tagMeppr

Robin H. van der Weide

2019-08-30

Hi there! Welcome!

TagMap is a very useful method for transposon mapping (Stern 2017), enabling researchers to map the insertion sites with ease and generate long sequencing reads. However, there is little to none automatisation and downstream analysis software available for these reads. TagMeppr is an easy to use, memory efficient fastq-to-figure package written in R.

Usage

The basic usage of tagMapper revolves around three clear steps:

1. index: a tagMapper-index is made once for a specific genome and protocol (e.g. hg19 and PigyBac).
2. align: a tagMapperSample-object is made and aligned to the index
3. analyse: determine and plot highly likely integraton-sites

```
library(tagMeppr)
```

tagMepprIndex

Remember: you only have to make a tagMepprIndex once per genome and protocol! This means that you can use it for however many samples for that protocol and organism (e.g. PiggyBac in human cells) you may have.

```
library("BSgenome.Hsapiens.UCSC.hg19")

reference_hg19_PB = makeIndex(indexPath = '../premadeReferences/',
                              bsgenome = BSgenome.Hsapiens.UCSC.hg19,
                              ITR = 'PiggyBac',
                              targetInsertionSite = 'TTAA', verbose = T)
```

Loading a tagMepprIndex is easy with loadIndex().

```
reference_hg19_PB = loadIndex('BSgenome.Hsapiens.UCSC.hg19_PiggyBac_tagMeppRindex.fa.gz')
```

You can use print() on this tagMepprIndex, which gives you a lot of information:

```
print(reference_hg19_PB)
#> tagMepprIndex
#>
#>  Protocol: PiggyBac
#>
#>  Directory: .
#>  Fasta: BSgenome.Hsapiens.UCSC.hg19_PiggyBac_tagMeppRindex.fa.gz
#>
#>  Target insertion site: TTAA
#>  TIS: BSgenome.Hsapiens.UCSC.hg19_PiggyBac_tagMeppRindex.tis
#>  Number of target insertion sites: 19228699
```

tagMepprSample

Creating a new sample

```
C91 = newTagMeppr(F1 = 'clone91_FWD_R1.fq.gz',
                  F2 = 'clone91_FWD_R2.fq.gz',
                  R1 = 'clone91_REV_R1.fq.gz',
                  R2 = 'clone91_REV_R2.fq.gz',
                  name = "clone9 (minimised)",
                  protocol = 'PiggyBac')
```

This is the print:

```
print(C91)
#> tagMepprSample
#>
#>  Name: clone9 (minimised)
#>  Protocol: PiggyBac
#>  Primer checked: no (highly recommended!!!)
#>  Aligned: FALSE
#>  Analysed: FALSE
```

checken

Normally, we expect the reverse-primer to be found at the start of the ITR- sequence. This can be checked with `checkPrimer()`. If it appears to be to other way around, the sample-object will have a `flipPrimer=T` flag.

Highly recommend this!

```
fwdPrimer = "CGTCAATTTTACGCAGACTATC"
revPrimer = "GTACGTCACAATATGATTATCTTTCTAG"

checkPrimer(fwdPrimer = fwdPrimer,
            revPrimer = revPrimer,
            exp = C91,
            ITR = 'PiggyBac')
```

```
print(C91)
#> tagMepprSample
#>
```

```
#>  Name: clone9 (minimised)
#>  Protocol: PiggyBac
#>  Primer checked: yes (flipped)
#>  Aligned: FALSE
#>  Analysed: FALSE
```

Align

To align the tagMeppr-sample to the index, just use the `align()`-function.

Omg. Mapping without leaving R! Please check if you have BWA and samtools intalled. Note no <- signs

```
align(exp = C91,
      ref = reference_hg19_PB,
      cores = 30,
      empericalCentre = T)
```

The object will be updated, which keeps your evironment tidy.

```
print(C91)
#> tagMepprSample
#>
#>  Name: clone9 (minimised)
#>  Protocol: PiggyBac
#>  Primer checked: yes (flipped)
#>  Alignment-folder: /tmp/DGFDY87778E
#>  Informative FWD-reads: 5189
#>  Informative REV-reads: 30250
#>  Analysed: FALSE
```

Analysis

Find insertions

```
findInsertions(exp = C91, ref = reference_hg19_PB, padding = 2)
```

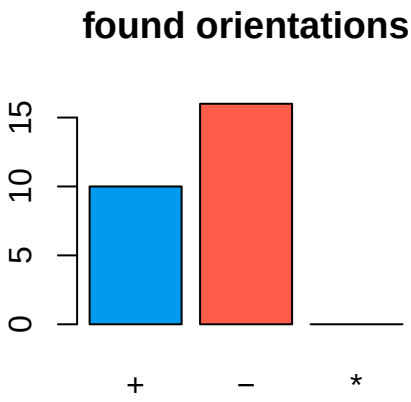
```
print(C91)
#> tagMepprSample
#>
#>  Name: clone9 (minimised)
#>  Protocol: PiggyBac
#>  Primer checked: yes (flipped)
#>  Alignment-folder: /tmp/DGFDY87778E
#>  Informative FWD-reads: 5189
#>  Informative REV-reads: 30250
#>  Unique TISs covered: 26
#>      p<0.05: 26
```

```
foundInsertions = results( C91 )

head(foundInsertions)
#>   seqnames     start       end strand fwdCount revCount fwdD revD  padj
#> 1     chr1  36377340  36377343      -      142     1360    1   -1 2e-16
```

```
#> 2     chr1 225144936 225144939      -      173      930    1   -1 2e-16
#> 3     chr3  75442959  75442962      -       41      750    1   -1 2e-16
#> 4     chr3 129776748 129776751      -      129     1753    1   -1 2e-16
#> 5     chr3 194542188 194542191      -      132     1071    1   -1 2e-16
#> 6     chr3 195352648 195352651      +      328     1482   -1    1 2e-16
```

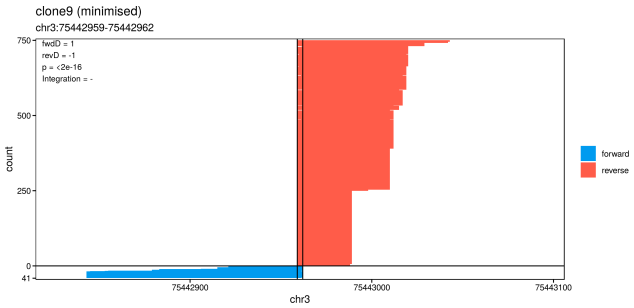
```
barplot(table(foundInsertions$strand), horiz = F,
        col = tagMepprCol(),
        main = 'found orientations')
```



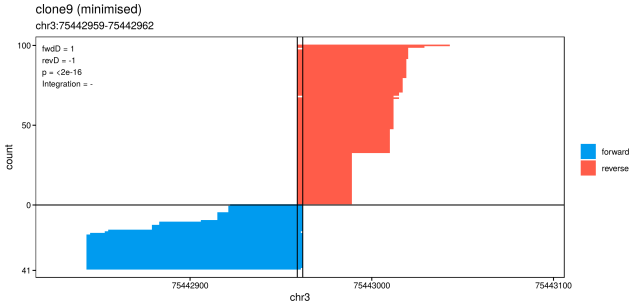
Plot

Single insertions

```
plotSite(C91,site = 3)
```

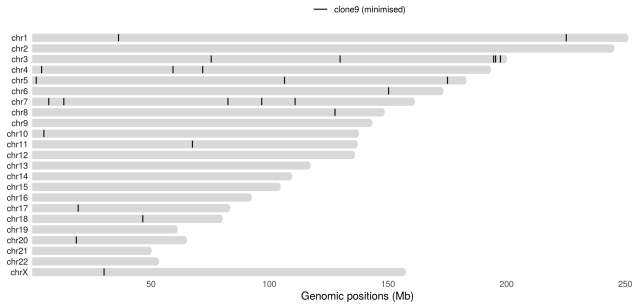



```
plotSite(C91,site = 3, maxReads = 100)
```



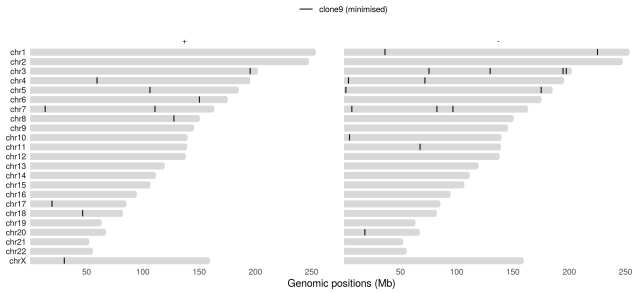
All insertions

```
plotInsertions(exp = C91)
```



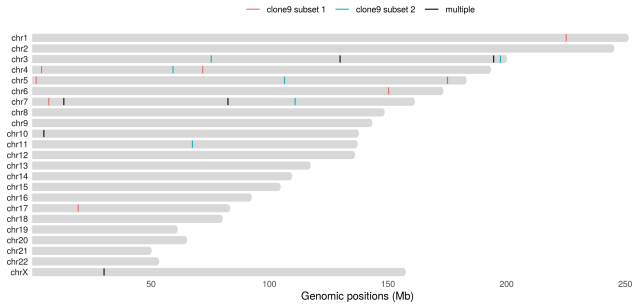
show orientation:

```
plotInsertions(exp = C91, showOrientation = T)
```

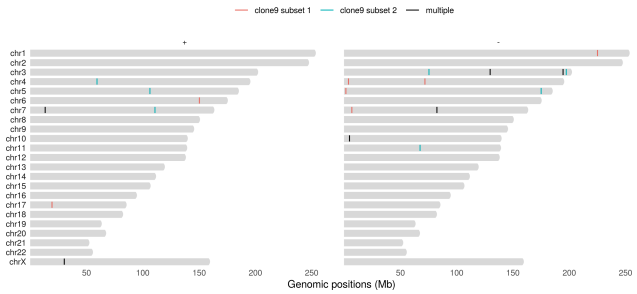


Plot muliple samples

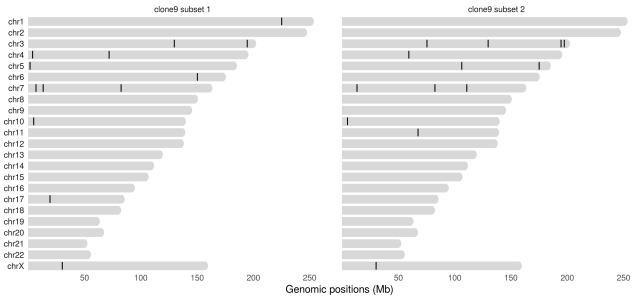
```
plotInsertions(tagMepprSampleList)
```




```
plotInsertions(tagMepprSampleList, showOrientation = T)
```




```
plotInsertions(tagMepprSampleList, sideBySide = T)
```



Biography

Stern, David L. 2017. "Tagmentation-Based Mapping (Tagmap) of Mobile Dna Genomic Insertion Sites." *bioRxiv*. Cold Spring Harbor Laboratory. https://doi.org/10.1101/037762.