iCNV

December 14, 2017

`bambaf_from_vcf` *Get BAM baf information from vcf*

Description

If your vcf follow the format in the example, you could use this function to extract NGS baf from vcf files. Remember to load library before hands. Save 6 lists, each list has N entry. N = # of individuals (or vcf file) ngs_baf.nm: name of the bamfiles; ngs_baf.chr: the chromosome; ngs_baf.pos: the position of the variants; ngs_baf: the BAF of the variants; ngs_baf.id: the ID of the variants; filenm:the file name

Usage

```
bambaf_from_vcf(dir = ".", vcf_list, chr = NULL, projectname = "")
```

Arguments

`dir`	The directory to all the vcf stored; default is right in this folder.
`vcf_list`	All the vcf names stored in vcf.list; could use command:"ls *.vcf > vcf.list" to generate.
`chr`	Specify the chromosome you want to generate. Must be of int from 1-22. If not specify, this function will generate all chromosomes.
`projectname`	Name of the project

Value

void

Examples

```
## Not run:
dir='PATH/TO/FOLDER'
bambaf_from_vcf(dir,'example_vcf.list')
bambaf_from_vcf(dir,'example_vcf.list',chr=22)
load('bambaf_22.rda')
str(ngs_baf)
str(ngs_baf.pos)

## End(Not run)
```

bed_generator *Generate BED file for WGS dataset.*

Description

Default position generated from USCS genome browser

Usage

```
bed_generator(chr, hg, start = NULL, end = NULL, by = 1000)
```

Arguments

chr	Specify the chromosome you want to generate. Must be of int from 1-22
hg	Specify the coordinate you want to generate from. Start and end position of hg19 and hg38 have been pre-implemented.
start	The start position of your BED file.
end	The end position of your BED file.
by	The chunk of your DNA for each bin. Default 1kb.

Value

void

Examples

```
bed_generator(chr=22,hg=38)
bed_generator(22,38,5001,10000,by=500)
```

get_array_intput *Get array information from given format*

Description

If your array input file follow the format in the example, you could use this function to extract array LRR and baf. Remember to load library before hands. Save 4*[# of chr] lists, each list has N entry. N = # of individuals snp_lrr: SNP LRR intensity; snp_lrr.pos: the position of the SNPs snp_baf: the BAF of the SNPs; snp_baf.pos: the position of the SNPs

Usage

```
get_array_intput(dir, pattern, chr = NULL, projectname = "")
```

Arguments

dir	A string. The directory path to the folder where store signal intensity file according to chr
pattern	A string. The pattern of all the intensity file
chr	Specify the chromosome you want to generate. Must be of int from 1-22. If not specify, this function will generate files for all chromosomes.
projectname	Name of the project

Value

void

Examples

```
## Not run:
dir='PATH/TO/FOLDER'
pattern=paste0('*.csv.arrayicnv$')
icnv_array_intput(dir,pattern,chr=22)
load('icnv_array_input_22.rda')
str(snp_lrr)
str(snp_lrr.pos)
str(snp_baf)
str(snp_baf.pos)

## End(Not run)
```

---

iCNV_detection *CNV detection*

Description

Copy number variation detection tool for germline data. Able to combine intensity and BAF from SNP array and NGS data.

Usage

```
iCNV_detection(ngs_plr = NULL, snp_lrr = NULL, ngs_baf = NULL,
  snp_baf = NULL, ngs_plr.pos = NULL, snp_lrr.pos = NULL,
  ngs_baf.pos = NULL, snp_baf.pos = NULL, maxIt = 50, visual = 0,
  projname = "iCNV", CN = 0, mu = c(-3, 0, 2), cap = FALSE)
```

Arguments

ngs_plr	A list of NGS intensity data. Each entry is an individual. If no NGS data, no need to specify.
snp_lrr	A list of SNP array intensity data. Each entry is an individual. If no SNP array data, no need to specify.
ngs_baf	A list of NGS BAF data. Each entry is an individual. If no NGS data, no need to specify.
snp_baf	A list of SNP array BAF data. Each entry is an individual. If no SNP array data, no need to specify.
ngs_plr.pos	A list of NGS intensity postion data. Each entry is an individual with dimension= (#of bins or exons, 2(start and end position)). If no NGS data, no need to specify.
snp_lrr.pos	A list of SNP array intensity postion data. Each entry is an individual with length=#of SNPs. If no SNP array data, no need to specify.
ngs_baf.pos	A list of NGS BAF postion data. Each entry is an individual with length=#of BAFs. If no NGS data, no need to specify.

snp_baf.pos	A list of SNP array BAF postion data. Each entry is an individual with length=#of BAFs. If no SNP array data, no need to specify.
maxIt	An integer number indicate the maximum number of EM iteration if not converged during parameter inference. Default 50.
visual	An indicator variable with value 0,1,2. 0 indicates no visualization, 1 indicates basic visualization, 2 indicates complete visualization (Note visual 2 only work for single platform and integer CN inferenced)
projname	A string as the name of this project. Default 'iCNV'
CN	An indicator variable with value 0,1 for whether wants to infer exact copy number. 0 no exact CN, 1 exact CN. Default 0.
mu	A length tree vectur specify means of intensity in mixture normal distribution (Deletion, Diploid, Duplification). Default c(-3,0,2)
cap	A boolean decides whether we cap insane intensity value due to double deletion or mutiple amplification. Default False

Value

(1) CNV inference, contains CNV inference, Start and end position for each inference, Conditional probability for each inference, mu for mixture normal, sigma for mixture normal, probability of CNVs, Z score for each inference.

(2) exact copy number for each CNV inference, if CN=1.

Examples

```
# icnv call without genotype (just infer deletion, duplication)
projname='icnv.demo'
icnv_res0=iCNV_detection(ngs_plr,snp_lrr,
                         ngs_baf,snp_baf,
                         ngs_plr.pos,snp_lrr.pos,
                         ngs_baf.pos,snp_baf.pos,
                         projname=projname,CN=0,mu=c(-3,0,2),cap=TRUE,visual = 1)
# icnv call with genotype inference and complete plot
## Not run:
projname='icnv.demo.geno'
icnv_res1=iCNV_detection(ngs_plr,snp_lrr,
                         ngs_baf,snp_baf,
                         ngs_plr.pos,snp_lrr.pos,
                         ngs_baf.pos,snp_baf.pos,
                         projname=projname,CN=1,mu=c(-3,0,2),cap=TRUE,visual = 2)

## End(Not run)
```

---

icnv_output_to_gb *Convert icnv.output to input for Genome Browser.*

Description

We could add the output to custom tracks on Genome Browser. Remeber to choose human assembly matches your input data. We color coded the CNVs to make it as consistant as IGV. To show color, click 'User Track after submission', and edit config to 'visibility=2 itemRgb="On"'. Color see Github page for more example.

Usage

```
icnv_output_to_gb(chr, icnv.output)
```

Arguments

chr	CNV chromosome
icnv.output	output from output_list_function

Value

matrix for Genome browser

Examples

```
icnv.output = output_list(icnv_res=icnv_res,sampleid=sampname_qc, CN=0, min_size=10000)
gb_input = icnv_output_to_gb(chr=22,icnv.output)
write.table(gb_input,file='icnv_res_gb_chr22.tab',quote=FALSE,col.names=FALSE,row.names=FALSE)
```

---

output_list *Generate ouput list.*

Description

Generate human readable output from result calculated by iCNV_detection function

Usage

```
output_list(icnv_res, sampleid = NULL, CN = 0, min_size = 0)
```

Arguments

icnv_res	CNV inference result. Output from iCNV_detection()
sampleid	the name of the sample, same order as the input
CN	An indicator variable with value 0,1 for whether exact copy number inferred in iCNV_detection. 0 no exact CN, 1 exact CN. Default 0.
min_size	A integer which indicate the minimum length of the CNV you are interested in. This could remove super short CNVs due to noise. Default 0. Recommend 1000.

Value

output CNV list of each individual

Examples

```
icnv.output = output_list(icnv_res=icnv_res,sampleid=sampname_qc, CN=0)
```

plot_intensity *plot out the NGS plr or array lrr.*

Description

For quality checking purpose during intermediate steps

Usage

```
plot_intensity(intensity, chr)
```

Arguments

intensity	Specify the ngs_plr object generated by CODEX or SNP array.
chr	Specify the chromosome you want to generate. Must be of int from 1-22

Value

void

Examples

```
plot_intensity(ngs_plr,chr)
plot_intensity(snp_lrr,chr)
```

plotHMMscore *Plot CNV inference score.*

Description

Plot out CNV inference score. Each row is a sample, each column is a SNP or, exon (WES) or bin (WGS). Red color indicate score favor duplication whereas blue favor deletion.

Usage

```
plotHMMscore(icnv_res, h = NULL, t = NULL, title = "score plot",
  output = NULL)
```

Arguments

icnv_res	CNV inference result. Result from iCNV_detection() (i.e. iCNV_detection(...))
h	start position of this plot. Default Start of the whole chromosome
t	end position of this plot. Default End of the whole chromosome
title	of this plot. Default "score plot"
output	generated from output_list_function. If it isn't null, only CNVs in output file will be highlighted. Default NULL

Value

void

Examples

```
pdf(file=paste0(projname,'.pdf'),width=13,height = 10)
plotHMMscore(icnv_res,h=20000000, t=30000000, title='my favorite subject')
dev.off()
```

`plotindi` *Individual sample plot*

Description

Plot relationship between platforms and features for each individual. Only work for muli-platform inference.

Usage

```
plotindi(ngs_plr, snp_lrr, ngs_baf, snp_baf, ngs_plr.pos, snp_lrr.pos,
  ngs_baf.pos, snp_baf.pos, icnvres, I, h = NULL, t = NULL)
```

Arguments

`ngs_plr`	A list of NGS intensity data. Each entry is an individual. If no NGS data, no need to specify.
`snp_lrr`	A list of SNP array intensity data. Each entry is an individual. If no SNP array data, no need to specify.
`ngs_baf`	A list of NGS BAF data. Each entry is an individual. If no NGS data, no need to specify.
`snp_baf`	A list of SNP array BAF data. Each entry is an individual. If no SNP array data, no need to specify.
`ngs_plr.pos`	A list of NGS intensity postion data. Each entry is an individual with dimension= (#of bins or exons, 2(start and end position)). If no NGS data, no need to specify.
`snp_lrr.pos`	A list of SNP array intensity postion data. Each entry is an individual with length=#of SNPs. If no SNP array data, no need to specify.
`ngs_baf.pos`	A list of NGS BAF postion data. Each entry is an individual with length=#of BAFs. If no NGS data, no need to specify.
`snp_baf.pos`	A list of SNP array BAF postion data. Each entry is an individual with length=#of BAFs. If no SNP array data, no need to specify.
`icnvres`	CNV inference result. The output from iCNV_detection()
`I`	Indicating the position of the individual to plot
`h`	start position of this plot. Default Start of the whole chromosome
`t`	end position of this plot. Default End of the whole chromosome

Value

void

Examples

```
pdf(file=paste0(projname,'.pdf'),width=13,height = 10)
plotindi(ngs_plr,snp_lrr,ngs_baf,snp_baf,
 ngs_plr.pos,snp_lrr.pos,ngs_baf.pos,snp_baf.pos,
 icnv_res,I=1,h=20000000, t=30000000)
dev.off()
```

Index