

R scripts for the additional analysis based on familiarity (included in the discussion)

Note: if any CSV file is imported in R as a data frame, it should be converted to a numeric matrix in order to use the following functions.

For example, if the HaFF_R_rpf CSV file is imported as a data frame using File/Import Dataset/From Text, it should be converted to a numeric matrix using `HaFF_R_rpf<-data.matrix(HaFF_R_rpf)`.

- 1) Function to obtain a matrix with recall and precision values of the features that have a familiarity value equal or greater than a threshold.

```
getRowsFromFamiliarity<-function(m,numberOfCols,numberOfRows,threshold)
{
  mymat=m
  out <- matrix(NA,nrow=0,ncol=numberOfCols,byrow = TRUE)
  i=1
  while(i<=numberOfRows) {
    row=i
    if(mymat[row,3]>=threshold) {
      out <- rbind(out, c(mymat[row,]))
    }
    i=i+1}
  colnames(out)<-c("Recall", "Precision", "Familiarity");
  return(out)
}
```

Example to execute the function:

```
HaFF_R_familiarity<-getRowsFromFamiliarity(HaFF_R_familiarity_all,3,58,4)
```

- 2) Function that prints the number of cases where the recall and precision values of a feature in the baseline outperforms the recall and precision values in HaFF when the familiarity value is equal or greater than a threshold. As a result, a matrix is obtained where each row (if any) contains number of the feature that fulfills the case.

```
getFamiliarRowsBaseOutperforms<-
function(matrixBase,matrixHaFFfamiliarity,numberOfRows,threshold)
{
  out <- matrix(NA,nrow=0,ncol=1,byrow = TRUE)
  i=1
  cases=0
  while(i<=numberOfRows) {
    row=i
    if(matrixHaFFfamiliarity[row,3]>=threshold) {
      if(matrixBase[row,1]>matrixHaFFfamiliarity[row,1] &&
matrixBase[row,2]>matrixHaFFfamiliarity[row,2]) {
        out <- rbind(out,row)
        cases=cases+1
      }
    }
    i=i+1}
}
```

```

    print(cases);
    colnames(out)<-c("FeatureId");
    return(out)
}

```

Example to execute the function:

```

FamiliarRowsBaseOutperforms<-getFamiliarRowsBaseOutperforms(Baseline_R_rpf,
HaFF_R_familiarity_all,58,4)

```

3) Function that prints the number of cases (i.e., features) that have the familiarity value lower than a threshold in HaFF and its recall and precision values outperforms the values of the baseline. As a result, a matrix is obtained where each row (if any) contains number of the feature that fulfills the case.

```

getNonFamiliarRowsOutperformsBase<-
function(matrixBase,matrixHaFFfamiliarity,numberOfRows,threshold)
{
    out <- matrix(NA,nrow=0,ncol=1,byrow = TRUE)
    i=1
    cases=0
    while(i<=numberOfRows) {
        row=i
        if(matrixHaFFfamiliarity[row,3]<threshold ) {
            if(matrixHaFFfamiliarity[row,1]>matrixBase[row,1]  &&
            matrixHaFFfamiliarity[row,2]>matrixBase[row,2]) {
                out <- rbind(out,row)
                cases=cases+1
            }
        }
        i=i+1}
    print(cases);
    colnames(out)<-c("FeatureId");
    return(out)
}

```

Example to execute the function:

```

NonFamiliarRowsOutperformsBase <-
getNonFamiliarRowsOutperformsBase(Baseline_R_rpf,
HaFF_R_familiarity_all,58,4)

```