BBCAnalyzer: A Visual Approach to Facilitate Variant Calling

Documentation of the web application: BBCAnalyzer_local

1 Left panel

The left panel corresponds to the R/Bioconductor package *BBCAnalyzer*-function "analyze-Bases()". It performs a complete analysis of the defined samples and target regions with BBCAnalyzer. The output (txt- and png-files) is saved in the defined output directory. Additionally, the application's progress as well as the generated plots are displayed in the right panel.

Samples to analyze	Names of the samples to analyze (one entry per line, no file name extensions).
Define folder containing bam- and bai-files	Input folder containing the alignment data of all samples to analyze. For each sample defined in "Samples to analyze" there must be a bam- and a bai-file with matching names in the defined folder. Important: Windows-User have to define the folder using double-backslashes (e.g. C:\\home\\BBCAnalyzer_local\\Scripts\\).
Target regions to analyze	Target regions to analyze. Single positions (chromosome and position) as well as longer regions (chromosome, start, end) are equally supported.
Define folder containing vcf files (optional)	Input folder containing vcf files for all samples to analyze. For each sample defined in "Samples to analyze" there must be a vcf-file with matching names in the defined folder.
Define output folder	Folder where output files shall be saved
Define tabix file containing known variants (optional)	Tabix file (format: file.vcf.gz.tbi) containing known variants. For the correct functioning of this option the file "file.vcf.gz" is equally necessary (same folder as "file.vcf.gz.tbi"; does not have to be defined anywhere).
Select reference genome for analysis	Select one of the available reference genomes for analysis. If a genome has not yet been installed, it is automatically downloaded and installed (this process can take a few minutes). Default: BSgenome.Hsapiens.UCSC.hg19.
Mapping quality threshold	A PHRED-scaled value to be used as a mapping quality threshold. All reads with a mapping quality below this threshold are excluded from analysis. Every base in an excluded read gets marked in the output. Default: 60.
Base quality threshold	A PHRED-scaled value to be used as a base quality threshold. All bases with a base quality below this threshold are excluded from analysis. Every excluded base gets marked in the output. The num-

	ber of excluded bases per position gets counted and reported. Default: 50.
Frequency threshold for variant reporting	A frequency to be used as a threshold for variants to be reported. Default: 0.01.
Lower- and upper mean quality bound for color-coding	The lower- and upper bound for the mean quality that shall be color-coded in the plots. All bases with a mean quality below the lower bound are colored with the lightest color defined for the corresponding base. All bases with a mean quality above the upper bound are colored with the darkest color defined for the corresponding base. If the bases shall not be color-coded according to their mean quality, the definable range has to be zero. Default: 58-63.
Select levels at which marks shall be drawn	Levels (relative number of reads) at which horizontal lines shall be drawn in the plots.
Plot number of reads	Relative or absolute number of reads can be plotted. Default: Relative.
Create one plot per	One plot per sample or one plot per position can be created. Default: Sample.

2 Middle panel

The middle panel corresponds to the R/Bioconductor package *BBCAnalyzer*-function "analyze-BasesPlotOnly()". It creates the plots for the defined samples and the previously analyzed target regions with BBCAnalyzer. "Create plots only" is only possible, if a complete analysis of the defined samples with BBCAnalyzer has already been performed and the samples share at least one targeted base that has been analyzed with BBCAnalyzer.

Creating plots with "Create plots only" is usually considerably faster than performing a complete analysis with BBCAnalyzer. Its main function is to change the previously created plots (e.g. change from relative to absolute number of reads or apply different quality bounds for color-coding). However, the function may also be used to combine the output of samples that have been analyzed separately so far.

The new plots are saved in the defined output directory. Additionally, the application's progress as well as the generated plots are displayed in the right panel.

Samples to analyze	Names of the samples to analyze (one entry per line, no file name extensions).
Consider vcf file information (only possible if evaluated in complete analysis)	If vcf file information was available for every sample analyzed with BBCAnalyzer and it was considered in the previously performed complete analysis, it can also be considered for "Create plots only". Default: No.
Define output folder	Folder where previously created output files are stored (sampleX.frequency.txt and sampleX.calling.txt) and new plots shall be saved.

Define tabix file containing known variants (optional)	Tabix file (format: file.vcf.gz.tbi) containing known variants. For the correct functioning of this option the file "file.vcf.gz" is equally necessary (same folder as "file.vcf.gz.tbi"; does not have to be defined anywhere).
Lower- and upper mean quality bound for color-coding	The lower- and upper bound for the mean quality that shall be color-coded in the plots. All bases with a mean quality below the lower bound are colored with the lightest color defined for the corresponding base. All bases with a mean quality above the upper bound are colored with the darkest color defined for the corresponding base. If the bases shall not be color-coded according to their mean quality, the definable range has to be zero. Default: 58-63.
Select levels at which marks shall be drawn	Levels (relative number of reads) at which horizontal lines shall be drawn in the plots.
Plot number of reads	Relative or absolute number of reads can be plotted. Default: Relative.
Create one plot per	One plot per sample or one plot per position can be created. Default: Sample.

3 Right panel

The right panel corresponds to the output screen. The progress as well as problems in the analysis with BBCAnalyzer are displayed. Furthermore, when the analysis is complete, the created plots are displayed as well.

4 Screenshots

BBCAnalyzer

Samples to analyze

Example_454
Example_IonTorrent

Define folder containing bam- and bai-files

/mnt/home2/share/Analyses/Nijmegen_MDS_sequen

Target regions to analyze

20 31024054
21 36206893
4   106157698

Define folder containing vcf files (optional)

Define output folder

/mnt/home2/share/Analyses/Nijmegen_MDS_sequen

Define tabix file containing known variants (optional)

Select reference genome for analysis

BSgenome.Hsapiens.UCSC.hg19

Mapping quality threshold

60

Base quality threshold

50

Frequency threshold for variant reporting

0.01

Samples to analyze

Sample_1
Sample_2

Consider vcf file information (only possible if evaluated in complete analysis)

○ Yes
● No

Define output folder

./output/

Define tabix file containing known variants (optional)

Lower- and upper mean quality bound for color-coding

58 — 63

Select levels at which marks shall be drawn

□ 0.01 □ 0.02
□ 0.05 □ 0.1
□ 0.15 □ 0.2
□ 0.25 □ 0.3
□ 0.35 □ 0.4
□ 0.45 □ 0.5
□ 0.55 □ 0.6
□ 0.65 □ 0.7
□ 0.75 □ 0.8
□ 0.85 □ 0.9
□ 0.95 □ 1

Plot number of reads

● Relative
○ Absolute

Create one plot per

Complete analysis with BBCAnalyzer starts...

Determine Target
Analyze Reads
  Sample 1 out of 2
    Position 1 out of 3
    Position 2 out of 3
    Position 3 out of 3
  Sample 2 out of 2
    Position 1 out of 3
    Position 2 out of 3
    Position 3 out of 3
Analyze Frequency
  Sample 1
  Sample 2
Report Variants
  Sample 1
  Sample 2

Complete analysis with BBCAnalyzer successful!

http://127.0.0.1:6920 | Open in Browser

BSgenome.Hsapiens.UCSC.hg19

Mapping quality threshold

60

Base quality threshold

50

Frequency threshold for variant reporting

0.01

Lower- and upper mean quality bound for color-coding

58 — 63

Select levels at which marks shall be drawn

0.01 0.02
0.05 0.1
0.15 0.2
0.25 0.3
0.35 0.4
0.45 0.5
0.55 0.6
0.65 0.7
0.75 0.8
0.85 0.9
0.95 1

Plot number of reads

● Relative
○ Absolute

Create one plot per

● Sample
○ Position

Start complete analysis with BBCAnalyzer

Plot number of reads

● Relative
○ Absolute

Create one plot per

● Sample
○ Position

Create plots only

Relative frequency

Del
Ins

Reference

chr20:31024054   chr21:36206893   chr4:106157698

Example_IonTorrent

Relative frequency

58   63 Mean quality

A
C
G
T
Del
Ins

Reference

chr20:31024054   chr21:36206893   chr4:106157698