The pwrEWAS User's Guide

Stefan Graw, Devin C. Koestler

1 July 2018

Abstract

pwrEWAS is a computationally efficient tool to estimate power in EWAS as a function of sample and effect size for two-group comparisons of DNAm (e.g., case vs control, exposed vs non-exposed, etc.). Detailed description of in-/outputs, instructions and an example, as well as interpretations of the example results are provided in the following vignette.

Package

pwrEWAS 1.0

Contents

Introduction

When designing an epigenome-wide association study (EWAS) to investigate the relationship be-tween DNA methylation (DNAm) and some exposure(s) or phenotype(s), it is critically important to assess the sample size needed to detect a hypothesized difference with adequate statistical power. However, the complex and nuanced nature of DNAm data make direct assessment of statistical power challenging. To circumvent these challenges and to address outstanding need for a user-friendly interface for EWAS power evaluation, we have developed pwrEWAS. The current imple-mentation of pwrEWAS accommodates power estimation for two-group comparisons of DNAm (e.g., case vs control, exposed vs non-exposed, etc.). Power is calculated by means of a Monte Carlo approach in which DNAm data are randomly generated from one of several different existing data datasets, chosen to cover the most common tissue-types used in EWAS. In addition to specify-ing the tissue type to be used for DNAm profiling, users are required to specify the sample size, number of differentially methylated CpGs, effect size(s) (e.g., Δ_β), target false discovery rate (FDR) and the number of simulated data sets, and have the option of selecting from several different sta-tistical methods to perform differential methylation analyses. pwrEWAS reports the marginal power, marginal type I error rate, marginal FDR, and false discovery cost (FDC) .The R-Shiny web inter-face allows for easy input of user-defined parameters and includes an advanced settings button that offers additional options pertaining to data generation and computation.

Dependencies

This document has the following dependencies

```
library(shiny)
library(shinyBS)
# library(car)
# library(CpGassoc)
# library(truncnorm)
# library(limma)
# library(genefilter)
library(ggplot2)
library(parallel)
```

Usage

```
out = pwrEWAS(minTotSampleSize = 10,
                    maxTotSampleSize = 50,
                    SampleSizeSteps = 10,
                    NcntPer = 0.5,
                    targetDelta = c(0.2, 0.5),
                    J = 10000,
                    targetDmCpGs = 100,
                    tissueType = "Saliva",
                    detectionLimit = 0.01,
```

```
                        DMmethod = "limma",
                        FDRcritVal = 0.05,
                        core = 4,
                        sims = 50)
```

Input parameter

Parameter	Decription
minTotSampleSize	Minimum total sample size
maxTotSampleSize	Maximum total sample size
SampleSizeSteps	Sample size increments
NcntPer	Percentage sample group 1
targetDelta	Target maximum difference in mean methylation
J	Number of CpGs tested/simulated
targetDmCpGs	Target number of DM CpGs
tissueType	Tissue type
detectionLimit	Detection Limit
DMmethod	Method of DM analysis
FDRcritVal	Target FDR
core	Threads
sims	Number of simulated data sets

Output parameter