

Manual for the R `casper` package

1 Introduction

The package `casper` implements statistical methodology to infer gene alternative splicing patterns from paired-end high-throughput sequencing data [Rossell, 2010].

Currently, we assume that splicing variants are known. We plan to implement the case of *de novo* variant discovery in the near future.

2 Estimating expression for a set of known variants

Below we show the code needed to load and analyze some sample data included with `casper`. Sample data was generated using real drosophila gene coordinates (ENTREZ id 43792) and preassigning the fraction of reads per variant to

tx	frac
99	0.0254
100	0.0708
101	0.1083
102	0.1453
103	0.1784
104	0.2202
105	0.2513

```
> .libPaths("/software/R_libs")
> library(casper)
> genomeName = "dm3"
> mc.cores = 8
> genomeDB <- procGenome(genomeName, mc.cores)
```

Processing Exons and Transcripts
 Finding non-overlapping exons
 Remapping transcript structure to new exons

```

> load("/Volumes/biostats/routines/R/casper_library/chr2data.RData")
> distrs <- getDistrs(genomeDB$txs, genomeDB$exons, data$frags,
+   mc.cores = mc.cores)
  
```

Calculating fragment length distribution
 Calculating fragment start distribution

```

> datadir <- system.file("data", package = "casper")
> bamFileName <- paste(datadir, "/simulated.sam", sep = "")
> data <- procBam(bamFileName, samtools = "", chrom = "", bam = 0,
+   parallel = FALSE)
  
```

Reading file with 2e+06 reads

```

> pathCounts <- pathCounts(data$reads, genomeDB$exonsNI)
  
```

Finding overlaps between reads and exons
 Preparing data
 Counting paths

```

> exp <- calcExp(distrs, genomeDB, pathCounts)
> geneExp <- genPlot(43792, genomeDB, data$reads, exp)
  
```

```

> cols <- rep(rainbow(min(length(geneExp$exp), 10)), ceiling(length(geneExp$exp)))
> barplot(geneExp$exp, col = cols)
  
```

```

> geneExp$exp
  
```

	99	100	101	102	103	104	105
2.545040	7.082151	10.832400	14.536399	17.841947	22.026992	25.135071	

References

- D. Rossell. QASPER: Quantifying Alternative Splicing from Paired End Reads. Technical report, Institute for Research in Biomedicine of Barcelona, 2010.

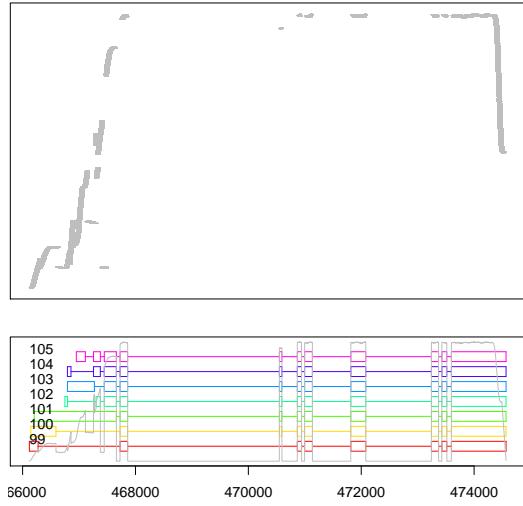


Figure 1: Aligned reads to variants from gene of interest

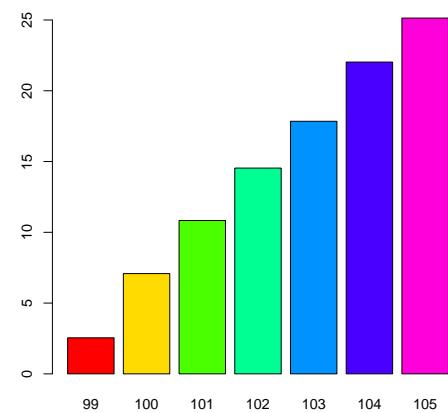


Figure 2: Estimated expression of variants