

PAE_plot_MRGPRX2

KimPankyung

2022-08-10

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com> (<http://rmarkdown.rstudio.com>).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
##Citation
#Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589 (2021). https://doi.org/10.1038/s41586-021-03819-2
#Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Židek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, Sameer Velankar, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, Nucleic Acids Research, Volume 50, Issue D1, 7 January 2022, Pages D439-D444, https://doi.org/10.1093/nar/gkab1061
```

```
#Install the package for read the json file
#install.packages("jsonlite")
#install.packages("curl")
#Load the package
library(jsonlite)
library(curl)
```

```
## Using libcurl 7.64.1 with Schannel
```

```
library(ggplot2)

#Check the directory
getwd()
```

```
## [1] "D:/R/PAE"
```

```
#Download the raw PAE file in alphafold website
#Set the file name
my="AF-A0A6J3EQU8-F1-predicted_aligned_error_v3.json"

#Read the json file in R
jdata<-fromJSON(my)

#Bring the PAE data in the first column
data<-jdata[1]

#Use unlist() to make PAE data in numeric value
#R read the json file in list format
L<-unlist(data)

#Convert it to dataframe
DL<-data.frame(L)
#Extract the numeric value of PAE
NL<-DL[,1]

#Prepare the matrix
X <- seq(1,330) #Use the seq() to make x value same as sequence number of protein
Y <- seq(1,330) #Use the seq() to make y value same as sequence number of protein
gd <- expand.grid(Scored_residue=X, Aligned_residue=Y) #Make grid coordinates
gd$PAE <-NL #Add PAE data

#Draw the plot
ggplot(gd, aes(x=Scored_residue, y=Aligned_residue, fill=PAE)) + #First value is dataframe, fill
#I will be the PAE data
  geom_tile()+ #Select the plot type
  scale_x_continuous(trans = "reverse",breaks = c(0,50,100,150,200,250,300),position = "bottom"
) + #Change the order with trans=
  #Change the range display with breaks=
  #Change the position with position=
  scale_y_continuous(breaks = c(0,50,100,150,200,250,300)) +
  scale_fill_continuous(limits=c(0,32),low = "darkgreen", high = "white")+ #Change the range of
#legend with limits=
  #Change the color with low=, high=
  coord_flip() #Change the x-axis and y-axis
```

