

# Package ‘DiffBind’

April 14, 2017

**Type** Package

**Version** 2.2.12

**Title** Differential Binding Analysis of ChIP-Seq peak data

**Author** Rory Stark<rory.stark@cruk.cam.ac.uk>, Gord Brown  
<gdbzork@gmail.com>

**Maintainer** Rory Stark<rory.stark@cruk.cam.ac.uk>

**Description** Compute differentially bound sites from multiple ChIP-seq experiments using affinity (quantitative) data. Also enables occupancy (overlap) analysis and plotting functions.

**License** Artistic-2.0

**LazyLoad** yes

**Depends** R (>= 3.3.0), GenomicRanges, SummarizedExperiment

**Imports** RColorBrewer, amap, edgeR, gplots, grDevices, limma,  
GenomicAlignments, locfit, stats, utils, IRanges, zlibbioc,  
lattice, systemPipeR, tools, Rcpp, dplyr, BiocParallel,  
parallel, S4Vectors, Rsamtools, DESeq2, methods

**Suggests** DESeq, BiocStyle, testthat

**Enhances** rgl, XLConnect

**LinkingTo** Rsamtools (>= 1.19.38), Rcpp

**Collate** core.R parallel.R counts.R contrast.R analyze.R io.R helper.R  
utils.R RcppExports.R cpp\_wrapper.R DBA.R

**biocViews** Sequencing, ChIPSeq, DifferentialPeakCalling

**NeedsCompilation** yes

## R topics documented:

DiffBind-package . . . . .	2
dba . . . . .	3
DBA object methods . . . . .	6
DBA tamoxifen resistance dataset . . . . .	7
dba.analyze . . . . .	8
dba.contrast . . . . .	10
dba.count . . . . .	12
dba.load . . . . .	15

dba.mask . . . . .	16
dba.overlap . . . . .	18
dba.peakset . . . . .	21
dba.plotBox . . . . .	25
dba.plotHeatmap . . . . .	28
dba.plotMA . . . . .	31
dba.plotPCA . . . . .	33
dba.plotVenn . . . . .	36
dba.report . . . . .	38
dba.save . . . . .	42
dba.show . . . . .	43
DiffBind – DBA global constant variables . . . . .	45

## Index 49

---

DiffBind-package	<i>Differential Binding Analysis of ChIP-seq peaksets</i>
------------------	---

---

### Description

Differential binding analysis of ChIP-seq peaksets

### Details

Computes differentially bound sites from multiple ChIP-seq experiments using affinity (quantitative) data. Also enables occupancy (overlap) analysis and plotting functions.

Entry Points:

<code>dba:</code>	Construct a dba object
<code>dba.peakset:</code>	Add a peakset to, or retrieve a peakset from, a dba object
<code>dba.overlap:</code>	Compute binding site overlaps and/or correlations
<code>dba.count:</code>	Count reads in binding sites
<code>dba.contrast:</code>	Establish contrast(s) for analysis
<code>dba.analyze:</code>	Execute affinity analysis
<code>dba.report:</code>	Generate report for a contrast analysis
<code>dba.plotHeatmap:</code>	Heatmap plot
<code>dba.plotPCA:</code>	Principal Components plot
<code>dba.plotBox:</code>	Boxplots
<code>dba.plotMA:</code>	MA/scatter plot
<code>dba.plotVenn:</code>	Venn diagram plot
<code>dba.show:</code>	Show dba metadata
<code>dba.mask:</code>	Mask samples or sites
<code>dba.save:</code>	Save dba object
<code>dba.load:</code>	Load dba object

**Author(s)**

Rory Stark <rory.stark @at@ cruk.cam.ac.uk> and Gord Brown <gdbzork @at@ gmail.com>

---

dba *Construct a DBA object*

---

**Description**

Constructs a new DBA object from a sample sheet, or based on an existing DBA object

**Usage**

```
dba(DBA,mask, minOverlap=2,
    sampleSheet="dba_samples.csv",
    config=data.frame(RunParallel=TRUE, reportInit="DBA", DataType=DBA_DATA_GRANGES,
                     AnalysisMethod=DBA_DESEQ2, minQCth=15, fragmentSize=125,
                     bCorPlot=FALSE, th=0.05, bUsePval=FALSE),
    peakCaller="raw", peakFormat, scoreCol, bLowerScoreBetter,
    filter, skipLines=0,
    bAddCallerConsensus=FALSE,
    bRemoveM=TRUE, bRemoveRandom=TRUE,
    bSummarizedExperiment=FALSE,
    bCorPlot, attributes)
```

**Arguments**

DBA	existing DBA object – if present, will return a fully-constructed DBA object based on the passed one, using criteria specified in the mask and/or minOverlap parameters. If missing, will create a new DBA object based on the sampleSheet.
mask	logical or numerical vector indicating which peaksets to include in the resulting model if basing DBA object on an existing one. See <a href="#">dba.mask</a> .
minOverlap	only include peaks in at least this many peaksets in the main binding matrix if basing DBA object on an existing one. If minOverlap is between zero and one, peak will be included from at least this proportion of peaksets.
sampleSheet	data frame containing sample sheet, or file name of sample sheet to load (ignored if DBA is specified). Columns names in sample sheet may include: <ul style="list-style-type: none"> <li>• SampleID: Identifier string for sample</li> <li>• Tissue: Identifier string for tissue type</li> <li>• Factor: Identifier string for factor</li> <li>• Condition: Identifier string for condition</li> <li>• Treatment: Identifier string for treatment</li> <li>• Replicate: Replicate number of sample</li> <li>• bamReads: file path for bam file containing aligned reads for ChIP sample</li> <li>• bamControl: file path for bam file containing aligned reads for control sample</li> <li>• ControlID: Identifier string for control sample</li> <li>• Peaks: path for file containing peaks for sample. format determined by PeakCaller field or caller parameter</li> </ul>

- **PeakCaller**: Identifier string for peak caller used. If Peaks is not a bed file, this will determine how the Peaks file is parsed. If missing, will use default peak caller specified in caller parameter. Possible values:
  - “raw”: text file file; peak score is in fourth column
  - “bed”: .bed file; peak score is in fifth column
  - “narrow”: default peak.format: narrowPeaks file
  - “macs”: MACS .xls file
  - “swembl”: SWEMBL .peaks file
  - “bayes”: bayesPeak file
  - “peakset”: peakset written out using pv.writepeakset
  - “fp4”: FindPeaks v4
- **PeakFormat**: string indicating format for peak files; see **PeakCaller** and [dba.peakset](#)
- **ScoreCol**: column in peak files that contains peak scores
- **LowerBetter**: logical indicating that lower scores signify better peaks
- **Counts**: file path for externally computed read counts; see [dba.peakset](#) (counts parameter)

For sample sheets loaded from a file, the accepted formats are comma-separated values (column headers, followed by one line per sample), or Excel-formatted spreadsheets (.xls or .xlsx extension). Leading and trailing white space will be removed from all values, with a warning.

config	data frame containing configuration options, or file name of config file to load when constructing a new DBA object from a sample sheet. NULL indicates no config file. Relevant fields include: <ul style="list-style-type: none"> <li>• <b>RunParallel</b>: logical indicating if counting and analysis operations should be run in parallel using multicore by default.</li> <li>• <b>DataType</b>: default class for peaks and reports (DBA_DATA_GRANGES, DBA_DATA_RANGEDDATA, or DBA_DATA_FRAME).</li> <li>• <b>ReportInit</b>: string to append to the beginning of saved report file names.</li> <li>• <b>AnalysisMethod</b>: either DBA_DESEQ2 or DBA_EDGER.</li> <li>• <b>bCorPlot</b>: logical indicating that a correlation heatmap should be plotted automatically</li> <li>• <b>th</b>: default threshold for reporting and plotting analysis results.</li> <li>• <b>bUsePval</b>: logical, default indicating whether to use FDR (FALSE) or p-values (TRUE).</li> <li>• <b>minQCth</b>: numeric, for filtering reads based on mapping quality score; only reads with a mapping quality score greater than or equal to this will be counted.</li> <li>• <b>fragmentSize</b>: numeric with mean fragment size. Reads will be extended to this length before counting overlaps. May be a vector of lengths, one for each sample.</li> </ul>
peakCaller	if a sampleSheet is specified, the default peak caller that will be used if the PeakCaller column is absent.
peakFormat	if a sampleSheet is specified, the default peak file format that will be used if the PeakFormat column is absent.
scoreCol	if a sampleSheet is specified, the default column in the peak files that will be used for scoring if the ScoreCol column is absent.

bLowerScoreBetter	if a sampleSheet is specified, the sort order for peak scores if the LowerBetter column is absent.
filter	if a sampleSheet is specified, a filter value if the Filter column is absent. Peaks with scores lower than this value (or higher if bLowerScoreBetter or LowerBetter is TRUE) will be removed.
skipLines	if a sampleSheet is specified, the number of lines (ie header lines) at the beginning of each peak file to skip.
bAddCallerConsensus	add a consensus peakset for each sample with more than one peakset (i.e. different peak callers) when constructing a new DBA object from a sample sheet.
bRemoveM	logical indicating whether to remove peaks on chrM (mitochondria) when constructing a new DBA object from a sample sheet.
bRemoveRandom	logical indicating whether to remove peaks on chrN_random when constructing a new DBA object from a sample sheet.
bSummarizedExperiment	logical indicating whether to return resulting object as a <a href="#">SummarizedExperiment</a> .
bCorPlot	logical indicating that a correlation heatmap should be plotted before returning. If DBA is NULL (a new DBA object is being created), and bCorPlot is missing, then this will take the default value (FALSE). However if DBA is NULL (a new DBA object is being created), and bCorPlot is specified, then the specified value will become the default value of bCorPlot for the resultant DBA object.
attributes	vector of attributes to use subsequently as defaults when generating labels in plotting functions: <ul style="list-style-type: none"> <li>• DBA_ID</li> <li>• DBA_TISSUE</li> <li>• DBA_FACTOR</li> <li>• DBA_CONDITION</li> <li>• DBA_REPLICATE</li> <li>• DBA_CONSENSUS</li> <li>• DBA_CALLER</li> <li>• DBA_CONTROL</li> </ul>

### Details

MODE: Construct a new DBA object from a samplesheet:

```
dba(sampleSheet, config, bAddCallerConsensus, bRemoveM, bRemoveRandom, attributes)
```

MODE: Construct a DBA object based on an existing one:

```
dba(DBA, mask, attributes)
```

MODE: Convert a DBA object to a SummarizedExperiment object:

```
dba(DBA, bSummarizedExperiment=TRUE)
```

### Value

DBA object

### Author(s)

Rory Stark and Gordon Brown

**See Also**

[dba.peakset](#), [dba.show](#)

**Examples**

```
# Create DBA object from a samplesheet
setwd(system.file("extra", package="DiffBind"))
tamoxifen <- dba(sampleSheet="tamoxifen.csv")
tamoxifen

tamoxifen <- dba(sampleSheet="tamoxifen_allfields.csv")
tamoxifen

tamoxifen <- dba(sampleSheet="tamoxifen_allfields.csv", config="config.csv")
tamoxifen

#Create a DBA object with a subset of samples
data(tamoxifen_peaks)
Responsive <- dba(tamoxifen, tamoxifen$masks$Responsive)
Responsive

# change peak caller but leave peak format the same
setwd(system.file("extra", package="DiffBind"))
tamoxifen <- dba(sampleSheet="tamoxifen.csv", peakCaller="macs",
                peakFormat="raw", scoreCol=5 )
dba.show(tamoxifen, attributes=c(DBA_TISSUE, DBA_CONDITION, DBA_REPLICATE, DBA_CALLER))

# Convert DBA object to SummarizedExperiment
data(tamoxifen_counts)
sset <- dba(tamoxifen, bSummarizedExperiment=TRUE)
sset
```

---

DBA object methods

*Standard S3 methods for DBA object*

---

**Description**

Standard S3 methods for DBA object.

**Usage**

```
## S3 method for class 'DBA'
print(x, ...)
## S3 method for class 'DBA'
summary(object, ...)
## S3 method for class 'DBA'
plot(x, ...)
```

**Arguments**

x	DBA object
object	DBA object
...	Arguments passed on to parent methods

**Details**

S3 methods for DBA object from the [DiffBind](#) package.

DBA objects are usually constructed using the [dba](#) function.

**Author(s)**

Rory Stark

**Examples**

```
data(tamoxifen_peaks)
tamoxifen
data(tamoxifen_counts)
tamoxifen
```

---

DBA tamoxifen resistance dataset

*Tamoxifen resistance dataset used for DBA examples*

---

**Description**

Tamoxifen resistance dataset used for DBA examples

**Usage**

```
data(tamoxifen_peaks)

data(tamoxifen_counts)

data(tamoxifen_analysis)
```

**Arguments**

```
tamoxifen_peaks
    load tamoxifen resistance dataset DBA object with peak (occupancy) data
tamoxifen_counts
    load tamoxifen resistance dataset DBA object with count (affinity) data
tamoxifen_analysis
    load tamoxifen resistance dataset DBA object with count (affinity) data and
    edgeR-based differential binding analysis results
```

**Details**

The tamoxifen resistance dataset is used for the DBA vignette and man page examples.

**Value**

loads a DBA object named tamoxifen

**Author(s)**

Rory Stark

**Examples**

```

data(tamoxifen_peaks)
tamoxifen
data(tamoxifen_counts)
plot(tamoxifen)
data(tamoxifen_analysis)
dba.plotMA(tamoxifen)

```

---

dba.analyze

*Perform differential binding affinity analysis*


---

**Description**

Performs differential binding affinity analysis

**Usage**

```

dba.analyze(DBA, method=DBA$config$AnalysisMethod,
            bSubControl=TRUE, bFullLibrarySize=TRUE, bTagwise=TRUE,
            filter=0, filterFun=max,
            bCorPlot=DBA$config$bCorPlot,
            bReduceObjects=TRUE,
            bParallel=DBA$config$RunParallel)

```

**Arguments**

DBA	DBA object. If no contrasts are specified (DBA\$contrast is NULL), default contrasts will be added via a call to <a href="#">dba.contrast</a> .
method	method, or vector of methods, by which to analyze differential binding affinity. Supported methods: <ul style="list-style-type: none"> <li>• <a href="#">DBA_EDGER</a></li> <li>• <a href="#">DBA_DESEQ2</a></li> </ul> also, for backward compatibility: <ul style="list-style-type: none"> <li>• <a href="#">DBA_DESEQ</a></li> </ul> Additionally, if this value is set to DBA_ALL_METHODS, this is equivalent to specifying c(DBA_EDGER, DBA_DESEQ2).
bSubControl	logical indicating whether Control read counts are subtracted for each site in each sample before performing analysis.
bFullLibrarySize	logical indicating if the full library size (total number of reads in BAM/SAM/BED file) for each sample is used for scaling normalization. If FALSE, the total number of reads present in the peaks for each sample is used (generally preferable if overall binding levels are expected to be similar between samples).
bTagwise	logical indicating if dispersion should be calculated on a tagwise (or per-condition) basis. If there are only a very few members of each group in a contrast (e.g. no replicates), this should be set to FALSE.
filter	value to use for filtering intervals with low read counts. Each contrast will be filtered separately. The filterFun will be applied to each interval, and any scores below the filter value will be removed prior to analysis.



filterFun	function that will be invoked for each interval with a vector of scores for each sample. Returns a score that will be evaluated against the filter value (only intervals with a score at least as high as filter will be kept). Default is max, indicating that at least one sample should have a score of at least filter; other useful values include sum (indicating that all the scores added together should be at least filter) and mean (setting a minimum mean normalized count level). Users can supply their own function as well.
bCorPlot	logical indicating whether to plot a correlation heatmap for the analyzed data (first contrast only). If no sites are significantly differentially bound using the default thresholds, no heatmap will be plotted.
bReduceObjects	logical indicating whether strip the analysis objects of unnecessary fields to save memory. If it is desired to use the DBA\$contrasts[[n]]\$edgeR and/or DBA\$contrasts[[n]]\$DESeq2 objects directly in the edgeR and/or DESeq2 packages, this should be set to FALSE.
bParallel	logical indicating that the analyses is to be done in parallel using multicore (one process for each contrast for each method, plus an additional process per method).

### Details

See the DBA User Guide for more details on how the edgeR and DESeq2 analyses are carried out.

### Value

DBA object with results of analysis added to DBA\$contrasts.

### Note

If there is a blocking factor for the contrast(s) specified using a previous call to [dba.contrast](#), a multi-factor analysis will automatically be carried out in addition to a single factor analysis.

### Author(s)

Rory Stark

### See Also

[dba.contrast](#), [dba.report](#)

### Examples

```
data(tamoxifen_counts)

tamoxifen <- dba.analyze(tamoxifen)
tamoxifen

data(tamoxifen_counts)
tamoxifen <- dba.contrast(tamoxifen, categories=DBA_CONDITION, block=tamoxifen$mask$MCF7)
tamoxifen <- dba.analyze(tamoxifen, method=DBA_ALL_METHODS)
tamoxifen
```

---

 dba.contrast

 Set up contrasts for differential binding affinity analysis
 

---

## Description

Sets up contrasts for differential binding affinity analysis

## Usage

```
dba.contrast(DBA, group1, group2=!group1, name1="group1", name2="group2",
             minMembers=3, block, bNot=FALSE,
             categories=c(DBA_TISSUE, DBA_FACTOR, DBA_CONDITION, DBA_TREATMENT))
```

## Arguments

DBA	DBA object with count data
group1	mask of samples in first group (when adding a specific contrast). See <a href="#">dba.mask</a> .
group2	mask of samples in second group (when adding a specific contrast). See <a href="#">dba.mask</a> .
name1	label for samples in first group (when adding a specific contrast).
name2	label for samples in second group (when adding a specific contrast).
minMembers	when automatically generating contrasts, minimum number of unique samples in a group. Must be at least 2, as replicates are strongly advised. If you wish to do an analysis with no replicates, you can set the group1 and group2 parameters explicitly.
bNot	include contrasts consisting of a group and all other samples not in that group (indicated by a ! in the contrast name).
categories	when automatically generating contrasts, attribute or vector of attributes to base contrasts on: <ul style="list-style-type: none"> <li>• DBA_ID</li> <li>• DBA_TISSUE</li> <li>• DBA_FACTOR</li> <li>• DBA_CONDITION</li> <li>• DBA_TREATMENT</li> <li>• DBA_REPLICATE</li> <li>• DBA_CALLER</li> </ul>
block	blocking attribute for multi-factor analysis. This may be specified as either a value, a vector, or a list. If block is a value, the specified metadata field is used to derive the blocking factor. One of: <ul style="list-style-type: none"> <li>• DBA_TISSUE</li> <li>• DBA_FACTOR</li> <li>• DBA_CONDITION</li> <li>• DBA_TREATMENT</li> <li>• DBA_REPLICATE</li> <li>• DBA_CALLER</li> </ul>

If block is a vector, it can either be a mask (logical vector) or a vector of peakset numbers. In this case, the peaksets indicated in the blocking vector are all given the same value (true), while any peaksets not included in the vector take the alternative value (false).

If block is a list, it should be a list of vectors (either logical masks or vectors of peakset numbers), with each indicating a set of peaksets that should share the same value. Each peakset should appear at most once, and any peaksets not specified will be given an default value (other).

### Details

MODE: Set up all possible contrasts:

```
dba.contrast(DBA, minMembers, categories)
```

MODE: Set up a specific contrast:

```
dba.contrast(DBA, group1, group2, name1, name2, block)
```

### Value

DBA object with contrast(s) set as DBA\$contrasts. Contrast list can be retrieved using dba.show(DBA, bContrasts=T).

### Note

Contrasts will only be set up for peaksets where DBA\_CALLER == "counts".

Contrasts can be cleared by DBA\$contrasts=NULL.

### Author(s)

Rory Stark

### See Also

[dba.analyze](#)

### Examples

```
data(tamoxifen_counts)
tamoxifen <- dba.contrast(tamoxifen, categories=DBA_CONDITION)
tamoxifen

# Another way to do the same thing
tamoxifen$contrasts=NULL
tamoxifen <- dba.contrast(tamoxifen, tamoxifen$mask$Responsive, tamoxifen$mask$Resistant,
                        "Responsive", "Resistant")
tamoxifen

# Add add default contrasts
tamoxifen$contrasts=NULL
tamoxifen <- dba.contrast(tamoxifen)
tamoxifen

# Specify a blocking factor
tamoxifen$contrasts=NULL
tamoxifen <- dba.contrast(tamoxifen, categories=DBA_CONDITION, block=DBA_TISSUE)
```

```

tamoxifen

tamoxifen$contrasts=NULL
tamoxifen <- dba.contrast(tamoxifen, categories=DBA_CONDITION, block=list(c(3,4,5,8,9),c(1,2,10,11)))
tamoxifen

tamoxifen$contrasts=NULL
tamoxifen <- dba.contrast(tamoxifen, categories=DBA_CONDITION, block=tamoxifen$masks$MCF7)
tamoxifen <- dba.analyze(tamoxifen)
tamoxifen

```

---

dba.count

*Count reads in binding site intervals*


---

## Description

Counts reads in binding site intervals. Files must be one of bam, bed and gzip-compressed bed. File suffixes must be ".bam", ".bed", or ".bed.gz" respectively.

## Usage

```

dba.count(DBA, peaks, minOverlap=2, score=DBA_SCORE_TMM_MINUS_FULL, bLog=FALSE,
          fragmentSize=DBA$config$fragmentSize,
          summits, filter=0, bRemoveDuplicates=FALSE, bScaleControl=TRUE,
          mapQCth=DBA$config$mapQCth,
          filterFun=max,
          bCorPlot=DBA$config$bCorPlot,
          bUseSummarizeOverlaps=FALSE, readFormat=DBA_READS_DEFAULT,
          bParallel=DBA$config$RunParallel)

```

## Arguments

DBA	DBA object
peaks	If GRanges, RangedData, dataframe, or matrix, this parameter contains the intervals to use for counting. If character string, it specifies a file containing the intervals to use (with the first three columns specifying chromosome, startpos, endpos). If missing or a mask, generates a consensus peakset using minOverlap parameter (after applying the mask if present). If NULL, the score, filter, and summits parameters are honored, updating the global binding matrix without re-counting in the cases of score and filter, and only counting after re-centering in the case of summits.
minOverlap	only include peaks in at least this many peaksets when generating consensus peakset (i.e. when peaks parameter is missing). If minOverlap is between zero and one, peak will be included from at least this proportion of peaksets.
score	which score to use in the binding affinity matrix. Note that all raw read counts are maintained for use by <a href="#">dba.analyze</a> , regardless of how this is set. One of:

DBA\_SCORE\_READS

DBA\_SCORE\_READS\_FOLD

DBA\_SCORE\_READS\_MINUS

raw read count for interval using only reads from CHIP

raw read count for interval from CHIP divided by read count for interval

raw read count for interval from CHIP minus read count for interval

DBA_SCORE_RPKM	RPKM for interval using only reads from ChIP
DBA_SCORE_RPKM_FOLD	RPKM for interval from ChIP divided by RPKM for interval from
DBA_SCORE_TMM_READS_FULL	TMM normalized (using edgeR), using ChIP read counts and Full
DBA_SCORE_TMM_READS_EFFECTIVE	TMM normalized (using edgeR), using ChIP read counts and Effec
DBA_SCORE_TMM_MINUS_FULL	TMM normalized (using edgeR), using ChIP read counts minus Co
DBA_SCORE_TMM_MINUS_EFFECTIVE	TMM normalized (using edgeR), using ChIP read counts minus Co
DBA_SCORE_TMM_READS_FULL_CPM	same as DBA_SCORE_TMM_READS_FULL, but reported in counts-per-r
DBA_SCORE_TMM_READS_EFFECTIVE_CPM	same as DBA_SCORE_TMM_READS_EFFECTIVE, but reported in count
DBA_SCORE_TMM_MINUS_FULL_CPM	same as DBA_SCORE_TMM_MINUS_FULL, but reported in counts-per-r
DBA_SCORE_TMM_MINUS_EFFECTIVE_CPM	same as DBA_SCORE_TMM_MINUS_EFFECTIVE, but reported in count
DBA_SCORE_SUMMIT	summit height (maximum read pileup value)
DBA_SCORE_SUMMIT_ADJ	summit height (maximum read pileup value), normalized to relative
DBA_SCORE_SUMMIT_POS	summit position (location of maximum read pileup)

bLog	logical indicating whether log2 of score should be used (only applies to DBA_SCORE_RPKM_FOLD and DBA_SCORE_READS_FOLD).
fragmentSize	This value will be used as the length of the reads. Each read will be extended from its endpoint along the appropriate strand by this many bases. If set to zero, the read size indicated in the BAM/BED file will be used. fragmentSize may also be a vector of values, one for each ChIP sample plus one for each unique Control library.
summits	if present, summit heights (read pileup) and locations will be calculated for each peak. The values can retrieved using <a href="#">dba.peakset</a> . The summits can also be used as a read score in the global binding matrix (see score). If the value of summits is TRUE (or 0), the summits will be calculated but the peaksets will be unaffected. If the value is greater than zero, all consensus peaks will be re-centered around a consensus summit, with the value of summits indicating how many base pairs to include upstream and downstream of the summit (so all consensus peaks will be of the same width, namely $2 * \text{summits}$ ). Note that if summits is greater than zero, the counting procedure will take twice as long, and bUseSummarizeOverlaps must be FALSE.
filter	value to use for filtering intervals with low read counts. The filterFun will be applied to the scores for each interval, and if it returns a value below the filter value, the interval will be removed from further analysis. If peaks is NULL, will remove sites from existing DBA object without recounting. If filter is a vector of values, dba.count will return a vector of the same length, indicating how many intervals will be retained for each specified filter level.
bRemoveDuplicates	logical indicating if duplicate reads (ones that map to exactly the same genomic position) should be removed. If TRUE, any location where multiple reads map will be counted as a single read. Note that if bLowMem is set, duplicates needs to have been already marked in all of the BAM files. The built-in counting code may not correctly handle certain cases when the bRemoveDuplicates parameter is set to TRUE. These cases include paired-end data and datasets where the read length may differ within a single BAM file. In these cases, see the bUseSummarizeOverlaps parameter.
bScaleControl	logical indicating if the Control reads should be scaled based on relative library sizes. If TRUE, and there are more reads in the Control library than in the ChIP library, the number of Control reads for each peak will be multiplied by a scaling factor determined by dividing the total number of reads in the ChIP library by

	the total number of reads in the Control library. If this value is not an integer, the number of Control reads for each peak will be the next highest integer.
mapQCth	for filtering by mapping quality (mapqc). Only alignments with mapping scores of at least this value will be included. Only applicable for bam files when bUseSummarizeOverlaps=FALSE (setting DBA\$config\$scanbamparam appropriately to filter on quality scores when using <a href="#">summarizeOverlaps</a> .)
filterFun	function that will be invoked for each interval with a vector of scores for each sample. Returns a score that will be evaluated against the filter value (only intervals with a score at least as high as filter will be kept). Default is max, indicating that at least one sample should have a score of at least filter; other useful values include sum (indicating that all the scores added together should be at least filter) and mean (setting a minimum mean normalized count level). Users can supply their own function as well.
bCorPlot	logical indicating whether to plot a correlation heatmap for the counted data
bUseSummarizeOverlaps	logical indicating that <a href="#">summarizeOverlaps</a> should be used for counting instead of the built-in counting code. This option is slower but uses the more standard counting function. If TRUE, all read files must be BAM (.bam extension), with associated index files (.bam.bai extension). The insertLength parameter must be absent. See notes for when the bRemoveDuplicates parameter is set to TRUE, where the built-in counting code may not correctly handle certain cases and bUseSummarizeOverlaps should be set to TRUE. Five additional parameters for <a href="#">summarizeOverlaps</a> may be specified in DBA\$config:
DBA\$config\$yieldSize	yieldSize indicating how many reads to process at one time; default is 5000000. The lower
DBA\$config\$intersectMode	mode indicating which overlap algorithm to use; default is "IntersectionNotEmpty"
DBA\$config\$singleEnd	logical indicating if reads are single end; default is TRUE
DBA\$config\$fragments	logical indicating how unmatched reads are counted; default is FALSE
DBA\$config\$scanbamparam	ScanBamParam object to pass to <a href="#">summarizeOverlaps</a> . If present, bRemoveDuplicates is
readFormat	Specify the file type of the read files, over-riding the file extension. Possible values:
DBA_READS_DEFAULT	use file extension (.bam, .bed, .bed.gz) to determine file type
DBA_READS_BAM	assume the file type is BAM, regardless of the file extension
DBA_READS_BED	assume the file type is BED (or zipped BED), regardless of the file extension.
	Note that if readFormat is anything other than DBA_READS_DEFAULT, all the read files must be of the same file type.
bParallel	if TRUE, use multicore to get counts for each read file in parallel

**Value**

DBA object with binding affinity matrix based on read count scores.

**Author(s)**

Rory Stark and Gordon Brown

**See Also**[dba.analyze](#)**Examples**

```

# These won't run unless you have the reads available in a BAM or BED file
data(tamoxifen_peaks)
## Not run: tamoxifen <- dba.count(tamoxifen)

# Count using a peakset made up of only peaks in all responsive MCF7 replicates
data(tamoxifen_peaks)
mcf7Common <- dba.overlap(tamoxifen,tamoxifen$mask$MCF7&tamoxifen$mask$Responsive)
## Not run: tamoxifen <- dba.count(tamoxifen,peaks=mcf7Common$inAll)
tamoxifen

#First make consensus peaksets from each set of replicates,
#then derive master consensus set for counting from those
data(tamoxifen_peaks)
tamoxifen <- dba.peakset(tamoxifen,consensus = -DBA_REPLICATE)
## Not run: tamoxifen <- dba.count(tamoxifen, peaks=tamoxifen$mask$Consensus)
tamoxifen

# Change binding affinity scores
data(tamoxifen_counts)
tamoxifen <- dba.count(tamoxifen,peaks=NULL,score=DBA_SCORE_READS)
dba.peakset(tamoxifen, bRetrieve=TRUE)
tamoxifen <- dba.count(tamoxifen,peaks=NULL,score=DBA_SCORE_RPKM_FOLD)
dba.peakset(tamoxifen, bRetrieve=TRUE)
tamoxifen <- dba.count(tamoxifen,peaks=NULL,score=DBA_SCORE_TMM_MINUS_FULL)
dba.peakset(tamoxifen, bRetrieve=TRUE)

# Plot effect of a range of filter values and then apply filter
data(tamoxifen_counts)
rate.max <- dba.count(tamoxifen, peaks=NULL, filter=0:250)
rate.sum <- dba.count(tamoxifen, peaks=NULL, filter=0:250,filterFun=sum)
plot(0:250,rate.max/rate.max[1],type='l',xlab="Filter Value",ylab="Proportion Retained Sites")
lines(0:250,rate.sum/rate.sum[1],col=2)
tamoxifen <- dba.count(tamoxifen,peaks=NULL,filter=125,filterFun=sum)
tamoxifen

```

---

`dba.load`*load DBA object*

---

**Description**

Reads in saved DBA object

**Usage**`dba.load(file='DBA', dir='.', pre='dba_', ext='RData')`

**Arguments**

file	main filename
dir	directory in which to save model
pre	string to pre-pend to filename
ext	file extension to use

**Value**

loaded DBA object

**Author(s)**

Rory Stark

**See Also**

[dba.save](#)

**Examples**

```
data(tamoxifen_peaks)
dba.save(tamoxifen, 'tamoxifenPeaks')
tamoxifen <- dba.load('tamoxifenPeaks')
```

---

dba.mask

*Derive a mask to define a subset of peaksets or sites for a DBA object*

---

**Description**

Derives a mask to define a subset of peaksets or sites for a DBA object.

**Usage**

```
dba.mask(DBA, attribute, value, combine='or', mask, merge='or', bApply=FALSE,
         peakset, minValue=-1)
```

**Arguments**

DBA	DBA object
attribute	when deriving a peakset mask, attribute to base mask on: <ul style="list-style-type: none"> <li>• DBA_ID</li> <li>• DBA_TISSUE</li> <li>• DBA_FACTOR</li> <li>• DBA_CONDITION</li> <li>• DBA_TREATMENT</li> <li>• DBA_REPLICATE</li> <li>• DBA_CONSENSUS</li> <li>• DBA_CALLER</li> <li>• DBA_CONTROL</li> </ul>



value	when deriving a peakset/sample mask, attribute value (or vector of attribute values) to match.
combine	when deriving a peakset/sample mask, if value is a vector, OR when deriving a site mask, and peaksets is a vector, this is method for combining result of each value: <ul style="list-style-type: none"> <li>• “or”</li> <li>• “and”</li> <li>• “nor”</li> <li>• “nand”</li> </ul>
mask	when deriving a peakset/sample mask, this specifies an existing mask to merge with; if missing, create new mask
merge	when deriving a peakset/sample mask, and an existing mask is supplied, this specifies the method for combining new mask with supplied mask: <ul style="list-style-type: none"> <li>• “or”</li> <li>• “and”</li> <li>• “nor”</li> <li>• “nand” note: if mask is missing, “nand” results in negative of mask</li> </ul>
bApply	when deriving a peakset/sample mask, a logical indicating that a new DBA object with the mask applied will be returned.
peakset	when deriving a peak/site mask, this specifies a peakset number, or a vector of peakset numbers. The resulting mask will indicate which of the overall sites were called as peaks in this peakset or set of peaksets. If a vector, the masks for each of the peaksets will be combined using the method specified in the combine parameter.
minValue	when deriving a peak/site mask, scores greater than this value will be considered as indicating that the site corresponds to a called peakset.

### Details

MODE: Derive a a mask of peaksets/samples:

dba.mask(DBA, attribute, value, combine, mask, merge, bApply)

MODE: Derive a mask of peaks/sites:

dba.mask(DBA, combine, mask, merge,bApply, peakset, minValue)

### Value

either a logical mask, or new DBA object if bApply is TRUE.

### Note

dba automatically generates masks for each unique value of DBA\_TISSUE, DBA\_FACTOR, DBA\_CONDITION, DBA\_TREATMENT, DBA\_CALLER, and DBA\_REPLICATE. These are accessible using masks field of the DBA object (DBA\$masks), and can be viewed using names(DBA\$masks).

### Author(s)

Rory Stark

**See Also**[dba.show](#)**Examples**

```

data(tamoxifen_peaks)

# Pre-made masks
names(tamoxifen$masks)
dba.show(tamoxifen, tamoxifen$masks$MCF7)

# New masks
mcf7Mask <- dba.mask(tamoxifen, DBA_TISSUE, "MCF7")
mcf7DerivedMask <- dba.mask(tamoxifen, DBA_TISSUE, "TAMR", mask=mcf7Mask)
mcf7Derived <- dba(tamoxifen, mcf7DerivedMask)
mcf7Derived

```

dba.overlap

*Compute binding site overlaps (occupancy analysis)***Description**

Computes binding overlaps and co-occupancy statistics

**Usage**

```

dba.overlap(DBA, mask, mode=DBA_OLAP_PEAKEs,
            contrast, method=DBA$config$AnalysisMethod, th=DBA$config$th,
            bUsePval=DBA$config$bUsePval,
            report, byAttribute, bCorOnly=TRUE, CorMethod="pearson",
            DataType=DBA$config$DataType)

```

**Arguments**

DBA	DBA object
mask	mask or vector of peakset numbers indicating a subset of peaksets to use (see <a href="#">dba.mask</a> ). When generating overlapping/unique peaksets, either two, three, or four peaksets may be specified. If the mode type is DBA_OLAP_ALL, and a contrast is specified, a value of TRUE (mask=TRUE) indicates that all samples should be included (otherwise only those present in one of the contrast groups will be included).
mode	indicates which results should be returned (see MODES below). One of: <ul style="list-style-type: none"> <li>• <a href="#">DBA_OLAP_PEAKEs</a></li> <li>• <a href="#">DBA_OLAP_ALL</a></li> <li>• <a href="#">DBA_OLAP_RATE</a></li> </ul>
contrast	contrast number to use. Only specified if contrast data is to be used when mode=DBA_OLAP_ALL. See <a href="#">dba.show</a> (DBA, bContrast=T) to get contrast numbers.
method	if contrast is specified and mode=DBA_OLAP_ALL, use data from method used for analysis:

	<ul style="list-style-type: none"> <li>• <a href="#">DBA_DESEQ2</a></li> <li>• <a href="#">DBA_DESEQ2_BLOCK</a></li> <li>• <a href="#">DBA_EDGER</a></li> <li>• <a href="#">DBA_EDGER_BLOCK</a></li> </ul>
th	if contrast is specified and mode= <code>DBA_OLAP_ALL</code> , significance threshold; all sites with FDR (or p-values, see <code>bUsePval</code> ) less than or equal to this value will be included. A value of 1 will include all binding sites, but only the samples included in the contrast.
bUsePval	if contrast is specified and mode= <code>DBA_OLAP_ALL</code> , logical indicating whether to use FDR ( <code>FALSE</code> ) or p-value ( <code>TRUE</code> ) for thresholding.
report	if contrast is specified and mode= <code>DBA_OLAP_ALL</code> , a report (obtained from <a href="#">dba.report</a> ) specifying the data to be used. If counts are included in the report (and a contrast is specified), the count data from the report will be used to compute correlations, rather than the scores in the global binding affinity matrix. If report is present, the method, th, and <code>bUsePval</code> parameters are ignored.
byAttribute	when computing co-occupancy statistics ( <code>DBA_OLAP_ALL</code> ), limit comparisons to peaksets with the same value for a specific attribute, one of: <ul style="list-style-type: none"> <li>• <a href="#">DBA_ID</a></li> <li>• <a href="#">DBA_TISSUE</a></li> <li>• <a href="#">DBA_FACTOR</a></li> <li>• <a href="#">DBA_CONDITION</a></li> <li>• <a href="#">DBA_TREATMENT</a></li> <li>• <a href="#">DBA_REPLICATE</a></li> <li>• <a href="#">DBA_CONSENSUS</a></li> <li>• <a href="#">DBA_CALLER</a></li> </ul>
bCorOnly	when computing co-occupancy statistics ( <code>DBA_OLAP_ALL</code> ), logical indicating that only correlations, and not overlaps, should be computed. This is much faster if only correlations are desired (e.g. to plot the correlations using <a href="#">dba.plotHeatmap</a> ).
CorMethod	when computing co-occupancy statistics ( <code>DBA_OLAP_ALL</code> ), method to use when computing correlations.
DataType	if mode= <code>DBA_OLAP_PEAKEKS</code> , the class of object that peaksets should be returned as: <ul style="list-style-type: none"> <li>• <a href="#">DBA_DATA_GRANGES</a></li> <li>• <a href="#">DBA_DATA_RANGEDDATA</a></li> <li>• <a href="#">DBA_DATA_FRAME</a></li> </ul> <p>Can be set as default behavior by setting <code>DBA\$config\$DataType</code>.</p>

## Details

MODE: Generate overlapping/unique peaksets:

```
dba.overlap(DBA, mask, mode=DBA_OLAP_PEAKEKS, minVal)
```

MODE: Compute correlation and co-occupancy statistics (e.g. for [dba.plotHeatmap](#)):

```
dba.overlap(DBA, mask, mode=DBA_OLAP_ALL, byAttribute, minVal, attributes, bCorOnly, CorMethod)
```

MODE: Compute correlation and co-occupancy statistics using significantly differentially bound sites (e.g. for [dba.plotHeatmap](#)):

```
dba.overlap(DBA, mask, mode=DBA_OLAP_ALL, byAttribute, minVal, contrast, method, th=, bUsePval, attributes, bCorOnly, CorMethod)
```

Note that the scores from the global binding affinity matrix will be used for correlations unless a report containing count data is specified.

MODE: Compute overlap rates at different stringency thresholds:

dba.overlap(DBA, mask, mode=DBA\_OLAP\_RATE, minVal)

### Value

Value depends on the mode specified in the mode parameter.

If mode = DBA\_OLAP\_PEAKEs, Value is an overlap record: a list of three peaksets for an A-B overlap, seven peaksets for a A-B-C overlap, and fifteen peaksets for a A-B-C-D overlap:

inAll	peaks in all peaksets
onlyA	peaks unique to peakset A
onlyB	peaks unique to peakset B
onlyC	peaks unique to peakset C
onlyD	peaks unique to peakset D
notA	peaks in all peaksets except peakset A
notB	peaks in all peaksets except peakset B
notC	peaks in all peaksets except peakset C
notD	peaks in all peaksets except peakset D
AandB	peaks in peaksets A and B but not in peaksets C or D
AandC	peaks in peaksets A and C but not in peaksets B or D
AandD	peaks in peaksets A and D but not in peaksets B or C
BandC	peaks in peaksets B and C but not in peaksets A or D
BandD	peaks in peaksets B and D but not in peaksets A or C
CandD	peaks in peaksets C and D but not in peaksets A or B

If mode = DBA\_OLAP\_ALL, Value is a correlation record: a matrix with a row for each pair of peaksets and the following columns:

A	peakset number of first peakset in overlap
B	peakset number of second peakset in overlap
onlyA	number of sites unique to peakset A
onlyB	number of sites unique to peakset B
inAll	number of peaks in both peakset A and B (merged)
R2	correlation value A vs B
Overlap	percentage overlap (number of overlapping sites divided by number of peaks unique to smaller peakset)

If mode = DBA\_OLAP\_RATE, Value is a vector whose length is the number of peaksets, containing the number of overlapping peaks at the corresponding minOverlaps threshold (i.e., Value[1] is the total number of unique sites, Value[2] is the number of unique sites appearing in at least two peaksets, Value[3] the number of sites overlapping in at least three peaksets, etc.).

### Author(s)

Rory Stark

**See Also**

[dba.plotVenn](#), [dba.plotHeatmap](#)

**Examples**

```
data(tamoxifen_peaks)
# default mode: DBA_OLAP_PEAKEs -- get overlapping/non overlapping peaksets
mcf7 <- dba.overlap(tamoxifen,tamoxifen$masks$MCF7&tamoxifen$masks$Responsive)
names(mcf7)
mcf7$inAll

# mode: DBA_OLAP_ALL -- get correlation record
mcf7 <- dba(tamoxifen,tamoxifen$masks$MCF7)
mcf7.corRec <- dba.overlap(mcf7,mode=DBA_OLAP_ALL,bCorOnly=FALSE)
mcf7.corRec

# mode: DBA_OLAP_RATE -- get overlap rate vector
data(tamoxifen_peaks)
rate <- dba.overlap(tamoxifen, mode=DBA_OLAP_RATE)
rate
plot(rate,type='b',xlab="# peaksets",ylab="# common peaks",
      main="Tamoxifen dataset overlap rate")
```

---

dba.peakset

*Add a peakset to, or retrieve a peakset from, a DBA object*

---

**Description**

Adds a peakset to, or retrieves a peakset from, a DBA object

**Usage**

```
dba.peakset(DBA=NULL, peaks, sampID, tissue, factor, condition, treatment, replicate,
            control, peak.caller, peak.format, reads=0, consensus=FALSE,
            bamReads, bamControl,
            scoreCol, bLowerScoreBetter, filter, counts,
            bRemoveM=TRUE, bRemoveRandom=TRUE,
            minOverlap=2, bMerge=TRUE,
            bRetrieve=FALSE, writeFile, numCols=4,
            DataType=DBA$config$DataType)
```

**Arguments**

DBA	DBA object. Required unless creating a new DBA object by adding an initial peakset.
peaks	When adding a specified peakset: set of peaks, either a GRanges or RangedData object, or a peak dataframe or matrix (chr,start,end,score), or a filename where the peaks are stored. When adding a consensus peakset: a sample mask or vector of peakset numbers to include in the consensus. If missing or NULL, a consensus is derived from all peaksets present in the model. See <a href="#">dba.mask</a> , or <a href="#">dba.show</a> to get peakset numbers.

	<p>When adding an empty peakset (zero peaks), set peaks=NA.</p> <p>When adding a set of consensus peaksets: a sample mask or vector of peakset numbers. Sample sets will be derived only from subsets of these peaksets.</p> <p>When adding all the peaks from one DBA object to another: a DBA object. In this case, the only other parameter to have an effect is minOverlap.</p> <p>When retrieving and/or writing a peakset: either a GRanges or RangedData object, or a peak dataframe or matrix (chr,start,end,score), or a peakset number; if NULL, retrieves/writes the full binding matrix.</p>
sampID	ID string for the peakset being added; if missing, one is assigned (a serial number for a new peakset, or a concatenation of IDs for a consensus peakset).
tissue	tissue name for the peakset being added; if missing, one is assigned for a consensus peakset (a concatenation of tissues).
factor	factor name for the peakset being added; if missing, one is assigned for a consensus peakset (a concatenation of factors).
condition	condition name for the peakset being added; if missing, one is assigned for a consensus peakset (a concatenation of conditions).
treatment	treatment name for the peakset being added; if missing, one is assigned for a consensus peakset (a concatenation of treatment).
replicate	replicate number for the peakset being added; if missing, one is assigned for a consensus peakset (a concatenation of replicate numbers).
control	control name for the peakset being added; if missing, one is assigned for a consensus peakset (a concatenation of control names).
peak.caller	<p>peak caller name string. If peaks is specified as a file, and peak.format is missing, a default file format for the caller will be used (see peak.format). Supported values:</p> <ul style="list-style-type: none"> <li>• “raw”: default peak.format: raw text file</li> <li>• “bed”: default peak.format: bed file</li> <li>• “narrow”: default peak.format: narrowPeaks file</li> <li>• “macs”: default peak.format: MACS .xls file</li> <li>• “bayes”: default peak.format: bayesPeak file</li> <li>• “tpic”: default peak.format: TPIC file</li> <li>• “sicer”: default peak.format: SICER file</li> <li>• “fp4”: default peak.format: FindPeaks v4 file</li> <li>• “swembl”: default peak.format: SWEMBL file</li> <li>• “csv”: default peak.format: comma separated value file</li> <li>• “report”: default peak.format: csv file saved via <a href="#">dba.report</a></li> </ul> <p>When adding a consensus peakset, a default value (a concatenation of peak caller names) is assigned if this is missing.</p>
peak.format	<p>peak format string. If specified, overrides the default file format for the specified peak caller. Supported formats (with default score column):</p> <ul style="list-style-type: none"> <li>• “raw”: raw text file file; scoreCol=4</li> <li>• “bed”: bed file; scoreCol=5</li> <li>• “narrow”: narrowPeaks file; scoreCol=8</li> <li>• “macs”: MACS .xls file; scoreCol=7</li> <li>• “bayes”: bayesPeak file; scoreCol=4, filter=0.5</li> <li>• “tpic”: TPIC file; scoreCol=0 (all scores=1)</li> </ul>

	<ul style="list-style-type: none"> <li>• “sicer”: SICER file; scoreCol=7</li> <li>• “fp4”: FindPeaks v4 file; scoreCol=5</li> <li>• “swembl”: SWEMBL file; scoreCol=4</li> <li>• “csv”: csv file; scoreCol=4</li> <li>• “report”: report file; scoreCol=9, bLowerScoreBetter=T</li> </ul>
reads	total number of ChIPed library reads for the peakset being added.
consensus	either the logical value of the consensus attribute when adding a specific peakset (set to TRUE for consensus peaksets generated by <a href="#">dba.peakset</a> ), or a metadata attribute or vector of attributes when generating a set of consensus peaksets. In the latter case, a consensus peakset will be added for each set of samples that have the same values for the specified attributes. Alternatively, attributes may be specified preceded by a negative sign, in which case a consensus peakset will be added for each set of samples that differ only in their values for those attributes. See examples. Allowable attributes: <ul style="list-style-type: none"> <li>• DBA_TISSUE; -DBA_TISSUE</li> <li>• DBA_FACTOR; -DBA_FACTOR</li> <li>• DBA_CONDITION; -DBA_CONDITION</li> <li>• DBA_TREATMENT; -DBA_TREATMENT</li> <li>• DBA_REPLICATE; -DBA_REPLICATE</li> <li>• DBA_CALLER; -DBA_CALLER</li> </ul>
bamReads	file path of the BAM/BED file containing the aligned reads for the peakset being added.
bamControl	file path of the BAM/BED file containing the aligned reads for the control used for the peakset being added.
scoreCol	peak column to normalize to 0...1 scale when adding a peakset; 0 indicates no normalization
bLowerScoreBetter	Logical indicating that lower scores indicate higher confidence peaks; default is that higher scores indicate better peaks.
filter	Numeric indicating a filter value for peaks. If present, any peaks with a score less than this value (or higher if bLowerScoreBetter==TRUE) will be removed from the peakset.
counts	Used for adding externally computed peak counts. Can be a filename or a dataframe. Can consist of a single column (or vector) with the counts, or two columns, with an ID for each interval in the first column and the counts in the second column, or four columns (chr, start, end, counts). When counts is specified, peaks and related parameters are ignored, and all peaksets in the DBA object must be specified in this way, all with exactly the same number of intervals.
bRemoveM	logical indicating whether to remove peaks on chrM when adding a peakset
bRemoveRandom	logical indicating whether to remove peaks on chrN_random when adding a peakset
minOverlap	the minimum number of peaksets a peak must be in to be included when adding a consensus peakset. When retrieving, if the peaks parameter is a vector (logical mask or vector of peakset numbers), a binding matrix will be retrieved including all peaks in at least this many peaksets. If minOverlap is between zero and one, peak will be included from at least this proportion of peaksets.

bMerge	logical indicating whether global binding matrix should be compiled after adding the peakset. When adding several peaksets via successive calls to <code>dba.peakset</code> , it may be more efficient to set this parameter to FALSE and call <code>dba(DBA)</code> after all the peaksets have been added.
bRetrieve	logical indicating that a peakset is being retrieved and/or written, not added.
writeFile	file to write retrieved peakset.
numCols	number of columns to include when writing out peakset. First four columns are chr, start, end, score; the remainder are maintained from the original peakset. Ignored when writing out complete binding matrix.
DataType	The class of object for returned peaksets: <ul style="list-style-type: none"> <li>• DBA_DATA_GRANGES</li> <li>• DBA_DATA_RANGEDDATA</li> <li>• DBA_DATA_FRAME</li> </ul> <p>Can be set as default behavior by setting <code>DBA\$config\$DataType</code>.</p>

## Details

MODE: Add a specified peakset:

`dba.peakset(DBA=NULL, peaks, sampID, tissue, factor, condition, replicate, control, peak.caller, reads, consensus, bamReads, bamControl, normCol, bRemoveM, bRemoveRandom)`

MODE: Add a consensus peakset (derived from overlapping peaks in peaksets already present):

`dba.peakset(DBA, peaks, minOverlap)`

MODE: Add a sets of consensus peaksets bases on sample sets that share or differ in specified attributes

`dba.peakset(DBA, peaks, consensus, minOverlap)`

MODE: Retrieve a peakset:

`dba.peakset(DBA, peaks, bRetrieve=T)`

MODE: Write a peakset out to a file:

`dba.peakset(DBA, peaks, bRetrieve=T, writeFile, numCols)`

## Value

DBA object when adding a peakset. Peakset matrix or RangedData object when retrieving and/or writing a peakset.

## Author(s)

Rory Stark

## See Also

to add peaksets using a sample sheet, see [dba](#).



**Examples**

```

# create a new DBA object by adding three peaksets
mcf7 <- dba.peakset(NULL,
  peaks=system.file("extra/peaks/MCF7_ER_1.bed.gz", package="DiffBind"),
  peak.caller="bed", sampID="MCF7.1",tissue="MCF7",
  factor="ER",condition="Responsive",replicate=1)
mcf7 <- dba.peakset(mcf7,
  peaks=system.file("extra/peaks/MCF7_ER_2.bed.gz", package="DiffBind"),
  peak.caller="bed", sampID="MCF7.2",tissue="MCF7",
  factor="ER",condition="Responsive",replicate=2)
mcf7 <- dba.peakset(mcf7,
  peaks=system.file("extra/peaks/MCF7_ER_3.bed.gz", package="DiffBind"),
  peak.caller="bed", sampID="MCF7.3",tissue="MCF7",
  factor="ER",condition="Responsive",replicate=3)

mcf7

#retrieve peaks that are in all three peaksets
mcf7.consensus <- dba.peakset(mcf7, 1:3, minOverlap=3, bRetrieve=TRUE)
mcf7.consensus

#add a consensus peakset -- peaks in all three replicates
mcf7 <- dba.peakset(mcf7, 1:3, minOverlap=3,sampID="MCF7_3of3")
mcf7

#add consensus peaksets for all sample types by combining replicates
data(tamoxifen_peaks)
tamoxifen <- dba.peakset(tamoxifen,consensus = -DBA_REPLICATE)
dba.show(tamoxifen,mask=tamoxifen$masks$Consensus)

#add consensus peaksets for all sample types by (same tissue and condition)
data(tamoxifen_peaks)
tamoxifen <- dba.peakset(tamoxifen,consensus = c(DBA_TISSUE,DBA_CONDITION))
dba.show(tamoxifen,mask=tamoxifen$masks$Consensus)
dba.plotVenn(tamoxifen,tamoxifen$masks$Responsive & tamoxifen$masks$Consensus)

#create consensus peaksets from sample type consensus for Responsive and Resistant sample groups
tamoxifen <- dba.peakset(tamoxifen,peaks=tamoxifen$masks$Consensus,consensus=DBA_CONDITION)
dba.show(tamoxifen,mask=tamoxifen$masks$Consensus)
dba.plotVenn(tamoxifen,17:18)

#retrieve the consensus peakset as RangedData object
mcf7.consensus <- dba.peakset(mcf7,mcf7$masks$Consensus,bRetrieve=TRUE)
mcf7.consensus

```

---

dba.plotBox

*Boxplots*


---

**Description**

Boxplots for read count distributions within differentially bound sites

**Usage**

```
dba.plotBox(DBA, contrast=1, method=DBA$config$AnalysisMethod,
            th=DBA$config$th, bUsePval=DBA$config$bUsePval,
            bNormalized=TRUE, attribute=DBA_GROUP,
            bAll=FALSE, bAllIncreased=FALSE, bAllDecreased=FALSE,
            bDB=TRUE, bDBIncreased=TRUE, bDBDecreased=TRUE,
            pvalMethod=wilcox.test, bReversePos=FALSE, attribOrder,
            vColors, varwidth=TRUE, notch=TRUE, ...)
```

**Arguments**

DBA	DBA object.
contrast	number of contrast to use for boxplot.
method	method used for analysis (used in conjunction with contrast): <ul style="list-style-type: none"> <li>• <a href="#">DBA_DESEQ2</a></li> <li>• <a href="#">DBA_DESEQ2_BLOCK</a></li> <li>• <a href="#">DBA_EDGER</a></li> <li>• <a href="#">DBA_EDGER_BLOCK</a></li> </ul>
th	significance threshold; all sites with FDR (or p-values, see bUsePval) less than or equal to this value will be included in the boxplot.
bUsePval	logical indicating whether to use FDR (FALSE) or p-value (TRUE) for thresholding.
bNormalized	logical indicating that normalized data (using normalization factors computed by differential analysis method) should be plotted. FALSE uses raw count data.
attribute	attribute to use for determining groups of samples. Default (DBA_GROUP) plots the two groups used in the contrast. Possible values: <ul style="list-style-type: none"> <li>• <a href="#">DBA_GROUP</a></li> <li>• <a href="#">DBA_ID</a></li> <li>• <a href="#">DBA_TISSUE</a></li> <li>• <a href="#">DBA_FACTOR</a></li> <li>• <a href="#">DBA_CONDITION</a></li> <li>• <a href="#">DBA_TREATMENT</a></li> <li>• <a href="#">DBA_REPLICATE</a></li> <li>• <a href="#">DBA_CONSENSUS</a></li> <li>• <a href="#">DBA_CALLER</a></li> </ul>
bAll	logical indicating if plot should include a set of boxplots using all counts, regardless of whether or not they pass the significance threshold.
bAllIncreased	logical indicating if plot should include a set of boxplots using all counts that increase in affinity, regardless of whether or not they pass the significance threshold.
bAllDecreased	logical indicating if plot should include a set of boxplots using all counts that decrease in affinity, regardless of whether or not they pass the significance threshold.
bDB	logical indicating if plot should include a set of boxplots using all counts in significantly differentially bound sites (i.e. those that pass the significance threshold), regardless of whether they increase or decrease in affinity.

bDBIncreased	logical indicating if plot should include a set of boxplots using all counts in significantly differentially bound sites that increase in affinity.
bDBDecreased	logical indicating if plot should include a set of boxplots using all counts in significantly differentially bound sites that decrease in affinity.
pvalMethod	method to use when computing matrix of p-values. If NULL, no matrix is computed, and NULL is returned; this may speed up processing if there are many boxplots.
bReversePos	logical indicating if the default definition of positive affinity (higher affinity in the second group of the contrast) should be reversed (i.e. positive affinity is defined as being higher in the first group of the contrast).
attribOrder	vector of group numbers used to change the order that groups are plotted. If NULL, default order is used (group order for DBA_GROUP, and the order the attribute values appear for other values of attribute).
vColors	vector of custom colors; if absent, default colors will be used.
varwidth	passed to boxplot
notch	passed to boxplot
...	other arguments passed to boxplot

### Details

Draws a boxplot showing distributions of read counts for various groups of samples under various conditions. In default mode, draws six boxes: one pair of boxes showing the distribution of read counts within all significantly differentially bound sites (one box for each sample group), one pair of boxes showing the distribution of read counts for significantly differentially bound sites that increase affinity in the second group, and a second pair of boxes showing the distribution of read counts for significantly differentially bound sites that have higher mean affinity in the first group.

### Value

if pvalMethod is not NULL, returns a matrix of p-values indicating the significance of the difference between each pair of distributions.

### Author(s)

Rory Stark

### Examples

```
data(tamoxifen_analysis)

#default boxplot includes all DB sites, then divided into those increasing
# affinity in each group
dba.plotBox(tamoxifen)

# plot non-normalized data for DB sites by tissue
# (changing order to place Resistant samples last)
dba.plotBox(tamoxifen, attribute=DBA_CONDITION, bDBIncreased=FALSE,
            bDBDecreased=FALSE, attribOrder=c(2,1), bNormalized=FALSE)
```

---

dba.plotHeatmap	<i>Draw a binding site heatmap</i>
-----------------	------------------------------------

---

### Description

Draws a binding site heatmap

### Usage

```
dba.plotHeatmap(DBA, attributes=DBA$attributes, maxSites=1000, minval, maxval,
  contrast, method=DBA$config$AnalysisMethod,
  th=DBA$config$th, bUsePval=DBA$config$bUsePval,
  report, score, bLog=TRUE, mask, sites, sortFun,
  correlations=TRUE, olPlot=DBA_COR,
  ColAttributes, RowAttributes, colSideCols, rowSideCols=colSideCols,
  margin=10, colScheme="Greens", distMethod="pearson",
  ...)
```

### Arguments

DBA	DBA object.
attributes	attribute or vector of attributes to use for column labels: <ul style="list-style-type: none"> <li>• <a href="#">DBA_ID</a></li> <li>• <a href="#">DBA_TISSUE</a></li> <li>• <a href="#">DBA_FACTOR</a></li> <li>• <a href="#">DBA_CONDITION</a></li> <li>• <a href="#">DBA_TREATMENT</a></li> <li>• <a href="#">DBA_REPLICATE</a></li> <li>• <a href="#">DBA_CONSENSUS</a></li> <li>• <a href="#">DBA_CALLER</a></li> </ul>
maxSites	maximum number of binding sites to use in heatmap. Only used when not drawing a correlation heatmap (correlations=FALSE)
minval	Set all scores less than this to minval
maxval	Set all scores greater than this to maxval
contrast	number of contrast to report on; if present, draws a heatmap based on a differential binding affinity analysis (see <a href="#">dba.analyze</a> ). Only significantly differentially bound sites will be used (subject to the th and bUsePval parameters). If mask is unspecified, only the samples in the contrast will be included. See <a href="#">dba.show</a> (DBA, bContrast=T) to get contrast numbers. If missing, uses scores in the main binding matrix.
method	analysis method (used in conjunction with contrast): <ul style="list-style-type: none"> <li>• <a href="#">DBA_DESEQ2</a></li> <li>• <a href="#">DBA_DESEQ2_BLOCK</a></li> <li>• <a href="#">DBA_EDGER</a></li> <li>• <a href="#">DBA_EDGER_BLOCK</a></li> </ul>

th	significance threshold; all sites with FDR (or p-values, see bUsePval) less than or equal to this value will be included in the report (subject to maxSites). Used in conjunction with contrast.
bUsePval	logical indicating whether to use FDR (FALSE) or p-value (TRUE) for thresholding. Used in conjunction with contrast.
report	report (obtained from <a href="#">dba.report</a> specifying the data to be used). If this is present, the method, th, and bUsePval parameters are ignored. Used in conjunction with contrast.
score	Score to use for count data. Only used when plotting the global binding matrix (no contrast specified). One of: <ul style="list-style-type: none"> <li>• <a href="#">DBA_SCORE_READS</a></li> <li>• <a href="#">DBA_SCORE_READS_MINUS</a></li> <li>• <a href="#">DBA_SCORE_READS_FOLD</a></li> <li>• <a href="#">DBA_SCORE_RPKM</a></li> <li>• <a href="#">DBA_SCORE_RPKM_FOLD</a></li> <li>• <a href="#">DBA_SCORE_TMM_READS_FULL</a></li> <li>• <a href="#">DBA_SCORE_TMM_READS_EFFECTIVE</a></li> <li>• <a href="#">DBA_SCORE_TMM_MINUS_FULL</a></li> <li>• <a href="#">DBA_SCORE_TMM_MINUS_EFFECTIVE</a></li> </ul>
bLog	Logical indicating that log2 values should be used. Only applicable to read count scores (not peak scores).
mask	mask indicating a subset of peaksets to use when using global binding matrix scores. If a contrast is specified, these peaksets will be included, but only the significantly differentially bound sites (using th, bUsePval, and/or report) will be included.
sites	logical vector indicating which sites to include; first maxSites of these. Only relevant when using global binding matrix (contrast is missing).
sortFun	function taking a vector of scores and returning a single value. Only relevant when using global binding matrix (contrast is missing). If present, the global binding matrix will be sorted (descending) on the results, and the first maxSites used in the heatmap. Recommended sort function options include sd, mean, median, min.
correlations	logical indicating that a correlation heatmap should be plotted (TRUE). If FALSE, a binding heatmap of scores/reads is plotted. This parameter can also be set to a correlation record; see <a href="#">dba.overlap(mode=DBA_OLAP_ALL)</a> , in which case a correlation heatmap is plotted based on the specified correlation record, using the statistic specified in olPlot.
olPlot	if correlations is specified as a dataframe returned by <a href="#">dba.overlap</a> , indicates which statistic to plot. One of: <ul style="list-style-type: none"> <li>• <a href="#">DBA_COR</a> Correlation</li> <li>• <a href="#">DBA_OLAP</a> Percentage overlap</li> <li>• <a href="#">DBA_INALL</a> number of peaks common to both samples</li> </ul>
ColAttributes	Attribute or vector of attributes to plot for column color bars. If missing, all attributes with two or more unique non-NA values will be plotted. (For correlation heatmaps, DBA_GROUP will be plotted in the column color bar by default when a contrast is specified). A value of NULL indicates that no column color bar should be drawn. Allowable attribute values include:

- [DBA\\_GROUP](#)
- [DBA\\_TISSUE](#)
- [DBA\\_FACTOR](#)
- [DBA\\_CONDITION](#)
- [DBA\\_TREATMENT](#)
- [DBA\\_REPLICATE](#)
- [DBA\\_CALLER](#)

RowAttributes	Attribute or vector of attributes for row color bars. Row color bars are only allowed for correlation heatmaps. Same values as for ColAttributes parameter. Default is to draw a row color bar only if a contrast is specified, in which case the plotted attribute is DBA_GROUP.
rowSideCols	Vector of colors to use in row color bars. Uses default colors if missing. Can also be a list of color vectors.
colSideCols	Vector of colors to use in column color bars. Uses default colors if missing. Can also be a list of color vectors.
margin	margin size of plot
colScheme	Color scheme; see colorRampPalette RColorBrewer
distMethod	distance method for clustering; see <a href="#">Dist</a>
...	passed on to heatmap.2 (gplots), e.g. scale etc.

### Details

MODE: Correlation Heatmap plot using statistics for global binding matrix:

```
dba.plotHeatmap(DBA, attributes=DBA$attributes, minval, maxval, correlations, olPlot, colScheme="Greens",
distMethod="pearson", ...)
```

MODE: Correlation Heatmap plot using statistics for significantly differentially bound sites:

```
dba.plotHeatmap(DBA, attributes=DBA$attributes, minval, maxval, contrast, method=DBA_DESEQ2,
th=0.05, bUsePval=F, mask, overlaps, olPlot=DBA_COR, colScheme="Greens", distMethod="pearson",
...)
```

MODE: Binding heatmap plot using significantly differentially bound sites:

```
dba.plotHeatmap(DBA, attributes, maxSites, minval, maxval, contrast, method, th, bUsePval, cor-
relations=FALSE, colScheme, distMethod, ...)
```

MODE: Binding heatmap plot using the global binding matrix:

```
dba.plotHeatmap(DBA, attributes, maxSites, minval, maxval, mask, sites, correlations=FALSE,
sortFun, colScheme, distMethod, ...)
```

### Value

if correlations is not FALSE, the overlap/correlation matrix is returned.

if correlations is FALSE, the sites used in the heatmap are returned in a [GRanges](#) object, in the row order they appear (top to bottom). The metadata contains a column for each sample (also in the order they appear in the clustering plot), with the values being the actual plotted values.

### Author(s)

Rory Stark

**See Also**[dba.overlap](#)**Examples**

```

data(tamoxifen_peaks)
# peak overlap correlation heatmap
dba.plotHeatmap(tamoxifen)

data(tamoxifen_counts)
# counts correlation heatmap
dba.plotHeatmap(tamoxifen)

data(tamoxifen_analysis)
#correlation heatmap based on all normalized data
dba.plotHeatmap(tamoxifen,contrast=1,th=1)

#correlation heatmap based on DB sites only
dba.plotHeatmap(tamoxifen,contrast=1)

#binding heatmap based on DB sites
dba.plotHeatmap(tamoxifen,contrast=1,correlations=FALSE)

#binding heatmap based on 1,000 sites with highest variance
sites <- dba.plotHeatmap(tamoxifen,contrast=1,th=1,
                        correlations=FALSE,sortFun=var)
sites

data(tamoxifen_counts)
#Examples of heatmaps using DB sites with different subsets of samples
#exclude T47D
tamoxifen <- dba.contrast(tamoxifen,tamoxifen$masks$Resistant,c(3:5,10:11))
tamoxifen <- dba.analyze(tamoxifen,bCorPlot=FALSE)
# regular heatmaps with two contrast groups
dba.plotHeatmap(tamoxifen, contrast=1)
#also include the T47D samples
dba.plotHeatmap(tamoxifen,contrast=1,mask=tamoxifen$masks$All)
#correlation heatmap without MCF7
plot(tamoxifen,contrast=1,mask=!tamoxifen$masks$MCF7)
# binding heatmap using only the MCF7 samples
dba.plotHeatmap(tamoxifen,contrast=1,mask=tamoxifen$masks$MCF7,correlations=FALSE)

```

dba.plotMA

*Generate MA and scatter plots of differential binding analysis results***Description**

Generates MA and scatter plots of differential binding analysis results.

**Usage**

```

dba.plotMA(DBA, contrast=1, method=DBA$config$AnalysisMethod,
           th=DBA$config$th, bUsePval=DBA$config$bUsePval,

```

```
fold=0, bNormalized=TRUE,
factor="", bFlip=FALSE, bXY=FALSE, dotSize=.45,
bSignificant=TRUE, bSmooth=TRUE,
xrange, yrange, ...)
```

### Arguments

DBA	DBA object, on which <code>dba.analyze</code> should have been successfully run.
contrast	number of contrast to report on. See <code>dba.show(DBA, bContrast=TRUE)</code> to get contrast numbers.
method	method or vector of methods to plot results for: <ul style="list-style-type: none"> <li>• <a href="#">DBA_DESEQ2</a></li> <li>• <a href="#">DBA_DESEQ2_BLOCK</a></li> <li>• <a href="#">DBA_EDGER</a></li> <li>• <a href="#">DBA_EDGER_BLOCK</a></li> </ul>
th	significance threshold; all sites with FDR (or p-values, see <code>bUsePval</code> ) less than or equal to this value will be colored red in the plot
bUsePval	logical indicating whether to use FDR (FALSE) or p-value (TRUE) for thresholding.
fold	will only include sites with fold change greater than this as significant (colored red).
bNormalized	logical indicating whether to plot normalized data using normalization factors computed by differential analysis method (TRUE) or raw read counts (FALSE).
factor	string to be prepended to plot main title; e.g. factor name.
bFlip	logical indicating that order of groups in contrast should be "flipped", allowing control of which sample group will have positive and which will have negative fold changes.
bXY	logical indicating whether to draw MA plot (FALSE) or XY scatter plot (TRUE).
dotSize	size of points on plot (cex).
bSignificant	Logical indicating if points corresponding to significantly differentially bound sites (based on contrast, th, bUsePval, and fold parameters) should be overlaid in red.
bSmooth	logical indicating that basic plot should be plotted using smooth Scatter. Note that overlaid significant sites will be not plotted using a smoothing function.
xrange	vector of length 2 containing the desired minimum and maximum concentrations to plot.
yrange	vector of length 2 containing the desired minimum and maximum fold changes to plot.
...	passed to plot.

### Author(s)

Rory Stark

### See Also

[dba.analyze](#)



**Examples**

```

data(tamoxifen_analysis)

# default MA plot
dba.plotMA(tamoxifen)

#XY plots (with raw and normalized data)
par(mfrow=c(1,2))
dba.plotMA(tamoxifen,bXY=TRUE,bSmooth=FALSE,bNormalized=FALSE)
dba.plotMA(tamoxifen,bXY=TRUE,bSmooth=FALSE,bNormalized=TRUE)

```

---

dba.plotPCA	<i>PCA plot</i>
-------------	-----------------

---

**Description**

Principal Component Analysis plot

**Usage**

```

dba.plotPCA(DBA, attributes, minval, maxval,
            contrast, method=DBA$config$AnalysisMethod,
            th=DBA$config$th, bUsePval=DBA$config$bUsePval,
            report, score, bLog=TRUE, mask, sites, label, cor=FALSE,
            b3D=FALSE, vColors, dotSize, labelSize, labelCols,
            components=1:3, ...)

```

**Arguments**

DBA	DBA object.
attributes	attribute or vector of attributes to use to color plotted points. Each unique combination of attribute values will be assigned a color. Chosen from: <ul style="list-style-type: none"> <li>• <a href="#">DBA_GROUP</a></li> <li>• <a href="#">DBA_ID</a></li> <li>• <a href="#">DBA_TISSUE</a></li> <li>• <a href="#">DBA_FACTOR</a></li> <li>• <a href="#">DBA_CONDITION</a></li> <li>• <a href="#">DBA_TREATMENT</a></li> <li>• <a href="#">DBA_REPLICATE</a></li> <li>• <a href="#">DBA_CONSENSUS</a></li> <li>• <a href="#">DBA_CALLER</a></li> </ul> <p>Note that <a href="#">DBA_GROUP</a> is a special attribute which will result in samples from each group in a contrast being colored separately.</p>
minval	Set all scores less than this to minval
maxval	Set all scores greater than this to maxval
contrast	number of contrast to use for PCA; if present, plots a PCA based on a differential binding affinity analysis (see <a href="#">dba.analyze</a> ). If mask is unspecified, only the samples in the contrast will be included. See <a href="#">dba.show</a> (DBA, bContrast=T) to get contrast numbers. If missing, uses scores in the main binding matrix.

method	method used for analysis (used in conjunction with contrast): <ul style="list-style-type: none"> <li>• <a href="#">DBA_DESEQ2</a></li> <li>• <a href="#">DBA_DESEQ2_BLOCK</a></li> <li>• <a href="#">DBA_EDGER</a></li> <li>• <a href="#">DBA_EDGER_BLOCK</a></li> </ul>
th	significance threshold; all sites with FDR (or p-values, see bUsePval) less than or equal to this value will be included in the PCA, subject to maxVal. Used in conjunction with contrast.
bUsePval	if TRUE, uses p-value instead of FDR for thresholding. Used in conjunction with contrast.
report	report (obtained from <a href="#">dba.report</a> ) specifying the data to be used . If this is present, the method, th, and bUsePval parameters are ignored.
score	Score to use for count data. Only used when plotting the global binding matrix (no contrast specified). One of: <ul style="list-style-type: none"> <li>• <a href="#">DBA_SCORE_READS</a></li> <li>• <a href="#">DBA_SCORE_READS_MINUS</a></li> <li>• <a href="#">DBA_SCORE_READS_FOLD</a></li> <li>• <a href="#">DBA_SCORE_RPKM</a></li> <li>• <a href="#">DBA_SCORE_RPKM_FOLD</a></li> <li>• <a href="#">DBA_SCORE_TMM_READS_FULL</a></li> <li>• <a href="#">DBA_SCORE_TMM_READS_EFFECTIVE</a></li> <li>• <a href="#">DBA_SCORE_TMM_MINUS_FULL</a></li> <li>• <a href="#">DBA_SCORE_TMM_MINUS_EFFECTIVE</a></li> </ul>
bLog	Logical indicating that log2 values should be used. Only applicable to read count scores (not peak scores).
mask	mask indicating a subset of peaksets to use when using global binding matrix scores. If a contrast is specified, these peaksets will be included, but only the significantly differentially bound sites (using th, bUsePval, and/or report) will be included. See <a href="#">dba.mask</a> .
sites	logical vector indicating which sites to include in PCA. Only relevant when using global binding matrix (contrast is missing).
label	A metadata field to use as a label in 2D plots. The value for this field will be written directly on the plot near the dot for each sample. Values can be any of those valid for the <code>attributes</code> parameter.
cor	a logical value indicating whether the calculation should use the correlation matrix or the covariance matrix. Passed into <code>princomp</code> .
b3D	logical indicating that three principal components should be plotted (requires <code>package{rgl}</code> ). If FALSE, the first two principal components are plotted.
vColors	vector of custom colors; is absent, default colors will be used.
dotSize	size of dots to plot; is absent, a default will be calculated.
labelSize	Scaling factor for labels if present. Default is 0.8.
labelCols	Vector of colors to use for labels. Default is "black".
components	Number(s) of the components to plot. Can be a vector of two or three component numbers, or a single integer. If an integer, that component, in addition to the succeeding one (b3D=FALSE) or two (b3D=TRUE) will be plotted.
...	arguments passed to <code>plot</code> or <code>plot3d</code> ( <code>rgl</code> ).

**Details**

MODE: PCA plot using significantly differentially bound sites:

```
dba.plotPCA(DBA, attributes, minval, maxval, contrast, method, th, bUsePval, b3D=F, vColors, dotSize, ...)
```

MODE: PCA plot using global binding matrix:

```
dba.plotPCA(DBA, attributes, minval, maxval, mask, sites, b3D=F, vColors, dotSize, ...)
```

**Value**

trellis plot from [lattice](#) package; see [xyplot](#)

**Note**

uses [rgl](#) package for 3D plots (if available)

**Author(s)**

Rory Stark

**See Also**

[dba.analyze](#), [dba.plotHeatmap](#)

**Examples**

```
data(tamoxifen_peaks)

# peakcaller scores PCA
dba.plotPCA(tamoxifen)

# raw count correlation PCA
data(tamoxifen_analysis)
dba.plotPCA(tamoxifen)

#PCA based on normalized data for all sites
dba.plotPCA(tamoxifen,contrast=1,th=1)

#PCA based on DB sites only
p <- dba.plotPCA(tamoxifen,contrast=1)
p <- dba.plotPCA(tamoxifen,contrast=1,attributes=DBA_TISSUE)
p <- dba.plotPCA(tamoxifen,contrast=1,attributes=DBA_TISSUE,label=DBA_CONDITION)
p <- dba.plotPCA(tamoxifen,contrast=1,attributes=DBA_CONDITION,label=DBA_TISSUE)
p <- dba.plotPCA(tamoxifen,contrast=1,attributes=c(DBA_TISSUE,DBA_CONDITION),
                 label=DBA_REPLICATE)
```

---

 dba.plotVenn

*Draw 2-way, 3-way, or 4-way Venn diagrams of overlaps*


---

### Description

Draws 2-way, 3-way, or 4-way Venn diagrams of overlaps

### Usage

```
dba.plotVenn(DBA, mask, overlaps, label1, label2, label3, label4, main, sub,
  contrast, method=DBA$config$AnalysisMethod,
  th=DBA$config$th, bUsePval=DBA$config$bUsePval,
  bDB=TRUE, bNotDB, bAll=TRUE, bGain=FALSE, bLoss=FALSE,
  labelAttributes, DataType=DBA$config$DataType)
```

### Arguments

DBA	DBA object; if present, only the mask parameter will apply.
mask	mask or vector of peakset numbers indicating which peaksets to include in Venn diagram. Only 2 or 3 peaksets should be included. See <a href="#">dba.mask</a> . Only one of mask or overlaps is used.
overlaps	overlap record, as computed by <a href="#">dba.overlap</a> (Report=DBA_OLAP_PEAKS). Only one of mask or overlaps is used.
label1	label for first peakset in diagram
label2	label for second peakset in diagram
label3	label for third peakset in diagram
label4	label for fourth peakset in diagram
main	main title for plot
sub	subtitle for plot
contrast	contrast number(s) to use for results-based plots. This can be a vector of contrast numbers. See <a href="#">dba.show</a> (DBA, bContrast=T) to get contrast numbers.
method	if contrast is specified, include results from analyses using this method or methods: <ul style="list-style-type: none"> <li>• <a href="#">DBA_DESEQ2</a></li> <li>• <a href="#">DBA_DESEQ2_BLOCK</a></li> <li>• <a href="#">DBA_EDGER</a></li> <li>• <a href="#">DBA_EDGER_BLOCK</a></li> <li>• <a href="#">DBA_ALL_METHODS</a></li> <li>• <a href="#">DBA_ALL_BLOCK</a></li> <li>• <a href="#">DBA_ALL_METHODS_BLOCK</a></li> </ul>
th	if contrast is specified, use this significance threshold; all sites with FDR (or p-values, see bUsePval) less than or equal to this value will be considered differentially bound (DB).
bUsePval	if contrast is specified, this logical indicates whether to use FDR (FALSE) or p-value (TRUE) for thresholding.

bDB	if contrast is specified, this logical indicates that peaksets should include Differentially Bound (DB) sites (respecting the th, bUsePval, and fold parameters).
bNotDB	if contrast is specified, this logical indicates that peaksets should include non-Differentially Bound (non-DB) sites (respecting the th, bUsePval, and fold parameters).
bAll	if contrast is specified, this logical indicates peaksets combining peaks with both positive and negative fold changes should be included.
bGain	if contrast is specified, this logical indicates that peaksets with only positive fold changes should be included.
bLoss	if contrast is specified, this logical indicates that peaksets with only negative fold changes should be included.
labelAttributes	<p>is labels are not specified, use these attributes to create default labels:</p> <ul style="list-style-type: none"> <li>• <a href="#">DBA_ID</a></li> <li>• <a href="#">DBA_TISSUE</a></li> <li>• <a href="#">DBA_FACTOR</a></li> <li>• <a href="#">DBA_CONDITION</a></li> <li>• <a href="#">DBA_TREATMENT</a></li> <li>• <a href="#">DBA_REPLICATE</a></li> <li>• <a href="#">DBA_CONSENSUS</a></li> <li>• <a href="#">DBA_CALLER</a></li> </ul> <p>Only specified attributes that differ between peaksets will be used for labels; the ones that have the same value for all peaksets will be used as the default subtitle.</p>
DataType	<p>if bReturnPeaksets is set to TRUE, the class of object that peaksets should be returned as:</p> <ul style="list-style-type: none"> <li>• <a href="#">DBA_DATA_GRANGES</a></li> <li>• <a href="#">DBA_DATA_RANGEDDATA</a></li> <li>• <a href="#">DBA_DATA_FRAME</a></li> </ul> <p>Can be set as default behavior by setting DBA\$config\$DataType. Alternatively, this can be set to:</p> <ul style="list-style-type: none"> <li>• <a href="#">DBA_DATA_DBAOBJECT</a></li> </ul> <p>to return a results-based DBA object, if a contrast is specified.</p>

**Value**

Either a list of peaksets is returned invisibly (as described in [dba.overlap](#)), or, if DataType=DBA\_DATA\_DBAOBJECT, a results-based DBA object.

**Note**

When working with results overlaps (a least one contrast is specified), and results-oriented DBA object is generated internally (as described in [dba.report](#)). In some cases, it may be better to generate the DBA object explicitly (using [dba.report](#) or setting bReturnPeaksets=TRUE and DataType=DBA\_DATA\_DBAOBJECT). This include the case where mseveral plots are being made of the same results set, and it takes a long time to generate the results-based DBA object, as well as the case where there are more than four results peaksets and a mask needs to be specified. I

This function relies on [vennPlot](#) in the systemPipeR package, written by Thomas Girke.

**Author(s)**

Rory Stark

**See Also**[dba.analyze](#), [dba.overlap](#), [dba.report](#), [dba.plotPCA](#), [vennPlot](#)**Examples**

```

data(tamoxifen_peaks)

par(mfrow=c(2,2))
# 2-way Venn
dba.plotVenn(tamoxifen,6:7)
dba.plotVenn(tamoxifen,tamoxifen$mask$ZR75)

# 3-way Venn (done two different ways)
dba.plotVenn(tamoxifen,tamoxifen$mask$MCF7&tamoxifen$mask$Responsive)
olaps <- dba.overlap(tamoxifen,tamoxifen$mask$MCF7&tamoxifen$mask$Responsive)
dba.plotVenn(tamoxifen,overlaps=olaps,
             label1="Rep 1",label2="Rep 2",label3="Rep 3",main="MCF7 (Responsive) Replicates")

#Venn of overlaps
Responsive=dba(tamoxifen,tamoxifen$mask$Responsive)
Responsive
Responsive <- dba.peakset(Responsive,1:3,sampID="MCF7")
Responsive <- dba.peakset(Responsive,4:5,sampID="T47D")
Responsive <- dba.peakset(Responsive,6:7,sampID="ZR75")
par(mfrow=c(1,1))
dba.plotVenn(Responsive,Responsive$mask$Consensus)

#4-way overlap
data(tamoxifen_peaks)
tamoxifen <- dba.peakset(tamoxifen, consensus=DBA_TISSUE)
par(mfrow=c(1,1))
dba.plotVenn(tamoxifen,tamoxifen$mask$Consensus,main="Tissue consensus overlaps")

#Venns of differentially bound sites
data(tamoxifen_analysis)
tamoxifen <- dba.contrast(tamoxifen,categories=DBA_CONDITION,block=tamoxifen$mask$MCF7)
tamoxifen <- dba.analyze(tamoxifen,method=c(DBA_EDGER,DBA_DESEQ2))
dba.plotVenn(tamoxifen,contrast=1,method=DBA_ALL_METHODS_BLOCK)
dba.plotVenn(tamoxifen,contrast=1,method=DBA_ALL_BLOCK,bAll=FALSE,bGain=TRUE,bLoss=TRUE)
par(mfrow=c(2,1))
dba.plotVenn(tamoxifen,contrast=1,method=DBA_ALL_BLOCK,bAll=FALSE,bGain=TRUE,bLoss=FALSE)
dba.plotVenn(tamoxifen,contrast=1,method=DBA_ALL_BLOCK,bAll=FALSE,bGain=FALSE,bLoss=TRUE)

```

---

dba.report*Generate a report for a differential binding affinity analysis*

---

**Description**

Generates a report for a differential binding affinity analysis

**Usage**

```
dba.report(DBA, contrast, method=DBA$config$AnalysisMethod,
          th=DBA$config$th, bUsePval=DBA$config$bUsePval,
          fold=0, bNormalized=TRUE, bFlip=FALSE,
          bCalled=FALSE, bCounts=FALSE, bCalledDetail=FALSE,
          bDB, bNotDB, bAll=TRUE, bGain=FALSE, bLoss=FALSE,
          file,initString=DBA$config$reportInit,ext='csv',DataType=DBA$config$DataType)
```

**Arguments**

DBA	DBA object. A differential binding affinity analysis needs to have been previously carried out (see <a href="#">dba.analyze</a> ).
contrast	contrast number to report on. When generating a report-based DBA object, this can be a vector of contrast numbers. If missing, defaults to first contrast for reports, and all contrasts when generating a report-based DBA object. See <a href="#">dba.show</a> (DBA, bContrast=T) to get contrast numbers.
method	method used for analysis: <ul style="list-style-type: none"> <li>• <a href="#">DBA_DESEQ2</a></li> <li>• <a href="#">DBA_DESEQ2_BLOCK</a></li> <li>• <a href="#">DBA_EDGER</a></li> <li>• <a href="#">DBA_EDGER_BLOCK</a></li> </ul> <p>When generating a report-based DBA object (see bDB and bNotDB parameters below), a list of methods may be supplied, including the shortcuts</p> <ul style="list-style-type: none"> <li>• <a href="#">DBA_ALL_METHODS</a></li> <li>• <a href="#">DBA_ALL_BLOCK</a></li> <li>• <a href="#">DBA_ALL_METHODS_BLOCK</a></li> </ul>
th	significance threshold; all sites with FDR (or p-values, see bUsePval) less than or equal to this value will be included in the report. A value of 1 will include all binding sites in the report.
bUsePval	logical indicating whether to use FDR (FALSE) or p-value (TRUE) for thresholding.
fold	only sites with an absolute Fold value greater than equal to this will be included in the report.
bNormalized	logical indicating that normalized data (using normalization factors computed by differential analysis method) should be reported. FALSE uses raw count data.
bFlip	logical indicating that order of groups in contrast should be "flipped", allowing control of which sample group will have positive and which will have negative fold changes.
bCalled	logical indicating that peak caller status should be included (if available from a previous call to <a href="#">dba.count</a> (bCalledMasks=TRUE)). This will add a column for each group, each indicating the number of samples in the group identified as a peak in the original peaksets.
bCounts	logical indicating that count data for individual samples should be reported as well as group statistics. Columns are added for each sample in the first group, followed by columns for each sample in the second group.

bCalledDetail	logical indicating that peak caller status should be included for each sample (if available). Columns are added for each sample in the first group, followed by columns for each sample in the second group.
bDB	logical indicating that a report-based DBA object should be generated, and that it should include Differentially Bound (DB) sites (respecting the th, bUsePval, and fold parameters).
bNotDB	logical indicating that a report-based DBA object should be generated, and that it should include non-Differentially Bound (non-DB) sites (respecting the th, bUsePval, and fold parameters).
bAll	logical indicating that a report-based DBA object should be generated, and that it should include peaksets combining peaks with both positive and negative fold changes.
bGain	logical indicating that a report-based DBA object should be generated, and that it should include peaksets with only positive fold changes.
bLoss	logical indicating that a report-based DBA object should be generated, and that it should include peaksets with only negative fold changes.
file	if present, also save the report to a comma separated value (csv) file, using this filename.
initString	if saving to a file, pre-pend this string to the filename.
ext	if saving to a file, append this extension to the filename.
DataType	The class of object for returned report: <ul style="list-style-type: none"> <li>• <a href="#">DBA_DATA_GRANGES</a></li> <li>• <a href="#">DBA_DATA_RANGEDDATA</a></li> <li>• <a href="#">DBA_DATA_FRAME</a></li> </ul> <p>If set to <a href="#">DBA_DATA_SUMMARIZED_EXPERIMENT</a>, the result will be a <a href="#">SummarizedExperiment</a> object, with all the count data and sample metadata for the experiment. The contrast statistics will be included as metadata columns in the rowData of the object. Can be set as default behavior by setting <code>DBA\$config\$DataType</code>.</p>

### Value

if neither bDB or bNotDB is set to TRUE, a report dataframe, GRanges, or RangedData object is returned, with a row for each binding site within the thresholding parameters, and the following columns:

Chr	Chromosome of binding site
Start	Starting base position of binding site
End	End base position of binding site
Conc	Concentration – mean (log) reads across all samples in both groups
Conc_group1	Group 1 Concentration – mean (log) reads across all samples first group
Conc_group2	Group 2 Concentration – mean (log) reads across all samples in second group
Fold	Fold difference – mean fold difference of binding affinity of group 1 over group 2 (Conc1 - Conc2). Absolute value indicates magnitude of the difference, and sign indicates which one is bound with higher affinity, with a positive value indicating higher affinity in the first group
p-value	p-value calculation – statistic indicating significance of difference (likelihood difference is not attributable to chance)



FDR                      adjusted p-value calculation – p-value subjected to multiple-testing correction

If bCalled is TRUE and caller status is available, two more columns will follow:

Called1                  Number of samples in group 1 that identified this binding site as a peak

Called2                  Number of samples in group 2 that identified this binding site as a peak

If bCounts is TRUE, a column will be present for each sample in group 1, followed by each sample in group 2. The Sample ID will be used as the column header. This column contains the read counts for the sample.

If bCalledDetail is TRUE, a column will be present for each sample in group 1, followed by each sample in group 2. The Sample ID will be used as the column header. This column contains a "+" to indicate for which sites the sample was called as a peak, and a "-" if it was not so identified.

If bDB or bNotDB is set to TRUE, a special DBA object is returned, containing peaksets based on sites determined to be differentially bound (or not) as specified using the bDB, bNotDB, bGain, bLoss, and bAll parameters. In this DBA object, the Tissue value will specify the direction of the change (Gain for positive fold changes, Loss for negative fold changes, and All for any fold change). The Factor value specifies if the peaks are differentially bound (DB) or not (!DB). The Condition value specifies the analysis method (e.g. edgeR), and the Treatment value is blank for unblocked analyses and set to block for blocked analyses.

### Author(s)

Rory Stark

### See Also

[dba.analyze](#)

### Examples

```
data(tamoxifen_analysis)

#Retrieve DB sites with FDR < 0.05
tamoxifen.DB <- dba.report(tamoxifen)
tamoxifen.DB

#Retrieve DB sites with p-value < 0.05 and Fold > 2
tamoxifen.DB <- dba.report(tamoxifen, th=.05, bUsePval=TRUE, fold=2)
tamoxifen.DB

#Retrieve all sites with confidence stats
# and how many times each site was identified as a peak
tamoxifen.DB <- dba.report(tamoxifen, th=1, bCalled=TRUE)
tamoxifen.DB

#Retrieve all sites with confidence stats and normalized counts
tamoxifen.DB <- dba.report(tamoxifen, th=1, bCounts=TRUE)
tamoxifen.DB

#Retrieve all sites with confidence stats and raw counts
tamoxifen.DB <- dba.report(tamoxifen, th=1, bCounts=TRUE, bNormalized=FALSE)
tamoxifen.DB

#Retrieve report as a SummarizedObject
```

```

tamoxifen.sset <- dba.report(tamoxifen, DataType=DBA_DATA_SUMMARIZED_EXPERIMENT)
tamoxifen.sset

#Retrieve report-based DBA object
data(tamoxifen_counts)
tamoxifen <- dba.contrast(tamoxifen, categories=DBA_CONDITION, block=tamoxifen$mask$MCF7)
tamoxifen <- dba.analyze(tamoxifen,bCorPlot=FALSE)
tamoxifen.DB <- dba.report(tamoxifen,method=c(DBA_DESEQ2,DBA_DESEQ2_BLOCK),
                          bDB=TRUE, bGain=TRUE, bLoss=TRUE)
dba.plotVenn(tamoxifen.DB,1:4,label1="Single Factor GAIN",label2="Single Factor LOSS",
             label3="Blocking Factor GAIN",label4="Blocking Factor LOSS")

```

---

dba.save	<i>save DBA object</i>
----------	------------------------

---

## Description

Writes out DBA object

## Usage

```
dba.save(DBA, file='DBA', dir='.', pre='dba_', ext='RData', bMinimize=FALSE)
```

## Arguments

DBA	DBA object
file	main filename
dir	directory to save model in
pre	string to pre-pend to filename
ext	extensions to use
bMinimize	logical indicating saved DBA object should be compressed as much as possible.

## Value

string containing full path and filename.

## Author(s)

Rory Stark

## See Also

[dba.load](#)

## Examples

```

data(tamoxifen_peaks)
savefile <- dba.save(tamoxifen,'tamoxifenPeaks')
savefile
tamoxifen <- dba.load('tamoxifenPeaks')
unlink(savefile)

```

---

dba.show *List attributes of peaksets of contrasts associated with a DBA object*

---

### Description

Returns attributes of peaksets and/or contrasts associated with a DBA object.

### Usage

```
dba.show(DBA, mask, attributes, bContrasts=FALSE,
         th=DBA$config$th, bUsePval=DBA$config$bUsePval)
```

### Arguments

DBA	DBA object
mask	mask of peaksets for which to get attributes (used when obtaining peakset attributes, i.e. bContrasts=FALSE).
attributes	attribute or vector of attributes to retrieve. Number of intervals is always shown. Used when obtaining peakset attributes, i.e. bContrasts=FALSE. Values: <ul style="list-style-type: none"> <li>• <a href="#">DBA_ID</a></li> <li>• <a href="#">DBA_TISSUE</a></li> <li>• <a href="#">DBA_FACTOR</a></li> <li>• <a href="#">DBA_CONDITION</a></li> <li>• <a href="#">DBA_TREATMENT</a></li> <li>• <a href="#">DBA_REPLICATE</a></li> <li>• <a href="#">DBA_CONSENSUS</a></li> <li>• <a href="#">DBA_CALLER</a></li> <li>• <a href="#">DBA_CONTROL</a></li> <li>• <a href="#">DBA_INTERVALS</a></li> <li>• <a href="#">DBA_FRIP</a></li> </ul>
bContrasts	logical indicating whether peaksets or contrast attributes are to be retrieved. TRUE retrieves a dataframe of contrast information instead of peakset attributes. If no contrasts are set, returns possible contrasts. See <a href="#">dba.contrast</a> .
th	if bContrasts is TRUE, then th is used as the threshold for determining how many significant sites there are for each contrast. Only relevant when obtaining contrast attributes (bContrasts=TRUE) and <a href="#">dba.analyze</a> has been run.
bUsePval	logical indicating that p-values will be used (along with th) to determine how many significant sites there are for each contrast; if FALSE, adjusted p-values (FDR) are used. Only relevant when obtaining contrast attributes (bContrasts=TRUE) and <a href="#">dba.analyze</a> has been run.

### Details

MODE: Return attributes of peaksets associated with a DBA object:

```
dba.show(DBA, mask, attributes)
```

MODE: Return contrasts associated with a DBA object:

```
dba.show(DBA,bContrasts=T, th, bUsePval)
```

**Value**

dataframe with peakset attributes.

If `bContrasts == FALSE`, each row represents a peakset, and each column is an attributes, with the final column, `Intervals`, indicating how many sites there are in the peakset.

If `bContrasts == TRUE`, each row represent a contrast, with the following columns:

Group1	Label for first group of contrast
Members1	Number of samples in first group of contrast
Group2	Label for first group of contrast
Members3	Number of samples in first group of contrast

if `dba.analyze` has been successfully run, there there will be up to four more columns showing the number of significant differentially bound (DB) sites identified for

DB.edgeR	Number of significantly differentially bound sites identified using edgeR
DB.DESeq	Number of significantly differentially bound sites identified using DESeq
DB.edgeR.block	Number of significantly differentially bound sites identified for blocking analysis using edgeR
DB.DESeq.block	Number of significantly differentially bound sites identified for blocking analysis using DESeq

**Author(s)**

Rory Stark

**See Also**

[dba](#), [dba.peakset](#), [dba.contrast](#), [dba.analyze](#)

**Examples**

```
data(tamoxifen_peaks)
dba.show(tamoxifen)
dba.show(tamoxifen, tamoxifen$mask$Responsive)
dba.show(tamoxifen, attributes=c(DBA_TISSUE, DBA_REPLICATE, DBA_CONDITION))

data(tamoxifen_counts)
tamoxifen <- dba.contrast(tamoxifen)
dba.show(tamoxifen, bContrasts=TRUE)

#alternatively:
tamoxifen
tamoxifen$config$th <- .05
tamoxifen
```

---

DiffBind -- DBA global constant variables  
*Constant variables used in DiffBind package*

---

## Description

Constant variables used in DiffBind package

## Usage

DBA\_ID  
DBA\_FACTOR  
DBA\_TISSUE  
DBA\_CONDITION  
DBA\_TREATMENT  
DBA\_REPLICATE  
DBA\_CALLER  
DBA\_CONSENSUS  
DBA\_CONTROL  
DBA\_ALL\_ATTRIBUTES

DBA\_INTERVALS  
DBA\_FRIP

DBA\_GROUP

DBA\_OLAP\_PEAKE  
DBA\_OLAP\_ALL  
DBA\_OLAP\_RATE

DBA\_COR  
DBA\_OLAP  
DBA\_INALL

DBA\_SCORE\_READS  
DBA\_SCORE\_READS\_MINUS  
DBA\_SCORE\_READS\_FOLD  
DBA\_SCORE\_RPKM  
DBA\_SCORE\_RPKM\_FOLD  
DBA\_SCORE\_TMM\_READS\_FULL  
DBA\_SCORE\_TMM\_READS\_EFFECTIVE  
DBA\_SCORE\_TMM\_MINUS\_FULL  
DBA\_SCORE\_TMM\_MINUS\_EFFECTIVE  
DBA\_SCORE\_TMM\_READS\_FULL\_CPM  
DBA\_SCORE\_TMM\_READS\_EFFECTIVE\_CPM  
DBA\_SCORE\_TMM\_MINUS\_FULL\_CPM  
DBA\_SCORE\_TMM\_MINUS\_EFFECTIVE\_CPM  
DBA\_SCORE\_SUMMIT  
DBA\_SCORE\_SUMMIT\_ADJ  
DBA\_SCORE\_SUMMIT\_POS

DBA\_READS\_DEFAULT  
 DBA\_READS\_BAM  
 DBA\_READS\_BED

DBA\_EDGER  
 DBA\_DESEQ  
 DBA\_DESEQ2  
 DBA\_EDGER\_BLOCK  
 DBA\_DESEQ\_BLOCK  
 DBA\_DESEQ2\_BLOCK  
 DBA\_EDGER\_CLASSIC  
 DBA\_DESEQ\_CLASSIC  
 DBA\_EDGER\_GLM  
 DBA\_DESEQ\_GLM  
 DBA\_ALL\_METHODS  
 DBA\_ALL\_BLOCK  
 DBA\_ALL\_METHODS\_BLOCK

DBA\_DATA\_FRAME  
 DBA\_DATA\_GRANGES  
 DBA\_DATA\_RANGEDDATA  
 DBA\_DATA\_SUMMARIZED\_EXPERIMENT  
 DBA\_DATA\_DBAOBJECT

### Arguments

DBA_ID	DBA peakset metadata: Peakset ID
DBA_FACTOR	DBA peakset metadata: Factor
DBA_TISSUE	DBA peakset metadata: Tissue
DBA_CONDITION	DBA peakset metadata: Condition
DBA_TREATMENT	DBA peakset metadata: Treatment
DBA_REPLICATE	DBA peakset metadata: Replicate
DBA_CALLER	DBA peakset metadata: Peak Caller
DBA_CONSENSUS	DBA peakset metadata: Is this a consensus peakset?
DBA_CONTROL	DBA peakset metadata: ID of Control sample
DBA_ALL_ATTRIBUTES	DBA peakset metadata: all attributes that can be used in certain plot labels (cf <a href="#">dba.plotVenn</a> ), equivalent to <code>c(DBA_ID, DBA_TISSUE, DBA_FACTOR, DBA_CONDITION, DBA_TREATMENT, DBA_REPLICATE, DBA_CALLER, DBA_CONSENSUS, DBA_CONTROL)</code>
DBA_INTERVALS	DBA peakset metadata: Number of intervals in peakset
DBA_FRIP	DBA peakset metadata: Fraction of Reads in Peaks (number of reads in intervals divided by total number of reads in library)
DBA_GROUP	DBA peakset metadata: color PCA plot using contrast groups
DBA_OLAP_PEAKS	dba.overlap mode: return overlapping/unique peaksets
DBA_OLAP_ALL	dba.overlap mode: return report of correlations/overlaps for each pair of samples
DBA_OLAP_RATE	dba.overlap mode: return overlap rates
DBA_COR	When plotting a heatmap from an overlap record, use the correlation value.
DBA_OLAP	When plotting a heatmap from an overlap record, use the percentage overlap value.

DBA_INALL	When plotting a heatmap from an overlap record, use the number of peaks in common to both samples.
DBA_SCORE_READS	dba.count score is number of reads in ChIP
DBA_SCORE_READS_FOLD	dba.count score is number of reads in ChIP divided by number of reads in Control
DBA_SCORE_READS_MINUS	dba.count score is number of reads in ChIP minus number of reads in Control
DBA_SCORE_RPKM	dba.count score is RPKM of ChIP
DBA_SCORE_RPKM_FOLD	dba.count score is RPKM of ChIP divided by RPKM of Control
DBA_SCORE_TMM_READS_FULL	dba.count score is TMM normalized (using edgeR), using ChIP read counts and Full Library size
DBA_SCORE_TMM_READS_EFFECTIVE	dba.count score is TMM normalized (using edgeR), using ChIP read counts and Effective Library size
DBA_SCORE_TMM_MINUS_FULL	dba.count score is TMM normalized (using edgeR), using ChIP read counts minus Control read counts and Full Library size
DBA_SCORE_TMM_MINUS_EFFECTIVE	dba.count score is TMM normalized (using edgeR), using ChIP read counts minus Control read counts and Effective Library size
DBA_SCORE_TMM_READS_FULL_CPM	dba.count score is TMM normalized (using edgeR), using ChIP read counts and Full Library size, reported in counts-per-million.
DBA_SCORE_TMM_READS_EFFECTIVE_CPM	dba.count score is TMM normalized (using edgeR), using ChIP read counts and Effective Library size, reported in counts-per-million.
DBA_SCORE_TMM_MINUS_FULL_CPM	dba.count score is TMM normalized (using edgeR), using ChIP read counts minus Control read counts and Full Library size, reported in counts-per-million.
DBA_SCORE_TMM_MINUS_EFFECTIVE_CPM	dba.count score is TMM normalized (using edgeR), using ChIP read counts minus Control read counts and Effective Library size, reported in counts-per-million.
DBA_SCORE_SUMMIT	dba.count score is summit height (highest pile-up).
DBA_SCORE_SUMMIT_ADJ	dba.count score is summit height (highest pile-up), adjusted for library size.
DBA_SCORE_SUMMIT_POS	dba.count score is summit location (position of highest pile-up).
DBA_READS_DEFAULT	When counting read files, use the file extension to determine the file type.
DBA_READS_BAM	When counting read files, assume the file type is BAM, regardless of the file extension.
DBA_READS_BED	When counting read files, assume the file type is BED (or zipped BED), regardless of the file extension.

DBA_EDGER	differential analysis method: edgeR (default: DBA_EDGER_GLM)
DBA_DESEQ2	differential analysis method: DESeq2 (using a single-factor GLM)
DBA_EDGER_BLOCK	differential analysis method: edgeR with blocking factors (GLM)
DBA_DESEQ2_BLOCK	differential analysis method: DESeq2 with blocking factors (GLM)
DBA_DESEQ	differential analysis method: DESeq (default: DBA_DESEQ_CLASSIC)
DBA_DESEQ_BLOCK	differential analysis method: DESeq with blocking factors (GLM)
DBA_EDGER_CLASSIC	differential analysis method: "classic" edgeR for two-group comparisons
DBA_DESEQ_CLASSIC	differential analysis method: "classic" DESeq for two-group comparisons
DBA_EDGER_GLM	differential analysis method: use GLM in edgeR for two-group comparisons
DBA_DESEQ_GLM	differential analysis method: use GLM in DESeq for two-group comparisons
DBA_ALL_METHODS	use both analysis methods: c(DBA_EDGER, DBA_DESEQ2)
DBA_ALL_BLOCK	report on block results for both analysis methods: c(DBA_EDGER_BLOCK, DBA_DESEQ2_BLOCK)
DBA_ALL_METHODS_BLOCK	report on block results for all analysis methods, both blocked and unblocked: c(DBA_ALL_METHODS, DBA_ALL_BLOCK)
DBA_DATA_GRANGES	Use GRanges class for peaksets and reports. This is the default (DBA\$config\$DataType = DBA_DATA_GRANGES).
DBA_DATA_RANGEDDATA	Use RangedData class for peaksets and reports. Can be set as default (DBA\$config\$DataType = DBA_DATA_RANGEDDATA).
DBA_DATA_FRAME	Use data.frame class for peaksets and reports. Can be set as default (DBA\$config\$DataType = DBA_DATA_FRAME).
DBA_DATA_SUMMARIZED_EXPERIMENT	Return report as a <a href="#">SummarizedExperiment</a> .
DBA_DATA_DBAOBJECT	Return a result-based DBA object from <a href="#">dba.plotVenn</a> .

**Note**

Variables with ALL CAP names are used as constants within DiffBind.

**Author(s)**

Rory Stark



# Index

## \*Topic **package**

- DiffBind-package, 2
- DBA (DBA object methods), 6
- dba, 2, 3, 7, 24, 44
- DBA object methods, 6
- DBA tamoxifen resistance dataset, 7
- dba.analyze, 2, 8, 11, 12, 15, 28, 32, 33, 35, 38, 39, 41, 43, 44
- dba.contrast, 2, 8, 9, 10, 43, 44
- dba.count, 2, 12, 39
- dba.load, 2, 15, 42
- dba.mask, 2, 3, 10, 16, 18, 21, 34, 36
- dba.overlap, 2, 18, 29, 31, 36–38
- dba.peakset, 2, 4, 6, 13, 21, 23, 24, 44
- dba.plotBox, 2, 25
- dba.plotHeatmap, 2, 19, 21, 28, 35
- dba.plotMA, 2, 31
- dba.plotPCA, 2, 33, 38
- dba.plotVenn, 2, 21, 36, 46, 48
- dba.report, 2, 9, 19, 22, 29, 34, 37, 38, 38
- dba.save, 2, 16, 42
- dba.show, 2, 6, 18, 21, 28, 32, 33, 36, 39, 43
- DBA\_ALL\_ATTRIBUTES (DiffBind -- DBA global constant variables), 45
- DBA\_ALL\_BLOCK, 36, 39
- DBA\_ALL\_BLOCK (DiffBind -- DBA global constant variables), 45
- DBA\_ALL\_METHODS, 36, 39
- DBA\_ALL\_METHODS (DiffBind -- DBA global constant variables), 45
- DBA\_ALL\_METHODS\_BLOCK, 36, 39
- DBA\_ALL\_METHODS\_BLOCK (DiffBind -- DBA global constant variables), 45
- DBA\_CALLER, 19, 26, 28, 30, 33, 37, 43
- DBA\_CALLER (DiffBind -- DBA global constant variables), 45
- DBA\_CONDITION, 19, 26, 28, 30, 33, 37, 43
- DBA\_CONDITION (DiffBind -- DBA global constant variables), 45
- DBA\_CONSENSUS, 19, 26, 28, 33, 37, 43
- DBA\_CONSENSUS (DiffBind -- DBA global constant variables), 45
- DBA\_CONTROL, 43
- DBA\_CONTROL (DiffBind -- DBA global constant variables), 45
- DBA\_COR, 29
- DBA\_COR (DiffBind -- DBA global constant variables), 45
- DBA\_DATA\_DBAOBJECT, 37
- DBA\_DATA\_DBAOBJECT (DiffBind -- DBA global constant variables), 45
- DBA\_DATA\_FRAME, 19, 37, 40
- DBA\_DATA\_FRAME (DiffBind -- DBA global constant variables), 45
- DBA\_DATA\_GRANGES, 19, 37, 40
- DBA\_DATA\_GRANGES (DiffBind -- DBA global constant variables), 45
- DBA\_DATA\_RANGEDDATA, 19, 37, 40
- DBA\_DATA\_RANGEDDATA (DiffBind -- DBA global constant variables), 45
- DBA\_DATA\_SUMMARIZED\_EXPERIMENT, 40
- DBA\_DATA\_SUMMARIZED\_EXPERIMENT (DiffBind -- DBA global constant variables), 45
- DBA\_DESEQ, 8
- DBA\_DESEQ (DiffBind -- DBA global constant variables), 45
- DBA\_DESEQ2, 8, 19, 26, 28, 32, 34, 36, 39
- DBA\_DESEQ2 (DiffBind -- DBA global constant variables), 45
- DBA\_DESEQ2\_BLOCK, 19, 26, 28, 32, 34, 36, 39
- DBA\_DESEQ2\_BLOCK (DiffBind -- DBA global constant variables), 45
- DBA\_DESEQ\_BLOCK (DiffBind -- DBA global constant variables), 45
- DBA\_DESEQ\_CLASSIC (DiffBind -- DBA global constant variables), 45
- DBA\_DESEQ\_GLM (DiffBind -- DBA global constant variables), 45
- DBA\_EDGER, 8, 19, 26, 28, 32, 34, 36, 39
- DBA\_EDGER (DiffBind -- DBA global constant variables), 45
- DBA\_EDGER\_BLOCK, 19, 26, 28, 32, 34, 36, 39
- DBA\_EDGER\_BLOCK (DiffBind -- DBA global constant variables), 45
- DBA\_EDGER\_CLASSIC (DiffBind -- DBA

- global constant variables), 45
- DBA\_EDGER\_GLM(DiffBind -- DBA global constant variables), 45
- DBA\_FACTOR, 19, 26, 28, 30, 33, 37, 43
- DBA\_FACTOR(DiffBind -- DBA global constant variables), 45
- DBA\_FRIP, 43
- DBA\_FRIP(DiffBind -- DBA global constant variables), 45
- DBA\_GROUP, 26, 30, 33
- DBA\_GROUP(DiffBind -- DBA global constant variables), 45
- DBA\_ID, 19, 26, 28, 33, 37, 43
- DBA\_ID(DiffBind -- DBA global constant variables), 45
- DBA\_INALL, 29
- DBA\_INALL(DiffBind -- DBA global constant variables), 45
- DBA\_INTERVALS, 43
- DBA\_INTERVALS(DiffBind -- DBA global constant variables), 45
- DBA\_OLAP, 29
- DBA\_OLAP(DiffBind -- DBA global constant variables), 45
- DBA\_OLAP\_ALL, 18
- DBA\_OLAP\_ALL(DiffBind -- DBA global constant variables), 45
- DBA\_OLAP\_PEAKS, 18
- DBA\_OLAP\_PEAKS(DiffBind -- DBA global constant variables), 45
- DBA\_OLAP\_RATE, 18
- DBA\_OLAP\_RATE(DiffBind -- DBA global constant variables), 45
- DBA\_READS\_BAM(DiffBind -- DBA global constant variables), 45
- DBA\_READS\_BED(DiffBind -- DBA global constant variables), 45
- DBA\_READS\_DEFAULT(DiffBind -- DBA global constant variables), 45
- DBA\_REPLICATE, 19, 26, 28, 30, 33, 37, 43
- DBA\_REPLICATE(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_READS, 29, 34
- DBA\_SCORE\_READS(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_READS\_FOLD, 29, 34
- DBA\_SCORE\_READS\_FOLD(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_READS\_MINUS, 29, 34
- DBA\_SCORE\_READS\_MINUS(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_RPKM, 29, 34
- DBA\_SCORE\_RPKM(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_RPKM\_FOLD, 29, 34
- DBA\_SCORE\_RPKM\_FOLD(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_SUMMIT(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_SUMMIT\_ADJ(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_SUMMIT\_POS(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_TMM\_MINUS\_EFFECTIVE, 29, 34
- DBA\_SCORE\_TMM\_MINUS\_EFFECTIVE(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_TMM\_MINUS\_EFFECTIVE\_CPM(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_TMM\_MINUS\_FULL, 29, 34
- DBA\_SCORE\_TMM\_MINUS\_FULL(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_TMM\_MINUS\_FULL\_CPM(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_TMM\_READS\_EFFECTIVE, 29, 34
- DBA\_SCORE\_TMM\_READS\_EFFECTIVE(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_TMM\_READS\_EFFECTIVE\_CPM(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_TMM\_READS\_FULL, 29, 34
- DBA\_SCORE\_TMM\_READS\_FULL(DiffBind -- DBA global constant variables), 45
- DBA\_SCORE\_TMM\_READS\_FULL\_CPM(DiffBind -- DBA global constant variables), 45
- DBA\_TISSUE, 19, 26, 28, 30, 33, 37, 43
- DBA\_TISSUE(DiffBind -- DBA global constant variables), 45
- DBA\_TREATMENT, 19, 26, 28, 30, 33, 37, 43
- DBA\_TREATMENT(DiffBind -- DBA global constant variables), 45
- DiffBind, 7
- DiffBind(DiffBind-package), 2
- DiffBind -- DBA global constant variables, 45
- DiffBind-package, 2
- Dist, 30
- GRanges, 30

lattice, [35](#)

plot.DBA (DBA object methods), [6](#)

print.DBA (DBA object methods), [6](#)

SummarizedExperiment, [5](#), [40](#), [48](#)

summarizeOverlaps, [14](#)

summary.DBA (DBA object methods), [6](#)

tamoxifen (DBA tamoxifen resistance dataset), [7](#)

tamoxifen\_analysis (DBA tamoxifen resistance dataset), [7](#)

tamoxifen\_counts (DBA tamoxifen resistance dataset), [7](#)

tamoxifen\_peaks (DBA tamoxifen resistance dataset), [7](#)

vennPlot, [37](#), [38](#)

xyplot, [35](#)