

Package ‘zfa’

January 29, 2024

Type Package

Title Zoom-Focus Algorithm

Version 1.1.0

Description Performs Zoom-Focus Algorithm (ZFA) to optimize testing regions for rare variant association tests in exome sequencing data.

License GPL

Imports SKAT (>= 1.1.2)

Encoding UTF-8

LazyData true

NeedsCompilation yes

RoxygenNote 7.2.3

Author Yexian Zhang [cre, aut],
Haoyi Weng [aut],
Maggie Wang [aut, cph]

Maintainer Yexian Zhang <yxzhang@bethbio.com>

Repository CRAN

Date/Publication 2024-01-29 08:30:02 UTC

R topics documented:

zfa	2
zfa.example	4
Index	6

Description

This function performs the Zoom-Focus Algorithm (ZFA) to locate optimal testing regions for rare variant association tests and performs the test based on the optimized regions. The package is suitable to be applied on sequencing data set that is composed of variants with minor allele frequency less than 0.01 (rare variants). The package calls existing rare variant test functions to conduct rare variant test.

ZFA consists of two steps: Zooming and Focusing. In the first step Zooming, a given genomic region is partitioned by an order of two, and the best partition is located using multiple testing corrected p-values returned by desired rare variant test. In the second step Focusing, the boundaries of the zoomed region are refined by allowing them to expand or shrink at micro-level. The computation complexity is linear to the number of variants for Zooming (when the option `fast.path=FALSE`); a fast-Zoom version can further reduce the complexity to the logarithm of data size (`fast.path=TRUE`, default).

Usage

```
zfa(
  data,
  y,
  bin = 256,
  fast.path = TRUE,
  filter.pval = 0.01,
  output.pval = 0.05,
  CommonRare_Cutoff = 0.01,
  test = c("SKAT", "SKATO", "burden", "wtest")
)
```

Arguments

<code>data</code>	a data frame or numeric matrix. Genotypes should be coded as 0,1 or 2.
<code>y</code>	a numeric vector with two levels. Phenotype values are coded as 0 or 1.
<code>bin</code>	a numeric integer taking value of power of two, namely, 2, 4, 8, 16, 32, 64, 128, 256, 512 etc. The bin size specifies the initial window size P to perform the Zoom-Focus Algorithm. Default <code>bin=256</code> .
<code>fast.path</code>	a logical value indicating whether or not to use the fast-Zoom approach. The fast-Zoom performs a binary search instead of exhaustive search, such that at each partition order, the region is divided into two parts, only the part with smaller p-value is continued for the next level search. Default = <code>TRUE</code> .
<code>filter.pval</code>	a p-value threshold to select zoomed regions for conducting focusing step. When specified, only zoomed regions with p-value smaller than the threshold will be passed to focusing step. Default=0.01. Set <code>filter.pval=NULL</code> for conducting focusing step with all the zoomed regions.

output.pval	a p-value threshold for filtering the output. If set NULL, all the results will be listed; otherwise, the function will only output the regions with p-values smaller than output.pval. Default=0.05.
CommonRare_Cutoff	MAF cutoff to define common and rare variants. Default=0.01.
test	a character to choose the rare variant method that combines with the Zoom-Focus Algorithm. If test = "SKAT", the SKAT of variance component test is applied. If test = "SKATO", the SKAT-O of combination method test is applied. If test = "burden", the weighted burden test is applied. If test = "wtest", the W-test of burden test category is applied.

Details

The algorithm divides sequencing data into multiple fixed genomic regions with a certain initial bin size. ZFA is conducted in each bin. Current version includes 4 existing rare variant tests. The SKAT, SKAT-O and weighted burden test are called from 'SKAT' package, and the W-test Collapsing method is self-contained.

Value

The "zfa" function returns a list with the following components:

n.regions	Total number of regions to which the input genotype data is divided by initial bin size P.
n.rare	Total number of rare variants used for the analysis.
n.common	Total number of common variants excluded in the analysis.
results	The testing results consist of several elements: 1) lower .bound and upper .bound represent variant information which indicates the lower and upper bound of optimized testing region; 2) opt .region .size denotes the size of optimal testing region after performing ZFA; 3) corrected .pvalue displays the multiple testing (Bonferroni) corrected p-value of the optimal testing region.
Bon.sig.level	Suggested Bonferroni corrected significance level for the input data at threshold alpha=0.05, which equals to 0.05 / # of regions.
variants	The variants contained in each output optimal region.

Note that the variants in the optimized region will not be printed in default. User can extract the information by calling details of results. See an example in **Examples** section.

The zfa optimizes the testing region according to input variant sequence and assumes that they are arranged by chromosome positions. If variants from non-adjacent genomic regions are input as one data, the zfa will still treat them as adjacent. In this case, the user should be careful in interpreting the results: when an optimized region consists of distant variants, the region may not be biologically meaningful; when an optimized region consists of variants from two neighboring genes, the results may be meaningful.

Author(s)

Haoyi Weng & Maggie Wang

References

- M. H. Wang., H. Weng., et al. (2017) A Zoom-Focus algorithm (ZFA) to locate the optimal testing region for rare variant association tests. *Bioinformatics*.
- M. H. Wang., R. Sun., et al. (2016) A fast and powerful W-test for pairwise epistasis testing. *Nucleic Acids Research*.doi:10.1093/nar/gkw347.
- R. Sun., H. Weng., et al. (2016) A W-test collapsing method for rare-variant association testing in exome sequencing data. *Genetic Epidemiology*, 40(7): 591-596.
- Wu, M. C., Lee, S., et al. (2011) Rare Variant Association Testing for Sequencing Data Using the Sequence Kernel Association Test (SKAT). *American Journal of Human Genetics*, 89, 82-93.
- Lee, S., Emond, M.J., et al. (2012) Optimal unified approach for rare variant association testing with application to small sample case-control whole-exome sequencing studies. *American Journal of Human Genetics*, 91, 224-237.
- Lee, S., Fuchsberger, C., et al. (2015) An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. *Biostatistics*, kxv033.

Examples

```
data(zfa.example)
attach(zfa.example)

# fast-zoom with wtest, all zoomed regions passed to focusing step, and output all results
zfa.result1<-zfa(X,y,bin = 32,fast.path = TRUE,filter.pval=NULL,output.pval=NULL,test = "wtest")

# zooming with wtest, select zoomed regions for focusing and output regions with both p-value<0.01
zfa.result2<-zfa(X,y,bin = 32,fast.path = FALSE,filter.pval=0.01,output.pval=0.01,test = "wtest")

## an example to view the detail of variants in each output optimal region
result1.detail<-zfa.result1$variants
```

zfa.example

Example data for zfa

Description

Example data for "zfa" function.

Usage

```
data(zfa.example)
```

Format

The `zfa.example` contains the following objects:

`x`

a numeric genotype matrix of 1800 individuals and 512 rare variants. Each row represents a different individual, and each column represents a different rare genetic variant. Genotypes are coded as 0,1 or 2.

`y`

a numeric vector of binary phenotypes, among which there are 343 cases and 1457 controls.

See Also

[zfa](#)

Index

* **datasets**

zfa.example, 4

zfa, 2, 5

zfa.example, 4