RAREsim Vignette V1

Megan Null

5/25/2020

This vigette describes how to use the RAREsim R package to simulate rare variant genetic data.

The RAREsim R package currently provides functions:

- Expected_variants

- Fit_AFS

- Total_variants

- Prune

Here we will walk through an example simulation of RAREsim using one cM block on chromosome 19 and four ancestry groups: African, East Asian, Non-Finnish European, and South Asian.

Install the package

```
#install.packages('RAREsim')
library(RAREsim)
```

Fitting the Total Number of Variants function

The first function is the Total Number of Variants function. For a given region, the Total Number of Variants function estimates the number of variants per Kb. This is done by estimating ϕ and ω to minimize the function $f_{Var}(x) = \phi x^{\omega}$.

Fitting the Allele Frequency Spectrum function

The AFS function inputs allele frequency information and outputs estimates for α, β, and b.

First, load in example data:

```
# load the data
data("AFS_target_data_NFE")
print(afs_nfe)
```

```
##     block Singletons Doubletons   MAC.3.5   MAC.6.10 MAC.11.20 MAC.21.MAF0.005
## 147    37  0.5230769  0.1664694 0.1573964 0.06232742  0.034714      0.03629191
##     MAF0.005.MAF0.01 Total pop  N_bp Zeros
## 147      0.003550296  2535 NFE 19029     0
```

Pruning Variants

In order to use RAREsim to prune simulated data, genetic data must be simulated with HAPGEN2, with all bp added to the input haplotypes. This will simulate an abudance of rare variants to allow for variant

pruning.

Pruning variants required a MAC file fo the simulated data, and the expected number of variants within each MAC bin.

Expected Number of Variants with Default Parameter

The default parameters can be used to estimate the number of variants within each MAC bin to use for simulation.

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.

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:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example: