Step 1: Install package

```
if (!require(devtools)){
    install.packages('devtools')
}
```

```
## Loading required package: devtools
```

```
## Loading required package: usethis
```

```
devtools::install_github('progenetix/pgxRpi')
```

```
## WARNING: Rtools is required to build R packages, but is not currently installed.
##
## Please download and install Rtools 4.0 from https://cran.r-project.org/bin/windows/Rtools/.
```

```
## Skipping install of 'pgxRpi' from a github remote, the SHA1 (ad1cf8a1) has not changed since last ins
##   Use `force = TRUE` to force installation
```

```
library(pgxRpi)
```

Step2: Search esophageal adenocarcinoma NCIt code

```
# query: esophageal adenocarcinoma, type of data: NCIt code, portal: progenetix.org - https://progeneti

# => NCIT:C4025
```

Step3: Access the CNV frequency data from samples with esophageal adenocarcinoma

```
freq <- pgxLoader(type='frequency', output='pgxseg',filters='NCIT:C4025',
                  codematches=T)
```

```
##
##  accessing IntervalFrequencies service from Progenetix
```

The retreived data is an object contaning two slots `meta` and `data`.

The `meta` slot looks like this (contains metadata, like sample count):

```
freq$meta
```

```
##         code                     label sample_count
## 1 NCIT:C4025 Esophageal Adenocarcinoma          862
```

The `data` slot includes two matrices.

```
names(freq$data)
```

```
## [1] "NCIT:C4025" "total"
```

In this case, these two matrices are the same. The columns gain_frequency and loss_frequency are of primary interest.

```
head(freq$data$`NCIT:C4025`)
```

```
##      filters reference_name   start     end gain_frequency loss_frequency index
## 1 NCIT:C4025              1       0 1000000         13.805          8.121     0
## 2 NCIT:C4025              1 1000000 2000000         10.093          5.800     1
```

```
## 3 NCIT:C4025              1 2000000 3000000          11.369           7.773     2
## 4 NCIT:C4025              1 3000000 4000000          12.645           9.861     3
## 5 NCIT:C4025              1 4000000 5000000          11.485          10.325     4
## 6 NCIT:C4025              1 5000000 6000000          11.485           9.977     5
```

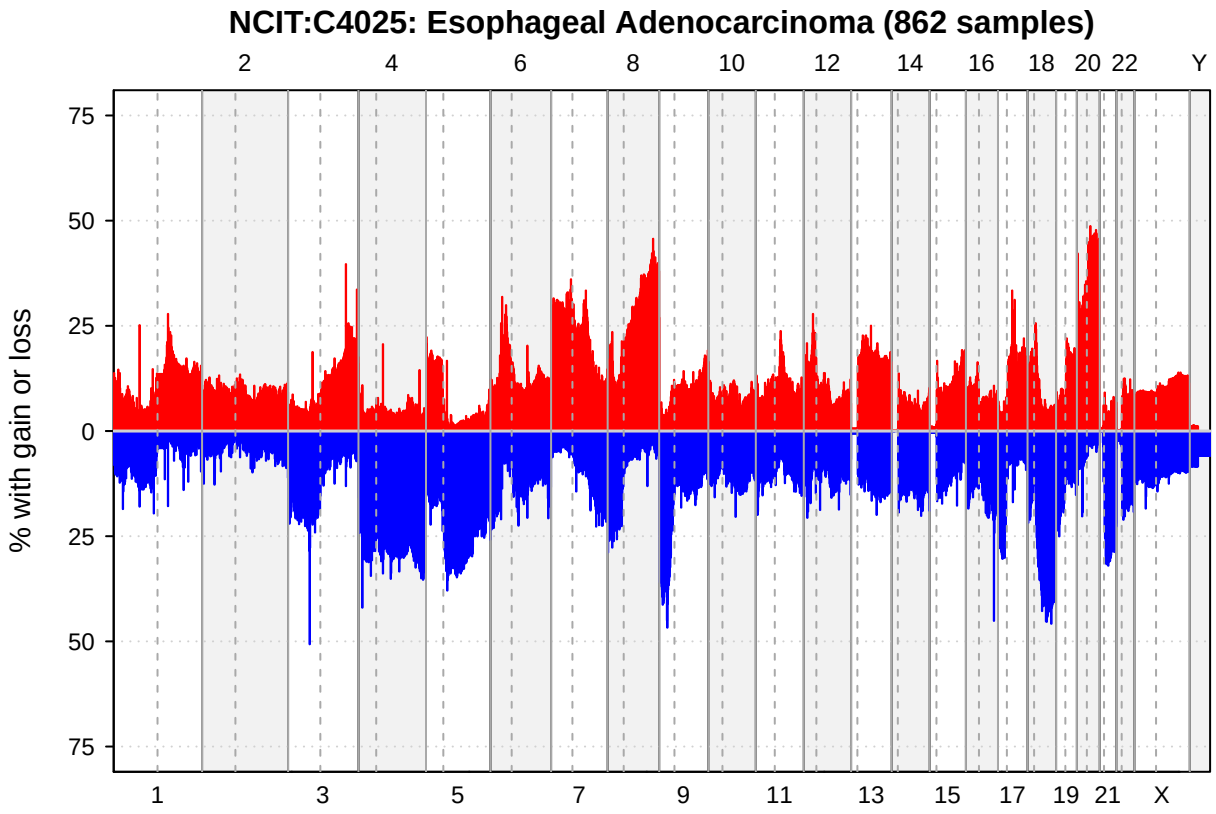
```
dim(freq$data$`NCIT:C4025`)
```

```
## [1] 3102    7
```

Step4: Visualize data

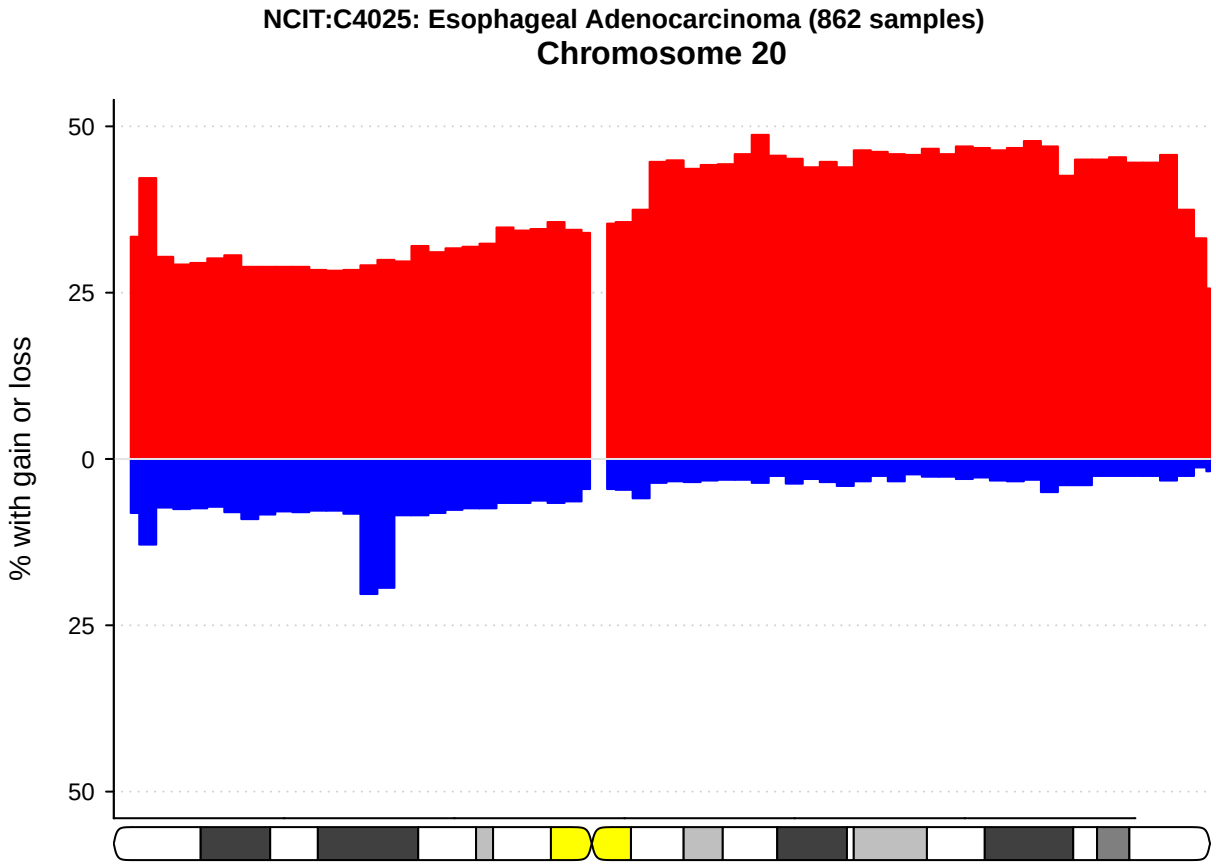
By genome

```
pgxFreqplot(freq)
```



By chromosome

```
pgxFreqplot(freq,chrom = 20)
```



Step5: Analyse the data

According the plot, we can see frequency gains on chromosome 7p, 8q, 20p, 20q and frequency losses on chromosome 4p, 4q, 5q, 9p, 17p, 18q, 21q.

There is a literature where the findings are consistent with the majority of mine. Here is the paper-link.

A more detailed use case see this link.