CNV of human Glioblastoma

Daniel Walther

10/01/2021

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 Glioblastoma NCIt code (see pdf on course page)

```
ncit_code <- "NCIT:C3058"
```

Step3: Access the CNV frequency data from samples with Glioblastoma

```
freq <- pgxLoader(type='frequency', output='pgxseg',filters=ncit_code,
                  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:C3058 Glioblastoma         4305
```

The `data` slot includes two matrices.

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

```
## [1] "NCIT:C3058" "total"
```

The columns gain_frequency and loss_frequency are of primary interest.

```
head(freq$data$`NCIT:C3058`)
```

```
##      filters reference_name   start     end gain_frequency loss_frequency index
## 1 NCIT:C3058              1       0 1000000          10.337           6.527     0
## 2 NCIT:C3058              1 1000000 2000000          11.638           6.620     1
## 3 NCIT:C3058              1 2000000 3000000          12.474          13.287     2
## 4 NCIT:C3058              1 3000000 4000000          16.400          16.330     3
## 5 NCIT:C3058              1 4000000 5000000          11.661          18.513     4
## 6 NCIT:C3058              1 5000000 6000000          10.848          18.931     5
```

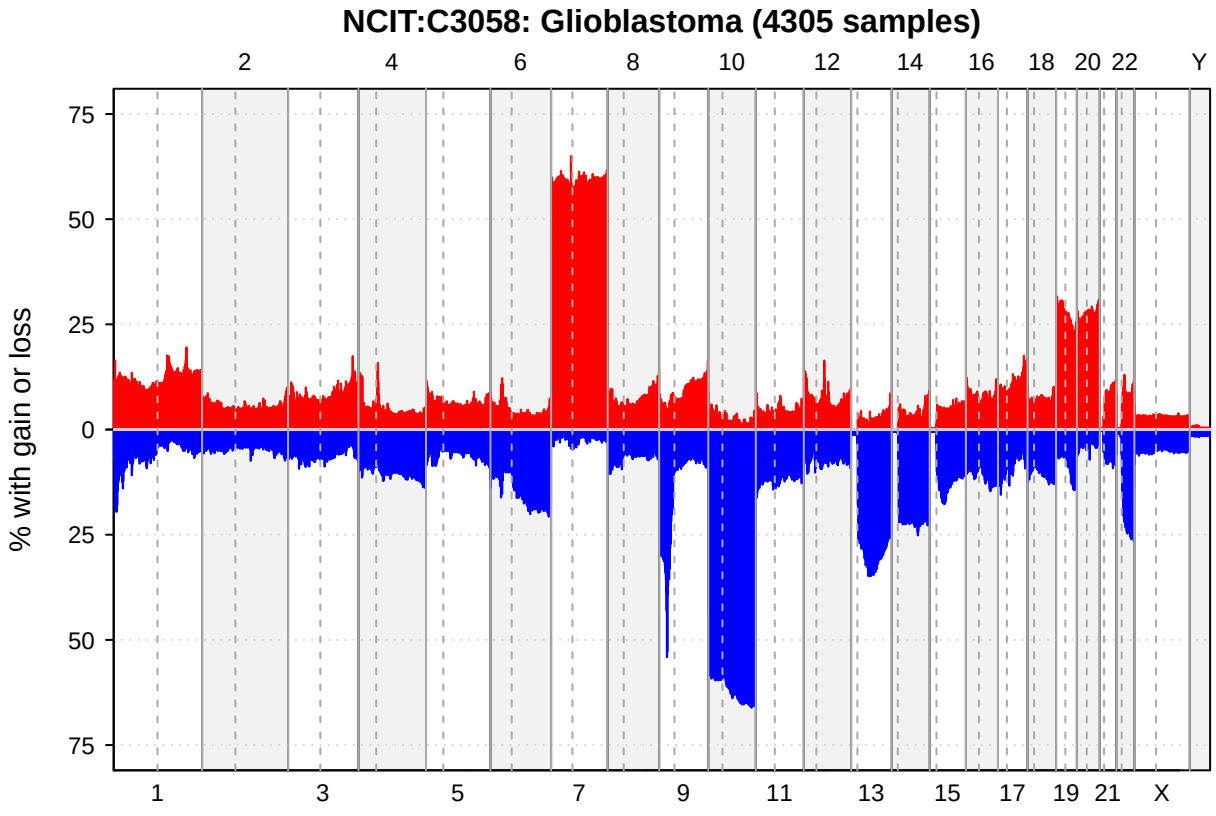
```
dim(freq$data$`NCIT:C3058`)
```

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

Step4: Visualize data

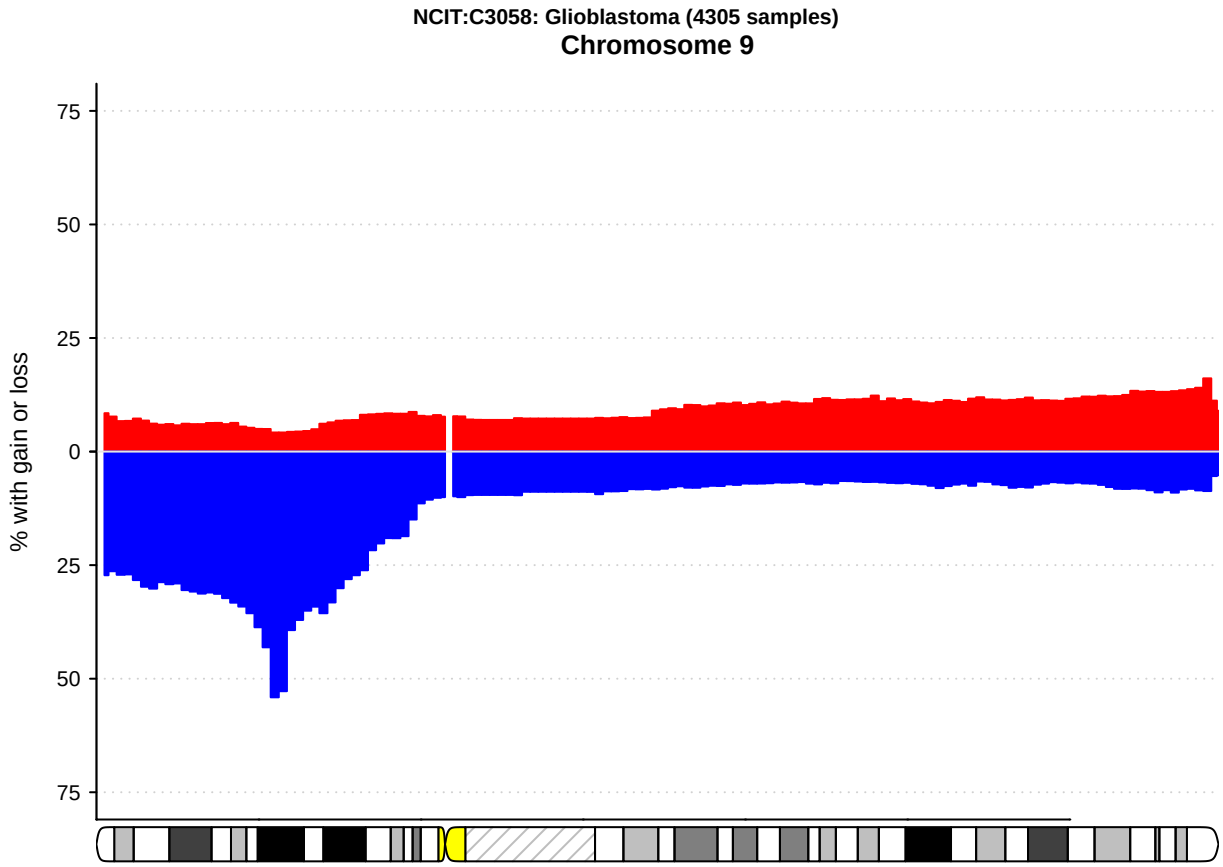
By genome

```
pgxFreqplot(freq)
```

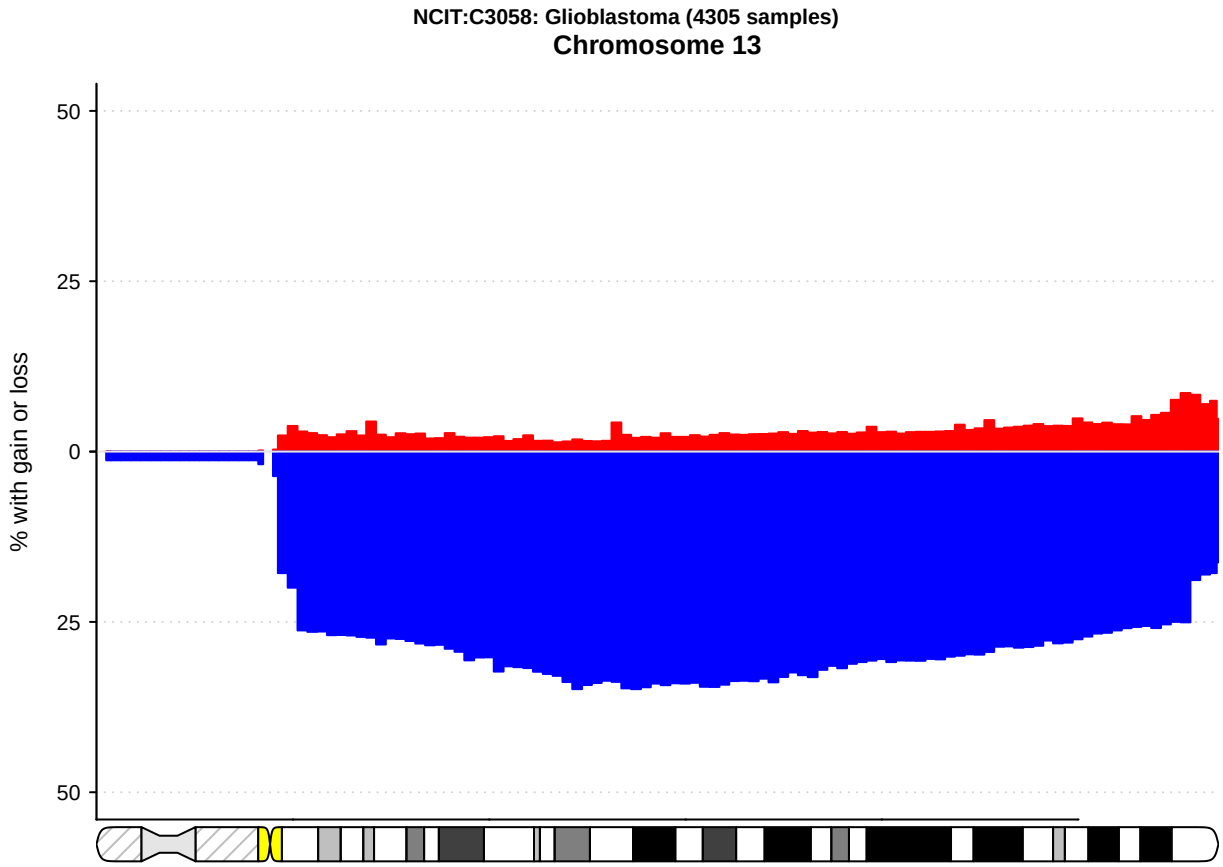


By chromosome

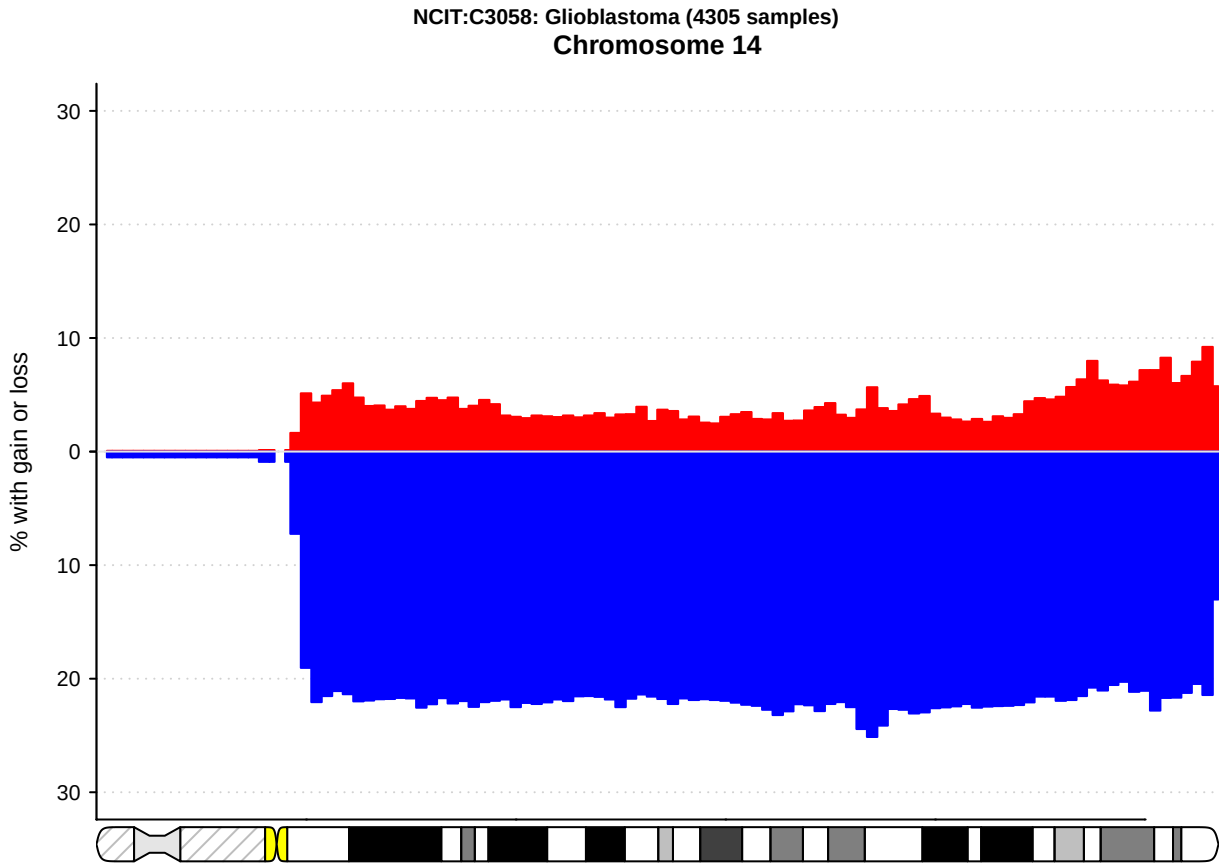
```
par(mfrow = c(2, 2))
pgxFreqplot(freq, chrom = 9)
```



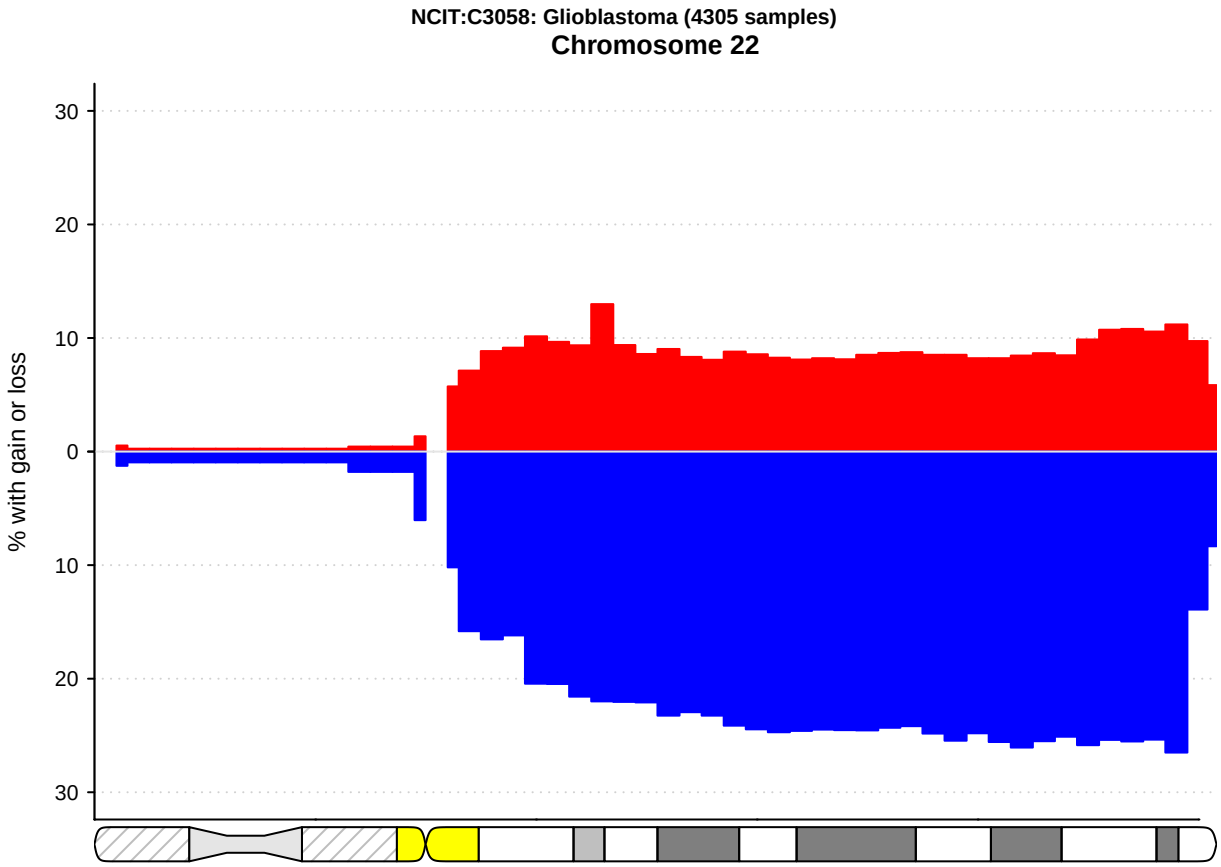

```
pgxFreqplot(freq, chrom = 13)
```




```
pgxFreqplot(freq, chrom = 14)
```




```
pgxFreqplot(freq, chrom = 22)
```



Step5: Analyse the data

According the plot, we can see frequency gains on chromosome 7p, 7q, 19p, 19q, 20p, 20q. and frequency losses on chromosome 9p, 10p, 10q, 13q, 14q, 15q, 17p, 18q, 22q.

The threshold for mentioning was arbitrarily chosen at roughly 20% gain or loss respectively (based on what sticks out in the genome visualization).

On chromosomes 13p, 14p and 22p there is nearly no gain at all and maybe 1% loss. This could indicate deletion of the whole p-half (potentially even both parental copies) of these chromosomes.

It would be interesting formulate a profile of gain versus loss CNV with respect to locus (in varying detail, e.g. which genes affected) and to compare this profile with other forms of malignant tumours.

Step6: Compare to peer studies

The proper workflow would be at this point to research literature regarding this topic to find supporting and/or contradicting result.