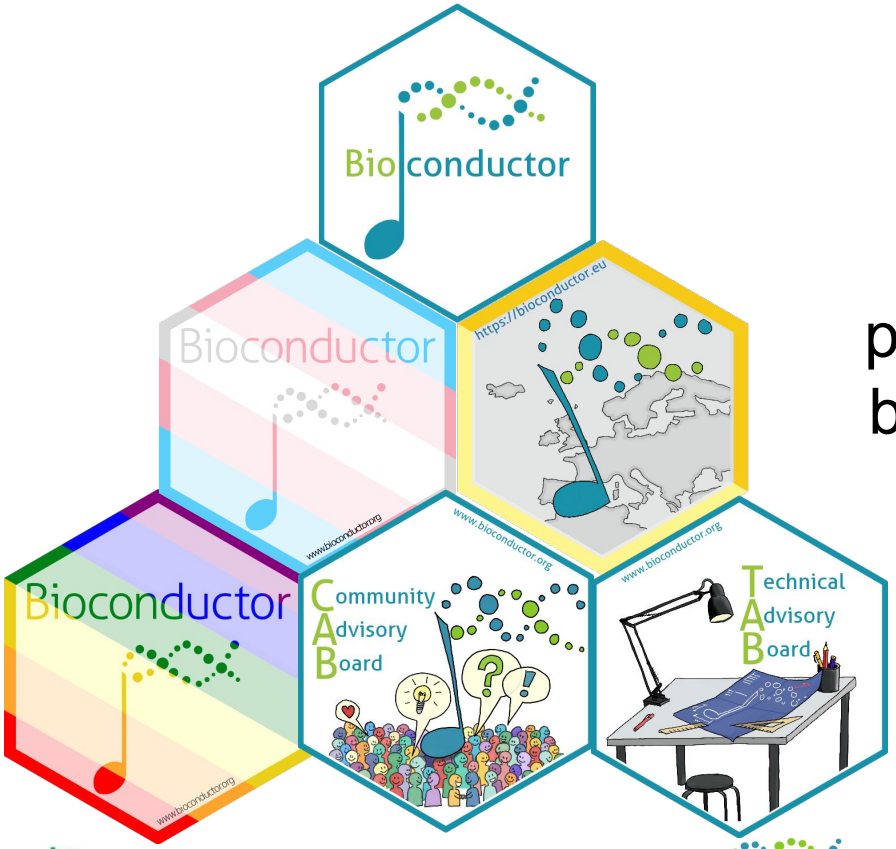


Bioconductor:

possibly the world's most impactful bioinformatics open source project

Prof Aedin Culhane
National Open Source Innovation Summit 2026

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

LERO

Dr Nicolas P Cooley
Global Developer Manager

Dr Maria Doyle
Global Community Manager

Temporary Slide with links

Background

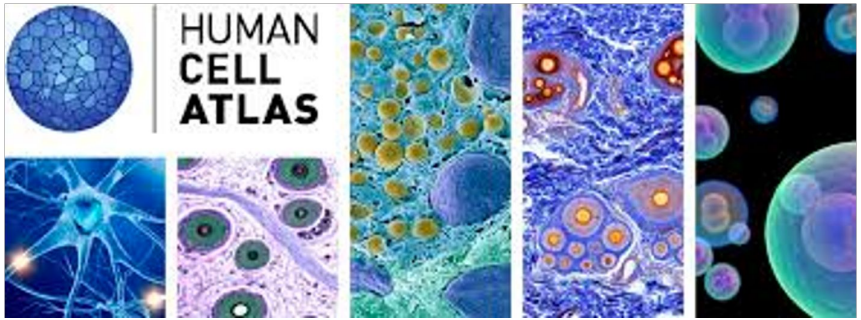
Bioconductor. (common R) - OHDSI (impossible to find packages)- Open not structured and not quality No review/Documentation/Many githubs

Bias in AI Danger of Poor Quality Software - Arifacts magnified
https://news.rice.edu/news/2024/breaking-mad-generative-ai-could-break-internet

AI continent and Data Spaces/Data Labs
Ethical, cost of AI - Better Software
RSQkit ELIXIR (organization thinking about data mgt/software quality)

Our body has 37,000,000,000,000 Cells

In the past 2 decades, life sciences has been revolutionized by the capacity to measure biology molecules in vast detail

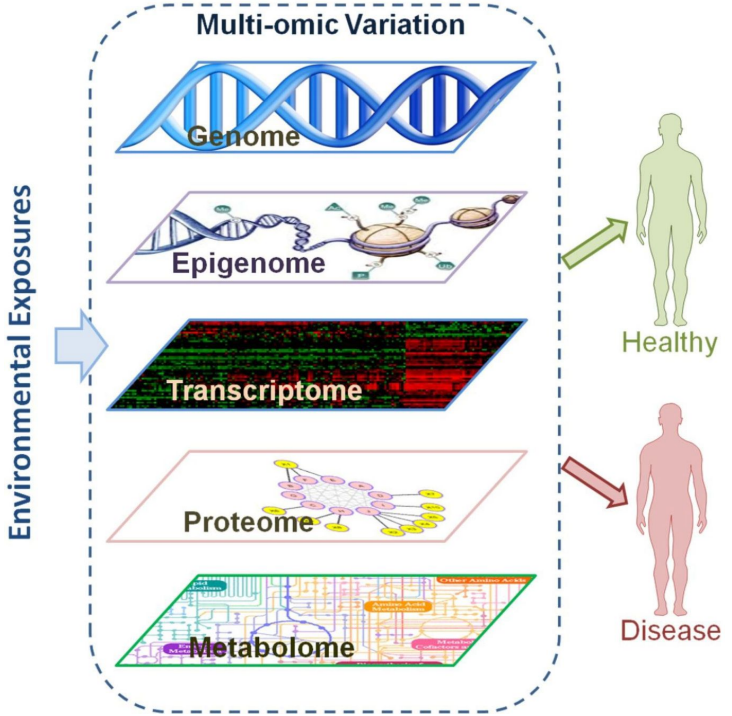


Why Bioconductor Matters

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

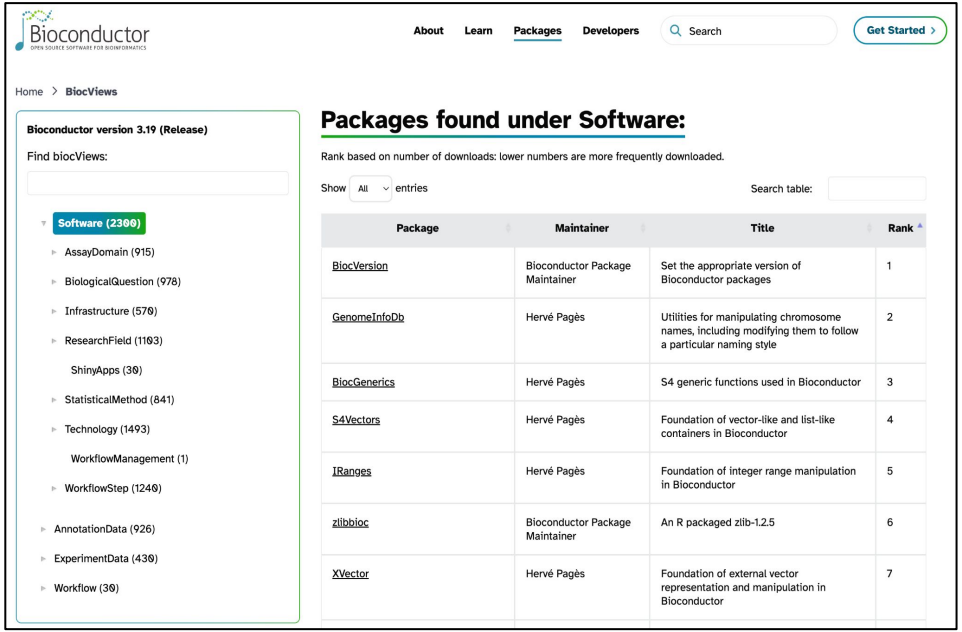
Large volumes of complex data require advanced, efficient, fast, user-friendly statistical & AI tools

- Biology data impact Medicine, Biodiversity, Agriculture, Marine
- Plants, Pathogens, Biodiversity, Microbiology, food
- Understanding Disease
- Discovery of biomarkers, new therapeutic targets
- 80% new cancer medicines are targeted



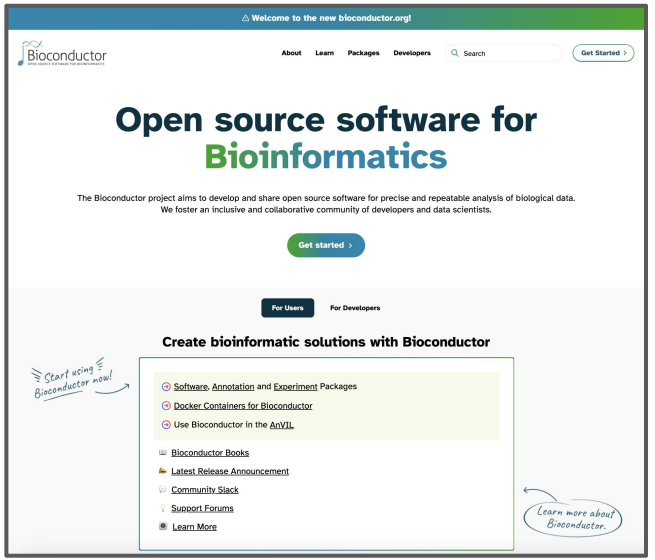
A ecosystem for health & life sciences

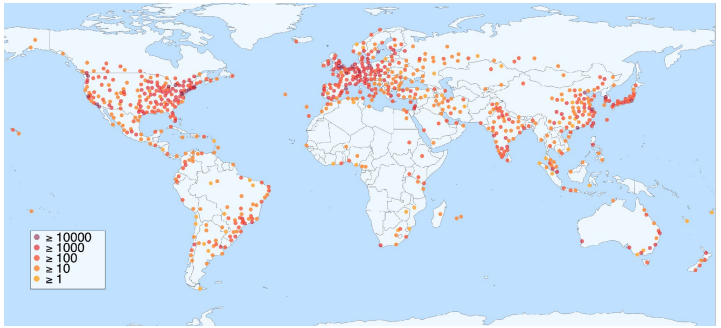
- Supports analysis of many biological data (genomics, transcriptomics, proteomics,
- >2,000 packages; data wrangling, statistical analysis & AI, reporting, visualization, interactive tools Written in R.
- Optimized underlying methods in multiple languages including C++, Fortran
- Strong methods for interoperability with other 4GL Python, Julia



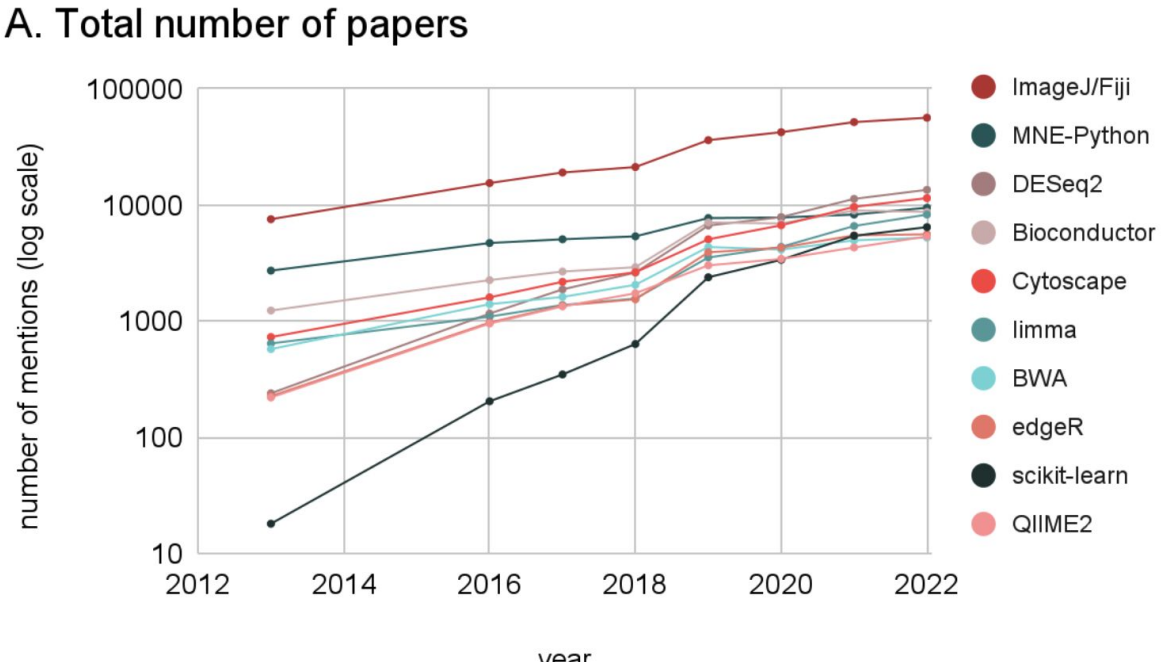
Bioconductor in 2026

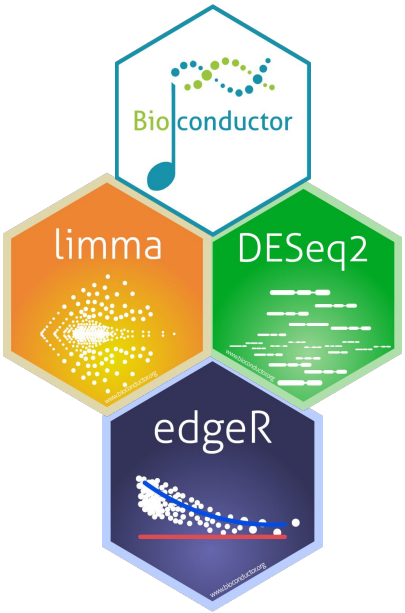
- Open source, open development
- Now version 3.22
- Community-driven
- Core value supporting users to become developers.
- 2,000+ contributed packages,
- Downloads >1 million per year
- Global





Data shows Bioconductor is essential software





4/10 top cited biomedical OSS are Bioconductor

CZI impact report : https://doi.org/10.5281/zenodo.11201216

Bioconductor as Software Infrastructure

Challenges at the intersection of biology and computer science

- Data scale - DNA sequencing data alone (in individual instution repositories!) is on the petabyte scale
- Complex hardware landscape
- Difficulty retaining talent and institutional knowledge

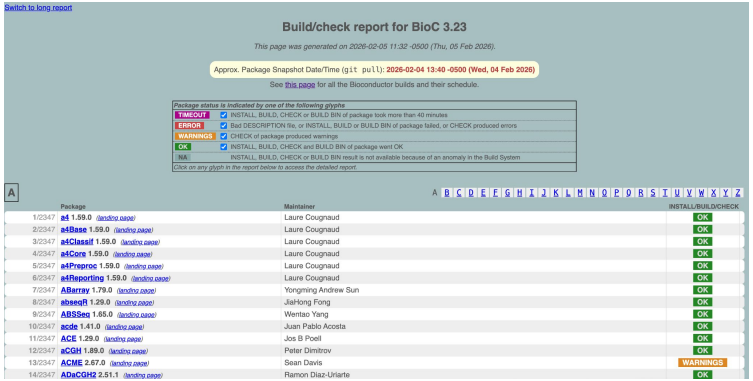
Bioconductor Strengths

- Historically robust maintenance and governance
- Metronomic versioning and release schedule emphasizing reliability
- Diverse community with strong and mixed technical backgrounds

Reproducibility, Reusability and Efficiency

- Bioconductor is package **repository** (SVN -> github)
- **Rigorous review** process* for packages (both automated and manual)
- **Daily build and check*** of all packages on multiple OS platforms. Issues reported to developers
- **Standardized data structures** and interoperability
- Rigour in **documentation** & vignette tutorial with examples*
- FAIR before FAIR
- Software quality -> Accepted by Pharma Regulatory
- **Connected** to Global OSS projects, Galaxy, Python scVerse, OHDSI, ELIXIR

* Stable, Quality, Trusted, Respected



Partner with us…

How can our learnings help you

- Build global community
- Increase Robustness, rigour and quality
- Software for vast data

Our Challenges

Distributed & Federated Compute

European Health Data Space (EHDS)

Software quality for EU AI Data Lab, AI Test Beds

GPU native

Funding is currently USA centred

- NIH U54, NCI
- CZI EOSS 4.0, 6.0
- Small PI grants

Seeking to diversity funding and engage with industry

- EU Horizon, Digital Europe
- IHI, DTIF