User manual for the signatureFit command line interface

signature.tools.lib version: 2.1.0
latest edit: 24/02/2022

Andrea Degasperi, University of Cambridge, UK
ad923@cam.ac.uk

1. Introduction

Mutational signature fit analysis attempts to identify the presence of a given set of mutational signatures in the somatic mutations of a cancer sample.

This document describes how to use the signatureFit command line script, which is a wrapper for the signatureFit_pipeline function in the signature.tools.lib R packages, which in turn is an interface for the Fit and FitMS mutational signature fit analysis functions.

The signatureFit_pipeline function is a flexible interface for mutational signature fit analysis. Users can provide mutation calls as input or a pre-built mutational catalogue, and then they can perform either an automated analysis, with very few options required such as the organ of origin of the sample, or use the options to perform a more tailored analysis.

Note that currently not all the options of signatureFit_pipeline are available in the signatureFit command line script, though we aim to have them all available as soon as possible.

2. Installation

The script signatureFit is included in the signature.tools.lib R package. Thus, in order to use it, one is required to install signature.tools.lib, which is available on GitHub:

https://github.com/Nik-Zainal-Group/signature.tools.lib

After the installation of signature.tools.lib, one can run the signatureFit script, which is located in the scripts folder in the github repository. For easy access, add a copy of the signatureFit script to a location in your command line PATH.

3. signatureFit options

The list of available options can be accessed by typing:

```
signatureFit --help
```

This is the current output:

```
This script runs the signature fit pipeline of the R package
signatures.tools.lib, using the Fit or FitMS functions.

Run this script as follows:

signatureFit [OPTIONS]

Available options:
  -o, --outdir=DIR          Name of the output directory. If omitted a name
                              will be given automatically.
  -b, --bootstrap          Request signature fit with bootstrap
  -x, --snvvcf=SNVVCF      SNVVCF is a tab separated file containing two
                              columns. The first column contains the sample
                              names, while the second column contains the
                              corresponding SNV vcf file names.
  -X, --snvtab=SNVTAB      SNVTAB is a tab separated file containing two
                              columns. The first column contains the sample
                              names, while the second column contains the
                              corresponding SNV tab file names. Each SNV tab
                              file should have a header with the following
                              columns: chr, position, REF, ALT.
  -s, --sigversion=SIGVERSION
                            Either COSMICv2, COSMICv3.2, RefSigv1 or RefSigv2.
                              If not specified SIGVERSION=RefSigv2.
  -O, --organ=ORGAN        When using RefSigv1 or RefSigv2 as SIGVERSION,
                              organ-specific signatures will be used.
                              If SIGVERSION is COSMICv2 or COSMICv3.2, then a
                              selection of signatures found in the given organ
                              will be used. Organ names depend on the selected
                              SIGVERSION. For RefSigv1 or RefSigv2: Biliary,
                              Bladder, Bone_SoftTissue, Breast, Cervix (v1 only),
                              CNS, Colorectal, Esophagus, Head_neck, Kidney,
                              Liver, Lung, Lymphoid, NET (v2 only),
                              Oral_Oropharyngeal (v2 only), Ovary, Pancreas,
                              Prostate, Skin, Stomach, Uterus.
  -l, --signames=SIGNAMES  If no ORGAN is specified, SIGNAMES can be used to
                              provide a comma separated list of signature names
                              to select from the COSMIC or reference signatures,
                              depending on the SIGVERSION requested. For example,
                              for COSMICv3.2 use: SBS1,SBS2,SBS3.
  -e, --genomev=GENOMEV    Genome version to be used: hg19, hg38 or mm10.
                              If not specified GENOMEV=hg19.
  -m, --fitmethod=FITMETHOD Either Fit or FitMS. If not specified FITMETHOD=FitMS
  -n, --nparallel=NPARALLEL Number of parallel CPUs to be used.
  -f, --nboot=NBOOT        Number of bootstrap to be used when bootstrap is
                              requested (-b), if not specified, NBOOT=200.
  -r, --randomSeed=SEED    Specify a random seed to obtain always the same
                              identical results.
  -h, --help               Show this explanation.
```

4. Examples

4.1 FitMS example

In this example we run a signature fit analysis with FitMS, using vcf single nucleotide variant (SNV) files as input, assuming that the vcf are obtained from breast cancer samples.

```
signatureFit --organ Breast -b -o outfolder -x snvvcf.tsv
```

Note that FitMS is the default fit method, so there is no need to specify –fitmethod=FitMS. FitMS will use the latest RefSigv2 signatures, which include the common and rare SBS signatures identified in the analysis of the Genomics England WGS cancer dataset. The flag -b requests a bootstrap analysis, -o indicates the output folder, and -x indicates the location of a tab separated file containing a list of sample names and corresponding vcf locations. The content of snvvcf.tsv could be as follows:

```
Sample1   sample1_snv.vcf
Sample2   sample2_snv.vcf
```

```
   Sample3   sample3_snv.vcf
   ...
```

Finally, note that all the mutations in the input vcf files will be used, so they should already be filtered, e.g. containing only PASS variants.

4.1 Fitting COSMIC v3.2 signatures

In this example we show how to fit a specific set of SBS COSMIC signatures from the COSMIC v3.2 set.

```
   signatureFit -b -o outfolder -x snvvcf.tsv -s COSMICv3.2 -m Fit -l
SBS1,SBS2,SBS3,SBS5,SBS8,SBS13,SBS17a,SBS17b,SBS18
```

In this case we did not specify an organ, but rather used the -s option to request the use of COSMIC v3.2 signatures and the -l option to supply a list of signatures to use. We also specified to use the Fit algorithm with the option -m.