Bioconductor

Software for orchestrating high-throughput biological data analysis

July 19, 2024,
BigCare 2024
UC Irvine

Sean Davis, MD, PhD
@seandavis12
https://seandavi.github.io

Bioconductor

*Software **& Community** for orchestrating high-throughput biological data analysis*

July 19, 2024,
BigCare 2024
UC Irvine

Sean Davis, MD, PhD
@seandavis12
https://seandavi.github.io

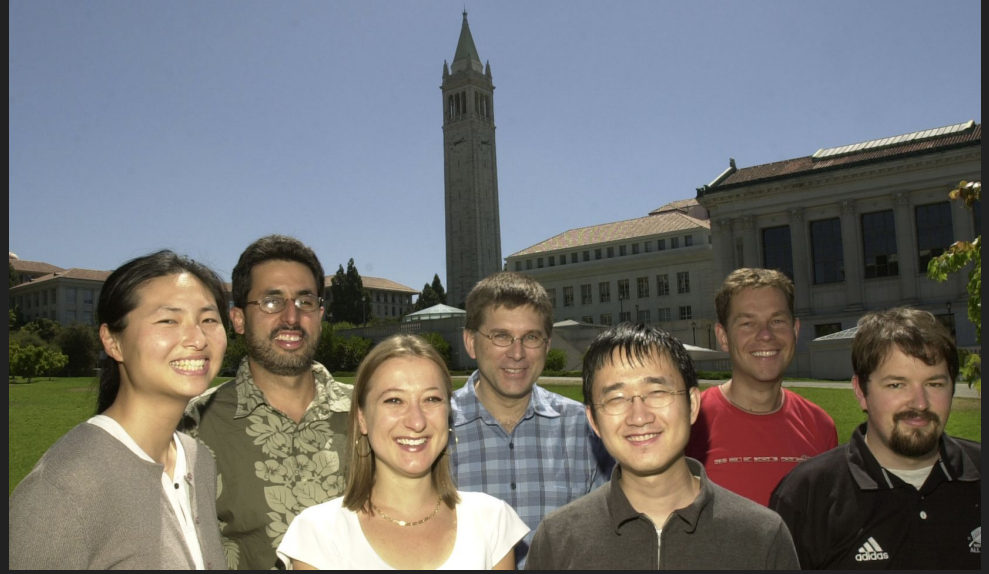
software

encodes

knowledge

Bioconductor is a large, NIH-funded open source software community dedicated to the analysis and comprehension of high throughput biological data.

2003: Changing the way genomics is done - Jean Yang, Rafa Irizarry, Sandrine Dudoit, Rob Gentleman, Cheng Li, Wolfgang Huber, Ben Bolstad -- foundations for:



2023+ Broadening competencies for **inclusive, collaborative science** worldwide, so more voices are heard and insights cultivated

VC

Bioconductor by the numbers

- Project started in 2002
- Approximately 100,000 unique downloads per month
- More than 22,000 PubmedCentral citations
- Bioconductor Release:
  - 2217 Software packages
  - 926 Annotation packages
  - 430 ExperimentHub packages
- Receiving submissions of 3-6 new packages *per week*
- More than 1000 active developers

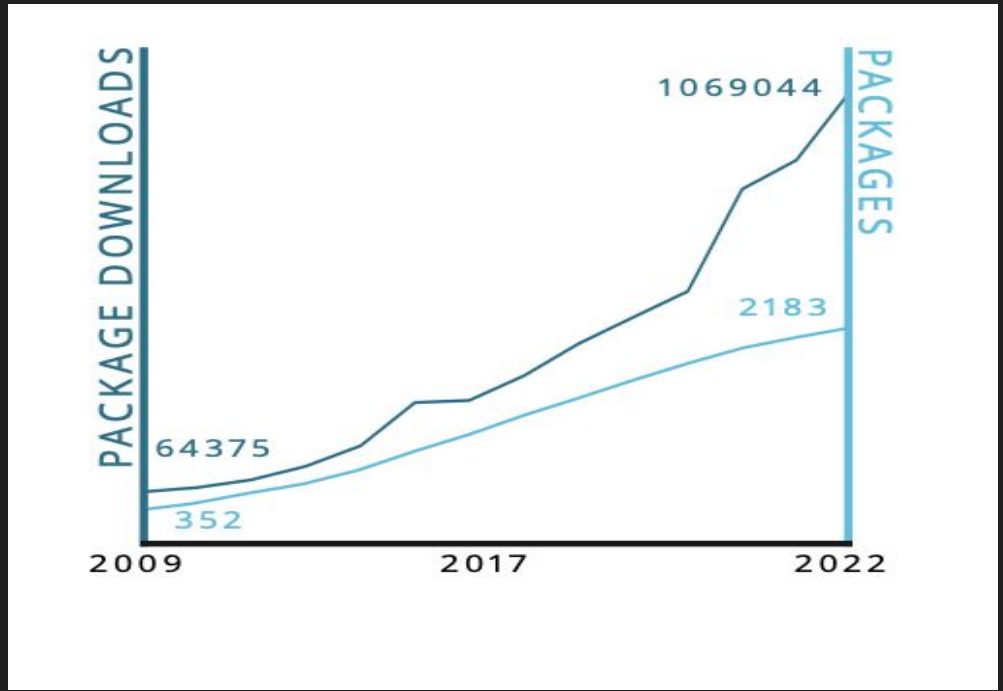
Just some of the many
Bioconductor available packages!!

Code Usage and Contributions

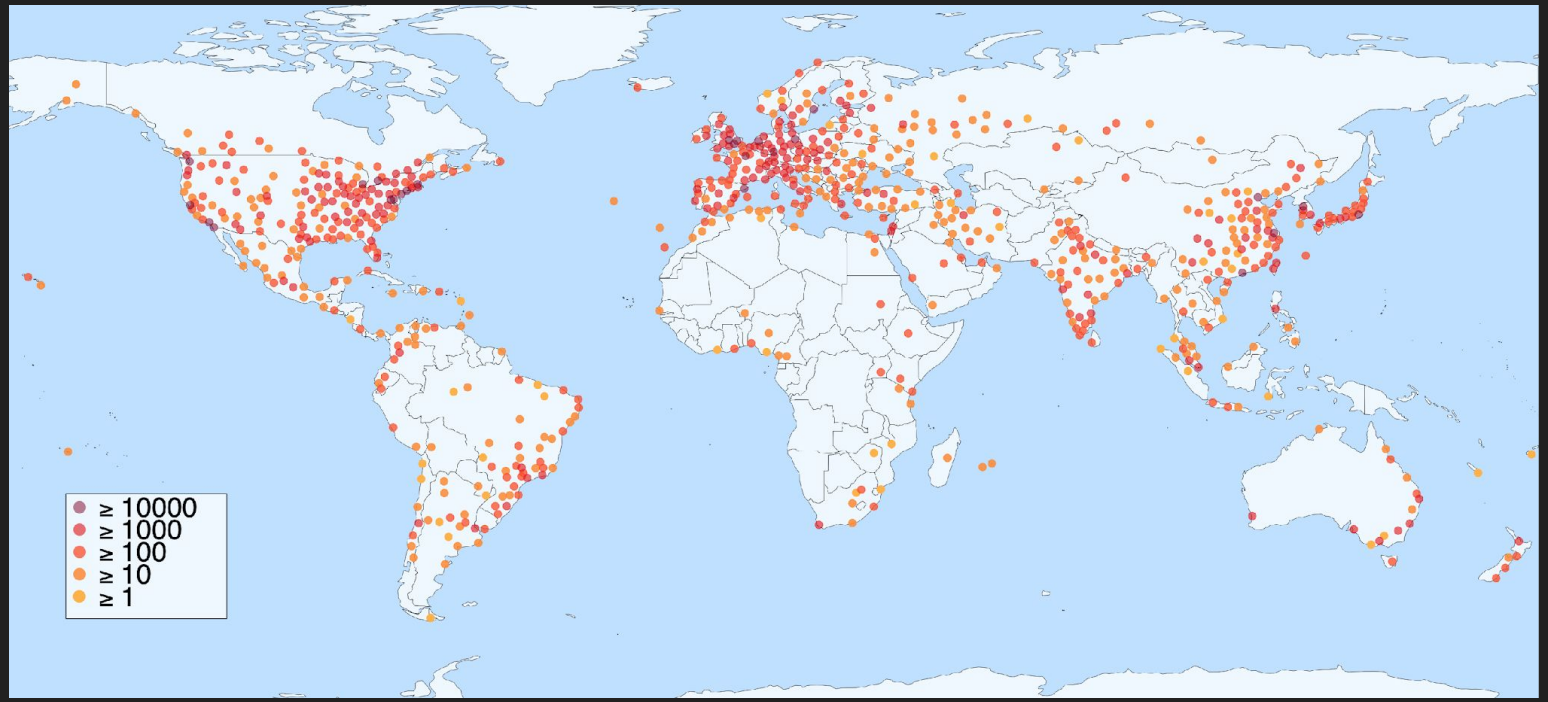
Software Download Rates
Indicator of usage and demand.

Community Code Contributions
Package contributions indicate collaborative development and innovation.

Challenges
Capturing the full picture. For example, downloads from mirrors



Community growth



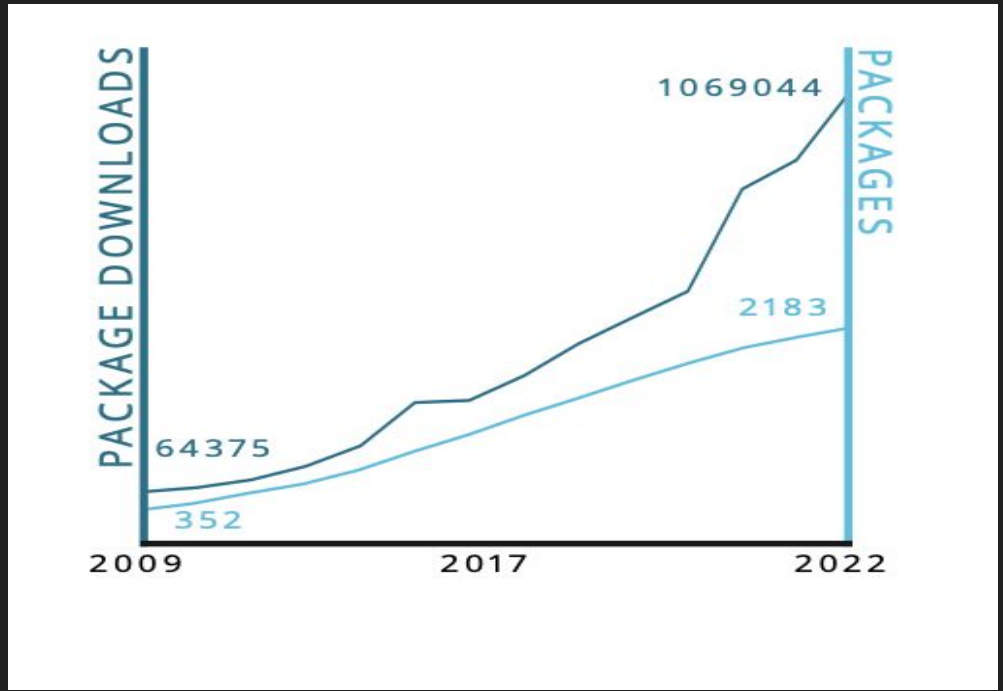
100,000s users, clustered by location

Code Usage and Contributions

Software Download Rates
Indicator of usage and demand.

Community Code Contributions
Package contributions indicate collaborative development and innovation.

Challenges
Capturing the full picture. For example, downloads from mirrors



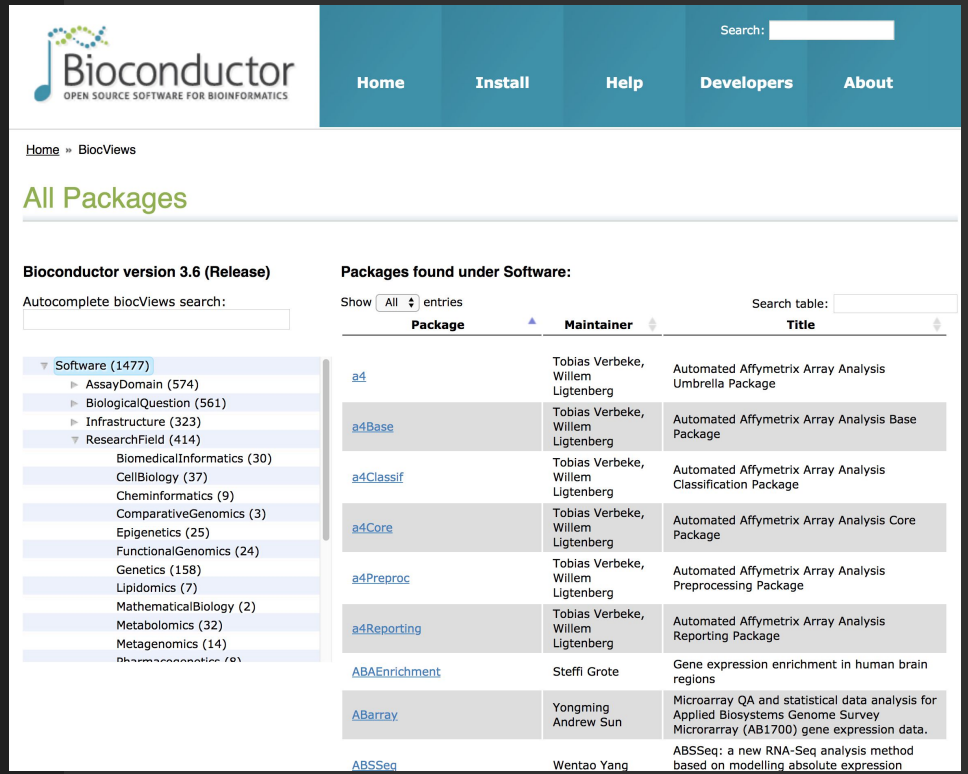
Capabilities

Bioconductor supports all phases of the biomedical data science workflow

- Specialized data import and export
- Data management leveraging FAIR data principles
- Data integration and interpretation, including access to millions of “public” datasets
- Context-specific analysis and statistics
- Visualization and publication-quality plotting
- Efficiency and reproducibility (human in the loop), even at scale
- Reproducible, literate reporting

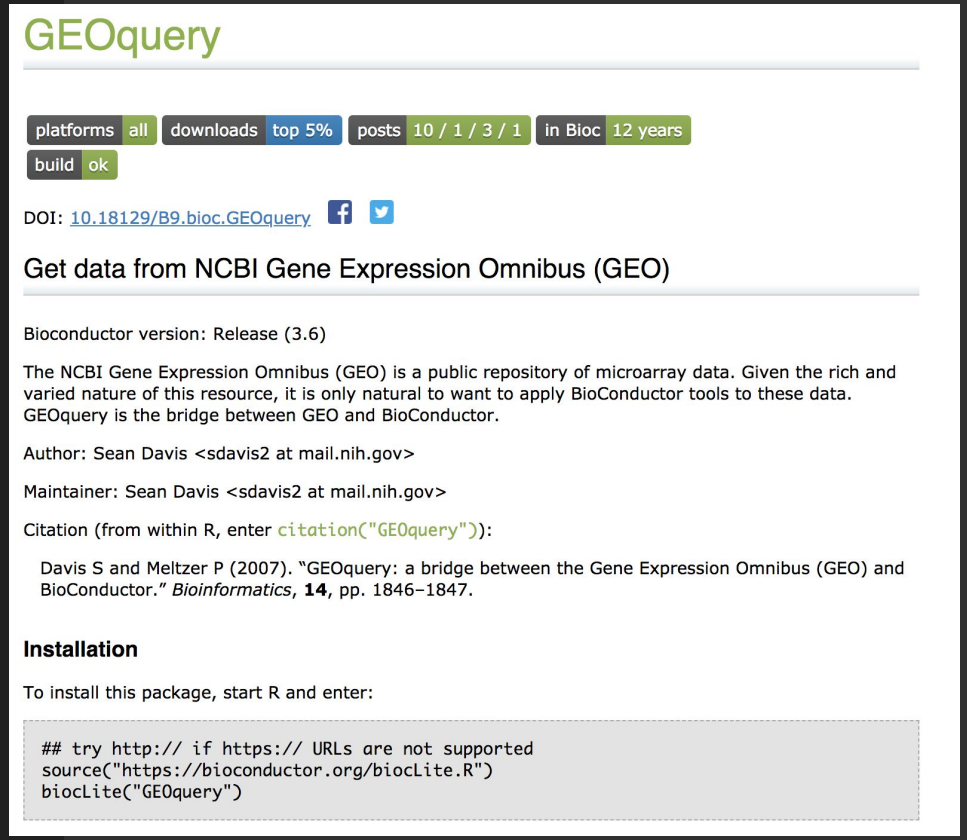
Qualities

- *Discoverable*
- Installable
- Reliable
- Documented
- Supported
- Integrated
- Connected
- Scalable
- State-of-the-art
- Community-driven



Qualities

- Discoverable
- *Installable*
- Reliable
- Documented
- Supported
- Integrated
- Connected
- Scalable
- State-of-the-art
- Community-driven



GEOquery

platforms all | downloads top 5% | posts 10 / 1 / 3 / 1 | in Bioc 12 years | build ok

DOI: 10.18129/B9.bioc.GEOquery

Get data from NCBI Gene Expression Omnibus (GEO)

Bioconductor version: Release (3.6)

The NCBI Gene Expression Omnibus (GEO) is a public repository of microarray data. Given the rich and varied nature of this resource, it is only natural to want to apply BioConductor tools to these data. GEOquery is the bridge between GEO and BioConductor.

Author: Sean Davis <sdavis2 at mail.nih.gov>

Maintainer: Sean Davis <sdavis2 at mail.nih.gov>

Citation (from within R, enter `citation("GEOquery")`):

Davis S and Meltzer P (2007). "GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor." *Bioinformatics*, **14**, pp. 1846–1847.

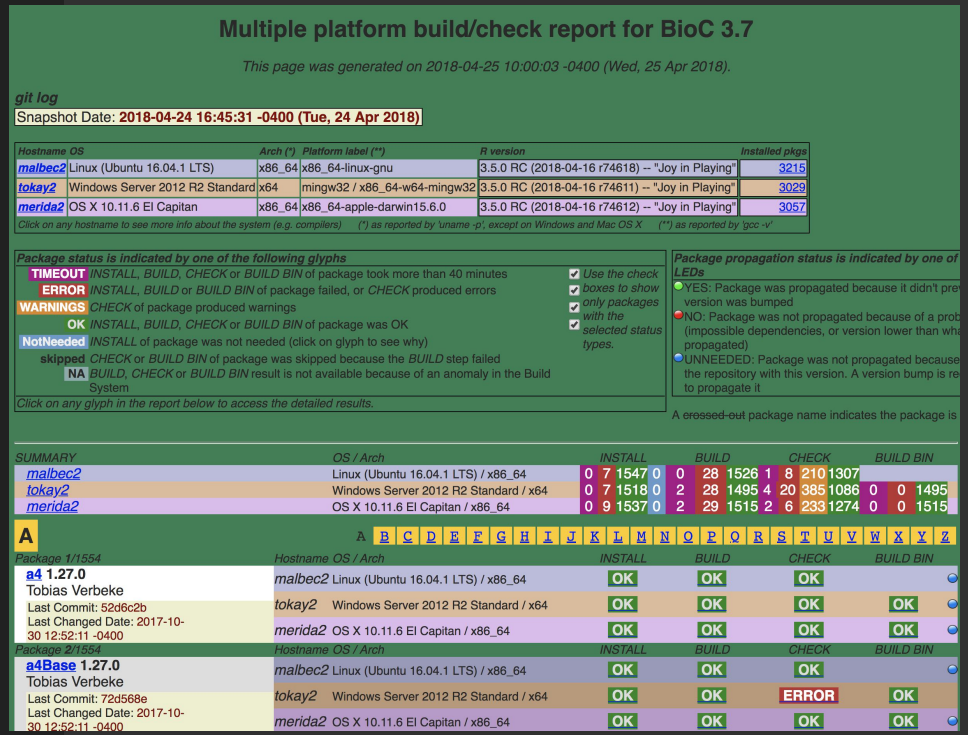
Installation

To install this package, start R and enter:

```
## try http:// if https:// URLs are not supported
source("https://bioconductor.org/biocLite.R")
biocLite("GEOquery")
```

Qualities

- Discoverable
- Installable
- *Reliable*
- Documented
- Supported
- Integrated
- Connected
- Scalable
- State-of-the-art
- Community-driven



Multiple platform build/check report for BioC 3.7

This page was generated on 2018-04-25 10:00:03 -0400 (Wed, 25 Apr 2018).

git log

Snapshot Date: **2018-04-24 16:45:31 -0400 (Tue, 24 Apr 2018)**

Hostname	OS	Arch (*)	Platform label (**)	R version	Installed pkgs
malbec2	Linux (Ubuntu 16.04.1 LTS)	x86_64	x86_64-linux-gnu	3.5.0 RC (2018-04-16 r74618) -- "Joy in Playing"	3215
tokay2	Windows Server 2012 R2 Standard	x64	mingw32 / x86_64-w64-mingw32	3.5.0 RC (2018-04-16 r74611) -- "Joy in Playing"	3029
merida2	OS X 10.11.6 El Capitan	x86_64	x86_64-apple-darwin15.6.0	3.5.0 RC (2018-04-16 r74612) -- "Joy in Playing"	3057

Click on any hostname to see more info about the system (e.g. compilers) (*) as reported by 'uname -p', except on Windows and Mac OS X (**) as reported by 'gcc -v'

Package status is indicated by one of the following glyphs

TIMEOUT *INSTALL, BUILD, CHECK* or *BUILD BIN* of package took more than 40 minutes
ERROR *INSTALL, BUILD* or *BUILD BIN* of package failed, or *CHECK* produced errors
WARNINGS *CHECK* of package produced warnings
OK *INSTALL, BUILD, CHECK* or *BUILD BIN* of package was OK
NotNeeded *INSTALL* of package was not needed (click on glyph to see why)
skipped *CHECK* or *BUILD BIN* of package was skipped because the *BUILD* step failed
NA *BUILD, CHECK* or *BUILD BIN* result is not available because of an anomaly in the Build System

Use the check boxes to show only packages with the selected status types.

Click on any glyph in the report below to access the detailed results.

Package propagation status is indicated by one of LEDs

YES: Package was propagated because it didn't pre version was bumped
NO: Package was not propagated because of a prob (impossible dependencies, or version lower than wh propagated)
UNNEEDED: Package was not propagated because the repository with this version. A version bump is re to propagate it

A ~~crossed-out~~ package name indicates the package is

SUMMARY	OS / Arch	INSTALL				BUILD				CHECK				BUILD BIN		
malbec2	Linux (Ubuntu 16.04.1 LTS) / x86_64	0	7	1547	0	0	28	1526	1	8	210	1307				
tokay2	Windows Server 2012 R2 Standard / x64	0	7	1518	0	2	28	1495	4	20	385	1086	0	0	1495	
merida2	OS X 10.11.6 El Capitan / x86_64	0	9	1537	0	2	29	1515	2	6	233	1274	0	0	1515	

A | A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Package 1/1554	Hostname	OS / Arch	INSTALL	BUILD	CHECK	BUILD BIN
a4 1.27.0 Tobias Verbeke Last Commit: 52d6c2b Last Changed Date: 2017-10-30 12:52:11 -0400	malbec2	Linux (Ubuntu 16.04.1 LTS) / x86_64	OK	OK	OK	
	tokay2	Windows Server 2012 R2 Standard / x64	OK	OK	OK	OK
	merida2	OS X 10.11.6 El Capitan / x86_64	OK	OK	OK	OK
Package 2/1554	**Hostname**	**OS / Arch**	**INSTALL**	**BUILD**	**CHECK**	**BUILD BIN**
a4Base 1.27.0 Tobias Verbeke Last Commit: 72d568e Last Changed Date: 2017-10-30 12:52:11 -0400	malbec2	Linux (Ubuntu 16.04.1 LTS) / x86_64	OK	OK	OK	
	tokay2	Windows Server 2012 R2 Standard / x64	OK	OK	ERROR	OK
	merida2	OS X 10.11.6 El Capitan / x86_64	OK	OK	OK	OK

Qualities

- Discoverable
- Installable
- Reliable
- *Documented*
- Supported
- Integrated
- Connected
- Scalable
- State-of-the-art
- Community-driven

The GenomicDataCommons Package

Sean Davis & Martin Morgan

Monday, October 30, 2017

Abstract

The National Cancer Institute (NCI) has established the Genomic Data Commons (GDC). The GDC provides the cancer research community with an open and unified repository for sharing and accessing data across numerous cancer studies and projects via a high-performance data transfer and query infrastructure. The *GenomicDataCommons* Bioconductor package provides basic infrastructure for querying, accessing, and mining genomic datasets available from the GDC. We expect that the Bioconductor developer and the larger bioinformatics communities will build on the *GenomicDataCommons* package to add higher-level functionality and expose cancer genomics data to the plethora of state-of-the-art bioinformatics methods available in Bioconductor.

Contents

1 What is the GDC?

2 Quickstart

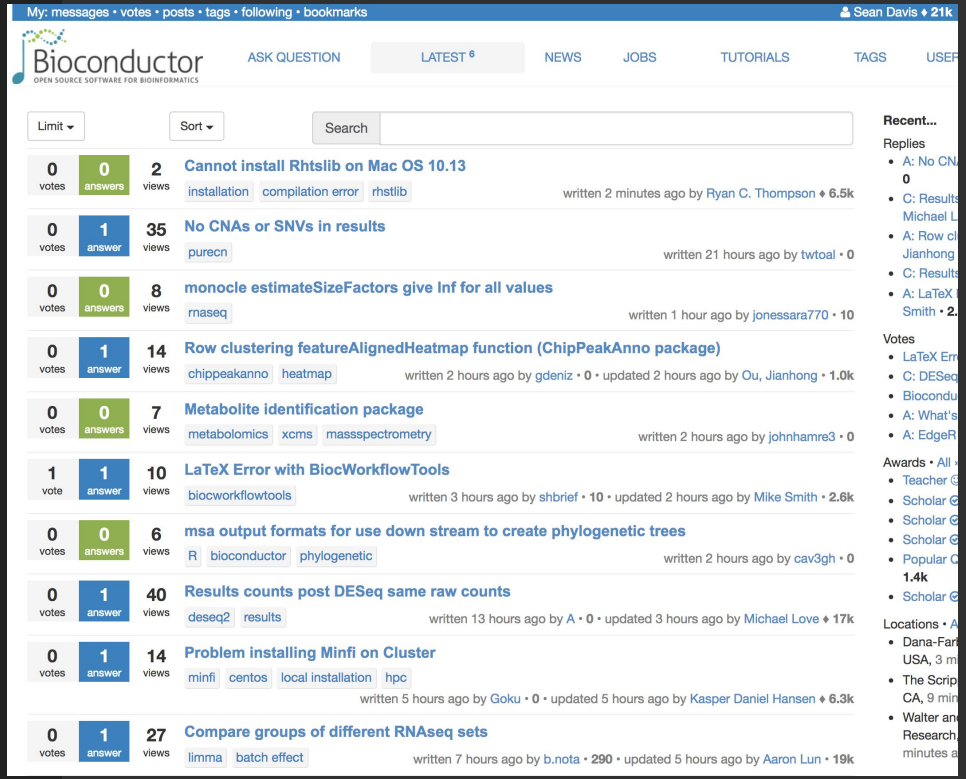
- 2.1 Installation
- 2.2 Check basic functionality
- 2.3 Find data
- 2.4 Download data
- 2.5 Metadata queries

3 Usage

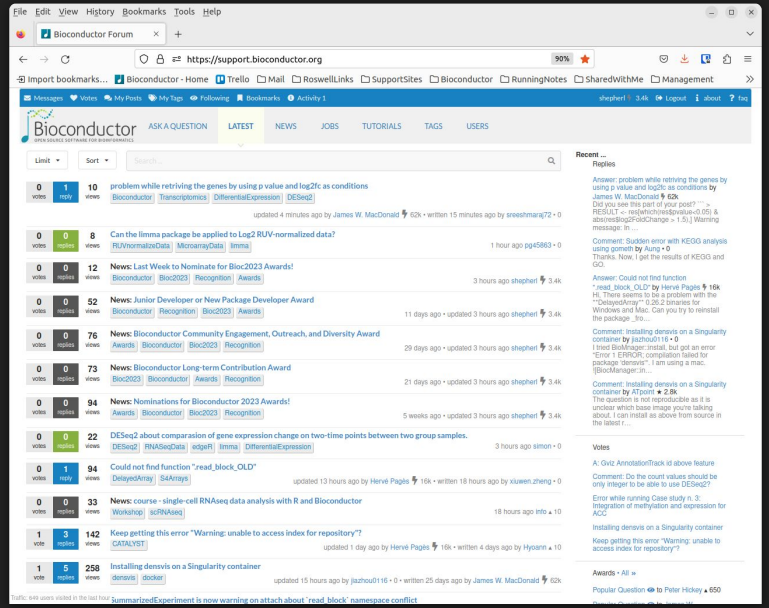
- 3.1 Querying metadata
  - 3.1.1 Creating a query
  - 3.1.2 Retrieving results
  - 3.1.3 Fields and Values
  - 3.1.4 Facets and aggregation
  - 3.1.5 Filtering
- 3.2 Authentication
- 3.3 Datafile access and download

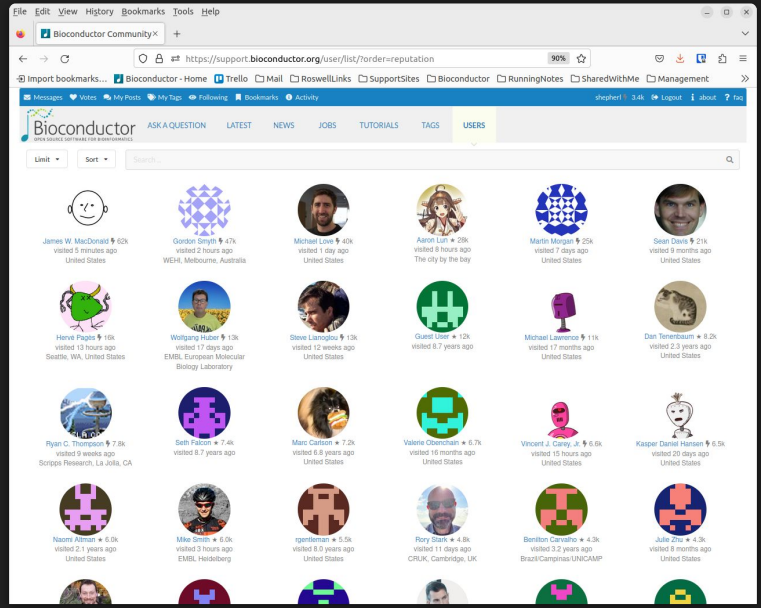
Qualities

- Discoverable
- Installable
- Reliable
- Documented
- *Supported*
- Integrated
- Connected
- Scalable
- State-of-the-art
- Community-driven



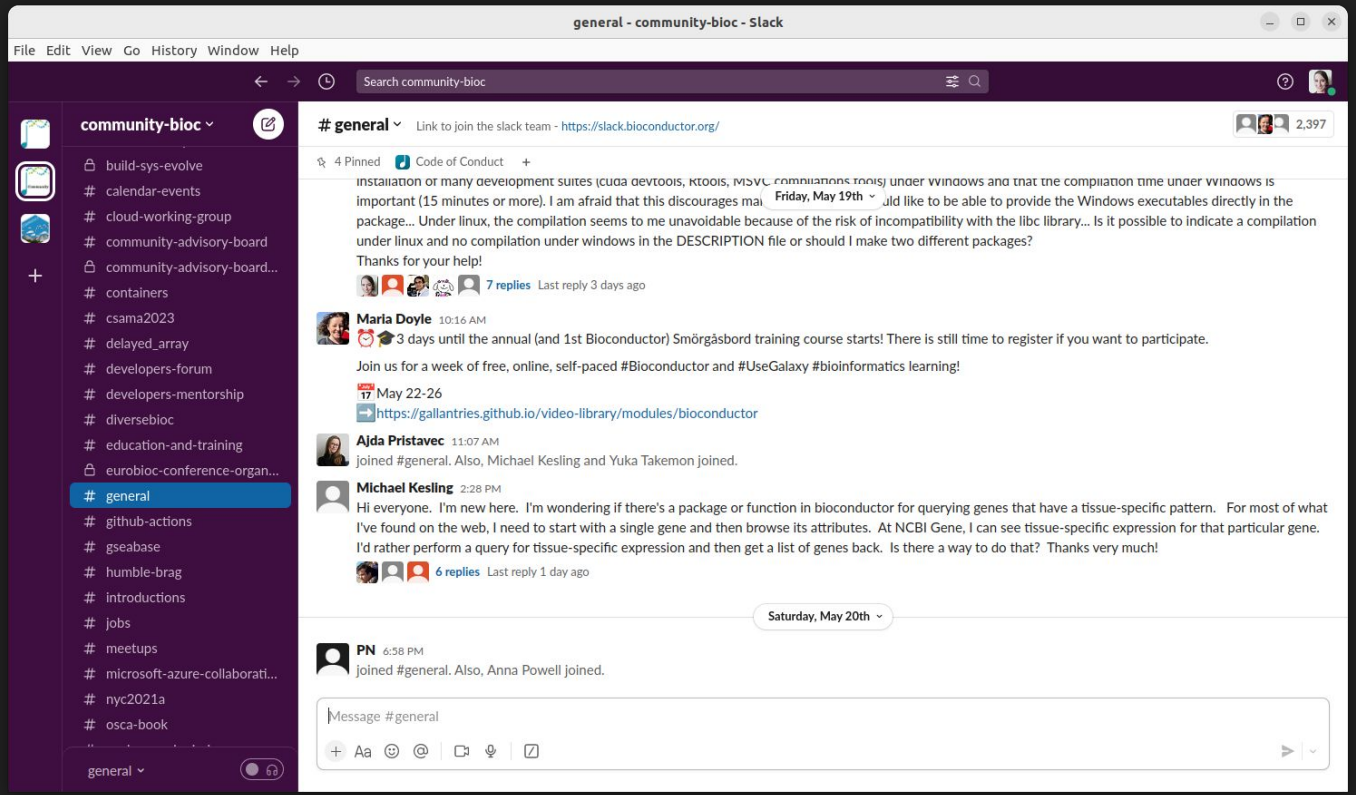
Support Site: https://support.bioconductor.org/





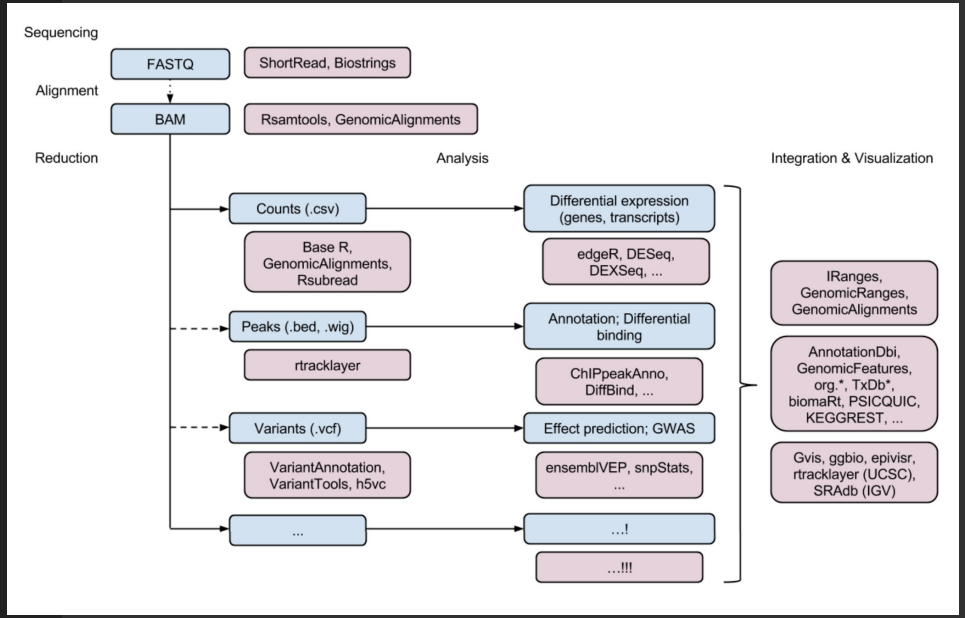
Ask questions about packages or data analysis
Thousands of Bioconductor users and maintainers are members

Community Slack: slack.bioconductor.org

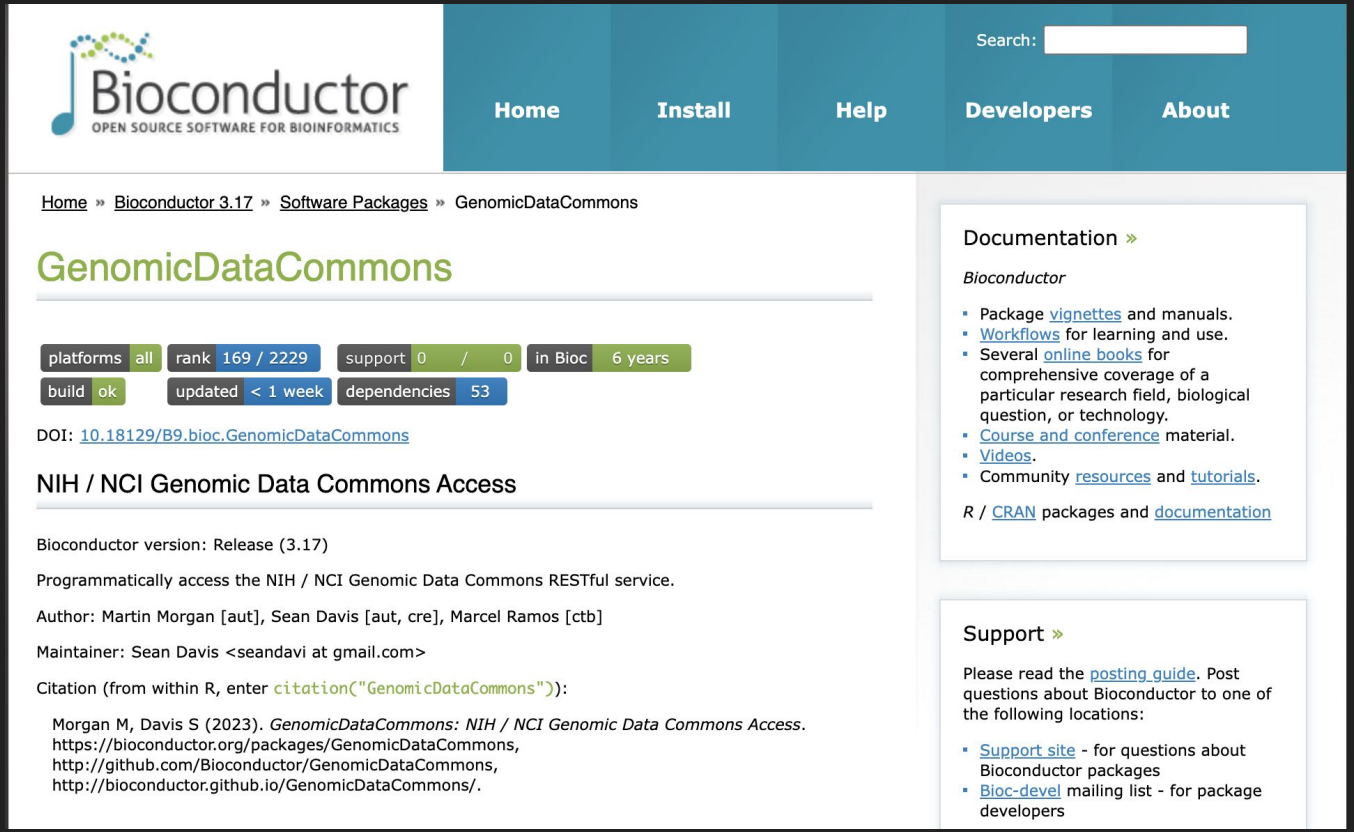


Qualities

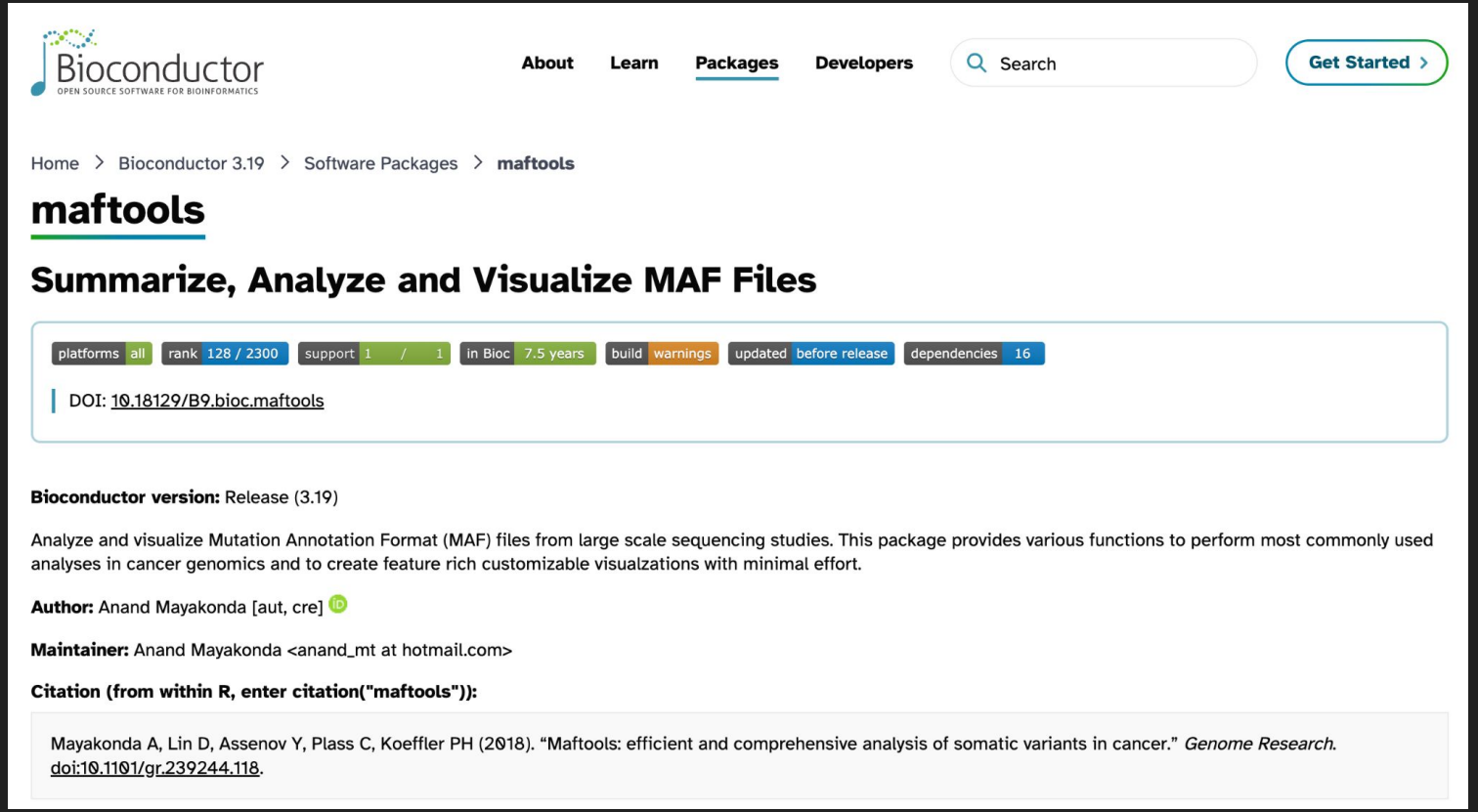
- Discoverable
- Installable
- Reliable
- Documented
- Supported
- *Integrated*
- Connected
- Scalable
- State-of-the-art
- Community-driven



Goal: Explore somatic variants seen in the TCGA cutaneous melanoma cohort in a reproducible, reusable way building on Bioconductor tools.



GenomicDataCommons


Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS
About
Learn
Packages
Developers
Search
Get Started
Home > Bioconductor 3.19 > Software Packages > maftools
maftools
Summarize, Analyze and Visualize MAF Files
platforms all
rank 128 / 2300
support 1 / 1
in Bioc 7.5 years
build warnings
updated before release
dependencies 16
DOI: 10.18129/B9.bioc.maftools
Bioconductor version: Release (3.19)
Analyze and visualize Mutation Annotation Format (MAF) files from large scale sequencing studies. This package provides various functions to perform most commonly used analyses in cancer genomics and to create feature rich customizable visualzations with minimal effort.
Author: Anand Mayakonda [aut, cre]
Maintainer: Anand Mayakonda <anand_mt at hotmail.com>
Citation (from within R, enter citation("maftools")):
Mayakonda A, Lin D, Assenov Y, Plass C, Koeffler PH (2018). "Maftools: efficient and comprehensive analysis of somatic variants in cancer." Genome Research. doi:10.1101/gr.239244.118.

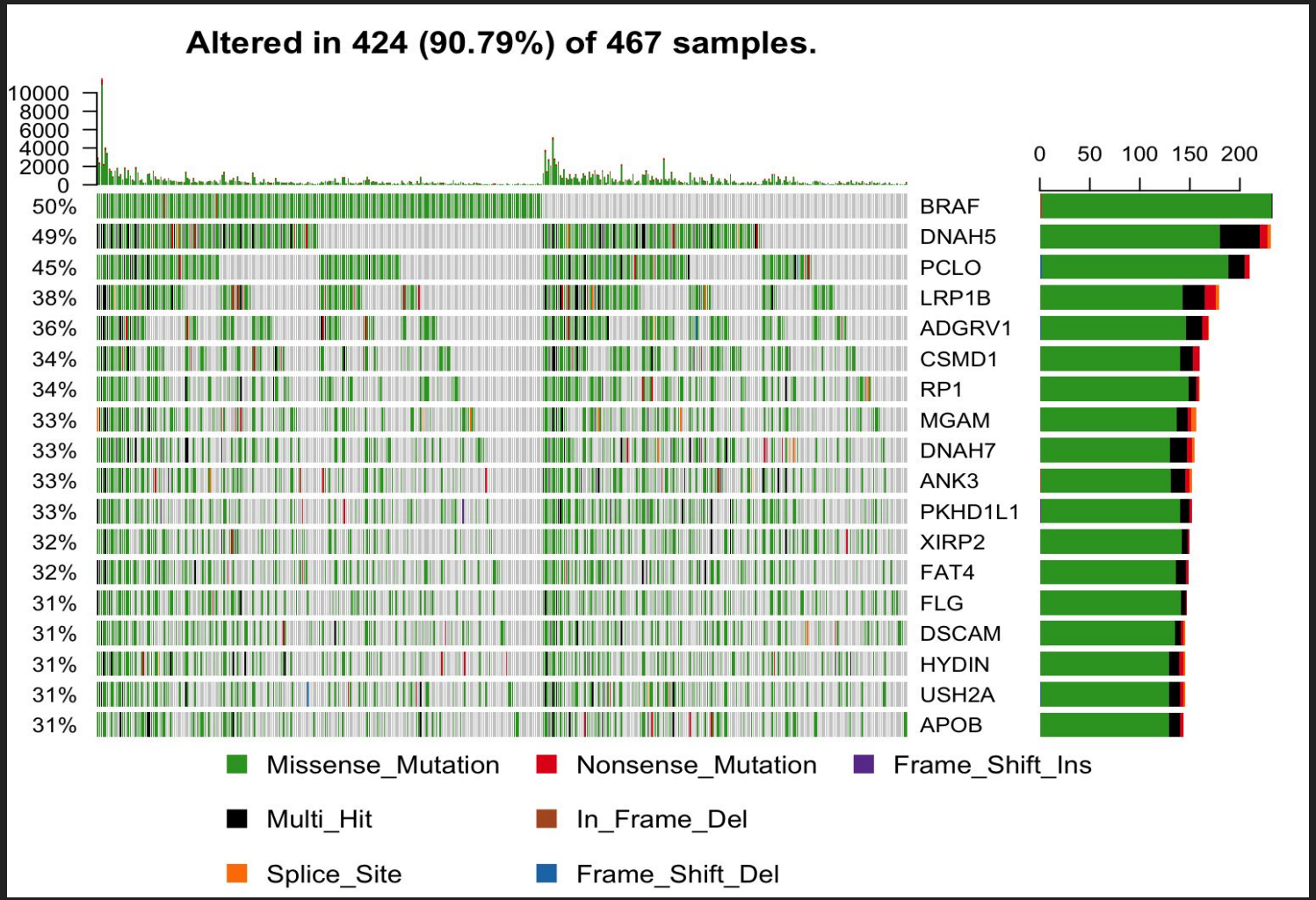
maftools

Use the GenomicDataCommons package to find and download variants from the TCGA cutaneous melanoma dataset.

```
library(GenomicDataCommons)
fnames = files() %>%
    GenomicDataCommons::filter(~ cases.project.project_id=='TCGA-SKCM' &
        data_type=='Masked Somatic Mutation' &
        data_format=='MAF' &
        analysis.workflow_type=='MuTect2 Variant Aggregation and Masking') %>%
    ids() %>%
    gdcdata()
```

And now take those data directly to maftools for analysis and visualization.

```
library(maftools)
melanoma = read.maf(maf = fnames[1])
```

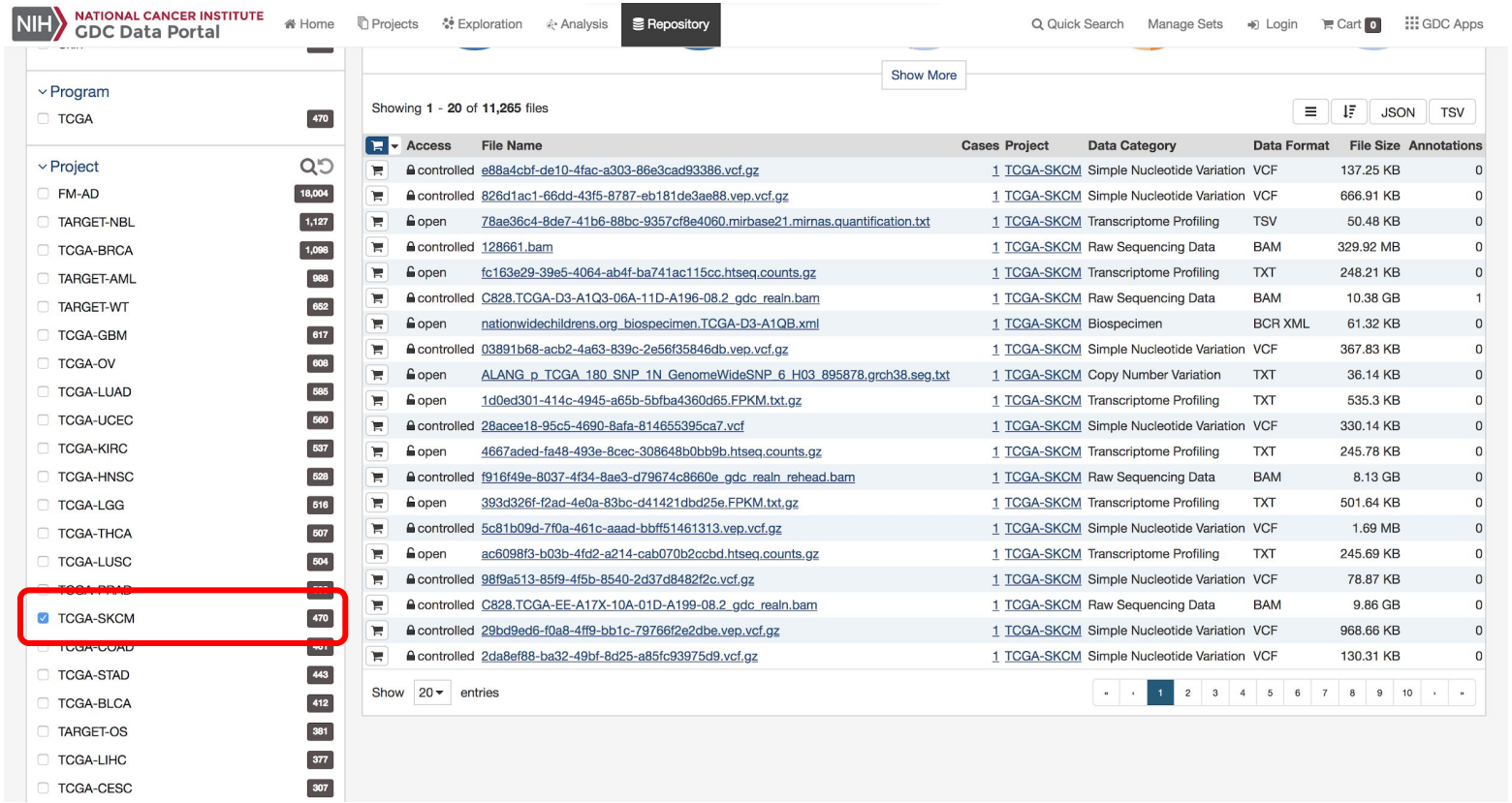

Altered in 424 (90.79%) of 467 samples.
10000
8000
6000
4000
2000
0
0
50
100
150
200
50%
49%
45%
38%
36%
34%
34%
33%
33%
33%
33%
32%
32%
31%
31%
31%
31%
31%
BRAF
DNAH5
PCLO
LRP1B
ADGRV1
CSMD1
RP1
MGAM
DNAH7
ANK3
PKHD1L1
XIRP2
FAT4
FLG
DSCAM
HYDIN
USH2A
APOB
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Ins
Multi_Hit
In_Frame_Del
Splice_Site
Frame_Shift_Del

Use the GenomicDataCommons package to find and download variants from the TCGA cutaneous melanoma dataset.

```
library(GenomicDataCommons)
fnames = files() %>%
    GenomicDataCommons::filter(~ cases.project.project_id=='TCGA-SKCM' &
        data_type=='Masked Somatic Mutation' &
        data_format=='MAF' &
        analysis.workflow_type=='MuTect2 Variant Aggregation and Masking') %>%
    ids() %>%
    gdcdata()
```

And now take those data directly to maftools for analysis and visualization.

```
library(maftools)
melanoma = read.maf(maf = fnames[1])
```



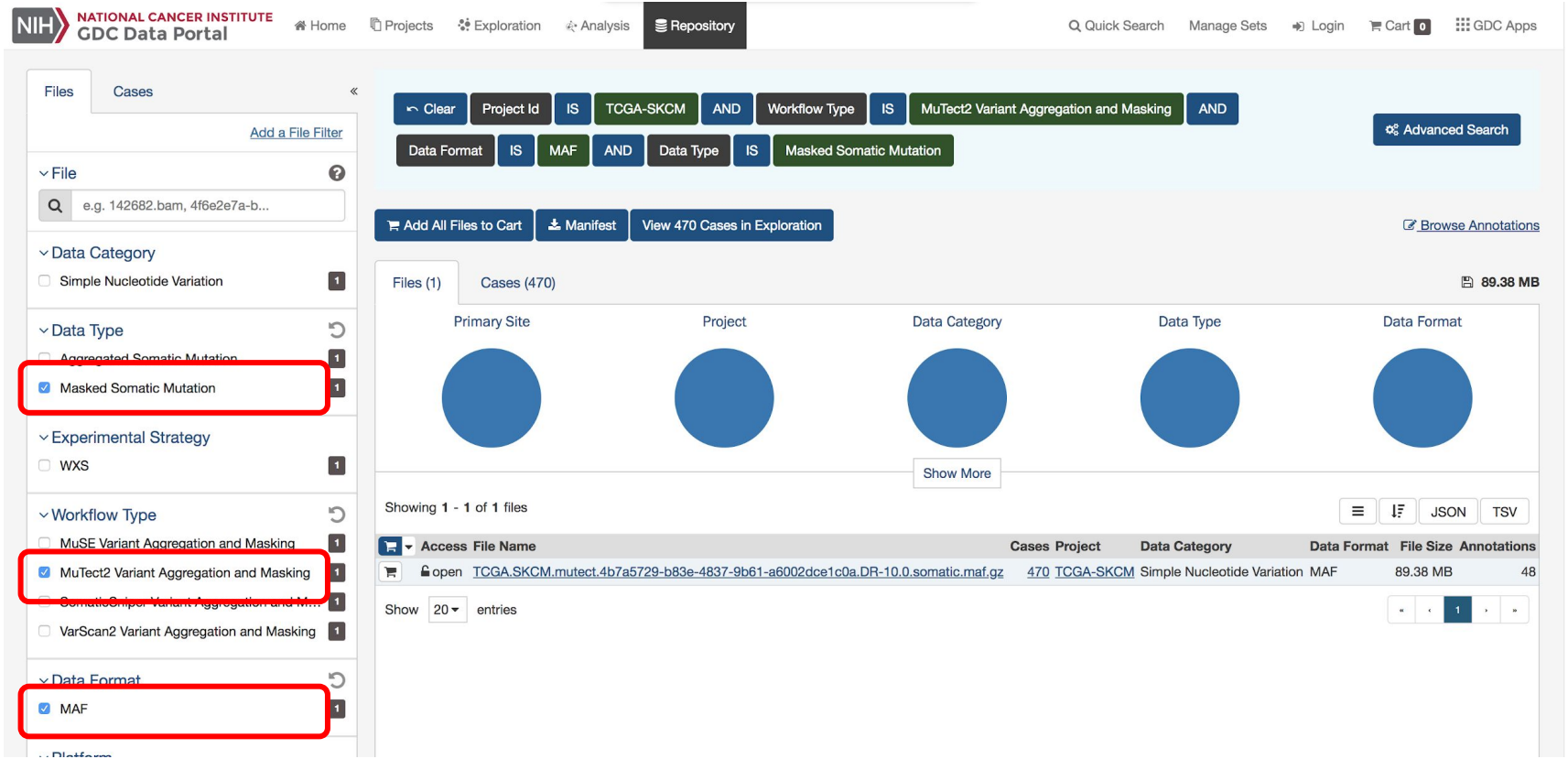
Choose patients based on project

Use the GenomicDataCommons package to find and download variants from the TCGA cutaneous melanoma dataset.

```
library(GenomicDataCommons)
fnames = files() %>%
    GenomicDataCommons::filter( ~ cases.project.project_id=='TCGA-SKCM' &
        data_type=='Masked Somatic Mutation' &
        data_format=='MAF' &
        analysis.workflow_type=='MuTect2 Variant Aggregation and Masking') %>%
    ids() %>%
    gdcdata()
```

And now take those data directly to maftools for analysis and visualization.

```
library(maftools)
melanoma = read.maf(maf = fnames[1])
```



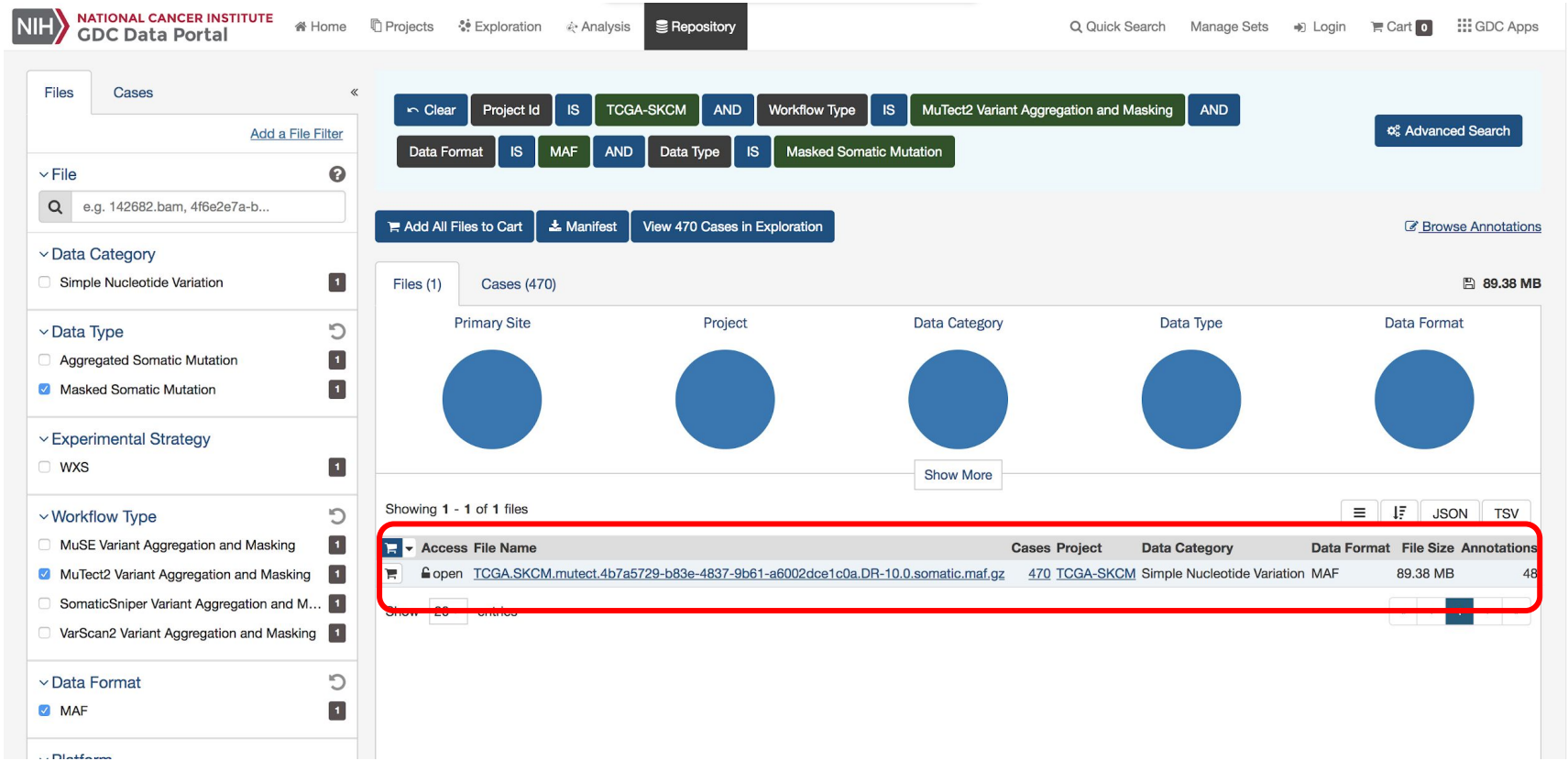
Choose Data Type and Workflow to select files

Use the GenomicDataCommons package to find and download variants from the TCGA cutaneous melanoma dataset.

```
library(GenomicDataCommons)
fnames = files() %>%
    GenomicDataCommons::filter(~ cases.project.project_id=='TCGA-SKCM' &
        data_type=='Masked Somatic Mutation' &
        data_format=='MAF' &
        analysis.workflow_type=='MuTect2 Variant Aggregation and Masking') %>%
    ids() %>%
    gdcdata()
```

And now take those data directly to maftools for analysis and visualization.

```
library(maftools)
melanoma = read.maf(maf = fnames[1])
```



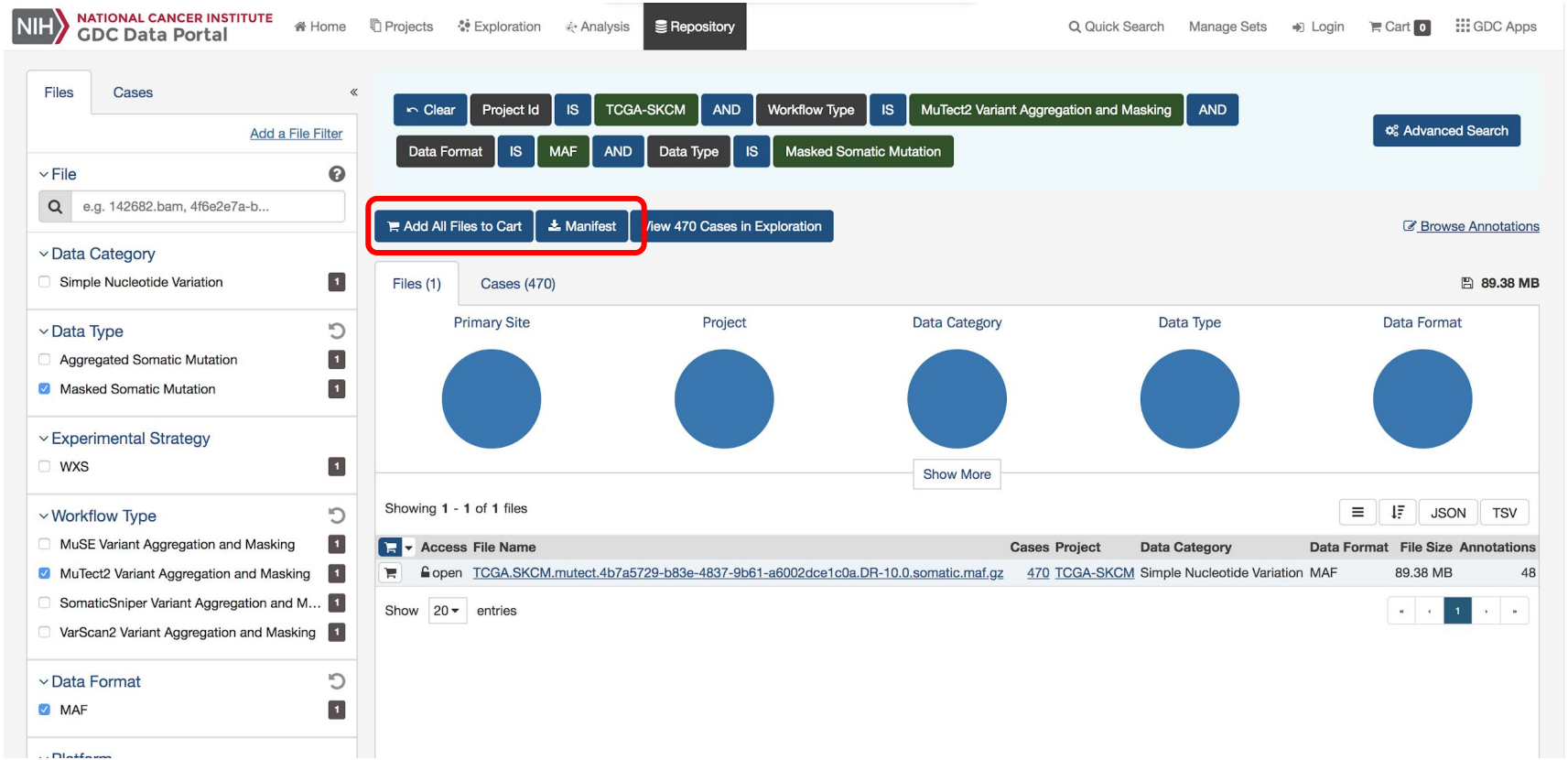
Files, like all entities in the GDC, have an associated UUID

Use the GenomicDataCommons package to find and download variants from the TCGA cutaneous melanoma dataset.

```
library(GenomicDataCommons)
fnames = files() %>%
    GenomicDataCommons::filter(~ cases.project.project_id=='TCGA-SKCM' &
        data_type=='Masked Somatic Mutation' &
        data_format=='MAF' &
        analysis.workflow_type=='MuTect2 Variant Aggregation and Masking') %>%
    ids() %>%
    gdcdata()
```

And now take those data directly to maftools for analysis and visualization.

```
library(maftools)
melanoma = read.maf(maf = fnames[1])
```

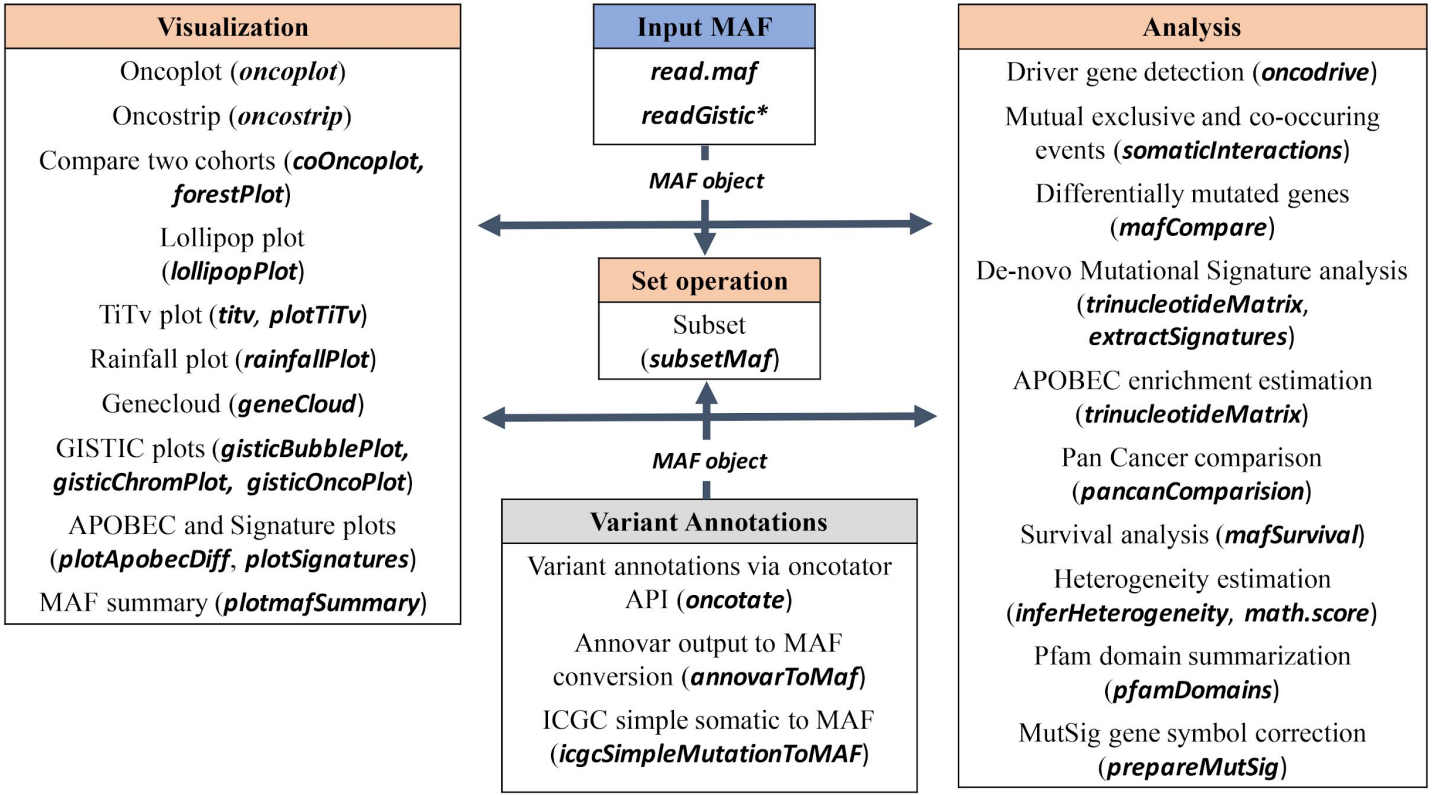

NIH NATIONAL CANCER INSTITUTE GDC Data Portal
Home
Projects
Exploration
Analysis
Repository
Quick Search
Manage Sets
Login
Cart 0
GDC Apps
Files
Cases
Add a File Filter
File
e.g. 142682.bam, 4f6e2e7a-b...
Data Category
Simple Nucleotide Variation 1
Data Type
Aggregated Somatic Mutation 1
Masked Somatic Mutation 1
Experimental Strategy
WXS 1
Workflow Type
MuSE Variant Aggregation and Masking 1
MuTect2 Variant Aggregation and Masking 1
SomaticSniper Variant Aggregation and M... 1
VarScan2 Variant Aggregation and Masking 1
Data Format
MAF 1
Clear
Project Id IS TCGA-SKCM AND Workflow Type IS MuTect2 Variant Aggregation and Masking AND
Data Format IS MAF AND Data Type IS Masked Somatic Mutation
Advanced Search
Add All Files to Cart
Manifest
View 470 Cases in Exploration
Browse Annotations
Files (1)
Cases (470)
89.38 MB
Primary Site
Project
Data Category
Data Type
Data Format
Show More
Showing 1 - 1 of 1 files
JSON
TSV
Access File Name Cases Project Data Category Data Format File Size Annotations
open TCGA.SKCM.mutect.4b7a5729-b83e-4837-9b61-a6002dce1c0a.DR-10.0.somatic.maf.gz 470 TCGA-SKCM Simple Nucleotide Variation MAF 89.38 MB 48
Show 20 entries
1

Use the GenomicDataCommons package to find and download variants from the TCGA cutaneous melanoma dataset.

```
library(GenomicDataCommons)
fnames = files() %>%
    GenomicDataCommons::filter(~ cases.project.project_id=='TCGA-SKCM' &
        data_type=='Masked Somatic Mutation' &
        data_format=='MAF' &
        analysis.workflow_type=='MuTect2 Variant Aggregation and Masking') %>%
    ids() %>%
    gdcdata()
```

And now take those data directly to maftools for analysis and visualization.

```
library(maftools)
melanoma = read.maf(maf = fnames[1])
```



GDC Programmatic access example:

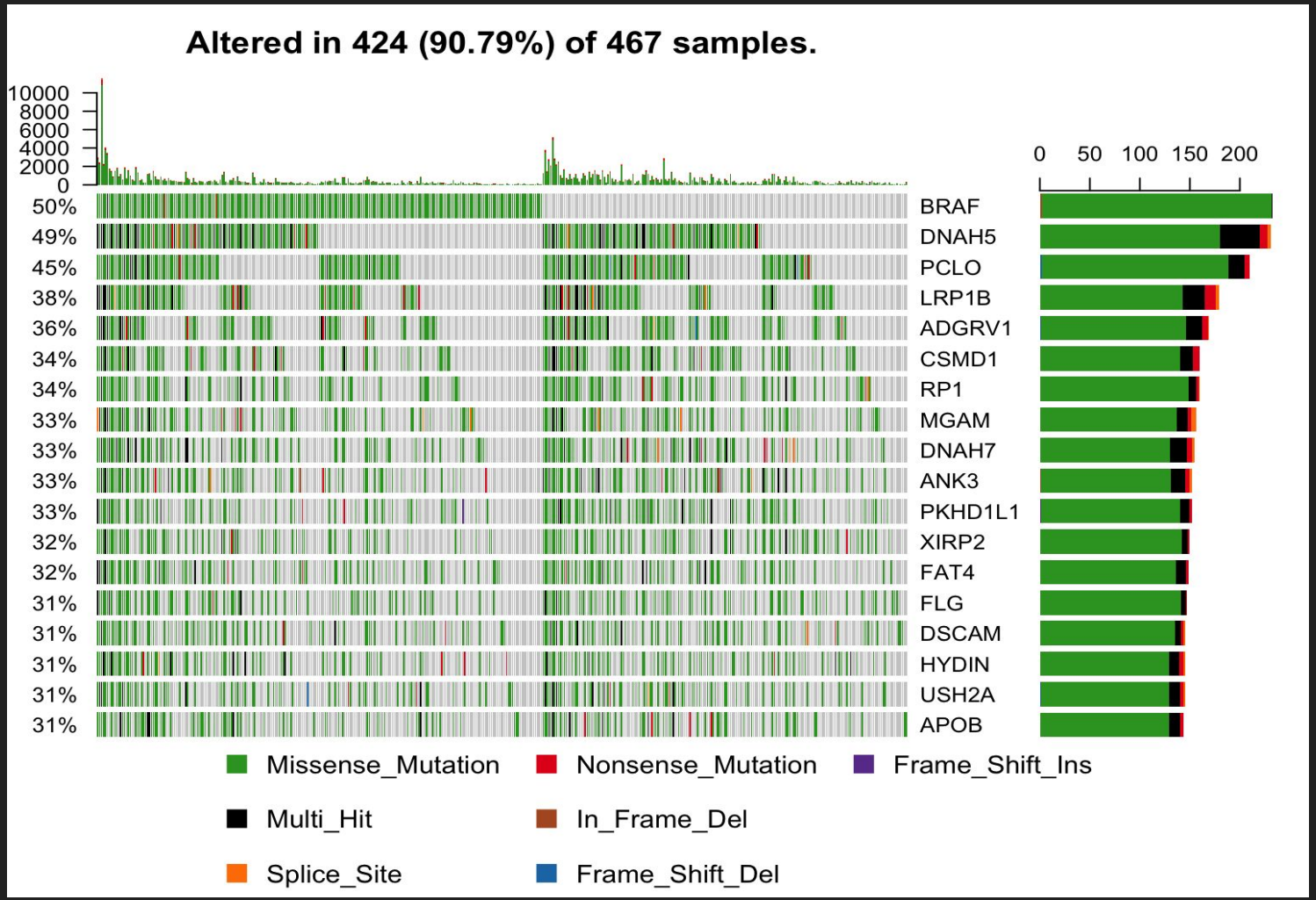
Somatic profiles from TCGA melanoma samples (8 lines of code)

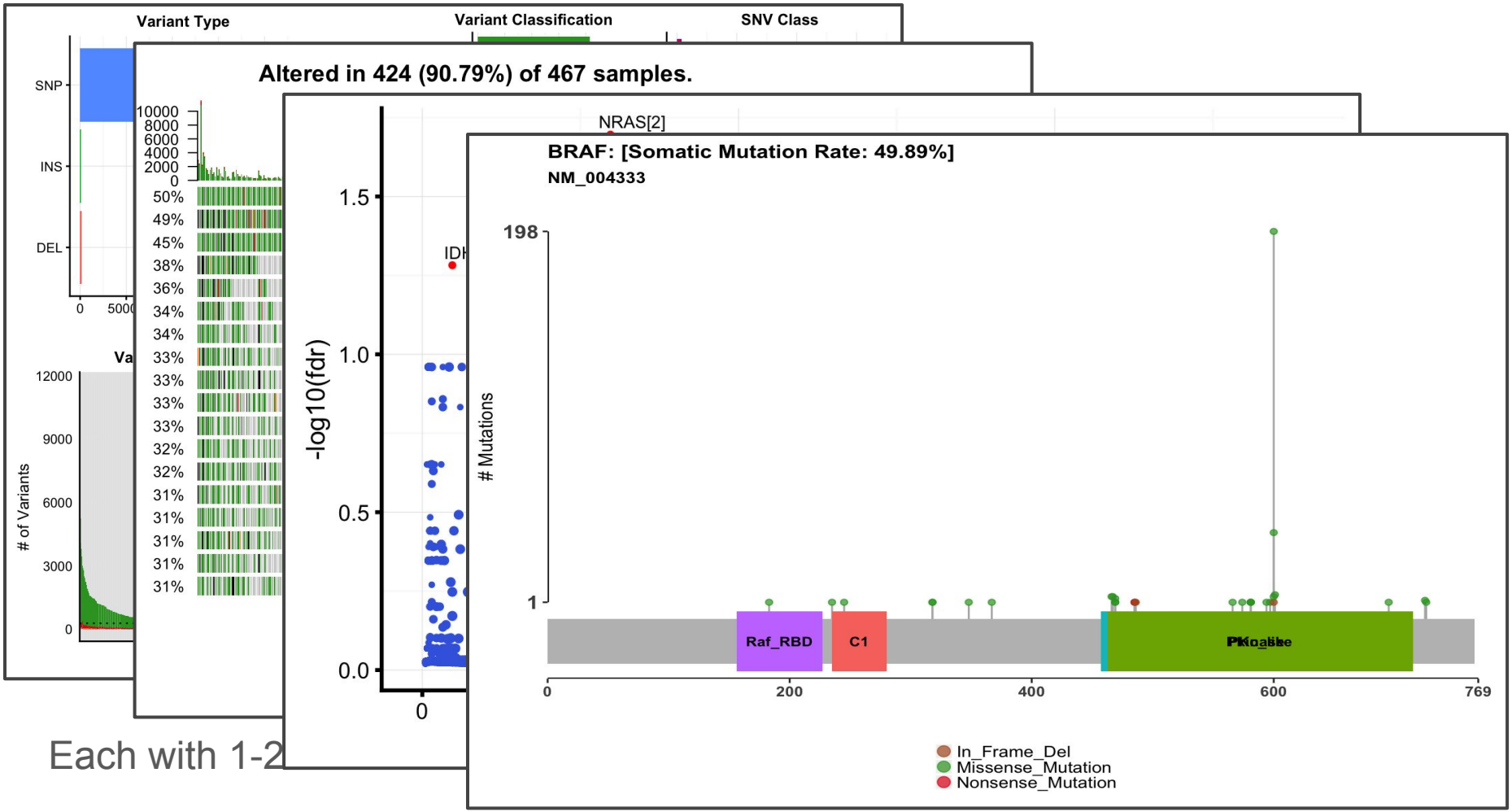
Use the GenomicDataCommons package to find and download variants from the TCGA cutaneous melanoma dataset.

```
library(GenomicDataCommons)
fnames = files() %>%
    GenomicDataCommons::filter(~ cases.project.project_id=='TCGA-SKCM' &
        data_type=='Masked Somatic Mutation' &
        data_format=='MAF' &
        analysis.workflow_type=='MuTect2 Variant Aggregation and Masking') %>%
    ids() %>%
    gdcdata()
```

And now take those data directly to maftools for analysis and visualization.

```
library(maftools)
melanoma = read.maf(maf = fnames[1])
```


Altered in 424 (90.79%) of 467 samples.
10000
8000
6000
4000
2000
0
0
50
100
150
200
50%
49%
45%
38%
36%
34%
34%
33%
33%
33%
33%
32%
32%
31%
31%
31%
31%
31%
BRAF
DNAH5
PCLO
LRP1B
ADGRV1
CSMD1
RP1
MGAM
DNAH7
ANK3
PKHD1L1
XIRP2
FAT4
FLG
DSCAM
HYDIN
USH2A
APOB
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Ins
Multi_Hit
In_Frame_Del
Splice_Site
Frame_Shift_Del


Variant Type
Variant Classification
SNV Class
SNP
INS
DEL
# of Variants
Altered in 424 (90.79%) of 467 samples.
NRAS[2]
IDH
-log10(fdr)
BRAF: [Somatic Mutation Rate: 49.89%]
NM_004333
Mutations
198
1
Raf_RBD
C1
0
200
400
600
769
In_Frame_Del
Missense_Mutation
Nonsense_Mutation

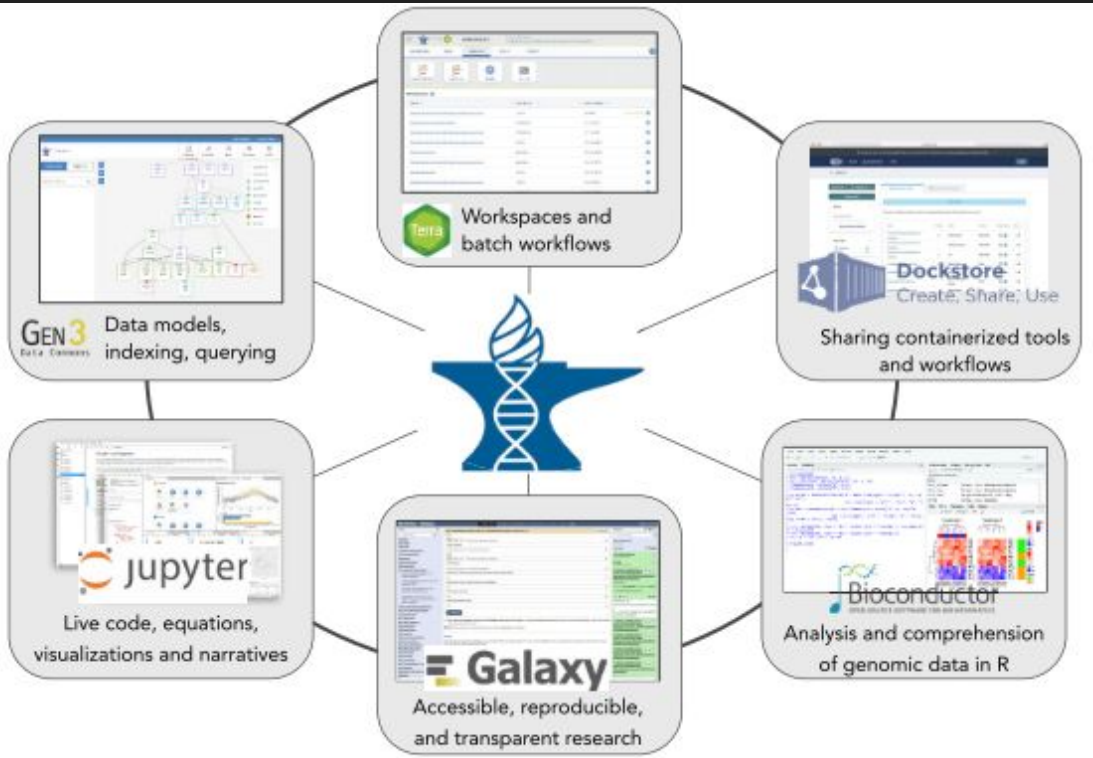
Each with 1-2

Qualities

- Discoverable
- Installable
- Reliable
- Documented
- Supported
- Integrated
- *Connected*
- Scalable
- State-of-the-art
- Community-driven

NHGRI's AnVIL – inverting the data sharing model

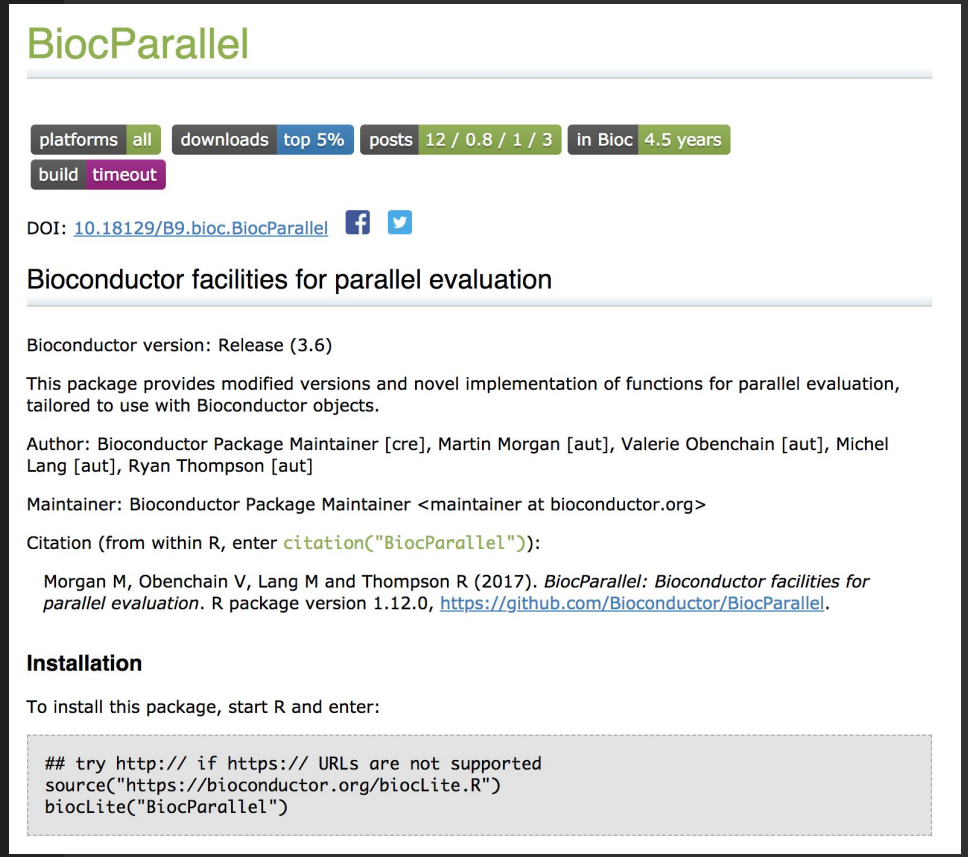
cost controls, security, data volumes, avoiding bitrot



Consortium	Datatypes	Cohorts	Samples	Participants	Size (TB)
1000 Genomes Project (1KGP)	WGS	1	3,202	3,202	72.98
Centers for Common Disease Genomics (CCDG)	WGS, WXS, Clinical Phenotypes	198	272,306	256,318	2,624.12
Centers for Mendelian Genomics (CMG)	WGS, Clinical Phenotypes	41	20,706	16,599	97.89
Convergent Neuroscience	WGS	2	304	300	5.32
Genotype-Tissue Expression (GTEx v8)	WGS, RNAseq	1	17,382	979	182.14
Human Pangenome Reference Consortium (HPRC)	Short & long-read WGS	1	57	47	223.47

Qualities

- Discoverable
- Installable
- Reliable
- Documented
- Supported
- Integrated
- Connected
- *Scalable*
- State-of-the-art
- Community-driven



BiocParallel

platforms all | downloads top 5% | posts 12 / 0.8 / 1 / 3 | in Bioc 4.5 years | build timeout

DOI: 10.18129/B9.bioc.BiocParallel

Bioconductor facilities for parallel evaluation

Bioconductor version: Release (3.6)

This package provides modified versions and novel implementation of functions for parallel evaluation, tailored to use with Bioconductor objects.

Author: Bioconductor Package Maintainer [cre], Martin Morgan [aut], Valerie Obenchain [aut], Michel Lang [aut], Ryan Thompson [aut]

Maintainer: Bioconductor Package Maintainer <maintainer at bioconductor.org>

Citation (from within R, enter `citation("BiocParallel")`):

Morgan M, Obenchain V, Lang M and Thompson R (2017). *BiocParallel: Bioconductor facilities for parallel evaluation*. R package version 1.12.0, https://github.com/Bioconductor/BiocParallel.

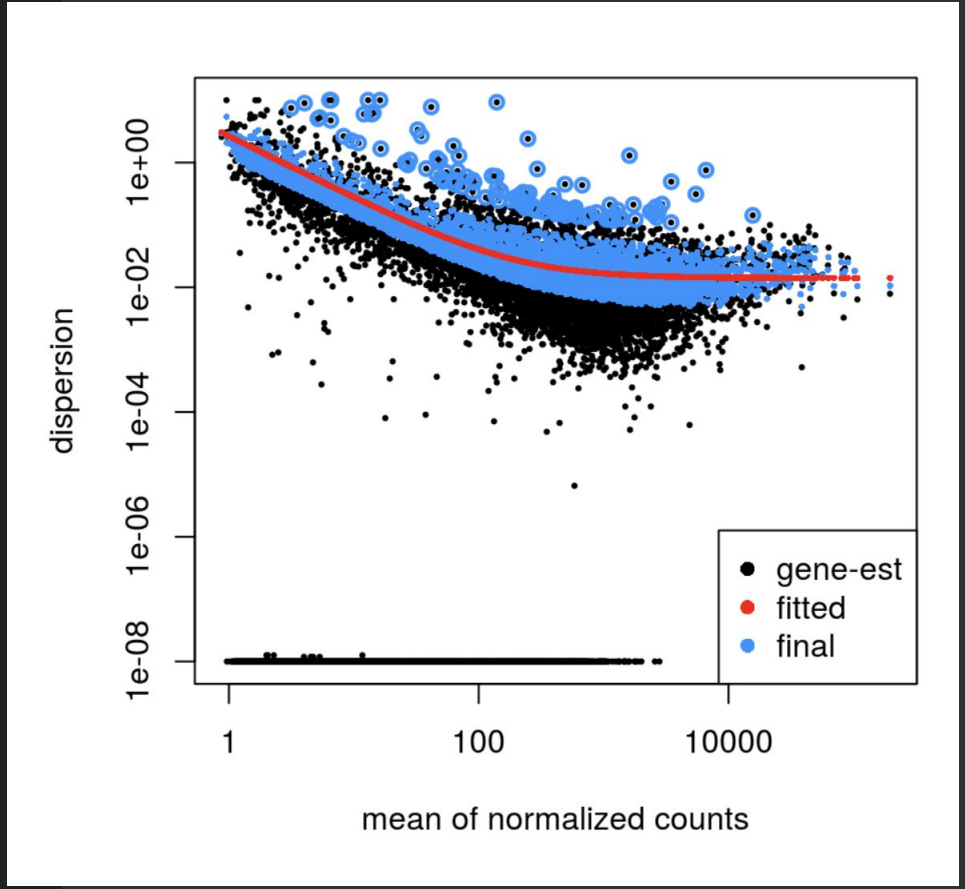
Installation

To install this package, start R and enter:

```
## try http:// if https:// URLs are not supported
source("https://bioconductor.org/biocLite.R")
biocLite("BiocParallel")
```

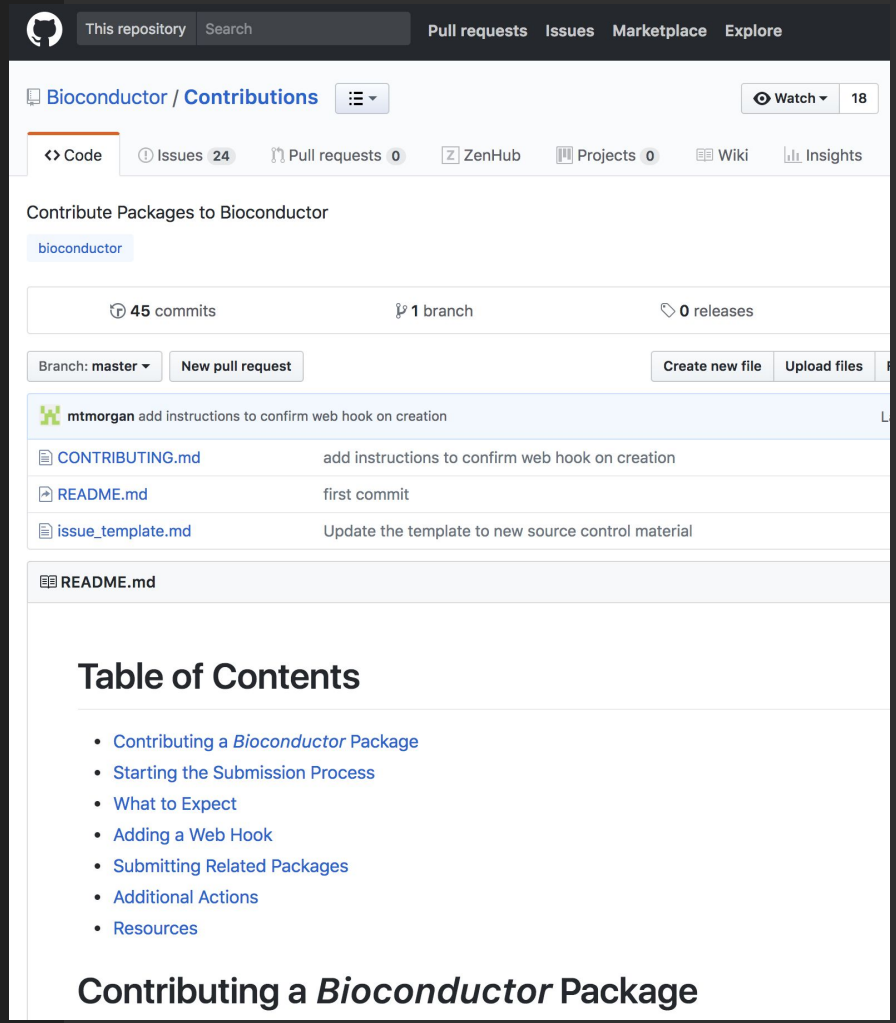
Qualities

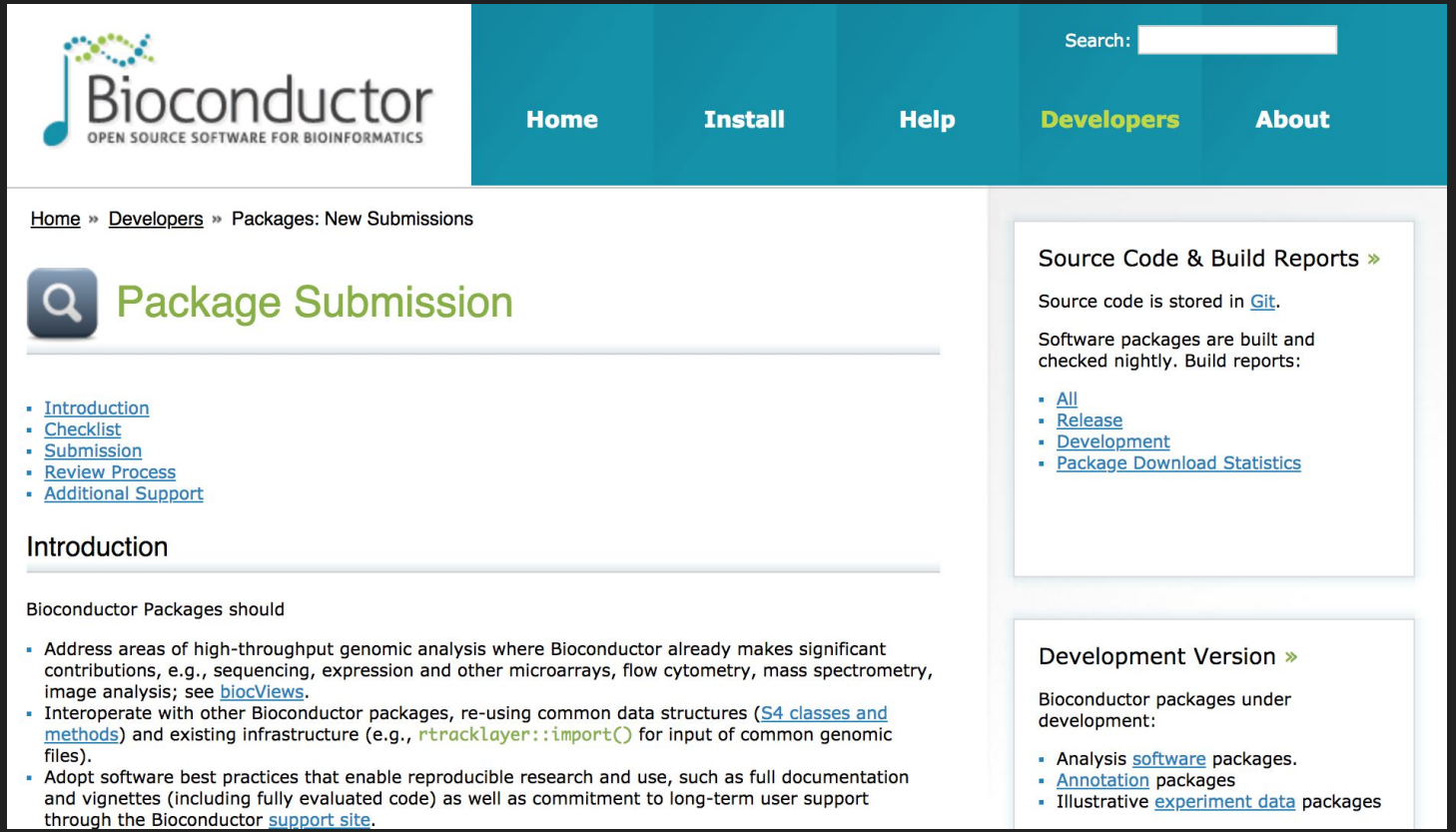
- Discoverable
- Installable
- Reliable
- Documented
- Supported
- Integrated
- Connected
- Scalable
- *State-of-the-art*
- Community-driven



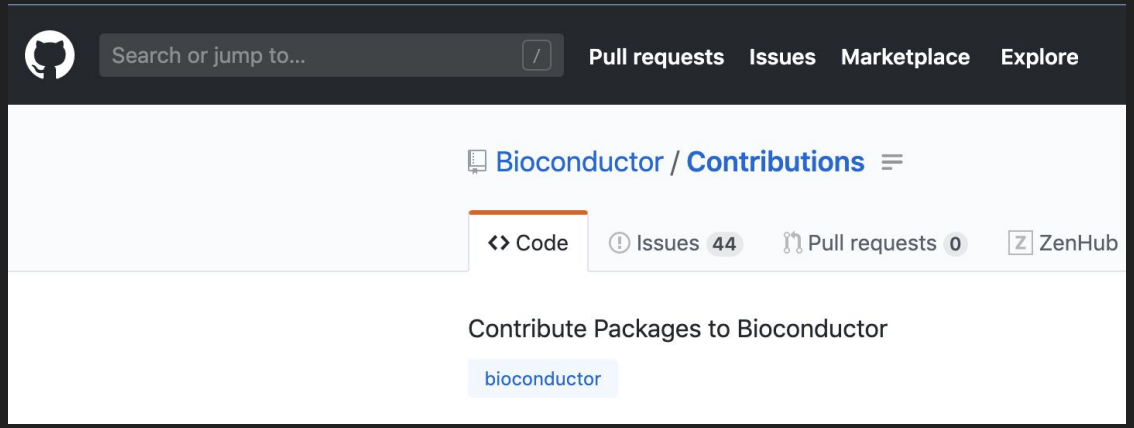
Qualities

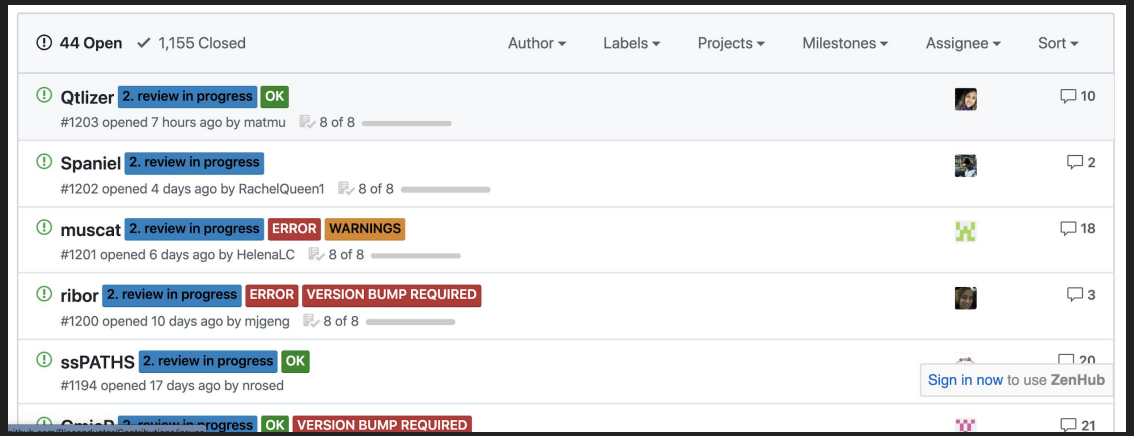
- Discoverable
- Installable
- Reliable
- Documented
- Supported
- Integrated
- Connected
- Scalable
- State-of-the-art
- *Community-driven*





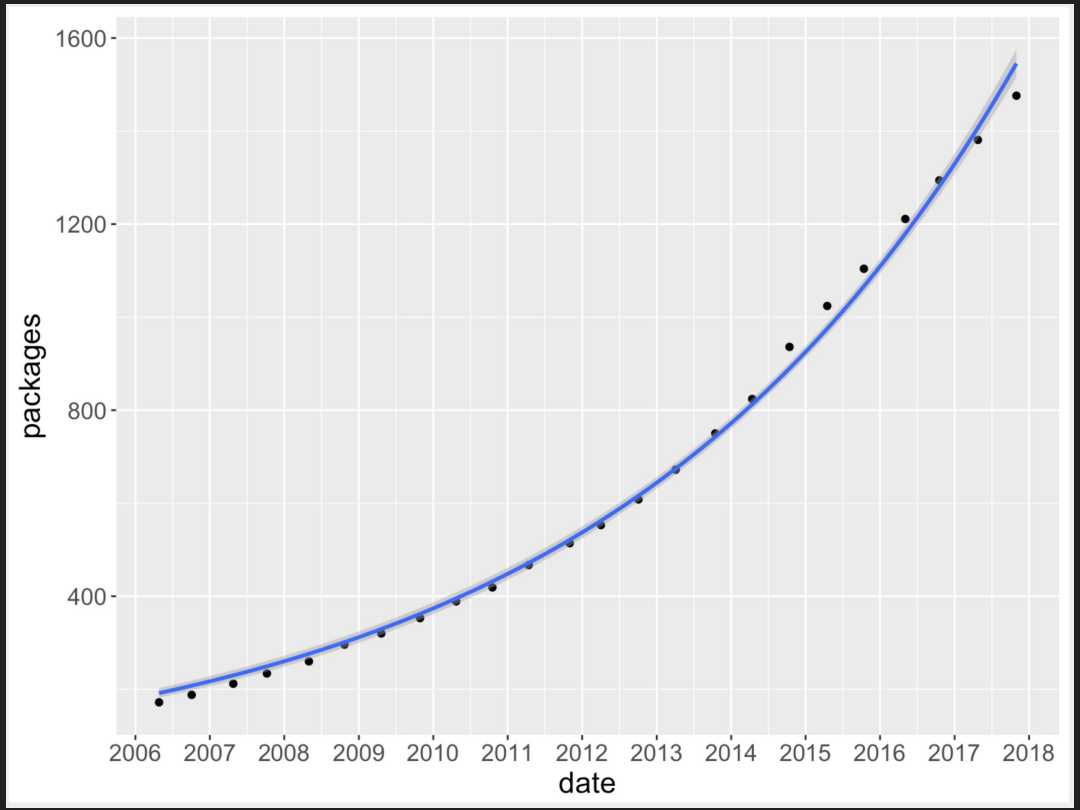
Core value: open and engaged





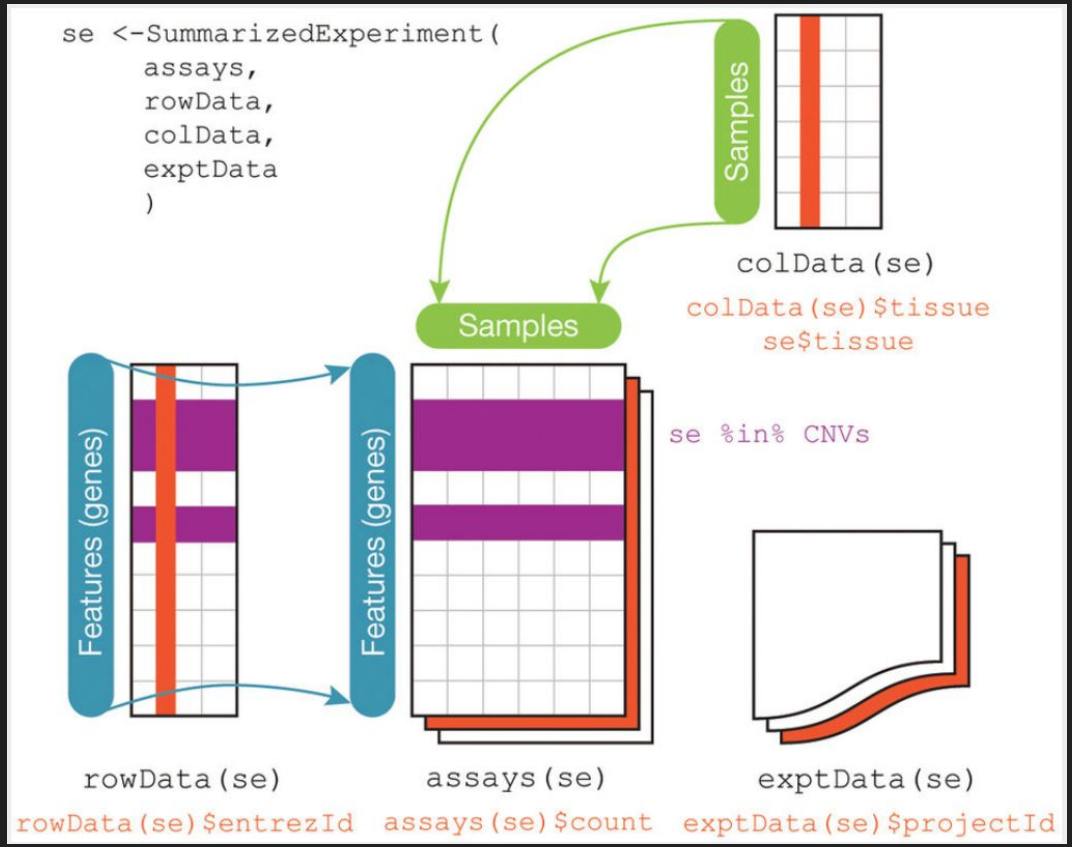
https://github.com/Bioconductor/Contributions

Bioconductor Contributions



Core Value: Reproducible research

- Recognize complexity in high-throughput biological data
- Version control everything
- Continuous testing and integration
- Text-based workflow (no GUI)
- Literate programming approaches and documentation
- Education on tooling
- Numerous mechanisms for FAIR data sharing

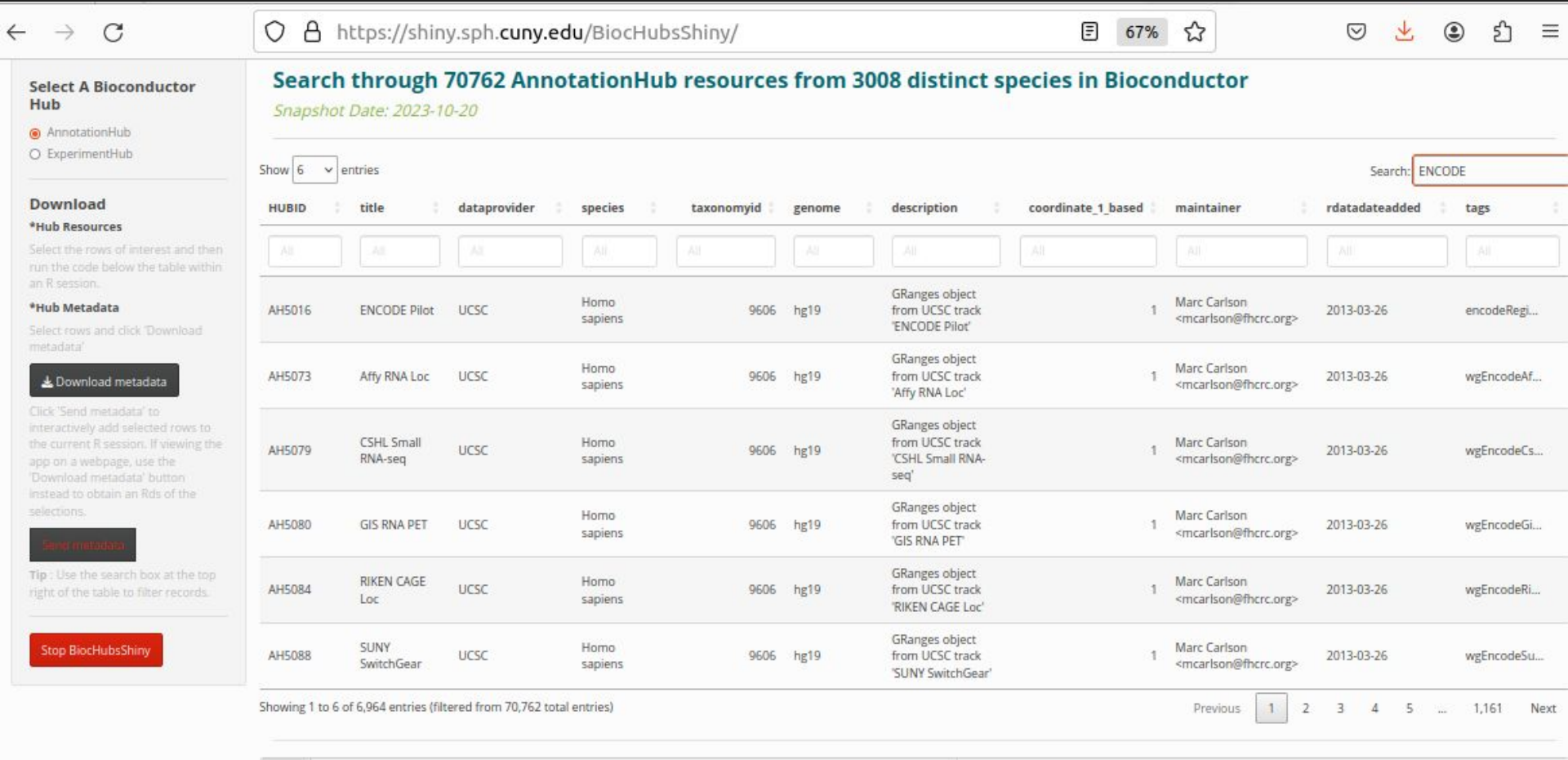


Not just analysis packages

- AnnotationHub: More than 30,000 curated public reference datasets
  - ENCODE
  - UCSC tracks
  - Organism databases for dozens of species
- ExperimentHub: User-submitted, curated data, code, documentation
  - CuratedTCGAData
  - CuratedMetagenomicsData
- Microarray annotation resources (more than 100 platforms, standardized)
- API access to dozens of cancer and biological databases

For free: versioning, information provenance, FAIR, marketing, interoperability across the project

Genomic annotation, ingestible via R function calls

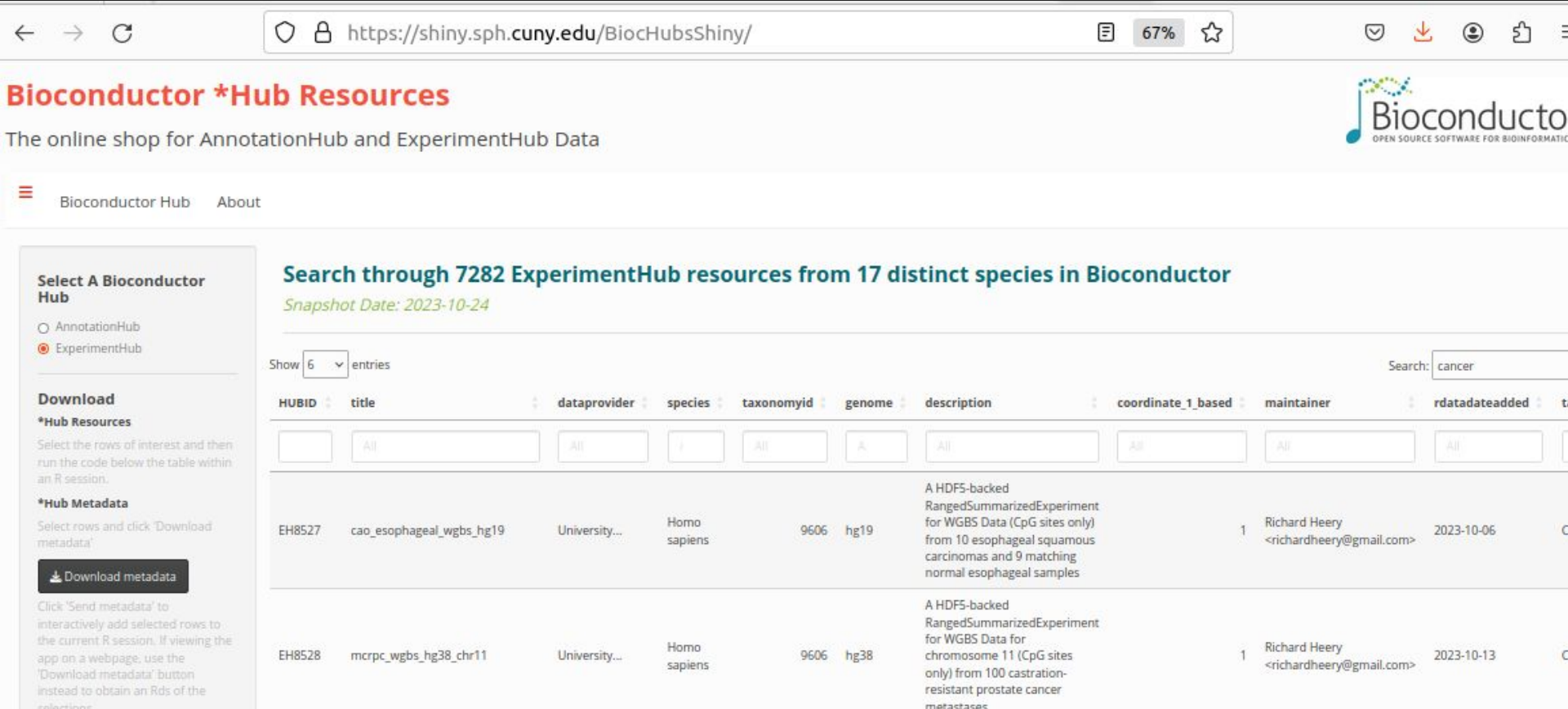


"CSHL small RNA-seq", a UCSC "track" to compute on

```
> library(AnnotationHub)
4/32 packages newly attached/loaded, see sessionInfo() for details.
> ah = AnnotationHub()
snapshotDate(): 2023-10-20
> smr = ah[["AH5079"]]
downloading 1 resources
retrieving 1 resource
  |======================================================================| 100%

loading from cache
require("GenomicRanges")
> smr
GRanges object with 32537 ranges and 2 metadata columns:
          seqnames              ranges strand |                name     score
             <Rle>           <IRanges>  <Rle> |         <character> <numeric>
      [1]     chr1     566845-566876      - | chr1:566844-566876(u..        20
      [2]     chr1     567544-567588      - | chr1:567543-567588(u..        28
      [3]     chr1     567549-567602      + | chr1:567548-567602(u..        45
      [4]     chr1     568187-568230      - | chr1:568186-568230(u..        96
```

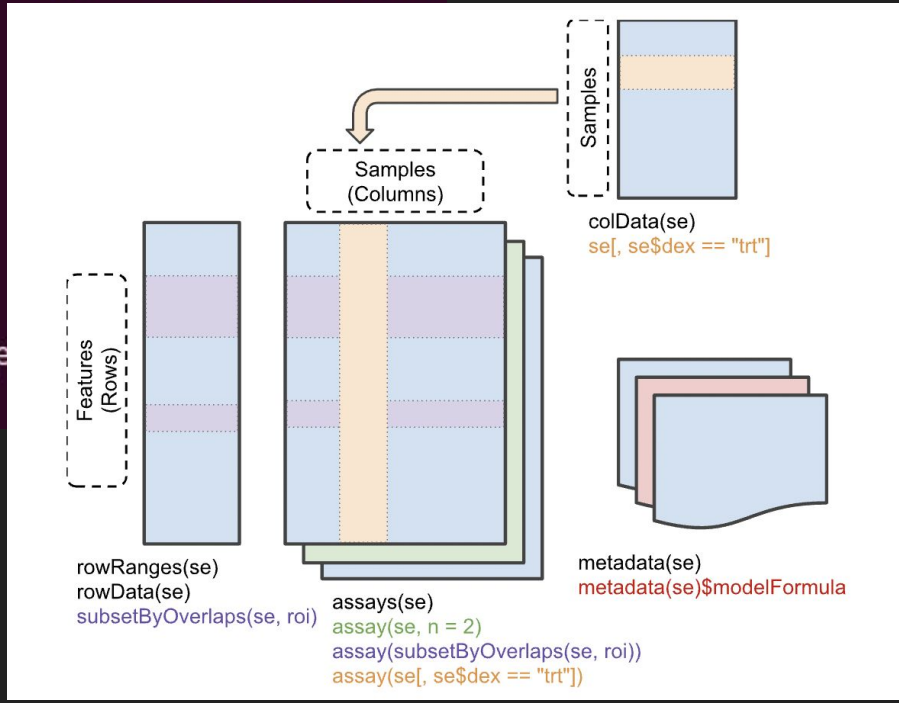
ExperimentHub … 'coherent' representations



A package in development … TumourMethData

```
> eso = TumourMethData::download_meth_dataset(dataset="cao_esophageal_wgbs_hg19")
[1] "A HDF5 SummarizedExperiment is already present in /home/vincent/TEMP/RtmpZeW
9 and is being returned"
> eso
class: RangedSummarizedExperiment
dim: 28217883 19
metadata(0):
assays(1): beta
rownames: NULL
rowData names(0):
colnames(19): N1 N2 ... T13 T15
colData names(6): patient_id disease_status ... tnm_stage
  survival_months
```

A major element in pursuit of coherence



Creating An ExperimentHub Package

Valerie Obenchain and Lori Shepherd

Modified: November 2017. Compiled: 02 May 2019

Contents

1 Overview

2 New resources

- 2.1 Notify `Bioconductor` team member
- 2.2 Building the data experiment package
- 2.3 Data objects
- 2.4 Metadata
- 2.5 Package review

3 Add additional resources

4 Bug fixes

- 4.1 Update the resource
- 4.2 Update the metadata

5 Remove resources

6 Uploading Data to S3

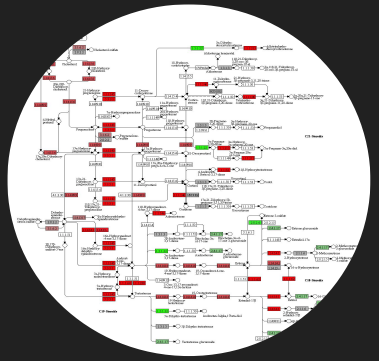
7 Validating

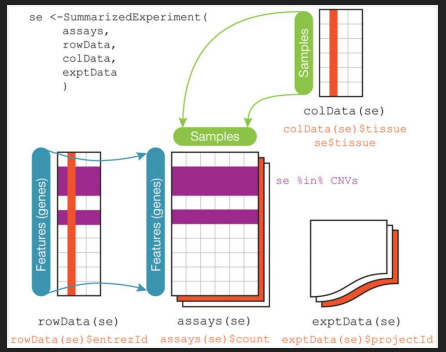
8 Example metadata.csv file and more information

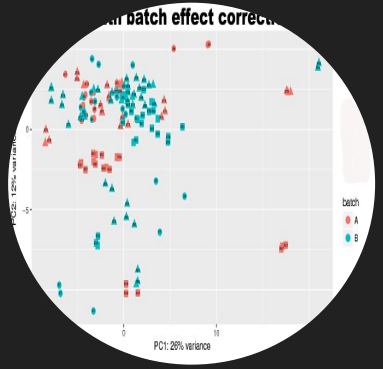
Typically WGBS data can't "fit in memory"; HDF5 is

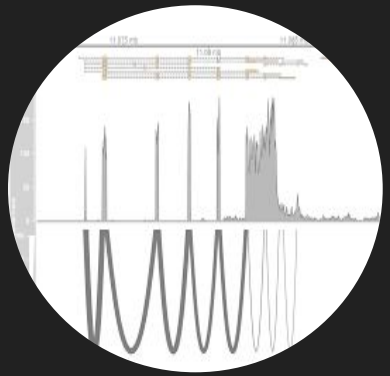
```
> str(assay(eso))
Formal class 'DelayedMatrix' [package "DelayedArray"] with 1 slot
  ..@ seed:Formal class 'DelayedSetDimnames' [package "DelayedArray"] with 2 slots
  .. .. ..@ dimnames:List of 2
  .. .. .. ..$ : int -1
  .. .. .. ..$ : chr [1:19] "N1" "N2" "N6" "N7" ...
  .. .. ..@ seed    :Formal class 'HDF5ArraySeed' [package "HDF5Array"] with 7 slots
  .. .. .. .. ..@ filepath : chr "/home/vincent/.cache/R/ExperimentHub/611bdc2bc8_8590"
  .. .. .. .. ..@ name     : chr "beta"
  .. .. .. .. ..@ as_sparse: logi FALSE
  .. .. .. .. ..@ type     : chr NA
  .. .. .. .. ..@ dim      : int [1:2] 28217883 19
  .. .. .. .. ..@ chunkdim : int [1:2] 1000000 1
  .. .. .. .. ..@ first_val: num NA
> head(rowRanges(eso))
GRanges object with 6 ranges and 0 metadata columns:
      seqnames    ranges strand
         <Rle> <IRanges>  <Rle>
  [1]     chr1     10469      *
  [2]     chr1     10471      *
  [3]     chr1     10484      *
  [4]     chr1     10489      *
```

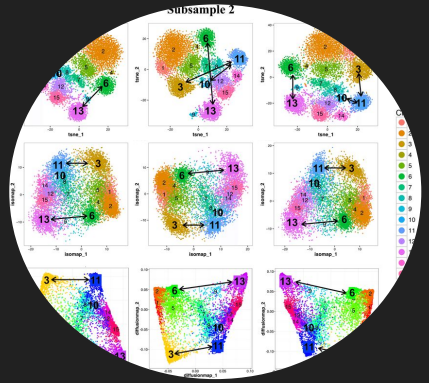
Core value: reuse and interoperability

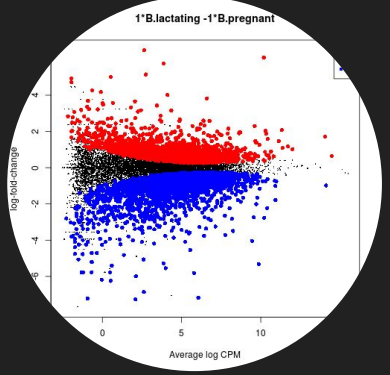




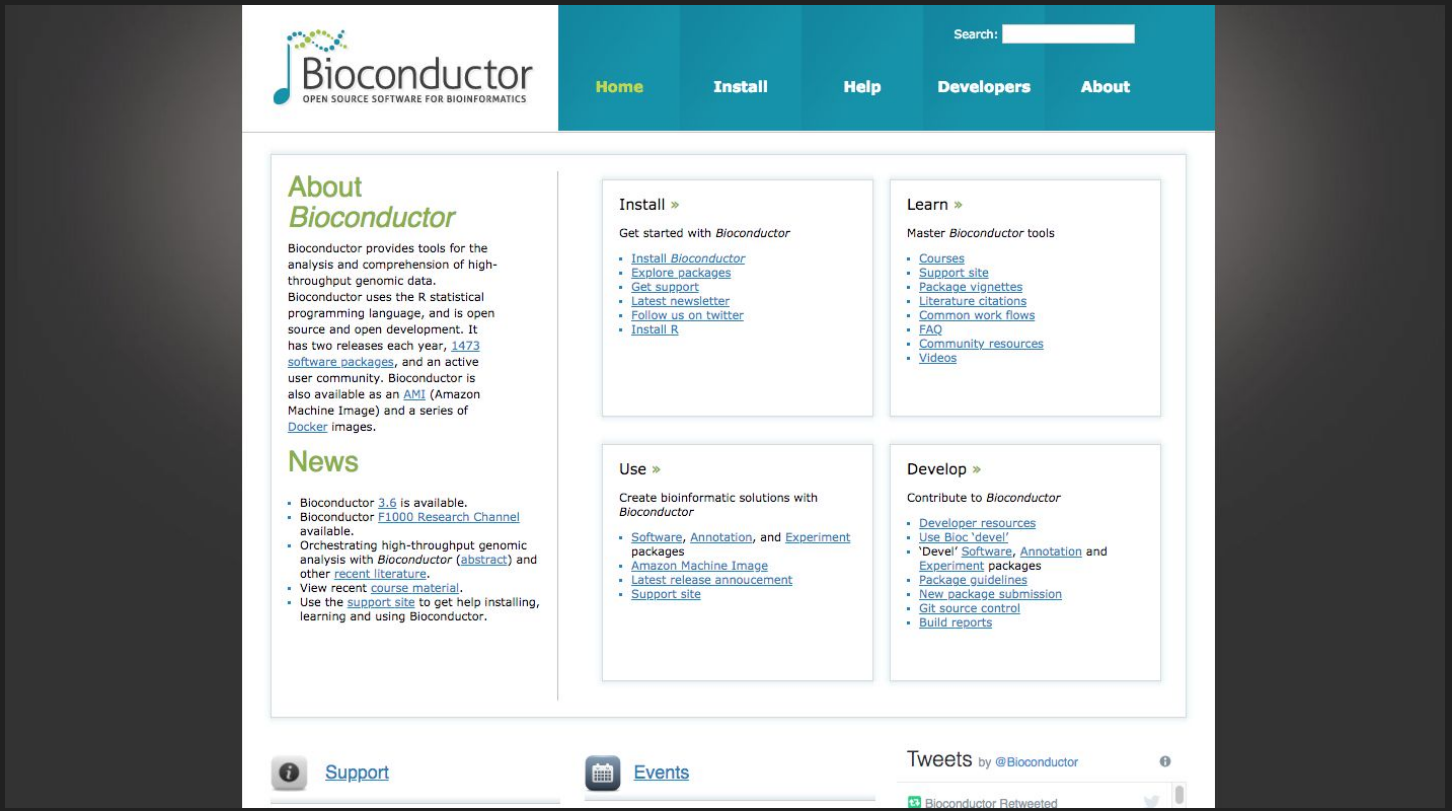








Bioconductor: Education, Training, and Community



Core infrastructure

Community contribution

Bioconductor Blog

Goals, but challenging to measure

Turning users into (increasingly experienced) developers

- Developer mailing list & Slack #developers-forum
- Package review
- Conferences and Workshops

"Care" of the community

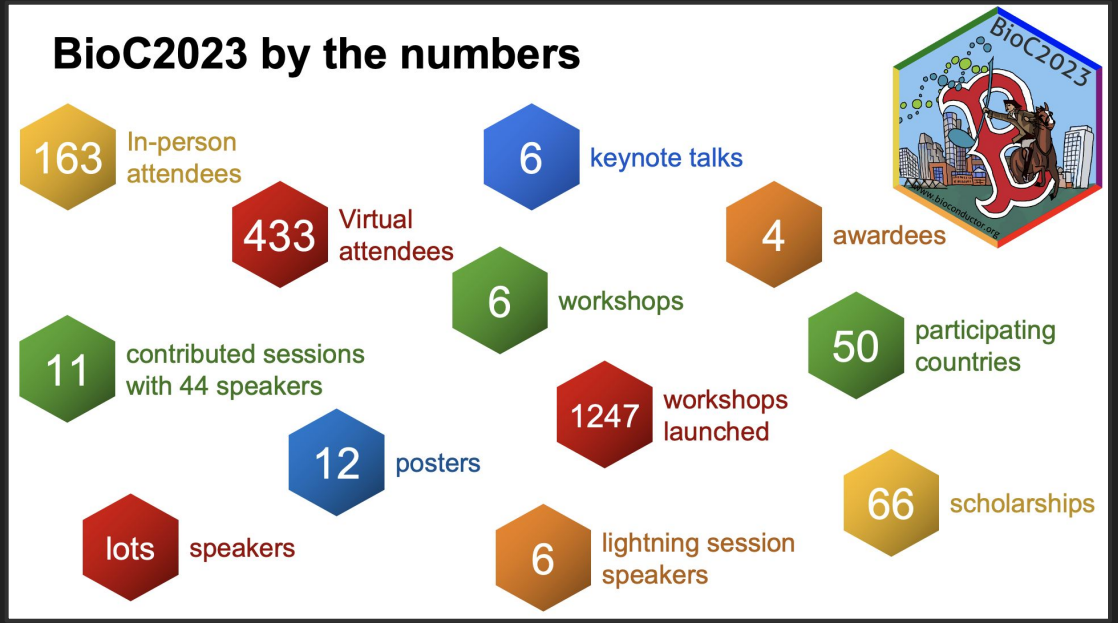
- Code of Conduct
- Outreach including to underrepresented areas
- Internships giving opportunities to learn

Events: 3 annual conferences

North American conference
BioC2024,
July 24-26,
Grand Rapids, USA

European conference
EuroBioC2024,
September 4-6,
Oxford, UK

Australasian conference
BioCAsia2024
late 2024,
Sydney, Australia



Website Tour

Questions?

https://bioconductor.org

https://seandavi.github.io