GRAPES

Manual

Version 0.9.2

GRAPES detect deletions, tandem duplications and inversions on whole-genome or exome/targeted Illumina's short-read sequencing data.

For questions, bugs, comments etc. feel free to contact me: bdelolmo@gencardio.com

0. Initial requirements

GRAPES will only work on Unix-based systems (Mac OS included).
Before building and installing GRAPES, make sure you have available on path:

- Perl
- Rscript
- g++ compiler
- Boost C++ library
- BEDtools (>= 1.17)
- SAMtools (>=1.19)
- GNU core utilities: wget, awk, sort, cat, grep, head, tail, sed, cut, paste, uniq, wc and mv

Perl modules

- Parallel::ForkManager
- Sort::Key::Natural
- Statistics::Descriptive
- PDF::API2 (Optional)
- PDF::Table (Optional)

1. Download and install

First, you need to download the latest release available at github:

```
git clone --recursive https://github.com/bdolmo/GRAPES.git
cd GRAPES
./INSTALL.PL
```

INSTALL.PL is a perl script that will compile all its C++ code along with the required R packages for segmentation and plotting.. In addition it will download both 75-mer and 100-mer mappability tracks for GRCH37 and GRCH38.

2. Analyze WGS

```
./GRAPES wgs -bam <BAM_FILE> -g <GENOME_FASTA> -o <OUTPUT_DIR>
```

Optional arguments:
-r INT Minimum number of break-reads (reads spanning a breakpoint). Default is 5.
-c INT Minimum number of discordant read-pairs (reads flanking a breakpoint). Default is

5.
-s INT Number of standard deviations from the mean to consider a discordant read pair. Default is 4.
-e STRING Exclusion regions. BED file used to exclude specific regions.

3. Analyze WES/Gene panels

```
./GRAPES wes [-pooled | -test -control] -b <BED> -g <GENOME_FASTA> -o
<OUTPUT_DIR> -t <NUM_CPUs> <PIPELINE_COMMANDS> <PLOTTING_OPTIONS>
<TUNING_PARAMS>
```

Pipeline commands:

-all Perform all steps below (including plots)
-breakpoint Perform Breakpoint analysis
-extract Extract Depth, GC and Mappability
-offtarget Perform Off-target analysis
-build-ref Build a reference from a pool of samples
-normalize Normalize read depth
-callCNV Perform Copy Ratio and segmentation
-reportPDF create a report in PDF (only for gene-panels)

Plotting options:

-plotLargeCNV Plot segmented CNVs
-plotBiases Plot read depth bias reduction before and after normalization
-plotScatter Plot genome-wide cnv scatter plot
-plotKaryotype Plot karyotype

Tuning parameters:

-**minCorrelation** Minimum pairwise-correlation to build a reference set [default = 0.90]
-**minSampleSizeCluster** Default minimum number of samples to build a single baseline [default = 4]
-**maxSampleSizeCluster** Default maximum number of samples to build a single baseline [default = 15]
-**lowerDelCutoff** Lower-bound deletion cuttof ratio [default = 0.35]
-**upperDelCutoff** Upper-bound deletion cuttof ratio [default = 0.71]
-**lowerDupCutoff** Lower-bound duplication cuttof ratio [default = 1.25
-**minSizeSV** Minimum SV size to report a breakpoint call [default = 15]

3.1. Pooled Analysis (creates a reference using all available samples):

```
./GRAPES wes -pooled <BAM_DIR> ...
```

3.1. Paired analysis: (tumor vs normal tissue):

```
./GRAPES wes -test <TEST.BAM> -control <CONTROL.BAM> ...
```