User manual for the genomeChart 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

This document describes how to use the `genomeChart` command line script from the signature.tools.lib R package.

A genome chart combines multiple visualisations to produce a comprehensive view of the provided somatic mutations. Somatic mutations that are accepted by the `genomeChart` script are: single nucleotide variants, indels, copy number variants, structural variants.

To use the `genomeChart` script, one needs to prepare an input table, where each row contains the location of the various somatic mutation files for each sample.

2. Installation

The script `genomeChart` 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 `genomeChart` 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 `genomeChart` script to a location in your command line PATH.

3. solutionSelectionForFitMS options

The list of available options can be accessed by typing:

```
genomeChart --help
```

This is the current output:

```
This script runs the genomeChart function of the signature.tools.lib R package.

Run this script as follows:

genomeChart [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 and of Indels_vcf_files
                              and Indels_tab_files are necessary
  -o, --outdir=OUTDIR       Name of the output directory. If omitted a name will
```

```
                              be given automatically.
  -e, --genomev=GENOMEV       Genome version to be used: hg19, hg38 or mm10. If not
                              specified GENOMEV=hg19.
  -h, --help                  Show this explanation.
```

4. Examples

4.1 Running genomeChart

We can run the `genomeChart` command line script as follows:

```
genomeChart -i inputTable.tsv -o outdir -e hg19
```

The content of the inputTable.tsv file can 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
...
```

Bear in mind that the somatic mutation files should be formatted according to the specifications of the function `genomeChart` in the R package signature.tools.lib, and in general already filtered according to user preference, as all mutations provided will be used.