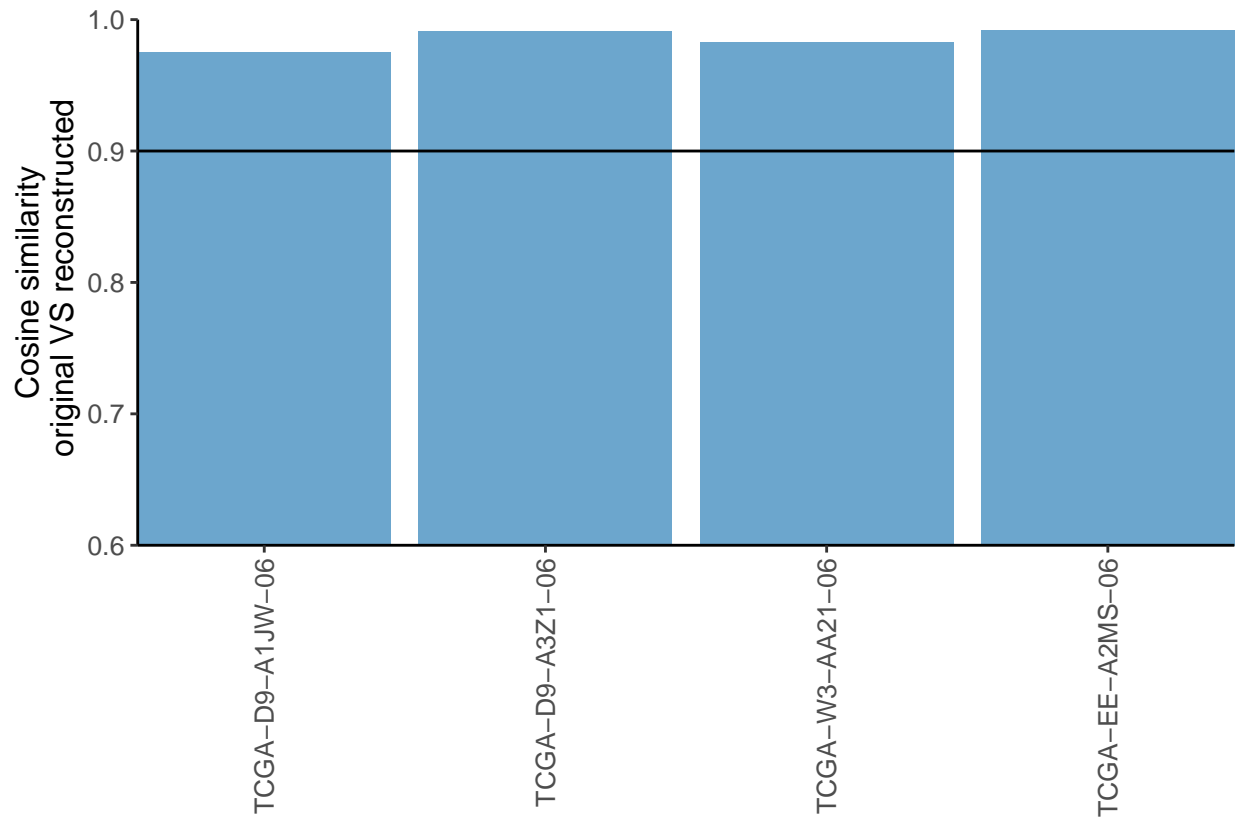
```
set.seed(11111)
samples_to_plot = sample(1:ncol(mm_skcm),4)

plot_original_vs_reconstructed(mm_skcm[,samples_to_plot],
                               fit_res$reconstructed[,samples_to_plot], y_intercept = 0.90)
```

```
plot_original_vs_reconstructed(mm_skcm[,samples_to_plot],
                               fit_res_strict$reconstructed[,samples_to_plot],
y_intercept = 0.90)
```

Discussion #2

Q4. Which is the etiology of the top absolute contributing signature for liver cancer?

Aristolochic Acid Exposure

Q5. Which is the most prominent mutational context for the top contributing signature in skin cancer?

C > T

Q6. The etiology of the top contributing signature for lung cancer corresponds to an endogenous cellular mechanism.

FALSE

Q7. SBS4 is one of the most common signatures found in lung cancer and is associated with tobacco smoking.

TRUE

Q8. SBS7d is one of the most common signatures in skin cancer and is associated with UV light exposure and high numbers of C>T mutations.

FALSE