BgeeCall Package
Expression calls from RNA-Seq data

Sara Fonseca Costa

Bioconductor Meeting
2019
Bgee database

Bgee is a database to retrieve and compare gene expression patterns in multiple animal species, produced from multiple data types (RNA-Seq, Affymetrix, in situ hybridization, and EST data) and from multiple data sets (including GTEx data).

SIMPLY NORMAL
Bgee is based exclusively on curated "normal", healthy, expression data (e.g., no gene knock-out, no treatment, no disease), to provide a comparable reference of normal gene expression.

COMPARABLE BETWEEN SPECIES
Bgee produces calls of presence/absence of expression, and of differential over-/under-expression, integrated along with information of gene ontology, and of homology between organs. This allows comparisons of expression patterns between species.
Which problems we want to solve with our approach?

1) Avoid arbitrary TPM thresholds to call expressed genes in RNA-Seq data.
2) Deal with bad genome annotations for some non-model organisms.
Method used to call expression genes in RNA-Seq data

Sara Fonseca Costa

BgeeCall Package

Bioconductor Meeting 2019
Looking in particular model and non-model organism

Mus musculus (133 libraries)

Macaca mulatta (90 libraries)

Mus musculus

Macaca Mulata

Intergenic TPM of 1% Intergenic TPM of 10% Intergenic TPM of 5% No deconvolution TPM >= 2

Intergenic TPM of 1% Intergenic TPM of 10% Intergenic TPM of 5% No deconvolution TPM >= 2
BgeeCall: reference intergenic regions

Proportion of N’s per species
Results

Call expressed genes for different species using TPM threshold

Call expressed genes for different species using reference intergenic regions
Results - GTEx data - TMP threshold vs Intergenic regions
Summary about the BgeeCall

BgeeCall package:

- is easy to use

```r
# init objects
user <- new("UserMetadata")

# edit attributes of the UserMetadata object
user@species_id <- "7227" # D. melanogaster
user <- setAnnotationFromPath(user, "path/to/annotation", "annot_name")
user <- setTranscriptomeFromPath(user, "path/to/transcriptome", "transcript_name")
user <- setRNASEqLibPath(user, "path/to/library/directory")

# generate present/absent calls
calls <- run_from_object(myUserMetadata = user)
```

- is highly tunable
- allow to call expressed genes from RNA-Seq data by using a robust statistical approach compared with standards cut-offs.
Summary about the BgeeCall

- can generate calls of present/absent genes for RNA-Seq data that correspond to 29 different species in Bgee by using Bgee reference intergenic regions.

```r
list_intergenic_release()
# create BgeeMetadata object and define one reference intergenic release
bgee <- new("BgeeMetadata", intergenic_release = "0.1")
# change the reference intergenic release of your BgeeMetadata object
bgee <- setIntergenicRelease(bgee, "0.2")

list_bgee_ref_intergenic_species(myBgeeMetadata = bgee)
```

- can use reference intergenic regions from species not established in Bgee through ZENODO.

```r
list_community_ref_intergenic_species()
## create a BgeeMetadata object using the community release
bgee <- new("BgeeMetadata", release = "community")
calls_output <- generate_calls_workflow(bgeeMetadata = bgee, userMetadata = user_BgeeCall)
```

- or you still can provide your own reference intergenic region

```r
bgee <- new("BgeeMetadata", release = "community", custom_intergenic_path = "path/to/custom/ref_intergenic.fa.gz")
```
Acknowledgment

Julien Wollbrett [aut] & Bgee team

Please visit my poster if you have questions!