

Today is November 28, 2012.

```
> rm(list = ls())
> library("vcf2R")
> library("GenomicRanges")
> library("Gviz")
```

1 Targeted Loci

```
> file <- system.file("extdata", "snvs.fixed_header.frq",
  package = "vcf2R")
> pos.df <- frq2pos(file)
> head(pos.df)

  chr      pos
1   1 18771811
2   1 18771817
3   1 18771849
4   1 18771875
5   1 18771893
6   1 18771920

> with(pos.df, table(chr))

chr
 1  10  14  17  20   4   8   9
44713 29857 1559 29486 18740 2444 27639 20751

> pos.gr <- with(pos.df, GRanges(seqnames = paste("chr",
  chr, sep = ""), ranges = IRanges(start = pos, end = pos),
  strand = "+"))
> seqlengths(pos.gr) <- seqlengths(Hsapiens)[levels(seqnames(pos.gr))]
> start(pos.gr) <- start(pos.gr) - 10000
> end(pos.gr) <- end(pos.gr) + 10000
> loci <- reduce(pos.gr)
> start(loci) <- start(loci) + 10000
> end(loci) <- end(loci) - 10000
```

```
> val <- data.frame(n.snps = countOverlaps(loci, pos.gr),
  kb = width(loci)/1000)
> values(loci) <- val
> names(loci) <- c("PAX7", "ABCA4", "IRF6", "VAX1", "FGFR2",
  "BMP4", "NTN1", "NOG", "MAFB", "MSX1", "8q24", "PTCH1",
  "FOXE1")
> loci
```

GRanges with 13 ranges and 2 metadata columns:

	seqnames	ranges	strand	n.snps	kb
	<Rle>	<IRanges>	<Rle>	<integer>	<numeric>
PAX7	chr1	[18771811, 19208536]	+	13309	436.726
ABCA4	chr1	[94324232, 95013570]	+	16310	689.339
IRF6	chr1	[209836758, 210468863]	+	15094	632.106
VAX1	chr10	[118421219, 119167920]	+	18546	746.702
FGFR2	chr10	[123095899, 123499252]	+	11311	403.354
BMP4	chr14	[54382326, 54445461]	+	1559	63.136
NTN1	chr17	[8754615, 9266552]	+	15212	511.938
NOG	chr17	[54402391, 54957844]	+	14274	555.454
MAFB	chr20	[38902239, 39614945]	+	18740	712.707
MSX1	chr4	[4824627, 4901842]	+	2444	77.216
8q24	chr8	[129295457, 130355340]	+	27639	1059.884
PTCH1	chr9	[98133217, 98413595]	+	7795	280.379
FOXE1	chr9	[100357256, 100877263]	+	12956	520.008

seqlengths:

chr1	chr10	chr14	...	chr4	chr8	chr9
249250621	135534747	107349540	...	191154276	146364022	141213431

```
> sum(width(loci))/1e+06
```

```
[1] 6.688949
```

```
> with(values(loci), sum(n.snps))
```

```
[1] 175189
```

```
> data("targets")
```

```
> targets.gr
```

GRanges with 13 ranges and 0 metadata columns:

	seqnames	ranges	strand
	<Rle>	<IRanges>	<Rle>
IRF6	chr1	[207903822, 208535029]	+
MAFB	chr20	[38336060, 39047927]	+
ABCA4	chr1	[94097248, 94785697]	+
8q24	chr8	[129365078, 130424128]	+
FOXE1	chr9	[99397513, 99916662]	+
PAX7	chr1	[18644887, 19080641]	+
VAX1	chr10	[118411615, 119157414]	+
NTN1	chr17	[8695839, 9206785]	+
NOG	chr17	[51757836, 52312389]	+

```

MSX1      chr4 [ 4876027,  4952286]      +
BMP4      chr14 [ 53452440, 53514803]      +
FGFR2     chr10 [123086364, 123488761]      +
PTCH1     chr9  [ 97173468,  97452983]      +
---
seqlengths:
chr1 chr10 chr14 chr17 chr20 chr4 chr8 chr9
NA    NA    NA    NA    NA    NA    NA    NA

```

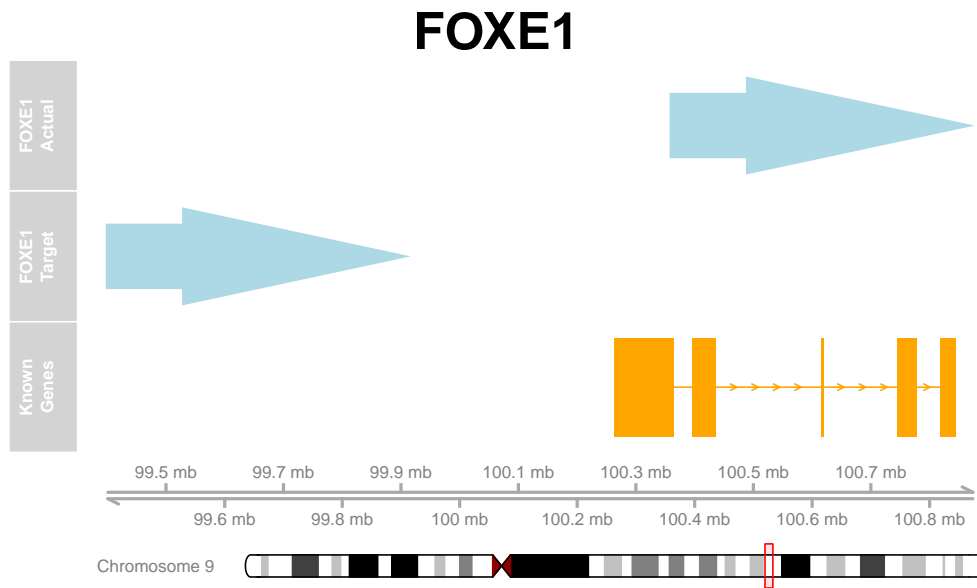
2 Regions

```

> known.gene.file <- system.file("extdata", "known-genes",
  package = "vcf2R")
> kg.df <- data.frame(scan(file = known.gene.file, sep = "\t",
  what = list("", 1L, 1L, "", "", "", 1L, 1L, 1L, 1L,
  "", "")))
> names(kg.df) <- c("chr", "start", "end", "name", "foo",
  "strand", "cds.start", "cds.end", "foo2", "num.exons",
  "exon.coods.start", "exon.coods.end")
> kg.gr <- GRanges(seqnames = kg.df$chr, ranges = IRanges(start = kg.df$start,
  end = kg.df$end), strand = kg.df$strand)

> ht <- 4
> wd <- ht * 1.618
> for (locus in names(loci)) {
  gr <- targets.gr[names(targets.gr) == locus, ]
  chr <- as.character(unique(seqnames(gr)))
  seqlengths(gr) <- seqlengths(Hsapiens)[levels(seqnames(gr))]
  atrack <- AnnotationTrack(gr, name = paste0(locus,
    " Target"), chromosome = chr, genome = "hg19",
    stacking = "squish")
  atrack2 <- AnnotationTrack(loci[names(loci) == locus,
    ], chromosome = chr, genome = "hg19", name = paste0(locus,
    " Actual"), stacking = "squish")
  gtrack <- GenomeAxisTrack()
  itrack <- IdeogramTrack(genome = "hg19", chromosome = chr)
  kg.sub.df <- as.data.frame(kg.gr[subjectHits(findOverlaps(loci[names(loci) ==
    locus, ], kg.gr)), ])
  if (nrow(kg.sub.df) != 0) {
    kg.sub.df <- kg.sub.df[, -c(1, 4, 5)]
    grtrack <- GeneRegionTrack(kg.sub.df, genome = "hg19",
      chromosome = as.character(seqnames(loci)[which(names(loci) ==
        locus)]), name = "Known Genes")
    pdf(file = paste0("./figures/loci/", locus, ".pdf"),
      height = ht, width = wd)
    plotTracks(list(atrack2, atrack, grtrack, gtrack,
      itrack), add = FALSE, main = locus)
    dev.off()
  }
  else {
    pdf(file = paste0("./figures/loci/", locus, ".pdf"),

```



```

    height = ht, width = wd)
  plotTracks(list(atrack2, atrack, gtrack, itrack),
    add = FALSE, main = locus)
  dev.off()
}
}

> kg.sub.df

      start      end
1 100263461 100364025
2 100263920 100364025
3 100395704 100436029
4 100615536 100618997
5 100745488 100778224
6 100840474 100844448
7 100818958 100845365

```

3 Hardy-Weinberg Equilibrium

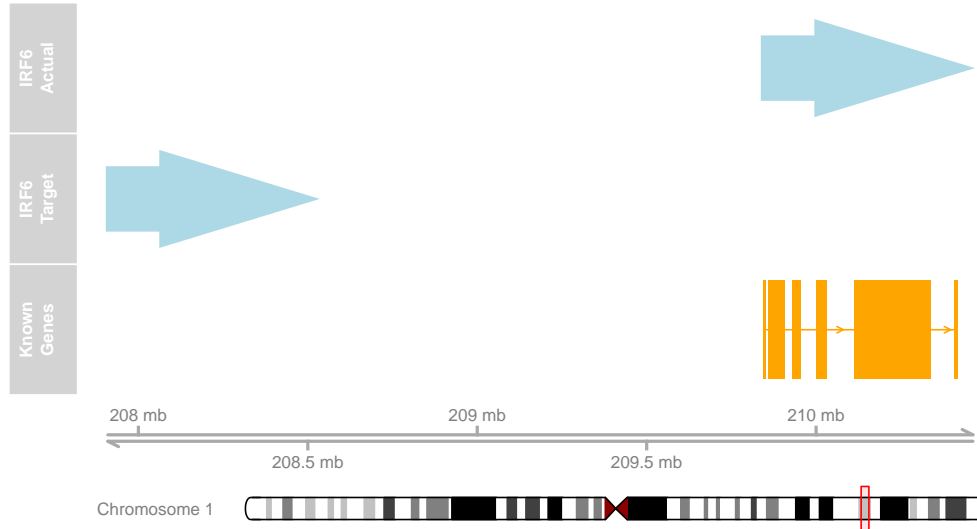
```

> file <- system.file("extdata", "snvs.fixed_header.hwe",
  package = "vcf2R")
> hwe.df <- hwe2R(file = file)
> head(hwe.df)

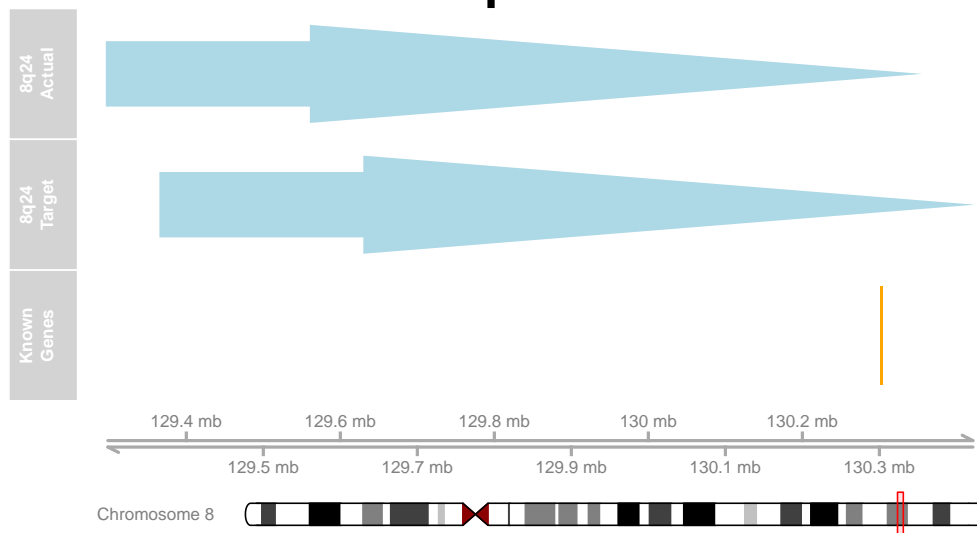
  chr    pos   chi.sq p.value homo.major het homo.minor
1   1 18772169 0.000056 1.000000      4494   1          0
2   1 18772190 0.000223 1.000000      4493   2          0
3   1 18772191 15.727310 0.053862      4464  30          1

```

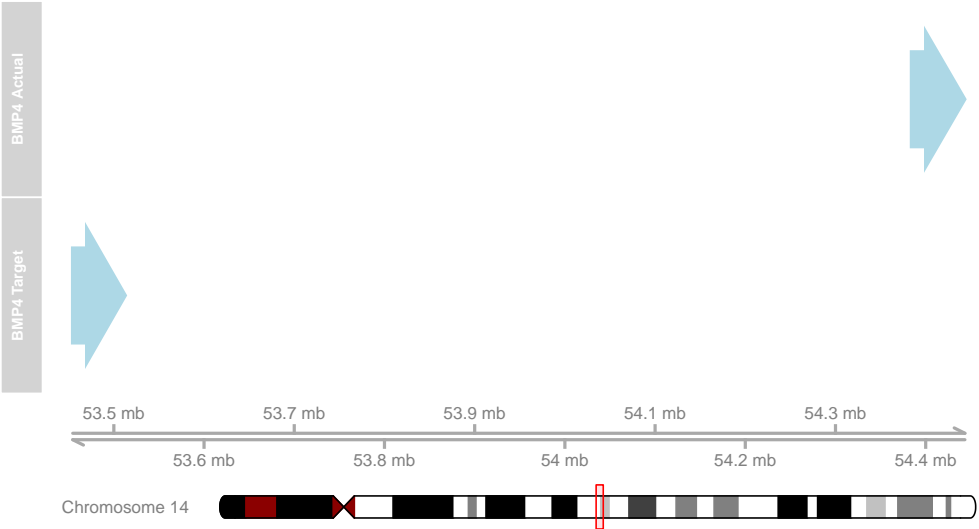
IRF6



8q24



BMP4



4	1	18772203	0.000223	1.000000	4493	2	0
5	1	18772218	0.000223	1.000000	4493	2	0
6	1	18772236	0.016134	1.000000	4478	17	0