Package 'EpipwR'

February 1, 2026

Type Package

Title Efficient Power Analysis for EWAS with Continuous or Binary Outcomes

Version 1.5.0

Description A quasi-simulation based approach to performing power analysis for EWAS (Epigenome-wide association studies) with continuous or binary outcomes. 'EpipwR' relies on empirical EWAS datasets to determine power at specific sample sizes while keeping computational cost low. EpipwR can be run with a variety of standard statistical tests, controlling for either a false discovery rate or a family-wise type I error rate.

License Artistic-2.0

Encoding UTF-8

LazyData false

URL https://github.com/jbarth216/EpipwR

BugReports https://github.com/jbarth216/EpipwR

Imports EpipwR.data, ExperimentHub (>= 2.10.0), ggplot2

Depends R (>= 4.4.0)

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Suggests knitr, rmarkdown, testthat (>= 3.0.0), sessioninfo

VignetteBuilder knitr

biocViews Epigenetics, ExperimentalDesign

Config/testthat/edition 3

git_url https://git.bioconductor.org/packages/EpipwR

git_branch devel

git_last_commit 409b2c6

git_last_commit_date 2025-10-29

Repository Bioconductor 3.23

Date/Publication 2026-02-01

Author Jackson Barth [aut, cre] (ORCID:
<https://orcid.org/0009-0009-6307-9928>),
Austin Reynolds [aut],
Mary Lauren Benton [ctb],
Carissa Fong [ctb]

Maintainer Jackson Barth <Jackson_Barth@Baylor.edu>

Contents

`EpipwR` *EpipwR: Efficient Power Analysis for EWAS with Continuous or Binary Outcomes*

Description

EpipwR: Efficient Power Analysis for EWAS with Continuous or Binary Outcomes

Details

A quasi-simulation based approach to performing power analysis for EWAS (Epigenome-wide association studies) with continuous or binary outcomes. 'EpipwR' relies on empirical EWAS datasets to determine power at specific sample sizes while keeping computational cost low. EpipwR can be run with a variety of standard statistical tests, controlling for either a false discovery rate or a family-wise type I error rate.

For a detailed overview, please use `vignette("EpipwR")`.

Author(s)

Maintainer: Jackson Barth <Jackson_Barth@Baylor.edu> (ORCID)

Authors:

- Austin Reynolds

Other contributors:

- Mary Lauren Benton [contributor]
- Carissa Fong [contributor]

See Also

Useful links:

- https://github.com/jbarth216/EpipwR
- Report bugs at https://github.com/jbarth216/EpipwR

EpipwR_plot *Plotting EpipwR Output*

Description

Plots 95% error bars of EpipwR output. Different sample sizes are displayed along the x-axis with lines representing distinct correlation means. This functions should only be used with EpipwR output.

Usage

```
EpipwR_plot(df)
```

Arguments

df	A data frame containing output from the get_power_cont() or get_power_cc() function.

Value

A line plot with sample size on the x-axis and power on the y-axis. Multiple lines are plotted for each correlation/effect size mean specified.

Examples

```
out <- get_power_cont(100,10000,c(70,80,90,100),.05,c(.3,.4,.5))
EpipwR_plot(out)
```

get_power_cc *Power Calculations for Case/Control EWAS*

Description

Calculates power for EWAS with a binary outcome for multiple sample sizes and/or effect sizes based on Barth and Reynolds (2025). Data sets are only simulated for the non-null tests; p-values are generated directly for the null tests. Rather than specifying the number of data sets to calculate power, you specify the precision level (MOE) and a maximum number of data sets (Nmax). After 20 data sets, the function terminates when the desired precision level is reached or if the number of tested data sets reaches Nmax. Researchers who are familiar with pwrEWAS can use pwrE_to_EpipwR to use pwrEWAS parameterization in EpipwR.

Usage

```
get_power_cc(
  dm,
  Total,
  n,
  fdr_fwer,
  delta_mu,
  delta_sd = 0,
  n1_prop = 0.5,
  Tissue = "Saliva",
  Nmax = 1000,
  MOE = 0.03,
  test = "pooled",
  use_fdr = TRUE,
  det_limit = 0,
  Suppress_updates = FALSE,
  emp_data = NULL
)
```

Arguments

`dm` Number of non-null tests.

`Total` The total number of tests (null and non-null).

`n` Sample size(s) for which power is calculated (accepts a vector). This is total sample size.

`fdr_fwer` Either the false discovery rate or the family-wise type I error rate, depending on `use_fdr`.

`delta_mu` Average effect size (difference in mean methylation between groups) for non-null tests (accepts a vector). If multiple `n` and `delta_mu` are specified, power is calculated for all unique settings under the Cartesian product of these vectors.

`delta_sd` Standard deviation of the effect size for non-null tests. If 0, all effect sizes are fixed at `delta_mu`.

`n1_prop` Indicates the proportion of the total sample size (`n`) in group 1 (rounded to the nearest integer).

`Tissue` Tissue type of Empirical EWAS to be used for data generation (see details for valid options).

`Nmax` The maximum number of data sets used to calculate power.

`MOE` The target margin of error of a 95% confidence interval for average power. This determines the stopping point of the algorithm.

`test` The type of t-test to be used. `"pooled"` indicates a pooled variance t-test while `"WS"` indicates Welch's t-test.

`use_fdr` If `TRUE`, uses `fdr_fwer` as the false discovery rate. If `FALSE`, uses the family-wise type I error rate.

`det_limit` The minimum mean difference for the effect size distribution. ignored if `delta_sd=0`.

`Suppress_updates`
If `TRUE`, blocks messages reporting the completion of each unique setting.

`emp_data` Reference data set in matrix or data frame format (Beta values with CpG sites as rows, samples as columns). Ignored unless `Tissue="Custom"`.

Details

Valid options for the `Tissue` argument are `"Saliva"`, `"Lymphoma"`, `"Placenta"`, `"Liver"`, `"Colon"`, `"Blood adult"`, `"Blood 5 year olds"`, `"Blood newborns"`, `"Cord-blood (whole blood)"`, `"Cord-blood (PBMC)"`, `"Adult (PBMC)"`, and `"Sperm"`. All data sets are publicly available on the gene expression omnibus (see EpipwR.data package for more details) and were identified by Graw, et al. (2019). Please note that, due to some extreme values in this data set, the Lymphoma option will occasionally throw a warning related to data generation. At this time, we recommend using one of the other tissue options. Users who would like to use their own reference data set should set `Tissue="Custom"` and provide the data in matrix or data frame format with `emp_data`. Users who would like to take advantage of this setting are responsible for formatting this data set correctly.

Unlike pwrEWAS (Graw et al., 2019), EpipwR enforces equality of precision (sum of the parameters) on the distributions of each group rather than equality of variance.

Value

A dataframe with rows equal to the number of `n` and `rho_mu` combinations

References

Barth, J., and Reynolds, A. W. (2025). EpipwR: Efficient power analysis for EWAS with continuous outcomes. *Bioinformatics Advances*, 5(1), vbaf150.

Graw, S., Henn, R., Thompson, J. A., and Koestler, D. C. (2019). pwrEWAS: A user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). *BMC Bioinformatics*, 20(1):218.

Examples

```
# This examples calculates power for 100 non-null tests out of 10,000 total
# with an FDR of 5%. Sample sizes of 40, 50, 60, 70 and fixed effect sizes
# of 0.01, 0.02, 0.05 are used. For improved accuracy, MOE=.01 to ensure
# that the 95% confidence interval for average power has a margin of error
# no larger than .01 (unless the maximum of 1,000 data sets is exceeded).
get_power_cc(100,10000,c(40,50,60,70),.05,c(.01,.02,.05), MOE=.01)
```

`get_power_cont` *Power Calculations for Continuous EWAS*

Description

Calculates power for EWAS with a continuous outcome for multiple sample sizes and/or correlations based on Barth and Reynolds (2025). Data sets are only simulated for the non-null tests; p-values are generated directly for the null tests. Rather than specifying the number of data sets to calculate power, you specify the precision level (`MOE`) and a maximum number of data sets (`Nmax`). After 20 data sets, the function terminates when the desired precision level is reached or if the number of tested data sets reaches `Nmax`.

Usage

```
get_power_cont(
  dm,
  Total,
  n,
  fdr_fwer,
  rho_mu,
  rho_sd = 0,
  Tissue = "Saliva",
  Nmax = 1000,
  MOE = 0.03,
  test = "pearson",
  use_fdr = TRUE,
  det_limit = 0.03,
  Suppress_updates = FALSE,
  emp_data = NULL,
  phenotype_data = NULL
)
```

Arguments

`dm` Number of non-null tests.

`Total` The total number of tests (null and non-null).

`n` Sample size(s) for which power is calculated (accepts a vector).

`fdr_fwer` Either the false discovery rate or the family-wise type I error rate, depending on `use_fdr`.

`rho_mu` Mean correlation(s) of methylation and phenotype for non-null tests (accepts a vector). If multiple `n` and `rho_mu` are specified, power is calculated for all unique settings under the Cartesian product of these vectors.

`rho_sd` Standard deviation of methylation and phenotype for non-null tests. If 0 all correlations are fixed at `rho_mu`.

`Tissue` Tissue type of Empirical EWAS to be used for data generation (see details for valid options).

`Nmax` The maximum number of datasets used to calculate power

`MOE` The target margin of error of a 95% confidence interval for average power. This determines the stopping point of the algorithm.

`test` The type of statistical test to be used. `"pearson"`, `"kendall"` and `"spearman"` are valid options.

`use_fdr` If `TRUE`, uses `fdr_fwer` as the false discovery rate. If `FALSE`, uses the family-wise type I error rate.

`det_limit` The minimum absolute correlation for the effect size distribution. Ignored if `rho_sd=0`.

`Suppress_updates`
If `TRUE`, blocks messages reporting the completion of each unique setting.

`emp_data` Reference data set in matrix or data frame format (Beta values with CpG sites as rows, samples as columns). Ignored unless `Tissue="Custom"`.

`phenotype_data` A sample of phenotype data to be used in the power calculation(s). Accepts a vector with length > 100, although we recommend that users make this at least as large as their maximum sample size. If left blank, a normal distribution is used to generate correlations.

Details

Valid options for the `Tissue` argument are `"Saliva"`, `"Lymphoma"`, `"Placenta"`, `"Liver"`, `"Colon"`, `"Blood adult"`, `"Blood 5 year olds"`, `"Blood newborns"`, `"Cord-blood (whole blood)"`, `"Cord-blood (PBMC)"`, `"Adult (PBMC)"`, and `"Sperm"`. All data sets are publicly available on the gene expression omnibus (see EpipwR.data package for more details) and were identified by Graw, et. al. (2019). Please note that, due to some extreme values in this data, the Lymphoma option will occasionally throw a warning related to data generation. At this time, we recommend using one of the other tissue options. Users who would like to use their own reference data set should set `Tissue="Custom"` and provide the data in matrix or data frame format with `emp_data`. Similarly, users can also now specify their own phenotype data, either from a real data set or by generating samples from a known distribution (i.e., `rt(1000,2)`). Users who would like to take advantage of either of these settings are responsible for the quality and formatting of the data provided.

Although this function only covers 3 types of statistical tests, its worth noting that tests run using software packages such as limma will yield the same results as a pearson correlation test in the absence of covariates or any dependence across CpG sites (as is the assumption here). Any users wanting to mimic an analysis done in limma should use `test="pearson"`.

Value

A data frame with rows equal to the number of `n` and `rho_mu` combinations

References

Barth, J., and Reynolds, A. W. (2025). EpipwR: Efficient power analysis for EWAS with continuous outcomes. *Bioinformatics Advances*, 5(1), vbaf150.

Graw, S., Henn, R., Thompson, J. A., and Koestler, D. C. (2019). pwrEWAS: A user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). *BMC Bioinformatics*, 20(1):218.

Examples

```
# This examples calculates power for 100 non-null tests out of 10,000 total
# with an FDR of 5%.
# Sample sizes of 70,80,90,100 and fixed correlations at 0.3,0.4,0.5 are
# used.
get_power_cont(100,10000,c(70,80,90,100),.05,c(.3,.4,.5))
```

pwrE_to_EpipwR *Using pwrEWAS effect-size distributions in EpipwR*

Description

A function that determines the EpipwR inputs (delta_mu, delta_sd, det_limit) for effect-size distribution using pwrEWAS methodology (Centered at 0, truncated around 0 +/- det_limit and the 99.99th percentile of the distribution as maximal_delta). The output can then be used as inputs to get_power_cc.

Usage

```
pwrE_to_EpipwR(maximal_delta, det_limit, quantile = 0.9999)
```

Arguments

maximal_delta	The desired 99.99th percentile (unless quantile is specified) of the effect size distribution. Must be between 0 and 1.
det_limit	The minimum effect size to be used in a power calculation.
quantile	The quantile that maximal_delta represents. Note that pwrEWAS uses quantile=0.9999.

Details

As described in Graw, et al. (2019), users specify the "maximal delta" and the detection limit to set the effect size distribution. The purpose of this function is to provide a map between the pwrEWAS inputs and the EpipwR inputs to generate the same effect size distribution.

The main purpose of the function is to calculate the standard deviation of the effect size distribution, since the mean is always 0 and the detection limit is user-sepcified. Specifically, the standard deviation is calculated such that the 99.99th percentile of a normal distribution (truncating out 0 +/- det_limit) is maximal_delta. For further justification, see Graw, et al. (2019). Note that this function does allow users to specify quantiles other than 0.9999. If users wish to use this function for continuous EpipwR, we recommend specifying a smaller quantile (i.e., 0.95, 0.99 etc.).

Value

A list with the calculated input values for the get_power_cc function.

References

Graw, S., Henn, R., Thompson, J. A., and Koestler, D. C. (2019). pwrEWAS: A user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). *BMC Bioinformatics*, 20(1):218.

Examples

```
#This example find the correct EpipwR settings for a maximal (99.99th
#percentile) difference of 0.25 and a dectection limit of .005.
pwrE_to_EpipwR(.25, .005)
```

Index