Program description

The program is implemented in shell scripts and Java.
The program requires external programs: BEDTOOLS (https://github.com/arq5x/bedtools2) and SAMtools (http://samtools.sourceforge.net).
It also requires genome (chromosome) sequence and annotation files; the annotation files in mouse are distributed with the program.

The program is tested only on Mac OS X and Linux platforms.

Installation

Installation of external programs
BEDTOOLS have to be installed.

Setting the paths to all programs
The program creates output files as well as several temporary directories/files in the current working directory. Thus the easiest way to test the program is to set the temporary paths, for example,

export PATH=$PATH:/…/bedtools2-2.20.0/bin
export PATH=$PATH:/…/samtools-1.2/
export CLASSPATH=/…/circRNA_detection/bin

then run the program from the directory where the output files will be created.

CircRNA detection steps

The program requires Tophat fusion outputs.
For the test "accepted_read.bam", which is the Tophat fusion output file from MEF sample SRR2038028 (Andergassen et al. 2015)

1. The code below remove secondary mapping, separate mapped reads to read1 and read2 and convert .bam file to .sam file.

```
samtools view –f 0x0180 -o accepted_hits_read1.sam accepted_hits.bam
samtools view –f 0x0140 -o accepted_hits_read2.sam accepted_hits.bam
```

2. Find reads that pairs are mapped to the genome but not on proper coordinates.
2-1.

```
java MappedPairs  NUMSEQ=xxx READ1=xxx.sam READ2=xxx.sam
```

(See detail in 'Java program commands').
2-2.

```
sort -k1,1 -k3,3 -k4,4n unmapped-read1.sam > unmapped-read1-sorted.sam
sort -k1,1 -k3,3 -k4,4n unmapped-read2.sam > unmapped-read2-sorted.sam
sort -k1,1 -k2,2 -k3,3n notProperMappedRead1.txt > notProperMappedRead1-sorted.txt
sort -k1,1 -k2,2 -k3,3n notProperMappedRead2.txt > notProperMappedRead2-sorted.txt
```

3. Find circRNA junctions

```
java Detection ANNOTATION=xxx.bed UNMAPPED1=unmapped-read1-sorted.sam
UNMAPPED2=unmapped-read2-sorted.sam MAPPED1= notProperMappedRead1-sorted.txt
MAPPED2= notProperMappedRead2-sorted.txt
```

Output files

Output file, ‘exon-backspliced-circRNAs.txt’ from the program ‘Detection’ is a list of predicted refSeq exon backspliced circRNAs.

Java program commands

To run MappedPairs:
java MappedPairs NUMSEQ=xxx READ1=xxx.sam READ2=xxx.sam

Required: NUMSEQ=xx - sequence read length
Required: READ1=xx - file name for read 1
Required: READ2=xx - file name for read 2

To run Detection:
java Detection ANNOTATION=xxx.bed UNMAPPED1=unmapped-read1-sorted.sam UNMAPPED2=unmapped-read2-sorted.sam MAPPED1=mappedRead1-sorted.txt MAPPED2=mappedRead2-sorted.txt

Required: ANNOTATION=xx.bed - .bed format file containing annotated location of exon junctions
Required: UNMAPPED1=unmapped-read1-sorted.sam - output file from the previous step
Required: UNMAPPED2=unmapped-read2-sorted.sam - output file from the previous step
Required: MAPPED1=mappedRead1-sorted.txt - output file from the previous step
Required: MAPPED2=mappedRead2-sorted.txt - output file from the previous step