

# Maria Helena Thomaz Maia

Sou aluna de doutorado da Pós-graduação em Biologia (Genética) orientada por Diogo Meyer. Meu projeto tem como foco a investigação de padrões de evolução na região promotora de genes da imunidade (genes HLA) em humanos e esta sendo desenvolvido no Laboratório de Genética e Biologia Evolutiva de Vertebrados do Instituto de Biociências da USP-SP.

## Meus exercícios

Aula 1 [thomazmaia\\_01\\_F](#)

Aula 2 [thomazmaia\\_02\\_F](#)

Aula 3 [thomazmaia\\_03\\_F](#)

Aula 4 [thomazmaia\\_04\\_F](#)

Aula 5 [thomazmaia\\_05\\_F](#)

Aula 6 [thomazmaia\\_06\\_F](#)

Aula 7 [thomazmaia\\_07\\_F](#)

Aula 8 [thomazmaia\\_08\\_F](#)

Aula 9 [thomazmaia\\_09\\_F](#)

## Propostas

### Principal

#### Contexto biológico

Os genes codificantes de receptores imunoglobulina símiles de células *natural killer* (*killer cell immunoglobulin-like receptor*) (KIR) ocupam um segmento de 150 kb do cromossomo 19 em humanos (Fig 1). O conteúdo gênico (a presença ou ausência de cada gene) é um dos fatores que conta na determinação do haplótipo *KIR* de um indivíduo. Basicamente, os haplótipos são classificados como A ou B, dependendo da presença ou ausência de um ou mais genes *KIR* estimulatórios. No caso do haplótipo A a definição clássica se dá pela ausência dos genes *KIR2DL2*, *KIR2DL5*, *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS5* e *KIR3DS1*, que são característicos do haplótipo B.



Fig.1: Estrutura do complexo de genes *KIR* no cromossomo 19 ([Single et al 2008](#))

#### Proposta:

Minha proposta é desenvolver uma função capaz identificar os haplótipos A e B e de calcular suas frequências a partir de bancos de genotipagem por presença e ausência de genes *KIR*. Além disso, calcular a frequência de cada gene. Para que essas operações sejam realizadas, deve-se assumir que as populações analisadas estarão em equilíbrio de Hardy-Weinberg.

### Argumentos:

X = objeto da classe `data.frame` contendo os resultados de presença e ausência de cada gene *kir*.

A função considerará todos os 16 genes *KIR* já descrito: *KIR2DL1*, *KIR2DL2*, *KIR2DL3*, *KIR2DL4*, *KIR2DL5*, *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS4*, *KIR2DS5*, *KIR3DS1*, *KIR3DL1*, *KIR3DL2*, *KIR3DL3*, *KIR3DP1*, *KIR2DP1*.

Os arquivos de entrada devem conter colunas de resultados para TODOS os genes, mesmo que eles não tenham sido testados, de tal forma que, para cada indivíduo, cujo gene não foi testado, será atribuído o valor 0 (ausência) na célula correspondente.

### Operações:

1- Selecionar os genótipos AA (homozigotos para o haplótipo A) pela exclusão de genes do haplótipo B: *KIR2DL2*, *KIR2DL5*, *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS5* e *KIR3DS1*.

2- Calcular a frequência de cada Haplótipo a partir do resultado da operação 1

Frequência do Haplótipo A =  $\sqrt{\text{Frequência do genótipo AA na população}}$   
Frequência do Haplótipo B = 1 - Frequência do Haplótipo A

3- Calcular a frequência observada (FO) e frequência esperada em equilíbrio (FE) de cada um dos genes presentes no banco de dados.

$F_{0i}$  = Frequência absoluta do gene *i* / N  
 $F_{Ei} = 1 - \sqrt{1 - F_{0i}}$   
Onde:  
N = Tamanho amostral

### Motivação:

Essas operações são rotineiras para quem trabalha com genes *KIR*. Criar uma função que incluía todas elas, agiliza o trabalho de análise de dados.

## Comentário

A contextualização biológica está ótima. Entretanto, não tenho como avaliar a dificuldade que irá enfrentar na elaboração da sua função. Não consigo vislumbrar os dados de entrada e como os haplótipos serão confrontados. Faltou dar uma ideia do formato dos dados de entrada e das operações que serão executadas. Uma boa partida é definir isso junto com os argumentos que a função terá e especificar claramente o objeto de saída. — [Alexandre Adalardo de Oliveira](#) 2010/03/31 11:31

## Página de Ajuda

KIR.summ  
Documentation

package:NA

R

KIR Summary

Description:

Função exclusiva para dados de presença e ausência de KIR.

Retorna um sumário com informações sobre frequência de Haplogrupos (A e B), frequência dos genes KIR e o número de diferentes perfis (combinações de presença e ausência de genes) e ainda mostra os perfis diferentes identificados.

As frequências, tanto de haplótipos quanto de genes, são calculadas sob a premissa de que as populações analisadas estão em equilíbrio de Hardy-Weinberg.

Usage:

KIR.summ(x)

Arguments:

x: objeto da classe data.frame

Details:

x: objeto da classe data.frame com 17 colunas: a primeira com a ID das amostras testadas e as outras 16 com os resultados de presença e ausência para cada um dos 16 genes KIR comumente testados, que são: KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR2DL5, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DS5, KIR3DS1, KIR3DL1, KIR3DL2, KIR3DL3, KIR3DP1, KIR2DP1.

IMPORTANTE: A ordem dos genes no data.frame deve ser EXATAMENTE essa!

Value:

A função KIR.summ retorna uma lista com 5 componentes:

GeneFreq: Uma tabela com as frequências dos genes carregados pelos indivíduos e as frequências esperadas para cada gene KIR sob a premissa de

que as populações estão em equilíbrio de Hardy-Weinberg.

HaplogroupFreq: A frequência absoluta dos genótipos dos haplogrupos (AA e AB/BB), a frequência relativa dos haplótipos A e B sob a premissa de que a população está em equilíbrio de Hardy-Weinberg. Além disso, mostra o intervalo de confiança de 95% das estimativas de frequência dos haplótipos A e B

AA\_Genotype\_Individuals: é uma lista contendo os indivíduos da população/amostra que são homozigotos para o haplogrupo A.

Profiles: mostra todas as combinações diferentes de presença e ausência de genes KIR encontradas na amostra

Profiles\_Count: Número de diferentes combinações encontradas.

Warning: A função retornará uma mensagem de erro caso o numero de colunas de x seja diferente de 17.

IMPORTANTE: A ordem dos genes no data.frame deve ser EXATAMENTE essa: KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR2DL5, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DS5, KIR3DS1, KIR3DL1, KIR3DL2, KIR3DL3, KIR3DP1, KIR2DP1.

PS: Sei que tem uma forma melhor de fazer isso, mas ainda não falo a lingua do R fluentemente....

Author(s):

Maria Helena T. Maia <mariahtmaia@usp.br>

Collaborator(s):

Barbara D. Bitarello

Rodrigo Ramalho

Diogo Meyer

Janaina Vasconcelos

Eduardo Santos

References:

SINGLE, R.M., MARTIN, M.P., MEYER, D., GAO, X., CARRINGTON, M., 2008 Methods for assessing gene content diversity of kir with examples from global set of

populations. Immunogenetics 60: 711-725.

Allele Frequencies in Worldwide Populations:  
<http://www.allelefrequencies.net/default.asp>

See Also:

Examples:

#Utilizar o arquivo de exemplo no fim da página.

KIR.summ(x)

## Código da Função

```
KIR.summ=function(x)

#Verificando se o formato do arquivo de entrada esta correto. Ele precisa
ter 17 colunas - Ver HELP)

{
  if(ncol(x)!=17)
  {
    cat("ERROR! O número de colunas de seu arquivo é diferente de 17|Ver
HELP\n")
    stop()
    break
  }

#=====
#  IDENTIFICANDO HAPLOGRUPOS
#=====

#1º: Criar objeto para colocar os resultados de if e else

  resulta=matrix(nrow=dim(x)[1],ncol=2)
  head(resulta)
  j=dim(x)[1]
  for (i in 1: j)

#2º: Selecionando homozigotos AA

  {
    if (x[i,3]==0 & x[i,6]==0 & x[i,7]==0 & x[i,8]==0 & x[i,9]==0 &
x[i,11]==0 & x[i,12]==0 & x[i,15]==0 & x[i,16]==0 & x[i,17]==0 )
    {
      resulta[i,1:2]=c(as.character(x[i,1]),"AA")    #Guardando o resultado
de if, ou seja, todas as combinações que atenderam a condição
```

```
}
else
  resulta[i,1:2]=c(as.character(x[i,1]),"AB/BB") #Guardando os
resultados que não atenderam a condição no mesmo objeto
}

#=====
#IDENTIFICAR OS INDIVÍDUOS AA
#=====

AAGenotype=as.character(as.data.frame(resulta)$V1[as.data.frame(resulta)$V2=
="AA"])

#=====
#CONTAR PERFIS PRESENTES NA AMOSTRA
#=====

Profile=unique(x[,2:17])
Profile_Number=dim(Profile)[1]

#=====
# FREQUÊNCIA DOS HAPLOGRUPOS AA E AB/BB
#=====

result=table(resulta[,2])
result=as.data.frame(result)
colnames(result)= c("Haplogroup", "HaplogroupFreq")

#=====
# FREQUÊNCIA DOS HAPLÓTIPOS A E B
#=====

total=result[1,2]+ result[2,2]
fa=sqrt(result[1,2]/total)
fb=1-fa
sda= 1.96*(sqrt((1-fa)/(4*total))) #Cálculo do Intervalo de Confiança de
95%
sdb= 1.96*(sqrt((1-fb)/(4*total)))
result$Haplotype= c("A", "B")
result$HaplotypeFreq=c(fa, fb)
result$CI=c(sda, sdb)

#=====
# FREQUÊNCIA DOS GENES KIR:
#=====

#1º: Criar objeto para colocar o resultado da ciclagem

obs=dim(x)[2]-1
for (i in 2:17)
```

```
{
  obs[i]=c(mean(x[,i], na.rm=T))
}
carrier=obs[2:17]
names=c("KIR2DL1", "KIR2DL2", "KIR2DL3", "KIR2DL4", "KIR2DL5",
"KIR2DS1", "KIR2DS2", "KIR2DS3", "KIR2DS4", "KIR2DS5", "KIR3DS1", "KIR3DL1",
"KIR3DL2", "KIR3DL3", "KIR3DP1", "KIR2DP1")
expected= 1-sqrt(1-carrier)
fgen=as.data.frame(carrier)
fgen$expected=expected
rownames(fgen)=names

#=====
# OBJETO FINAL
#=====

  final.obj<-list(GeneFreq=fgen,HaplogroupFreq=result,
AA_Genotype_Individuals=AAGenotype, Profiles=Profile,
Profiles_Count=Profile_Number)
  return(final.obj)
}
```

## Arquivos para Teste da Função

Arquivo para ser usado em testes [kird.csv](#)

PS:O separador é";" e não ","

From:

<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:

[http://ecor.ib.usp.br/doku.php?id=05\\_curso\\_antigo:r2010:alunos:trabalho\\_final:mariahtmaia:start](http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2010:alunos:trabalho_final:mariahtmaia:start)



Last update: **2020/08/12 06:04**