Bioc 2014

Sonali Arora

Developer Day, July 30, 2014

Outline

- Website organization
- biocViews
- Videos
- GenomeInfoDb

Website Organisation

- 4 squares for 'install' ,'learn', 'use', 'develop' motivated by the idea of different Bioconductor Users.
- Another motivation was to make the most accessed material easily accessed from the main landing page/ Home page.

Video

- YouTube Channel - `http://www.youtube.com/user/bioconductor`
- Instructional videos which complement traditional vignettes
- Package Contributors can add a Video: Tag in DESCRIPTION file containing link of their video.

Update biocViews

- enrich the biocViews vocabulary
- autocomplete biocViews search box - find packages faster
- find by package name / author name / keywords.

GenomeInfoDb

- This package provides an interface to access seqlevelsStyle() and their supported mappings for various organisms.
- It supports 10 organisms which can be found with

```
library(GenomeInfoDb)
names(genomeStyles())
```

```
##  [1] "Arabidopsis_thaliana"     "Caenorhabditis_elegans"
##  [3] "Cyanidioschyzon_merolae"  "Drosophila_melanogaster"
##  [5] "Homo_sapiens"             "Mus_musculus"
##  [7] "Oryza_sativa"             "Populus_trichocarpa"
##  [9] "Saccharomyces_cerevisiae" "Zea_mays"
```

keepStandardChromosomes

- Subsetting operation that returns only the 'standard' Chromosomes.
- We define 'standard Chromosomes' as those Chromosomes which represent sequences in the assembly that are not scaffolds.
- Applicable when X has a Seqinfo object.
- multiple organisms supported.

demo

```
library(parathyroidSE)
data(exonsByGene)
```

```
GRangesList of length 100:
$ENSG00000000003
GRanges with 17 ranges and 2 metadata columns:
       seqnames              ranges strand   |   exon_id       exon_name
          <Rle>           <IRanges>  <Rle>   | <integer>     <character>
   [1]        X [99883667, 99884983]      -   |    664095 ENSE00001459322
   [2]        X [99885756, 99885863]      -   |    664096 ENSE00000868868
   [3]        X [99887482, 99887565]      -   |    664097 ENSE00000401072
   [4]        X [99887538, 99887565]      -   |    664098 ENSE00001849132
   [5]        X [99888402, 99888536]      -   |    664099 ENSE00003554016
   ...      ...                 ...    ... ...       ...               ...
  [13]        X [99890555, 99890743]      -   |    664106 ENSE00003512331
  [14]        X [99891188, 99891686]      -   |    664108 ENSE00001886883
  [15]        X [99891605, 99891803]      -   |    664109 ENSE00001855382
  [16]        X [99891790, 99892101]      -   |    664110 ENSE00001863395
  [17]        X [99894942, 99894988]      -   |    664111 ENSE00001828996

...
<99 more elements>
---
seqlengths:
                   1                   2 ...              LRG_98              LRG_99
           249250621           243199373 ...               18750               13294
```

demo

```
std_exons <- keepStandardChromosomes(exonsByGene)
```

```
GRangesList of length 100:
$ENSG00000000003
GRanges with 17 ranges and 2 metadata columns:
       seqnames               ranges strand   |   exon_id       exon_name
          <Rle>            <IRanges>  <Rle>   | <integer>     <character>
   [1]        X [99883667, 99884983]      -   |    664095 ENSE00001459322
   [2]        X [99885756, 99885863]      -   |    664096 ENSE00000868868
   [3]        X [99887482, 99887565]      -   |    664097 ENSE00000401072
   [4]        X [99887538, 99887565]      -   |    664098 ENSE00001849132
   [5]        X [99888402, 99888536]      -   |    664099 ENSE00003554016
   ...      ...                  ...     ... ...       ...             ...
  [13]        X [99890555, 99890743]      -   |    664106 ENSE00003512331
  [14]        X [99891188, 99891686]      -   |    664108 ENSE00001886883
  [15]        X [99891605, 99891803]      -   |    664109 ENSE00001855382
  [16]        X [99891790, 99892101]      -   |    664110 ENSE00001863395
  [17]        X [99894942, 99894988]      -   |    664111 ENSE00001828996

...
<99 more elements>
---
seqlengths:
        1         2         3         4         5        MT         X ...        18        19        20        21        22         Y
249250621 243199373 198022430 191154276 180915260     16569 155270560 ...  78077248  59128983  63025520  48129895  51304566  59373566
```

demo

```
std_exons <- sortSeqlevels(std_exons)
```

```
GRangesList of length 100:
$ENSG00000000003
GRanges with 17 ranges and 2 metadata columns:
       seqnames              ranges strand   |   exon_id       exon_name
          <Rle>           <IRanges>  <Rle>   | <integer>     <character>
   [1]        X [99883667, 99884983]      -   |    664095 ENSE00001459322
   [2]        X [99885756, 99885863]      -   |    664096 ENSE00000868868
   [3]        X [99887482, 99887565]      -   |    664097 ENSE00000401072
   [4]        X [99887538, 99887565]      -   |    664098 ENSE00001849132
   [5]        X [99888402, 99888536]      -   |    664099 ENSE00003554016
   ...      ...                 ...    ... ...       ...             ...
  [13]        X [99890555, 99890743]      -   |    664106 ENSE00003512331
  [14]        X [99891188, 99891686]      -   |    664108 ENSE00001886883
  [15]        X [99891605, 99891803]      -   |    664109 ENSE00001855382
  [16]        X [99891790, 99892101]      -   |    664110 ENSE00001863395
  [17]        X [99894942, 99894988]      -   |    664111 ENSE00001828996

...
<99 more elements>
---
seqlengths:
        1         2         3         4         5         6         7 ...        20        21        22         X         Y        MT
249250621 243199373 198022430 191154276 180915260 171115067 159138663 ...  63025520  48129895  51304566 155270560  59373566     16569
```

demo

```
reads <- GRanges(seqnames=paste0("chr",1:20),
                 ranges=IRanges(1:20, width=5))
cnt1 <- countOverlaps(std_exons, reads)
```

```
ENSG00000000003 ENSG00000000005 ENSG00000000419 ENSG00000000457 ENSG00000000460 ENSG00000000938 ENSG00000000971 ENSG00000001036 ENSG00000001084
              0               0               0               0               0               0               0               0               0
ENSG00000001460 ENSG00000001461 ENSG00000001497 ENSG00000001561 ENSG00000001617 ENSG00000001626 ENSG00000001629 ENSG00000001630 ENSG00000001631
              0               0               0               0               0               0               0               0               0
ENSG00000002079 ENSG00000002330 ENSG00000002549 ENSG00000002586 ENSG00000002587 ENSG00000002726 ENSG00000002745 ENSG00000002746 ENSG00000002822
              0               0               0               0               0               0               0               0               0
ENSG00000002919 ENSG00000002933 ENSG00000003056 ENSG00000003096 ENSG00000003137 ENSG00000003147 ENSG00000003249 ENSG00000003393 ENSG00000003400
              0               0               0               0               0               0               0               0               0
ENSG00000003436 ENSG00000003509 ENSG00000003756 ENSG00000003987 ENSG00000003989 ENSG00000004059 ENSG00000004139 ENSG00000004142 ENSG00000004399
              0               0               0               0               0               0               0               0               0
ENSG00000004468 ENSG00000004478 ENSG00000004487 ENSG00000004534 ENSG00000004660 ENSG00000004700 ENSG00000004766 ENSG00000004776 ENSG00000004777
              0               0               0               0               0               0               0               0               0
ENSG00000004799 ENSG00000004809 ENSG00000004838 ENSG00000004846 ENSG00000004848 ENSG00000004864 ENSG00000004866 ENSG00000004897 ENSG00000004939
              0               0               0               0               0               0               0               0               0
ENSG00000004961 ENSG00000004975 ENSG00000005001 ENSG00000005007 ENSG00000005020 ENSG00000005022 ENSG00000005059 ENSG00000005073 ENSG00000005075
              0               0               0               0               0               0               0               0               0
ENSG00000005102 ENSG00000005108 ENSG00000005156 ENSG00000005175 ENSG00000005187 ENSG00000005189 ENSG00000005194 ENSG00000005206 ENSG00000005238
              0               0               0               0               0               0               0               0               0
ENSG00000005249 ENSG00000005302 ENSG00000005339 ENSG00000005379 ENSG00000005381 ENSG00000005421 ENSG00000005436 ENSG00000005448 ENSG00000005469
              0               0               0               0               0               0               0               0               0
Warning message:
In .Seqinfo.mergexy(x, y) :
  Each of the 2 combined objects has sequence levels not in the other:
  - in 'x': 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, Y, MT
  - in 'y': chr1, chr2, chr3, chr4, chr5, chr6, chr7, chr8, chr9, chr10, chr11, chr12, chr13, chr14, chr15, chr16, chr17, chr18, chr19, chr20
  Make sure to always combine/compare objects based on the same reference
  genome (use suppressWarnings() to suppress this warning).
```

demo

```
seqlevelsStyle(std_exons) <- seqlevelsStyle(reads)
cnt2 <- countOverlaps(std_exons, reads)
```

ENSG00000000003	ENSG00000000005	ENSG00000000419	ENSG00000000457	ENSG00000000460	ENSG00000000938	ENSG00000000971	ENSG00000001036
0	0	0	0	0	0	0	0
ENSG00000001461	ENSG00000001497	ENSG00000001561	ENSG00000001617	ENSG00000001626	ENSG00000001629	ENSG00000001630	ENSG00000001631
0	0	0	0	0	0	0	0
ENSG00000002549	ENSG00000002586	ENSG00000002587	ENSG00000002726	ENSG00000002745	ENSG00000002746	ENSG00000002822	ENSG00000002834
0	0	0	0	0	0	0	0
ENSG00000003096	ENSG00000003137	ENSG00000003147	ENSG00000003249	ENSG00000003393	ENSG00000003400	ENSG00000003402	ENSG00000003436
0	0	0	0	0	0	0	0
ENSG00000003989	ENSG00000004059	ENSG00000004139	ENSG00000004142	ENSG00000004399	ENSG00000004455	ENSG00000004468	ENSG00000004478
0	0	0	0	0	0	0	0
ENSG00000004700	ENSG00000004766	ENSG00000004776	ENSG00000004777	ENSG00000004779	ENSG00000004799	ENSG00000004809	ENSG00000004838
0	0	0	0	0	0	0	0
ENSG00000004866	ENSG00000004897	ENSG00000004939	ENSG00000004948	ENSG00000004961	ENSG00000004975	ENSG00000005001	ENSG00000005007
0	0	0	0	0	0	0	0
ENSG00000005073	ENSG00000005075	ENSG00000005100	ENSG00000005102	ENSG00000005108	ENSG00000005156	ENSG00000005175	ENSG00000005187
0	0	0	0	0	0	0	0
ENSG00000005238	ENSG00000005243	ENSG00000005249	ENSG00000005302	ENSG00000005339	ENSG00000005379	ENSG00000005381	ENSG00000005421
0	0	0	0	0	0	0	0

Refrences

1. Arora S, Morgan M, Carlson M and Pages H. GenomeInfoDb: Utilities for manipulating chromosome and other 'seqname' identifiers. R package version 1.1.17.
2. Haglund, F., et al (2012). Evidence of a functional estrogen receptor in parathyroid adenomas. J. Clin. Endocrinol. Metab., 97(12), pp. 4631-4639
3. Carey V, Harshfield B and Falcon S (2014). biocViews: Categorized views of R package repositories. R package version 1.32.1, `http://www.bioconductor.org/packages/release/BiocViews.html`

Bioconductor is supported in part by grants NHGRI / NIH U41HG004059 and NSF 1247813

Thank You

Questions?

Thank you!