How to use shatterSeek

Ruibin Xi

July 25, 2016

Overview

The *shatterSeek* package provides the functionality for detecting chromothripsis with copy number variation and structural variation data. In this vignette, we give detailed guide about how to use this package.

Main functions

The main function of the *shatterSeek* package is the function `shatterseek`. This function requires a `SVs` object and a `CNVsegs` object. The `SVs` object contains all structural variations of a sample and the `CNVsegs` object contains the copy number segmentations of the sample. Currently, we only consider detecting chromothripsis for Homo sapiens and we only consider structural variations located on chromosome $1, 2, \cdots, 22, X$. We also assume that all structural variation and copy number variation coordinates are based on the reference genome GRCh37/hg19. If this is not the case, one may use *LiftOver* tool (e.g. the LiftOver tool as availble at the UCSC genome browser) to convert these coordinates to GRCh37/hg19.

Given a data frame `SV` that contains the breakpoint informtion of structrual variations of a sample, we may use the following command to create a `SVs` object,

```
> SVSample <- SVs(chrom1=SV$chrom1,pos1=SV$pos1,
+ chrom2=SV$chrom2,pos2=SV$pos2,SVtype=SV$SVtype)
```

where the option *SVtype* is optional. If all structrual variations are intra-chromosomal events (two breakpoints are on the same chromosome), we may also use the following commands

```
> SVSample <- SVs(chrom1=SV$chrom1,
 + pos1=SV$pos1,pos2=SV$pos2,SVtype=SV$SVtype)
```

or

```
> SVSample <- SVs(pos1=SV$pos1,chrom2=SV$chrom2,
+ pos2=SV$pos2,SVtype=SV$SVtype)
```

.

To create a `CNVsegs` object, one may use the following command

```
> segSample <- CNVsegs(chrom=seg$chrom,start=seg$start,end=seg$end,
+ log2copyratio=seg$log2copyratio)
```

, where `seg` is a data frame containing all segmentation of the sample.

After creating `SVSample` and `segSample`, one can now test if there is any evidence of a chromothripsis event. Chromothripsis events can be called with both SV and CNV data or with only SV data. If it is more conservative to use both SV and CNV data. One may using the following commands to call chromothripsis events

```
> chromthripsis <-  shatterseek(SVSample,B=B, cores=cores) ## only with SV data
```

or

```
> chromthripsis <-  shatterseek(SVSample,segSample,B=B, cores=cores) ## with both CNV
```

where B is the number of permutations for performing the permutation test and *cores* is the number cores to use for parrallel processing. This will give a `chromoth` object. The `chromoth` contains p-values for each chromosome; One p-value is based on structrual variation data and another p-value (if CNV data is used) is based on copy number data. The p-value based on structural variations is looking for if there is an enrichment of large clustered structural variations on a chromosome. The p-value based on copy number variation is looking for alternating pattern of copy number states. A score is also given based on the p-values, which is defined as -log10(SVpvalue) (only with SV) or -log10(SVpvalue)-log10(CNVpvalue) (with both SV and CNV). A larger score would be more likely to be a chromothripsis event. To call chromothripsis, one may require the score larger than a threshold (e.g. 2). To display these p-values, one may use

```
> chromthripsis
```

One can also convert this `chromoth` object to a data frame with

```
> as(chromthripsis,"data.frame")
```

The *shatterSeek* package also provides a plot function for inspecting structural variations and copy number variations of a chromosome. Note that this function is still in its primitive stage and we did not put effort to make it produce pretty plots. A refined version of this function will be released in later version of this package. To plot, one can use

```
> plot.chromoth(chromthripsis,chrom=chrom.name,sample=sample.name)
```

where *chrom.name* is the chromosome name to be plotted and *sample.name* is the sample name (optional).

```