

Vignette - The use of hoardeR

Daniel Fischer

June 28, 2016

Installation

The installation of hoardeR is straight forward, it is located on Cran and can be installed with the command

```
install.packages("hoardeR")
```

There are no special requirements for the installation, however there are a few package dependencies that have to be met. Normally, these dependencies should be installed automatically, if not, missing Cran packages can be installed as above, with

```
install.packages("packageName")
```

and missing Bioconductor packages (e.g. the package *Biostrings*) with

```
source("https://bioconductor.org/biocLite.R")
biocLite("Biostrings")
```

The latest developer version of hoardeR is located on GitHub, here

<https://github.com/fischuu/hoardeR>

Packages from GitHub can be installed directly in R, using the *devtools* package

```
install.packages("devtools")
library("devtools")
install_github("fischuu/hoardeR")
```

Typical workflow prior the use of hoardeR

The common way to use hoardeR is to search cross-species orthologs of unannotated regions in a present organism. For that, typically, an RNA-seq experiment has been conducted, the reads have been mapped to a reference genome and gene expressions have been estimated using some annotation.

Further, reads from non-annotated regions have been analysed and a set of novel gene candidate regions has been identified. This can be done either across the whole data set, or individually for each sample.

From this analysis, the user has either a gtf file with novel loci (e.g. Cufflinks provides this), or then a bed file with the new loci.

The first rows of a typical gtf file would like this

1	Cufflinks	exon	242203	242862	.	+	.	gene_id "XL0C_000002"; transcript_id "TC0NS_00000002"; exon_number "1"; oId "CUFF.2.1"; class_code "u"; tss_id "TSS2";
1	Cufflinks	exon	242203	242646	.	+	.	gene_id "XL0C_000002"; transcript_id "TC0NS_00000003"; exon_number "1"; oId "CUFF.1.1"; class_code "u"; tss_id "TSS2";
1	Cufflinks	exon	254559	256717	.	+	.	gene_id "XL0C_000002"; transcript_id "TC0NS_00000003"; exon_number "2"; oId "CUFF.1.1"; class_code "u"; tss_id "TSS2";
1	Cufflinks	exon	254240	256717	.	+	.	gene_id "XL0C_000002"; transcript_id "TC0NS_00000004"; exon_number "1"; oId "CUFF.3.1"; class_code "u"; tss_id "TSS3";
1	Cufflinks	exon	341982	343630	.	+	.	gene_id "XL0C_000003"; transcript_id "TC0NS_00000005"; exon_number "1"; oId "CUFF.10.1"; class_code "u"; tss_id "TSS4";
1	Cufflinks	exon	342113	342607	.	+	.	gene_id "XL0C_000003"; transcript_id "TC0NS_00000006"; exon_number "1"; oId "CUFF.11.1"; class_code "u"; tss_id "TSS5";
1	Cufflinks	exon	342961	343494	.	+	.	gene_id "XL0C_000003"; transcript_id "TC0NS_00000006"; exon_number "2"; oId "CUFF.11.1"; class_code "u"; tss_id "TSS5";
1	Cufflinks	exon	3312599	3313720	.	+	.	gene_id "XL0C_000024"; transcript_id "TC0NS_00000073"; exon_number "1"; oId "CUFF.75.1"; class_code "u"; tss_id "TSS37";
1	Cufflinks	exon	3446776	3447142	.	+	.	gene_id "XL0C_000024"; transcript_id "TC0NS_00000073"; exon_number "2"; oId "CUFF.75.1"; class_code "u"; tss_id "TSS37";
1	Cufflinks	exon	9347375	9347527	.	+	.	gene_id "XL0C_000047"; transcript_id "TC0NS_00000118"; exon_number "1"; oId "CUFF.115.1"; class_code "u"; tss_id "TSS63";

The first lines of a bed file containing the same information would look like this

```
1 242203 242862 XL0C_000002.1
1 242203 242646 XL0C_000002.2
1 254559 256717 XL0C_000002.3
1 254240 256717 XL0C_000002.4
1 341982 343630 XL0C_000003.1
1 342113 342607 XL0C_000003.2
1 342961 343494 XL0C_000003.3
1 3312599 3313720 XL0C_000024.1
1 3446776 3447142 XL0C_000024.2
1 9347375 9347527 XL0C_000047.1
```