DNA Motif Search Example

Anatoly Dryga

9/22/2015

Biostrings package has function **matchDNAPattern** that solves the problem for us. For more details on see http://ugrad.stat.ubc.ca/R/library/Biostrings/html/matchDNAPattern.html.

Note that now it is called **matchPattern**!

And the pattern is represented by a standard DNA alphabet, see http://ugrad.stat.ubc.ca/R/library/Biostrings/html/DNAPatternAlphabet.html.

Here is an example for exact match:

```
pattern <- 'GATC'
contig <- DNAString('AAAGATCCCC')
matchPattern(pattern, contig, fixed=FALSE)
```

```
##   Views on a 10-letter DNAString subject
## subject: AAAGATCCCC
## views:
##     start end width
## [1]     4   7     4 [GATC]
```

Note that by default reverse complement is NOT searched:

```
pattern <- 'GGGG'
contig <- DNAString('AAACCCC')
matchPattern(pattern, contig, fixed=FALSE)
```

```
##   Views on a 7-letter DNAString subject
## subject: AAACCCC
## views: NONE
```

And here is an example with Ns:

```
pattern <- 'GGGNNNAAA'
contig <- DNAString('GGGTTTAAA')
matchPattern(pattern, contig, fixed=FALSE)
```

```
##   Views on a 9-letter DNAString subject
## subject: GGGTTTAAA
## views:
##     start end width
## [1]     1   9     9 [GGGTTTAAA]
```

We have to use `fixed` argument to do proper match with ambiguity nucleotides and we will use it for all cases for consistency.