

BRM5012 R practical: Bioconductor

Flavours of R

Base R

Functions and types that are always available in R.

CRAN packages

```
# Example  
install.packages("glmnet")
```

Tidyverse

```
install.packages("tidyverse")
```

Bioconductor

```
# Setup  
source("https://bioconductor.org/biocLite.R")  
biocLite()  
# Install a specific package  
biocLite("limma")
```

What Bioconductor is for

Differential expression analysis from microarray or RNA-seq

- ▶ limma, edgeR, DESeq2, ...
- ▶ Further packages to normalize, impute, batch correct, check quality, ...

Visualization

- ▶ ComplexHeatmap, Gviz, ggbio, ...

Peaks, regions, motifs in genome (eg ChIP-seq)

- ▶ GenomicRanges, GenomicFeatures, ...

Single-cell sequencing

Statistical methods for $p \gg n$ data

...

More types

Up to now we have been using vectors and data frames (specifically Tidyverse “tibbles”).

To use Bioconductor, several further types are important.

- ▶ `matrix`
- ▶ `DNASTring`, `DNASTringSet`
- ▶ `GRanges`
- ▶ Others: `SeqInfo`, `TxDb`, `EnsDb`, `OrgDb`, `SummarizedExperiment`,
...

Getting reference data

Analysis usually occurs in the context of:

- ▶ a genome assembly: DNA sequences of chromosomes
 - ▶ DNASTringSet, or BSgenome or TwoBitFile
- ▶ gene annotations: ranges of positions in chromosome and strand
 - ▶ TxDb or EnsDb

In Bioconductor:

- ▶ AnnotationHub package gives access to reference data from various sources
 - ▶ access specific versions, good for reproducibility
- ▶ rtracklayer package to load files

Others

- ▶ biomaRt to access BioMart servers
- ▶ Data packages

Documentation

Most packages have “vignettes” explaining normal usage.

Information about types available with, eg:

```
? "GRanges-class"
```