

Bioconductor Developer Day


Martin Morgan (mtmorgan@fhcrc.org)
Fred Hutchinson Cancer Research Center
Seattle, WA

17 July 2013

Orientation

- ▶ Campus map¹
- ▶ Arnold Building
- ▶ Developer Day²

¹<http://www.fhcrc.org/content/dam/public/contact-us/Visit-Us/campusmap.pdf>

²<http://bioconductor.org/developers/developer-day/> 

Acknowledgements & Introductions

- ▶ Abby Stimmel
- ▶ *Bioconductor* team: Marc Carlson, Valerie Obenchain, Hervé Pagès, Paul Shannon, Dan Tenenbaum
- ▶ Technical advisory council: Vincent Carey, Wolfgang Huber, Robert Gentleman, Rafael Irizzary, Sean Davis, Kasper Hansen
- ▶ Scientific advisory board: Simon Tavaré, Vivian Bonazzi, Vincent Carey, Wolfgang Huber, Robert Gentleman, Rafael Irizzary, Paul Flicek, Simon Urbanek.
- ▶ NIH / NHGRI U41HG0004059
- ▶ ...and the *Bioconductor* community!

Core team highlights

BioInstaller

- ▶ `useDevel()`
- ▶ `biocValid()`
- ▶ Encouraging updates

```
> R.version.string  
[1] "R version 2.15.3 Patched (2013-03-13 r62579)"  
> source("http://bioconductor.org/biocLite.R")  
Your Bioconductor is out-of-date, upgrade to version  
2.12 by following instructions at  
http://bioconductor.org/install.
```

BiocStyle

- ▶ L^AT_EX (and other. . .) style for consistent, easy-to-use Sweave documents
- ▶ DESCRIPTION file – Suggests: BiocStyle
- ▶ Vignette.Rnw preamble:

```
<<style, eval=TRUE, echo=FALSE, results=tex>>=
BiocStyle::latex()
@
```
- ▶ Example in the package vignette³

³<http://bioconductor.org/packages/devel/bioc/vignettes/BiocStyle/inst/doc/LatexStyle.pdf>

BiocParallel

“Modified versions and novel implementation of functions for parallel evaluation, tailored to use with Bioconductor objects.”

```
> library(BiocParallel)
```

```
> registered()
```

```
> bplapply(1:8, function(i) Sys.sleep(1))
```

```
> bpvec(1:32, sqrt, AGGREGATE=list)
```

- ▶ Params: Serial, Multicore, Snow, DoPar, (BatchJobs)
- ▶ register, bpschedule (e.g., recursive=FALSE), (bpstart, bpstop), bpvectorize
- ▶ Example benefit: bplapply subsets rather than coerces to list

Cloud-aided Computation

The AMI

- ▶ Pre-configured, easy to deploy & use

AnnotationHub

- ▶ Genome-scale resources
- ▶ Centralized, readily updated & available

Future scenarios

- ▶ Data proximity, e.g., Ensembl servers in same zone
- ▶ Light-weight computation, e.g.,
 - ▶ Overlap regions-of-interest with each ENCODE track
 - ▶ Re-dispatch *ensemblVEP* annotation tasks