BIO392-cnv-freq

Hangjia Zhao

9/28/2022

Step 1: Install package

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

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

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

```
if (!require(pgxRpi)){
    devtools::install_github('progenetix/pgxRpi')
}
```

```
## Loading required package: pgxRpi
```

```
library(pgxRpi)
```

Step2: Search esophageal adenocarcinoma NCIt code

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

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

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

The `meta` slot looks like this:

```
freq$meta
```

```
##         code                       label sample_count
## 1 NCIT:C4025 Esophageal Adenocarcinoma          865
## 2      total                                    865
```

The `data` slot includes two matrices.

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

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

In this case, these two matrices are the same.

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

```
##      filters reference_name   start     end gain_frequency loss_frequency no
## 1 NCIT:C4025              1       0  400000          8.902          6.127  1
## 2 NCIT:C4025              1  400000 1400000         13.642          5.665  2
## 3 NCIT:C4025              1 1400000 2400000          9.827          6.243  3
## 4 NCIT:C4025              1 2400000 3400000         13.179         10.983  4
## 5 NCIT:C4025              1 3400000 4400000         12.717         10.058  5
## 6 NCIT:C4025              1 4400000 5400000         10.867         10.636  6
```

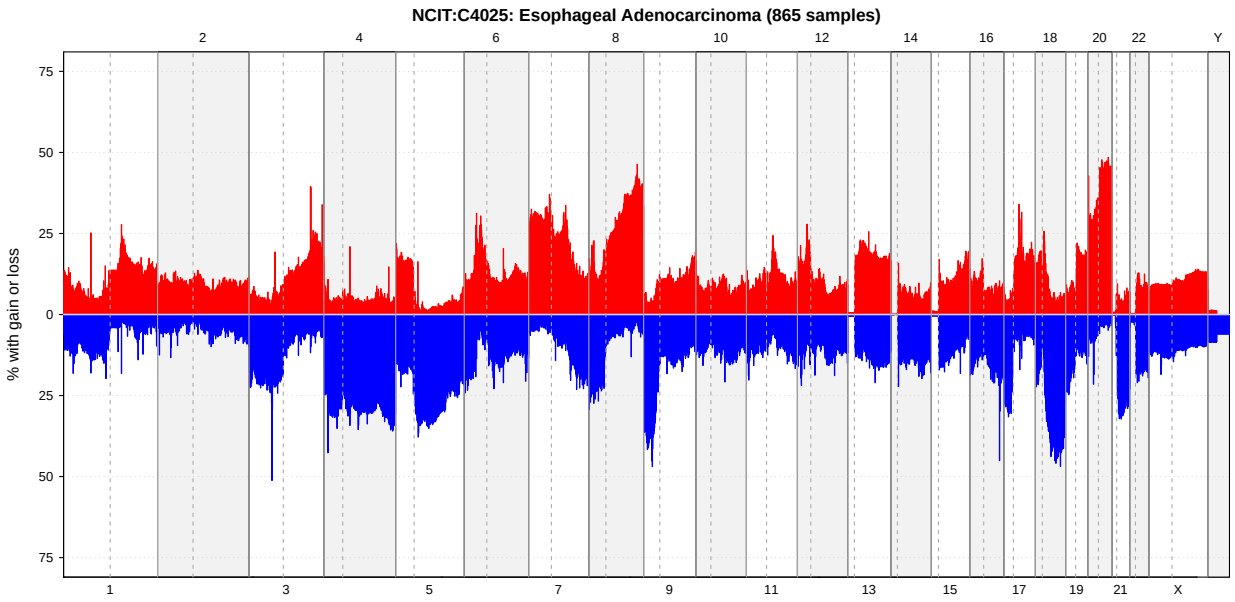
```
dim(freq$data$`NCIT:C4025`)
```

```
## [1] 3106    7
```

Step4: Visualize data

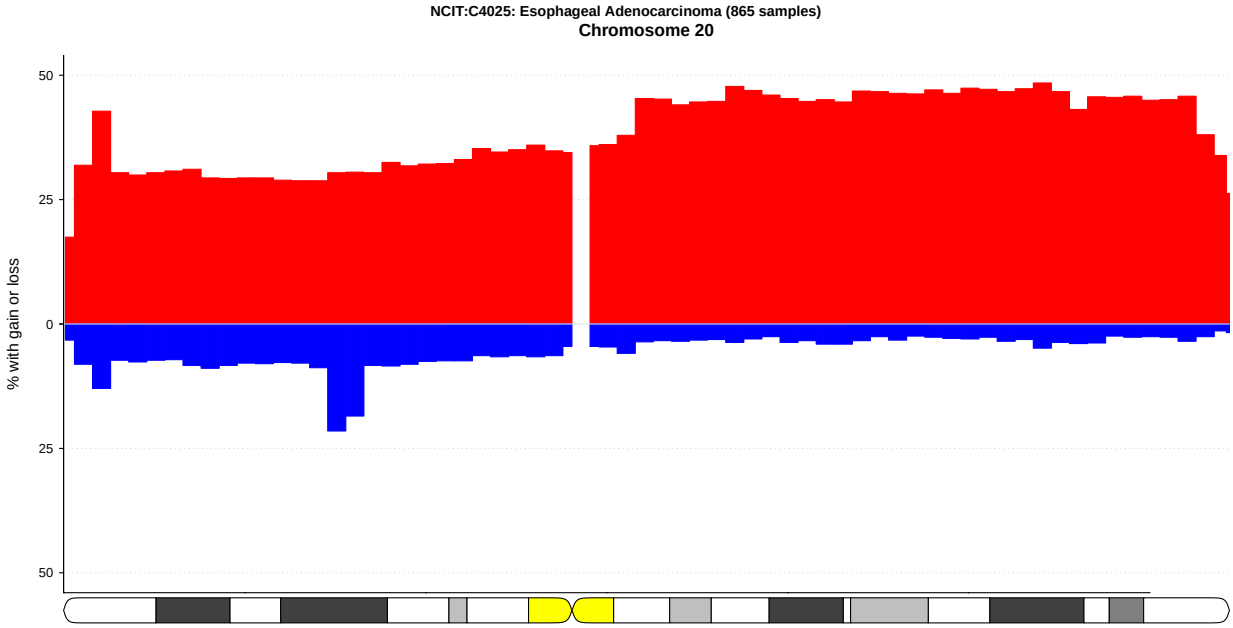
By genome

```
pgxFreqplot(freq)
```



By chromosome

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



Step5: Analyse the data

According the plot, we can see frequenct gains on chromosome 7p, 8q, 20p,20q and frequenct 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.