User manual for the solutionSelectionForFitMS command line script

signature.tools.lib version: 2.4.1
latest edit: 15/05/2023

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.

The `solutionSelectionForFitMS` script described in this document assumes that mutational signature analysis has already been performed using FitMS and that the FitMS results are stored into an rData file. Such rData file can be obtained from the `signatureFit` command line script, or in R using the `saveFitToFile` function.

When running FitMS, it is possible that for each sample there exist multiple alternative solutions, where different rare signatures might explain the sample mutations equally well. In this case, FitMS would normally select the best solution according to the MaxCosSim criteria, that is select the solution that produces the maximum cosine similarity between the input mutational catalogue and the reconstructed catalogue obtained as a linear combination of the fitted signatures.

The `solutionSelectionForFitMS` script can be used to change the selection criteria to MinError, thus selecting the solution that produces the minimum error, computed as sum of absolute deviations, between the input mutational catalogue and the reconstructed catalogue obtained as a linear combination of the fitted signatures.

The `solutionSelectionForFitMS` script also allows to select solutions manually. These must be rare signature solutions already identified and stored in the input rData file, or, alternatively, the common signatures only solution can also be specified.

2. Installation

The script `solutionSelectionForFitMS` 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 `solutionSelectionForFitMS` script, which is located in the scripts folder in the github repository. For easy access, add a copy of, or a symbolic link to, the `solutionSelectionForFitMS` script to a location in your command line PATH.

3. solutionSelectionForFitMS options

The list of available options can be accessed by typing:

```
solutionSelectionForFitMS --help
```

This is the current output:

```
This script selects FitMS solutions according to a given selection criteria,
and also allows for manual selection of solutions.

Run this script as follows:

solutionSelectionForFitMS [OPTIONS]

Available options:
  -i, --infile=IFILE        Name of the input rData file. The required IFILE is
                              an rData file containing an R object obtained from
                              running FitMS and saved using the saveFitToFile
                              function, or obtained using the signatureFit
                              command line script.
  -o, --outdir=OUTDIR       Name of the output directory. This directory will
                              contain the new plots, as well as a fitData.rData file
                              with the updated object loaded from IFILE.
                              If omitted OUTDIR=solutionSelectionOutput.
  -t, --selectiontable=STABLE
                            Name of tab separated file. There should be no column
                              names. The first column should contain the names
                              of the samples where the solution should be manually
                              changed, while the second column should contain the
                              name of the solution to use, which could be the name
                              of a signature, or a list of signature names separated
                              by the : character, or the text "common".
  -c, --selectioncriteria=SELCRIT
                            SELCRIT can be either MaxCosSim or MinError, default is
                              MaxCosSim.
  -J, --writejson           Write the updated signature fit object to a JSON file,
                              which will save all fit results and options used.
  -h, --help                Show this explanation.
```

4. Examples

4.1 Manual selection of a different rare signature solution

Let us assume we are running FitMS on a cohort of breast cancer samples, using the `signatureFit` command line script:

```
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.

Now, let us assume that we want to change the rare signature solution selection for Sample2, because multiple rare signatures solutions were found, and we have reasons to believe that the correct solution is among the alternative solutions. We can use the `solutionSelectionForFitMS` script for this. To pass this manual selection to the script, we need to prepare a tab separated file, where sample names are in the first column and solution names are in the second column. For example, if the solution we want for Sample2 is signature SBS44, we need to create a tab separated file containing the following:

```
   Sample2   SBS44
```

We can save that file as solutionSelection.tsv, and then run the `solutionSelectionForFitMS` script as follows:

```
   solutionSelectionForFitMS -i outfolder/fitData.rData -o outfolder_new -t
solutionSelection.tsv
```

The above command line will reorganise the selected solutions in the fitData.rData object and replot everything in the outfolder_new folder.

4.2 Manual selection of the common signatures only solution

Following from the example above, let us assume that a rare signature was identified for Sample3, but that we have reasons to believe this is a false positive assignment, and that the solution including only the common signatures is more likely to be the correct one.

False positive assignment of rare signatures by FitMS can be resolved in many ways, before resorting to manual selection. For example, one can curate a set of mutational signatures, removing rare signatures that are unlikely to be present in a dataset. Alternatively, one can tune false positives by changing the FitMS parameters, thus increasing specificity. If these still fail and some samples require a manual selection, we can write a solutionSelection.tsv file where we can use the keyword “common” to specify the common signatures only solution. For example, to change the solution of Sample3 to common signatures only, along with changing Sample2 to SBS44 from the previous example, we can write the solutionSelection.tsv file as follows:

```
   Sample2   SBS44
   Sample3   common
```

And once again run the `solutionSelectionForFitMS` script as follows:

```
   solutionSelectionForFitMS -i outfolder/fitData.rData -o outfolder_new -t
solutionSelection.tsv
```