Tag and Word Clouds

January Weiner january.weiner@gmail.com

2025-06-23

tagcloud.R

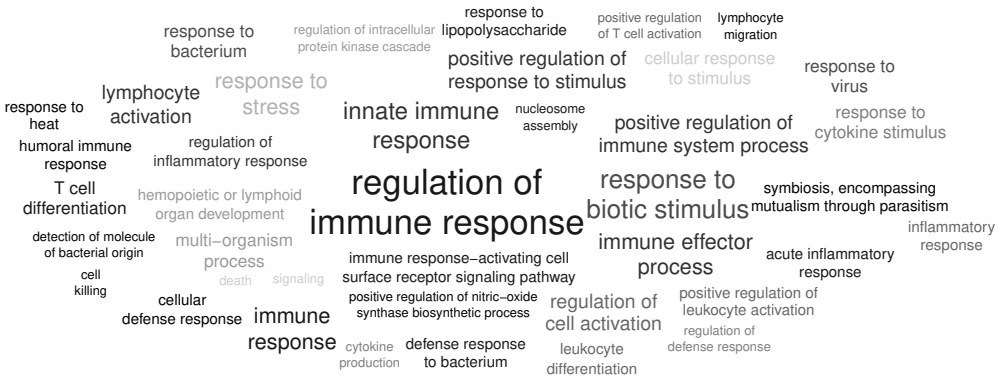
Introduction

`tagcloud` command creates various styles of tag and word clouds. In it simplest form, it takes a character vector (vector of tags) as an argument. Optionally, one can add different weights, colors and layouts. Here is an advanced example of a typical GO-Term cloud, where colors and weights (font size) which correspond to the effect size and P-value, respectively:

```
library(tagcloud)
```

```
## Loading required package: Rcpp
```

```
data(gambia)
tags <- strmultline(gambia$Term)[1:40]
weights <- -log(gambia$Pvalue)[1:40]
or <- gambia$OddsRatio[1:40]
colors <- smoothPalette(or, max=4)
tagcloud(tags, weights=weights, col=colors)
```



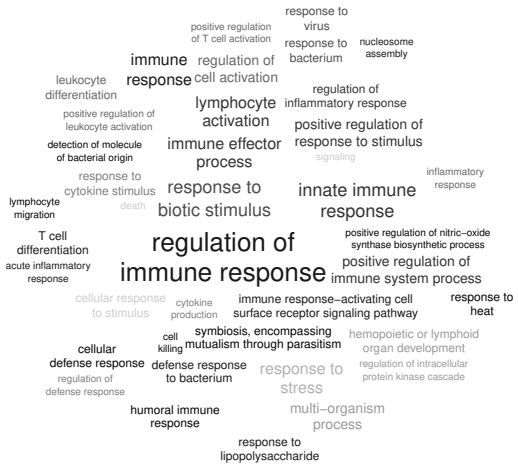
tagcloud.R

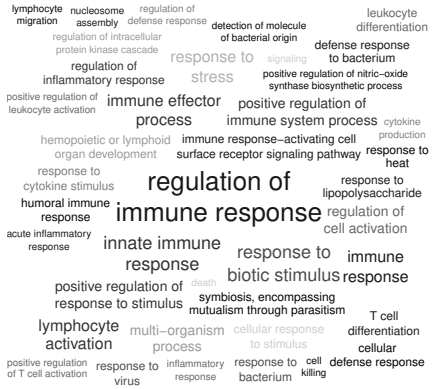
Notes. The geometry of the cloud will reflect the geometry of the plotting area: simply resize the plot and re-run `tagcloud` to get a different look. `smoothPalette` automagically converts a numeric vector into a vector of a color gradient of the same length. `strmultline` breaks long, multi-word lines, which otherwise mess up the figure.

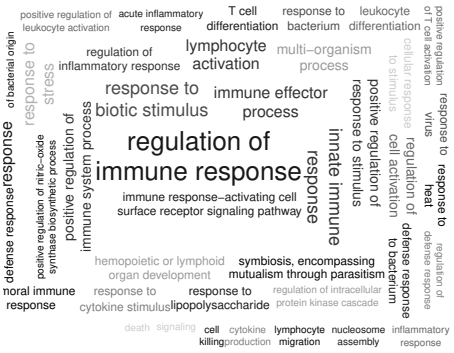
Layouts

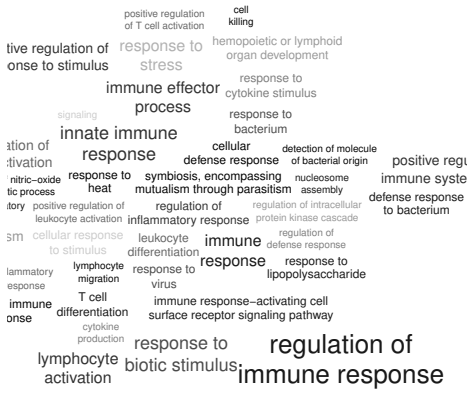
There is a number of algorithms that allow you to create different layouts.

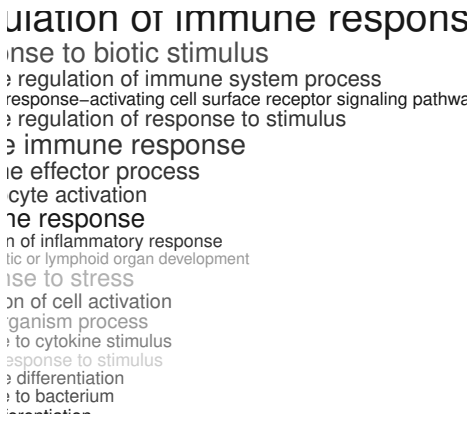
```
par( mfrow=c( 3, 2 ) )
tagcloud(tags, weights=weights, col=colors, algorithm="oval")
tagcloud(tags, weights=weights, col=colors, algorithm="fill")
tagcloud(tags, weights=weights, col=colors, algorithm="snake")
tagcloud(tags, weights=weights, col=colors, algorithm="random")
tags2 <- gambia$Term[1:20]
cols2 <- colors[1:20]
wei2 <- weights[1:20]
tagcloud(tags2, weights=wei2, col=cols2, algorithm="list")
tagcloud(tags2, weights=wei2, col=cols2, algorithm="clist")
```

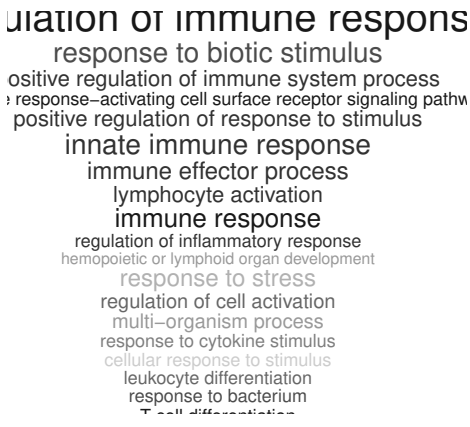

response to virus
positive regulation of T cell activation
response to bacterium
nucleosome assembly
immune response
regulation of cell activation
leukocyte differentiation
regulation of inflammatory response
lymphocyte activation
positive regulation of leukocyte activation
positive regulation of response to stimulus
detection of molecule of bacterial origin
immune effector process
signaling
inflammatory response
response to cytokine stimulus
response to biotic stimulus
innate immune response
lymphocyte migration
death
T cell differentiation
regulation of immune response
positive regulation of nitric−oxide synthase biosynthetic process
acute inflammatory response
positive regulation of immune system process
cellular response to stimulus
cytokine production
immune response−activating cell surface receptor signaling pathway
response to heat
cell killing
symbiosis, encompassing mutualism through parasitism
hemopoietic or lymphoid organ development
cellular defense response
defense response to bacterium
regulation of intracellular protein kinase cascade
response to stress
regulation of defense response
humoral immune response
multi−organism process
response to lipopolysaccharide


lymphocyte migration
nucleosome assembly
regulation of defense response
leukocyte differentiation
detection of molecule of bacterial origin
regulation of intracellular protein kinase cascade
defense response to bacterium
response to stress
signaling
regulation of inflammatory response
positive regulation of nitric−oxide synthase biosynthetic process
positive regulation of leukocyte activation
immune effector process
positive regulation of immune system process
cytokine production
hemopoietic or lymphoid organ development
immune response−activating cell surface receptor signaling pathway
response to heat
response to cytokine stimulus
regulation of immune response
response to lipopolysaccharide
humoral immune response
regulation of cell activation
acute inflammatory response
innate immune response
response to biotic stimulus
immune response
positive regulation of response to stimulus
death
symbiosis, encompassing mutualism through parasitism
T cell differentiation
lymphocyte activation
multi−organism process
cellular response to stimulus
cellular defense response
positive regulation of T cell activation
response to virus
inflammatory response
response to bacterium
cell killing


positive regulation of leukocyte activation
acute inflammatory response
T cell differentiation
response to bacterium
leukocyte differentiation
regulation of inflammatory response
lymphocyte activation
multi−organism process
response to stress
response to biotic stimulus
immune effector process
positive regulation of response to stimulus
regulation of immune response
innate immune response
immune response−activating cell surface receptor signaling pathway
cellular response to stimulus
cell activation
response to virus
response to heat
defense response to bacterium
hemopoietic or lymphoid organ development
symbiosis, encompassing mutualism through parasitism
response to cytokine stimulus
response to lipopolysaccharide
regulation of protein kinase cascade
death
signaling
cell killing
cytokine production
lymphocyte migration
nucleosome assembly
inflammatory response


cell killing
positive regulation of T cell activation
hemopoietic or lymphoid organ development
response to stress
response to cytokine stimulus
immune effector process
response to bacterium
signaling
innate immune response
cellular defense response
detection of molecule of bacterial origin
response to heat
symbiosis, encompassing mutualism through parasitism
nucleosome assembly
regulation of inflammatory response
regulation of intracellular protein kinase cascade
positive regulation of leukocyte activation
cellular response to stimulus
leukocyte differentiation
immune response
lymphocyte migration
response to virus
response to lipopolysaccharide
T cell differentiation
immune response−activating cell surface receptor signaling pathway
cytokine production
response to biotic stimulus
regulation of immune response
lymphocyte activation


ulation of immune respons
nse to biotic stimulus
regulation of immune system process
response−activating cell surface receptor signaling pathwa
regulation of response to stimulus
immune response
e effector process
cyte activation
ne response
n of inflammatory response
tic or lymphoid organ development
ise to stress
on of cell activation
ganism process
to cytokine stimulus
esponse to stimulus
differentiation
to bacterium

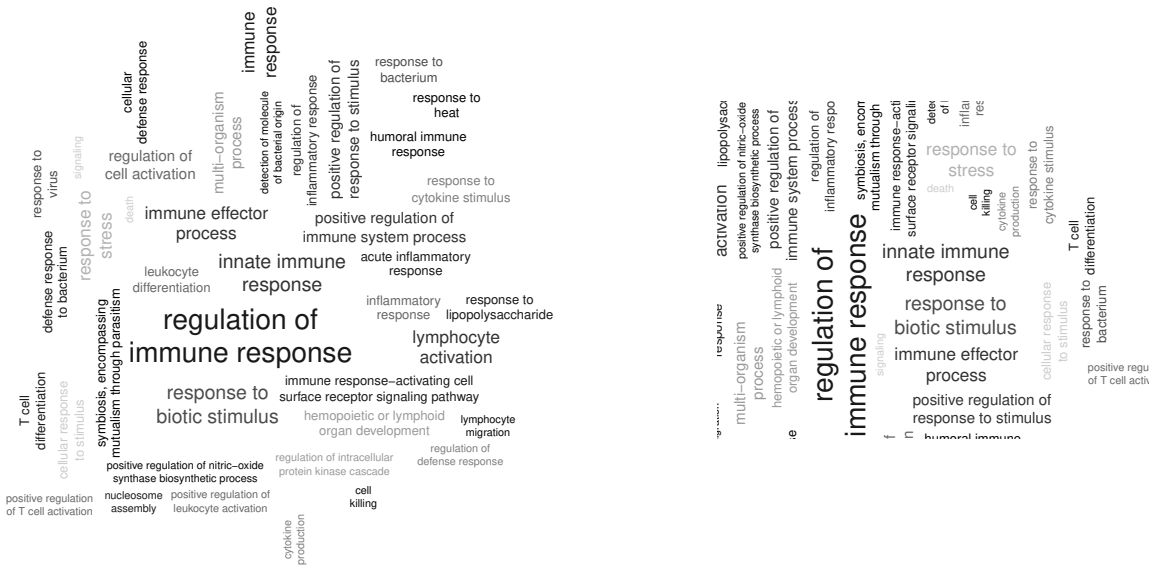

ulation of immune respons
response to biotic stimulus
ositive regulation of immune system process
response−activating cell surface receptor signaling pathw
positive regulation of response to stimulus
innate immune response
immune effector process
lymphocyte activation
immune response
regulation of inflammatory response
hemopoietic or lymphoid organ development
response to stress
regulation of cell activation
multi−organism process
response to cytokine stimulus
cellular response to stimulus
leukocyte differentiation
response to bacterium

tag-

cloud.R

Another parameter to tune is `fvert`, the proportion of tags that are displayed vertically (which is 0 by default).

```
par(mfrow=c(1, 2))
tagcloud(tags, weights=weights, col=colors, fvert=0.3)
tagcloud(tags, weights=weights, col=colors, fvert=0.7)
```



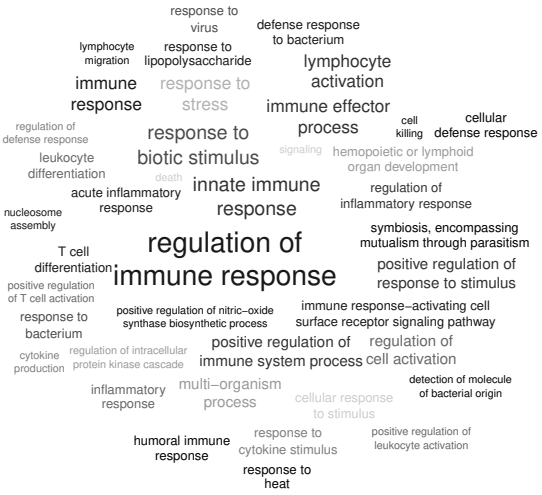
tagcloud.R

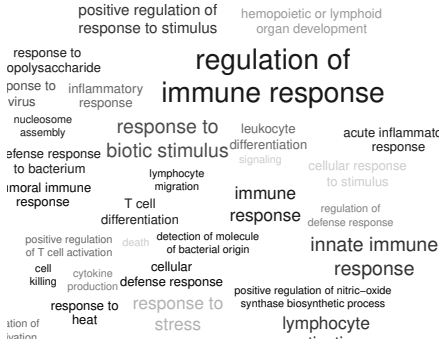
Finally, using the parameter `order` you can also influence the layout of the word cloud:

- `size`: tags are ordered by size, that is, their effective width multiplied by their effective height. Default.
- `keep`: keep the order from the list of words provided
- `random`: randomize the tag list
- `width`: order by effective screen width
- `height`: order by effective screen height

Starting with the tag with the largest weight typically makes this tag at the center of the cloud. Sometimes, however, a randomized order results in a more interesting output.

```
par(mfrow=c(1, 2))
tagcloud(tags, weights=weights, col=colors, order="size")
tagcloud(tags, weights=weights, col=colors, order="random")
```





tagcloud.R

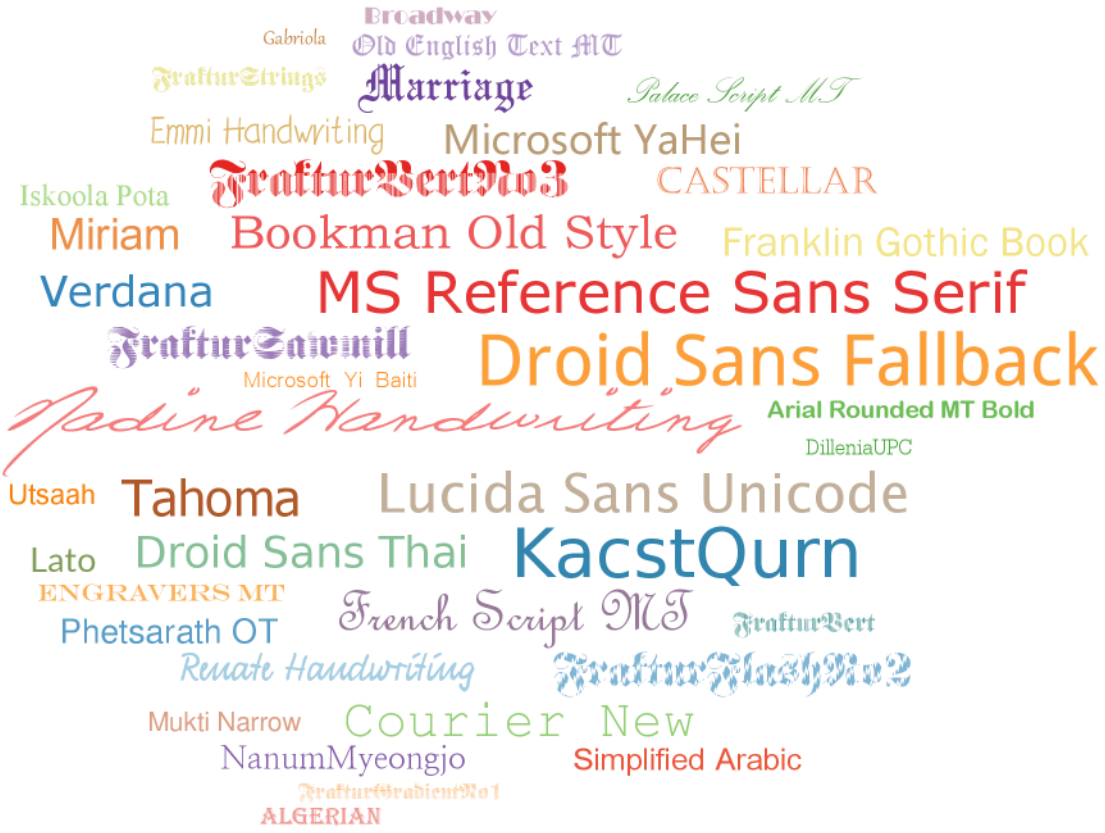
Fonts

Using the parameter `family`, you can specify the font family to be used. In the following, we use the excellent `extrafont` package[1]. However note that to produce correct PDFs, you should use the cairo engine, for example with `dev.copy2pdf( out.type="cairo", ... )`. Alternatively, use the `png()` device.

```
library(extrafont)
library(RColorBrewer)
fnames <- sample(fonts(), 40)
fweights <- rgamma(40, 1)
fcolors <- colorRampPalette( brewer.pal( 12, "Paired" ) )( 40 )
tagcloud( fnames, weights=fweights, col=fcolors, family=fnames )
```

tagcloud.R

[1] After installing the package, run `font_import` to import the fonts installed on the system

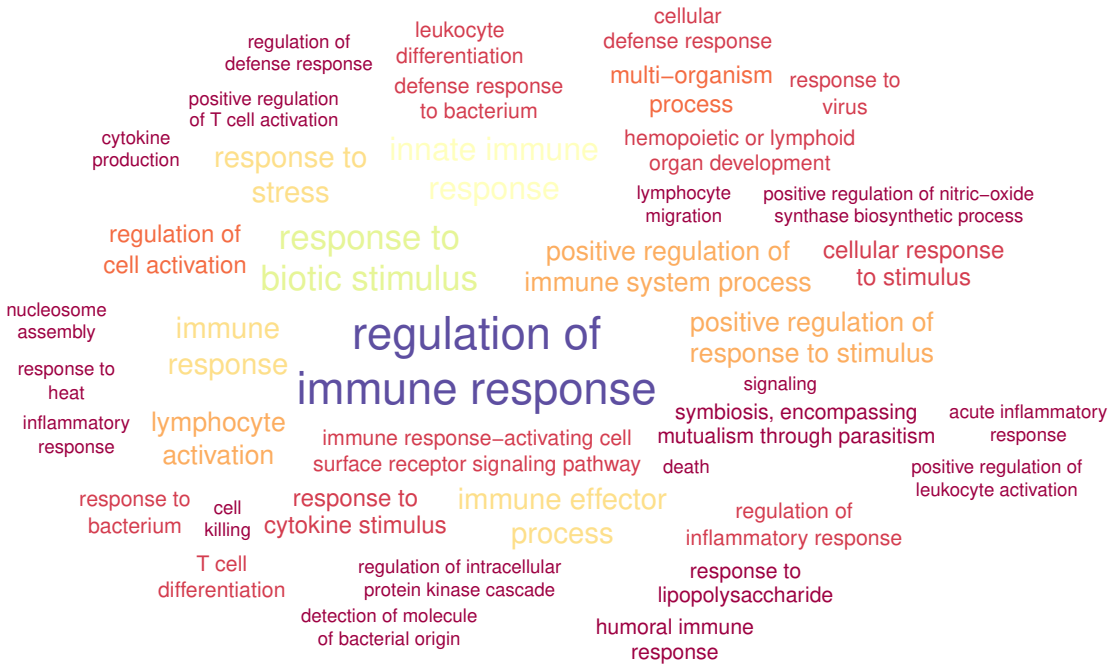


Colors

Using the tools `smoothPalette`, you can easily map a numeric vector onto colors. `smoothPalette` by default produces a grey-black gradient, but anything goes with the help of `RColorBrewer`. `smoothPalette` either takes a predefined palette (it will not expand it, however, so if you define three colors, three colors will be on the figure, no extrapolated colors in between), or an RColorBrewer palette.

In the example below, the weights are on purpose correlated to the color.

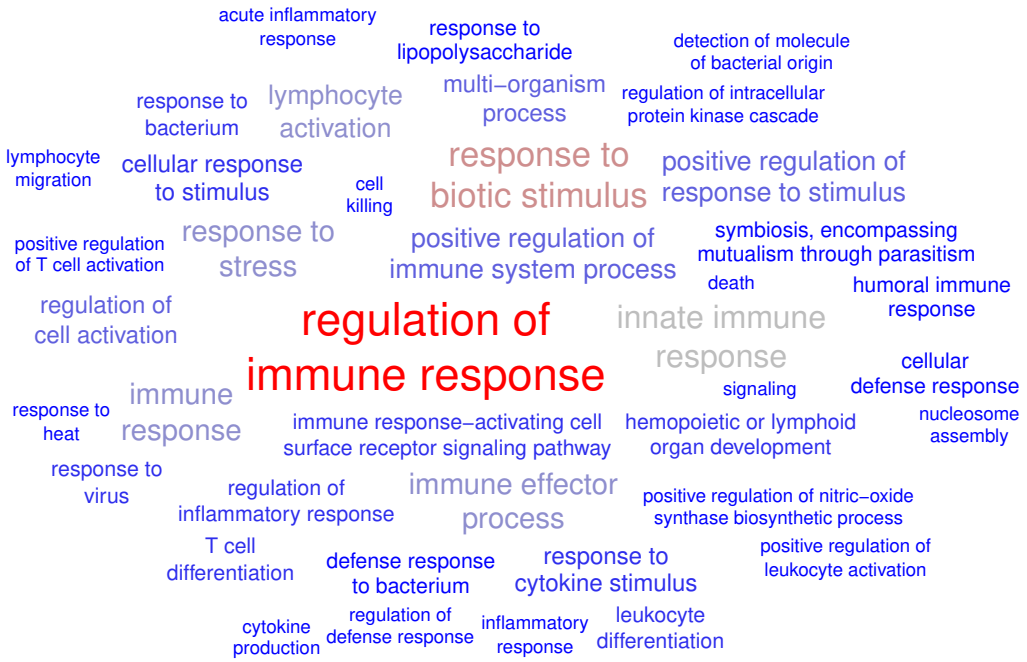
```
library(RColorBrewer)
colors <- smoothPalette(weights, pal= brewer.pal( 11, "Spectral" ) )
tagcloud(tags, weights=weights, col=colors, order="size")
```



tagcloud.R

Alternative way to specify the colors is to provide a function that can generate a palette – for example, the return value of `colorRampPalette`. This has the advantage that `smoothPalette` will generate, with the palette function, as many color steps as necessary.

```
palf <- colorRampPalette( c( "blue", "grey", "red" ) )
colors <- smoothPalette(weights, palfunc= palf )
tagcloud(tags, weights=weights, col=colors, order="size")
```


acute inflammatory response
response to lipopolysaccharide
detection of molecule of bacterial origin
response to bacterium
lymphocyte activation
multi−organism process
regulation of intracellular protein kinase cascade
lymphocyte migration
cellular response to stimulus
cell killing
response to biotic stimulus
positive regulation of response to stimulus
positive regulation of T cell activation
response to stress
positive regulation of immune system process
symbiosis, encompassing mutualism through parasitism
death
humoral immune response
regulation of cell activation
regulation of immune response
innate immune response
cellular defense response
signaling
response to heat
immune response
immune response−activating cell surface receptor signaling pathway
hemopoietic or lymphoid organ development
nucleosome assembly
response to virus
regulation of inflammatory response
immune effector process
positive regulation of nitric−oxide synthase biosynthetic process
T cell differentiation
defense response to bacterium
response to cytokine stimulus
positive regulation of leukocyte activation
cytokine production
regulation of defense response
inflammatory response
leukocyte differentiation

tagcloud.R