Working with Genomic Ranges

Mireia Ramos-Rodriguez

March 14, 2019

Introduction to GenomicRanges[1]

The GenomicRanges serves as the foundation for representing genomic locations within the Bioconductor project. This package lays a foundation for genomic analysis by introducing three classes (GRanges, GPos, and GRangesList), which are used to represent genomic ranges, genomic positions, and groups of genomic ranges.

```r
library(GenomicRanges)
z <- GRanges("chr1",IRanges(1000001,1001000),strand="+")
start(z)
end(z)
width(z)
strand(z)
mcols(z) # the 'metadata columns', any information stored alongside each range
ranges(z) # gives the IRanges
seqnames(z) # the chromosomes for each ranges
seqlevels(z) # the possible chromosomes
seqlengths(z) # the lengths for each chromosome
```

Intra-range methods

Affects ranges independently

function	description
shift	moves left/right
narrow	narrows by relative position within range
resize	resizes to width, fixing start for +, end for -
flank	returns flanking ranges to the left +, or right -
promoters	similar to flank
restrict	restricts ranges to a start and end position
trim	trims out of bound ranges
+/-	expands/contracts by adding/subtracting fixed amount
*	zooms in (positive) or out (negative) by multiples

Inter-range methods

Affects ranges as a group

function	description
range	one range, leftmost start to rightmost end

---

[1]Reference card extracted from mikelove/bioc-refcard

function	description
reduce	cover all positions with only one range
gaps	uncovered positions within range
disjoin	breaks into discrete ranges based on original starts/ends

Nearest methods

Given two sets of ranges, `x` and `subject`, for each range in `x`, returns...

function	description
nearest	index of the nearest neighbor range in subject
precede	index of the range in subject that is directly preceded by the range in x
follow	index of the range in subject that is directly followed by the range in x
distanceToNearest	distances to its nearest neighbor in subject (Hits object)
distance	distances to nearest neighbor (integer vector)

A Hits object can be accessed with `queryHits`, `subjectHits` and `mcols` if a distance is associated.

set methods

If `y` is a GRangesList, then use `punion`, etc. All functions have default `ignore.strand=FALSE`, so are strand specific.

```
union(x,y)
intersect(x,y)
setdiff(x,y)
```

Overlaps

```
x %over% y  # logical vector of which x overlaps any in y
fo <- findOverlaps(x,y) # returns a Hits object
queryHits(fo)    # which in x
subjectHits(fo) # which in y
```

Seqnames and seqlevels

GenomicRanges and GenomeInfoDb

```
gr.sub <- gr[seqlevels(gr) == "chr1"]
seqlevelsStyle(x) <- "UCSC" # convert to 'chr1' style from "NCBI" style '1'
```