

Browsing and searching the Bioconductor codebase

BioC Developers' Forum - 20/08/2020

Mike Smith

  @grimbough

History lesson

- Bioconductor used to use Subversion (SVN) for version control
 - <https://hedgehog.fhcrc.org/bioconductor/trunk/madman/Rpacks/mypackage/>
- Git <-> subversion bridge (I never used this)
 - Local git, remote SVN
- Mid 2017 started the transition to `git.bioconductor.org`
- SVN server retired (2018?)

Subversion + git

- Subversion was mirrored to a read-only Github collection
- Both confusing and very helpful
 - You couldn't actually use the Github repo
 - But searching, browsing etc was available
- Easy to checkout all packages
 - <https://hedgehog.fhcrc.org/bioconductor/trunk/madman/Rpacks/mypackage/>
 - Far fewer packages!
 - Browse & search locally

Current status

- People have been asking for a way to view git server forever
 - e.g BioC2020 BOF, emails from Feb 2018, ...
- Many packages are on Github
 - Some are active, others only for submission
 - Not easy to know if you're looking at 'current' code
- Hard (impossible?) to search for across all packages
 - e.g. every package calling a specific function

<http://bioc-code-tools.msmith.de/>

- Prototype website to provide browsing & searching of all BioC packages
- Git browser using Gitlist
 - <https://github.com/patrikx3/gitlist/>
- Code search using Zoekt
 - <https://github.com/google/zoekt>
- Currently hosted on de.NBI cloud & my domain for convenience
- Clone of (most) BioC git repos, updates every 2 hours.

Show 25 entriesSearch:

Package	Latest Commit
tximeta	2020-08-20 12:18:32 by mikelove to master ' <i>fix readme</i> '
MouseFM	2020-08-20 10:07:13 by matmu to master ' <i>illegal version bump</i> '
Qtlizer	2020-08-20 09:23:17 by Matthias Munz to master ' <i>Update DESCRIPTION</i> '
batchelor	2020-08-20 04:28:33 by LTLA to master ' <i>Eliminate need for batch= when design= is set in regressBatches.</i> '
MSstats	2020-08-20 02:05:18 by MeenaChoi to RELEASE_3_11 ' <i>fix error</i> '
cBioPortalData	2020-08-19 23:04:26 by LiNk-NY to master ' <i>version bump</i> '
flowFP	2020-08-19 18:37:23 by Herb Holyst to master ' <i>Fixing compatibility with R 4.0</i> '
rWikiPathways	2020-08-19 18:17:15 by Alexander Pico to master ' <i>release date fix</i> '
flowCL	2020-08-19 17:35:22 by jmeskas to master ' <i>changed maintainer's email</i> '
BiocCheck	2020-08-19 17:31:04 by Martin Morgan to master ' <i>mention instructions for removing large files from git</i> '
rcellminer	2020-08-19 14:48:58 by Elloumi to RELEASE_3_11 ' <i>Merged origin master with local master</i> '
ENmix	2020-08-19 14:42:11 by Zongli Xu to master ' <i>bugfix</i> '
POMA	2020-08-19 13:37:44 by pcastellanoescuder to master ' <i>authors updated</i> '
InTAD	2020-08-19 07:18:11 by Konstantin Okonechnikov to master ' <i>Include correlations from loops: step2</i> '
signatureSearch	2020-08-19 06:42:49 by Yuzhu Duan to RELEASE_3_11 ' <i>note msigdb as special case, jump to 1.3.5</i> '



browsing: **master** ▾

 Files

 Commits

 Log

 Graph











 Stats

tximeta



RSS

Download ▾

 R	fix typo	5 weeks ago
 inst	adding GENCODE 35 and Ensembl 101	46 minutes ago
 man	fix typo	5 weeks ago
 tests	fix alevin skipMeta bug	3 months ago
 vignettes	adding tx_ids after summarization	2 months ago
 .gitignore	42B update to gencode 30/M21, ensembl 96	1 year, 3 months ago
 CODE_OF_CONDUCT.md	3kB add code of cond. and package man page	9 months ago
 DESCRIPTION	1kB adding GENCODE 35 and Ensembl 101	46 minutes ago
 NAMESPACE	2kB Add splitSE function	5 months ago
 NEWS	6kB adding tx_ids after summarization	2 months ago
 README.md	7kB fix readme	41 minutes ago



Search tree...

browsing: **master** ▾

Files

Commits

Log

Graph

★ Stats

tximeta / DESCRIPTION

Raw

Blame

History

RSS

Download ▾

Scroll mode

Full height

```
1 Package: tximeta
2 Version: 1.7.8
3 Title: Transcript Quantification Import with Automatic Metadata
4 Description: Transcript quantification import from Salmon and
5   alevin with automatic attachment of transcript ranges and
6   release information, and other associated metadata. De novo
7   transcriptomes can be linked to the appropriate sources with
8   linkedTxomes and shared for computational reproducibility.
9 Authors@R: c(
10   person("Michael", "Love", email="michaelisaiahlove@gmail.com", role = c("aut","cre")),
11   person("Charlotte", "Soneson", role = c("aut","ctb")),
12   person("Peter", "Hickey", role = c("aut","ctb")),
13   person("Rob", "Patro", role = c("aut","ctb")))
14 Maintainer: Michael Love <michaelisaiahlove@gmail.com>
15 License: GPL-2
16 VignetteBuilder: knitr
17 Imports:
18   SummarizedExperiment, tximport, ica, lite, S4Vectors
```


browsing: **master** ▾

Files

Commits

Log

Graph

★ Stats

tximeta / Commit 36b8ee0b5f8381845dd3a676d4f1c51f52803bea

File changed: 1



RSS

Download ▾

Browse code



fix readme

mikelove authored on 20/08/2020 12:18:32

README.md

Diff lines: 11 Type: text



Diff



History



View

Old	---	a/README.md
New	+++	b/README.md
...	...	@@ -5,8 +5,8 @@ For a reference and citation for `tximeta`:
5	5	> Michael I. Love, Charlotte Sonesson, Peter F. Hickey, Lisa K. Johnson,
6	6	> N. Tessa Pierce, Lori Shepherd, Martin Morgan, Rob Patro.
7	7	> "Tximeta: reference sequence checksums for provenance
8		-> identification in RNA-seq" *bioRxiv*, September 2019.
9		-> [doi: 10.1101/777888](https://doi.org/10.1101/777888)
8		+> identification in RNA-seq" *PLOS Computational Biology* (2020)
9		+> [doi: 10.1371/journal.pcbi.1007664](https://doi.org/10.1371/journal.pcbi.1007664)
10	10	
11	11	# Idea in diagram
12	12	

Still a work in progress

- Definitely still a prototype
 - Don't bookmark the URL yet!
- Would love to get feedback on whether these seem like a useful tool people would use
- Also interested in if things break, several other tools failed when scaling from 20 packages to 2000