Practical6

aungmyatphyo(22221184)

2023-03-15

Part1 (1) Mutational signatures are patterns of DNA mutation due to different genetic instability processes (environmental factors such as smoking & carcinogenic viruses or specific biological processes). They can give the hints about the cause of mutation. They are identified by some bioinformatic tools (mutSignatures) in which mutation count matrix is prepared and non-negative matrix factorization (NMF) is carried out to detect the signatures.

2.

```
library(kableExtra)
#install.packages('devtools')
library(devtools)
```

```
## Loading required package: usethis
```

```
#devtools::install_github("dami82/mutSignatures", force = TRUE, build_vignettes = FALSE)
library(mutSignatures)
```

```
## Loading required package: foreach
```

```
library(BiocManager)
```

```
## Bioconductor version '3.15' is out-of-date; the current release version '3.16'
##   is available with R version '4.2'; see https://bioconductor.org/install
```

```
##
## Attaching package: 'BiocManager'
```

```
## The following object is masked from 'package:devtools':
##
##     install
```

```
#BiocManager::install("BSgenome.Hsapiens.UCSC.hg19")
library(BSgenome.Hsapiens.UCSC.hg19)
```

```
## Loading required package: BSgenome
```

```
## Loading required package: BiocGenerics
```

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

```
## The following object is masked from 'package:mutSignatures':
##
##     as.data.frame
```

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

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

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
##
## Attaching package: 'S4Vectors'
```

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

```
## Loading required package: IRanges
```

```
## Loading required package: GenomeInfoDb
```

```
## Loading required package: GenomicRanges
```

```
## Loading required package: Biostrings
```

```
## Loading required package: XVector
```

```
##
## Attaching package: 'Biostrings'
```

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

```
## Loading required package: rtracklayer
```

```
setwd("~/Downloads")
#2a
maf.dset <- read.delim("somatic(1).mafplus", header = TRUE, as.is = TRUE)
#2b
maf.dset <- filterSNV(dataSet = maf.dset,
                      seq_colNames = c("Reference_Allele",
                                       "Tumor_Seq_Allele1",
                                       "Tumor_Seq_Allele2"))
#2c
hg19 <- BSgenome.Hsapiens.UCSC.hg19::BSgenome.Hsapiens.UCSC.hg19

a<-rep('chr',50172)
Chromosome<-paste(a,maf.dset$Chromosome,sep ="")
maf.dset<- maf.dset[,-5]
maf.dset<-cbind(maf.dset[,1:4],Chromosome,maf.dset)
maf.dset<-maf.dset[,-c(6,7,8,9)]
maf.dset <- attachContext(mutData = maf.dset,
                          chr_colName =  "Chromosome",
                          start_colName = "Start_position",
                          end_colName = "End_position",
                          nucl_contextN = 3,
                          BSGenomeDb = hg19)
```

```
## 358 rows were excluded from analysis
```

```
## Removing out-of-bounds positions...  0 records were removed.
## Done!
```

```
#2d
maf.dset <- removeMismatchMut(mutData = maf.dset,
                              refMut_colName = "Reference_Allele",
                              context_colName = "context",
                              refMut_format = "N")
#2e
maf.dset <- attachMutType(mutData = maf.dset,
                          ref_colName = "Reference_Allele",
                          var_colName = "Tumor_Seq_Allele1",
                          var2_colName = "Tumor_Seq_Allele2",
                          context_colName = "context")
```

```
## Assigning mutation types ..................... Done!
## Now applying RevCompl transformation. Done!
## Final formatting. Done!
```

```
#2f
maf.dset$TCGAid <- substr(maf.dset$Tumor_Sample_Barcode, 1, 15)
maf.counts <- countMutTypes(mutTable = maf.dset, sample_colName = "TCGAid",
                            mutType_colName = "mutType")
maf.counts
```

```
##  Mutation Counts object - mutSignatures
##
##  Total num of MutTypes: 96
##  MutTypes: A[C>A]A, A[C>A]C, A[C>A]G, A[C>A]T, A[C>G]A ...
##
##  Total num of Samples: 9
##  Sample Names: TCRBOA5-T-WEX, TCRBOA4-T-WEX, TCRBOA7-T-WEX, TCRBOA7-T-WGS, TCRBOA6-T-WEX ...
```

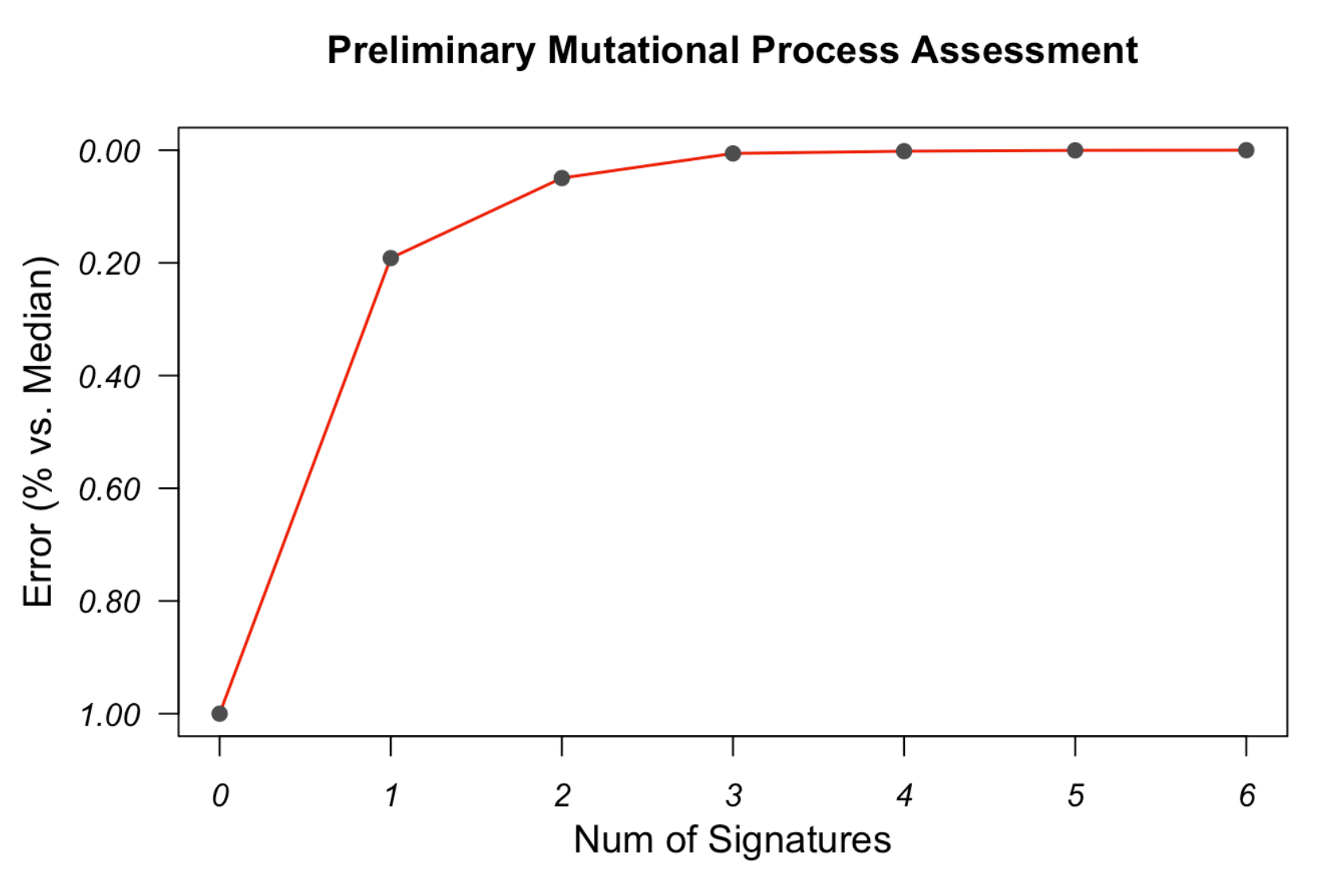
(3)In the method of NMF,researchers construct the matrix with the rows (96 possible 3 nucleotide context) and columns (SNVs in the set of sample). Then the matrix factorized into 2 matrices(one with mutational signatures and other describes the relationship of each signature to each sample).

Total number of mutation types depend on the specific mutations observed in the samples.

#part2

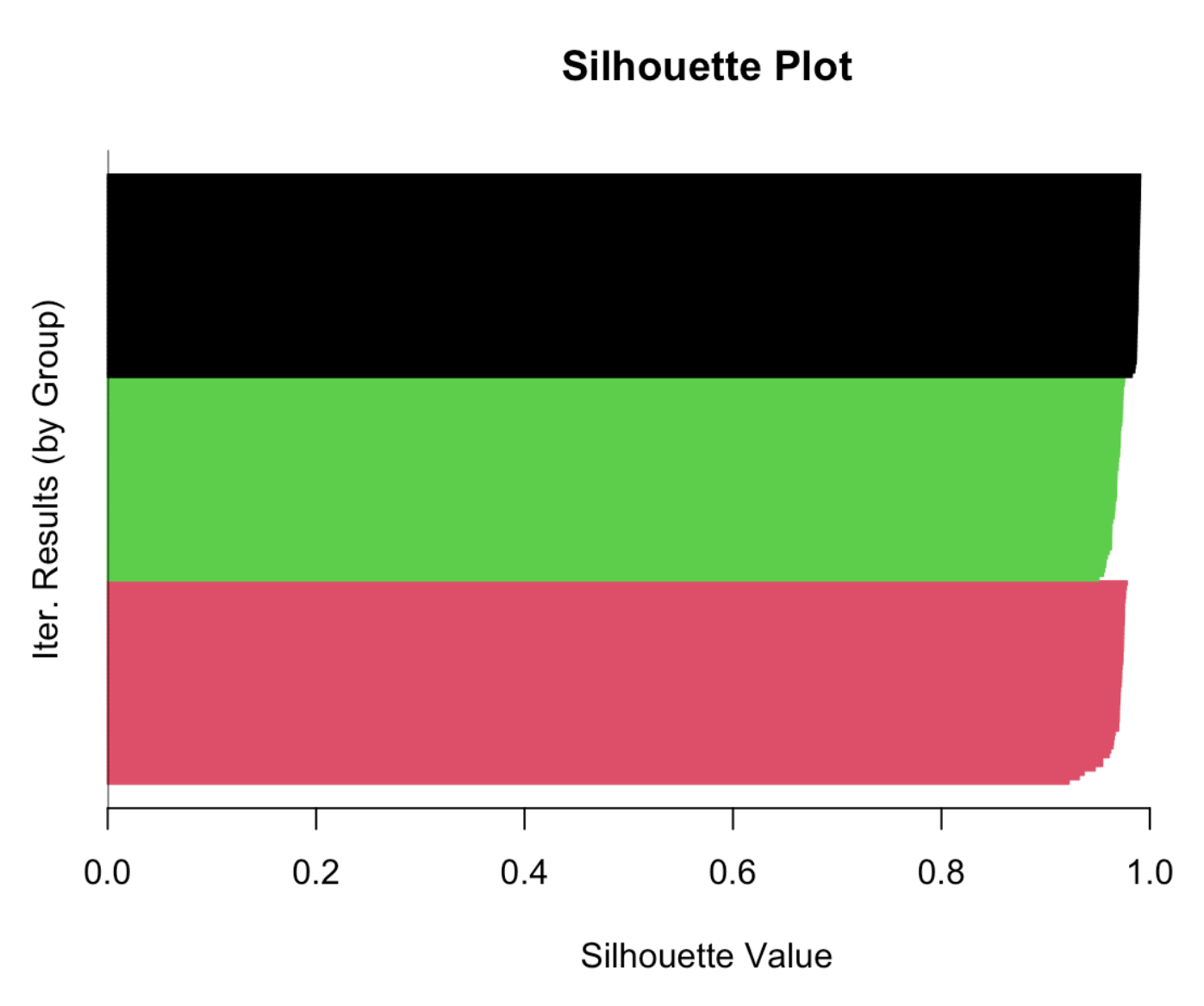
```
mouCancer.assess <- prelimProcessAssess(input = maf.counts, approach = "counts")
```

```
## Preliminary Mutational Process Assessment: ......
```



```
maf.counts.params<-setMutClusterParams(num_processesToExtract = 3,
                                       approach = 'counts',
                                       num_totIterations = 50,
                                       num_parallelCores = 1,
                                       debug = FALSE,
                                       algorithm = 'alexa'
                                       )
```

```
maf.analysis<-decipherMutationalProcesses(input = maf.counts,
                                          params = maf.counts.params)
```



The silhouette plot tell that the chosen number of parameters is appropriate.

```
maf.signs<-maf.analysis$Results$signatures
print(maf.signs)
```

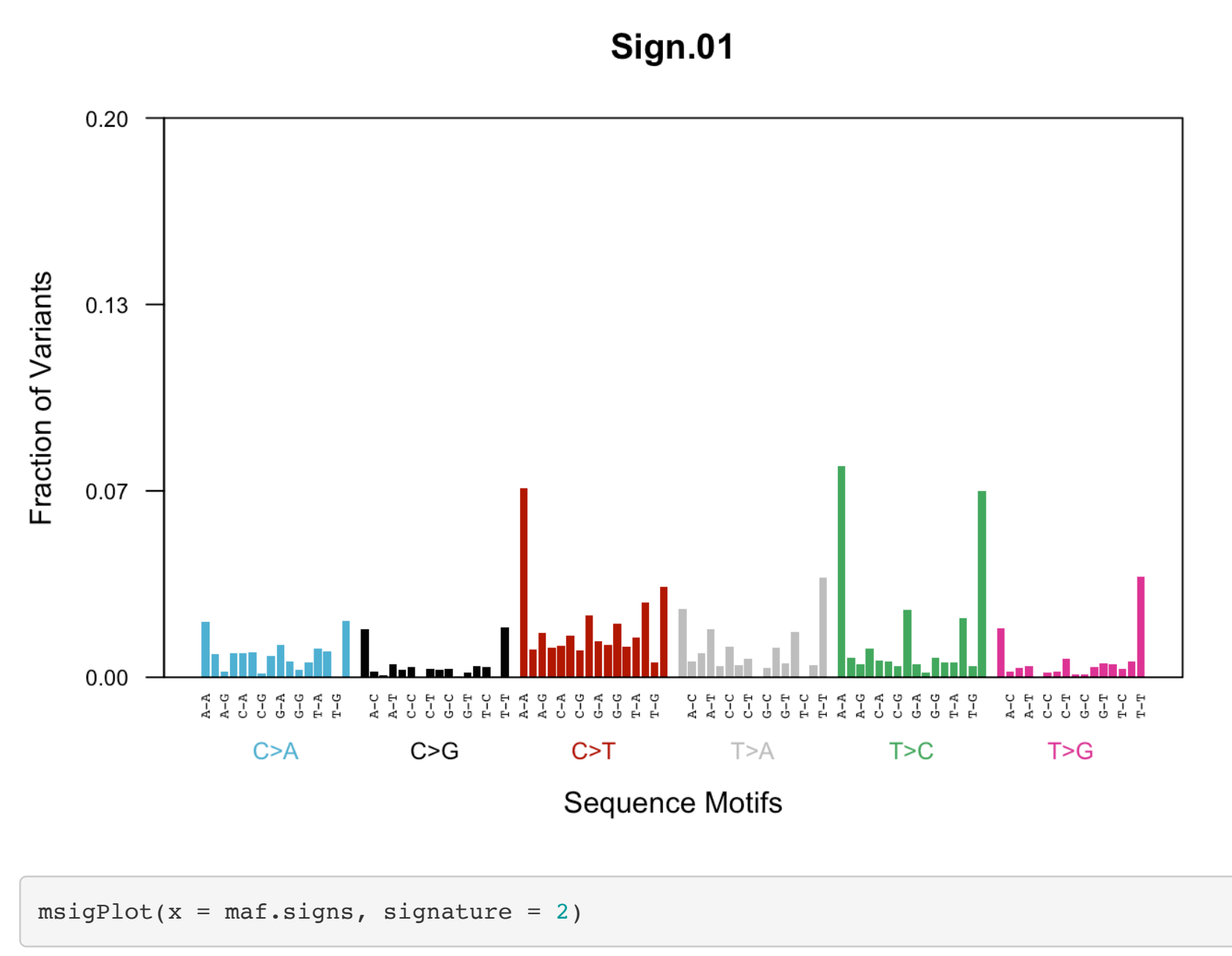
```
##  Mutation Signatures object - mutSignatures
##
##  Total num of Signatures: 3
##  Total num of MutTypes: 96
##
##     Sign.1   Sign.2   Sign.3
##     ------   ------   ------
##   + 0.0199   0.0054   0.0029   +  A[C>A]A
##   + 0.0082   0.0016   0.0044   +  A[C>A]C
##   + 0.0020   0.0052   0.0042   +  A[C>A]G
##   + 0.0085   0.0028   0.0025   +  A[C>A]T
##   + 0.0170   0.0189   0.0100   +  A[C>G]A
##   + 0.0022   0.0119   0.0069   +  A[C>G]C
##   + 0.0008   0.0085   0.0090   +  A[C>G]G
##   + 0.0045   0.0160   0.0068   +  A[C>G]T
##   + 0.0678   0.0138   0.0062   +  A[C>T]A
##   + 0.0100   0.0077   0.0097   +  A[C>T]C
##     ......   ......   ......
```

```
maf.expos<- maf.analysis$Results$exposures
maf.expos
```

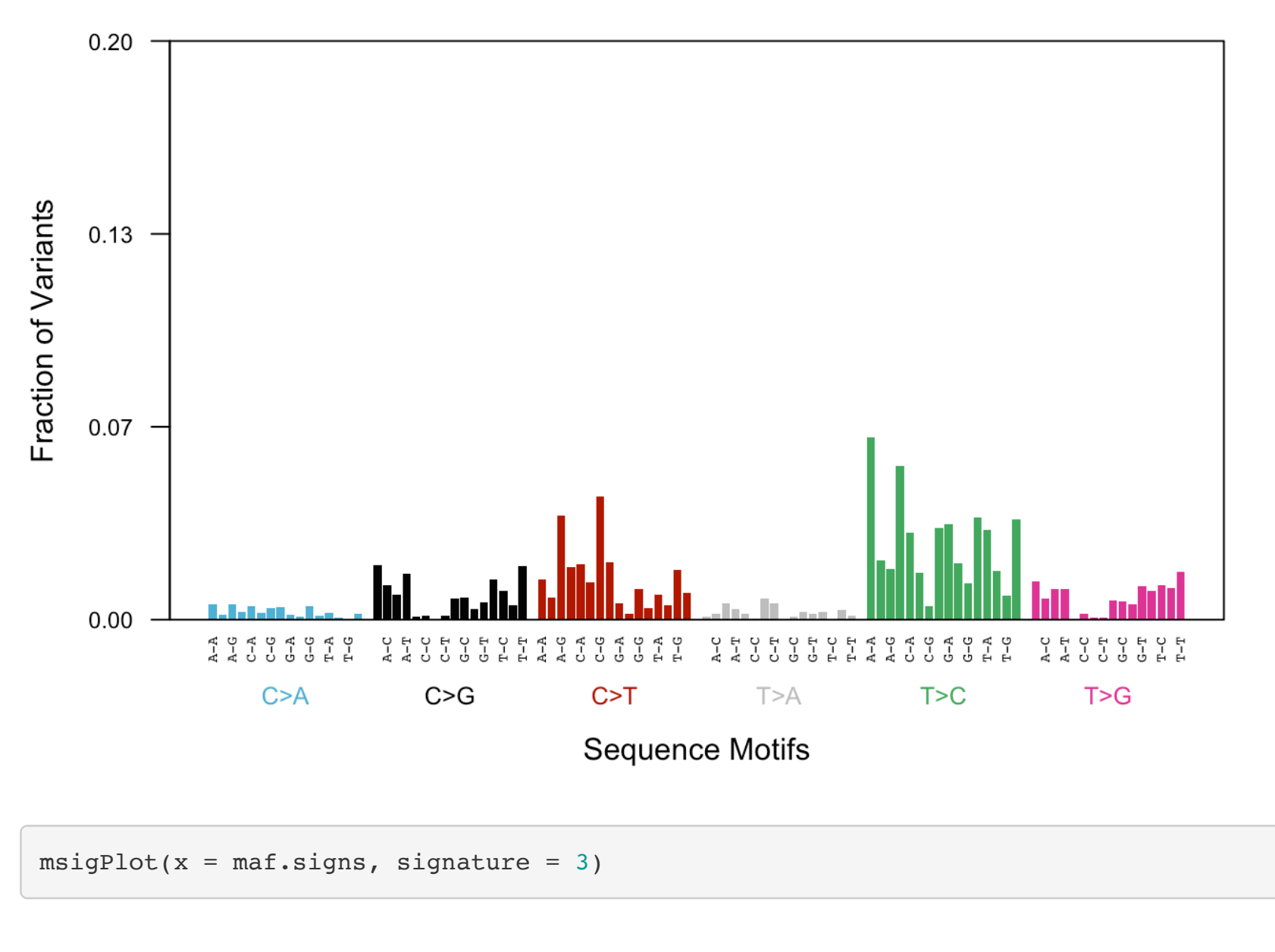
```
##  MutSignature Exposures object - mutSignatures
##
##  Total num of Samples: 9
##  Total num of Signatures: 3  { first 3 signatures are displayed }
##  Signature names:  Sign.01, Sign.02, Sign.03
##
##     Sign.1   Sign.2   Sign.3
##     ------   ------   ------
##   +     99      677      935   +  TCRBOA5-T-WEX
##   +    154      538     >999   +  TCRBOA4-T-WEX
##   +   >999     >999     >999   +  TCRBOA7-T-WEX
##   +   >999      719      116   +  TCRBOA7-T-WGS
##   +     57      811     >999   +  TCRBOA6-T-WEX
##   +   >999      456      349   +  TCRBOA6-T-WGS
##   +     56     >999      755   +  TCRBOA1-T-WEX
##   +     35      130      113   +  TCRBOA3-T-WEX
##   +    173      882      985   +  TCRBOA2-T-WEX
```

#part3

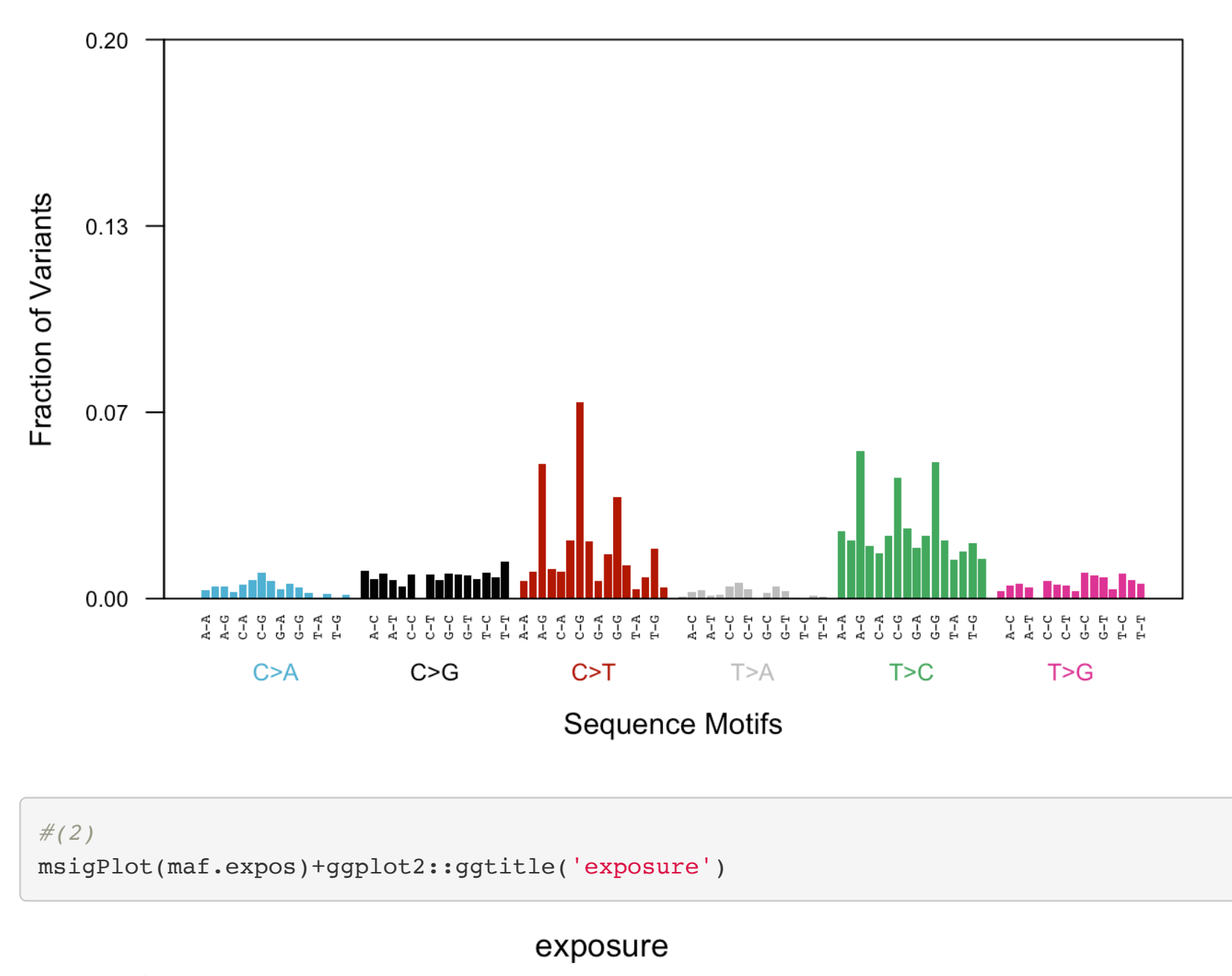
```
#(1)
msigPlot(x = maf.signs, signature = 1)
```



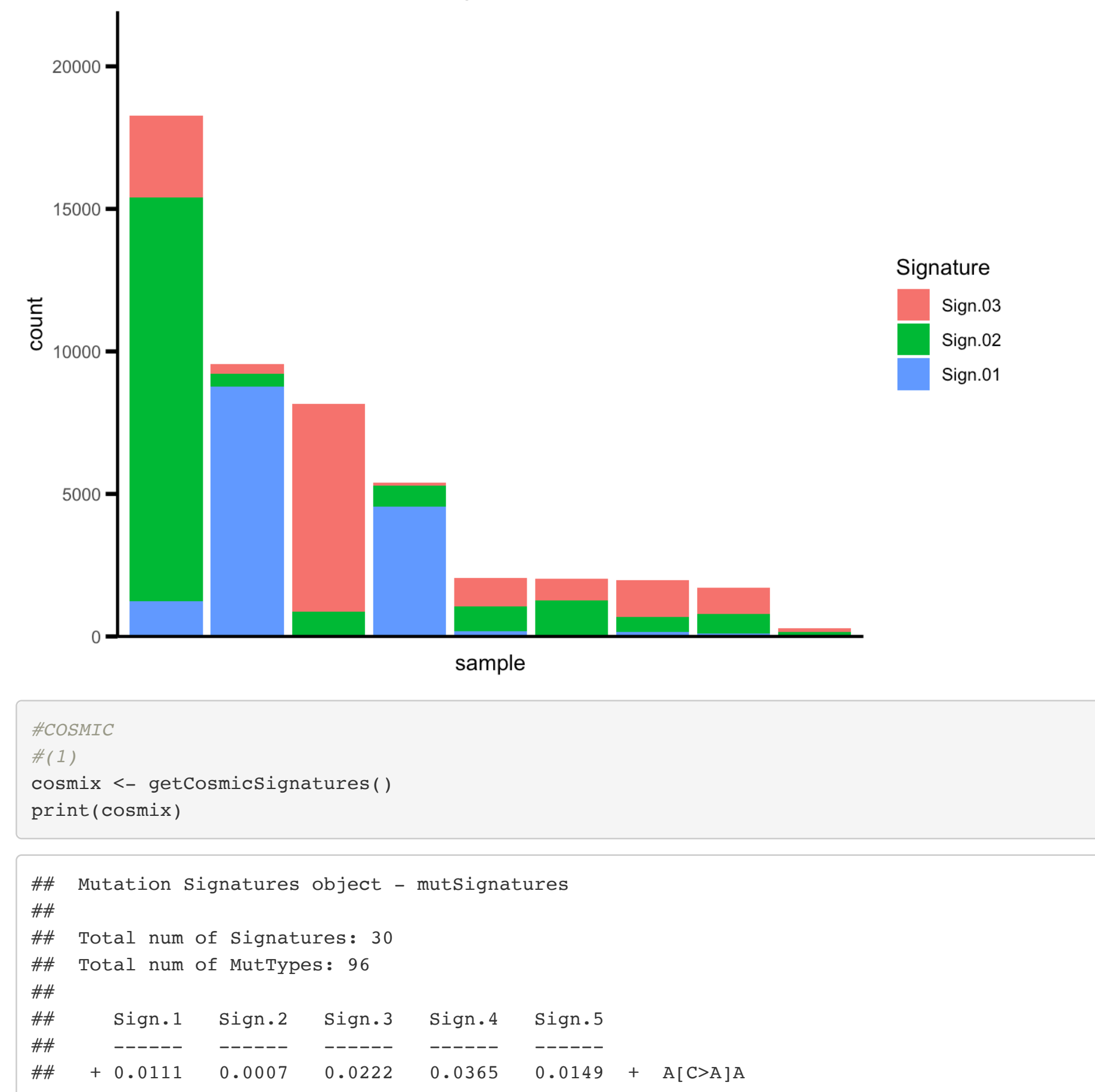
```
msigPlot(x = maf.signs, signature = 2)
```



```
msigPlot(x = maf.signs, signature = 3)
```



```
#(2)
msigPlot(maf.expos)+ggplot2::ggtitle('exposure')
```



```
#COSMIC
#(1)
cosmix <- getCosmicSignatures()
print(cosmix)
```

```
##  Mutation Signatures object - mutSignatures
##
##  Total num of Signatures: 30
##  Total num of MutTypes: 96
##
##     Sign.1   Sign.2   Sign.3   Sign.4   Sign.5
##     ------   ------   ------   ------   ------
##   + 0.0111   0.0007   0.0222   0.0365   0.0149   +  A[C>A]A
##   + 0.0091   0.0006   0.0179   0.0309   0.0090   +  A[C>A]C
##   + 0.0015   0.0001   0.0021   0.0183   0.0022   +  A[C>A]G
##   + 0.0062   0.0003   0.0163   0.0243   0.0092   +  A[C>A]T
##   + 0.0018   0.0003   0.0240   0.0097   0.0117   +  A[C>G]A
##   + 0.0026   0.0003   0.0122   0.0054   0.0073   +  A[C>G]C
##   + 0.0006   0.0002   0.0053   0.0031   0.0023   +  A[C>G]G
##   + 0.0030   0.0006   0.0233   0.0054   0.0117   +  A[C>G]T
##   + 0.0295   0.0074   0.0179   0.0120   0.0218   +  A[C>T]A
##   + 0.0143   0.0027   0.0089   0.0075   0.0128   +  A[C>T]C
##     ......   ......   ......   ......   ......
```

```
# 30 signatures are loaded
hm<-matchSignatures(mutSign = maf.signs,reference = cosmix)
print(hm$plot)
```

(3)Active signature 2 is similar with Cosmic 16 and Cosmic 5 while signature 1 is similar with Cosmic 5. For these Cosmic Signatures, we need to know about the Cosmic 16 and 5 and then we will receive the clues for the potential causes and associations of active signatures 1 & 2.