User manual for the hrDetect command line interface

signature.tools.lib version: 2.1.1
latest edit: 06/03/2022

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

1. Introduction

HRDetect is a classifier for homologous recombination deficiency in tumour samples that uses mutational signatures as input features.

This document describes how to use the hrDetect command line script, which is a wrapper for the HRDetect_pipeline function in the signature.tools.lib R package.

The HRDetect_pipeline function is a flexible interface for the HRDetect classifier. HRDetect scores can be computed directly from input features, or the input features can be computed by the function providing the somatic mutations of the tumours, including single nucleotide variants, indels, copy number variants and structural variants.

2. Installation

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

3. hrDetect options

The list of available options can be accessed by typing:

```
hrDetect --help
```

This is the current output:

```
This script runs the HRDetect pipeline of the signature.tools.lib R package.

Run this script as follows:

hrDetect [OPTIONS]

Available options:
  -i, --input=INPUTTABLE    Tab separate input table with the list of files for
                              each sample. Columns of INPUTTABLE should be:
                              sample, SNV_vcf_files, SNV_tab_files,
                              Indels_vcf_files, Indels_tab_files, CNV_tab_files,
                              SV_bedpe_files. Note that only one column of
                              SNV_vcf_files and SNV_tab_files is necessary
```

```
-o, --outdir=OUTDIR       Name of the output directory. If omitted a name will
                            be given automatically.
-O, --organ=ORGAN         When using RefSigv1 or RefSigv2 as SNVSV or SVSV,
                            organ-specific signatures will be used.
                            If SNVSV is COSMICv2 or COSMICv3.2, then a
                            selection of signatures found in the given organ
                            will be used. Available organs depend on the selected
                            SNVSV and SVSV. 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.
-s, --snvsigversion=SNVSV Either COSMICv2, COSMICv3.2, RefSigv1 or RefSigv2.
                            When SNVSV=RefSigv2 and an organ is specified,
                            signature fit for SNVs will be performed with FitMS
-S, --svsigversion=SVSV   Currently only RefSigv1 is available for SV signatures
-l, --snvsignames=SNVSN   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.
-L, --svsignames=SVSN     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.
-b, --bootstrap           Request HRDetect with bootstrap
-e, --genomev=GENOMEV     Genome version to be used: hg19, hg38 or mm10. If not
                            specified GENOMEV=hg19.
-n, --nparallel=NPARALLEL Number of parallel CPUs to be used
-f, --nbootFit=NBOOTFIT   Number of bootstrap to be used in signature fit. If
                            not specified NBOOTFIT=100.
-r, --randomSeed=SEED     Specify a random seed to obtain always the same
                            identical results.
-h, --help                Show this explanation.
```

4. Examples

4.1 Using organ-specific mutational signatures and bootstrap HRDetect

In this example, we compute the HRDetect score for two breast cancer samples using mutation files, and request HRDetect with boostrap.

```
hrDetect -O Breast -b -o outfolder -i inputTable.tsv
```

Using the -O option to specify an organ, the HRDetect pipeline will use organ-specific mutational signatures for estimating SNV and SV signature exposures.

Note that FitMS is the default fit method for SNV organ-specific signatures, while the SV organ-specific signatures are fitted with the simpler Fit function. The flag -b requests a bootstrap HRDetect score, -o indicates the output folder, and -i indicates the location of a tab separated file containing a list of sample names and corresponding mutation files locations. The content of inputTable.tsv could be as follows:

```
sample    SNV_vcf_files     Indels_vcf_files  CNV_tab_files     SV_bedpe_files
Sample1   s1_snv.vcf        s1_id.vcf         s1_cnv.tsv        s1_sv.bedpe
Sample2   s2_snv.vcf        s2_id.vcf         s2_cnv.tsv        s2_sv.bedpe
```

```
Sample3   s3_snv.vcf        s3_id.vcf         s3_cnv.tsv        s3_sv.bedpe
...
```

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.

Below is an example of the pipeline output: