Remove softclipped bases from BAM file

Toshihiro Arae

Load self-made package

```
import Pkg: activate, status
activate("../script_julia/RemoveSoftclip")
include("../script_julia/Fs.jl")
```

```
status()
```

```
Project RemoveSoftclip v0.1.0
Status `~/Dropbox/ngs_analysis/chiba_riboseq/script_julia/RemoveSoftclip/Project.toml`
  [6e4b80f9] BenchmarkTools v1.3.2
  [00701ae9] BioAlignments v3.1.0
  [336ed68f] CSV v0.10.10
  [a93c6f00] DataFrames v1.5.0
  [c2308a5c] FASTX v2.1.0
  [953ef90b] Fmt v0.1.0 `https://github.com/bicycle1885/Fmt.jl#main`
  [295af30f] Revise v3.5.2
  [2913bbd2] StatsBase v0.34.0
  [d759349c] XAM v0.3.1
```

Import some packages

```
import RemoveSoftclip; const rs = RemoveSoftclip
import XAM.BAM; const bam = BAM
import StatsBase; const sb = StatsBase
import DataFrames; const df = DataFrames
import CSV; const csv = CSV
```

Setup directory

```
work_dir = Fs.pwd() |> Fs.path_dir
in_dir = work_dir * "/data_preproc/mapped_by_star"
out_dir = work_dir * "/data_preproc/rm_softclip"
Fs.dir_create(out_dir)
```

Setup file paths

```
infs = Fs.dir_ls(in_dir)
infs = Fs.path_filter.(infs, r"_ribo.sort.bam$") |> x -> x[.!isnothing.(x)]
labels = Fs.path_ext_rm.(Fs.path_ext_rm.(infs))
infs = (in_dir * "/") .* Fs.path_ext_set.(labels, ".sort.bam")
outfs = (out_dir * "/") .* Fs.path_ext_set.(labels, ".sam")
```

Remove softclips from reads

```
# 1. read a bam record
# 2. remove softclips from CIGAR string
# 3. truncate sequence and quality string
```

```
# 4. add the lengths of softclip as auxiliary tags (e.g. LS:i:11\tRS:i:0)
# 5. write record to sam file
for i in 1:length(infs)
    @time rs.bam_rm_softclip(infs[i], outfs[i])
    temp_in_path = "data_preproc/rm_softclip/" * Fs.path_file(outfs[i])
    temp_out_path = Fs.path_ext_rm_set(temp_in_path, ".sort.bam")
    cmd_sort = "samtools view -@4 -buS $(temp_in_path) | samtools sort -@4 > $(temp_out_path)"
    cmd_idx = "samtools index -@8 $(temp_out_path)"
    run(`bash -c "$(cmd_sort)"`)
    run(`bash -c "$(cmd_idx)"`)
    run(`rm $(outfs[i])`)
end
```