, 2024, pp. 1–7

doi: DOI HERE
Advance Access Publication Date: Day Month Year
Paper

PAPER

EpipwR: Efficient Power Analysis for EWAS with Continuous Outcomes

Jackson Barth[1]* and Austin W. Reynolds[2,3]

[1]Department of Statistical Science, Baylor University, One Bear Place 97140, TX, 76798, USA, [2]Department of Anthropology, Baylor University, One Bear Place 97173, TX, 76798, USA and [3]Department of Microbiology, Immunology, and Genetics, , School of Biomedical Sciences, University of North Texas Health Science Center, Fort Worth, TX, USA
*Corresponding author. Jackson_Barth@Baylor.edu
FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

Motivation: Epigenome-wide association studies (EWAS) have emerged as a popular way to investigate the pathophysiology of complex diseases and to assist in bridging the gap between genotypes and phenotypes. Despite the increasing popularity of EWAS, very few tools exist to aid researchers in power estimation and those are limited to case-control studies. The existence of user-friendly tools, expanding power calculation functionality to additional study designs would be a significant aid to researchers planning EWAS.
Results: We introduce EpipwR, an open-source R package that can efficiently estimate power for EWAS with continuous outcomes. EpipwR uses a quasi-simulated approach, meaning that data is generated only for CpG sites with methylation associated with the outcome, while p-values are generated directly for those with no association (when necessary). Like existing EWAS power calculators, reference datasets of empirical EWAS are used to guide the data generation process. Two simulation studies show the effect of the selected empirical dataset on the generated correlations and the relative speed of EpipwR compared to similar approaches.
Availability and Implementation: The EpipwR R-package is currently available for download at github.com/jbarth216/EpipwR.

Key words: EWAS, Power Anaylsis, R package

1. Introduction

Over the past decade, epigenome-wide association studies (EWAS) have emerged as the dominant way for researchers to investigate the relationship between epigenetic markers and phenotypes at the genome-wide level. DNA methylation has become the most widely studied epigenetic mechanism because of its ease to assay through existing microarray and sequencing technology, its ubiquity across the genome, and its impact on gene expression. DNA methylation, the process whereby a methyl group is added to a nucleotide changing its availability for transcription, occurs most often in the context of cytosine-phosphate-guanine (CpG) dinucleotides in millions of locations around the genome. To assay this variation at a population scale, most EWAS use microarray-based platforms for assessing DNAm, such as the Illumina Infinium HumanMethylation450 BeadChip (covering ˜450,000 CpG sites in the human genome) or the Illumina Infinium MethylationEPIC Array (covering >850,000 CpGs). Both arrays assess methylation at single-CpG resolution, quantified using a methylation β-value; an approximately continuously-distributed measure that reflects the level of methylation at a specific locus, ranging from 0 (unmethylated) to 1 (methylated) (Bibikova et al., 2006). The development of these arrays has precipitated a huge interest in studying DNA methylation in the context of human health and disease (Murphy and Mill, 2014; Rakyan et al., 2011; Wei et al., 2021). The need for power calculations to inform EWAS study design has often been reiterated since their advent (Mansell et al., 2019; Michels et al., 2013; Tsai and Bell, 2015), but few tools have been created to aid researchers.

It is now widely accepted that formal power calculation and sample size justification is an essential part of genomic study design to ensure meaningful findings are reported. This has motivated the development of numerous statistical power evaluation tools for genome- and transcriptome-wide association studies, including: the GAS Power Calculator (Johnson and Abecasis, 2017), GWAPower (Feng et al., 2011), RnaSeqSampleSize (Zhao et al., 2018), and RNAseqPS (Guo et al., 2014). However, surprisingly few tools for EWAS power evaluation have been developed, despite substantial work in the areas of DNA methylation data QC, normalization, and analysis (Aryee et al., 2014; Chakravarthy et al., 2018; Fortin et al., 2014; Houseman et al., 2012; Jaffe et al., 2012; Morris et al., 2014; Peters et al., 2015; Song and Kuan, 2022; Teschendorff et al., 2013, 2017). The pwrEWAS package (Graw et al., 2019) is perhaps the only tool currently available

to researchers for easily estimating power under a variety of conditions (e.g. sample size, effect size, false discovery rate threshold). Even with the availability of this tool, most EWAS continue to be conducted without formal power analyses, resulting in potentially under- and over-powered studies (Tsai and Bell, 2015). Furthermore, pwrEWAS is currently limited to case-control study designs, limiting the ability of researchers to easily estimate sample size requirements during study design.

The scope of EWAS, however, have expanded beyond two-group comparison designs into the realm of complex traits. For example, the relationship between age and CpG methylation has been extensively studied over the past decade (Hannum et al., 2013; Horvath, 2013; Horvath and Raj, 2018), with many applications to other health outcomes (Belsky et al., 2020, 2022; Levine et al., 2018; Lu et al., 2019a,b, 2022). More recently, EWAS have investigated CpG methylation correlated with a number of quantitative traits; including blood metabolites (Gomez-Alonso et al., 2021; Jhun et al., 2021), body mass (Aslibekyan et al., 2015; Demerath et al., 2015; Wahl et al., 2017), lung function (Herrera-Luis et al., 2022), and even behavioral health (Barbu et al., 2022; Montalvo-Ortiz et al., 2022; Raffington et al., 2023). As the number and complexity of phenotypes being studied with EWAS continues to grow, so too does the research community's need for appropriate and easily accessible power calculation tools.

Here we introduce EpipwR, a publicly available tool for comprehensive EWAS power evaluation in the context of continuous outcome variables. EpipwR is a quasi-simulated approach to power analysis, in that data based on empirical EWAS are generated only for non-null distributions while false positive p-values are generated directly using classic statistical theory. In this way, EpipwR leverages the unique form of EWAS data while running much faster than a fully-simulated method. EpipwR was written using the R statistical programming language (R Core Team, 2021) and the package is available at https://github.com/jbarth216/EpipwR.

2. Methods

Suppose an EWAS has K CpG sites and n distinct patient samples. Let β_{ik} be the proportion of DNA methylation from patient i, CpG k, where it is assumed that β_{ik} from a common CpG site are independently and identically distributed from some beta distribution, $\beta_{\cdot k} \sim Beta(\alpha_k, \beta_k)$. Of interest to researchers is a continuous phenotype Y_i which is assumed to be correlated with the DNA methylation of some CpG sites, $\rho_k = cor(Y, M_{\cdot k})$, where $M_{ik} = logit(\beta_{ik})$. While M_{ik} is usually closer to normality than β_{ik} (Du et al., 2010), it should be noted that skewed distributions of β_{ik} can still yield non-normal M_{ik}. Finally, K can be split into true nulls (K_n of K have $\rho_k = 0$) and false nulls (K_m of K have $\rho_k \neq 0$). The goal of EpipwR is to evaluate the power of this analysis, defined as the proportion of false null hypotheses among K_m that are correctly identified. The remainder of this section outlines and justifies the methodology implemented in EpipwR.

2.1. Data Generation

The first step in calculating power for the scenario described above is to generate methylation data, β_{ik}. In an effort to maintain consitency across tools used in epigenomics, EpipwR applies the same method outlined by Graw et al. (2019) in pwrEWAS to generate the methylated proportions. To determine plausible distributions for each non-null CpG site (which will then be used to generate β_{ik}), EpipwR relies on empirical EWAS reference data. One of twelve publicly available EWAS of different tissue types is selected by the user based on relevance to the study being planned. Each dataset corresponds to a different list of beta distribution parameters $\hat{\alpha}_k$, $\hat{\beta}_k$ estimated through method of moments on every CpG site in the selected dataset. For every truly correlated CpG site in the planned study, EpipwR samples K_m pairs of these parameters with replacement to be used to generate the methylation data. Unlike pwrEWAS which is a fully simulated approach, methylation datasets are not generated for the true nulls in EpipwR (see section 2.2), which significantly reduces computation time and memory load.

Once the beta distributions have been identified, the next step is to generate n samples of $(Y, \beta_{\cdot 1}, \beta_{\cdot 2}, ..., \beta_{\cdot K_m})$. EpipwR assumes that all β_{ik} are conditionally independent on Y such that $Cov(\beta_{\cdot k}, \beta_{\cdot l}|Y) = 0$ for the k^{th} and l^{th} CpG sites. This is a common simplifying assumption in power analysis, and one that removes the responsibility from the researcher of identifying a valid covariance structure for potentially thousands of $\beta_{\cdot k}$. This assumption allows Y to be sampled marginally, and then the subsequent $\beta_{ik}|Y$ can be sampled concurrently. EpipwR also assumes that Y is normally distributed and assumes without loss of generality that it follows a standard normal distribution. Therefore, EpipwR initially generates n samples of $Y \sim N(0, 1)$. To conditionally sample β_{ik}, EpipwR utilizes a version of rank correlation closely based on Ruscio and Kaczetow (2008). First, n samples of $X_k|Y$ are generated, where X_k has a marginal standard normal distribution with $\rho_k = cor(X_k, Y)$. The percentiles of each X_{ik} are then used to systematically generate β_{ik}, such that $\Phi(X_{ik}) = F_{\beta_{\cdot k}}(\beta_{ik})$ where Φ is the standard normal cumulative distribution function (CDF) and $F_{\beta_{\cdot k}}$ is the CDF of $\beta_{ik} \sim Beta\left(\hat{\alpha}_k, \hat{\beta}_k\right)$ with parameters sampled from the empirical EWAS data. For example, suppose x_{ik} is generated to be -0.8416. Since this is the 20th percentile value of the standard normal distribution, then β_{ik} is assigned the 20th percentile value of its corresponding beta distribution. The β_{ik} are then converted to M_{ik} via the logit transformation.

Taking the approach described in Ruscio and Kaczetow (2008) for generating non-normal correlated data would involve randomly generated beta samples, which are then reordered based on the ranks of X_{ik}. EpipwR leverages the inverse CDF of the beta distribution to directly convert X_{ik} to β_{ik}. While the computational resources of each approach is similar, using the inverse CDF tends to be slightly more accurate and has a standard error closer to what would be expected under normal conditions (see section 3.1). Additionally, the method in Ruscio and Kaczetow (2008) is an iterative approach, adjusting ρ_k and repeating the process until r_k (the sample correlation) reaches the desired level. EpipwR does not use an iterative approach but rather accepts the first generated correlation for each CpG. This is not neccesarily because EpipwR will always produce $\rho_k = r_k$ (although it is more accurate on average) but because this is a more accurate reflection of reality. When employing Pearson correlation (or any parametric measure) to assess the association between two quantitative variables, it is essential to acknowledge the underlying assumption of linearity in their relationship. However, this assumption may not always hold true, particularly when one of the distributions exhibits skewness (as is the case with the reference data- see figure 1). In such instances, only a segment of the relationship may adhere to linearity, leaving the remaining portion inaccurately estimated by the test. Therefore, if ρ_k is meant to describe the

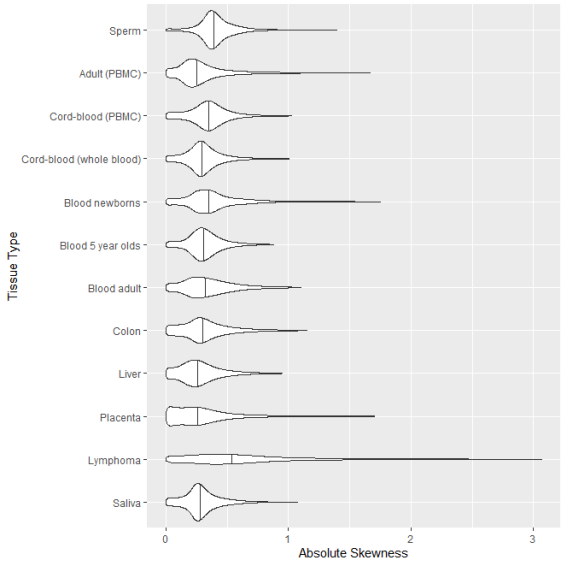

Fig. 1. The violin plots show the distributions of absolute skewness across each of the 12 reference datasets used by EpipwR. For ease of viewing, the plots only show up to the 99th percentile of each dataset. The vertical lines display the median of each dataset

entrie relationship between Y and the DNA methylation, the generation of a linear correlation coefficient r_k slightly lower than anticipated under normal conditions is not a flaw but rather a distinctive feature of the method. Simulation results in Section 3.1 shows that heightened skewness in $\beta_{\cdot k}$ corresponds with a more pronounced underestimation in r_k compared to perfectly normal conditions.

In setting the target correlation ρ_k, researchers have the option of choosing a fixed correlation for all CpG sites such that $\mu_\rho = \rho_k$ or of identifying parameters of a truncated normal distribution to randomly generate the correlation of each associated CpG site:

$$\rho_k \sim N(\mu_p, \sigma_\rho), \quad \rho \in \left\{ \begin{array}{ll} \text{if } \mu_\rho > 0, & [.03, 1] \\ \text{if } \mu_\rho < 0, & [-1, -.03] \end{array} \right\}.$$

For the non-null tests, ρ_k cannot be in the range $(-.03, .03)$, preserving practical significance.

2.2. Calculating p-values

For the K_m associated CpG sites, the p-value calculations are standard. The sample correlation for each CpG site, r_k, is first calculated based on the chosen method (Pearson, Kendall or Spearman) for each dataset, and then p-values are similarly determined. For the non-parametric tests, exact p-values are calculated when the data have no ties (this is often the case) and are estimated with parametric methodology when ties are present. Becasue EpipwR (like most power analysis software) does not take into account dependence between CpG sites unrelated to Y, using EpipwR with the pearson test would yield the same results if the analysis was instead run using specialized software such as limma (Ritchie et al., 2015).

If a bonferonni adjustment is used to control for the family-wise Type I error rate, then p-values for the K_n null tests can be safely ignored. If instead users wish to control for a false discovery rate (FDR), then additional p-values for the non-significant associations must be considered. For the K_n CpG sites with a true null, computational resources are reduced significantly by generating p-values from a uniform distribution on the interval $(0, 1)$, a well-established result in statistical theory. However, even this approach can become computationally taxing if K_n is large and the sampling is repeated on several datasets ($N > 100$). To further improve computational efficiency, EpipwR then calculates the largest order statistic from the K_n p-values with a probability of at least 1×10^{-10} of moving the cutoff and altering the power calculation (call this ξ). For each dataset, EpipwR then draws only the first ξ order statistics from K_m $U(0, 1)$ samples, each of which can be sampled directly from a beta distribution (For a large number of tests, xi is around 30-40). All other p-values are assumed to be too large to alter the FDR cutoff.

2.3. Calculating Power

EpipwR uses a traditional calculation to evaluate power, where power is the percentage of the K_m truly associated CpG sites found to be significant. False positives are ignored in the power calculation (although if FDR is chosen, the false positives can effect the threshold). This calculation is repeated for each of the N datasets. The average across all N datasets is accepted as the average power.

Rather than specifying a specific number of datasets N, EpipwR uses a dynamic process to determine N as it runs. First, EpipwR calculates power for 20 different datasets (the mininum value for N). Beginning at the 20th dataset, it then calculates the margin of error of a 95% confidence interval for average power using the below formula,

$$m_{(j)} = 1.96 \cdot \frac{s_{(j)}}{\sqrt{N_{(j)}}} \tag{1}$$

where $s_{(j)}$ is the sample standard deviation of power through the first $N_{(j)}$ generated datasets. If $m_{(j)}$ is less than a user-specified margin of error (the default is .03), the process stops and uses the results from the $N_{(j)}$ generated datasets to calculate average power. Otherwise, a new dataset is generated and the algorithm repeats, this time with $N_{(j+1)} = N_{(j)} + 1$ datasets. The process continues until $m_{(j)}$ is small enough or if $N_{(j)}$ reaches a user-specified maximum (default is 1,000) to limit the strain on computational resources.

2.4. EpipwR in Practice

There are two main functions in the EpipwR package, one that calculates power under various scenarios (EpipwR_cont) and one that plots the results of the power analysis (EpipwR_plot). EpipwR_cont takes 12 inputs, as indicated in table 1. Since the sample size and μ_ρ inputs can take on multiple values, EpipwR calculates power separately for all scenarios under the cartesian product of these two vectors. The output of EpipwR_cont is a dataframe with 1 row for every combination of n and rho_mu, and columns tracking the average power, number of datasets used and the standard error of the average power. EpipwR_plot takes this dataframe and plots 95% confidence intervals using ggplot2 (Wickham, 2016).

Figure 2 demonstrates the EpipwR workflow. In this scenario, researchers plan an EWAS with 100,000 CpG sites, of which 500 are expected to have non-null correlations with some phenotype of interest. The researchers wish to test sample sizes of 100-200 at denominations of 25 and $\mu_\rho \in (0.3, 0.35, 0.4)$ where ρ_k are fixed at these values. At a 5% false discovery rate, the plots on the right-hand side of figure 2 show that sample sizes of 125 and 150 are high enough to acheive 80% and 90%

Table 1. EpipwR_cont inputs

Name	Description	Notes
dm	Amount of non-null tests	
Total	Amount of all tests	
n	Sample size(s) at which power is calculated	Accepts a vector of sample sizes
fdr_fwer	family-wise type I error or false discovery rate (FDR)	Depends on use_fdr
rho_mu	Average value of ρ for non-null tests	Accepts a vector of averages
rho_sd = 0	Standard deviation of ρ for non-null tests	0 implies that ρ_k= rho_mu for all non-null tests
Tissue = "Saliva"	Reference data used to generate β_{ik}	See package for valid options
Nmax = 1000	Maximum number of datasets generated	Minimum number is fixed at 20
MOE = .03	Target margin of error for a 95% confidence interval	See section 2.3
test = "pearson"	Type of correlation/test utilized	Can be pearson, kendall or spearman
use_fdr = T	Indication that fdr_fwer should be treated as FDR	If false, fdr_fwer is treated as FWER
Suppress_updates = F	If T, removes intermediate status updates	Has no impact on the results

```
> out <- get_power_cont(
dm=500,
Total=100000,
n=c(100,125,150,175,200),
fdr_fwer=.05,
rho_mu=c(0.3,0.35,0.4),
rho_sd=0, Tissue="Saliva",
Nmax=1000,
MOE=.03,
test="pearson",
use_fdr=T,
Suppress_updates=F)
> out


> EpipwR_plot(out)
```

```
sample_size rho_mu avg_power   sd_power  N    se_power
        100   0.30 0.1403333 0.08294258 30 0.015143174
        125   0.30 0.2834091 0.10035896 44 0.015129683
        150   0.30 0.4852500 0.11408804 56 0.015245656
        175   0.30 0.5902963 0.11076639 54 0.015073397
        200   0.30 0.7403871 0.08369535 31 0.015032129
        100   0.35 0.3679444 0.12881474 72 0.015180962
        125   0.35 0.6158689 0.11830236 61 0.015147065
        150   0.35 0.7489565 0.10367534 46 0.015286095
        175   0.35 0.8796000 0.05303266 20 0.011858464
        200   0.35 0.9232000 0.04748141 20 0.010617166
        100   0.40 0.6177113 0.14950251 97 0.015179680
        125   0.40 0.8202667 0.10062044 45 0.014999609
        150   0.40 0.9253000 0.05495750 20 0.012288870
        175   0.40 0.9580000 0.02690334 20 0.006015769
        200   0.40 0.9873000 0.01342856 20 0.003002718
```

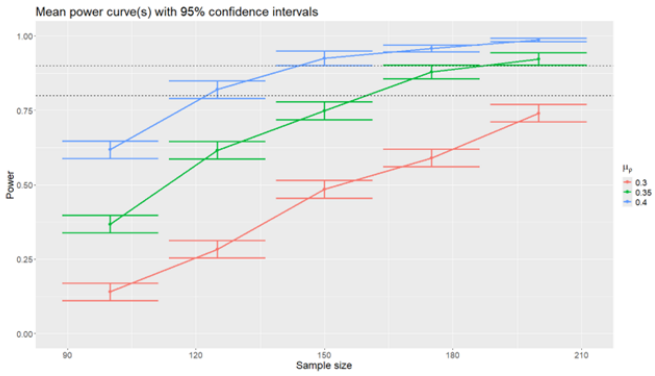

Fig. 2. Code and output of the primary EpipwR functions. The get_power_cont() function produces the R-console table on the right-hand side, while EpipwR_plot() produces the error bar plot.

power (respectively) if $\mu_\rho = 0.4$, whereas larger sample sizes are needed for smaller correlations.

3. Results

3.1. Generation of correlated datasets

To evaluate the efficacy of our method to produce dependent data, a small simulation was conducted to observe the large-sample behavior of the sample correlations under differing beta distributions. Sample datasets for CpG methylation were generated based on 800 $\hat{\alpha}, \hat{\beta}$ estimates from the saliva dataset, drawn using stratified sampling based on the estimated skewness (γ) of the distribution. Specifically, 100 pairs of estimates were drawn such that $|\gamma| \in (0, 0.01)$, another 100 were drawn such that $|\gamma| \in (0.01, 0.5)$ and so on for a variety of ranges (see the x-axis of figure 3). Correlated data were then generated based on sample sizes $n \in (10, 30, 50, 100, 150, 200)$ and "true" correlation ρ from 0.1 to 0.9 in increments of 0.1. For each simulation setting, 1,000 correlated datasets were generated based on the method described in section 2.1. The average and standard deviation of the sample correlations were then stored.

Figure 3 shows the results of this simulation. Unsurprisingly, the bias for each setting is always negative, confirming that this method tends to underestimate correlation. When the skewness of the beta distribution is low ($\gamma < 1$), the bias is very close to 0, particularly for large sample sizes. As the skewness increases, the bias becomes more severe, reaching as high as -2%, -4% and -6% for small, medium and large correlations, respectively. It should be noted, however, that absolute skewness levels above 2 or 3 are very rare in the reference data (0.61% and

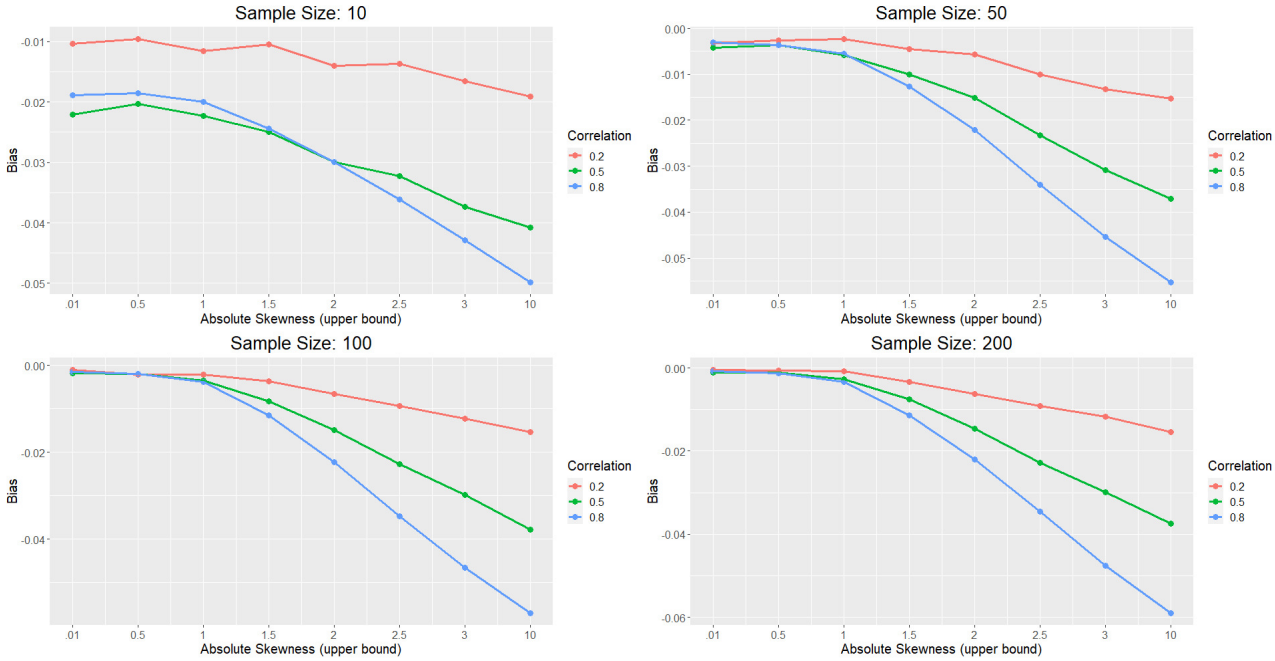

Fig. 3. Correlation estimation results using the inverse CDF method. When skewness is low, there is very little bias in the correlation of the generated datasets. However as absolute skewness grows, so does the bias, getting as low as $-.06$ for the most extreme scenarios.

0.14% frequency, respectively), while levels above 10 are nearly non-existent ($<$ .01% frequency). Therefore, average bias on the whole tends to be quite small.

Absolute skewness levels of each dataset are reported in figure 1. While the median of each dataset is well below 1, certain datasets (lymphoma, newborn blood and sperm) have a small amount of CpG sites with relatively high (> 1) skewness levels. These distributions tend to generate slightly smaller linear correlations consistent with the results of the simulation, which in turn produce lower power levels. Given the rarity of these cases, the effect of these high-skewness distributions are marginal and tend to be washed out by the high frequency of distributions with low or moderate skewness.

3.2. Computation time

The second simulation compares the computation time required to calculate power across various sample sizes and settings. All settings assumed 100,000 CpG sites, with $K_m \in (10, 100, 1000)$ of those having a non-zero correlation with the phenotype (i.e.,). Sample sizes of $n \in (10, 50, 100, 200)$ and correlations $\rho \in (0.1, 0.3, 0.5, 0.7, 0.9)$ were considered. The simulation also ran with a static ρ ($\sigma_\rho = 0$), and again with $\sigma_\rho = 0.05$. For all settings, an FDR of 5% was used with a target power margin of error of 1%. Because of this, each simulation setting used a different number of datasets (minimum is 20 and maximum is 1,000).

Partial results for the simulation are shown below in Table 2. Only simulation settings with 1,000 significantly associated CpG sites and with $\sigma_\rho = 0$ are shown (relaxing the latter restriction to $\sigma_\rho = 0.05$ increases the computation time by an average of 3 seconds). The number of datasets used in each setting is also provided in the parentheses, as this influences the overall computation time. Generally speaking, larger sample sizes and larger K_m correspond to longer run times. The same is not true for K_n. In cases where the sample size is considerably small or large based on ρ, almost all datasets will lead to estimated power of 0 or 1, and the minimum number of datasets ($N = 20$) will typically suffice. This highlights another major advantage of EpipwR: the minimal cost of specifying a wide range of sample sizes, as the algorithm produces results from these 'trivial' power scenarios in just a few seconds. In cases where the average power is closer to 0.5, more datasets are needed to produce reliable estimates of average power, leading to longer computation times. Alternatively, increasing the target margin of errror will decrease the number of datasets required and therefore reduce the total computation time. At the recommended level of 3%, N is almost always under 300. Regardless of the scenario, the standard for EpipwR is to use anywhere from a few seconds to a few minutes to produce results, whereas other simulation-based tools can take hours or even days.

Table 2. Computation time in seconds of EpipwR for various settings with $K_m = 1000$.

	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
$n = 10$	1.32 ($N = 20$)	1.70 (26)	28.39 (452)	62.56 (1000)	1.37 (20)
$n = 50$	4.64 (20)	39.22 (172)	87.37 (382)	4.60 (20)	4.66 (20)
$n = 100$	8.73 (20)	170.20 (393)	8.72 (20)	8.67 (20)	8.72 (20)
$n = 200$	17.49 (20)	69.61 (82)	17.07 (20)	17.02 (20)	17.09 (20)

4. Discussion

EpipwR is the first open-source tool specifically designed for sample size determination of EWAS with continuous phenotypes. Similar to many power analysis tools for large-scale studies, it employs simulations to replicate empirical studies. But because these are limited to non-null and a small number of null tests, EpipwR is much faster than existing, fully-simulated EWAS power analysis tools (Graw et al., 2019). It is important

to note that, like all power analysis tools, EpipwR makes several simplifying assumptions, and its outputs should be viewed as preliminary estimates rather than exact calculations.

Although EpipwR is introduced in this manuscript as a power calculator for continuous phenotypes, a future version of this package will contain tools to estimate power under a case control study design as well. Additionally, the methodology of EpipwR can be applied to much more sophisticated scenarios, such as the inclusion of one or more confounding variables, estimating correlation across CpG sites that can not be explained by the phenotype, and more complicated study designs (ANOVA, mixed effect models, etc.).

References

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., and Irizarry, R. A. (2014). Minfi: A flexible and comprehensive bioconductor package for the analysis of infinium dna methylation microarrays. *Bioinformatics*, 30(10):1363–1369.

Aslibekyan, S., Demerath, E. W., Mendelson, M., Zhi, D., Guan, W., Liang, L., Sha, J., Pankow, J. S., Liu, C., Irvin, M. R., Fornage, M., Hidalgo, B., Lin, L.-A., Stanton Thibeault, K., Bressler, J., Tsai, M. Y., Grove, M. L., Hopkins, P. N., Boerwinkle, E., others, and Arnett, D. K. (2015). Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity*, 23(7):1493–1501.

Barbu, M. C., Amador, C., Kwong, A. S. F., Shen, X., Adams, M. J., Howard, D. M., Walker, R. M., Morris, S. W., Min, J. L., Liu, C., Van Dongen, J., Ghanbari, M., Relton, C., Porteous, D. J., Campbell, A., Evans, K. L., Whalley, H. C., and McIntosh, A. M. (2022). Complex trait methylation scores in the prediction of major depressive disorder. *eBioMedicine*, 79:104000. Article 104000.

Belsky, D. W., Caspi, A., Arseneault, L., Baccarelli, A., Corcoran, D. L., Gao, X., Hannon, E., Harrington, H. L., Rasmussen, L. J., Houts, R., Huffman, K., Kraus, W. E., Kwon, D., Mill, J., Pieper, C. F., Prinz, J. A., Poulton, R., Schwartz, J., Sugden, K., others, and Moffitt, T. E. (2020). Quantification of the pace of biological aging in humans through a blood test, the dunedinpoam dna methylation algorithm. *eLife*, 9. Article e54870.

Belsky, D. W., Caspi, A., Corcoran, D. L., Sugden, K., Poulton, R., Arseneault, L., Baccarelli, A., Chamarti, K., Gao, X., Hannon, E., Harrington, H. L., Houts, R., Kothari, M., Kwon, D., Mill, J., Schwartz, J., Vokonas, P., Wang, C., Williams, B. S., and Moffitt, T. E. (2022). Dunedinpace, a dna methylation biomarker of the pace of aging. *eLife*, 11. Article e73420.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E. W., Wu, B., Doucet, D., Thomas, N. J., Wang, Y., Vollmer, E., Goldmann, T., Seifart, C., Jiang, W., Barker, D. L., Chee, M. S., Floros, J., and Fan, J.-B. (2006). High-throughput dna methylation profiling using universal bead arrays. *Genome Research*, 16(3):383–393.

Chakravarthy, A., Furness, A., Joshi, K., Ghorani, E., Ford, K., Ward, M. J., King, E. V., Lechner, M., Marafioti, T., Quezada, S. A., Thomas, G. J., Feber, A., and Fenton, T. R. (2018). Pan-cancer deconvolution of tumour composition using dna methylation. *Nature Communications*, 9(1):3220.

Demerath, E. W., Guan, W., Grove, M. L., Aslibekyan, S., Mendelson, M., Zhou, Y.-H., Hedman, K., Sandling, J. K., Li, L.-A., Irvin, M. R., Zhi, D., Deloukas, P., Liang, L., Liu, C., Bressler, J., Spector, T. D., North, K., Li, Y., Absher, D. M., others, and Boerwinkle, E. (2015). Epigenome-wide association study (ewas) of bmi, bmi change and waist circumference in african American adults identifies multiple replicated loci. *Human Molecular Genetics*, 24(15):4464–4479.

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W. A., Hou, L., and Lin, S. M. (2010). Comparison of beta-value and m-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics*, 11(1):587.

Feng, S., Wang, S., Chen, C.-C., and Lan, L. (2011). Gwapower: A statistical power calculation software for genome-wide association studies with quantitative traits. *BMC Genetics*, 12(1):12.

Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., Greenwood, C. M., and Hansen, K. D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology (Online Edition)*, 15(12):503.

Gomez-Alonso, M. D. C., Kretschmer, A., Wilson, R., Pfeiffer, L., Karhunen, V., Seppälä, I., Zhang, W., Mittelstra, K., Wahl, S., Matias-Garcia, P. R., Prokisch, H., Horn, S., Meitinger, T., Serrano-Garcia, L. R., Sebert, S., Raitakari, O., Loh, M., Rathmann, W., Müller-Nurasyid, M., others, and Waldenberger, M. (2021). Dna methylation and lipid metabolism: An ewas of 226 metabolic measures. *Clinical Epigenetics*, 13(1):7.

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.

Guo, Y., Zhao, S., Li, C.-I., Sheng, Q., and Shyr, Y. (2014). Rnaseqps: A web tool for estimating sample size and power for rnaseq experiment. *Cancer Informatics*, 13s6.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., and Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 49(2):359–367.

Herrera-Luis, E., Li, A., Mak, A. C. Y., Perez-Garcia, J., Elhawary, J. R., Oh, S. S., Hu, D., Eng, C., Keys, K. L., Huntsman, S., Beckman, K. B., Borrell, L. N., Rodriguez-Santana, J., Burchard, E. G., and Pino-Yanes, M. (2022). Epigenome-wide association study of lung function in latino children and youth with asthma. *Clinical Epigenetics*, 14(1):9.

Horvath, S. (2013). Dna methylation age of human tissues and cell types. *Genome Biology*, 14(10):R115.

Horvath, S. and Raj, K. (2018). Dna methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, 19(6):371–384.

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., Wiencke, J. K., and Kelsey, K. T. (2012). Dna methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, 13(1):86.

Jaffe, A. E., Feinberg, A. P., Irizarry, R. A., and Leek, J. T. (2012). Significance analysis and statistical dissection of variably methylated regions. *Biostatistics*, 13(1):166–178.

Jhun, M.-A., Mendelson, M., Wilson, R., Gondalia, R., Joehanes, R., Salfati, E., Zhao, X., Braun, K. V. E., Do, A. N., Hedman, K., Zhang, T., Carnero-Montoro, E., Shen,

J., Bartz, T. M., Brody, J. A., Montasser, M. E., O'Connell, J. R., Yao, C., Xia, R., others, and Assimes, T. L. (2021). A multi-ethnic epigenome-wide association study of leukocyte dna methylation and blood lipids. *Nature Communications*, 12(1):3987.

Johnson, J. L. and Abecasis, G. R. (2017). Gas power calculator: Web-based power calculator for genetic association studies. *bioRxiv*.

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Hou, L., Baccarelli, A. A., Stewart, J. D., Li, Y., Whitsel, E. A., Wilson, G., Reiner, A. P., Aviv, A., Lohman, K., Liu, Y., and Ferrucci, L. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, 10(4).

Lu, A. T., Binder, A. M., Zhang, J., Yan, Q., Reiner, A. P., Cox, S. R., Corley, J., Harris, S. E., Kuo, P.-L., Moore, A. Z., Bandinelli, S., Stewart, J. D., Wang, C., Hamlat, E. J., Epel, E. S., Schwartz, J. D., Whitsel, E. A., Correa, A., Ferrucci, L., others, and Horvath, S. (2022). Dna methylation grimage version 2. *Aging*.

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., Whitsel, E. A., Assimes, T. L., Ferrucci, L., and Horvath, S. (2019a). Dna methylation grimage strongly predicts lifespan and healthspan. *Aging*, 11(2):303–327.

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., Whitsel, E. A., Assimes, T. L., Ferrucci, L., and Horvath, S. (2019b). Dna methylation grimage strongly predicts lifespan and healthspan. *Aging*, 11(2):303–327.

Mansell, G., Gorrie-Stone, T. J., Bao, Y., Kumari, M., Schalkwyk, L. S., Mill, J., and Hannon, E. (2019). Guidance for dna methylation studies: Statistical insights from the illumina epic array. *BMC Genomics*, 20(1):366.

Michels, K. B., Binder, A. M., Dedeurwaerder, S., Epstein, C. B., Greally, J. M., Gut, I., Houseman, E. A., Izzi, B., Kelsey, K. T., Meissner, A., Milosavljevic, A., Siegmund, K. D., Bock, C., and Irizarry, R. A. (2013). Recommendations for the design and analysis of epigenome-wide association studies. *Nature Methods*, 10(10):949–955.

Montalvo-Ortiz, J. L., Gelernter, J., Cheng, Z., Girgenti, M. J., Xu, K., Zhang, X., Gopalan, S., Zhou, H., Duman, R. S., Southwick, S. M., Krystal, J. H., Traumatic Stress Brain Research Study Group, Friedman, M. J., Duman, R. S., Girgenti, M. J., Krystal, J. H., Montalvo-Ortiz, J. L., and Pietrzak, R. H. (2022). Epigenome-wide association study of posttraumatic stress disorder identifies novel loci in u.s. military veterans. *Translational Psychiatry*, 12(1):65.

Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., and Beck, S. (2014). Champ: 450k chip analysis methylation pipeline. *Bioinformatics*, 30(3):428–430.

Murphy, T. M. and Mill, J. (2014). Epigenetics in health and disease: Heralding the ewas era. *The Lancet*, 383(9933):1952–1954.

Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samaras, K., V. Lord, R., Clark, S. J., and Molloy, P. L. (2015). De novo identification of differentially methylated regions in the human genome. *Epigenetics & Chromatin*, 8(1):6.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raffington, L., Tanksley, P. T., Vinnik, L., Sabhlok, A., Patterson, M. W., Mallard, T., Malanchini, M., Ayorech, Z., Tucker-Drob, E. M., and Paige Harden, K. (2023). Associations of dna-methylation measures of biological aging with social disparities in child and adolescent mental health. *Clinical Psychological Science*.

Rakyan, V. K., Down, T. A., Balding, D. J., and Beck, S. (2011). Epigenome-wide association studies for common human diseases. *Nature Reviews Genetics*, 12(8):529–541.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, 43(7):e47–e47.

Ruscio, J. and Kaczetow, W. (2008). Simulating multivariate nonnormal data using an iterative algorithm. *Multivariate behavioral research*, 43(3):355–381.

Song, J. and Kuan, P.-F. (2022). A systematic assessment of cell type deconvolution algorithms for dna methylation data. *Briefings in Bioinformatics*, 23(6).

Teschendorff, A. E., Breeze, C. E., Zheng, S. C., and Beck, S. (2017). A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. *BMC Bioinformatics*, 18(1):105.

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in illumina infinium 450 k dna methylation data. *Bioinformatics*, 29(2):189–196.

Tsai, P.-C. and Bell, J. T. (2015). Power and sample size estimation for epigenome-wide association scans to detect differential dna methylation. *International Journal of Epidemiology*, 44(4):1429–1441.

Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W. R., Kunze, S., Tsai, P.-C., Ried, J. S., Zhang, W., Yang, Y., Tan, S., Fiorito, G., Franke, L., Guarrera, S., Kasela, S., Kriebel, J., Richmond, R. C., Adamo, M., Afzal, U., others, and Chambers, J. C. (2017). Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*, 541(7635):81–86.

Wei, S., Tao, J., Xu, J., Chen, X., Wang, Z., Zhang, N., Zuo, L., Jia, Z., Chen, H., Sun, H., Yan, Y., Zhang, M., Lv, H., Kong, F., Duan, L., Ma, Y., Liao, M., Xu, L., Feng, R., others, and Jiang, Y. (2021). Ten years of ewas. *Advanced Science*, 8(20):2100727.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Zhao, S., Li, C.-I., Guo, Y., Sheng, Q., and Shyr, Y. (2018). Rnaseqsamplesize: Real data based sample size estimation for rna sequencing. *BMC Bioinformatics*, 19(1):191.