

Package ‘zalpha’

February 22, 2020

Type Package

Title Run a Suite of Selection Statistics

Version 0.1.0

Author Clare Horscroft

Maintainer Clare Horscroft <c.horscroft@soton.ac.uk>

Description A suite of statistics from the Zalpha family for identifying areas of the genome under selective pressure.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 2.10)

LazyData true

RoxygenNote 6.1.1

Suggests testthat (>= 2.1.0)

R topics documented:

LDprofile	2
LR	2
L_plus_R	3
snps	4
Zalpha	5
Zalpha_all	6
Zalpha_BetaCDF	8
Zalpha_expected	9
Zalpha_log_rsqr_over_expected	11
Zalpha_rsqr_over_expected	12
Zalpha_Zscore	14
Zbeta	15
Zbeta_BetaCDF	16
Zbeta_expected	18
Zbeta_log_rsqr_over_expected	19
Zbeta_rsqr_over_expected	21
Zbeta_Zscore	22
Index	24

LDprofile

Dataset containing an example LD profile

Description

A simulated LD profile, containing example LD statistics for genetic distances of 0 to 0.0049, in bins of size 0.0001.

Usage

```
data(LDprofile)
```

Format

A data frame with 50 rows and 5 variables:

bin the lower bound of each bin

rsq the expected rsq value for a pair of snps, where the genetic distance between them falls in the given bin

sd the standard deviation of the expected rsq value

Beta_a the first shape parameter for the Beta distribution fitted for this bin

Beta_b the second shape parameter for the Beta distribution fitted for this bin

LR

Runs the LR function

Description

Returns the $|L|/|R|$ value for each SNP location supplied to the function. For more information about the $|L|/|R|$ diversity statistic please see Jacobs (2016).

Usage

```
LR(pos, ws, X = NULL)
```

Arguments

pos A numeric vector of SNP locations

ws The window size which the LR statistic will be calculated over. This should be on the same scale as the pos vector.

X Optional. Specify a region of the chromosome to calculate LR for in the format `c(startposition,endposition)`. The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate LR for every SNP in the pos vector.

Value

A list containing the SNP positions and the LR values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps example dataset
data(snps)
## run LR over all the SNPs with a window size of 3000 bp
LR(snps$positions,3000)
## only return results for SNPs between locations 600 and 1500 bp
LR(snps$positions,3000,X=c(600,1500))
```

L_plus_R	<i>Runs the L_plus_R function</i>
----------	-----------------------------------

Description

Returns the $\binom{|L|}{2} + \binom{|R|}{2}$ value for each SNP location supplied to the function. For more information about the L_plus_R diversity statistic please see Jacobs (2016).

Usage

```
L_plus_R(pos, ws, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
ws	The window size which the L_plus_R statistic will be calculated over. This should be on the same scale as the pos vector.
X	Optional. Specify a region of the chromosome to calculate L_plus_R for in the format c(startposition,endposition). The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate L_plus_R for every SNP in the pos vector.

Value

A list containing the SNP positions and the L_plus_R values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps example dataset
data(snps)
## run L_plus_R over all the SNPs with a window size of 3000 bp
L_plus_R(snps$positions,3000)
## only return results for SNPs between locations 600 and 1500 bp
L_plus_R(snps$positions,3000,X=c(600,1500))
```

snps

Dataset containing details on simulated SNPs

Description

A dataset containing the positions, genetic distances and alleles for 20 SNPs, across 10 simulated chromosomes.

Usage

```
data(snps)
```

Format

A data frame with 20 rows and 12 variables:

positions location of the SNP on the chromosome

distances genetic distance of the SNP from the start of the chromosome

chrom_1 allele of the SNP on the first example chromosome

chrom_2 allele of the SNP on the second example chromosome

chrom_3 allele of the SNP on the third example chromosome

chrom_4 allele of the SNP on the fourth example chromosome

chrom_5 allele of the SNP on the fifth example chromosome

chrom_6 allele of the SNP on the sixth example chromosome

chrom_7 allele of the SNP on the seventh example chromosome

chrom_8 allele of the SNP on the eighth example chromosome

chrom_9 allele of the SNP on the ninth example chromosome

chrom_10 allele of the SNP on the tenth example chromosome

Zalpha

*Runs the Zalpha function***Description**

Returns a Z_α value for each SNP location supplied to the function. For more information about the Z_α statistic please see Jacobs (2016). The Z_α statistic is defined as:

$$Z_\alpha = \frac{\binom{|L|}{2}^{-1} \sum_{i,j \in L} r_{i,j}^2 + \binom{|R|}{2}^{-1} \sum_{i,j \in R} r_{i,j}^2}{2}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws , and r^2 is equal to the squared correlation between a pair of SNPs

Usage

```
Zalpha(pos, x, ws, minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
ws	The window size which the Z_α statistic will be calculated over. This should be on the same scale as the pos vector.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate Z_α for in the format <code>c(startposition,endposition)</code> . The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate Z_α for every SNP in the pos vector.

Value

A list containing the SNP positions and the Z_α values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps example dataset
data(snps)
## run Zalpha over all the SNPs with a window size of 3000 bp
Zalpha(snps$positions,as.matrix(snps[,3:12]),3000)
## only return results for SNPs between locations 600 and 1500 bp
Zalpha(snps$positions,as.matrix(snps[,3:12]),3000,X=c(600,1500))
```

Zalpha_all

Runs all the statistics in the zalpha package

Description

Returns every statistic for each SNP location, given the appropriate parameters. See Details for more information.

Usage

```
Zalpha_all(pos, x = NULL, ws, dist = NULL, LDprofile_bins = NULL,
  LDprofile_rsq = NULL, LDprofile_sd = NULL, LDprofile_Beta_a = NULL,
  LDprofile_Beta_b = NULL, minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	Optional. A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
ws	The window size which the statistics will be calculated over. This should be on the same scale as the pos vector.
dist	Optional. A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
LDprofile_bins	Optional. A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_rsq	Optional. A numeric vector containing the expected r^2 values for the corresponding bin in the LD profile. Must be between 0 and 1.
LDprofile_sd	Optional. A numeric vector containing the standard deviation of the r^2 values for the corresponding bin in the LD profile.
LDprofile_Beta_a	Optional. A numeric vector containing the first estimated Beta parameter for the corresponding bin in the LD profile.
LDprofile_Beta_b	Optional. A numeric vector containing the second estimated Beta parameter for the corresponding bin in the LD profile.
minRandL	Minimum number of SNPs in each set R and L for the statistics to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate the statistics for in the format c(startposition,endposition). The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate the statistics for every SNP in the pos vector.

Details

Not all statistics will be returned, depending on the parameters supplied to the function.

If `x` is not supplied, only `Zalpha_expected`, `Zbeta_expected`, `LR` and `L_plus_R` will be calculated.

For any of the statistics which use an expected r^2 value, the parameters `dist`, `LDprofile_bins` and

`LDprofile_rsq` must be supplied. This includes the statistics: `Zalpha_expected`, `Zalpha_rsq_over_expected`, `Zalpha_log_rsq_over_expected`, `Zalpha_Zscore`, `Zalpha_BetaCDF`, `Zbeta_expected`, `Zbeta_rsq_over_expected`, `Zbeta_log_rsq_over_expected`, `Zbeta_Zscore` and `Zbeta_BetaCDF`.

- For `Zalpha_Zscore` and `Zbeta_Zscore` to be calculated, the parameter `LDprofile_sd` must also be supplied.
- For `Zalpha_BetaCDF` and `Zbeta_BetaCDF` to be calculated, the parameters `LDprofile_Beta_a` and `LDprofile_Beta_b` must also be supplied.

For more information about the statistics please see Jacobs (2016).

Value

A list containing the SNP positions and the statistics for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. *Genetics*, 2016. **203**(4): p. 1807

See Also

`Zalpha` `Zalpha_expected` `Zalpha_rsq_over_expected` `Zalpha_log_rsq_over_expected` `Zalpha_Zscore` `Zalpha_BetaCDF` `Zbeta` `Zbeta_expected` `Zbeta_rsq_over_expected` `Zbeta_log_rsq_over_expected` `Zbeta_Zscore` `Zbeta_BetaCDF` `LR` `L_plus_R`

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zalpha_all over all the SNPs with a window size of 3000 bp
## will return all 15 statistics
Zalpha_all(snps$positions,as.matrix(snps[,3:12]),3000,snps$distances,
  LDprofile$bin,LDprofile$rsq,LDprofile$sd,LDprofile$Beta_a,LDprofile$Beta_b)
## only return results for SNPs between locations 600 and 1500 bp
Zalpha_all(snps$positions,as.matrix(snps[,3:12]),3000,snps$distances,
  LDprofile$bin,LDprofile$rsq,LDprofile$sd,LDprofile$Beta_a,LDprofile$Beta_b,X=c(600,1500))
## will only return statistics not requiring an LD profile
Zalpha_all(snps$positions,as.matrix(snps[,3:12]),3000)
```

Zalpha_BetaCDF	<i>Runs the Zalpha function using a cumulative beta distribution function on the r-squared values for the region</i>
----------------	--

Description

Returns a $Z_{\alpha}^{BetaCDF}$ value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the $Z_{\alpha}^{BetaCDF}$ statistic please see Jacobs (2016). The $Z_{\alpha}^{BetaCDF}$ statistic is defined as:

$$Z_{\alpha}^{BetaCDF} = \frac{\binom{|L|}{2}^{-1} \sum_{i,j \in L} \frac{B(r_{i,j}^2; a, b)}{B(a, b)} + \binom{|R|}{2}^{-1} \sum_{i,j \in R} \frac{B(r_{i,j}^2; a, b)}{B(a, b)}}{2}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws , r^2 is equal to the squared correlation between a pair of SNPs, and $\frac{B(r_{i,j}^2; a, b)}{B(a, b)}$ is the cumulative distribution function for the Beta distribution given the estimated a and b parameters from the LD profile.

Usage

```
Zalpha_BetaCDF(pos, x, dist, ws, LDprofile_bins, LDprofile_Beta_a,
  LDprofile_Beta_b, minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the $Z_{\alpha}^{BetaCDF}$ statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_Beta_a	A numeric vector containing the first estimated Beta parameter for the corresponding bin in the LD profile.
LDprofile_Beta_b	A numeric vector containing the second estimated Beta parameter for the corresponding bin in the LD profile.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate $Z_{\alpha}^{BetaCDF}$ for in the format c(startposition, endposition). The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate $Z_{\alpha}^{BetaCDF}$ for every SNP in the pos vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the $Z_{\alpha}^{BetaCDF}$ values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zalpha_BetaCDF over all the SNPs with a window size of 3000 bp
Zalpha_BetaCDF(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$Beta_a,LDprofile$Beta_b)
## only return results for SNPs between locations 600 and 1500 bp
Zalpha_BetaCDF(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$Beta_a,LDprofile$Beta_b,X=c(600,1500))
```

Zalpha_expected	<i>Runs the Zalpha function on the expected r-squared values for the region</i>
-----------------	---

Description

Returns a $Z_{\alpha}^{E[r^2]}$ value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the $Z_{\alpha}^{E[r^2]}$ statistic please see Jacobs (2016). The $Z_{\alpha}^{E[r^2]}$ statistic is defined as:

$$Z_{\alpha}^{E[r^2]} = \frac{\binom{|L|}{2}^{-1} \sum_{i,j \in L} E[r_{i,j}^2] + \binom{|R|}{2}^{-1} \sum_{i,j \in R} E[r_{i,j}^2]}{2}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws , and $E[r^2]$ is equal to the expected squared correlation between a pair of SNPs, given an LD profile.

Usage

```
Zalpha_expected(pos, dist, ws, LDprofile_bins, LDprofile_rsq,
  minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the $Z_{\alpha}^{E[r^2]}$ statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_rsq	A numeric vector containing the expected r^2 values for the corresponding bin in the LD profile. Must be between 0 and 1.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate $Z_{\alpha}^{E[r^2]}$ for in the format <code>c(startposition,endposition)</code> . The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate $Z_{\alpha}^{E[r^2]}$ for every SNP in the pos vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the $Z_{\alpha}^{E[r^2]}$ values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zalpha_expected over all the SNPs with a window size of 3000 bp
Zalpha_expected(snps$positions,snps$distances,3000,LDprofile$bin,LDprofile$rsq)
## only return results for SNPs between locations 600 and 1500 bp
Zalpha_expected(snps$positions,snps$distances,3000,LDprofile$bin,LDprofile$rsq,X=c(600,1500))
```

Zalpha_log_rsq_over_expected

Runs the Zalpha function on the log of the r-squared values over the expected r-squared values for the region

Description

Returns a $Z_{\alpha}^{\log_{10}(r^2/E[r^2])}$ value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the $Z_{\alpha}^{\log_{10}(r^2/E[r^2])}$ statistic please see Jacobs (2016). The $Z_{\alpha}^{\log_{10}(r^2/E[r^2])}$ statistic is defined as:

$$Z_{\alpha}^{\log_{10}(r^2/E[r^2])} = \frac{\binom{|L|}{2}^{-1} \sum_{i,j \in L} \log_{10}(r_{i,j}^2/E[r_{i,j}^2]) + \binom{|R|}{2}^{-1} \sum_{i,j \in R} \log_{10}(r_{i,j}^2/E[r_{i,j}^2])}{2}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws , r^2 is equal to the squared correlation between a pair of SNPs, and $E[r^2]$ is equal to the expected squared correlation between a pair of SNPs, given an LD profile.

Usage

```
Zalpha_log_rsq_over_expected(pos, x, dist, ws, LDprofile_bins,
                             LDprofile_rsq, minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the $Z_{\alpha}^{\log_{10}(r^2/E[r^2])}$ statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_rsq	A numeric vector containing the expected r^2 values for the corresponding bin in the LD profile. Must be between 0 and 1.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate $Z_{\alpha}^{\log_{10}(r^2/E[r^2])}$ for in the format c(startposition, endposition). The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate $Z_{\alpha}^{\log_{10}(r^2/E[r^2])}$ for every SNP in the pos vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the $Z_{\alpha}^{\log_{10}(r^2/E[r^2])}$ values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zalpha_log_rsq_over_expected over all the SNPs with a window size of 3000 bp
Zalpha_log_rsq_over_expected(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsq)
## only return results for SNPs between locations 600 and 1500 bp
Zalpha_log_rsq_over_expected(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsq,X=c(600,1500))
```

Zalpha_rsq_over_expected

Runs the Zalpha function on the r-squared values over the expected r-squared values for the region

Description

Returns a $Z_{\alpha}^{r^2/E[r^2]}$ value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the $Z_{\alpha}^{r^2/E[r^2]}$ statistic please see Jacobs (2016). The $Z_{\alpha}^{r^2/E[r^2]}$ statistic is defined as:

$$Z_{\alpha}^{r^2/E[r^2]} = \frac{\binom{|L|}{2}^{-1} \sum_{i,j \in L} r_{i,j}^2 / E[r_{i,j}^2] + \binom{|R|}{2}^{-1} \sum_{i,j \in R} r_{i,j}^2 / E[r_{i,j}^2]}{2}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws , r^2 is equal to the squared correlation between a pair of SNPs, and $E[r^2]$ is equal to the expected squared correlation between a pair of SNPs, given an LD profile.

Usage

```
Zalpha_rsq_over_expected(pos, x, dist, ws, LDprofile_bins, LDprofile_rsq,
  minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the $Z_{\alpha}^{r^2/E[r^2]}$ statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_rsqa	A numeric vector containing the expected r^2 values for the corresponding bin in the LD profile. Must be between 0 and 1.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate $Z_{\alpha}^{r^2/E[r^2]}$ for in the format c(startposition,endposition). The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate $Z_{\alpha}^{r^2/E[r^2]}$ for every SNP in the pos vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the $Z_{\alpha}^{r^2/E[r^2]}$ values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zalpha_rsq_over_expected over all the SNPs with a window size of 3000 bp
Zalpha_rsq_over_expected(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsqa)
## only return results for SNPs between locations 600 and 1500 bp
Zalpha_rsq_over_expected(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsqa,X=c(600,1500))
```

Zalpha_Zscore	<i>Runs the Zalpha function using the Z score of the r-squared values for the region</i>
---------------	--

Description

Returns a Z_{α}^{Zscore} value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the Z_{α}^{Zscore} statistic please see Jacobs (2016). The Z_{α}^{Zscore} statistic is defined as:

$$Z_{\alpha}^{Zscore} = \frac{\binom{|L|}{2}^{-1} \sum_{i,j \in L} \frac{r_{i,j}^2 - E[r_{i,j}^2]}{\sigma[r_{i,j}^2]} + \binom{|R|}{2}^{-1} \sum_{i,j \in R} \frac{r_{i,j}^2 - E[r_{i,j}^2]}{\sigma[r_{i,j}^2]}}{2}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws , r^2 is equal to the squared correlation between a pair of SNPs, $E[r^2]$ is equal to the expected squared correlation between a pair of SNPs, given an LD profile, and $\sigma[r^2]$ is the standard deviation.

Usage

```
Zalpha_Zscore(pos, x, dist, ws, LDprofile_bins, LDprofile_rsq,
  LDprofile_sd, minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the Z_{α}^{Zscore} statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_rsq	A numeric vector containing the expected r^2 values for the corresponding bin in the LD profile. Must be between 0 and 1.
LDprofile_sd	A numeric vector containing the standard deviation of the r^2 values for the corresponding bin in the LD profile.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate Z_{α}^{Zscore} for in the format <code>c(startposition, endposition)</code> . The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate Z_{α}^{Zscore} for every SNP in the pos vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the Z_{α}^{Zscore} values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zalpha_Zscore over all the SNPs with a window size of 3000 bp
Zalpha_Zscore(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsq,LDprofile$sd)
## only return results for SNPs between locations 600 and 1500 bp
Zalpha_Zscore(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsq,LDprofile$sd,X=c(600,1500))
```

Zbeta	<i>Runs the Zbeta function</i>
-------	--------------------------------

Description

Returns a Z_{β} value for each SNP location supplied to the function. For more information about the Z_{β} statistic please see Jacobs (2016). The Z_{β} statistic is defined as:

$$Z_{\beta} = \frac{\sum_{i \in L, j \in R} r_{i,j}^2}{|L||R|}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws , and r^2 is equal to the squared correlation between a pair of SNPs

Usage

```
Zbeta(pos, x, ws, minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
ws	The window size which the Z_β statistic will be calculated over. This should be on the same scale as the pos vector.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate Z_β for in the format c(startposition,endposition). The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate Z_β for every SNP in the pos vector.

Value

A list containing the SNP positions and the Z_β values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps example dataset
data(snps)
## run Zbeta over all the SNPs with a window size of 3000 bp
Zbeta(snps$positions,as.matrix(snps[,3:12]),3000)
## only return results for SNPs between locations 600 and 1500 bp
Zbeta(snps$positions,as.matrix(snps[,3:12]),3000,X=c(600,1500))
```

Zbeta_BetaCDF	<i>Runs the Zbeta function using a cumulative beta distribution function on the r-squared values for the region</i>
---------------	---

Description

Returns a $Z_\beta^{BetaCDF}$ value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the $Z_\beta^{BetaCDF}$ statistic please see Jacobs (2016). The $Z_\beta^{BetaCDF}$ statistic is defined as:

$$Z_\beta^{BetaCDF} = \frac{\sum_{i \in L, j \in R} \frac{B(r_{i,j}^2; a, b)}{B(a, b)}}{|L||R|}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws, r^2 is equal to the squared correlation between a pair of SNPs, and $\frac{B(r_{i,j}^2; a, b)}{B(a, b)}$ is the cumulative distribution function for the Beta distribution given the estimated a and b parameters from the LD profile.

Usage

```
Zbeta_BetaCDF(pos, x, dist, ws, LDprofile_bins, LDprofile_Beta_a,
               LDprofile_Beta_b, minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the $Z_{\beta}^{BetaCDF}$ statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_Beta_a	A numeric vector containing the first estimated Beta parameter for the corresponding bin in the LD profile.
LDprofile_Beta_b	A numeric vector containing the second estimated Beta parameter for the corresponding bin in the LD profile.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate $Z_{\beta}^{BetaCDF}$ for in the format c(startposition,endposition). The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate $Z_{\beta}^{BetaCDF}$ for every SNP in the pos vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the $Z_{\beta}^{BetaCDF}$ values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zbeta_BetaCDF over all the SNPs with a window size of 3000 bp
Zbeta_BetaCDF(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$Beta_a,LDprofile$Beta_b)
## only return results for SNPs between locations 600 and 1500 bp
Zbeta_BetaCDF(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$Beta_a,LDprofile$Beta_b,X=c(600,1500))
```

Zbeta_expected	<i>Runs the Zbeta function on the expected r-squared values for the region</i>
----------------	--

Description

Returns a $Z_{\beta}^{E[r^2]}$ value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the $Z_{\beta}^{E[r^2]}$ statistic please see Jacobs (2016). The $Z_{\beta}^{E[r^2]}$ statistic is defined as:

$$Z_{\beta}^{E[r^2]} = \frac{\sum_{i \in L, j \in R} E[r_{i,j}^2]}{|L||R|}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the window ws , and $E[r^2]$ is equal to the expected squared correlation between a pair of SNPs, given an LD profile.

Usage

```
Zbeta_expected(pos, dist, ws, LDprofile_bins, LDprofile_rsq,
  minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the $Z_{\beta}^{E[r^2]}$ statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_rsq	A numeric vector containing the expected r^2 values for the corresponding bin in the LD profile. Must be between 0 and 1.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.

- X Optional. Specify a region of the chromosome to calculate $Z_{\beta}^{E[r^2]}$ for in the format `c(startposition,endposition)`. The start position and the end position should be within the extremes of the positions given in the `pos` vector. If not supplied, the function will calculate $Z_{\beta}^{E[r^2]}$ for every SNP in the `pos` vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the $Z_{\beta}^{E[r^2]}$ values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zbeta_expected over all the SNPs with a window size of 3000 bp
Zbeta_expected(snps$positions,snps$distances,3000,LDprofile$bin,LDprofile$rsqr)
## only return results for SNPs between locations 600 and 1500 bp
Zbeta_expected(snps$positions,snps$distances,3000,LDprofile$bin,LDprofile$rsqr,X=c(600,1500))
```

Zbeta_log_rsqr_over_expected

Runs the Zbeta function on the log of the r-squared values over the expected r-squared values for the region

Description

Returns a $Z_{\beta}^{\log_{10}(r^2/E[r^2])}$ value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the $Z_{\beta}^{\log_{10}(r^2/E[r^2])}$ statistic please see Jacobs (2016). The $Z_{\beta}^{\log_{10}(r^2/E[r^2])}$ statistic is defined as:

$$Z_{\beta}^{\log_{10}(r^2/E[r^2])} = \frac{\sum_{i \in L, j \in R} \log_{10}(r_{i,j}^2/E[r_{i,j}^2])}{|L||R|}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window `ws`, r^2 is equal to the squared correlation between a pair of SNPs, and $E[r^2]$ is equal to the expected squared correlation between a pair of SNPs, given an LD profile.

Usage

```
Zbeta_log_rsq_over_expected(pos, x, dist, ws, LDprofile_bins,
  LDprofile_rsq, minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the $Z_{\beta}^{\log_{10}(r^2/E[r^2])}$ statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_rsq	A numeric vector containing the expected r^2 values for the corresponding bin in the LD profile. Must be between 0 and 1.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate $Z_{\beta}^{\log_{10}(r^2/E[r^2])}$ for in the format c(startposition, endposition). The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate $Z_{\beta}^{\log_{10}(r^2/E[r^2])}$ for every SNP in the pos vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the $Z_{\beta}^{\log_{10}(r^2/E[r^2])}$ values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zbeta_log_rsq_over_expected over all the SNPs with a window size of 3000 bp
Zbeta_log_rsq_over_expected(snps$positions, as.matrix(snps[, 3:12]), snps$distances, 3000,
```

```
LDprofile$bin,LDprofile$rsq)
## only return results for SNPs between locations 600 and 1500 bp
Zbeta_log_rsq_over_expected(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsq,X=c(600,1500))
```

Zbeta_rsq_over_expected

Runs the Zbeta function on the r-squared values over the expected r-squared values for the region

Description

Returns a $Z_{\beta}^{r^2/E[r^2]}$ value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the $Z_{\beta}^{r^2/E[r^2]}$ statistic please see Jacobs (2016). The $Z_{\beta}^{r^2/E[r^2]}$ statistic is defined as:

$$Z_{\beta}^{r^2/E[r^2]} = \frac{\sum_{i \in L, j \in R} r_{i,j}^2 / E[r_{i,j}^2]}{|L||R|}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws , r^2 is equal to the squared correlation between a pair of SNPs, and $E[r^2]$ is equal to the expected squared correlation between a pair of SNPs, given an LD profile.

Usage

```
Zbeta_rsq_over_expected(pos, x, dist, ws, LDprofile_bins, LDprofile_rsq,
  minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the $Z_{\beta}^{r^2/E[r^2]}$ statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_rsq	A numeric vector containing the expected r^2 values for the corresponding bin in the LD profile. Must be between 0 and 1.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate $Z_{\beta}^{r^2/E[r^2]}$ for in the format <code>c(startposition,endposition)</code> . The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate $Z_{\beta}^{r^2/E[r^2]}$ for every SNP in the pos vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the $Z_{\beta}^{r^2/E[r^2]}$ values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zbeta_rsq_over_expected over all the SNPs with a window size of 3000 bp
Zbeta_rsq_over_expected(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsq)
## only return results for SNPs between locations 600 and 1500 bp
Zbeta_rsq_over_expected(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsq,X=c(600,1500))
```

Zbeta_Zscore	<i>Runs the Zbeta function using the Z score of the r-squared values for the region</i>
--------------	---

Description

Returns a Z_{β}^{Zscore} value for each SNP location supplied to the function, based on the expected r^2 values given an LD profile and genetic distances. For more information about the Z_{β}^{Zscore} statistic please see Jacobs (2016). The Z_{β}^{Zscore} statistic is defined as:

$$Z_{\beta}^{Zscore} = \frac{\sum_{i \in L, j \in R} \frac{r_{i,j}^2 - E[r_{i,j}^2]}{\sigma[r_{i,j}^2]}}{|L||R|}$$

where $|L|$ and $|R|$ are the number of SNPs to the left and right of the current locus within the given window ws , r^2 is equal to the squared correlation between a pair of SNPs, $E[r^2]$ is equal to the expected squared correlation between a pair of SNPs, given an LD profile, and $\sigma[r^2]$ is the standard deviation.

Usage

```
Zbeta_Zscore(pos, x, dist, ws, LDprofile_bins, LDprofile_rsq, LDprofile_sd,
  minRandL = 4, minRL = 25, X = NULL)
```

Arguments

pos	A numeric vector of SNP locations
x	A matrix of SNP values. Columns represent chromosomes, rows are SNP locations. Hence, the number of rows should equal the length of the pos vector. SNPs should all be biallelic.
dist	A numeric vector of genetic distances (e.g. cM, LDU). This should be the same length as pos.
ws	The window size which the Z_{β}^{Zscore} statistic will be calculated over. This should be on the same scale as the pos vector.
LDprofile_bins	A numeric vector containing the lower bound of the bins used in the LD profile. These should be of equal size.
LDprofile_rsq	A numeric vector containing the expected r^2 values for the corresponding bin in the LD profile. Must be between 0 and 1.
LDprofile_sd	A numeric vector containing the standard deviation of the r^2 values for the corresponding bin in the LD profile.
minRandL	Minimum number of SNPs in each set R and L for the statistic to be calculated. Default is 4.
minRL	Minimum value for the product of the set sizes for R and L. Default is 25.
X	Optional. Specify a region of the chromosome to calculate Z_{β}^{Zscore} for in the format <code>c(startposition,endposition)</code> . The start position and the end position should be within the extremes of the positions given in the pos vector. If not supplied, the function will calculate Z_{β}^{Zscore} for every SNP in the pos vector.

Details

The LD profile describes the expected correlation between SNPs at a given genetic distance, generated using simulations or real data. Care should be taken to utilise an LD profile which is representative of the population in question. The LD profile should consist of evenly-sized bins of distances (for example 0.00001 cM per bin), where the value given is the (inclusive) lower bound of the bin.

Value

A list containing the SNP positions and the Z_{β}^{Zscore} values for those SNPs

References

Jacobs, G.S., T.J. Sluckin, and T. Kivisild, *Refining the Use of Linkage Disequilibrium as a Robust Signature of Selective Sweeps*. Genetics, 2016. **203**(4): p. 1807

Examples

```
## load the snps and LDprofile example datasets
data(snps)
data(LDprofile)
## run Zbeta_Zscore over all the SNPs with a window size of 3000 bp
Zbeta_Zscore(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsq,LDprofile$sd)
## only return results for SNPs between locations 600 and 1500 bp
Zbeta_Zscore(snps$positions,as.matrix(snps[,3:12]),snps$distances,3000,
  LDprofile$bin,LDprofile$rsq,LDprofile$sd,X=c(600,1500))
```

Index

*Topic **datasets**

LDprofile, [2](#)

snps, [4](#)

L_plus_R, [3](#), [7](#)

LDprofile, [2](#)

LR, [2](#), [7](#)

snps, [4](#)

Zalpha, [5](#), [7](#)

Zalpha_all, [6](#)

Zalpha_BetaCDF, [7](#), [8](#)

Zalpha_expected, [7](#), [9](#)

Zalpha_log_rsqr_over_expected, [7](#), [11](#)

Zalpha_rsqr_over_expected, [7](#), [12](#)

Zalpha_Zscore, [7](#), [14](#)

Zbeta, [7](#), [15](#)

Zbeta_BetaCDF, [7](#), [16](#)

Zbeta_expected, [7](#), [18](#)

Zbeta_log_rsqr_over_expected, [7](#), [19](#)

Zbeta_rsqr_over_expected, [7](#), [21](#)

Zbeta_Zscore, [7](#), [22](#)