

SYSKA: Bioconductor

IRanges, GenomicRanges, and AnnotationHub

Murray Cadzow

Bioconductor

www.bioconductor.org

Online repository for biologically focussed R packages

Higher requirement for package documentation

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

```
> biocLite()
```

IRanges - biocLite("IRanges")

Package for manipulating intervals on sequences

Contains start, end and width - specify 2+ on creation

lots of built in methods such as:

start(), end(), widths()

shift(), narrow(), flank(), resize(), restrict()

findOverlaps(), subsetByOverlaps(), reduce(), disjoin(), identical(), intersect(),
union()

```
> library("IRanges")
```

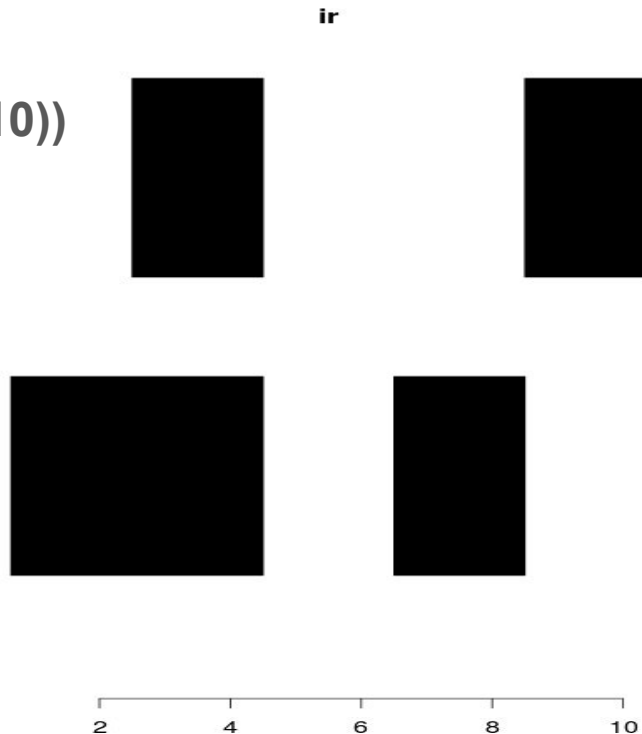
```
> ir <- IRanges(start = c(1,3,7,9), end = c(4,4,8,10))
```

```
> ir
```

IRanges of length 4

start end width

[1]	1	4	4
[2]	3	4	2
[3]	7	8	2
[4]	9	10	2



> **reduce(ir)**

IRanges of length 2

start end width

[1] 1 4 4

[2] 7 10 4

reduce(ir)



> **disjoin(ir)**

IRanges of length 4

start end width

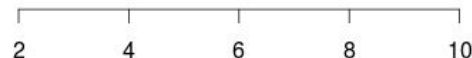
[1] 1 2 2

[2] 3 4 2

[3] 7 8 2

[4] 9 10 2

disjoin(ir)



```
> df1 <- DataFrame(iranges = ir)
```

```
> df1
```

DataFrame with 4 rows and 1 column

iranges

<IRanges>

1 [1, 4]

2 [3, 4]

3 [7, 8]

4 [9, 10]

```
> df1$iranges
```

IRanges of length 4

start end width

[1] 1 4 4

[2] 3 4 2

[3] 7 8 2

[4] 9 10 2

GenomicRanges - biocLite("GenomicRanges")

IRanges plus metadata -> GRanges

Has some extra methods e.g. promoters()

Need to now include seqnames (chromosome) and strand

```
> library("GenomicRanges")  
> gr <- GRanges(seqnames = "chr1", strand = c("+", "-", "+"),  
+               ranges = IRanges(start = c(1,3,5), width = 3))  
> values(gr) <- DataFrame(score = c(0.1, 0.5, 0.3))  
> gr
```

GRanges object with 3 ranges and 1 metadata column:

	seqnames	ranges	strand	score
	<Rle>	<IRanges>	<Rle>	<numeric>
[1]	chr1	[1, 3]	+	0.1
[2]	chr1	[3, 5]	-	0.5
[3]	chr1	[5, 7]	+	0.3

seqinfo: 1 sequence from an unspecified genome; no seqlengths

AnnotationHub - `biocLite("AnnotationHub")`

Provides access to multiple annotation sources

e.g. NCBI, ensembl, encode, UCSC

Includes metadata about each resource you can query and download


```
>ah <- AnnotationHub()
```

```
> ah
```

```
AnnotationHub with 35345 records
```

```
# snapshotDate(): 2015-08-26
```

```
# $dataprovder: BroadInstitute, UCSC, Ensembl, NCBI, Haemcode, Inparanoid8, Pazar, d...
```

```
# $species: Homo sapiens, Mus musculus, Bos taurus, Pan troglodytes, Danio rerio, Rat...
```

```
# $rdataclass: GRanges, FaFile, BigWigFile, OrgDb, ChainFile, Inparanoid8Db, TwoBitFi...
```

```
# additional mcols(): taxonomyid, genome, description, tags, sourceurl,
```

```
# sourcetype
```

```
# retrieve records with, e.g., 'object[["AH2"]]'
```

```
title
```

```
AH2 | Ailuropoda_melanoleuca.ailMel1.69.dna.toplevel.fa
```

```
AH3 | Ailuropoda_melanoleuca.ailMel1.69.dna_rm.toplevel.fa
```

```
AH4 | Ailuropoda_melanoleuca.ailMel1.69.dna_sm.toplevel.fa
```

```
AH5 | Ailuropoda_melanoleuca.ailMel1.69.ncrna.fa
```

```
AH6 | Ailuropoda_melanoleuca.ailMel1.69.pep.all.fa
```

```
... ..
```

```
AH49436 | Xiphophorus_maculatus.Xipmac4.4.2.dna_rm.toplevel.fa
```

```
AH49437 | Xiphophorus_maculatus.Xipmac4.4.2.dna_sm.toplevel.fa
```

```
AH49438 | Xiphophorus_maculatus.Xipmac4.4.2.dna.toplevel.fa
```

```
AH49439 | Xiphophorus_maculatus.Xipmac4.4.2.ncrna.fa
```

```
AH49440 | Xiphophorus_maculatus.Xipmac4.4.2.pep.all.fa
```

```
> ah <- subset(ah, species == "Homo sapiens")
> query(ah, "GRanges")
AnnotationHub with 14100 records
# snapshotDate(): 2015-08-26
# $dataProvider: BroadInstitute, UCSC, Ensembl, EncodeDCC
# $species: Homo sapiens
# $rdataclass: GRanges
# additional mcols(): taxonomyid, genome, description, tags, sourceurl,
#  sourcetype
# retrieve records with, e.g., 'object[["AH3166"]]'
```

title

```
AH3166 | wgEncodeRikenCageSkinshraCellPapTssHmm
AH3912 | wgEncodeUwDgfTregwb78495824Hotspots
AH3913 | wgEncodeUwDgfTregwb78495824Pk
AH4368 | wgEncodeUwDnaseWi38PkRep1
AH4369 | wgEncodeUwDnaseWi38PkRep2
...
AH46980 | E127_15_coreMarks_mnemonics.bed.gz
AH46981 | E128_15_coreMarks_mnemonics.bed.gz
AH46982 | E129_15_coreMarks_mnemonics.bed.gz
AH47066 | Homo_sapiens.GRCh38.80.gtf
AH47963 | Homo_sapiens.GRCh38.81.gtf
```