

Bioconductor



Bioconductor – informacje podstawowe

- <https://www.bioconductor.org/>
- Aktualna wersja 3.6
- Dwa wydania rocznie
- 1473 pakiety (stan na marzec 2018)

Bioconductor – informacje podstawowe

- Wykorzystuje język statystyczne R
- Rozwijane jako otwarte oprogramowanie
- Społeczność developerów i użytkowników

Bioconductor – core team

- Hervé Pagès
- Michael Lawrence
- Qian Liu
- James MacDonald
- Martin Morgan (Project Lead)
- Valerie Obenchain
- Andrzej Oleś
- Marcel Ramos
- Lori Ann Shepherd
- Nitesh Turaga
- Daniel van Twisk

Bioconductor – cele projektu

- Zapewnienie dostępu do potężnego narzędzia statystycznego oraz metod graficznego prezentowania danych genomowych
- Ułatwienie integracji metadanych biologicznych (GenBank, GO, LocusLink, PubMed) w analiza danych eksperymentalnych

Bioconductor – cele projektu

- Pozwala na szybki rozwój rozszerzalnego, interoperacyjnego i skalowalnego oprogramowanie
- Promuj wysokiej jakości dokumentację i powtarzalność badań
- Zapewnij szkolenie z zakresu metod obliczeniowych i statystycznych

Bioconductor - instalacja

- Własny system instalacji, częściowo niezależny od standardowego mechanizmu R - czyli funkcji `install.packages()`
- Umożliwia koordynację między pakietami i lepszą kontrolę błędów - (usuwa potencjalne niezgodności między wersjami R'a i pakietów oraz pakietów między sobą)
- Dostarczona przez Bioconductora funkcja `biocLite()` jest nakładką na `install.packages()` zapewniającą instalację poprawnej wersji pakietów zgodnej z wersją Bioconductora

Bioconductor - instalacja

- Instalacja Bioconductor:

```
source("https://bioconductor.org/biocLite.R") biocLite()
```

- Instalacja pakietów Bioconductor

```
biocLite(c("GenomicFeatures",  
"AnnotationDbi"))
```


Bioconductor - pakiety

Bioconductor zapewnia dwie główne klasy pakiety oprogramowania:

- Pakiety „End-User”:
 - skierowane do użytkowników niezaznajomionych z R lub programowaniem
 - dopracowane i łatwe w użyciu interfejsy do szerokiej gamy metod obliczeniowych i statystycznych dla analiza danych genomowych
- Pakiety dla deweloperów: oprogramowanie ukierunkowane na programistów, w tym sensie, że zapewniają oprogramowanie do pisania oprogramowania.

Bioconductor - pakiety

- Pakiety danych:
 - Biologiczne metadane: odwzorowania między różnymi identyfikatorami genów (np. AffyID, GO, LocusID, PMID), CDF i informacje o sekwencji sond dla macierzy Affy
 - Dane eksperymentalne: kod, dane i dokumentacja dla konkretnych eksperymentów lub projektów.
 - Pakiety kursów: kod, dane, dokumentacja, i przykłady do nauki konkretnego kursu

Bioconductor - zastosowania

Pakiety Bioconductora są najczęściej zorganizowane w potoki przetwarzania (workflows). Pakiety wykonują dobrze zdefiniowane zadania cząstkowe i łączą się z innymi używając wspólnych struktur danych

Przykładowe zastosowania

- Sequence Analysis
- Oligonucleotide Arrays
- Annotation Resources
- Annotating Genomic Ranges
- Annotating Genomic Variants
- RNA seq
 - edgeR
 - DESeq2