

# Package ‘imsig’

May 8, 2026

**Type** Package

**Title** Immune Cell Gene Signatures for Profiling the Microenvironment of Solid Tumours

**Version** 1.1.3

**Author** Ajit Johnson Nirmal

**Maintainer** Ajit Johnson Nirmal <ajitjohnson.n@gmail.com>

**Description** Estimate the relative abundance of tissue-infiltrating immune subpopulations abundances using gene expression data.

**License** GPL-3

**URL** <https://github.com/ajitjohnson/imsig/>

**BugReports** <https://github.com/ajitjohnson/imsig/issues>

**Encoding** UTF-8

**LazyData** true

**Imports** HiClimR (>= 1.2), RColorBrewer (>= 1.1), igraph (>= 1.2), ggplot2 (>= 2.2), gridExtra (>= 2.3), survival (>= 2.4)

**RoxygenNote** 7.1.1

**Suggests** testthat

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2021-01-10 01:00:02 UTC

## Contents

corr_matrix	2
example_cli	3
example_data	3
feature_select	4
gene_stat	4
imsig	5
imsig_survival	6

plot_abundance . . . . .	7
plot_network . . . . .	8
plot_survival . . . . .	9
pp_exp . . . . .	10
pp_sig . . . . .	11
sig . . . . .	11

<b>Index</b>	<b>12</b>
--------------	-----------

---

corr_matrix	<i>Correlation matrix</i>
-------------	---------------------------

---

## Description

Creates a correlation matrix of ImSig signature genes.

## Usage

```
corr_matrix(exp, r)
```

## Arguments

exp	Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- head(example_data): <a href="#">example_data</a> .
r	Use a value between 0 and 1. Default is 0.6. This is a user defined correlation cut-off to perform feature selection ( <a href="#">feature_select</a> ). Feature selection aids to enrich the prediction of relative abundance of immune cells by filtering off poorly correlated ImSig genes. To get an idea of what cut-off to use check the results of ( <a href="#">gene_stat</a> ) and choose a cut-off that displays high median correlation and maintains a high proportion of genes after feature selection.

## Value

Gene-gene correlation matrix of ImSig genes.

---

`example_cli`*Example clinical data file for survival analysis with ImSig*

---

**Description**

An example clinical data file. Minimum required informations are the sample name (same as that of the expression matrix), event (dead or alive) and time to event (days, months or years).

**Usage**`example_cli`**Format**`dataframe`

---

`example_data`*Example transcriptomics data*

---

**Description**

Example expression data matrix. The data is preferred to be in natural scale with genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- `head(example_data)`

**Usage**`example_data`**Format**`dataframe`

---

feature\_select                      *Feature selection of signature genes*

---

### Description

ImSig genes were designed to be co-expressed in tissue transcriptomic data. However, depending on the dataset some of the genes may not co-express with the dominant module. In order to remove such deviant genes, a feature selection can be carried out based on correlation. This function removes genes that exhibit a poor correlation (less than the defined r value) with the dominant ImSig module. This step of feature selection is recommended to enrich the prediction of relative abundance of immune cells.

### Usage

```
feature_select(exp, r = 0.6)
```

### Arguments

exp	Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- <code>head(example_data)</code> : <a href="#">example_data</a> .
r	Use a value between 0 and 1. Default is 0.6. This is a user defined correlation cut-off to perform feature selection. To get an idea of what cut-off to use check the results of ( <a href="#">gene_stat</a> ) and choose a cut-off that displays high median correlation and maintains a high proportion of genes after feature selection.

### Value

Returns a list of 'feature selected' genes based on the set r value.

### Examples

```
feature_select (exp = example_data, r = 0.7)
```

---

gene\_stat                              *General statistics of ImSig analysis*

---

### Description

[Total genes in ImSig]: The total number of genes in ImSig list. [No. of ImSig genes in user dataset]: The number of ImSig genes found in user's dataset. Like all signatures, ImSig works best when this overlap is high, preferably over 75

**Usage**

```
gene_stat(exp, r = 0.6)
```

**Arguments**

**exp** Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- `head(example_data)`: [example\\_data](#).

**r** Use a value between 0 and 1. Default is 0.6. This is a user defined correlation cut-off to perform feature selection ([feature\\_select](#)). Feature selection aids to enrich the prediction of relative abundance of immune cells by filtering off poorly correlated ImSig genes. To get an idea of what cut-off to use check the results of ([gene\\_stat](#)) and choose a cut-off that displays high median correlation and maintains a high proportion of genes after feature selection.

**Value**

Dataframe of general statistics of ImSig analysis.

**See Also**

[feature\\_select](#)

**Examples**

```
gene_stat (exp = example_data, r = 0.7)
```

---

imsig	<i>Estimate the relative abundance of tissue-infiltrating immune subpopulations abundances using gene expression data</i>
-------	---

---

**Description**

Estimates the relative abundance of immune cells across patients/samples.

**Usage**

```
imsig(exp, r = 0.6, sort = TRUE, sort_by = "T cells")
```

**Arguments**

**exp** Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- `head(example_data)`: [example\\_data](#).

<code>r</code>	Use a value between 0 and 1. Default is 0.6. This is a user defined correlation cut-off to perform feature selection ( <a href="#">feature_select</a> ). Feature selection aids to enrich the prediction of relative abundance of immune cells by filtering off poorly correlated ImSig genes. To get an idea of what cut-off to use check the results of ( <a href="#">gene_stat</a> ) and choose a cut-off that displays high median correlation and maintains a high proportion of genes after feature selection.
<code>sort</code>	Sort the samples based on abundance of a particular cell type. 'Set sort = FALSE' if you wish not to apply sorting. By default the function sorts by abundance of T cells. The cell type of interest for sorting can be controlled by the 'sort_by' parameter.
<code>sort_by</code>	Can be used to sort the samples by predicted abundance of a particular cell type. All other cell types follow this sorting. By default it is sorted by 'T cells'

**Value**

Relative abundance of immune cells across samples. Returns a dataframe.

**See Also**

[feature\\_select](#), [example\\_data](#)

**Examples**

```
cell_abundance = imsig (exp = example_data, r = 0.7, sort=TRUE, sort_by='T cells')
head(cell_abundance)
```

---

<code>imsig_survival</code>	<i>Survival analysis based on relative abundance of immune infiltration estimated by ImSig</i>
-----------------------------	--

---

**Description**

Patients are split into two groups based on their immune cell abundance (median abundance value) and a regular survival analysis is carried out.

**Usage**

```
imsig_survival(exp, cli, time = "time", status = "status", r = 0.6)
```

**Arguments**

<code>exp</code>	Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- <code>head(example_data)</code> : <a href="#">example_data</a> .
<code>cli</code>	Clinical metadata containing the event data (dead or alive) and time to event data. Samples names should be in rownames and same as that in the expression file. Check <code>head()</code> of <a href="#">example_cli</a> for an example clinical data.

time	Column name of time-to-event parameter.
status	Column name of event (dead or alive) parameter.
r	Use a value between 0 and 1. Default is 0.6. This is a user defined correlation cut-off to perform feature selection ( <a href="#">feature_select</a> ). Feature selection aids to enrich the prediction of relative abundance of immune cells by filtering off poorly correlated ImSig genes. To get an idea of what cut-off to use check the results of ( <a href="#">gene_stat</a> ) and choose a cut-off that displays high median correlation and maintains a high proportion of genes after feature selection.

**Value**

Hazard Ratio

**See Also**[feature\\_select](#), [example\\_data](#), [example\\_cli](#)**Examples**

```
survival = imsig_survival (exp = example_data, cli = example_cli)
head(survival)
```

---

plot_abundance	<i>Plot relative abundance of immune cells</i>
----------------	--

---

**Description**

Barplots of relative abundance of immune cells across samples. The order of the samples are the same as that of [imsig](#).

**Usage**

```
plot_abundance(exp, r = 0.6)
```

**Arguments**

exp	Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- <code>head(example_data)</code> : <a href="#">example_data</a> .
r	Use a value between 0 and 1. Default is 0.6. This is a user defined correlation cut-off to perform feature selection ( <a href="#">feature_select</a> ). Feature selection aids to enrich the prediction of relative abundance of immune cells by filtering off poorly correlated ImSig genes. To get an idea of what cut-off to use check the results of ( <a href="#">gene_stat</a> ) and choose a cut-off that displays high median correlation and maintains a high proportion of genes after feature selection.

**Value**

ggplot

**See Also**

[feature\\_select](#), [example\\_data](#)

**Examples**

```
plot_abundance (exp = example_data, r = 0.7)
```

---

plot_network	<i>Network graph of ImSig genes</i>
--------------	-------------------------------------

---

**Description**

A Network visualization displays undirected graph structures and highlights the relationships between entities. The nodes are ImSig genes and the edges represent the correlation between them. The nodes are coloured based on cell type. Try using a correlation cut-off of '0' to get a complete picture.

**Usage**

```
plot_network(  
  exp,  
  r = 0.6,  
  pt.cex = 2,  
  cex = 1,  
  inset = 0,  
  x.intersp = 2,  
  vertex.size = 3,  
  vertex.label = NA,  
  layout = layout_with_fr  
)
```

**Arguments**

exp	Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- <code>head(example_data)</code> : <a href="#">example_data</a> .
r	Use a value between 0 and 1. Default is 0.6. This is a user defined correlation cut-off to perform feature selection ( <a href="#">feature_select</a> ). Feature selection aids to enrich the prediction of relative abundance of immune cells by filtering off poorly correlated ImSig genes. To get an idea of what cut-off to use check the results of ( <a href="#">gene_stat</a> ) and choose a cut-off that displays high median correlation and maintains a high proportion of genes after feature selection.

pt.cex	expansion factor(s) for the points.
cex	character expansion factor relative to current par("cex"). Used for text, and provides the default for pt.cex.
inset	inset distance(s) from the margins as a fraction of the plot region when legend is placed by keyword.
x.intersp	character interspacing factor for horizontal (x) spacing.
vertex.size	Node size of network graph
vertex.label	Add gene names to the network graph. Default set to NA.
layout	Layout algorithm to be used for building network. Default set to force-directed layout algorithm by Fruchterman and Reingold. Read documentation of 'igraph' for other available algorithms.

**Value**

Network graph

**See Also**

[feature\\_select](#)

**Examples**

```
plot_network (exp = example_data, r = 0.7)
```

---

plot_survival	<i>Forest plot of survival analysis by ImSig</i>
---------------	--

---

**Description**

Patients are split into two groups based on their immune cell abundance (median abundance value) and a regular survival analysis is carried out. Raw values can be obtained from [imsig\\_survival](#).

**Usage**

```
plot_survival(exp, cli, time = "time", status = "status", r = 0.6)
```

**Arguments**

exp	Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- head(example_data): <a href="#">example_data</a> .
cli	Clinical metadata containing the event data (dead or alive) and time to event data. Samples names should be in rownames and same as that in the expression file. Check head() of <a href="#">example_cli</a> for an example clinical data.

time	Column name of time-to-event parameter.
status	Column name of event (dead or alive) parameter.
r	Use a value between 0 and 1. Default is 0.6. This is a user defined correlation cut-off to perform feature selection ( <a href="#">feature_select</a> ). Feature selection aids to enrich the prediction of relative abundance of immune cells by filtering off poorly correlated ImSig genes. To get an idea of what cut-off to use check the results of ( <a href="#">gene_stat</a> ) and choose a cut-off that displays high median correlation and maintains a high proportion of genes after feature selection.

**Value**

Forest plot

**See Also**

[feature\\_select](#), [example\\_data](#), [example\\_cli](#)

**Examples**

```
plot_survival (exp = example_data, r = 0.7, cli = example_cli, time = 'time', status= 'status')
```

---

pp\_exp                      *Pre-processing expression matrix*

---

**Description**

Subsets the user's dataset based on the genes that are common to the users dataset and ImSig.

**Usage**

```
pp_exp(exp)
```

**Arguments**

exp	Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- <code>head(example_data)</code> : <a href="#">example_data</a> .
-----	--

**Value**

Expression dataframe

---

pp_sig	<i>Pre-processing ImSig file</i>
--------	----------------------------------

---

**Description**

Subsets ImSig genes based on the genes that are common to the users dataset and ImSig

**Usage**

```
pp_sig(exp)
```

**Arguments**

exp	Dataframe of transcriptomic data (natural scale) containing genes as rows and samples as columns. Note: Gene names should be set as row names and duplicates are not allowed. Missing values are not allowed within the expression matrix. Check example- <code>head(example_data)</code> : <a href="#">example_data</a> .
-----	--

**Value**

ImSig dataframe

---

sig	<i>ImSig genes</i>
-----	--------------------

---

**Description**

ImSig signature genes and the cell type they represent

**Usage**

```
sig
```

**Format**

dataframe

# Index

## \* datasets

- example\_cli, 3
- example\_data, 3
- sig, 11

corr\_matrix, 2

example\_cli, 3, 6, 7, 9, 10

example\_data, 2, 3, 4–11

feature\_select, 2, 4, 5–10

gene\_stat, 2, 4, 4, 5–8, 10

imsig, 5, 7

imsig\_survival, 6, 9

plot\_abundance, 7

plot\_network, 8

plot\_survival, 9

pp\_exp, 10

pp\_sig, 11

sig, 11