

Package **CompSign**

Lena Morrill

October 2022

CompSign is a toolkit for differential abundance analysis of mutational signatures using a mixed effects Dirichlet-multinomial model (or simpler variations). The compositional nature of mutational signature exposures has often been overlooked but has important implications, as the analyses must be done in relative terms.

Contents

1	Installation	1
2	Datasets	1

1 Installation

CompSign can be installed as usual from github:

```
library(devtools)
devtools::install_github("lm687/CompSign")
```

```
library(CompSign)
library(gridExtra)
library(TMB)

## Warning: package 'TMB' was built under R version 4.0.5
```

2 Datasets

```
## if the folder data/ is not in github
for(i in list.files("../inst/extdata/", pattern = "*RDA", full.names = TRUE)){load(i)}
```

The package contains the following datasets of exposures of mutational signatures and metadata of the corresponding samples. These datasets are:

- **PancEndocrine_signaturesMSE**: Signature exposures for early and late mutations, in the PCAWG Panc-Endocrine cohort
- **ProstAdenoCA_chrom**: Signature exposures for each chromosome, in the PCAWG Prost-AdenoCA cohort

PancEndocrine_signaturesMSE is an object of class sign

```
PancEndocrine_signaturesMSE = load_PCAWG("../inst/extdata/roo/Panc-Endocrine_signaturesMSE_R00.RDS",
                                         read_directly = T,
                                         typedata = "signaturesMSE", override_warning_X_Z = T)

## [1] "../inst/extdata/roo/Panc-Endocrine_signaturesMSE_R00.RDS"
## Reading file ../inst/extdata/roo/Panc-Endocrine_signaturesMSE_R00.RDS

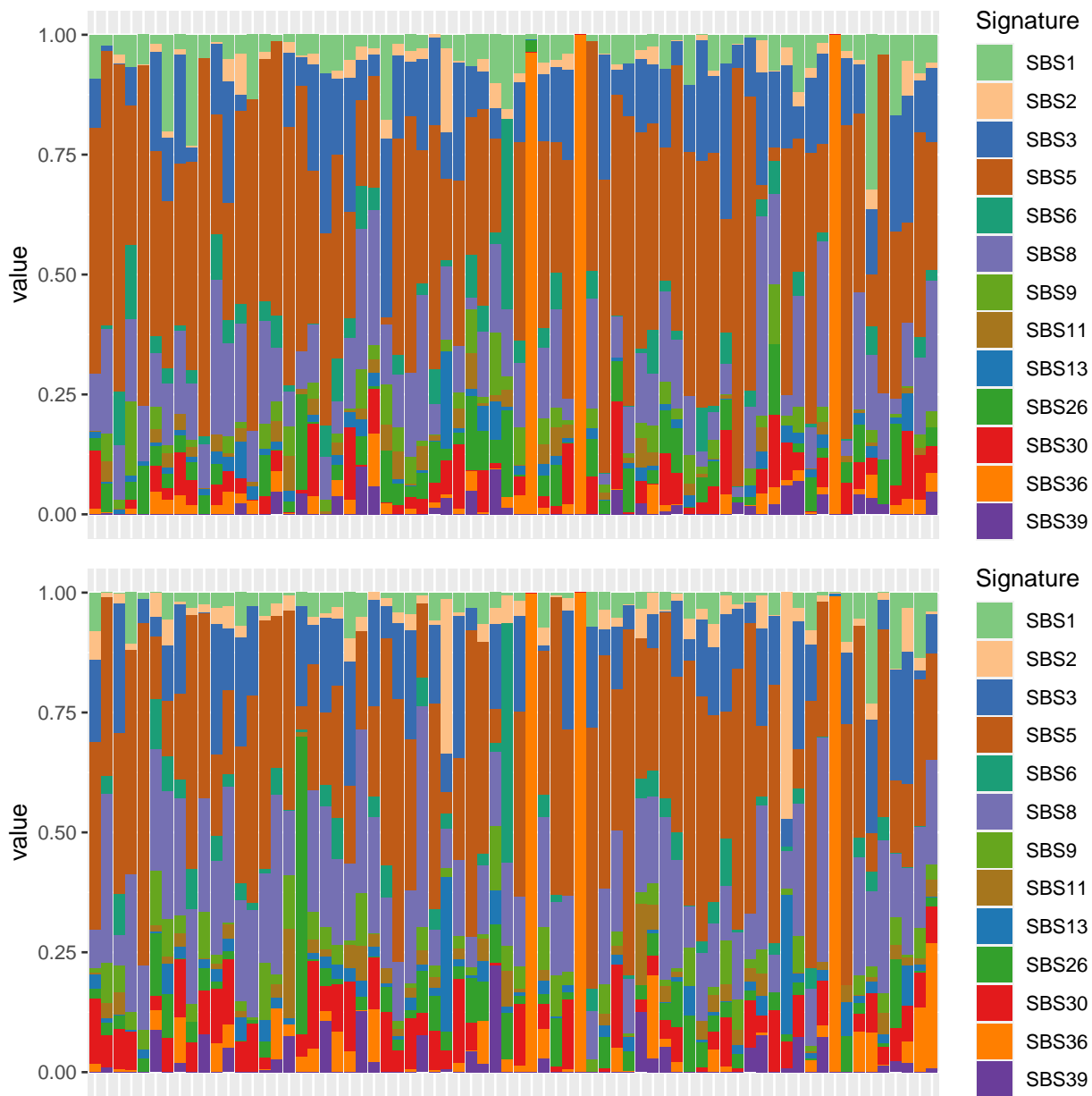
PancEndocrine_signaturesMSE_v2 = load_PCAWG(ct = "Panc-Endocrine", typedata = "signaturesMSE", path_to_da
```

```
# PancEndocrine_signaturesMSE
```

```
do.call('grid.arrange', lapply(split_matrix_in_half(PancEndocrine_signaturesMSE$Y), function(i) createBar

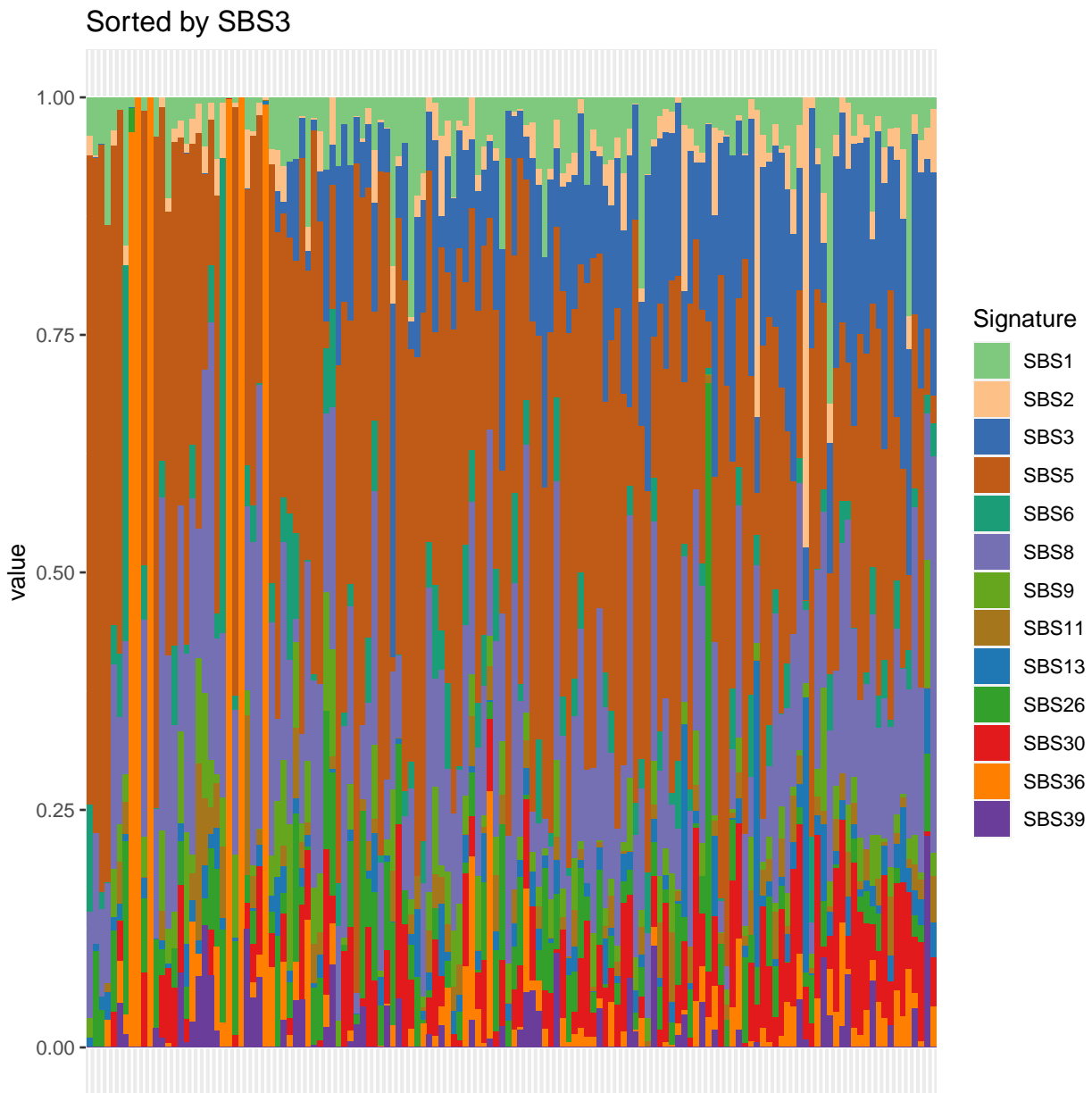
## Loading required package: reshape2
## Loading required package: ggplot2
## Warning: package 'RColorBrewer' was built under R version 4.0.5

## Creating plot... it might take some time if the data are large. Number of samples: 70
## Creating plot... it might take some time if the data are large. Number of samples: 70
```



```
createBarplot(normalise_rw(non_duplicated_rows(PancEndocrine_signaturesMSE$Y)),
              order_labels = names(sort(non_duplicated_rows(PancEndocrine_signaturesMSE$Y)[,'SBS3'],
              decreasing = F)), remove_labels=T)+ggtitle('Sorted by SBS3')
```

```
## Creating plot... it might take some time if the data are large. Number of samples: 140
```



```
fullDM_no_small_sigs <- wrapper_run_TMB(object = give_subset_sigs_TMBobj(PancEndocrine_signaturesMSE,
  sigs_to_remove = c('SBS13', 'SBS17a', 'SBS17b', 'SBS30')),
  model = "fullRE_DM", use_nllminb=T, smart_init_vals=F)

## Error in .Call("getParameterOrder", data, parameters, new.env(), NULL, : "getParameterOrder"
not available for .Call() for package "fullRE_ME_dirichletmultinomial"
```