Package 'JctSeqData'

October 17, 2019

Version 1.14.0

Title Example Junction Count data for use with JunctionSeq

Depends R (>= 3.3)

Suggests knitr, BiocStyle, DESeq2, DEXSeq, edgeR, JunctionSeq

Description Junction count data from an example dataset taken from a subset of the RNA-seq reads from six samples. Data was subsampled and modified to provide edge cases for testing and to reduce file sizes.

License file LICENSE

VignetteBuilder knitr

NeedsCompilation no

LazyData true

URL http://hartleys.github.io/JunctionSeq/

BugReports http://github.com/hartleys/JunctionSeq/issues

biocViews Rattus_norvegicus_Data, ExperimentData, Genome, RNASeqData, RepositoryData, GEO

Author Stephen Hartley [aut, cre] (PhD)

Maintainer Stephen Hartley <JunctionSeq-contact@list.nih.gov>

git_url https://git.bioconductor.org/packages/JctSeqData

git_branch RELEASE_3_9

git_last_commit 39064c8

git_last_commit_date 2019-05-02

Date/Publication 2019-10-17

R topics documented:

---

exampleDataSet   *Example JunctionSeq Dataset*

Description

fullExampleDataSet contains a small JunctionSeqCountSet. exampleDataSet contains an *extremely* small JunctionSeqCountSet.

Usage

```
data("exampleDataSet")
data("fullExampleDataSet")
```

Format

JunctionSeqCountSet

Value

(Not Applicable.)

Note

IMPORTANT: THIS DATASET IS INTENDED FOR TESTING ONLY.

The original dataset had 24 samples which were multiplexed and sequenced across 6 lanes. Only 6 of those samples were selected for use in this example dataset, and only the first three lanes were used. In addition, only chromosome 14 was included in its entirety. We also included a small number of genes randomly from chromosomes 15, X, and M.

A number of minor modifications were made to the dataset to provide artificial edge cases for code testing.

The fullExampleDataSet dataset contains 14851 counting bins, 11365 of which are testable (ie. nonzero). Of these, 6154 are exonic regions, 5072 are known splice junctions, and 139 are novel splice junctions.

The exampleDataSet dataset is cut down even further to provide an extremely small dataset for testing purposes. It contains 1464 counting bins, 1103 of which are testable (ie. nonzero). Of these, 600 are exonic regions, 492 are known splice junctions, and 11 are novel splice junctions. These features span 58 genes.

NONE of these datasets are intended to be representative of the original data.

IMPORTANT: THIS DATASET IS INTENDED FOR TESTING ONLY.

Source

A complete and comprehensive analysis pipeline walkthrough that includes the generation of this dataset can be found online here.

The bam files can be found here, and the walkthrough with processed results can be found here

The original, full-size datasets are available on the NCBI Gene Expression Omnibus (GEO), accession number GSE63309.

References

Hartley, S. W., Coon, S. L., Savastano, L. E., Mullikin, J. C., Fu, C., Klein, D. C., & NISC Comparative Sequencing Program. (2015). *Neurotranscriptomics: The Effects of Neonatal Stimulus Deprivation on the Rat Pineal Transcriptome*. PloS one, 10(9), e0137548.

Examples

```
library("JunctionSeq");

#The smaller example dataset was generated using the code:
decoder.file <- system.file(
        "extdata/annoFiles/decoder.bySample.txt",
        package="JctSeqData");
decoder <- read.table(decoder.file,
        header=TRUE,
        stringsAsFactors=FALSE);
gff.file <- system.file(
        "extdata/tiny/withNovel.forJunctionSeq.gff.gz",
        package="JctSeqData");
countFiles <- system.file(paste0("extdata/tiny/",
        decoder$sample.ID,
        "/QC.spliceJunctionAndExonCounts.withNovel.forJunctionSeq.txt.gz"),
        package="JctSeqData");
jscs <- runJunctionSeqAnalyses(sample.files = countFiles,
        sample.names = decoder$sample.ID,
        condition=factor(decoder$group.ID),
        flat.gff.file = gff.file,
        analysis.type = "junctionsAndExons");
```

Index