Gviz visualization: Copy Number Variations

Iris Raes

4-05-2020

Installation of BiocManager and Gviz (= Bioconductor package)

```
# Install BiocManager if not yet installed
# Install Gviz package if not yet installed
if (!requireNamespace("BiocManager", quietly = TRUE)){
  install.packages("BiocManager")}
if (! requireNamespace("Gviz", quietly = TRUE)) {
  BiocManager::install("Gviz")
}
# Load BiocManager
# library(BiocManager)
```

Loading Gviz package

```
# Load Gviz package
library(Gviz) # Version 1.30.3
# Loading required package: S4Vectors
# Loading required package: stats4
# Loading required package: BiocGenerics
# Loading required package: parallel
# Loading required package: IRanges
# Loading required package: GenomicRanges
# Loading required package: GenomeInfoDb
# Loading required package: grid

# Set working directory (path/to/folder)
setwd("C:/Users/irisr/Desktop/CSV")
```

Function to create a table with GRCh37 coordinates

```
## separate function requires tidyr package
library(tidyr)
Gviz_table <- function(filename) {
  X <- read.csv(file=filename, header = TRUE, sep=";")
  ##################### GRCh37 coordinates in first assembly
  tablepart1 <- X[grep("^GRCh37", X$assembly1),]
  id1 <- tablepart1[1]
```

```
  if(any(names(tablepart1) == 'variant_region_id')){regionid1 <- tablepart1[2]}
  if(any(names(tablepart1) == 'study_ID')){studyid1 <- tablepart1[4]}
  Chr_1a <- tablepart1[["Chr_1"]]
  Chr_1a <- gsub("[A-Za-z]", "", Chr_1a)
  Chr_1a <- sub("", "chr", Chr_1a)
  assembly1 <- tablepart1[["assembly1"]]
  assembly1 <- gsub(".*?:", "", assembly1)
  assembly1 <- separate(data = as.data.frame(assembly1), col = assembly1,
                        into =   c("start", "end"), sep = "-")
  if(any(names(tablepart1) == 'variant_region_id') &
     any(names(tablepart1) == 'study_ID')){
    GRCh37_assembly1 <- data.frame(id1,Chr_1a,assembly1,regionid1,studyid1)
  } else {
    GRCh37_assembly1 <- data.frame(id1,Chr_1a,assembly1)
  }
  names(GRCh37_assembly1)[names(GRCh37_assembly1) == "assembly1"] <- "assembly"
  names(GRCh37_assembly1)[names(GRCh37_assembly1) == "Chr_1a"] <- "chr"
  #GRCh37_assembly1["genome"] <- "hg19"
  ##################### GRCh37 coordinates in second assembly
  tablepart2 <- X[grep("^GRCh37", X$assembly2),]
  id2 <- tablepart2[1]
  if(any(names(tablepart2) == 'variant_region_id')){regionid2 <- tablepart2[2]}
  if(any(names(tablepart2) == 'study_ID')){studyid2 <- tablepart2[4]}
  Chr_1b <- tablepart2[["Chr_2"]]
  Chr_1b <- gsub("[A-Za-z]", "", Chr_1b)
  Chr_1b <- sub("", "chr", Chr_1b)
  assembly2 <- tablepart2[["assembly2"]]
  assembly2 <- gsub(".*?:", "", assembly2)
  assembly2 <- separate(data = as.data.frame(assembly2), col = assembly2,
                        into = c("start", "end"), sep = "-")
  if(any(names(tablepart2) == 'variant_region_id') &
     any(names(tablepart2) == 'study_ID')){
    GRCh37_assembly2 <- data.frame(id2,Chr_1b,assembly2,regionid2,studyid2)
  } else {
    GRCh37_assembly2 <- data.frame(id2,Chr_1b,assembly2)
  }
  names(GRCh37_assembly2)[names(GRCh37_assembly2) == "assembly2"] <- "assembly"
  names(GRCh37_assembly2)[names(GRCh37_assembly2) == "Chr_1b"] <- "chr"
  #GRCh37_assembly2["genome"] <- "hg19"
  ##### Gviz table
  GRCh37_Gviz <- rbind(GRCh37_assembly1,GRCh37_assembly2)
  GRCh37_Gviz <- GRCh37_Gviz[order(-GRCh37_Gviz[1]),]
}
```

Input tables and GRanges Object

```
### results-CNV-dbVar.csv
csvfile <- read.csv(file="results-CNV-dbVAR.csv", header = TRUE, sep=";")
csvfile
```

```
##    CNV_variant_id variant_region_id                      type study_ID
```

```
## 1        49623411        nsv4457776 ['copy number variation']  nstd102
## 2        49353191        nsv4358278 ['copy number variation']  nstd102
## 3        49353005        nsv4358092 ['copy number variation']  nstd102
## 4        49350830        nsv4355917 ['copy number variation']  nstd102
## 5        49349701        nsv4354788 ['copy number variation']  nstd102
## 6        49349293        nsv4354380 ['copy number variation']  nstd102
## 7        49345450        nsv4350537 ['copy number variation']  nstd102
## 8        49344315        nsv4349402 ['copy number variation']  nstd102
## 9        48468240        nsv3904885 ['copy number variation']  nstd102
## 10       48466558        nsv3903203 ['copy number variation']  nstd102
## 11       48466447        nsv3903092 ['copy number variation']  nstd102
## 12       48453939        nsv3890584 ['copy number variation']  nstd102
## 13       45807136        nsv2779094 ['copy number variation']   nstd37
## 14       17813982                                         []  nstd102
## 15       17813734                                         []  nstd101
## 16        3740775         nsv533414 ['copy number variation']   nstd37
## 17        3739972         nsv532611 ['copy number variation']  nstd101
## 18        3738955         nsv531594 ['copy number variation']  nstd101
## 19        3738954         nsv531593 ['copy number variation']  nstd101
## 20        3738649         nsv531288 ['copy number variation']  nstd101
## 21        1212838         nsv497563 ['copy number variation']   nstd37
## 22        1137112                                         []   nstd37
##    clinical_assertion Chr_1             assembly1 Chr_2
## 1      ['Pathogenic']    19  GRCh37:260911-4788357    19
## 2      ['Pathogenic']    19  NCBI36:184565-4650484    19
## 3      ['Pathogenic']    19  NCBI36:210395-6746622    19
## 4      ['Pathogenic']    19 NCBI36:3959558-4714171    19
## 5      ['Pathogenic']    19 NCBI36:3505633-4641977    19
## 6      ['Pathogenic']    19 NCBI36:1923244-9620555    19
## 7      ['Pathogenic']    19 GRCh37:3076808-4796782    19
## 8      ['Pathogenic']    19 GRCh37:3338022-4833151    19
## 9      ['Pathogenic']    19  GRCh37:68029-59110290    19
## 10     ['Pathogenic']    19 GRCh37:260912-59097160    19
## 11     ['Pathogenic']    19 GRCh37:260912-58956888    19
## 12     ['Pathogenic']    19 GRCh37:3120160-9732820    19
## 13     ['Pathogenic']    19 GRCh37:260912-58956888    19
## 14                 []    19 GRCh37:260912-58956888    19
## 15                 []    19 GRCh37:260912-58956888    19
## 16     ['Pathogenic']    19  NCBI36:210395-6746622    19
## 17     ['Pathogenic']    19 NCBI36:3505633-4641977    19
## 18     ['Pathogenic']    19 NCBI36:3959558-4714171    19
## 19     ['Pathogenic']    19 NCBI36:1923244-9620555    19
## 20     ['Pathogenic']    19  NCBI36:184565-4650484    19
## 21     ['Pathogenic']    19 GRCh37:3338022-4833151    19
## 22                 []    19 GRCh37:3338022-4833151    19
##                     assembly2
## 1   GRCh38.p12:260911-4788345
## 2   GRCh37.p13:233565-4699484
## 3   GRCh37.p13:259395-6795622
## 4  GRCh37.p13:4008558-4763171
## 5  GRCh37.p13:3554633-4690977
## 6  GRCh37.p13:1972244-9759555
## 7  GRCh38.p12:3076810-4796770
## 8  GRCh38.p12:3338024-4833139
```

```
## 9   GRCh38.p12:68029-58598923
## 10 GRCh38.p12:260912-58585793
## 11 GRCh38.p12:260912-58445521
## 12 GRCh38.p12:3120162-9622144
## 13 GRCh38.p12:260912-58445521
## 14 GRCh38.p12:260912-58445521
## 15 GRCh38.p12:260912-58445521
## 16  GRCh37.p13:259395-6795622
## 17 GRCh37.p13:3554633-4690977
## 18 GRCh37.p13:4008558-4763171
## 19 GRCh37.p13:1972244-9759555
## 20  GRCh37.p13:233565-4699484
## 21 GRCh38.p12:3338024-4833139
## 22 GRCh38.p12:3338024-4833139
```

```
### dataframe to convert
CNVdbVar <- Gviz_table("results-CNV-dbVAR.csv")
CNVdbVar
```

```
##    CNV_variant_id   chr   start      end variant_region_id study_ID
## 1        49623411 chr19  260911  4788357        nsv4457776  nstd102
## 2        49353191 chr19  233565  4699484        nsv4358278  nstd102
## 3        49353005 chr19  259395  6795622        nsv4358092  nstd102
## 4        49350830 chr19 4008558  4763171        nsv4355917  nstd102
## 5        49349701 chr19 3554633  4690977        nsv4354788  nstd102
## 6        49349293 chr19 1972244  9759555        nsv4354380  nstd102
## 7        49345450 chr19 3076808  4796782        nsv4350537  nstd102
## 8        49344315 chr19 3338022  4833151        nsv4349402  nstd102
## 9        48468240 chr19   68029 59110290        nsv3904885  nstd102
## 10       48466558 chr19  260912 59097160        nsv3903203  nstd102
## 11       48466447 chr19  260912 58956888        nsv3903092  nstd102
## 12       48453939 chr19 3120160  9732820        nsv3890584  nstd102
## 13       45807136 chr19  260912 58956888        nsv2779094   nstd37
## 14       17813982 chr19  260912 58956888                    nstd102
## 15       17813734 chr19  260912 58956888                    nstd101
## 16        3740775 chr19  259395  6795622         nsv533414   nstd37
## 17        3739972 chr19 3554633  4690977         nsv532611  nstd101
## 18        3738955 chr19 4008558  4763171         nsv531594  nstd101
## 19        3738954 chr19 1972244  9759555         nsv531593  nstd101
## 20        3738649 chr19  233565  4699484         nsv531288  nstd101
## 21        1212838 chr19 3338022  4833151         nsv497563   nstd37
## 22        1137112 chr19 3338022  4833151                     nstd37
```

```
### reduce lengths of CNVs, so features/ids can be visualized
CNVdbVar$start[as.numeric(CNVdbVar$start)<4669000] <- 4669000
CNVdbVar$end[as.numeric(CNVdbVar$end)>4724100] <- 4724100

### convert dataframe CNVdbVar to GRanges Object
CNVdbVar_GR <- makeGRangesFromDataFrame(CNVdbVar, keep.extra.columns=TRUE)
CNVdbVar_GR
```

```
## GRanges object with 22 ranges and 3 metadata columns:
##        seqnames          ranges strand | CNV_variant_id variant_region_id
```

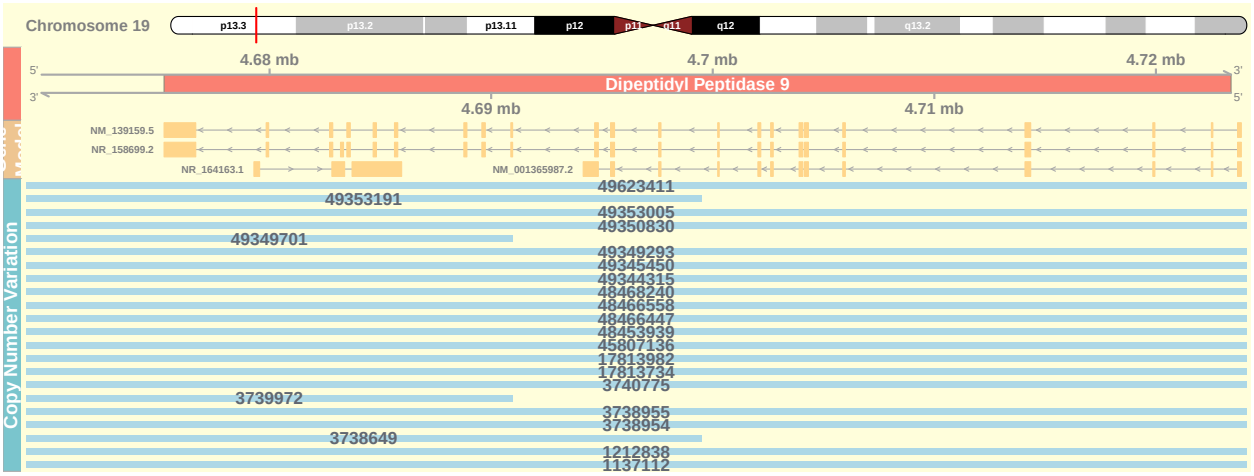
```
##           <Rle>       <IRanges>  <Rle> |      <integer>          <factor>
##     [1]    chr19 4669000-4724100      * |       49623411        nsv4457776
##     [2]    chr19 4669000-4699484      * |       49353191        nsv4358278
##     [3]    chr19 4669000-4724100      * |       49353005        nsv4358092
##     [4]    chr19 4669000-4724100      * |       49350830        nsv4355917
##     [5]    chr19 4669000-4690977      * |       49349701        nsv4354788
##     ...      ...             ...    ... .           ...               ...
##    [18]    chr19 4669000-4724100      * |        3738955         nsv531594
##    [19]    chr19 4669000-4724100      * |        3738954         nsv531593
##    [20]    chr19 4669000-4699484      * |        3738649         nsv531288
##    [21]    chr19 4669000-4724100      * |        1212838         nsv497563
##    [22]    chr19 4669000-4724100      * |        1137112
##         study_ID
##         <factor>
##     [1]  nstd102
##     [2]  nstd102
##     [3]  nstd102
##     [4]  nstd102
##     [5]  nstd102
##     ...      ...
##    [18]  nstd101
##    [19]  nstd101
##    [20]  nstd101
##    [21]   nstd37
##    [22]   nstd37
##    -------
##    seqinfo: 1 sequence from an unspecified genome; no seqlengths
```

Gene Model with CNVs - DPP9

Show the figure in new window. Variant ids are shown.

```
chr1 <- as.character(unique(seqnames(CNVdbVar_GR)))
itrack1 <- IdeogramTrack(genome = "hg19", chromosome = chr1,
                         background.panel = "#FFFEDB",
                         background.title = "#FFFEDB")
GeneModel <- read.csv(file="results-transcripts-UCSC.csv", header = TRUE, sep=";")
grtrack1 <- GeneRegionTrack(GeneModel, genome = "hg19", chromosome = "chr19",
                            name = "Gene Model",
                            transcriptAnnotation = "symbol",
                            background.title = "#EEC591",
                            background.panel = "#FFFEDB",
                            col.border.title = "dark gray",
                            cex.title = 1.1, showId = TRUE)
gtrack1 <- GenomeAxisTrack(background.panel = "#FFFEDB",
                           background.title = "#FA8072",
                           range=IRanges(start = 4675239,
                                         end = 4723855,
                                         names = "Dipeptidyl Peptidase 9"),
                           col.border.title = "dark gray", showId = TRUE)
atrack1 <- AnnotationTrack(CNVdbVar_GR,
                           name = "Copy Number Variation",
                           background.title = "#7AC5CD",
```

```
                          background.panel = "#FFFEDB",
                          col.border.title = "dark gray",
                          cex.title = 1.1,
                          feature = CNVdbVar$CNV_variant_id,
                          showFeatureId = T, cex.feature = 1,
                          fontcolor.feature = "#616771")
plotTracks(list(itrack1,gtrack1,grtrack1,atrack1),
           from = 4669000, to = 4724100,
           col = NULL,
           add53 = TRUE, add35 = TRUE,
           cex = 1.1,
           cex.id = 1.2,
           col.id = "white",
           fill.range = "#FA8072",
           showBandId = TRUE, cex.bands = 0.7,
           fontface = 2, title.width = 0.3, sizes = c(0.3,0.5,0.4,2),
           margin = 0)
```



Documentation to create nice plots

Usage: https://www.bioconductor.org/packages/devel/bioc/vignettes/Gviz/inst/doc/Gviz.html
Color Chart: https://github.com/EarlGlynn/colorchart/wiki/Color-Chart-in-R