

Visualizing_genome

Paulyna Magana

19/05/2022

Libraries

```
library(Gviz)
library(rtracklayer)
library(GenomicRanges)
library(biomaRt)
```

```
session <- browserSession()
genome(session) <- "hg38"
trackNames(session)
```

##	Base Position	Fix Patches
##	"ruler"	"fixSeqLiftOverPsl"
##	Alt Haplotypes	Assembly
##	"altSeqLiftOverPsl"	"gold"
##	Centromeres	Chromosome Band
##	"centromeres"	"cytoBand"
##	Clone Ends	Exome Probesets
##	"cloneEndSuper"	"exomeProbesets"
##	FISH Clones	Gap
##	"fishClones"	"gap"
##	GC Percent	GRC Contigs
##	"gc5BaseBw"	"ctgPos2"
##	GRC Incident	Hg19 Diff
##	"grcIncidentDb"	"hg38ContigDiff"
##	INSDC	LiftOver & ReMap
##	"ucscToINSDC"	"liftHg19"
##	LRG Regions	Mappability
##	"lrg"	"mappability"
##	RefSeq Acc	Restr Enzymes
##	"ucscToRefSeq"	"cutters"
##	Scaffolds	Short Match
##	"scaffolds"	"oligoMatch"
##	STS Markers	GENCODE V41
##	"stsMap"	"knownGene"
##	NCBI RefSeq	All GENCODE
##	"refSeqComposite"	"wgEncodeGencodeSuper"
##	CCDS	CRISPR Targets
##	"ccdsGene"	"crisprAllTargets"
##	HGNC	IKMC Genes Mapped

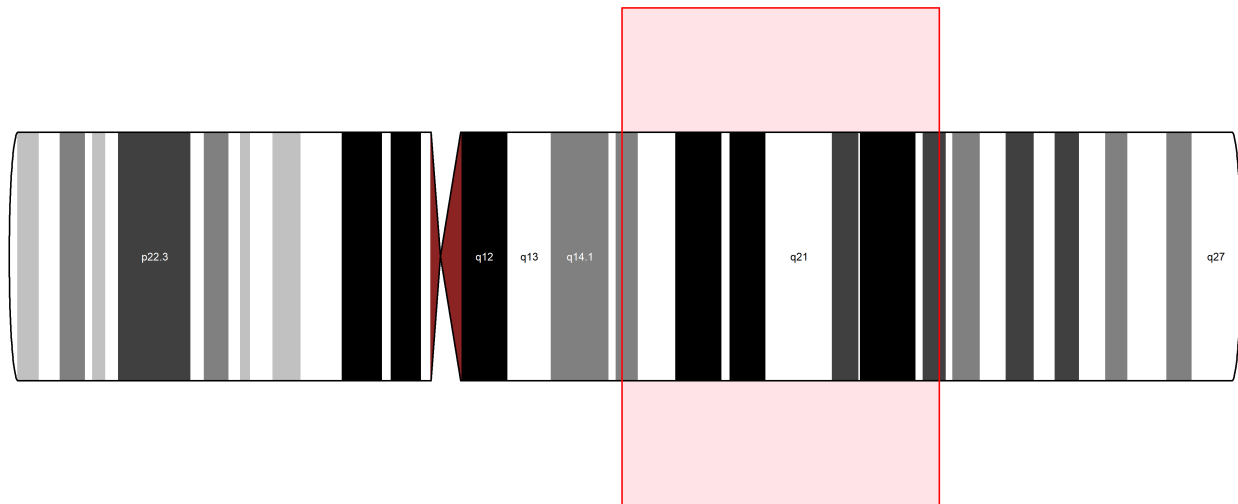
##	"hgnc"	"hgIkmc"
##	LRG Transcripts	MANE v1.0
##	"lrgTranscriptAli"	"mane"
##	MGC Genes	Non-coding RNA
##	"mgcFullMrna"	"nonCodingRNAs"
##	Old UCSC Genes	ORFeome Clones
##	"knownGeneOld12"	"orfeomeMrna"
##	Other RefSeq	Pfam in GENCODE
##	"xenoRefGene"	"ucscGenePfam"
##	Prediction Archive	RetroGenes V9
##	"genePredArchive"	"ucscRetroAli9"
##	TransMap V5	UCSC Alt Events
##	"transMapV5"	"knownAlt"
##	UniProt	OMIM Alleles
##	"uniprot"	"omimAvSnp"
##	CADD	Cancer Gene Expr
##	"caddSuper"	"cancerExpr"
##	ClinGen	ClinGen CNVs
##	"clinGenComp"	"iscaComposite"
##	ClinVar Variants	Constraint scores
##	"clinvar"	"constraintSuper"
##	Coriell CNVs	COSMIC Regions
##	"coriellDelDup"	"cosmicRegions"
##	DECIPHER CNVs	DECIPHER SNVs
##	"decipher"	"decipherSnvs"
##	Development Delay	GenCC
##	"cnvDevDelay"	"genCC"
##	Gene Interactions	GeneReviews
##	"interactions"	"geneReviews"
##	GWAS Catalog	HGMD Variants
##	"gwasCatalog"	"hgmd"
##	LOVD Variants	OMIM Cyto Loci
##	"lovdComp"	"omimLocation"
##	OMIM Genes	Orphanet
##	"omimGene2"	"orphadata"
##	PanelApp	REVEL Scores
##	"panelApp"	"revel"
##	SNPedia	TCGA Pan-Cancer
##	"snpedia"	"gdcCancer"
##	UniProt Variants	Variants in Papers
##	"spMut"	"varsInPubs"
##	Rare Harmful Vars	Blood (PBM) Hao
##	"covidMuts"	"bloodHao"
##	Colon Wang	Cortex Velmeshev
##	"colonWang"	"cortexVelmshv"
##	Fetal Gene Atlas	Heart Cell Atlas
##	"fetalGeneAtlas"	"heartCellAtlas"
##	Ileum Wang	Kidney Stewart
##	"ileumWang"	"kidneyStewart"
##	Liver MacParland	Lung Travaglini
##	"liverMacParland"	"lungTravaglini2020"
##	Merged Cells	Muscle De Micheli
##	"singleCellMerged"	"muscleDeMicheli"
##	Pancreas Baron	Placenta Vento-Tormo

##	"pancreasBaron"	"placentaVentoTormo"
##	Rectum Wang	Skin Sole-Boldo
##	"rectumWang"	"skinSoleBoldo"
##	Tabula Sapiens	Human ESTs
##	"tabulaSapiens"	"est"
##	Human mRNAs	Other ESTs
##	"mrna"	"xenoEst"
##	Other mRNAs	SIB Alt-Splicing
##	"xenoMrna"	"sibTxGraph"
##	Spliced ESTs	GTEx Gene V8
##	"intronEst"	"gtexGeneV8"
##	GTEx RNA-Seq Coverage	Affy Archive
##	"gtexCov"	"affyArchive"
##	EPDnew Promoters	GNF Atlas 2
##	"epdNew"	"gnfAtlas2"
##	GTEx Gene	GTEx Transcript
##	"gtexGene"	"gtexTranscExpr"
##	GWIPS-viz Riboseq	miRNA Tissue Atlas
##	"gwipsvizRiboseq"	"miRnaAtlas"
##	ENCODE cCREs	ENCODE Regulation
##	"encodeCcreCombined"	"wgEncodeReg"
##	CpG Islands	GeneHancer
##	"cpgIslandSuper"	"geneHancer"
##	GTEx cis-eQTLs	Hi-C and Micro-C
##	"gtexEqtlHighConf"	"hicAndMicroC"
##	JASPAR Transcription Factors	ORegAnno
##	"jaspar"	"oreganno"
##	RefSeq Func Elems	ReMap ChIP-seq
##	"refSeqFuncElems"	"ReMap"
##	Conservation	Cactus 241-way
##	"cons100way"	"cons241way"
##	Cons 30 Primates	Primate Chain/Net
##	"cons30way"	"primateChainNet"
##	Placental Chain/Net	Vertebrate Chain/Net
##	"placentalChainNet"	"vertebrateChainNet"
##	CHM13 alignments	Human Chain/Net
##	"chm13LiftOver"	"humanChainNet"
##	dbSNP 155	dbSNP 153
##	"dbSnp155Composite"	"dbSnp153Composite"
##	1000G Archive	Array Probesets
##	"tgpArchive"	"genotypeArrays"
##	dbSNP Archive	dbVar Common Struct Var
##	"dbSnpArchive"	"dbVarSv"
##	DGV Struct Var	Genome In a Bottle
##	"dgvPlus"	"giab"
##	gnomAD Variants	Platinum Genomes
##	"gnomadVariants"	"platinumGenomes"
##	RepeatMasker	Interrupted Rpts
##	"rmsk"	"nestedRepeats"
##	Microsatellite	RepeatMasker Viz.
##	"microsat"	"joinedRmsk"
##	Segmental Dups	Self Chain
##	"genomicSuperDups"	"chainSelf"
##	Simple Repeats	WM + SDust

```
##                                "simpleRepeat"                "windowmaskerSdust"

gtrack <- GenomeAxisTrack() #to add a genomic axis to the plot
ideoTrack <- IdeogramTrack(genome = "hg38", chromosome = "chr6") # ideogram

#plot
# showbandId - chromosome bands in the ideogram
# turn off the explicit plotting of the chromosome name by setting the showId display parameter to FALSE
plotTracks(ideoTrack, from = 85e6, to = 129e6, showId = FALSE,
            showBandId = TRUE, cex.bands = 0.5)
```



SLC22A1

```
###Simple annotation tracks:
fromA1 <- 160115734
toA1 <- 160160664

##Gene model type tracks:

ncbi_ref <- UscsTrack(genome="hg38", chromosome="chr6", track="NCBI RefSeq",
                      from=fromA1, to=toA1, trackType="GeneRegionTrack",
                      rstarts="exonStarts", rends="exonEnds", gene="name",
                      symbol="name", transcript="name", strand="strand",
                      fill="#767676", stacking="squish", name="NCBI RefSeq",
                      showId=TRUE, geneSymbol=TRUE)

biomTrack <- BiomartGeneRegionTrack(genome="hg38", chromosome="chr6",
```

```

start=fromA1, end=toA1, name="Ensembl")

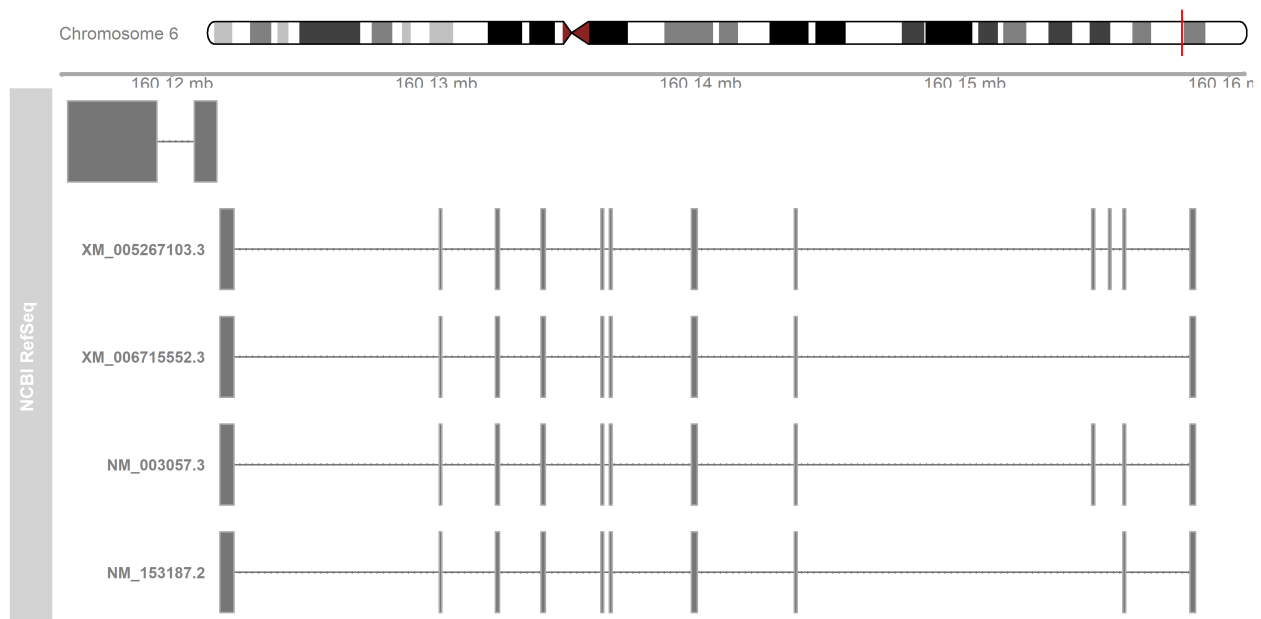
###Data tracks

###Other tracks and plotting:

axTrack <- GenomeAxisTrack(labelPos="below", width=6)
idxTrack <- IdeogramTrack(genome="hg38", chromosome="chr6")

##PLOTING
plotTracks(list(idxTrack, axTrack, ncbi_ref), from=fromA1, toA1, showTitle=TRUE, col.line=NULL,
collapseTranscripts="longest", transcriptAnnotation = "symbol")

```



highlight CRISPR-Cas9 region

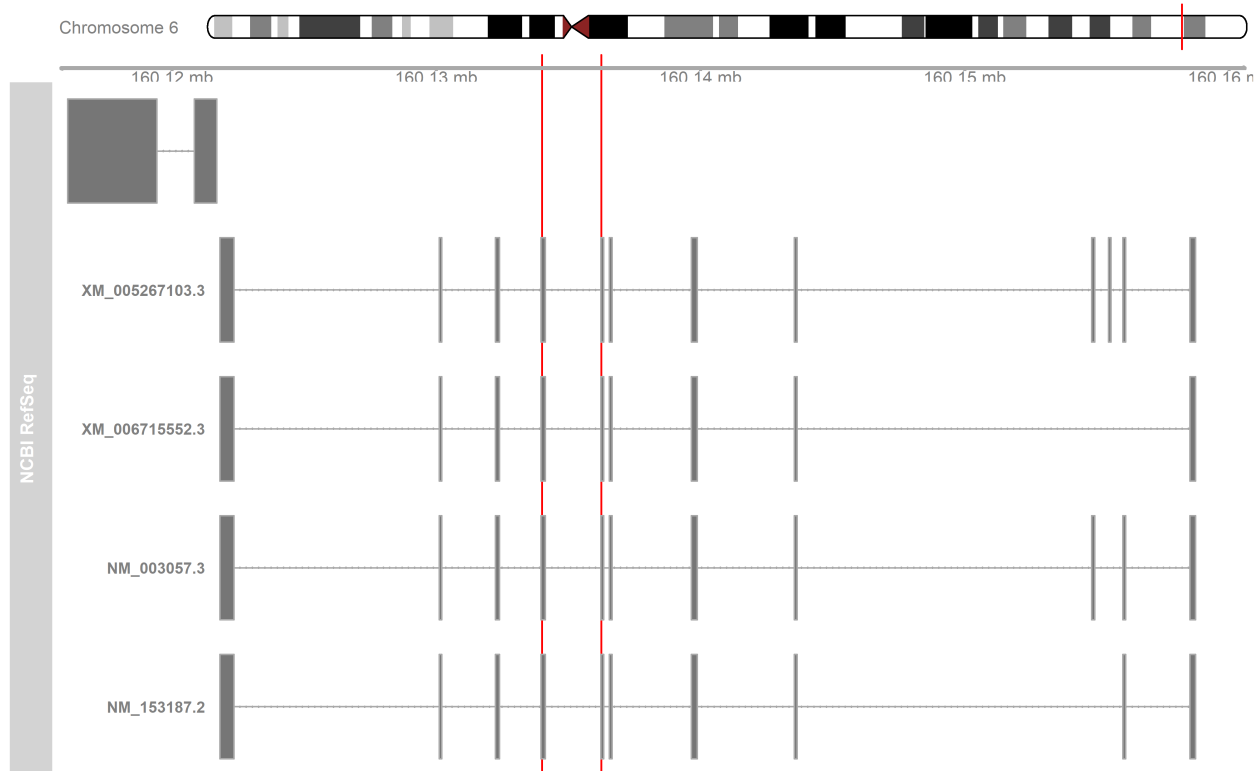
```

ht <- HighlightTrack(trackList = list(axTrack, ncbi_ref),
start = c(160136234, 160133991), width = 10,
chromosome = 6)

lim <- c(fromA1, toA1)

plotTracks(list(idxTrack, ht), from = lim[1], to = lim[2])

```



luciferase regions

```
library(tidyverse)
```

```
luciferase_plasmids <- read_csv("./luciferase_plasmids.csv")
```

```
## Rows: 12 Columns: 5
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr (2): gene, chromosome
```

```
## dbl (3): start, end, width
```

```
##
```

```
## i Use 'spec()' to retrieve the full column specification for this data.
```

```
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
as.data.frame(luciferase_plasmids)
```

```
##      gene      start      end chromosome width
## 1  pGL3-A1_PF1 160119315 160121872      chr6   2558
## 2  pGL3- A1_PF2 160121691 160121872      chr6    182
## 3  pGL3- A1_PF3 160121762 160121872      chr6    111
## 4  pGL3- A1_PF4 160121795 160121872      chr6     78
## 5  pGL3- A1_INT1 160122625 160123645      chr6   1021
## 6  pGL3- A1_INT2 160123849 160125370      chr6   1522
## 7  pGL3- A1_INT3 160124767 160125370      chr6    604
```

```
## 8    pGL3-A4_ENH 132293804 132293985      chr5    182
## 9    pGL3-A4_PF1 132294073 132294419      chr5    347
## 10   pGL3-A4_PF2 132294344 132294419      chr5     76
## 11   pGL3-A5_PF3 132369253 132369470      chr5    218
## 12   pGL3-A5_INT 132378284 132379321      chr5   1038
```

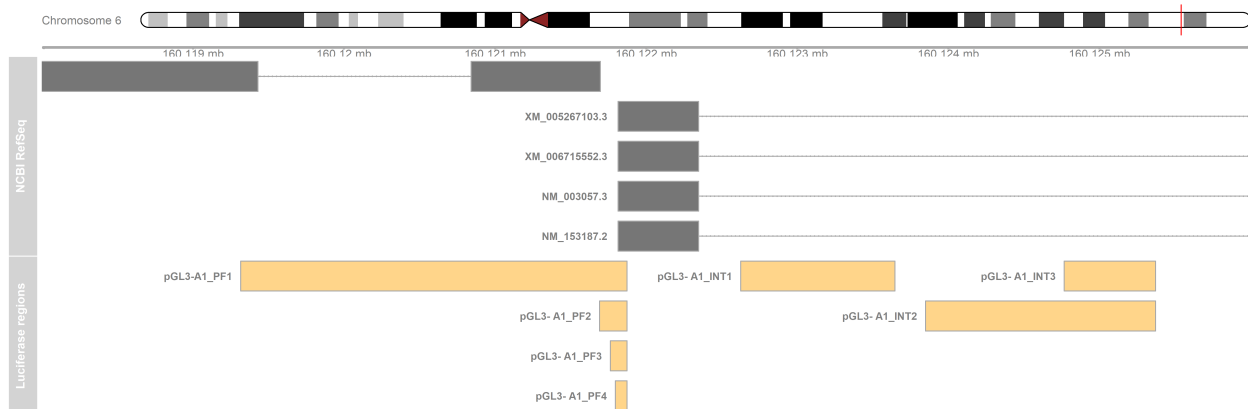
```
luciferase_plasmids$start <- as.integer(luciferase_plasmids$start)
luciferase_plasmids$end <- as.integer(luciferase_plasmids$end)
luciferase_plasmids$width <- as.integer(luciferase_plasmids$width)
luciferase_plasmids$chromosome <- as.factor(luciferase_plasmids$chromosome)
luciferase_plasmids$gene <- as.factor(luciferase_plasmids$gene)
```

```
head(luciferase_plasmids)
```

```
## # A tibble: 6 x 5
##   gene          start      end chromosome width
##   <fct>        <int>    <int>    <fct>    <int>
## 1 pGL3-A1_PF1  160119315 160121872 chr6      2558
## 2 pGL3- A1_PF2  160121691 160121872 chr6       182
## 3 pGL3- A1_PF3  160121762 160121872 chr6       111
## 4 pGL3- A1_PF4  160121795 160121872 chr6        78
## 5 pGL3- A1_INT1 160122625 160123645 chr6     1021
## 6 pGL3- A1_INT2 160123849 160125370 chr6     1522
```

```
grtrack <- GeneRegionTrack(luciferase_plasmids, genome = "hg38",
                           chromosome = "chr6",
                           name = "Luciferase regions")
```

```
plotTracks(list(idxTrack, axTrack, ncbi_ref, grtrack), from = 160118000, to = 160126000, transcriptAnnotat
```



SLC22A5

```
###Simple annotation tracks:
fromA5 <- 132366800
toA5 <- 132396017
```

```
##Gene model type tracks:
```

```
ncbi_ref <- UscsTrack(genome="hg38", chromosome="chr5", track="NCBI RefSeq",
                      from=fromA5, to=toA5, trackType="GeneRegionTrack",
                      rstarts="exonStarts", rends="exonEnds", gene="name",
                      symbol="name", transcript="name", strand="strand",
                      fill="#767676", stacking="squish", name="NCBI RefSeq",
                      showId=TRUE, geneSymbol=TRUE)
```

```
biomTrack <- BiomartGeneRegionTrack(genome="hg38", chromosome="chr5",
                                     start=fromA5, end=toA5, name="Ensembl")
```

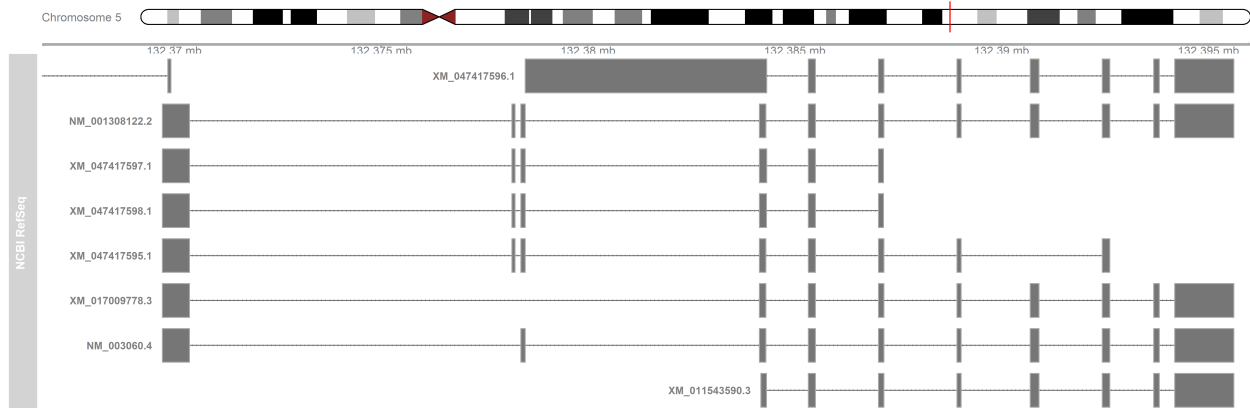
```
###Data tracks
```

```
###Other tracks and plotting:
```

```
axTrack <- GenomeAxisTrack(labelPos="below")
idxTrack <- IdeogramTrack(genome="hg38", chromosome="chr5")
```

```
##PLOTING
```

```
plotTracks(list(idxTrack, axTrack, ncbi_ref), from=fromA5, to=toA5, showTitle=TRUE, col.line=NULL,
            collapseTranscripts="longest", transcriptAnnotation = "symbol")
```



highlight CRISPR-Cas9 region

```
ht <- HighlightTrack(trackList = list(axTrack, ncbi_ref),
                     start = c(132384160, 132378383), width = 10,
                     chromosome = 5)
```

```
lim <- c(fromA5, toA5)
```

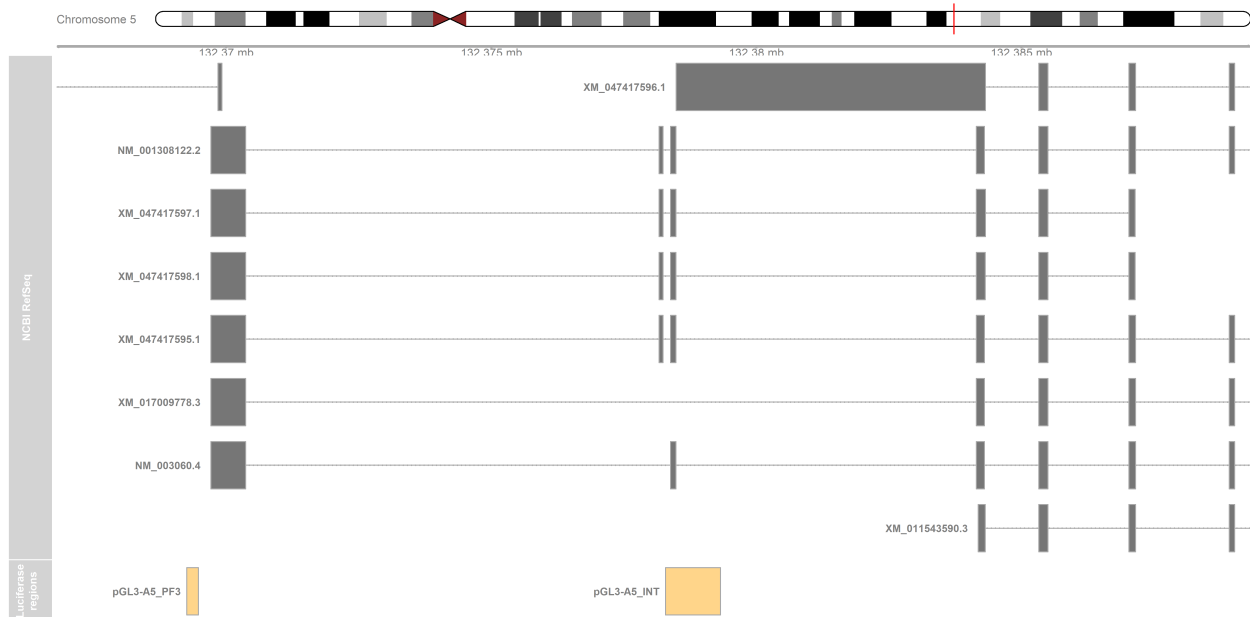
```
plotTracks(list(idxTrack, ht), from = lim[1], to = lim[2])
```




luciferase regions

```
grtrack <- GeneRegionTrack(luciferase_plasmids, genome = "hg38",
                           chromosome = "chr5",
                           name = "Luciferase regions")
```

```
plotTracks(list(idxTrack, axTrack, ncbi_ref, grtrack), from = fromA5, to = 132389321, transcriptAnnotation = FALSE)
```



SLC22A4

```

###Simple annotation tracks:
fromA4 <- 132290928
toA4 <- 132345343

##Gene model type tracks:

ncbi_ref <- UscsTrack(genome="hg38", chromosome="chr5", track="NCBI RefSeq",
                      from=fromA4, to=toA4, trackType="GeneRegionTrack",
                      rstarts="exonStarts", rends="exonEnds", gene="name",
                      symbol="name", transcript="name", strand="strand",
                      fill="#767676", stacking="squish", name="NCBI RefSeq",
                      showId=TRUE, geneSymbol=TRUE)

biomTrack <- BiomartGeneRegionTrack(genome="hg38", chromosome="chr5",
                                     start=fromA4, end=toA4, name="Ensembl")

###Data tracks

##Other tracks and plotting:

axTrack <- GenomeAxisTrack(labelPos="below")
idxTrack <- IdeogramTrack(genome="hg38", chromosome="chr5")

##PLOTTING
plotTracks(list(idxTrack, axTrack, ncbi_ref), from=fromA4, to=toA4, showTitle=TRUE, col.line=NULL,
           collapseTranscripts="longest", transcriptAnnotation = "symbol")

```



highlight CRISPR-Cas9 region

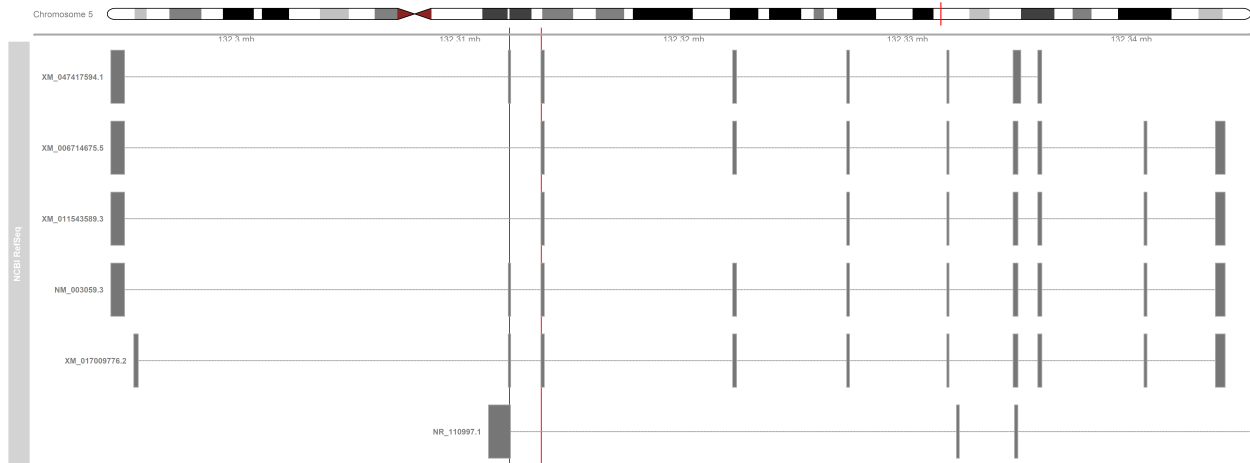
```

ht <- HighlightTrack(trackList = list(axTrack, ncbi_ref),
                     start = c(132313626, 132312204), width = 7,
                     chromosome = 5)

lim <- c(fromA4, toA4)

```

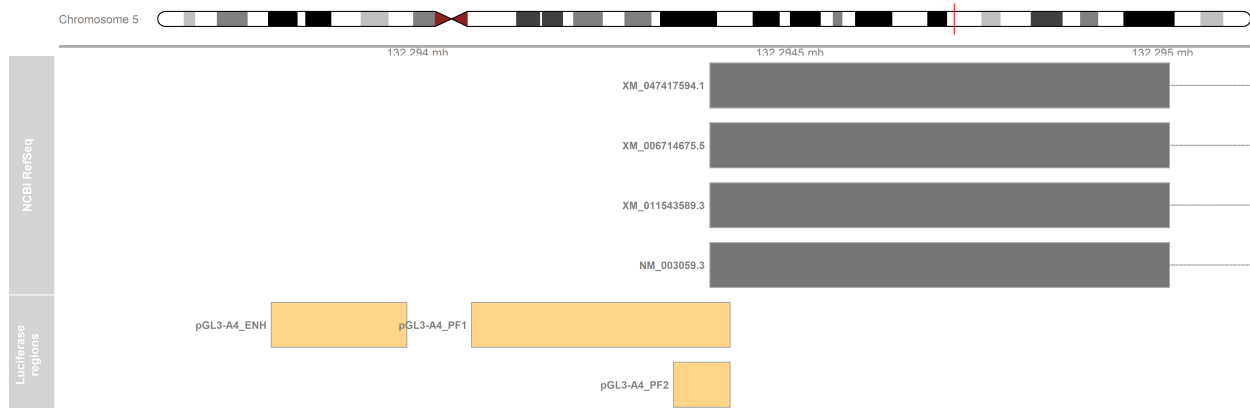
```
plotTracks(list(idxTrack, ht), from = lim[1], to = lim[2])
```



luciferase regions

```
grtrack <- GeneRegionTrack(luciferase_plasmids, genome = "hg38",
                           chromosome = "chr5",
                           name = "Luciferase regions")
```

```
plotTracks(list(idxTrack, axTrack, ncbi_ref, grtrack), from = 132293519, to = 132295119, transcriptAnnotat
```



Session Info

```
## R version 4.1.1 (2021-08-10)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 18363)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United Kingdom.1252
```

```

## [2] LC_CTYPE=English_United Kingdom.1252
## [3] LC_MONETARY=English_United Kingdom.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.1252
##
## attached base packages:
## [1] grid      stats4      stats      graphics  grDevices  utils      datasets
## [8] methods  base
##
## other attached packages:
## [1] forcats_0.5.2      stringr_1.4.1      dplyr_1.0.10
## [4] purrr_0.3.5        readr_2.1.3        tidyr_1.2.1
## [7] tibble_3.1.8        ggplot2_3.3.6      tidyverse_1.3.2
## [10] biomaRt_2.50.3      rtracklayer_1.54.0 Gviz_1.36.2
## [13] GenomicRanges_1.46.1 GenomeInfoDb_1.30.1 IRanges_2.28.0
## [16] S4Vectors_0.32.4    BiocGenerics_0.40.0
##
## loaded via a namespace (and not attached):
## [1] googledrive_2.0.0      colorspace_2.0-3
## [3] rjson_0.2.21           deldir_1.0-6
## [5] ellipsis_0.3.2         biovizBase_1.42.0
## [7] htmlTable_2.4.1        XVector_0.34.0
## [9] fs_1.5.2               base64enc_0.1-3
## [11] dichromat_2.0-0.1      rstudioapi_0.14
## [13] bit64_4.0.5            lubridate_1.8.0
## [15] AnnotationDbi_1.56.2   fansi_1.0.3
## [17] xml2_1.3.3             codetools_0.2-18
## [19] splines_4.1.1          cachem_1.0.6
## [21] knitr_1.40             jsonlite_1.8.2
## [23] Formula_1.2-4          Rsamtools_2.10.0
## [25] broom_1.0.1            cluster_2.1.2
## [27] dbplyr_2.2.1           png_0.1-7
## [29] compiler_4.1.1         http_1.4.4
## [31] backports_1.4.1        assertthat_0.2.1
## [33] Matrix_1.5-1           fastmap_1.1.0
## [35] lazyeval_0.2.2         gargle_1.2.1
## [37] cli_3.4.1              htmltools_0.5.3
## [39] prettyunits_1.1.1      tools_4.1.1
## [41] gtable_0.3.1           glue_1.6.2
## [43] GenomeInfoDbData_1.2.7 rappdirs_0.3.3
## [45] Rcpp_1.0.9             Biobase_2.54.0
## [47] cellranger_1.1.0       vctrs_0.4.2
## [49] Biostrings_2.62.0      xfun_0.33
## [51] rvest_1.0.3            lifecycle_1.0.3
## [53] restfulr_0.0.15        ensemblDb_2.18.4
## [55] googlesheets4_1.0.1    XML_3.99-0.10
## [57] zlibbioc_1.40.0        scales_1.2.1
## [59] vroom_1.6.0            BSgenome_1.62.0
## [61] VariantAnnotation_1.40.0 hms_1.1.2
## [63] MatrixGenerics_1.6.0   ProtGenerics_1.26.0
## [65] parallel_4.1.1         SummarizedExperiment_1.24.0
## [67] AnnotationFilter_1.18.0 RColorBrewer_1.1-3
## [69] yaml_2.3.5             curl_4.3.3
## [71] memoise_2.0.1          gridExtra_2.3

```

## [73] rpart_4.1-15	latticeExtra_0.6-30
## [75] stringi_1.7.6	RSQLite_2.2.17
## [77] highr_0.9	BiocIO_1.4.0
## [79] checkmate_2.1.0	GenomicFeatures_1.46.5
## [81] filelock_1.0.2	BiocParallel_1.28.3
## [83] rlang_1.0.6	pkgconfig_2.0.3
## [85] matrixStats_0.62.0	bitops_1.0-7
## [87] evaluate_0.17	lattice_0.20-44
## [89] GenomicAlignments_1.30.0	htmlwidgets_1.5.4
## [91] bit_4.0.4	tidyselect_1.2.0
## [93] magrittr_2.0.3	R6_2.5.1
## [95] generics_0.1.3	Hmisc_4.7-1
## [97] DelayedArray_0.20.0	DBI_1.1.3
## [99] withr_2.5.0	haven_2.5.1
## [101] pillar_1.8.1	foreign_0.8-81
## [103] survival_3.2-11	KEGGREST_1.34.0
## [105] RCurl_1.98-1.8	nnet_7.3-16
## [107] modelr_0.1.9	crayon_1.5.2
## [109] interp_1.1-3	utf8_1.2.2
## [111] BiocFileCache_2.2.1	tzdb_0.3.0
## [113] rmarkdown_2.17	jpeg_0.1-9
## [115] progress_1.2.2	readxl_1.4.1
## [117] data.table_1.14.2	blob_1.2.3
## [119] reprex_2.0.2	digest_0.6.29
## [121] munsell_0.5.0	