Alluvian plot of Protein Isoform Localization Prediction

Vi Dang

2022-09-15

```r
#Load packages
library(tidyverse)
library(ggalluvial)
```

```r
#Load data of TSS clusters and protein isoform localization predicted by Deeploc 2.0
loc_H99<-read.csv("D:/phD/TSS cluster/Localization signals/manually analysis with Deeploc/220 H99 aATG
```

```r
#Filter for genes where the Long and Short isoforms have different localization
loc_H99_diff<-loc_H99%>%filter(Long!=Short)
#Number of gene
nrow(loc_H99_diff)
```

```
## [1] 87
```

```r
#Construct a contingency table
loc_contingency_table<-table(loc_H99_diff$Long,loc_H99_diff$Short)%>%as.data.frame()
colnames(loc_contingency_table)<-c("LONG","SHORT","Frequency")
head(loc_contingency_table)
```

```
##                         LONG          SHORT Frequency
## 1              Cytoplasm Cell membrane         0
## 2 Endoplasmic reticulum Cell membrane         1
## 3          Extracellular Cell membrane         1
## 4        Golgi apparatus Cell membrane         0
## 5       Lysosome/Vacuole Cell membrane         0
## 6          Mitochondrion Cell membrane         0
```

```r
#Plot the alluvian plot
ggplot(data = loc_contingency_table,
       aes(axis1 = LONG,    # First variable on the X-axis
           axis2 = SHORT,
           y = Frequency)) +
  geom_alluvium(
    aes(fill = LONG),
    #    width        = 0,
    #    knot.pos     = 0,
    reverse      = FALSE)  +
  geom_stratum(width        = 1/3.5  ,
               reverse      = FALSE,
               color        = '#222222') +
```

```r
geom_text(stat = "stratum", aes(label = after_stat(stratum)),
          reverse       =  FALSE) +
scale_x_continuous(breaks        =  1:2,
                   labels        =  c("LONG isoform","SHORT isoform")
) +                                                           # Uncomment to plot from bottom to top
theme_minimal()+
theme(axis.text.x.bottom = element_text(size=17),
      axis.title.y.left = element_text(size=20),
      legend.position = "None")
```

```r
geom_text(stat = "stratum", aes(label = after_stat(stratum)),
          reverse       =  FALSE) +
scale_x_continuous(breaks        =  1:2,
```