**COPS version1.1**

**MANUAL**

**Prerequisites:**

1. Operating System : Linux 64bit
2. RAM : 4 GB
3. Samtools-0.1.12a or above
4. R programming language version 2.12.1
5. Perl module Distribution.pm

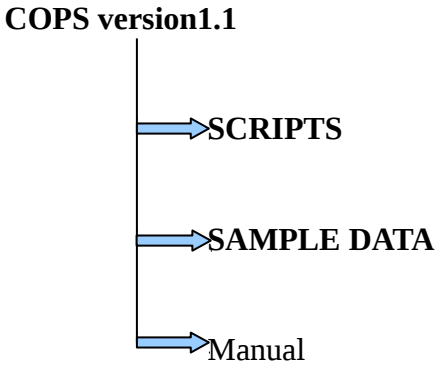
**Installation:**

Decompress the given file to a suitable location. Avoid placing any other files into the extracted folder.

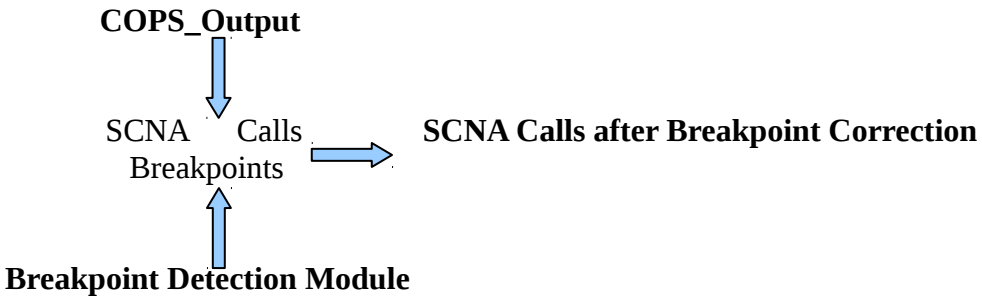
The hierarchy of COPS folder is as follows:

**COPS version1.1.zip**

After decompressing it will give following folders:




Upon successfully running COPS, an output folder (COPS_Output) will be created within the same COPS folder. It will contain SCNAs for each chromosome. A sample R script is provided to plot the SCNAs for individual chromosomes.




**Instructions:**

Locate the following files within the Scripts subdirectory, List_test.name & List_ref.name. These files should have all the name of chromosomes in your input sam/bam files (one per line) as per the

third field of your input sam/bam file.
Eg: chr1 or c1.fa or chr1.fa or c1 etc...

**NOTE :**

1. Only the chromosomes specified in the above files will be processed.
2. Ensure both the files should have the same chromosome name and number of chromosomes, in the same sequence.

Once the files have been filled with the required data, we can run the main script.
To run the main script,

% bash COPS.sh <*input file-type*> <*test file-name*> <*ref file-name*>

| | | |
|---|---|---|
| input file-type | : | 0 for ".bam" file and 1 for ".sam" |
| test file-name | : | File name of test/cancer sample (with full path) |
| ref file-name | : | File name of reference/normal sample (with full path) |

**NOTE:** *PLEASE PROVIDE ARGUMENTS IN THE SAME ORDER*
*PLEASE provide a co-ordinate sorted SAM/BAM file.*

Upon successful completeion, the Script will generate an output folder within the /COPS version1.1 directory called /COPS_output. The final output file ' Test.Specific.SCNAs' contains all the SCNAs and associated statistics.

**Caution : DO NOT PROCESS MULTIPLE SAMPLE PAIRS AT THE SAME TIME IN THE SAME FOLDER.**

**NOTE:** This tool was tested under the following conditions:

1. OS : Linux-2.6.35 (Ubuntu 10.10)
2. RAM : 4 GB DDR-3 @ 1333MHz
3. Hard disk : 2 TB
4. Processor : Quad Core(intel i3-3GHz)

**COPS OUTPUT FILE (Test.Specific.SCNAs) FORMAT:**

Coloumn 1 : Chromosome name
Coloumn 2 : Start position of SCNA
Coloumn 3 : End position of SCNA
Coloumn 4 : Cumulative $\log_2$ ratio
Coloumn 5 : t-statistic
Coloumn 6 : P-Value

**<u>Breakpoint Detection Module:</u>**

The user needs to provide additional parameters such as the mean and standard deviation of the library insert size.

% bash extractBreakPoints.sh <*input file-type*> <*test file-name*> <*ref file-name*> <*insert size mean*> <*insert size stdev*> <*genome build*>

| | | |
|---|---|---|
| input file-type | : | 0 for ".bam" file and 1 for ".sam" |
| test file-name | : | File name of test/cancer sample (with full path) |

| | | |
|---|---|---|
| ref file-name | : | File name of reference/normal sample (with full path) |
| insert size mean | : | mean library insert size in nucleotides |
| insert size stdev | : | standard deviation of library insert size in nucleotides |
| genome build | : | 0 for hg18, 1 for hg19 |

The script processes the chromosomes mentioned in List_test.name & List_ref.name, as described earlier with COPS workflow. In addition, the user must provide List_chr_hg18.size or List_chr_hg19.size, depending on the genome build used for alignment. This file consists of the sizes of chromosomes (in nucleotides) specified in List_test.name/List_ref.name in the same order. For the user's convenience, two files List_chr_hg18.size.orig and List_chr_hg19.size.orig are provided in the Scripts sub-directory, containing sizes of chromosomes 1-22, X, Y and M, in that order. The user may choose the sizes corresponding to his chromosomes of interest from these .orig files, as per the genome build, to create List_chr_hg18.size or List_chr_hg19.size, respectively.

Successful execution of the program will generate two output files within the /COPS_Output sub-directory: Test.Specific.Ins.BPs and Test.Specific.Del.BPs, corresponding to the Insertion and Deletion type breakpoints respectively.

The format of the breakpoint output file in event of Deletions is as follows:

X Copy Deletion BP [A – B]: Ratio

where X = 0.5 for mono-allelic deletion and 1 for full deletion;
A = chromosomal breakpoint boundary start
B = chromosomal breakpoint boundary end
Ratio = ratio of anomalous reads between the boundary end and start.

for e.g.

0.5 Copy Deletion BP [1911585 - 1911586]: 0.5
0.5 Copy Deletion BP [17805677 - 17805678]: 0.5
0.5 Copy Deletion BP [23407061 - 23407062]: 0.5
0.5 Copy Deletion BP [25721107 - 25721108]: 0.5
0.5 Copy Deletion BP [48806393 - 48806394]: 0.419753086419753
0.5 Copy Deletion BP [50302395 - 50302396]: 0.5
0.5 Copy Deletion BP [50613002 - 50613003]: 0.5
0.5 Copy Deletion BP [50636994 - 50636995]: 0.5
0.5 Copy Deletion BP [50646758 - 50646759]: 0.5
0.5 Copy Deletion BP [50651120 - 50651121]: 0.428571428571429

The format of the breakpoint output file in event of Insertions is as follows:

X Copy Insertion BP [A – B]: Ratio

where X = 0.5 for mono-allelic insertion, 1 for 1 copy insertion, ...
A = chromosomal breakpoint boundary start
B = chromosomal breakpoint boundary end
Ratio = ratio of anomalous reads between the boundary end and start.

The module can currently detect insertions upto 4 copies.

for e.g.

3 Copy Insertion BP [54640 - 54641]: 4
3 Copy Insertion BP [55568 - 55569]: 4
0.5 Copy Insertion BP [58387 - 58388]: 1.5
1 Copy Insertion BP [59126 - 59127]: 2
1 Copy Insertion BP [61540 - 61541]: 2
0.5 Copy Insertion BP [61541 - 61542]: 1.5
1 Copy Insertion BP [62027 - 62028]: 2
2 Copy Insertion BP [62556 - 62557]: 3
1 Copy Insertion BP [64318 - 64319]: 2
1 Copy Insertion BP [65462 - 65463]: 2

**Correction of COPS detected SCNA boundaries:**

The SCNA boundaries detected using COPS are mapped against the Insertion and Deletion category breakpoints. In regions of proximity within 1kb to the breakpoints, the COPS SCNA boundaries are corrected to reflect the breakpoint boundaries. Such COPS SCNAs with corrected breakpoints are reported in a file, 'Test.Specific.BPS.COPS.map', within the /COPS_output sub-directory.