

# Package: blaseRdata (via r-universe)

August 25, 2024

**Title** Supporting Data for the blaseRtools Package  
**Version** 0.0.0.9013  
**Description** What the package does (one paragraph).  
**License** MIT + file LICENSE  
**Encoding** UTF-8  
**Roxygen** list(markdown = TRUE)  
**RoxygenNote** 7.1.2  
**URL** <https://github.com/blaserlab/blaseRdata>  
**BugReports** <https://github.com/blaserlab/blaseRdata/issues>  
**LazyData** true  
**Imports** BSgenome, monocle3  
**Repository** <https://blaserlab.r-universe.dev>  
**RemoteUrl** <https://github.com/blaserlab/blaseRdata>  
**RemoteRef** HEAD  
**RemoteSha** 8ddc401c80dd0d78d9e8efbb647c1ba77bf64a66

## Contents

dr11_remove_genes . . . . .	2
dr11_tss . . . . .	2
Drerio . . . . .	3
hg38_granges . . . . .	3
hg38_granges_reduced . . . . .	4
hg38_remove_genes . . . . .	4
hg38_tss . . . . .	5
Hsapiens . . . . .	5
meme . . . . .	6
mm39_remove_genes . . . . .	6
msigdb_genesets . . . . .	7
msigdb_geneset_metadata . . . . .	7

vignette_cds . . . . .	8
vignette_pseudobulk_res . . . . .	9
vignette_top_markers . . . . .	9
vig_qc_res . . . . .	10
wordhash . . . . .	11
zfin_granges . . . . .	11
zfin_granges_reduced . . . . .	12

<b>Index</b>	<b>13</b>
--------------	-----------

---

dr11_remove_genes	<i>Genes to Remove from ScRNA-seq Analysis: Zebrafish</i>
-------------------	---

---

### Description

A precompiled list of mitochondrial and ribosomal gene names to remove.

### Usage

```
dr11_remove_genes
```

### Format

A character vector

### Details

See data-raw/remove\_genes

### Source

Ensembl

---

dr11_tss	<i>Zebrafish Gene TSS</i>
----------	---------------------------

---

### Description

From TxDb.Drerio.UCSC.danRer11.refGene

### Usage

```
dr11_tss
```

### Format

A GRanges object:

**Details**

see data-raw/tss.R

---

Drerio

*Drerio*

---

**Description**

This is a copy of the Drerio object from BSgenome.Drerio.UCSC.danRer11. It is here for more convenient use with blaseRtools functions and to reduce dependencies.

**Usage**

Drerio

**Format**

A BSgenome object:

**Details**

see data-raw/grcz11\_objects.R

**Source**

UCSC danRer11 genome build

---

hg38\_granges

*Human Gene Model*

---

**Description**

Full Genome Model for humans. Based on GRCh38 genome build.

**Usage**

hg38\_granges

**Format**

A GRanges object:

**Details**

see data-raw/hg38\_objects.R

**Source**

[http://ftp.ensembl.org/pub/release-104/gff3/homo\\_sapiens/](http://ftp.ensembl.org/pub/release-104/gff3/homo_sapiens/)

---

hg38\_granges\_reduced *Reduced Human Gene Model*

---

**Description**

Similar to the Human Gene Model "hg38\_granges" but with different metadata columns and only with 5', CDS, and 3' UTR ranges. Also has principle isoform data.

**Usage**

hg38\_granges\_reduced

**Format**

A GRanges object:

**Details**

see data-raw/trace\_objects.R

**Source**

[http://ftp.ensembl.org/pub/release-104/gff3/homo\\_sapiens/](http://ftp.ensembl.org/pub/release-104/gff3/homo_sapiens/)

---

hg38\_remove\_genes *Genes to Remove from ScRNA-seq Analysis: Human*

---

**Description**

A precompiled list of mitochondrial and ribosomal gene names to remove.

**Usage**

hg38\_remove\_genes

**Format**

A character vector

**Details**

See data-raw/remove\_genes

**Source**

Ensembl

---

`hg38_tss`*Human Gene TSS*

---

**Description**

From TxDb.Hsapiens.UCSC.hg38.knownGene

**Usage**

```
hg38_tss
```

**Format**

A GRanges object:

**Details**

see data-raw/tss.R

---

`Hsapiens`*Hsapiens*

---

**Description**

This is a copy of the Hsapiens object from BSgenome.Hsapiens.UCSC.hg38. It is here for more convenient use with blaseRtools functions and to reduce dependencies.

**Usage**

```
Hsapiens
```

**Format**

A BSgenome object:

**Details**

see data-raw/hg38\_objects.R

**Source**

UCSC hg38 genome build

---

meme

*Jaspar 2020 TF Motifs*

---

**Description**

JASPAR 2020 TF Motif database in character string format. Mostly for internal use.

**Usage**

meme

**Format**

A character vector.

**Details**

see data-raw/meme.R

**Source**

JASPAR2020\_CORE\_vertbrates\_non-redundant\_pfms\_jaspar.txt

---

mm39\_remove\_genes

*Genes to Remove from ScRNA-seq Analysis: Mouse*

---

**Description**

A precompiled list of mitochondrial and ribosomal gene names to remove.

**Usage**

mm39\_remove\_genes

**Format**

A character vector

**Details**

See data-raw/remove\_genes

**Source**

Ensembl

---

`msigdb_genesets`*Broad Institute MSIGDB Gene Sets*

---

**Description**

List of MSIGDB gene sets. Most gene sets are commonly known by "STANDARD\_NAME". This is a unique value and is provided as the name for each of the elements in this list. If you want to select a single gene set by name you can use the usual R syntax for selecting named elements: `msigdb_genesets$<gene set name>`. If you want to select by other metadata features you should use `bb_extract_msig`.

**Usage**`msigdb_genesets`**Format**

A list.

**Details**Use with `bb_extract_msig`**Source**<https://www.gsea-msigdb.org/gsea/msigdb/> accessed January 12 2022.

---

`msigdb_geneset_metadata`*Broad Institute MSIGDB Gene Set Metadata*

---

**Description**

Gene set metadata for the whole MSIGDB. Most gene sets are known by "STANDARD\_NAME". You can filter the gene set list by supplying a named filter list to the `bb_extract_msig` function. The name of each list element should be one of the column names in this metadata table and the list element contents should be the values to filter for. Filtering works in an additive way, meaning if you supply a filter list with two elements it will extract gene sets passing filters 1 AND 2.

**Usage**`msigdb_geneset_metadata`

**Format**

A data frame with 32796 rows and 18 variables:

STANDARD\_NAME character  
SYSTEMATIC\_NAME character  
HISTORICAL\_NAME character  
ORGANISM character  
PMID character  
AUTHORS character  
GEOID character  
EXACT\_SOURCE character  
GENESET\_LISTING\_URL character  
EXTERNAL\_DETAILS\_URL character  
CHIP character  
CATEGORY\_CODE character  
SUB\_CATEGORY\_CODE character  
CONTRIBUTOR character  
CONTRIBUTOR\_ORG character  
DESCRIPTION\_BRIEF character  
DESCRIPTION\_FULL character  
TAGS character

**Details**

Use with `bb_extract_msig`

**Source**

<https://www.gsea-msigdb.org/gsea/msigdb/> accessed January 12 2022.

---

vignette\_cds

*ScRNA-seq Vignette CDS*

---

**Description**

A small example CDS for the `blaseRtools` vignette

**Usage**

`vignette_cds`



**Format**

A cellDataSet object

**Details**

See blaseRtools scRNAseq Vignette

**Source**

<https://www.10xgenomics.com/resources/datasets/500-human-pbm-cs-3-1t-v-3-1-chromium-controller-3-1-low-6-1-0> and <https://www.10xgenomics.com/resources/datasets/500-human-pbm-cs-3-1t-v-3-1-chromium-x-3-1-low-6-1-0>

---

vignette\_pseudobulk\_res

*BlaseRtools ScRNA-seq Vignette CDS Pseudobulk Results*

---

**Description**

Pseudobulk header, results and qc plots.

**Usage**

vignette\_pseudobulk\_res

**Format**

A list of pseudobulk results

**Details**

See blaseRtools scRNAseq Vignette

---

vignette\_top\_markers

*BlaseRtools ScRNA-seq Vignette CDS Top Markers*

---

**Description**

Top Markers for the vignette CDS

**Usage**

vignette\_top\_markers

**Format**

A data frame with 1100 rows and 11 variables:

gene\_id character Ensembl gene identifier  
gene\_short\_name character Gene Name  
cluster\_method character Clustering Method  
cell\_group character Cell Cluster or Partition  
marker\_score double Composite score of fraction expressing and level of expression.  
mean\_expression double Mean marker expression  
fraction\_expressing double Fraction expressing  
specificity double Marker specificity  
pseudo\_R2 double Regression pseudo R squared  
marker\_test\_p\_value double Wald test P value  
marker\_test\_q\_value double q value

**Details**

See blaseRtools scRNAseq Vignette

---

vig\_qc\_res

*Vignette QC Results*

---

**Description**

Standard bb\_qc function output

**Usage**

vig\_qc\_res

**Format**

a list

**Details**

See blaseRtools scRNAseq Vignette

**Source**

vignette\_cds

---

wordhash

*Wordlist Hashtable*

---

### Description

Wordlist from the Grady Parts of Speech database. In order to reuse, 1.) Calculate the md5sum of the object you want to hash, 2.) Take the first 5 or more hex digits and fill them in this expression: `as.integer(as.hexmode("<hex digits>") %% 226857) + 1`. This will return an integer between 1 and 226857 which is the length of the wordlist hashtable. Then select the word according to the integer/index value using something like: `wordlist %>% filter(index == integer) %>% pull(word)`.

### Usage

wordhash

### Format

A data frame with 226857 rows and 2 variables:

index integer

word character

### Details

see `data-raw/wordhash.R` for construction

### Source

<https://www.gutenberg.org/ebooks/3201>

---

zfin\_granges

*Zebrafish Gene Model*

---

### Description

This is a granges object containing the full genome model from the zfin database. Based on GRCz11 genome build.

### Usage

zfin\_granges

### Format

A GRanges object:

**Details**

see data-raw/grcz11\_objects.R

**Source**

[https://zfin.org/downloads/zfin\\_genes.gff3](https://zfin.org/downloads/zfin_genes.gff3) release April 2018

---

zfin\_granges\_reduced *Reduced Zebrafish Gene Model*

---

**Description**

Similar to the Zebrafish Gene Model but with different metadata columns and only with 5', CDS and 3' UTR ranges. Also has principle isoform data.

**Usage**

```
zfin_granges_reduced
```

**Format**

A GRanges object:

**Details**

see data-raw/trace\_objects.R

**Source**

[https://zfin.org/downloads/zfin\\_genes.gff3](https://zfin.org/downloads/zfin_genes.gff3) release April 2018

# Index

## \* datasets

- dr11\_remove\_genes, [2](#)
  - dr11\_tss, [2](#)
  - Drerio, [3](#)
  - hg38\_granges, [3](#)
  - hg38\_granges\_reduced, [4](#)
  - hg38\_remove\_genes, [4](#)
  - hg38\_tss, [5](#)
  - Hsapiens, [5](#)
  - meme, [6](#)
  - mm39\_remove\_genes, [6](#)
  - msigdb\_geneset\_metadata, [7](#)
  - msigdb\_genesets, [7](#)
  - vig\_qc\_res, [10](#)
  - vignette\_cds, [8](#)
  - vignette\_pseudobulk\_res, [9](#)
  - vignette\_top\_markers, [9](#)
  - wordhash, [11](#)
  - zfin\_granges, [11](#)
  - zfin\_granges\_reduced, [12](#)
- 
- dr11\_remove\_genes, [2](#)
  - dr11\_tss, [2](#)
  - Drerio, [3](#)
- 
- hg38\_granges, [3](#)
  - hg38\_granges\_reduced, [4](#)
  - hg38\_remove\_genes, [4](#)
  - hg38\_tss, [5](#)
  - Hsapiens, [5](#)
- 
- meme, [6](#)
  - mm39\_remove\_genes, [6](#)
  - msigdb\_geneset\_metadata, [7](#)
  - msigdb\_genesets, [7](#)
- 
- vig\_qc\_res, [10](#)
  - vignette\_cds, [8](#)
  - vignette\_pseudobulk\_res, [9](#)
  - vignette\_top\_markers, [9](#)