f-circRNA is an integrated tool for the identification of fusion circular RNAs based on rRNA-depleted RNA-seq data.

System requirements

- OS: Linux x86 **64bit** system

- Script: Perl ,Python 2.4.3 or higher

Prerequisites

Software :

Fastqc(test is 0.11.4);
STAR-Fusion(test is 1.7.0) and STAR(test is 2.7.2a) [We recommend STAR version 2.7.2a or higher, the latest STAR library was builded by STAR-2.7.2a or higher]
Bwa(test is 0.7.16a)
Bedtools (test is 2.25)
Cutadapt (test is 1.16)
Samtools [0.1.19 is required if you use ericscript or find_circ]
Seqtk (test is 1.2)
Bowtie (test is 2.3.3)
BLAT

Samtools-0.1.19, Seqtk and BLAT is required for fusion software ericscript, if you don't use ericscript to detect fusion genes, you don't need to install them.

You need to add the executable file of these software to your environment path.

For better compatibility and user's convenience, ericscript-0.5.5 , trim_galore-0.5.0, CIRI2, find_circ-1.2, samtools-1.6 and Trimmomatic-0.38 are included in the package, users do not need to install and make these softwares. Just use the following command to unzip the package:

```
cd src
unzip ericscript-0.5.5.zip
unzip MapSplice-v2.2.1.zip
unzip samtools-1.6.zip
unzip Trimmomatic-0.38.zip
cd ..
```

Reference genome and annotation files.

We recommend download the reference and gtf files from Genecode(https://www.gencodegenes.org/).

Libraray:

1. If you want to use STAR-Fusion to detect fusion genes, you should prepare STAR-Fusion reference library:

The latest reference lib can be download from :
https://data.broadinstitute.org/Trinity/CTAT_RESOURCE_LIB/GRCh38_gencode_v31_CTAT_lib_Oct012019.plug-n-play.tar.gz

2. If you want to use ericscript to detect fusion genes, you should prepare ericscript_db,
The latest ericscript required reference lib can be download from :
https://sites.google.com/site/bioericscript/home

3. If you want to use mapsplice2 to detect fusion genes , you should use bowtie1 to index reference. Bowtie 2 index is not supported. A reference directory should be prepared as follows:

(1). In "FASTA" format, with '.fa' extension.
(2). One chromosome per sequence file.
(3). Chromosome name in the header line ('>' not included) is the same as the sequence file base name, and does not contain any blank space
E.g. If the header line is '>chr1', then the sequence file name should be 'chr1.fa'.
We provide split_reference.pl in the script folder to generate this format.

f-circRNA Usage:

```
perl $SCRIPT_PATH/f_cirRNA.pl \
   -input fastq.list \
   -gtf_annotation_file   $PATH/gencode.v31.annotation.gtf   \
   -qc_method             trim_galore \
   -fusion_method         STAR-fusion
   -STAR_lib              $path_to_STAR_lib   \
   -cir_method            CIRI2
```

Arguments:
Input Options:

-input	input fastq list, users need to prepare a file to tell f-circRNA the name and path of fastq, which separate with tab. example: Test1 test1_R1.fastq test1_R2.fastq Test2 test2_R1.fastq test2_R2.fastq
-gtf_annotation_file	gene annotation file, required

Output Options:

-outdir	output dir, default is output.

Analysis Options:

-qc_method	qc method, only support “trimmomatic” or “trim_galor”, default is “trim_galore”
-fusion_method	fusion method, only support “STAR-fusion”, “mapsplice2” or “ericscript”, default is “STAR-fusion”
-cir_method	cirRNA identify method, only support “CIRI2” or “find_circ”, default is “CIRI2”
-STAR_lib	STAR-fusion lib file，if fusion method is

	“STAR-fusion”,this file is required.
-mapsplice2_reference_dir	Reference file, users must prepare it as mapsplice2 required format, if fusion method is “mapslice2”, this file is required
-mapsplice2_bowtie1_index	Reference index file, bowtie1 is required to build this index, if fusion method is "mapslice2", this file is required
-ericscript_db_lib	ericscript lib file, if fusion method is "ericscript", this file is required

Notice: If you already have the fusion results, you can use the scripts blew to generate the fusion library instead of running the full pipeline which can save your time.

Command line reference : **get_fusion_sequence.pl**

Usage:

```
perl get_fusion_sequence.pl \
   -F          your_fusion_result.txt \
   -gtf        $PATH/gencode.v31.annotation.gtf   \
   -R          $PATH/GRCh38.primary_assembly.genome.fa \
   -L          out_fuison_location.txt        \
   -O          fusion_new_reference.fasta
```

Arguments:

-F, --fusion_file	fusion location file of genes,extract from the fusion software directly. [required]
-gtf, --gtf_annotation_file	gene annotation file, usually download from Genecode. [required]
-R, --reference	reference file, usually download from Genecode. Reference index file build by samtools "faidx" command also needed. [required]
-L, --out_fuison_location	the bed file we will generate,this file will be used in the script "get_fusion_new_gtf.pl" . [required]
-O, --output_prefix	the prefix of fasta output. [required]
-H , --help	output help information to screen

<fusion_file> should contain 8 colums as the blew format which separate with tab.

```
gene1     chr1 100  +     gene2     chr12     123  -
gene3     chr3 134  -     gene4     chr7 111  +
```

if your fusion results generate from STAR-fusion, mapsplice2 or ericscript, you can use the blow script directly to format the result.

```
perl progress_star_fusion_result.pl your_fusion_result.txt    temp
perl progress_mapsplice_result.pl your_fusion_result.txt    temp
perl progress_eriscript_result.pl your_fusion_result.txt    temp
```

Command line reference: **get_fusion_new_gtf.pl**

Usage:

```
perl get_fusion_new_gtf.pl \
   -F          out_fuison_location.txt   \   #this file is generated by get_fusion_sequence.pl
   -gtf        $PATH/gencode.v31.annotation.gtf      \
   -O          fusion_library.gtf
```

Arguments:

-F,	--fusion_file	Fusion file generate by get_fusion_sequence.pl [required]
-gtf,	--gtf_annotation_file	gene annotation file, usually download from Genecode. [required]
-O,	--output_prefix	the prefix of new gtf file output. [required]
-H ,	--help	output help information to screen

<fusion_file> is generated by script get_fusion_sequence.pl, contain 8 colums as the blew format which separate with tab

```
chr1 100  200  gene1_gene2_130_350   1     +
chr1 400  900  gene1_gene2_130_350   2     -
```

After get the fusion library, you can use the circRNA detecting software to find the f_circRNA.

Results

f-circRNA will generate 4 folders, the fusion libraries will be generated in the ‘fusion_results’ folder. The fusion circRNA results will be generated in the ‘cirRNA_results’ folder, which contains positions and sequences of fusion circRNAs.