

viewing alignments for structural rearrangements

Eitan Halper-Stromberg

February 4, 2013

Abstract

This package is designed to evaluate structural rearrangement calls from a candidate list, the output for tools such as HYDRA, GASV, VariationHunter, etc. The user should have a text file with one row per candidate structural rearrangement. For each candidate rearrangement, readpairs from the two loci will be read in and realigned three different ways. One of these realignments supports the structural variant, with readpairs realigned to a sequence representing the rearranged sequence (the sequence of the two loci concatenated together). The other two realignments support no structural rearrangement, with readpairs realigned to the two sequences representing contiguous fragments of the reference genome taken from each of the two loci.

1 Simple Usage

```
> library(targetSeqView)
> library(grid)
```

Perform realignment on 2 candidates that failed to validate

Our text file indicates the loci allegedly involved in two structural variants, one structural variant per row. For each row we will load reads aligning to the two loci involved in the structural variant from our bam file, realign these reads, calculate our likelihood score, and visualize the realignments. These particular structural variants failed PCR validation. This will become evident once we view the realignments (we will do this for the one of the structural variants) and when we inspect the likelihood scores (we will compare these to the likelihood scores from candidate structural variants that passed PCR validation)

```
> path <- system.file("extdata", package="targetSeqView")
> nodes=1
> registerDoMC(nodes)
> filename=file.path(path, "twoSVJunctionsFailed.txt")
> retfail=ViewAndScore(filename=filename, bamFilePath=path, estimateIndelRate=FALSE,
+ estimateMmRate=FALSE, getReadLength=FALSE, build="hg19", verbose=TRUE)
```

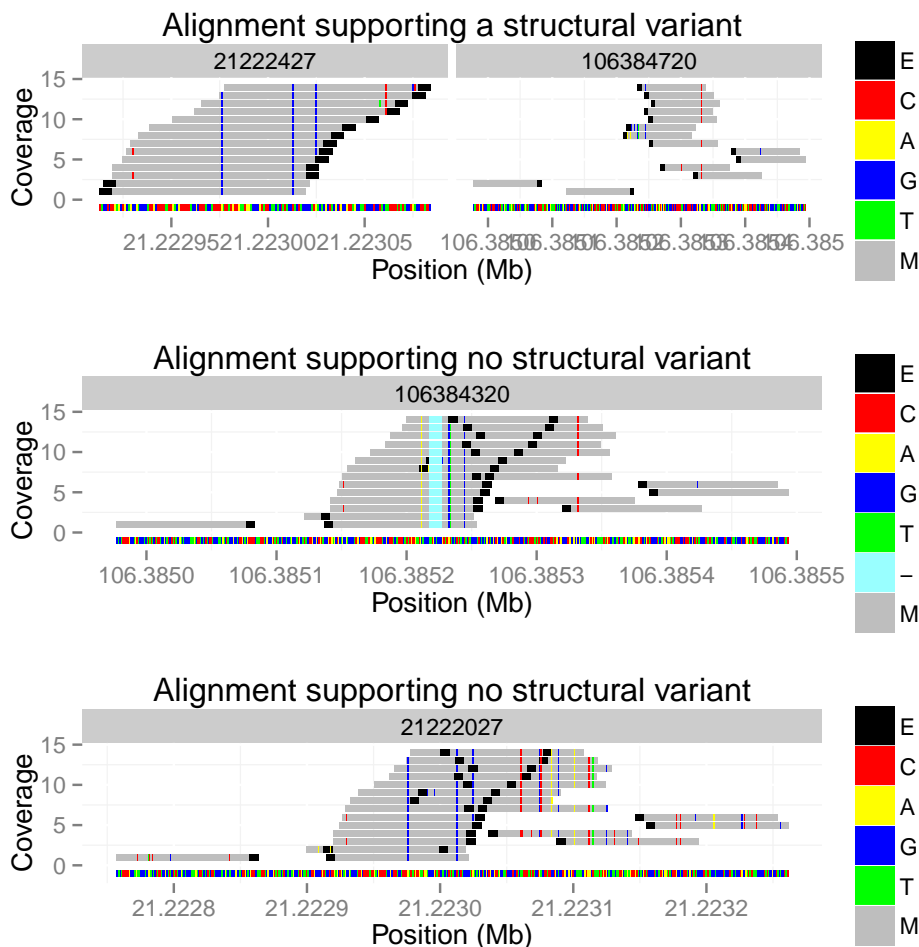
```
[1] "Working on event 1 of 2"
[1] "primary alignment for event 1 done"
[1] "secondary alignment (1 of 2) for event 1 done"
[1] "secondary alignment (2 of 2) for event 1 done"
[1] "Working on event 2 of 2"
[1] "primary alignment for event 2 done"
[1] "secondary alignment (1 of 2) for event 2 done"
[1] "secondary alignment (2 of 2) for event 2 done"
```

```
> ## The likelihood scores for the events failing validation
> print(retfail[[2]])
```

```
[1] 26.69819 132.74742
```

View the 3 realignment configurations for one of these failing events

```
> p1=formatPlot(retfail[[1]][[1]][[1]][[2]],title='Alignment supporting a structural variant')
> p2=formatPlot(retfail[[1]][[1]][[2]][[2]],title='Alignment supporting no structural variant')
> p3=formatPlot(retfail[[1]][[1]][[3]][[2]],title='Alignment supporting no structural variant')
> grid.newpage()
> pushViewport(viewport(layout = grid.layout(3, 1)))
> print(p1, vp = viewport(layout.pos.row = 1, layout.pos.col=1))
> print(p2, vp = viewport(layout.pos.row = 2, layout.pos.col=1))
> print(p3, vp = viewport(layout.pos.row = 3, layout.pos.col=1))
```



Explanation of Pictures

The top picture shows read pairs aligning to the two loci allegedly involved in the structural variant. This alignment supports the existence of a structural variant as each paired end has one end aligned to the left locus and one end aligned to the right locus. There appear to be 4 snps identified consistently by the reads in the locus on the left side of the top picture. The picture in the middle represents the same read pairs as in the top picture aligned only to locus on the right side of the top picture. The absence of white space between some read pairs indicates paired ends with overlapping alignments, i.e fragments that are shorter than 200 bp. The read ends on the left of the middle picture indicate 4 snps and one indel. The similarity in alignment qualities between the alignment supporting the structural variant and one of the two alignments supporting no structural variant is an indication that this is a false positive.

Perform realignment on 2 candidates that passed validation

```
> filename=file.path(path,"twoSVJunctionsPassed.txt")
> retpass=ViewAndScore(filename=filename,bamFilePath=path,estimateIndelRate=FALSE,
+ estimateMmRate=FALSE,getReadLength=FALSE,build="hg19",verbose=TRUE)

[1] "Working on event 1 of 2"
[1] "primary alignment for event 1 done"
[1] "secondary alignment (1 of 2) for event 1 done"
[1] "secondary alignment (2 of 2) for event 1 done"
[1] "Working on event 2 of 2"
[1] "primary alignment for event 2 done"
[1] "secondary alignment (1 of 2) for event 2 done"
[1] "secondary alignment (2 of 2) for event 2 done"

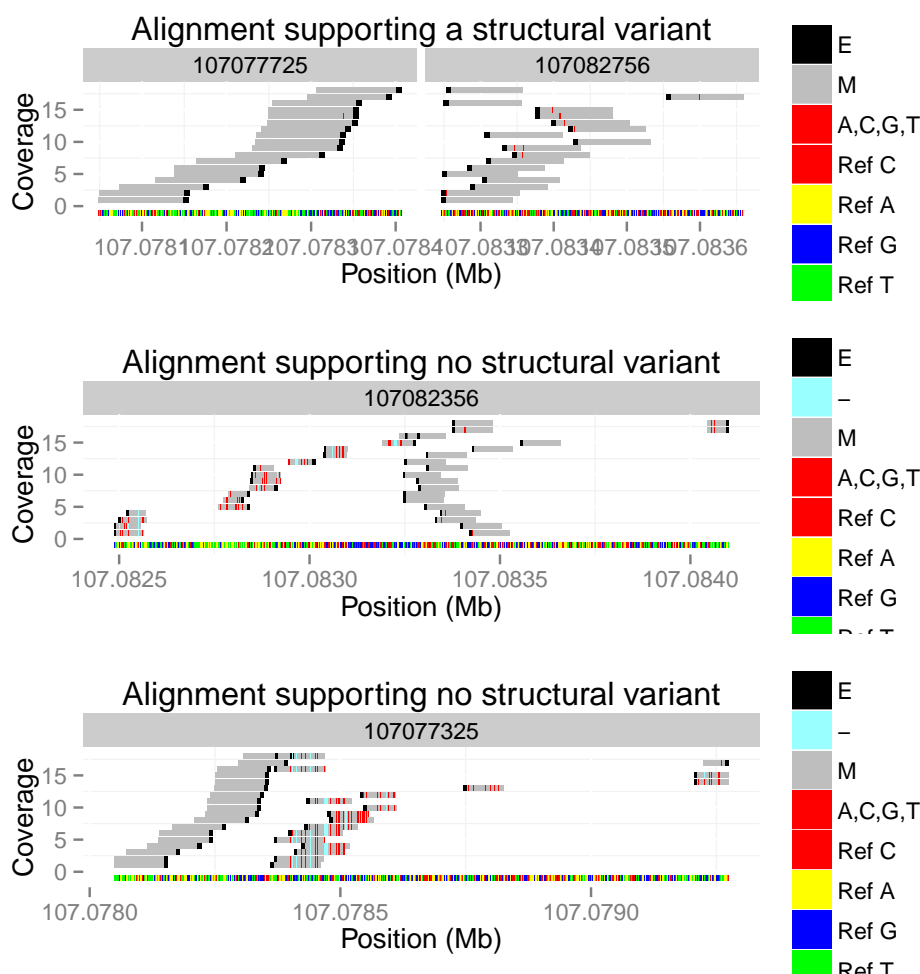
> ## The likelihood scores for the events passing validation:
> print(retpass[[2]])

[1] 2041.7801 463.7081
```

If we compare the likelihood scores from the validated events to the unvalidated events we see that the validation event likelihoods are substantially greater

View the 3 realignment configurations for one of these passing events

```
> p1=formatPlot(retpass[[1]][[1]][[1]][[2]],title='Alignment supporting a structural variant',
+ mismatchcolor='red')
> p2=formatPlot(retpass[[1]][[1]][[2]][[2]],title='Alignment supporting no structural variant',
+ mismatchcolor='red')
> p3=formatPlot(retpass[[1]][[1]][[3]][[2]],title='Alignment supporting no structural variant',
+ mismatchcolor='red')
> grid.newpage()
> pushViewport(viewport(layout = grid.layout(3, 1)))
> print(p1, vp = viewport(layout.pos.row = 1, layout.pos.col=1))
> print(p2, vp = viewport(layout.pos.row = 2, layout.pos.col=1))
> print(p3, vp = viewport(layout.pos.row = 3, layout.pos.col=1))
```



Explanation of Pictures

As with the earlier pictures, the top picture shows alignments support a structural variant while the middle and bottom pictures show alignments supporting no structural variant. The top picture shows read pairs with both ends aligning well to each of the two loci. The middle and bottom pictures have one read from each pair aligning well and one read aligning poorly (mismatches are all indicated in red here for clarity, a viewing option available in the formatPlot function). This pattern is indicative of a true structural rearrangement, as we are unable to make both pairs of the paired end reads align well to only one locus of the two involved in the junction.

```
> toLatex(sessionInfo())
```

- R version 2.15.2 Patched (2012-10-28 r61038), x86_64-unknown-linux-gnu
- Locale: LC_CTYPE=en_US.iso885915, LC_NUMERIC=C, LC_TIME=en_US.iso885915, LC_COLLATE=C, LC_MONETARY=en_US.iso885915, LC_MESSAGES=en_US.iso885915, LC_PAPER=C, LC_NAME=C, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.iso885915, LC_IDENTIFICATION=C
- Base packages: base, datasets, grDevices, graphics, grid, methods, stats, utils
- Other packages: BSgenome 1.26.1, BSgenome.Hsapiens.UCSC.hg19 1.3.19, BiocGenerics 0.4.0, Biostrings 2.26.2, GenomicRanges 1.10.5, IRanges 1.16.4, Rsamtools 1.10.2, doMC 1.2.5, foreach 1.4.0, ggplot2 0.9.2.1, iterators 1.0.6, multicore 0.1-7, targetSeqView 1.0

- Loaded via a namespace (and not attached): MASS 7.3-22, RColorBrewer 1.0-5, bitops 1.0-5, codetools 0.2-8, colorspace 1.2-0, compiler 2.15.2, dichromat 1.2-4, digest 0.6.0, gtable 0.1.1, labeling 0.1, memoise 0.1, munsell 0.4, parallel 2.15.2, plyr 1.7.1, proto 0.3-9.2, reshape2 1.2.1, scales 0.2.2, stats4 2.15.2, stringr 0.6.1, tools 2.15.2, zlibbioc 1.4.0