FocalCall: An R package for the annotation of focal copy number aberrations.

Oscar Krijgsman

October 17, 2016

Department of Pathology VU University Medical Center the Netherlands
Current adres: Department of Molecular Oncology Netherlands Cancer Institute the Netherlands

oscarkrijgsman@gmail.com o.krijgsman@nki.nl

Contents

1 Overview

To identify somatic focal copy number aberrations (CNAs) in cancer specimens and distinguish them from germ-line copy number variations (CNVs) we developed a software package named, focalCall. focalCall permits user-defined size cut-offs to recognize focal aberrations and builds on established array CGH segmentation and calling algorithms. To differentiate CNAs from CNVs the algorithm can either make use of a matched normal reference signal or, if this is not available, an external CNV track. focalCall furthermore differentiates between homozygous and hemizygous deletions as well as gains and amplifications and is applicable to high-resolution array and sequencing data.

2 Example

In this section we will use focalCall to analyse the dataset previously published by (**?**) Bierkens et al. (2013) and preprocessed using (**?**) by Wiel et

al. (2007). The example set used here only contains complete chromosome 2. For the other chromosomes only a small portion of the CGH probes are included. The complete dataset can be downloaded from the NCBI Gene Expression Omnibus (GEO), accession number GSE34575. First, we load the required packages and the data:

```
> library(focalCall)
> data(BierkensCNA)
```

Next, we apply the `focalCall` function which:

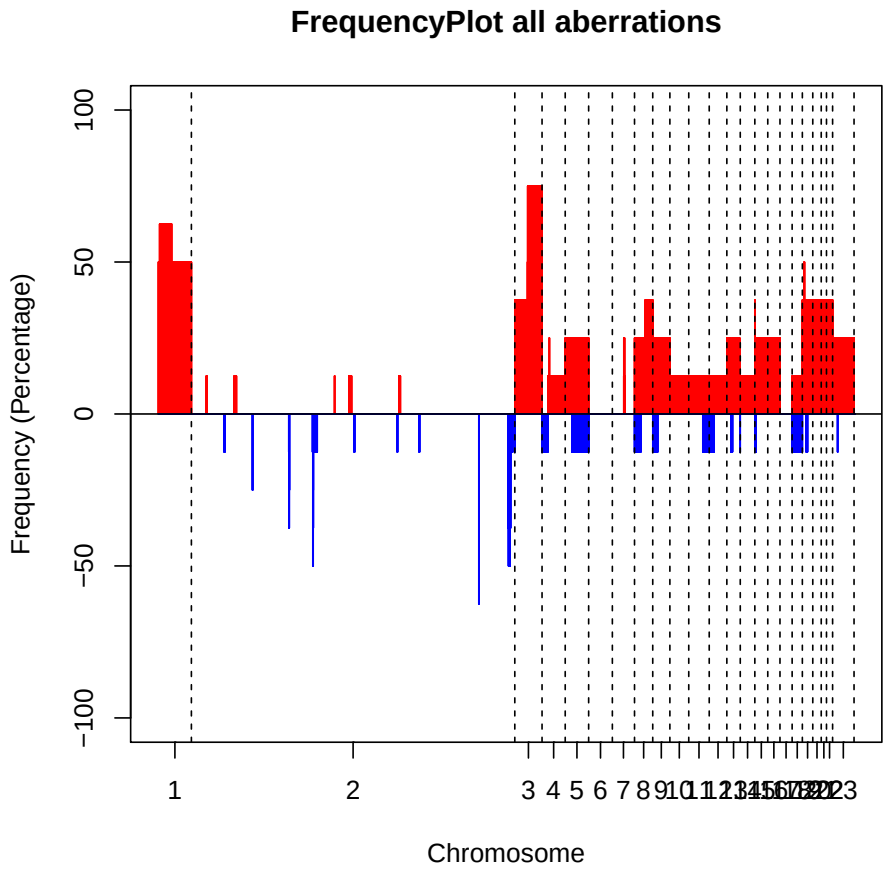
- detects all aberrations smaller than `focalSize` in Mb in the tumor data (CGHset)
- calculates peak regions for each genomic region where recurrent focal CNAs are detected
- compare each peak region to known copy number variants (CNVset)
- returns calls object with calls for focal CNA included
- report all focal CNAs as figures and tables

```
> calls_focals<-focalCall(CGHset, CNVset, focalSize = 3, minFreq=2)

Array resolution too low for calling aberrations smaller than 3MB.
Counted number of segments for each tumor sample
Generated matrix with detected aberrations < 3 Mb...
Start detection of peak regions...
Simple_ 1
Complex_ 2
Complex_ 3
Complex_ 4
Matching CNV list to copynumber data ...
Generating output files...
Written text file to workDir...
Written 'focalCall.RData' to workDir...
FINISHED!
Total time: 0 minutes
```

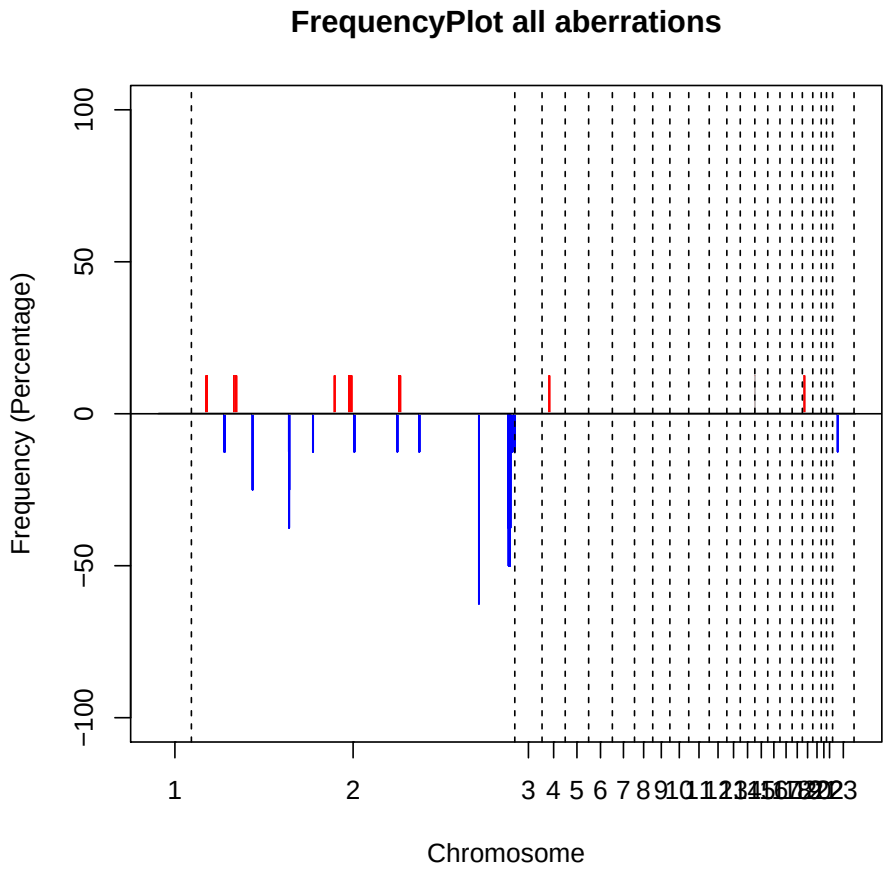
A frequency plot of all aberrations and all focal aberrations can be generated using `FreqPlot` and `FreqPlotfocal`.

```
> FreqPlot(calls_focals, header="FrequencyPlot all aberrations")
```




```
> FreqPlotfocal(calls_focals, header="FrequencyPlot all aberrations")

n
```


FrequencyPlot all aberrations
Frequency (Percentage)
100
50
0
−50
−100
1
2
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Chromosome

Alternatively, we can generate files for visualisation in IGV with `igvFiles`. The files will be written to the home directory and can be loaded in IGV directly.

```
> igvFiles(calls_focals)
```

```
Generated .SEG file for loading in IGV
Generated .IGV file for loading in IGV
Generated .IGV file for loading in IGV
```

3 Session Information

The version number of R and packages loaded for generating the vignette were:

```
> sessionInfo()

R version 3.3.1 (2016-06-21)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.9.5 (Mavericks)

locale:
[1] C/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] parallel  stats     graphics  grDevices utils     datasets  methods
[8] base

other attached packages:
 [1] focalCall_1.8.0     CGHcall_2.36.0      snowfall_1.84-6.1
 [4] snow_0.4-2          CGHbase_1.34.0      marray_1.52.0
 [7] limma_3.30.0        Biobase_2.34.0      BiocGenerics_0.20.0
[10] DNAcopy_1.48.0      impute_1.48.0

loaded via a namespace (and not attached):
[1] tools_3.3.1
```