Package 'patchworkCG'

May 13, 2013

Type Package

Title Allele-specific Copy Number Analysis of CompleteGenomics Whole Genome data

Version 2.0

Date 2012-11-01

Author Markus Rasmussen, Sebastian DiLorenzo

Maintainer Markus Rasmussen <Markus.Mayrhofer@medsci.uu.se>

Description Visualizes CompleteGenomics Whole genome data

License GPL-2

R topics documented:

`CG_KaCh` *Karyotype plotting of each chromosome*

Description

Visualises the calculated data of patchwork.CG.plot() for each chromosome. See details for a walk-through of the plot.

Usage

```
CG_KaCh(chr,start,end,int,ai,mchr,mpos,mval,sval,schr,spos,name,xlim,ylim)
```

Arguments

chr	Chromosomes name of segm object, as generated from somaticCnvSegmentsNonDiplod file.
start	Start position of segment.
end	End position of segment.
int	Ratio between relative coverage and segment length.
ai	Allelic imbalance as calculated from (1 - (lesser readcount / greater readcount)) per segment.
mchr	Chromosomes name of depcov (depthOfCoverage file).
mpos	Start position from depcov.
mval	Ratio between averagenormalizedcoverage and mean(averagenormalizedcoverage) from depthOfCoverage file.
sval	Inverse ratio between allele maximum and allele minimum per position.
schr	Chromosomes name of mastervar.
spos	Start position from mastervar.
name	Name that you want output to have. Default is ' ' which will yield files named as '_KaCh_chr#.png'. When method is run by patchwork.CG.plot() it will inherit the name parameter thereof.
xlim	Default is c(0,2.4). Limit of x axis for whole genome plot.
ylim	Default is c(0,1). Limit of y axis for whole genome plot.

Details

Walkthrough of the plot:

TOP
Vertical axis: Allelic Imbalance
Horizontal axis: Total Intensity

The chromosome plotted against the complete genome background. The separation between clusters within the plot are due to the fluctuating intensity and allelic imbalance and as such display the varying allele counts and copy numbers. Longer/larger segments have bigger circles. The varying coloration of the segments correspond to the amount of segments, since greater color scale variation is needed for a greater number of segments, and where they are located on the chromosome. They coincide and should be compared with the middle plot. Darker grey circles show more content as they are ontop of eachother.

MIDDLE
Vertical axis: Total Intensity
Horizontal axis: Chromosomal coordinate

The chromosome in questions total intensity plotted against the position on the chromosome.

BOTTOM
Vertical axis: Allelic Imbalance
Horizontal axis: Chromosomal coordinate

The chromosome in questions allelic imbalance plotted against the position on the chromosome.

Author(s)

Markus Mayrhofer, <markus.mayrhofer@medsci.uu.se>
Sebastian DiLorenzo, <sebastian.dilorenzo@medsci.uu.se>

See Also

patchwork.CG.plot
CG_karyotype

CG_KaChCN *Karyotype plotting for each chromosome post patchwork.CG.copynumbers().*

Description

Visualizes the calculated data of patchwork.CG.plot() + patchwork.CG.copynumbers() for each chromosome. See details for a walkthrough of the plot.

Usage

```
CG_KaChCN(chr,start,end,int,ai,Cn,mCn,mchr,mpos,mval,sval,schr,spos,name='',xlim=c(0,2.4),ylim=c
```

Arguments

chr	Chromosomes name of segm object.
start	Start position of segment.
end	End position of segment.
int	Ratio of relative coverage from somaticCnvSegmentsNonDiplod file and segment length.
ai	Allelic imbalance.
Cn	Copy number.
mCn	mCn.
mchr	Chromosomes name of depcov (depthOfCoverage file).
mpos	Start position from depcov.
mval	Ratio between averagenormalizedcoverage and mean(averagenormalizedcoverage) from depthOfCoverage file.
sval	Inverse ratio between allele maximum and allele minimum per position.
schr	Chromosomes name of mastervar.
spos	Start position from mastervar.
name	Name that you want output to have. Default is '' which will yield files named as '_KaChCN_chr#.png'. When method is run by patchwork.CG.copynumbers() it will that functions name parameter.
xlim	Default is c(0,2.4). Limit of x axis.
ylim	Default is c(0,1). Limit of y axis.
maxCn	Default is 8. The highest copynumber of the sample to check for.

Details

Walkthrough of the plot:

TOP
Vertical axis: Allelic Imbalance
Horizontal axis: Total Intensity

The chromosome plotted against the complete genome background. The separation between clusters within the plot are due to the fluctuating intensity and allelic imbalance and as such display the varying allele counts and copy numbers. Longer/larger segments have bigger circles. Darker circles show more content as they are ontop of eachother.

TOP MIDDLE
Vertical axis: Copynumber
Horizontal axis: Chromosomal coordinate

Displays the total and minor copynumbers for different segments of the chromosome in question.

LOWER MIDDLE
Vertical axis: Total Intensity
Horizontal axis: Chromosomal coordinate

The chromosome in questions total intensity plotted against the position on the chromosome.

BOTTOM
Vertical axis: Allelic Imbalance
Horizontal axis: Chromosomal coordinate

The chromosome in questions allelic imbalance plotted against the position on the chromosome.

Author(s)

Markus Mayrhofer, <markus.mayrhofer@medsci.uu.se> Sebastian DiLorenzo, <sebastian.dilorenzo@medsci.uu.

See Also

patchwork.CG.copynumbers

`CG_karyotype` *CompleteGenomics karyotype plotting function*

Description

Plots each, color coded by chromosomal coordinate, chromosome against a background of the complete genome.

Usage

```
CG_karyotype(chr,start,end,int,ai,name,xlim,ylim)
```

Arguments

chr	Chromosomes name of segm object, as generated from somaticCnvSegmentsNonDiplod file.
start	Start position of segment.
end	End position of segment.
int	Ratio between relative coverage and segment length.
ai	Allelic imbalance.
name	Default is ''. Plot generated as "name_karyotype.png" in your working directory.
xlim	Default is c(0,2.4). Limit of x axis.
ylim	Default is c(0,1). Limit of y axis.

Details

Vertical axis: Allelic imbalance.
Horizontal axis: Total intensity.

The plot is a overview, for a closer look see the plots generated by CG_KaCh().

Author(s)

Markus Mayrhofer, <markus.mayrhofer@medsci.uu.se>
Sebastian DiLorenzo, <sebastian.dilorenzo@medsci.uu.se>

See Also

patchwork.CG.plot
CG_KaCh

CG_Ka_check *Karyotype check function*

Description

Plots the whole genome coverage vs allelic imbalance with the approximated areas copynumbers and allele constitution. CG_Ka_check is called from the patchwork.CG.copynumbers() function. It would be hard to run this function outside of patchwork.CG.copynumbers().

Usage

```
CG_Ka_check(chr,start,end,int,ai,Cn,mCn,t,name='',xlim=c(0,2.4),ylim=c(0,1))
```

Arguments

chr	Chromosome.
start	Start position of segment.
end	End position of segment.
int	Ratio between relative coverage and segment length.
ai	Allelic imbalance as calculated from (1 - (lesser readcount / greater readcount)) per segment.
Cn	Copy number object created in patchwork.CG.copynumbers().
mCn	mCn, object created in patchwork.CG.copynumbers().
t	A list between int and ai. (ratio and allelic imbalance)
name	Default is ". When method is run by patchwork.CG.copynumbers() it will inherit that functions name parameter.
xlim	Default is c(0,2.4). Limit of x axis.
ylim	Default is c(0,1). Limit of y axis.

Details

Vertical axis: Allelic Imbalance.
Horizontal axis: Relative coverage.

The naming scheme is Copynumber-m-LesserAlleleDistribution So for example 2m0 means copynumber = 2, both allels are the same whereas 2m1 means copynumber = 2, 1 allele each.

Another example: 4m0, copynumber = 4, All allels are the same. (Loss of heterozygosity). 4m1, copynumber = 4, 3 alleles are the same, one is different. 4m2, copynumber = 4, 2 alleles each.

The total number of alleles present are always the copynumber.

Author(s)

Markus Mayrhofer, <markus.mayrhofer@medsci.uu.se>
Sebastian DiLorenzo, <sebastian.dilorenzo@medsci.uu.se>

See Also

patchwork.CG.copynumbers

---

ideogram *Ideogram data for the chromosomes.*

Description

Ideogram data for the chromosomes.

Usage

```
ideogram
```

Format

A data frame with the following 6 variables.

chr Which chromosome

length Length of chromosome

start Start of centromere

mid Middle of centromere

end End of centromere

c Easy integer of chromosome number

Details

Ideogram data for the chromosomes. Used in many functions.

See Also

The functions `patchwork.CG.plot` and `patchwork.CG.copynumbers` plot functions use this data.

`patchwork.CG.copynumbers` *Function plot allele-specific copy numbers of the genome*

Description

Input data for this function should be assessed from any of the CG_KaCh() plots.

See argument details.

Very important: For this function to work you must have the file CG.Rdata, generated using patchwork.CG.plot(), in your working directory!

Usage

```
patchwork.CG.copynumbers(cn2,delta,het,hom,maxCn=8,ceiling=1,name="copynumbers_",
CGfile=NULL,forcedelta=F)
```

Arguments

`cn2`	The approximate position of copy number 2, diploid, on total intensity axis.
`delta`	The difference in total intensity between consecutive copy numbers. For example 1 and 2 or 2 and 3. If copy number 2 has total intensity ~0.6 and copy number 3 har total intensity ~0.8 then delta would be 0.2.
`het`	Allelic imbalance ratio of heterozygous copy number 2.
`hom`	Allelic imbalance ratio of Loss-of-Heterozygosity copy number 2.
`maxCn`	Highest copy number to calculate for. Default is 8.
`ceiling`	Default is 1.
`name`	Default is "copynumbers_". The name you want attached to generated plots.

CGfile	Default is NULL. If your CG.Rdata file is not in your working directory, and you dont wish to move it to your working directory, you can simply input the path here as CGfile = "path/to/file/CG.Rdata" so patchwork.CG.copynumbers() can find its data.
forcedelta	Default is FALSE. If TRUE the delta value will be absolute and not subject to adjustments.

Details

1. Run the patchwork.CG.plot correctly. This will generate CG.Rdata.
2. View your plots and decide the correct parameters for the other arguments of this function.
3. Make sure CG.Rdata is in your working directory, or set the CGfile parameter to point to CG.Rdata.
4. Run the function.

Upon completion two files will be generated, copynumbers.Rdata which contains the objects used to generate the plots and Copynumbers.csv. Other than this plots from CG_Ka_check and CG_KaChCN will be generated.

Author(s)

Markus Mayrhofer, <markus.mayrhofer@medsci.uu.se>
Sebastian DiLorenzo, <sebastian.dilorenzo@medsci.uu.se>

See Also

patchwork.CG.plot
subfunctions_copynumbers
CG_Ka_check
CG_KaChCN

Examples

```
## Not run:
>library(patchworkCG)

#With CG.Rdata in working directory. Values for parameters estimated from
#HCC2218 chromosome 1.
#See plot patchworkCG/example_plots/HCC2218_KaCh_chr1.png
>patchwork.CG.copynumbers(cn2=0.85,delta=0.35,het=0.25,hom=0.95,name="HCC2218_")

#Pointing to location of CG.Rdata file
>patchwork.CG.copynumbers(cn2=0.85,delta=0.35,het=0.25,hom=0.95,name="HCC2218_",
CGfile="path/to/CG.Rdata")

#Load the objects generated from patchwork.CG.copynumbers
#For comparison purposes only
>load("copynumbers.Rdata")

> head(ai)
$cn2m1
[1] 0.2460513

$cn2m0
[1] 0.95
```

```
$cn1m0
[1] 0.9690196

$cn0m0
[1] NA

$cn3m1
[1] 0.2677295

$cn3m0
[1] 0.9671712

> head(int)
$cn2
[1] 0.85

$cn1
[1] 0.5

$cn0
[1] 0.15

$cn3
[1] 1.2

$cn4
[1] 1.55

$cn5
[1] 1.9

> head(regions)
   chr  start    end avgnormcov relcov     ratio snvs        ai np Cn mCn
1 chr1  10000 177417       79.0   1.05 0.9660335   24 0.5531915 17  2   1
2 chr1 227417 267719       79.0   1.05 0.9660335    2 0.3243243  4  2   1
3 chr1 317719 471368       79.0   1.05 0.9660335    2 0.7826087 15  2   0
4 chr1 521368 600000      170.8   2.26 2.0792721    2 0.1944444  8  6   3
5 chr1 600000 700000       74.6   0.99 0.9108316    2 0.5333333 10  2   1
6 chr1 700000 900000       51.9   0.69 0.6348220  109 0.4960325 20  1   0
  fullCN
1  cn2m1
2  cn2m1
3  cn2m0
4  cn6m3
5  cn2m1
6  cn1m0

## End(Not run)
```

`patchwork.CG.plot` *The main function of patchworkCG.*

Description

Reads files and performs calculations later used in the plotting functions.

In details there is some very important information, highly recommended!

Usage

```
patchwork.CG.plot(path=NULL,name='CG_sample',manual_file_input=FALSE,
masterVarBeta=NULL,somaticCnvSegments=NULL,
depthOfCoverage=NULL)
```

Arguments

path	Path to the ASM folder. One of the subfolders of your completegenomics directory.
name	Default is 'CG_sample'. The name you wish associated with the plots that will be generated.
manual_file_input	Default is FALSE. If you set it to TRUE you will be prompted to provide path and filename for the masterVarBeta, somaticCnvSegmentsNondiploid and depthOfCoverage file.
masterVarBeta	Path to and COMPLETE NAME of the masterVarBeta file, should you wish to implement it directly. Useful if you want to run the program on multiple samples and wish to create a script going from file to file. This saves you from having to conserve the CompleteGenomics file structure. Default is NULL.
somaticCnvSegments	Path to and COMPLETE NAME of the somaticCnvSegments file, should you wish to implement it directly. Useful if you want to run the program on multiple samples and wish to create a script going from file to file. This saves you from having to conserve the CompleteGenomics file structure. Default is NULL.
depthOfCoverage	Path to and COMPLETE NAME of the depthOfCoverage file, should you wish to implement it directly. Useful if you want to run the program on multiple samples and wish to create a script going from file to file. This saves you from having to conserve the CompleteGenomics file structure. Default is NULL.

Details

Before execution make sure that you have unpacked the necessary files. masterVarBeta, located in your ASM/ folder,somaticCnvSegmentsNonDiploidBeta, located in your ASM/CNV/ folder, and depthOfCoverage, also located in your ASM/CNV/ folder.

If you wish to use this function in a script so it performs its function on multiple genome entries, remember to also loop over the "name" parameter or you run the risk of the plots replacing eachother. The easiest way to script this to run over multiple files is to use the masterVarBeta, somaticCnvSegments and depthOfCoverage parameters as opposed to the path parameter.

It is prudent not do run multiple instances of patchwork.CG.plot in the same directory. This can cause issues with the generated file, read below, and also with plot names if you have kept the default value.

During execution of patchwork.CG.plot a file named CG.Rdata will be built then saved in your working directory. This file is used for plotting purposes in patchwork.CG.copynumbers(). As such, patchwork.CG.copynumbers() should either be executed in the same working directory as this file is located or using the CGfile parameter of patchwork.CG.copynumbers().

Author(s)

Markus Mayrhofer, <markus.mayrhofer@medsci.uu.se>
Sebastian DiLorenzo, <sebastian.dilorenzo@medsci.uu.se>

Examples

```
## Not run:

#Load patchworkCG
>library(patchworkCG)

#An example of running patchwork.CG.plot. Note that the path is to the ASM folder
#of the default CompleteGenomics File structure.
>patchwork.CG.plot(path="Data/HCC2218/GS00258-DNA_B03/GS00258-DNA_B03/ASM",
name="HCC2218_")

Reading files from ASM folder
File input complete
Performing calculations
Calculations complete
Saving objects to CG.Rdata
Initiating Plotting
Plotting Complete
patchwork.CG.plot Complete
Please read documentation on running patchwork.CG.copynumbers

#The values used to generate the plots are saved in three objects,
#mastervar,segs and depcov, and can be viewed loading CG.Rdata.
>load("CG.Rdata")

> head(mastervar)
      chr begin   end vartype ref_count tot_count ref_countN tot_countN
876  chr1 46669 46670     snp         1         1          2          3
912  chr1 49313 49315     sub        69       129         71        123
1030 chr1 55387 55388     snp        64       211         78        247
1042 chr1 55815 55816     snp         5        12          3         10
1054 chr1 56484 56485     snp         0        11          2          6
1082 chr1 57989 57990     snp        58       187         56        177
     mut_count max min      ratio
876          0   1   0 0.01635786
912         60  69  60 2.11016381
1030       147 147  64 3.45150825
1042         7   7   5 0.19629431
1054        11  11   0 0.17993645
1082       129 129  58 3.05891963

> head(segs)
   chr   start     end avgnormcov relcov     ratio snvs        ai
1 chr1   10000  177417       88.2   1.36 1.0434479   42 0.5449591
2 chr1  227417  267719       88.2   1.36 1.0434479    0 0.0000000
3 chr1  317719  471368       88.2   1.36 1.0434479    2 1.0000000
4 chr1  521368  600000       61.4   0.95 0.8015226    1 1.0000000
5 chr1  600000 2634220       93.9   1.45 1.0965535  895 0.6904948
6 chr1 2684220 3845268       93.9   1.45 1.0965535  570 0.8107859

> head(depcov)
```

```
   chr  begin    end avgnormcov     ratio
1 chr1  10000 100000       94.8 1.0897044
2 chr1 100000 177417      102.9 1.1622234
3 chr1 227417 267719      124.4 1.3547120
4 chr1 317719 400000       99.5 1.1317833
5 chr1 400000 471368       83.9 0.9921172
6 chr1 521368 600000       70.1 0.8685663

## End(Not run)
```

---

patchworkCG.readme *Welcome to patchworkCG*

Description

PatchworkCG allows you to obtain allele-specific copy number information from Complete Genomics data.

To use patchwork on other data formats,bam and pileups, download patchwork and patchworkData.

//
http://patchwork.r-forge.r-project.org/
//

It is highly recommended that you visit the webpage for this project as it contains the most recently updated information and instructions regarding everything patchwork.

Details

//
Installation Guide: PatchworkCG
//

Start the latest version of R.(as of writing 2.14.2)

Text with a ">" infront is R executions.

> install.packages("patchworkCG", repos="http://R-Forge.R-project.org")

If for some reason that does not work add the 'type="source"' to it, as so:

> install.packages("patchworkCG", repos="http://R-Forge.R-project.org",type="source")

The source codes can also be downloaded outside of R at https://r-forge.r-project.org/R/?group_id=1250

//////////////////
Tutorial:
//////////////////

//////////////////
Requirements:
//////////////////

An unpacked CompleteGenomics tumor genome data ASM directory. Specifically the files ASM/masterVarBeta, CNV/depthOfCoverage and CNV/somaticCnvSegmentsNondiploidBeta.

Some helpful commands for unpacking your CompleteGenomics data (UNIX):
To unpack .tar files:tar -zxvf filename.tar
To unpack .bz2 files:bunzip2 filename.bz2

////////////////////////////////
Execution: patchwork.CG.plot()
////////////////////////////////

Start R in your working directory. (Or use setwd() after starting R wherever you choose)

> library(patchworkCG)

See

> ?patchwork.CG.plot

Excerpt from ?patchwork.CG.plot:

//
Usage:

patchwork.CG.plot(path=NULL,name='CG_sample',manual_file_input=FALSE, masterVarBeta=NULL,somaticCnvSe
depthOfCoverage=NULL)

Arguments:

path: Default is NULL. Path to the ASM folder. One of the subfolders of your completegenomics directory.

name: Default is 'CG_sample'. The name you wish associated with the plots that will be generated.

manual_file_input: Default is FALSE. If you set it to TRUE you will be prompted to provide path and filename for the masterVarBeta, somaticCnvSegmentsNondiploid and depthOfCoverage file.

masterVarBeta: Path to AND COMPLETE name of the masterVarBeta file, should you wish to implement it directly. Useful if you want to run the program on multiple samples and wish to create a script going from file to file. This saves you from having to conserve the CompleteGenomics file structure. Default is NULL.

somaticCnvSegments: Path to AND COMPLETE name of the somaticCnvSegments file, should you wish to implement it directly. Useful if you want to run the program on multiple samples and wish to create a script going from file to file. This saves you from having to conserve the CompleteGenomics file structure. Default is NULL.

depthOfCoverage: Path to AND COMPLETE name of the depthOfCoverage file, should you wish to implement it directly. Useful if you want to run the program on multiple samples and wish to create a script going from file to file. This saves you from having to conserve the CompleteGenomics file structure. Default is NULL. //

Probably the easiest way of running patchwork.CG.plot is pointing to the ASM folder:

> patchwork.CG.plot(path="path/to/ASM/")
Reading files from ASM folder
File input complete
Performing calculations
Calculations complete
Saving objects to CG.Rdata
Initiating Plotting
Plotting Complete
patchwork.CG.plot Complete
Please read documentation on running patchwork.CG.copynumbers

Once it is complete there should be several plots in your working directory as well as the file "CG.Rdata".

Execution should take no longer half an hour.

//
Execution: patchwork.CG.copynumbers()
//

For this function to work you will need to have obtained the CG.Rdata file from patchwork.CG.plot(). If it is located in your working directory you will not need to point to it, otherwise you can input it with the argument "CGfile".

See

> ?patchwork.CG.copynumbers

Excerpt from ?patchwork.CG.copynumbers:

//
Usage:

patchwork.CG.copynumbers(cn2,delta,het,hom,maxCn=8,ceiling=1,name="copynumbers_", CGfile=NULL,forcedelta=

Arguments:

cn2: The approximate position of copy number 2, diploid, on total intensity axis.

delta: The difference in total intensity between consecutive copy numbers. For example 1 and 2 or 2 and 3. If copy number 2 has total intensity ~0.6 and copy number 3 har total intensity ~0.8 then delta would be 0.2.

het: Allelic imbalance ratio of heterozygous copy number 2.

hom: Allelic imbalance ratio of Loss-of-Heterozygosity copy number 2.

maxCn: Highest copy number to calculate for. Default is 8.

ceiling: Default is 1.

name: Default is "copynumbers_". The name you want attached to generated plots.

CGfile: Default is NULL. If your CG.Rdata file is not in your working directory, and you dont wish to move it to your working directory, you can simply input the path here as CGfile = "path/to/file/CG.Rdata" so patchwork.CG.copynumbers() can find its data.

forcedelta: Default is FALSE. If TRUE the delta value will be absolute and not subject to adjustments. //

On our webpage there will also be a picture accompanying this portion pointing to the argument values should the documentation for the function not be sufficient. (http://patchwork.r-forge.r-project.org/)

> patchwork.CG.copynumbers(cn2=0.8, delta=0.3, het=0.2, hom=0.8, name="Example_", CGfile="path/to/CG.Rdata")

This should have generated several plots in your working directory.

//////////////////////////////////
Interpreting plots
//////////////////////////////////

It is recommended that you visit our project homepage for guidance in interpreting the plots you have generated using patchworkCG.

http://patchwork.r-forge.r-project.org/

Other aids in this is the documentation

>?CG_Ka_check
>?CG_KaCh
>?CG_KaChCN
>?CG_karyotype

If you have any questions please feel free to contact us and we will help you to the best of our extent!

Author(s)

Sebastian DiLorenzo, <sebastian.dilorenzo@medsci.uu.se>
Markus Mayrhofer, <markus.mayrhofer@medsci.uu.se>

`subfunctions_copynumbers` *A number of subfunctions used within patchwork.CG.copynumbers().*

Description

A number of subfunctions used within patchwork.CG.copynumbers(). These functions are identical in patchwork package, so should that package be installed before patchworkCG or vice versa you will get a non-consequential message saying:

The following object(s) are masked from 'package:patchwork':

chrom_ucsc, deChrom_ucsc, is.autosome, weightedMean, weightedMedian

Usage

```
weightedMedian(data,weights)
weightedMean(data,weights)
is.autosome(vector)
deChrom_ucsc(data)
chrom_ucsc(data)
```

Arguments

`data`	Object to be handled by the internal function.
`weights`	Weights to be applied.
`vector`	The segm objects chr column.

Author(s)

Markus Mayrhofer, <markus.mayrhofer@medsci.uu.se>
Sebastian DiLorenzo, <sebastian.dilorenzo@medsci.uu.se>

See Also

`patchwork.CG.copynumbers`

Index