

ASMS 2019
ANNUAL
CONFERENCE
WORKSHOP



MASS SPEC PACKAGES FOR R

BEFORE YOU START, YOU NEED TO CONVERT YOUR DATA!

- ▶ Most mass spec instrument vendors save data files in proprietary formats that R can't directly read
- ▶ There are a variety of open formats for mass spec data, **mzML** is the main one currently in use
- ▶ Fortunately, open source tools exist that can convert most proprietary formats to mzML
- ▶ Unfortunately, this adds an extra processing step and mzML files can be large

PROTEOWIZARD TO THE RESCUE

- ▶ Proteowizard is an open source project with lots of tools for mass spectrometry data processing and analysis
<http://proteowizard.sourceforge.net>
- ▶ Of particular note: **MSConvert**
A GUI (and command line) tool for converting among MS data formats including proprietary → open formats
- ▶ Important: you have to use the *Windows* version to convert from proprietary to open formats

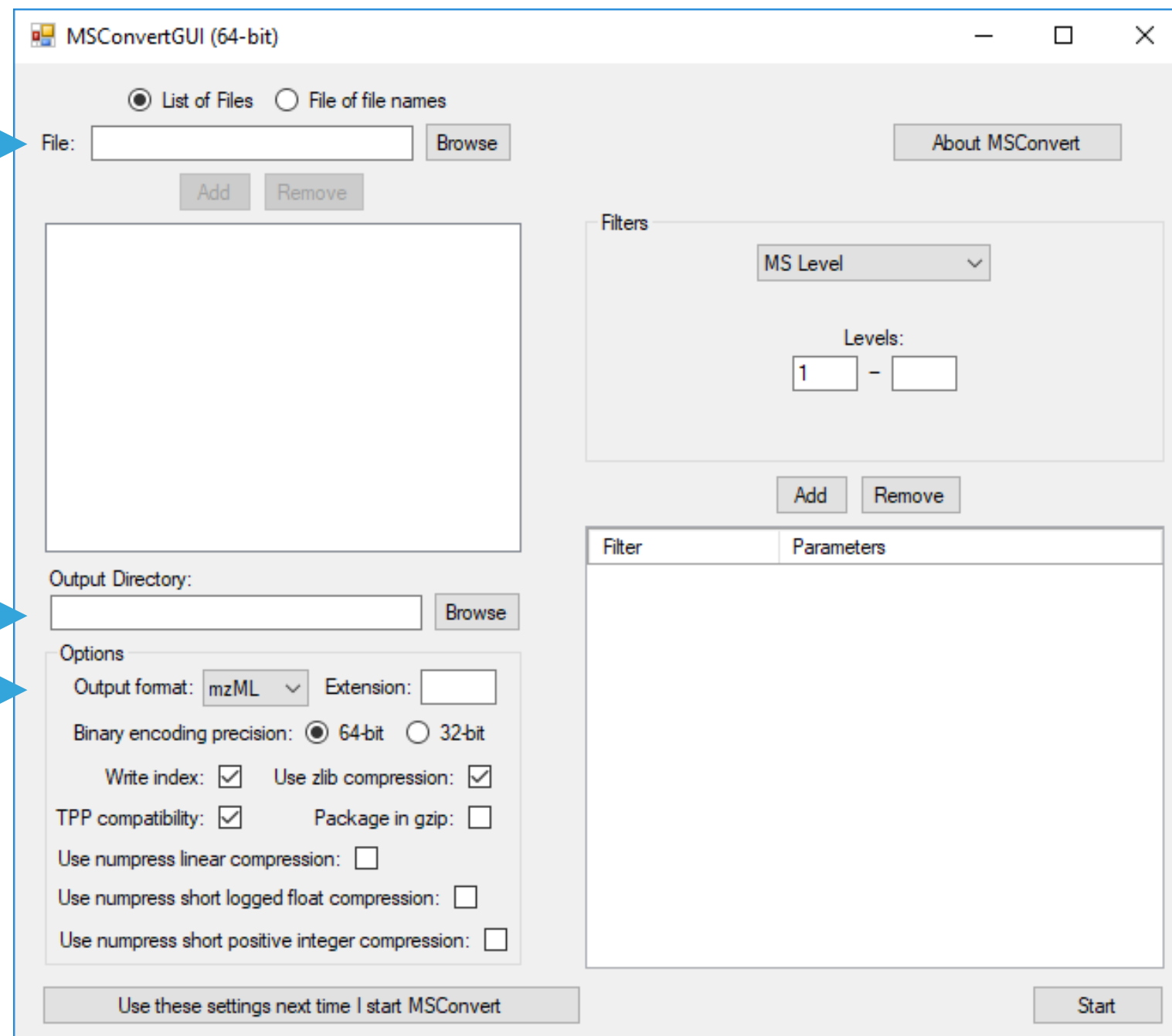
THE MS CONVERT GUI IS EASY TO USE

1. Specify input files

2. Specify the output directory

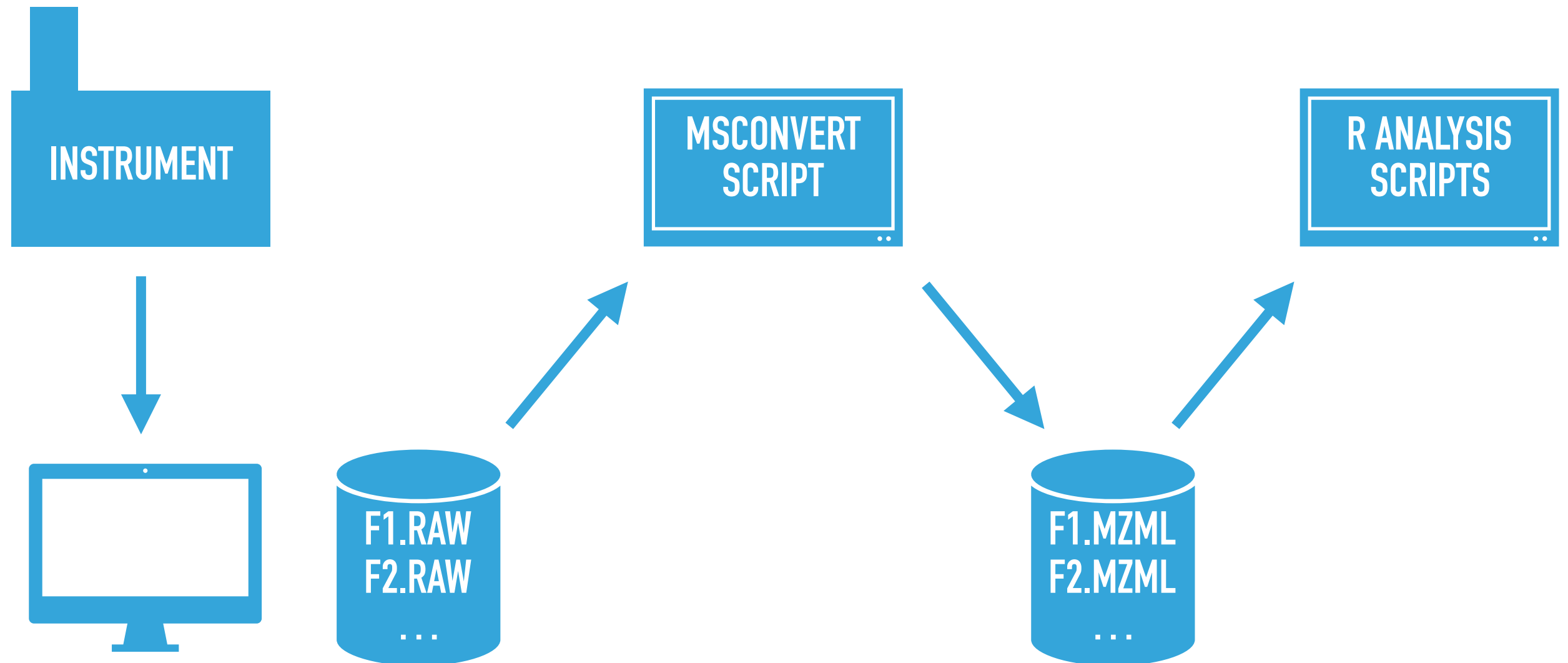
3. Select the output format

4. Click to convert



There's also a command line version, but you must use Windows

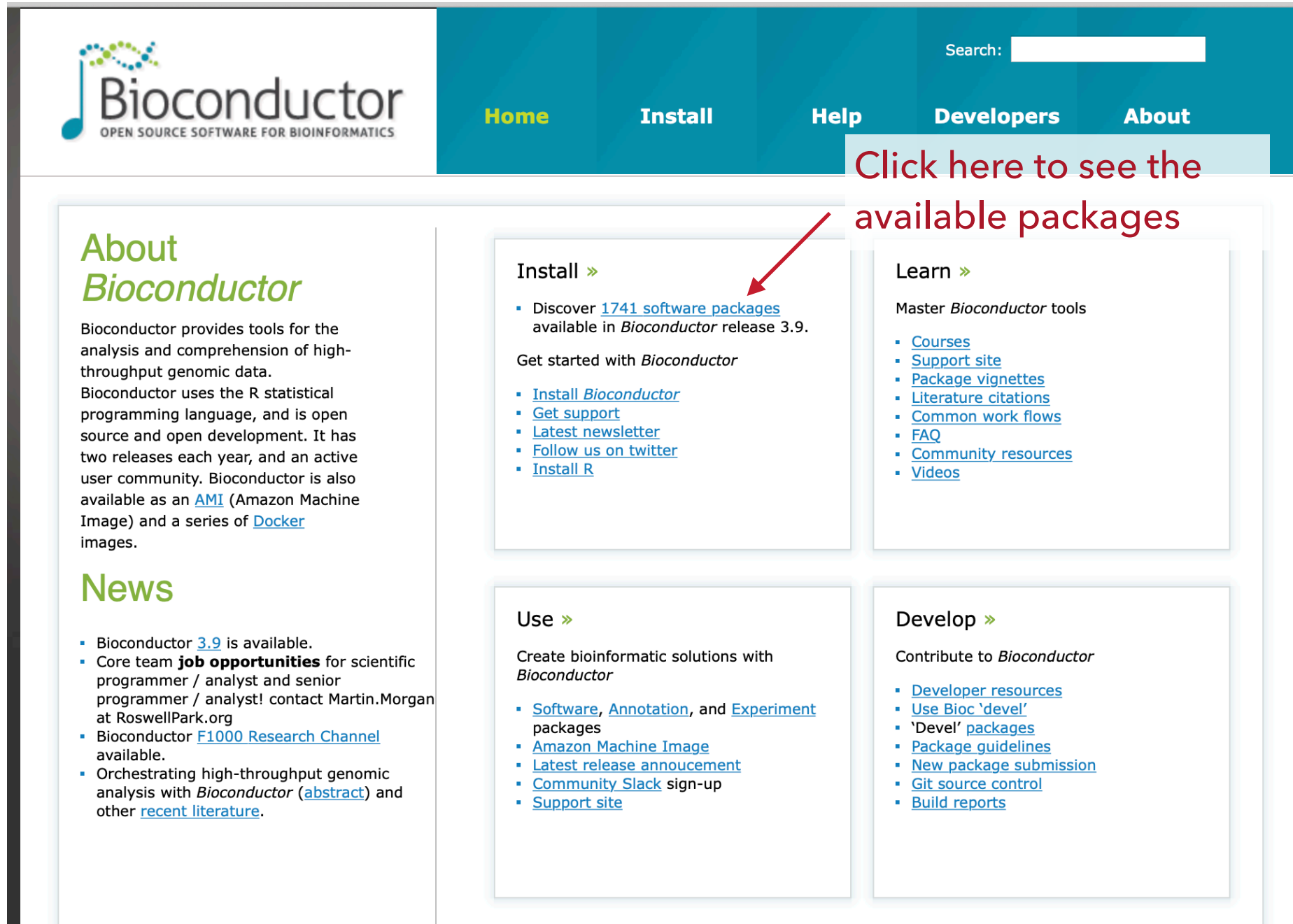
EXAMPLE DATA CONVERSION WORKFLOW



BIOCONDUCTOR IS A GREAT RESOURCE FOR MS PACKAGES

- ▶ **Bioconductor** is a project and repository for 1700+ R packages for biological data analysis
<https://bioconductor.org>
- ▶ Includes lots of great packages for mass spectrometry data analysis
- ▶ Packages typically have reference documentation and examples (vignettes)

WORKING WITH BIOCONDUCTOR

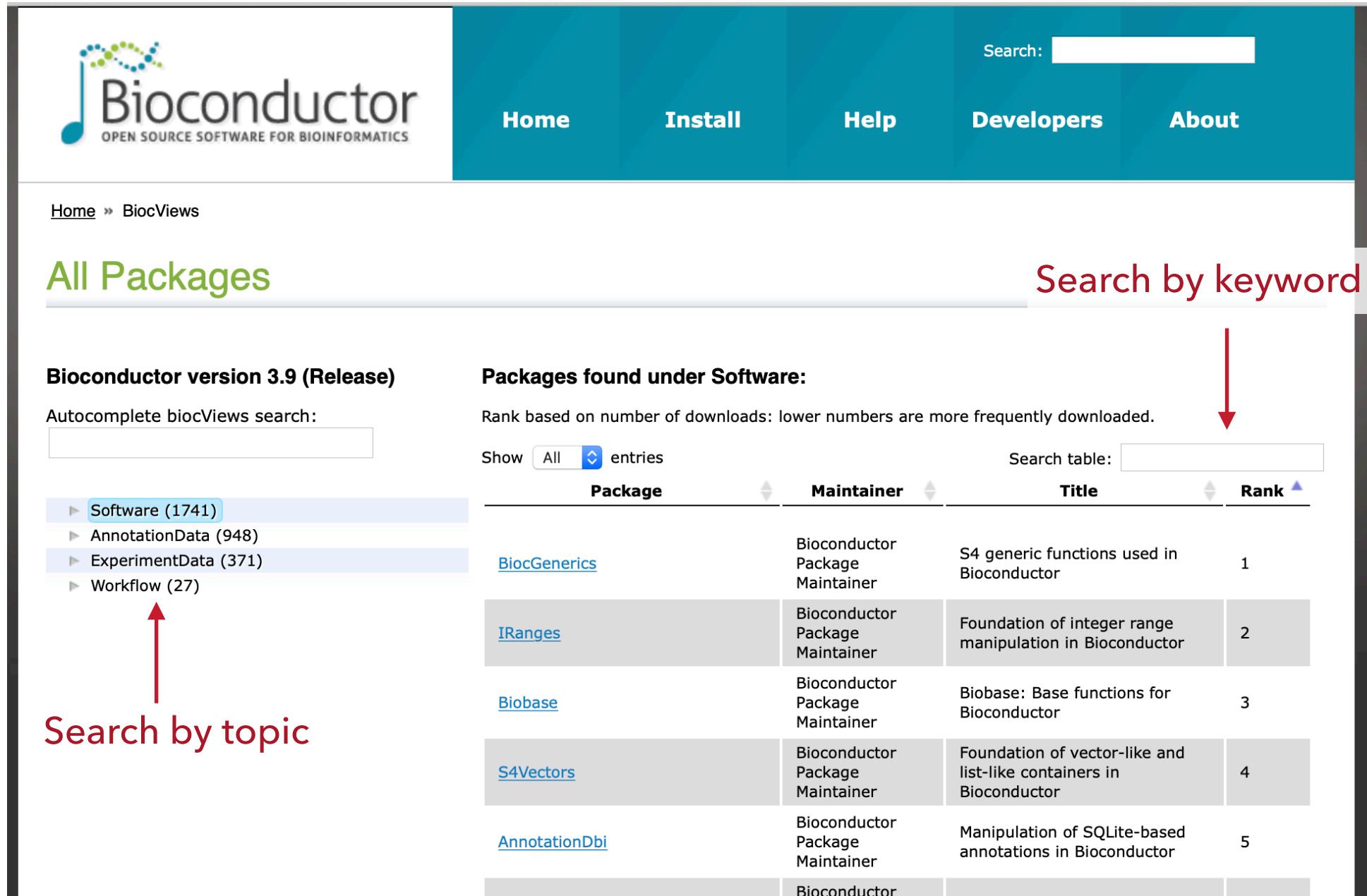


The screenshot shows the Bioconductor website homepage. The header includes the Bioconductor logo, a search bar, and navigation links: Home, Install, Help, Developers, and About. The main content area is divided into several sections:

- About Bioconductor**: A paragraph describing Bioconductor as open source software for bioinformatics, using R, and available as an AMI and Docker images.
- News**: A list of recent updates, including Bioconductor 3.9 availability and core team job opportunities.
- Install »**: A section with a red arrow pointing to the link "Discover [1741 software packages](#) available in *Bioconductor* release 3.9." Below this are links for "Get started with *Bioconductor*", including "Install *Bioconductor*", "Get support", "Latest newsletter", "Follow us on twitter", and "Install R".
- Learn »**: A section titled "Master *Bioconductor* tools" with links for "Courses", "Support site", "Package vignettes", "Literature citations", "Common work flows", "FAQ", "Community resources", and "Videos".
- Use »**: A section titled "Create bioinformatic solutions with *Bioconductor*" with links for "Software, Annotation, and Experiment packages", "Amazon Machine Image", "Latest release announcement", "Community Slack sign-up", and "Support site".
- Develop »**: A section titled "Contribute to *Bioconductor*" with links for "Developer resources", "Use Bioc 'devel'", "'Devel' packages", "Package guidelines", "New package submission", "Git source control", and "Build reports".

<https://bioconductor.org>

WORKING WITH BIOCONDUCTOR



Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home Install Help Developers About

Search:

Home » BiocViews

All Packages

Bioconductor version 3.9 (Release)

Autocomplete biocViews search:

- ▶ Software (1741)
- ▶ AnnotationData (948)
- ▶ ExperimentData (371)
- ▶ Workflow (27)

Packages found under Software:

Rank based on number of downloads: lower numbers are more frequently downloaded.

Show entries

Search table:

Package	Maintainer	Title	Rank
BiocGenerics	Bioconductor Package Maintainer	S4 generic functions used in Bioconductor	1
IRanges	Bioconductor Package Maintainer	Foundation of integer range manipulation in Bioconductor	2
Biobase	Bioconductor Package Maintainer	Biobase: Base functions for Bioconductor	3
S4Vectors	Bioconductor Package Maintainer	Foundation of vector-like and list-like containers in Bioconductor	4
AnnotationDbi	Bioconductor Package Maintainer	Manipulation of SQLite-based annotations in Bioconductor	5
	Bioconductor		

<https://bioconductor.org>

INSTALLING BIOCONDUCTOR PACKAGES

- ▶ Installing packages from Bioconductor is different than the standard R method: don't use `install.packages()`

- ▶ First you need to install the Bioconductor framework;

Within an R session, type:

```
if (!requireNamespace("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")  
BiocManager::install()
```

you only need to do this once (or when you upgrade)

- ▶ To install individual packages, type:

```
BiocManager::install("package_name")
```

- ▶ Refer to “Install” on the Bioconductor site for more info

NOTABLE BIOCONDUCTOR MS PACKAGES

- ▶ **MSnBase**

infrastructure for reading, processing and analyzing MS data

- ▶ **mzR**

unified API for reading a variety of MS data formats

- ▶ **MassSpecWavelet**

MS spectrum processing tools

- ▶ **xcms**

comprehensive set of tools for MS analysis

XCMS HAS A GREAT TUTORIAL ON LCMS PROCESSING/ANALYSIS

- ▶ LCMS data preprocessing and analysis with xcms
<https://bioconductor.org/packages/release/bioc/vignettes/xcms/inst/doc/xcms.html>
- ▶ Covers
 - ▶ Loading data ("on disk")
 - ▶ High-level data review
 - ▶ Chromatographic peak detection
 - ▶ Retention time alignment, and cross-experiment feature grouping

Demo Time

Using example data from the OpenMS Tutorial data sets

<https://www.openms.de/tutorials/>

Download the Example Data

Example_Data/Labelfree/datasets/lfq_spikein_dilution_1.mzML