

William Bertani Torres



Mestrando em Aconselhamento Genético pelo IB-USP

O título do meu trabalho é “Sequenciamento de nova geração e sua aplicação no aconselhamento genético na Síndrome de Waardenburg”

<http://lattes.cnpq.br/2401488457657302>

[Meus Exercícios](#)

Proposta A:

A função irá calcular a probabilidade de nascer uma criança afetada a partir de Herança Autossômica Recessiva, utilizando-se da frequências do gene (x), levando-se em conta também os possíveis cruzamentos entre genótipos heterozigotos e homozigotos.

Ao final será dado uma lista com a probabilidade para cada cruzamento, o risco final único e uma tabela exemplificando o risco em função de valores pre-estabelecidos de q e s (coeficiente de seleção) para observar a mudança do risco de acordo com a mudança desses dois parametros.

Essa função é útil no aconselhamento genético de famílias onde há risco de ocorrência/recorrência de síndrome genética com herança autossômica recessiva pois fornece de forma rápida calculos de probabilidade muito avançados e complexos que levariam muito tempo para serem feitos à mão.

A idéia para essa função foi baseada na aula do Professor Dr. Paulo Otto, onde ele discorre sobre este assunto no livro (http://www.ib.usp.br/~otto/downloads/pdf_otto/riesgos-otto.pdf) (página 6)

Planejamento da função:

Entrada: a função ira requerer a frequencia do gene e o coeficiente de seleção.

Argumentos: x = Frequencia do gene recessivo (se 1%, então $x=1$) s = Coeficiente de seleção (varia de 0 a 1)

Pseudo - codigo:

1- insere em “ q ” o valor da frequencia do gene em porcentagem ($q=x/100$) 2- tira-se o valor de “ p ” a partir do valor de “ q ” ($p=1-q$) 3- é adicionado as formulas para o valor adaptativo (w) e taxa de mutação (μ) ($w=1-s$ e $\mu=s*q^2$) 4- a seguir são utilizadas as formulas do Professor Paulo Otto para

estipular as probabilidades desejadas para cada cruzamento: $AA \times AA = (p^4)(\mu^2)$ $AA \times Aa = (2p^3)q\mu$ $AA \times aa = 2(p^2)(q^2)\mu w$ $Aa \times Aa = (p^2)(q^2)$ $Aa \times aa = 2p(q^3)w$ $aa \times aa = q^4w^2$ 5-o resultado final faz o calculo das probabilidades de haver criança afetada para cada tipo de cruzamento, fornece o risco final geral, e será gerado uma tabela exemplificando a variação do risco quando se variam os dois parametros da função.

Saída: - probabilidades de haver criança afetada para cada tipo de cruzamento - risco final geral - tabela exemplificando a variação do risco quando se variam os dois parametros da função.

Proposta B:

A função irá filtrar variantes em tabelas geradas pela tecnica do sequenciamento do exoma **a partir do modo de herança** (dominante ou recessivo (0/1 e 1/1 respectivamente)) e **da localização cromossomica** (se em autossomos ou se em cromossomos sexuais). Como default, a função filtra os **tipos de mutação** mais comumente pertinentes (mutações em exons e regioao de splicing). Ainda é possível filtrar a **frequencia** dessas variantes nos bancos de dados, que na função é representado pelo banco exac.

Para sua utilização, é necessario que o usuario ajuste seu arquivo de input ao que é requerido pela função.

Essa função é uma versão resumida do que uma função para filtrar exoma faria, já que a tabela gerada pela técnica é muito extensa e seria inviável a sua conclusão no curto espaço de tempo disponível pela matéria, mas demonstra que a função pode ser expandida para os outros parâmetros que utilizariam codigos muito similares, alterando-se apenas os caracteres os quais desejam-se filtrar (um exemplo disso são os diversos outros bancos de dados de frequencia, cuja filtragem seria também abaixo de 1%, logo, pouco seria alterado em termos de código para essas outras colunas).

A função é de extrema importância na busca por variantes patogênicas que expliquem quadros sindrômicos não solucionados por técnicas tanto mais simples quanto mais baratas, mas que no entanto, sua interpretação é dificultada pela grande abrangência da técnica, que retorna todas as variantes encontradas no indivíduo, inclusive as comuns não patogênicas sem relevância para a clínica.

Planejamento da função:

Entrada: a função irá requerer o nome do arquivo ou o caminho dos diretorios até o arquivo, a indentificação do que se quer filtrar quanto aos cromossomos, se autossomico ou sexual, o modo de herança, se dominante ou recessivo, e ainda pode-se escolher o valor que se quer filtrar para a frequencia da variante nos bancos de dados.

Argumentos:

tabela: requer o nome do arquivo ou o caminho dos diretorios até o arquivo e deve estar entre áspas cromossomo deve ser indicado se a doença que se busca é autossômica ou ligada aos cromossomos sexuais heranca deve ser indicado o modo de herança da doença, se dominante ou recessiva exac valor a ser dado para a filtragem da frequencia das variantes nos bancos de dados (padronizado em abaixo de 1%)

Pseudo código:

1- le o arquivo onde se encontram as variantes 2- filtra a coluna cromossomo a partir de autossomico ou sexual (para autossomico, pega-se aqueles de 1 a 22, para sexual, pega-se X ou Y apenas) 3- filtra a coluna do paciente, que indica a presença em um dos alelos (dominante 0/1) ou em ambos os alelos (recessiva 1/1) ou a ausência (0/0) da variante. 4- estabelecido como default, os tipos de mutação a serem filtradas são as mutações que ocorrem em exons e em região de splicing. 5- a filtragem da frequência das variantes nos bancos de dados também está estabelecida em abaixo de 1%, mas a função permite que esse valor seja alterado quando a função é requisitada

Saída:

Tabela com as linhas que preenchem os critérios de filtragem estabelecidos. Referencia:

ERGUN, M. A new method for analysis of whole exome sequencing data (SELIM) depending on variant prioritization. Informatics in Medicine Unlocked 8 (2017) 51-53.

Achei a proposta A simples demais, e também não entendi como ela pode ser generalizável para vários tipos de usuários.

Meu conselho é que vc torne o plano A mais amplo (por exemplo, fornecendo opções de mais análises e interpretações -que vão exigir manipulações diferentes e mais complexas -, mas que sejam relevantes aos mesmos usuários e que usem dados parecidos ou os mesmos) e apresente algum plano B, que está ausente no arquivo .otf que vc linkou em sua wiki.

É muito importante também que você poste seus planos diretamente na wiki, e não na forma de um arquivo .otf (cheque a página do [trabalho final](#) para instruções completas e detalhadas)

Para ajudar um pouco mais nesse detalhamento, no dia 15 ou 16 vou entrar aqui de novo pra olhar as modificações do seu pseudo código e te dar mais feedbacks. Depois desse prazo, vc também pode tentar me contatar por whatsapp (por favor não mande áudio) (11) 9-9199-3842

Matheus Januario

Trabalho final

Minha função >

```
#tabela de exemplo
tabela <- data.frame("cromossomo" = 1:8, "mutacao" = c("exonica", "exonica",
"splicing", "intronica", "intronica", "exonica", "splicing", "intronica"),
"exac" = c(0, NA, 0.001, 0.0000001, 0.05, 0.1, 0, NA), "paciente" = c("1/1",
"0/1", "0/1", "1/1", "0/0", "0/0", "0/1", "0/1"))
tabela[7,1]="y"
tabela[8,1]="x"
```

```
write.table(tabela, "tabela", sep = "\t", quote = F)

filtrar = function(tabela, cromossomo, heranca, exac = 0.01)
{
  #warning para input correto do nome do arquivo e consequente leitura
  if(typeof(tabela) != "character") warning("TABELA PRECISA CONTER O CAMINHO DO ARQUIVO ENTRE ÁSPAS")
  tabela <- read.table(tabela)
  #warning para input correto dos argumentos
  if(cromossomo != "autossomico" & cromossomo != "sexual"){warning("'cromossomo' tem que ser igual a 'autossomico' ou 'sexual'")}
```

#filtrando o cromossomo

```
  if(cromossomo == "autossomico"){
    tabela = tabela[tabela$cromossomo %in% 1:22, ]
  } else {
    tabela = tabela[tabela$cromossomo == "x" | tabela$cromossomo == "y", ]
  }
```

#warning para input correto dos argumentos

```
  if(heranca != "dominante" & heranca != "recessiva"){warning("'heranca' tem que ser igual a 'dominante' ou 'recessiva'")}
```

#filtrando a heranca (se a variante esta presente no paciente em um alelo ou em ambos)

```
  if(heranca == "dominante"){
    tabela = tabela[tabela$paciente == "0/1", ]
  } else {
    tabela = tabela[tabela$paciente == "1/1", ]
  }
```

####DEFAULT####

#filtrando o tipo de mutação

```
  tabela = tabela[tabela$mutacao == "exonica" | tabela$mutacao == "splicing", ]
```

#filtrar exac

```
  tabela = tabela[tabela$exac <= exac | is.na(tabela$exac), ]
```

```
return(tabela)

}

filtrar(tabela = "tabela", cromossomo = "autossomico", heranca =
"dominante")
```

Help da função >

modelo package:unknown R Documentation

Função para filtrar variantes possivelmente patogênicas em arquivos de Exoma

Description:

A função filtra as variantes de arquivos de exoma baseando-se pela variável de a doença ser ligada aos cromossomos autossômicos, ou ligada aos cromossomos sexuais e pelo modo de herança, se recessivo, então a coluna do paciente mostrará 1/1, se dominante 0/1, se a variante estiver ausente nesse indivíduo, 0/0 e portanto não será de interesse. A coluna "exac" representa os bancos de dados que fornecem a frequência dessa variante. Se frequente, ela é descartada por se tratar de polimorfismo, se rara e abaixo de 1% (0.01) então é mantida.

Usage:

```
filtrar(tabela = "nome do arquivo ou o caminho dos diretorios até o
arquivo", cromossomo = "autossomico ou sexual", heranca = "dominante ou
recessiva")
```

Arguments:

tabela: requer o nome do arquivo ou o caminho dos diretorios até o arquivo e deve estar entre áspas
cromossomo deve ser indicado se a doença que se busca é autossômica ou ligada aos cromossomos sexuais
heranca deve ser indicado o modo de herança da doença, se dominante ou recessiva
exac valor a ser dado para a filtragem da frequência das variantes nos bancos de dados (padronizado em abaixo de 1%)

Value:

Tabela contendo somente as linhas pertinentes à filtragem feita.
Na filtragem, o valor dos bancos de dados (exac) podem ser alterados, para permitir que variantes com frequências mais altas também apareçam na filtragem, caso desejado.

Warning:

É exibido um **warning** para que o nome correto do arquivo a ser filtrado seja colocado **de** forma correta na função.

Outros dois **warnings** podem surgir caso os argumentos para cromossomo e herança sejam digitados **de** forma errônea.

Author(s):

William Bertani Torres
e-mail: wiillliam@hotmail.com

References:

ERGUN, M. A **new** method **for** analysis of whole exome sequencing **data** (SELIM) depending on variant prioritization. **Informatics in Medicine Unlocked** 8 (2017) 51-53.

Examples:

```
filtrar(tabela = "tabela", cromossomo = "autossomico", heranca =  
"dominante")  
filtrar(tabela = "tabela", cromossomo = "sexual", heranca =  
"dominante")  
filtrar(tabela = "tabela", cromossomo = "autossomico", heranca =  
"recessivo")  
filtrar(tabela = "tabela", cromossomo = "sexual", heranca = "recessivo")
```

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

Permanent link:
http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2018:alunos:trabalho_final:willbertani:start

Last update: 2020/09/23 20:16

