

Expression calls from RNA-Seq data using BgeeCall R package



Sara Fonseca Costa^{1,2} & Bgee Team

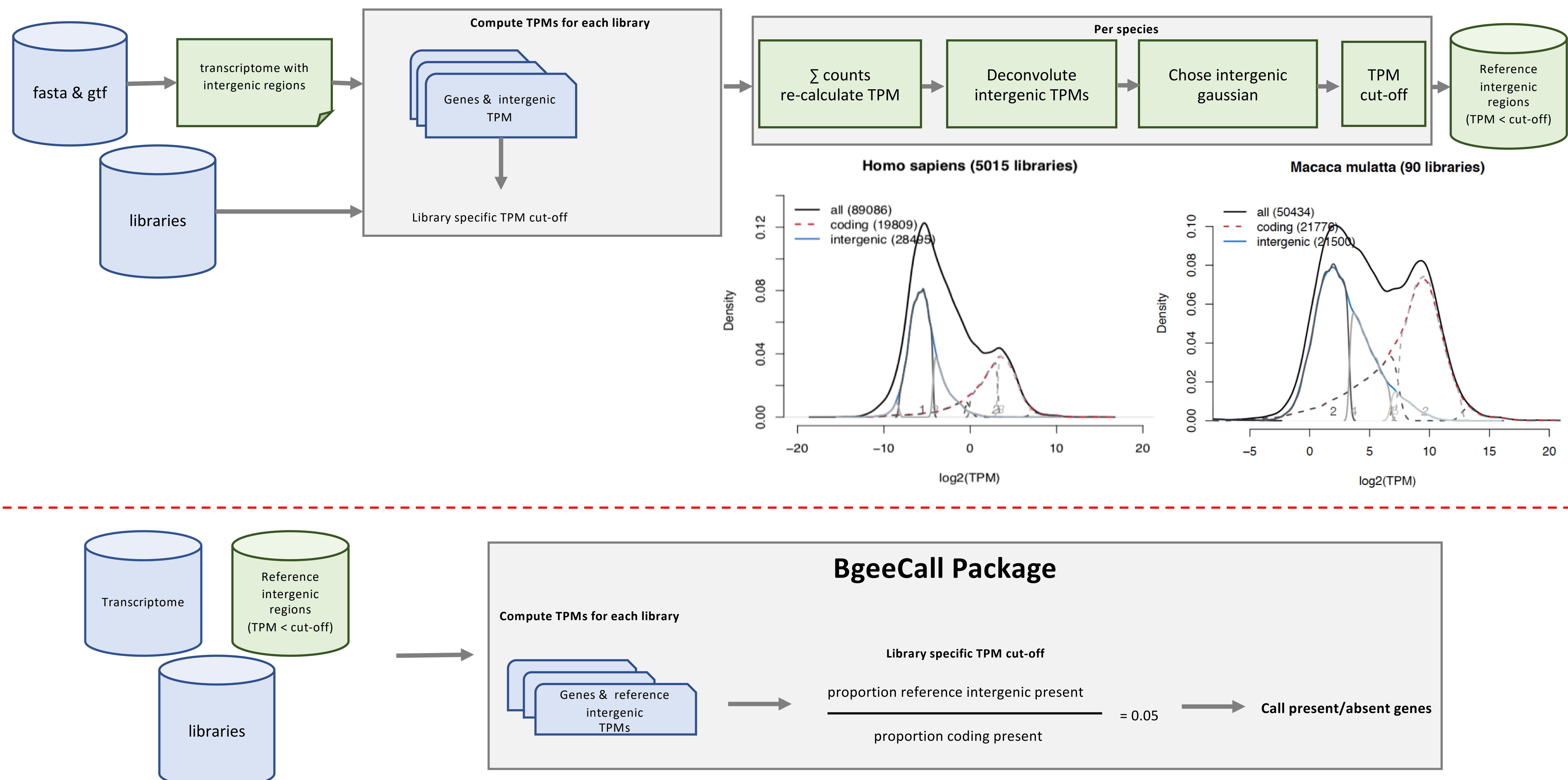
¹ Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland.

² SIB Swiss Institute of Bioinformatics, Quartier Sorge, Bâtiment Génopode, 1015 Lausanne, Switzerland

BgeeCall allows researchers to classify in an accurate way the presence or absence of gene expression from their own RNA-Seq data. The approach integrated in the BgeeCall to make this classification is based on the set of reference intergenic regions.

Workflow

Use reference intergenic sequences to distinguish signal from noise in your RNA-Seq libraries



Summary

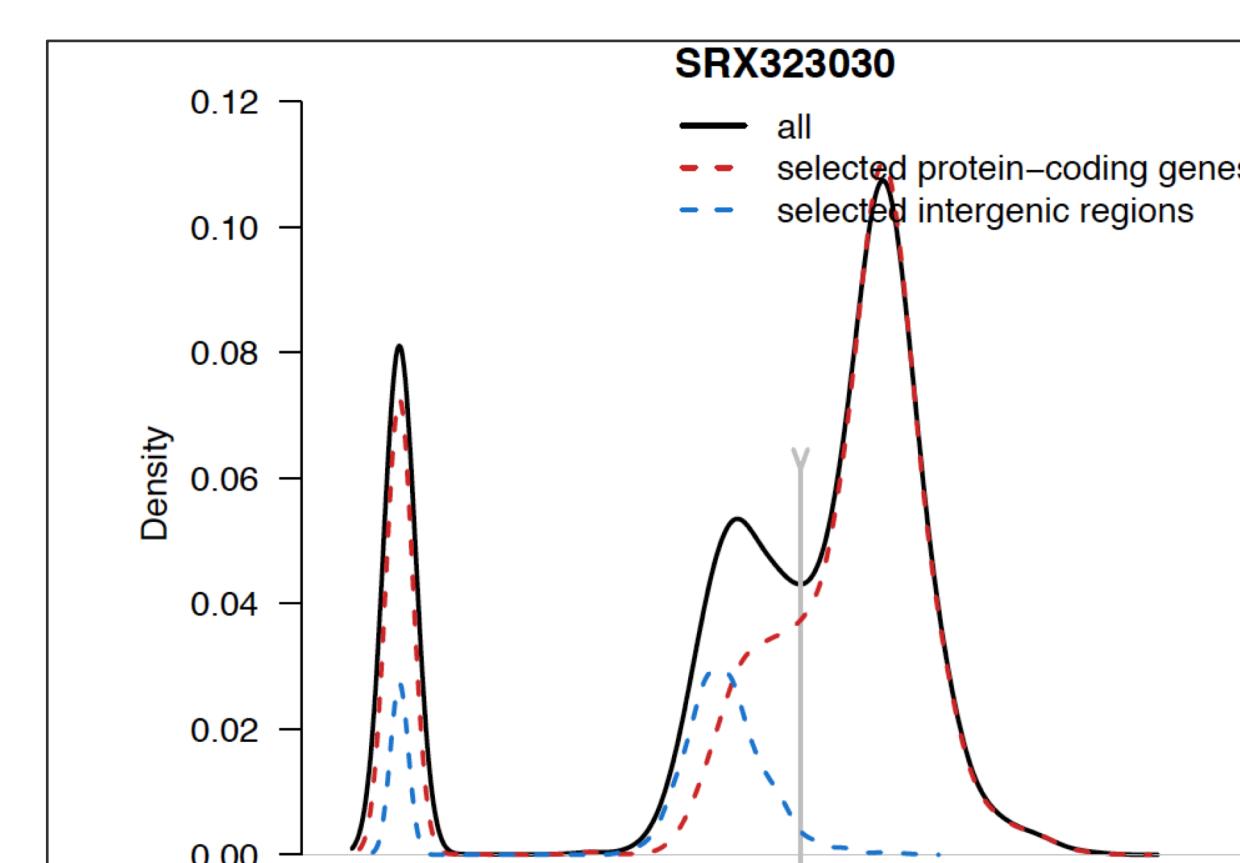
How to run BgeeCall

Use the reference intergenic regions from Bgee

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

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

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



1	libraryId	SRX323030
2	cutoffTPM	0.839872
3	proportionGenicPresent	65.9964726631393
4	numberGenicPresent	9355
5	numberGenic	14175
6	proportionCodingPresent	66.7912056329932
7	numberPresentCoding	9296
8	numberCoding	13918
9	proportionIntergenicPresent	3.30639235855988
10	numberIntergenicPresent	90
11	numberIntergenic	2722
12	ratioIntergenicCodingPresent	0.05

id	abundance	counts	length	biotype	type	call
FBgn0000008	7.518832	943.5161	4916.5497782037	protein_coding	genic	present
FBgn0000014	0.0382818	5.01889	5136.61	protein_coding	genic	absent
FBgn0000015	0.0526116	5.00835	3729.70591382889	protein_coding	genic	absent
FBgn0000017	27.7385471	5745.663	8115.52610622933	protein_coding	genic	present
FBgn0000018	2.26879	92	1588.75	protein_coding	genic	present

Use the reference intergenic regions from the community

The screenshot shows the Zenodo interface with a dataset titled 'Intergenic sequences' for 'Solenopsis invicta'. It includes a search bar, upload button, and communities section. The dataset details are: uploaded on September 19, 2019 (v1), dataset, open access, created by Carlos Martinez-Ruiz and Yannick Wurm, Bgee team, and part of the Solenopsis ATLAS project. A red arrow points from this section to the command-line code below.

```
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)
```