PASSion Manual

December 15, 2011
Version 1.2

PASSion is a pattern growth algorithm based pipeline for splice site detection in paired-end RNA-Seq reads. It can discover differential and shared splicing patterns between different samples.

Availability: The code and utilities can be freely downloaded from https://trac.nbic.nl/passion and ftp://ftp.sanger.ac.uk/pub/zn1/passion

Contact: y.zhang@lumc.nl; k.ye@lumc.nl

1 Getting started

1.1 Installation

PASSion requires gcc >=4.3, SMALT, perl, and SAMtools to be pre-installed.

Go to PASSION directory to compile the component.

```
cd  path_to_passion/PASSION
sh compile.sh
```

Go to SMALT directory, set the right binary version of smalt (according to your system)

```
cp smalt_version smalt
(example:) cp smalt_x86_64 smalt
```

Install Samtools

Set SMALT, SAMtools and PASSION directory to PATH

Details see Section 2.

1.2 Index the reference

```
smalt index -k 13 -s 6 refindex reference.fa
samtools faidx reference.fa
```

1.3 Run PASSion

For one library

```
passion.pl -s insert_size -r read1.fq -f read2.fq -R reference.fa -I
refindex
```

To detect shared and sample specific junctions between multiple samples

1. Apply PASSion individually
2. Edit sample configure file. For example sample.txt

```
./s8_passion_output      s8
./s1_passion_output      s1
./s2_passion_output      s2
```

3. Run passion_diff.pl

```
passion_diff.pl -f sample.txt
```

Example: discovery of junctions at chromosome 17

```
cd  path_to_testset/testset
smalt index -k 13 -s 6 17 17.fa
samtools faidx 17.fa
passion.pl -s 200 -r read1.50.fq -f read2.50.fq -R 17.fa -I 17 -o s1_passion_output
passion.pl -s 300 -r read1.75.fq -f read2.75.fq -R 17.fa -I 17 -o s2_passion_output
passion.pl -s 500 -r read1.100.fq -f read2.100.fq -R 17.fa -I 17 -o s8_passion_output
passion_diff.pl -f sample.txt
```

Details about how to set the parameters, please see Section 3.

1.4 Output

Junctions.bed: Detected junctions in bed format

Detected junctions in bed format: column 7 and 8 record the breakpoint range.
Junction start = Column2+Column11[1]; Junction end = Column3-Column11[2]+1

```
track name=junctions description= "PASSION junctions"
17      12920399        12921087        JUNC_1  1       -       12920412        12921055        255,0,0 2       16,34   0,654
17      41959870        41960311        JUNC_2  1       -       41959878        41960272        255,0,0 2       9,41    0,400
17      79205244        79205395        JUNC_3  1       -       79205273        79205378        255,0,0 2       32,18   0,133
17      74141605        74157944        JUNC_4  1       +       74141643        74157935        255,0,0 2       40,10   0,16329
```

Junctions.detail: Details about how split reads align to exon-exon boundaries

1st Summary: Due to the existance of microhomology, breakpoint ranges are also reported. For example, at the following breaking point, "AG" can be either aligned to the left or the right. In this format, we report the leftmost breakpoint and the range. The final breakpoint will be design with the assistance of splicing motifs. LL/RL: left/right hanging length on the exons. +/-: downstream/upperstream reads for support
2nd line: Reference
3rd line: Alignment

```
################################################################################################
401     D 1331  ChrID 17        BP 1678492      1679824 BP_range 1678492        1679826 Supports 5      + 3     - 2     S1 12   S2 123.658
GGACCCTAAGGCTGTTTTACGCTATGGCTTGGATTCAGATCTCAGCTGCAaggtctgtag<1311>cacttgtctcAGATTGCCCAGCTGCCCTTGACCG
                                CAGATCTCTGCTGCA                         AGATTGCCCAGCTGCCCTTGACCGGA                              +       167554
                    TACGCTATGGCTTGGATTCAGATCTCAGCTGCA                    AGATTGCCCAGCTGCCC                                      -       167993
                                TTCAGATCTCAGCGGCA                       AGATTGCCCAGCTGCCCTTGACCCGAAGCATGA                       -       167992
                        TATGGCTTGGATTCAGATCTCAGCTGCA                    AGATTGCCCAGCTGCCCTTGAC                                 +       167531
                                        CTCAGCTGCA                      AGATTGCCCAGCTGCCCTTGACCGGAAGCATGAGTATCAT               +       167534
```

orignal.sam: exonic reads alignment in SAM format

final.sam: exnoic reads and split reads alignment in SAM format

Junctions_mix.final:"Lenth \t Ref \t BP_S \t BP_E \t Range_S \t Range_E \t Support \t Sup+ \t Sup- \t LL \t RL \t Marker \t Start \t End \t Lefexon_cov \t Rightexon_cov\sample_count\details\n";

13755	17	35804869	35818625	35804869	35818626	3	3	0	17	39	GTAG
1202	17	73204713	73205916	73204713	73205918	2	1	1	33	24	GTAG
1092	17	48601143	48602236	48601143	48602238	2	0	2	39	18	GTAG
566	17	73567181	73567748	73567181	73567749	2	1	1	35	22	GTAG
25061	17	13980370	14005432	13980370	14005435	1	1	0	26	24	GTAG
1471	17	29686031	29687503	29686031	29687505	8	4	4	37	40	GTAG
1069	17	62125345	62126415	62125345	62126416	1	0	1	11	39	CTAC
183	17	74936628	74936812	74936628	74936817	6	3	3	38	34	GTAG
649	17	62020456	62021106	62020456	62021109	1	0	1	29	21	CTAC

2 Full Installation

2.1 g++

g++ >=4.3

2.2 Perl

Perl needs to be installed at /usr/bin/perl

2.3 SMALT

Download smalt-0.4.3 and choose the right binary of SMALT. If your system is linux x86_64

```
cp smalt_x86_64 smalt
```

2.4 Samtools

Download samtools-0.1.8

Follow INSTALL instruction. *** PASSion currently does not support the samtools versions with *mpileup*.

2.5 PASSion

```
cd path_to_passion/PASSION/
chmod +x *
sh compile.sh
```

2.6 Set PATH environment variable

Set SMALT, Samtools and PASSion's path to PATH environment variable.

For linux users:

```
vim ~/.bashrc or vim ~/.cshrc
export PATH=$PATH:path_to_samlt/smalt: path_to_samtools/samtools:
path_to_passion/PASSION
```

3 User Manual

3.1 For one library

```
Usage: passion.pl Arguments Options
Arguments:
    -s/--insert_size       insert size is the length of the two reads together with
                           the non-sequenced part.
    -r/--read1             read1 file in fasta/fastq format
    -f/--read2             read2 file in fasta/fastq format
    -R/--ref               the reference sequence in fasta format
    -I/--refindex          the reference index using SMALT
```

```
Options:
    -c/--cutoff           cutoff=(number of support reads)/(coverage of higher
                          expressed flanking exon) [default 0.1]
    -d/--divide2files     divide bed file according to split sites (GT-AG, GC-AG,
                          AT-AC and unknown motifs) [default F]
    -x/--exonisland_file  user defined exon islands file in GFF or GTF format (the
                          1st, 4th, and 5th fields the coordinates in the genome
                          instead of relative location in a gene) [default F]
    -S/--max_SNP          max number of SNP allowed [default 2]
    -M/--max_intron_index maximum intron index [default 7]
                          [1]=100; [2]=400; [3]=1600; [4]=6400; [5]=25600;
                          [6]=102400; [7]=409600; [8]=1638400; [9]=6553600;
                          [10]=26214400; [11]=104857600; [12]=419430400;
    -m/--min_intron       minimum intron size [default 20]
    -o/--output_folder    output folder [default ./passion_output]
    -e/--sequence_error   sequence error rate [default 0.05]
    -T/--thread           number of thread [default 1]
    -w/--window_size      window size [default 5000000]
    -h/--help             help
    -v/--version          version
```

3.2 Detect differential splicing pattern in multiple samples:

```
Usage: passion_diff.pl Arguments Options
Arguments:
    -f/--configurefile    configuration file format: passion_output_path tag
                          (separate by \t)
Options:
    -c/--cutoff           cutoff=(number of support reads)/(coverage of higher
                          expressed flanking exon) [default 0.1]
    -d/--dividebysite     divide bed file according to split sites (GT-AG, GC-AG,
                          AT-AC and unknown motifs)  [default F]
    -o/--outputfolder     output folder [default ./passion_diff_output]
    -h/--help             help
    -v/--version          version
```