

**Wabico : Multi-resolution GC bias correction and its application to copy number alteration identification**

version 1.0

Ho Jang

Data Mining and Computational Biology Lab (DMCB)

Gwangju Institute of Science and Technology (GIST)

**February 25, 2018**

# Contents

<b>1</b>	<b>Installation</b>	<b>1</b>
1.1	Requirements . . . . .	1
1.2	Installation details . . . . .	1
<b>2</b>	<b>Workflow</b>	<b>2</b>
2.1	Quantification of sequencing depth . . . . .	2
2.2	Generation of artificial control DOC . . . . .	2
2.3	Signal denosing using translation invariant wavelet transform . . . . .	3
2.4	DOC signal drawing . . . . .	3
2.5	Example . . . . .	4

# 1 Installation

## 1.1 Requirements

This software runs in Linux system. R program should be installed on your system.

## 1.2 Installation details

Before running provided programs, following R packages should be installed.

- snowfall R package for using multi processing
- data.table R package
- limma R package
- fields R package
- Rsamtools R package

Also following c codes should be compiled.

- Shell>R CMD SHLIB d150520008\_aggregate.c
- Shell>R CMD SHLIB d160810390\_calc\_num\_of\_overlaps\_between\_two\_sets\_of\_ranges.c
- Shell>R CMD SHLIB d160811216\_calc\_overlaps\_between\_two\_sets\_of\_ranges.c
- Shell>R CMD SHLIB d170626121\_get\_gc\_from\_starts\_and\_ends.c
- Shell>R CMD SHLIB d180218116\_get\_bin\_counts.c
- Shell>R CMD SHLIB d180218121\_get\_bin\_ranges.c
- Shell>R CMD SHLIB d180219121\_count\_doc\_within\_ranges.c

## 2 Workflow

### 2.1 Quantification of sequencing depth

This is the R function for quantifying copy number aberrations.

```
d180218106_make_DOC_bin(path.in.bam = '/home/tom/A.bam',  
                        paths.in.unique_pos = c('/home/tom/hg19.50mer.CRC.chr9.txt',  
                                              '/home/tom/hg19.50mer.CRC.chr21.txt'),  
                        cnts_ump_within_window = 100,  
                        arg.chrs = c('9', '21'))
```

- **path.in.bam** File path of WGS bam file
- **paths.in.unique\_pos** File paths of uniquely mappable positions
- **cnts\_ump\_within\_window** The number of uniquely mappable positions within the DOC window
- **arg.chrs** Reference sequence name for DOC quantification

### 2.2 Generation of artificial control DOC

This is the R function for generating artificial control DOC.

```
d180207106_self11(arg.list.doc = list.res$doc,  
                 arg.list.marker = list.res$marker,  
                 arg.path.umpositions = c('/home/tom/hg19.50mer.CRC.chr9.txt',  
                                         '/home/tom/hg19.50mer.CRC.chr21.txt'),  
                 arg.path.ref = c('chr9.fa',  
                                  'chr21.fa'),  
                 arg.thres.stop = 0.5,  
                 arg.max_decomposition_level = 10)
```

- **arg.list.doc** List of DOC signals of chromosomes
- **arg.list.marker** List of DOC markers of chromosomes
- **arg.path.umpositions** File paths of uniquely mappable positions
- **arg.path.ref** File paths of fasta reference sequences
- **arg.thres.stop** Threshold for DOC change direction ratio
- **arg.max\_decomposition** The maximum decomposition level

## 2.3 Signal denosing using translation invariant wavelet transform

This is the R function for denosing raw signal using translation invariant wavelet transform.

```
d180101331_denoising(arg.input = raw.signal,  
                    arg.level = 15,  
                    arg.C = 2)
```

- **arg.input** Input vector of signal
- **arg.level** Decomposition level
- **arg.C** Custom thresholding parameter

## 2.4 DOC signal drawing

This is the R function for drawing DOC signal.

```
d180109151_draw_DOC(arg.starts = starts,  
                   arg.ends = ends,  
                   arg.values = values,  
                   arg.lwd = 1,  
                   path.out = '/home/tom/my.png')
```

- **arg.starts** Start locations of the DOC bin
- **arg.ends** End locations of the DOC bin
- **arg.values** Values of the DOC bin
- **path.out** Output PNG file path

## 2.5 Example

The R function 'd180220101\_from\_bam\_to\_plot' provides whole functions from bam file quantification to DOC signal drawing based on the R functions described above. Also the R function 'myexample' provides an example code for running the code 'd180220101\_from\_bam\_to\_plot'. You can run the procedure by running the following codes in R,

```
># Modify myexample() in 'd180225101.r' file to your needs (such as bam file path, uniquely mappable position  
file path,...)  
>source('d180225101.r')  
>myexample()
```