

RESEARCH

Open Access



# DSLE2 random-effects meta-analysis model for high-throughput methylation data

Nan Wang<sup>1,2</sup>, Yang Zhou<sup>1</sup>, Fengping Zhu<sup>3\*</sup> and Shuilin Jin<sup>1\*</sup>

## Abstract

**Background** With the rapid development of high-throughput sequencing technology, high-throughput sequencing data has grown on a massive scale, leading to the emergence of multiple public databases, such as EBI and GEO. Conducting secondary mining of high-throughput sequencing data in these databases can yield more valuable insights. Meta-analysis can quantitatively combine high-throughput sequencing data from the the same topic. It increases the sample size for data analysis, enhances statistical power, and results in more consistent and reliable conclusions.

**Results** This study proposes a new between-study variance estimator  $E_m$ . We prove that  $E_m$  is non-negative and  $E_m(\hat{\tau}_m^2)$  increases with the increase of  $\hat{\tau}_m^2$ , satisfying the general conditions of the between-study variance estimator. We get the DSLE2 (two-step estimation starting with the DSL estimate and the  $E_m$  in the second step) random-effects meta-analysis model based on the between-study variance estimator  $E_m$ . The accuracy and a series of evaluation metrics of the DSLE2 model are better than those of the other 6 meta-analysis models. DSLE2 model is applied to lung cancer and Parkinson's methylation data. Significantly differentially methylated sites identified by DSLE2 model and the genes with significantly differentially methylated sites are closely related to two diseases, indicating the effectiveness of DSLE2 random-effects model.

**Conclusions** This paper propose the DSLE2 random-effects meta-analysis model based on new between-study variance estimator  $E_m$ . The DSLE2 model performs well for methylation data.

**Keywords** Methylation sequencing data, Meta-analysis, Between-study variance, Random-effects model

## Background

DNA methylation refers to the addition of a methyl group at the CpG site of the DNA chain [1]. DNA methylation occurs at millions of CpG dinucleotide positions in the genome and changes with age and the external

environment [2]. It is an important epigenetic modification [3]. In the correlation analysis between methylation and traits, HM450K (Infinium Human Methylation 450 K BeadChip) or other Illumina technologies use the methylated fluorescence signal intensity ( $M$ ) and the unmethylated fluorescence signal intensity ( $U$ ) of the CpG site to calculate DNA methylation levels [4]. Two commonly used calculation methods for methylation levels are  $\beta = M/(M + U)$  and  $M = \log_2(M/U)$  [5]. In the minfi R package, the getBeta function and getM function can be used directly to calculate the  $\beta$  value and  $M$  value of every probe [6].

With the development of high-throughput sequencing technology, multiple public databases have been formed, such as the GEO (Gene Expression Omnibus) database and TCGA (The Cancer Genome Atlas database)

\*Correspondence:

Fengping Zhu

fpzhu@fudan.edu.cn

Shuilin Jin

jinsl@hit.edu.cn

<sup>1</sup> School of Mathematics, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China

<sup>2</sup> School of Advanced Manufacturing, Guangdong University of Technology, Jieyang, Guangdong, China

<sup>3</sup> Department of Neurosurgery, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai 200040, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

database [7]. People can obtain methylation sequencing data of the same disease from different databases [8]. Due to differences in experimental conditions and sample processing procedures, the results of methylation-trait association analysis of the same disease in different studies are different [9]. Meta-analysis can combine methylation data on the same topic to provide a more reliable list of significantly different methylation sites [10]. In meta-analysis, datasets from independent studies with the same purpose often have heterogeneity [11]. In the random-effects model, the comprehensive effect-size result depends on the quantification of heterogeneity between studies [12]. Therefore, the estimation of heterogeneity is the most critical step in the random-effects meta-analysis model [13, 14]. In addition to quantifying heterogeneity, investigating differences in experimental designs, participants, and interventions across studies can help to understand the causes of heterogeneity [15, 16]. This paper aims to estimate the size of the between-study variance in a random-effects meta-analysis model.

We propose the DSLE2 meta-analysis model based on new between-study variance estimator  $E_m$  for methylation high-throughput sequencing data. Under three hypothesis testing conditions, this model is compared with DSLR2 (two-step estimation starting with the DSL estimate and the  $R^2$  in the second step), DL (DerSimonian and Laird estimate), EB (Empirical Bayes estimate), HO (Hedges and Olkin estimate), RML (restricted maximum likelihood estimate), and SJ (Sidik and Jonkman estimate) random-effects meta-analysis models using sensitivity, and other evaluation metrics. The results show that the DSLE2 methylation meta-analysis model performs well under the first hypothesis testing condition. We apply the DSLE2 random-effects model to lung cancer and Parkinson's methylation data, further demonstrating the reliability of the DSLE2 meta-analysis model.

The proposed between-study variance estimator is designed to accurately quantify the true variability in effect sizes across studies, distinguishing it from random error. This advancement enhances the precision and reliability of meta-analysis models, offering significant benefits for research methodology, personalized medicine, study design, and resource allocation. By implementing and validating this new estimator, we can achieve a clearer understanding of effect consistency across studies, ultimately leading to more robust and generalizable conclusions. The primary purpose of the proposed between-study variance estimator is to improve the modeling of heterogeneity in methylation high-throughput sequencing data. FEM meta-analysis models often struggle to accurately capture the complex variability inherent in such data due to differences in study design, populations, and technical procedures.

Accurate effect size estimation is critical for identifying biologically significant methylation patterns and their potential clinical implications. To refine the estimation of true effect sizes by more accurately accounting for between-study variance is very important. Reliable identification of consistent methylation markers across studies is also crucial for developing personalized medical interventions. The improved model supports the discovery of biomarkers that are consistently reproducible, aiding in disease diagnosis, prognosis, and treatment customization.

## Methods

The random-effects model is a generalization of the fixed-effects model. The fixed effects model assumes that all studies included in the same meta-analysis have the same true effect size, while the random-effects model assumes that the true effect sizes of different studies included in the same meta-analysis obey a normal distribution. The between-study variance is a statistic used by the random-effects model to measure the heterogeneity between studies from the same topic. If the between study variance is 0, the random-effects model degenerates into the fixed-effects model. The between-study variance estimator of DSL (DerSimonian and Laird estimate) random-effects model is simple and the most commonly used method for estimating the between-study heterogeneity. In addition to being easy to calculate, the between-study variance estimator is also suitable for effect sizes of different dimensions. However, in practice, due to the possible occurrence of negative values, DSL between-study variance estimators are often truncated. In this paper, we present a non-truncated estimator of between-study variance  $E_m$ .

### DSLE2 random-effects meta-analysis model

We propose a general heterogeneity variance estimator for methylation sequencing data that is applicable to effect sizes at any scale and is non-negative. Assume that  $y_{1m}, y_{2m}, \dots, y_{km}$  are the effect sizes of  $k$  independent studies of methylation site  $m$ ;  $Q_m$  is a heterogeneity statistic for methylation site  $m$ ;  $\omega_{im}$  is the weight of methylation site  $m$  for  $i$ th study in the fixed-effects meta-analysis model;  $\omega_{im}^*$  is the weight of methylation site  $m$  for  $i$ th study in the random-effects meta-analysis model;  $\sigma_{im}^2$  is the within-study variability representing sampling errors of  $i$ th study;  $\hat{\tau}_{m,DSL}^2$  is the between-study variance estimator in DSL random-effects model;  $\hat{\mu}_{m,DSL}$  is the mean of effects in DSL random-effects model. And the between-study variance estimator  $E_m$  is

$$E_m = \frac{Q_m - S_{MM,m}}{\sum_{i=1}^k \omega_{im}^* - \sum_{i=1}^k (\omega_{im}^*)^2 / \sum_{i=1}^k \omega_{im}^*}$$

where

$$Q_m = \sum_{i=1}^k \omega_{im} (y_{im} - \hat{\mu}_m)^2$$

$$\hat{\mu}_m = \frac{\sum_{i=1}^k \omega_{im} y_{im}}{\sum_{i=1}^k \omega_{im}}$$

$$\omega_{im} = \frac{1}{\sigma_{im}^2}, i = 1, 2, \dots, k$$

$$S_{MM,m} = \sum_{i=1}^k \omega_{im}^* (y_{im} - \hat{\mu}_{m,DSL})^2$$

$$\hat{\mu}_{m,DSL} = \frac{\sum_{i=2}^k \omega_{im}^* y_{im}}{\sum_{i=1}^k \omega_{im}^*}$$

$$\omega_{im}^* = \frac{1}{\sigma_{im}^2 + \hat{\tau}_{m,DSL}^2}, i = 1, 2, \dots, k$$

The algorithm of DLSE2 meta-analysis model is as follows:

First, we use the fixed-effects model to calculate the weight and comprehensive effect value of each study

$$\omega_{im} = \frac{1}{\sigma_{im}^2}, i = 1, 2, \dots, k$$

and

$$\hat{\mu}_m = \frac{\sum_{i=1}^k \omega_{im} y_{im}}{\sum_{i=1}^k \omega_{im}}$$

Then, we calculate the heterogeneity statistic  $Q_m$

$$Q_m = \sum_{i=1}^k \omega_{im} (y_{im} - \hat{\mu}_m)^2$$

(1) We use the DSL random-effects model to calculate the between-study variance estimator, the weight of each study, and the corresponding comprehensive effect size

$$\hat{\tau}_{m,DSL}^2 = \max \left\{ 0, \frac{Q_m - (k - 1)}{S_{1m} - S_{2m}/S_{1m}} \right\}$$

$$S_{rm} = \sum_{i=1}^k \omega_{im}^r, r = 1, 2$$

$$\omega_{im}^* = \frac{1}{\sigma_{im}^2 + \hat{\tau}_{m,DSL}^2}, i = 1, 2, \dots, k$$

$$\hat{\mu}_{m,DSL} = \frac{\sum_{i=1}^k \omega_{im}^* y_{im}}{\sum_{i=1}^k \omega_{im}^*}$$

We further calculate  $S_{MM,m}$

$$S_{MM,m} = \sum_{i=1}^k \omega_{im}^* (y_{im} - \hat{\mu}_{m,DSL})^2$$

then

$$E_m = \frac{Q_m - S_{MM,m}}{\sum_{i=1}^k \omega_{im}^* - \sum_{i=1}^k \omega_{im}^{*2} / \sum_{i=1}^k \omega_{im}^*}$$

The weight of each study based on the between-study variance  $E_m$  is

$$\omega_{im,DSLE2} = \frac{1}{\sigma_{im}^2 + E_m}, i = 1, 2, \dots, k$$

The comprehensive effect is

$$\hat{\mu}_{m,DSLE2} = \frac{\sum_{i=1}^k \omega_{im,DSLE2} y_{im}}{\sum_{i=1}^k \omega_{im,DSLE2}}$$

The random-effect for each study is

$$\hat{\xi}_{ig} = \frac{E_m}{\sigma_{im}^2 + E_m} (y_{im} - \hat{\mu}_{m,DSLE2} (\hat{\tau}_{m,DSLE2}^2)), i = 1, 2, \dots, k$$

The variance of comprehensive effect is

$$\text{Var}(\hat{\mu}_{m,DSLE2}) = \frac{1}{\sum_{i=1}^k \omega_{im,DSLE2}}$$

The  $z$  statistic of comprehensive effect is

$$z_m(\hat{\mu}_{m,DSLE2}) = \frac{\hat{\mu}_{m,DSLE2}}{\sqrt{\text{Var}(\hat{\mu}_{m,DSLE2})}} \sim N(0, 1)$$

The lower and upper bounds of the  $100(1 - \alpha)\%$  confidence interval are

$$LL[\hat{\mu}_{m,DSLE2}] = \hat{\mu}_{m,DSLE2} - z_{(1-\alpha)/2(m)} \times \sqrt{\text{Var}[\hat{\mu}_{m,DSLE2}]}$$

and

$$UL[\hat{\mu}_{m,DSLE2}] = \hat{\mu}_{m,DSLE2} + z_{(1-\alpha)/2(m)} \times \sqrt{\text{Var}[\hat{\mu}_{m,DSLE2}]}$$

the  $p$ -value of one-sided test is

$$p = 1 - \Phi(z_m(\hat{\mu}_{m,DSLE2}))$$

the  $p$ -value of two-sided test is

$$p = 2[1 - \Phi(|z_m(\hat{\mu}_{m,DSLE2})|)]$$

where  $\Phi(\cdot)$  is the cumulative function of the standard normal distribution.

**Theorem 1** Assume that  $y_{1m}, y_{2m}, \dots, y_{nm}$  are the effect sizes of  $n$  independent studies of methylation site  $m$ , and the between-study variance estimator  $E_m$  is

$$E_m = \frac{Q_m - S_{MM,m}}{A} \tag{2}$$

where

$$Q_m = \sum_{i=1}^n \frac{1}{\sigma_{im}^2} \left( y_{im} - \frac{\sum_{i=1}^n \frac{1}{\sigma_{im}^2} y_i}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2}} \right)^2 \tag{3}$$

$$S_{MM,m} = \sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} \left( y_{im} - \frac{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} y_{im}}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}} \right)^2 \tag{4}$$

$$A = \sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} - \frac{\sum_{i=1}^n \left( \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} \right)^2}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}} \tag{5}$$

Then,  $E_m(\hat{\tau}_m^2)$  is monotone and non-decreasing with respect to  $\hat{\tau}_m^2$

**Proof** It can be obtained from (3)

$$\frac{dQ_m}{d\hat{\tau}_m^2} = 0 \tag{6}$$

It can be obtained from (4)

$$\frac{dS_{MM,m}(\hat{\tau}_m^2)}{d\hat{\tau}_m^2} = - \sum_{i=1}^n \left( \left( \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} \right) \left( y_{im} - \frac{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} y_{im}}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}} \right) \right)^2 \leq 0 \tag{7}$$

It can be obtained from (5)

$$\frac{dA(\hat{\tau}_m^2)}{d\hat{\tau}_m^2} = - \sum_{i=1}^n \frac{1}{(\sigma_{im}^2 + \hat{\tau}_m^2)^2} + \frac{2 \sum_{i=1}^n \frac{1}{(\sigma_{im}^2 + \hat{\tau}_m^2)^3} \sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} - \left( \sum_{i=1}^n \frac{1}{(\sigma_{im}^2 + \hat{\tau}_m^2)^2} \right)^2}{\left( \sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} \right)^2}$$

let

$$x_i = \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}$$

Then

$$x_i > 0$$

$$\begin{aligned} \frac{dA(\hat{\tau}_m^2)}{d\hat{\tau}_m^2} &= - \sum_{i=1}^n x_i^2 + \frac{2 \sum_{i=1}^n x_i^3 \sum_{i=1}^n x_i - \left( \sum_{i=1}^n x_i^2 \right)^2}{\left( \sum_{i=1}^n x_i \right)^2} \\ &= \frac{- \sum_{i=1}^n x_i^2 \left( \sum_{i=1}^n x_i \right)^2 + 2 \sum_{i=1}^n x_i^3 \sum_{i=1}^n x_i - \left( \sum_{i=1}^n x_i^2 \right)^2}{\left( \sum_{i=1}^n x_i \right)^2} \\ &\leq \frac{2 \left( \sum_{i=1}^n x_i^3 \sum_{i=1}^n x_i - \left( \sum_{i=1}^n x_i^2 \right)^2 \right)}{\left( \sum_{i=1}^n x_i \right)^2} \end{aligned}$$

Let

$$f(y) = \sum_{i=1}^n x_i^{(4-y)} \sum_{i=1}^n x_i^y$$

Then

$$\frac{df}{dy} = - \sum_{i=1}^n x_i^{(4-y)} \ln x_i \sum_{i=1}^n x_i^y + \sum_{i=1}^n x_i^{(4-y)} \sum_{i=1}^n x_i^y \ln x_i$$

So, when  $y = 2, \frac{df}{dy} = 0$ .

We can get that  $y = 2$  is the stagnation point of  $f(y)$ .

Then, we can get

$$\sum_{i=1}^n x_i^3 \sum_{i=1}^n x_i \leq \left(\sum_{i=1}^n x_i^2\right)^2$$

So

$$\frac{dA(\hat{\tau}_m^2)}{d\hat{\tau}_m^2} \leq \frac{2\left(\sum_{i=1}^n x_i^3 \sum_{i=1}^n x_i - \left(\sum_{i=1}^n x_i^2\right)^2\right)}{\left(\sum_{i=1}^n x_i\right)^2} \leq 0 \quad (8)$$

$$Q_m = \sum_{i=1}^n \frac{1}{\sigma_{im}^2} \left( y_{im} - \frac{\sum_{i=1}^n \frac{1}{\sigma_{im}^2} y_i}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2}} \right)^2$$

$$S_{MM,m} = \sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} \left( y_{im} - \frac{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} y_{im}}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}} \right)^2$$

It can be obtained from (2)

$$\frac{d(E_m(\hat{\tau}_m^2))}{d\hat{\tau}_m^2} = \frac{d\left(\frac{Q_m - S_{MM,m}(\hat{\tau}_m^2)}{A(\hat{\tau}_m^2)}\right)}{d\hat{\tau}_m^2} = \frac{-\left(A(\hat{\tau}_m^2) \frac{dS_{MM,m}(\hat{\tau}_m^2)}{d\hat{\tau}_m^2} + (Q_m - S_{MM,m}) \frac{dA(\hat{\tau}_m^2)}{d\hat{\tau}_m^2}\right)}{\left(A(\hat{\tau}_m^2)\right)^2} \quad (9)$$

It can be obtained from (3) and (4)

$$Q_m = S_{MM,m}(0)$$

It is attainable via (7)

$$Q_m - S_{MM,m}(\hat{\tau}_m^2) \geq 0 \quad (10)$$

It is available through (6), (7), (8) and (10)

$$\frac{d(E_m(\hat{\tau}_m^2))}{d\hat{\tau}_m^2} \geq 0$$

$$A = \sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} - \frac{\sum_{i=1}^n \left(\frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}\right)^2}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}}$$

Then,  $E_m \geq 0$ .

**Proof** Because  $A(\hat{\tau}_m^2) = \sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} - \frac{\sum_{i=1}^n \left(\frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}\right)^2}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}} > 0$

$$Q_m - S_{MM,m}(\hat{\tau}_m^2) \geq 0$$

we have

$$E_m(\hat{\tau}_m^2) = \frac{\sum_{i=1}^n \frac{1}{\sigma_{im}^2} \left( y_{im} - \frac{\sum_{i=1}^n \frac{1}{\sigma_{im}^2} y_i}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2}} \right)^2 - \sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} \left( y_{im} - \frac{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} y_{im}}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}} \right)^2}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2} - \frac{\sum_{i=1}^n \left(\frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}\right)^2}{\sum_{i=1}^n \frac{1}{\sigma_{im}^2 + \hat{\tau}_m^2}}} = \frac{Q_m - S_{MM,m}(\hat{\tau}_m^2)}{A(\hat{\tau}_m^2)} \geq 0$$

So,  $E_m(\hat{\tau}_m^2)$  increases with the increase of  $\hat{\tau}_m^2$ .

**Theorem 2** Assume that  $y_{1m}, y_{2m}, \dots, y_{nm}$  are the effect sizes of  $n$  independent studies of methylation site  $m$ , and the between-study variance estimator  $E_m$  is

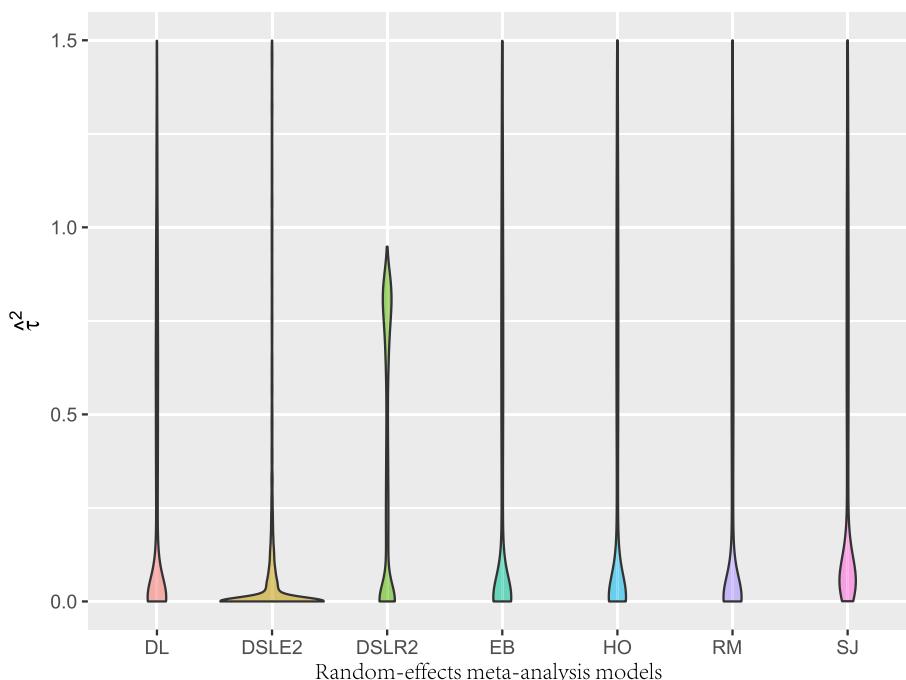
$$E_m = \frac{Q_m - S_{MM,m}}{A}$$

where

## Results

### Simulation study

We simulated five sets of meta-analysis data. Each set of data has  $K$  ( $K = 4, 6, 8, 10, 12$ ) studies. Each study contains 20,000 methylation sites and 30 sample data. The first 15 samples are control data, and the last 15 sample data are experimental data. The methylation sites of each sample in each study are divided into 100 categories ( $c = 1, 2, \dots, 100$ ), with 200 methylation sites



**Fig. 1** Violin plot of between-study variance for seven random-effects meta-analysis models

in each category. The top 9000 methylation sites of each study are divided into  $K$  groups ( $k_m = 1, 2, \dots, K$ ). The first  $9000/K$  methylation sites belong to the first group ( $k_m = 1$ ), and the  $9000/K + 1$  to  $9000 * 2/K$  methylation sites belong to the second group ( $k_m = 2$ ), and so on, the methylation site from  $9000(K - 1)/K + 1$  to 9000 methylation sites belong to the  $K$ -th group ( $k_g = K$ ), the 9001 to 20,000 methylation sites belong to the 0th group ( $k_m = 0$ ). The algorithm for simulating data is summarized as follows:

First, we randomly sampled the methylation sites of the  $c(1 \leq c \leq 100)$  th class of the  $k(1 \leq k \leq K)$  th study  $k(1 \leq k \leq K)$ , where  $k(1 \leq k \leq K)$ ,  $I_{200 \times 200}$  is identity matrix,  $J_{200 \times 200}$  is a matrix with all elements 1,  $W^{-1}$  is the inverse of the Wishart distribution, and  $\sum_{c_k}^{\prime}$  is normalized to  $\sum_{c_k}$ , such that all its diagonal elements are 1. Then, sample the methylation level of the  $c(1 \leq c \leq 100)$  th class site of the  $n(1 \leq n \leq 30)$  sample of the  $k$  th study  $(X'_{mc1nk}, X'_{mc2nk}, \dots, X'_{mc200nk})^T \sim MVN(0, \sum_{c_k})$ ,  $1 \leq k \leq K$ .

We performed differential methylation settings on the first 9000 methylation sites, randomly sampling  $\delta_{mk} \in (0, 1)$ , so that  $\sum_{k=1}^K \delta_{mk} = k_m (k_m = 1, 2, \dots, K)$ . When  $\delta_{mk} = 1$ , site  $m$  of the  $k$  th study is a significantly differentially methylated site. When  $\delta_{mk} = 0$ , site  $m$  of the  $k$  th study is a non-significantly differential methylation site. This paper conducts random sampling  $\mu_{mk} \sim U(0.5, 3)$ . The methylation level of the control group remains

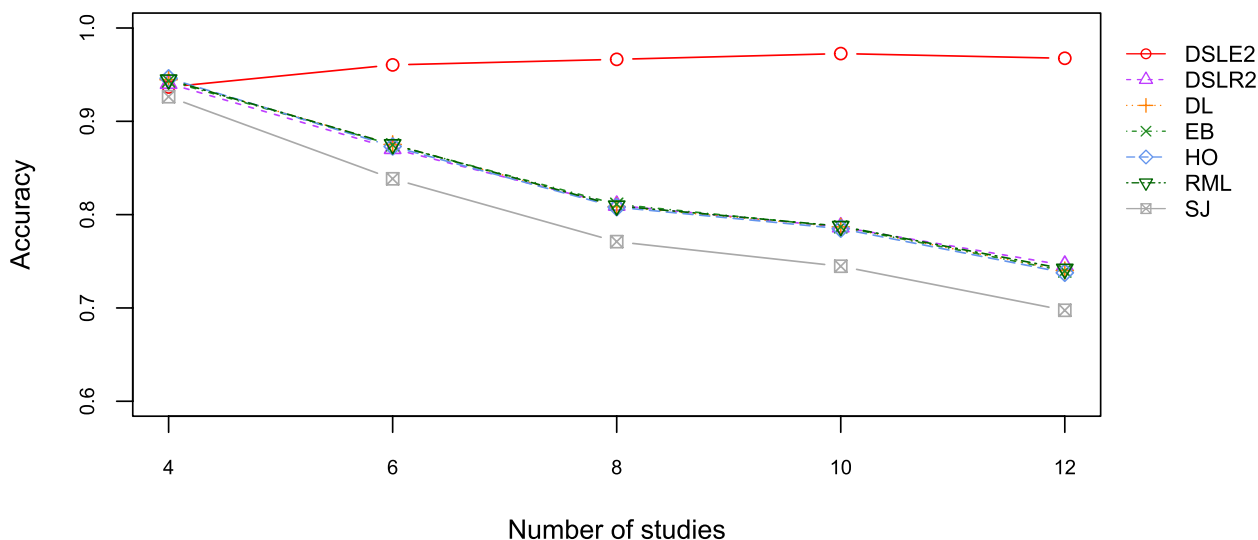
unchanged, and the methylation expression level of the experimental group is  $Y_{mjk} = X'_{m(n+N)k} + \mu_{mk} * \delta_{mk}$ , where  $1 \leq m \leq 9000, 1 \leq n \leq 15, 1 \leq k \leq K$ .

We compared the distribution of between study variance estimators of the DSLE2 random-effects model and six other meta-analysis models. As can be seen from Fig. 1, compared with the other six random-effects models, the between-study variance estimators of the DSLE2 meta-analysis model is relatively concentrated. The between-study variance estimators of DSLE2, DL, EB, HO, RML, and SJ random-effects models are mostly below 0.5. Compared with the other 6 random-effects models, the DSLE2 random-effects model has a relatively larger number of between-study variances distributed around 0.

We also compared the accuracy, false negative rate, negative prediction rate, recall rate, Matthews correlation coefficient, PCMiss (Prediction-conditioned miss) value, SAR value, and F value of the DSLE2 meta-analysis model with DL, EB, HO, RML, and SJ meta-analysis models under three hypothesis.

**Accuracy**

We compared the accuracy of seven random-effects models under three hypothesis (see Fig. 2, Figure S1 and Figure S2). Under the first hypothesis, the DSLE2 meta-analysis model has the highest accuracy, the SJ random-effects model has the lowest accuracy, and the DSLR2,



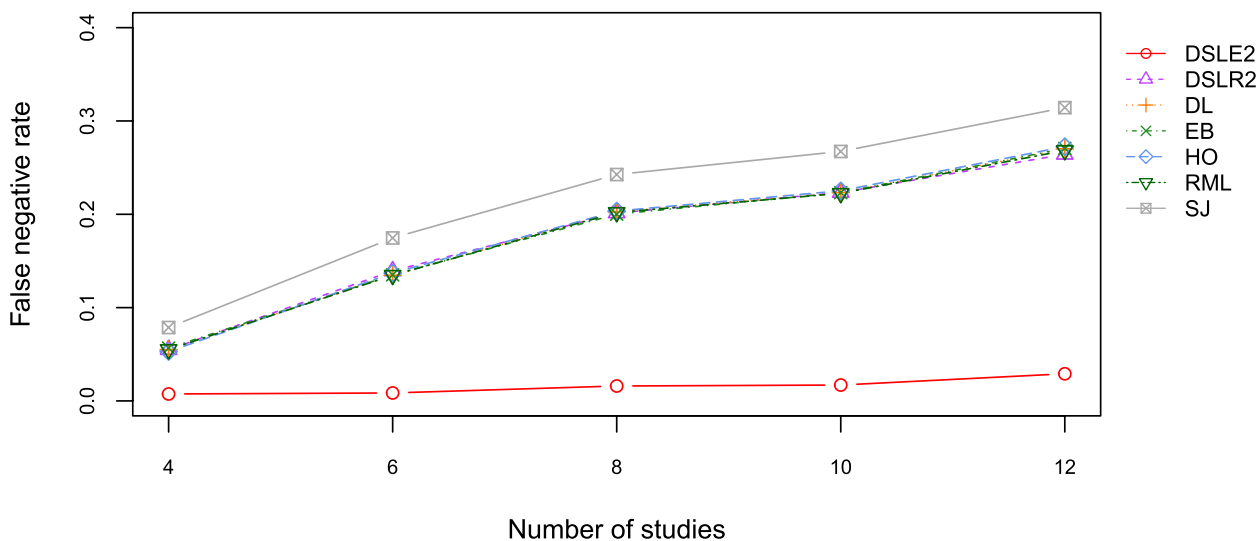
**Fig. 2** Plot of accuracy for 7 random-effects meta-analysis models under the first hypothesis

DL, EB, HO, and RML meta-analysis models have the similar accuracy. The accuracy of the DSLR2, DL, EB, HO, RML, and SJ meta-analysis models decreases as the number of studies increases. There is no obvious decreasing trend as the number increases for DSLE2 random-effects model. Under the second hypothesis, the DSLE2 meta-analysis model has the lowest accuracy, the SJ meta-analysis model has the highest accuracy. The accuracy of the DSLR2, DL, EB, HO, RML, and SJ meta-analysis models increases with the number of studies increases. Under the third hypothesis, the DSLE2 random-effects model has the lowest accuracy, and the DSLR2 meta-analysis model has the highest accuracy.

**False negative rate**

The false negative rate, also known as the second type error rate, refers to the proportion of the number of significantly differential methylation sites predicted by the model as non-significantly differential methylation sites to the number of all significantly differential methylation sites. We compared the false negative rate of seven meta-analysis models under three hypothesis (see Fig. 3, Figure S3 and Figure S4).

Under the first hypothesis, the false negative rate of the DSLE2 model is the lowest among the seven random-effects models; the false negative rate of the SJ model is the highest among the seven random-effects



**Fig. 3** Plot of false negative rate curves for seven random-effects meta-analysis models under the first hypothesis



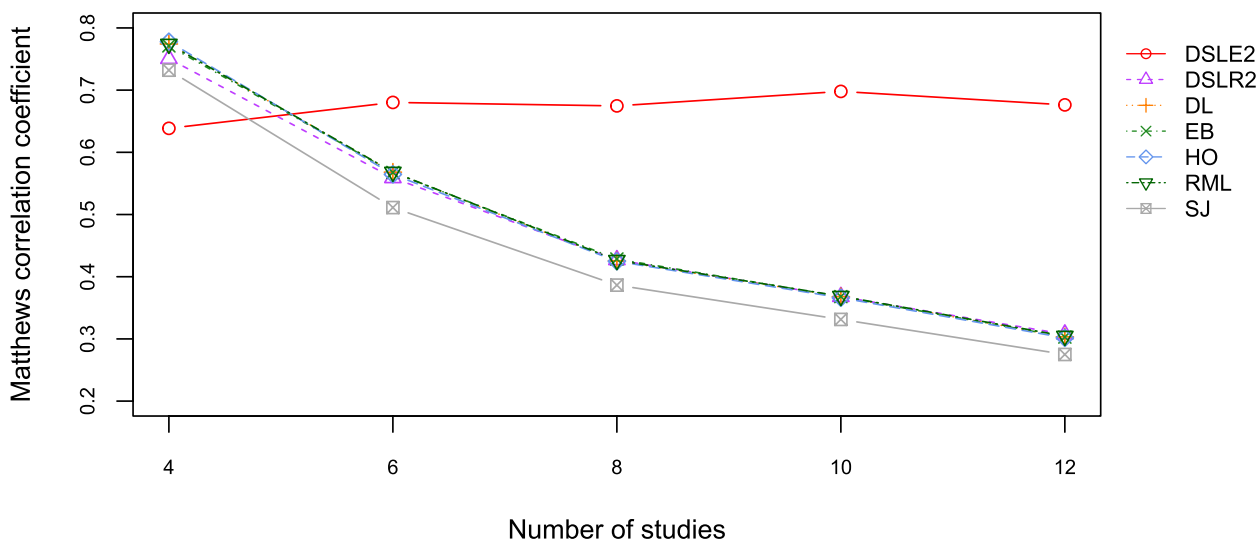
models. The false negative rates of the DSLR2, DL, EB, HO, and RML meta-analysis models are relatively close, and they are all lower than the false-negative rates of the SJ meta-analysis model, and they are all higher than the false negative rate of the DSLE2 meta-analysis model. The false negative rate of the DSLR2, DL, EB, HO, RML, and SJ meta-analysis models increases with the increase of the number of studies. Under the second hypothesis, the false negative rate of the DSLE2 meta-analysis model is the lowest among the seven random-effects models, and the false negative rates of the DL, EB, HO, and RML random-effects models are relatively close, and the false negative rate of the SJ model is the highest among the seven meta-analysis models.

Under the third hypothesis, the false negative rate of the DSLE2 meta-analysis model is the lowest among the seven meta-analysis models; the false negative rate of the SJ random-effects model is the highest among the seven random-effects models. The false negative rates of the DSLR2, DL, EB, HO, and RML meta-analysis models are relatively close, and they are all lower than the false negative rates of the SJ random-effects model, and they are all higher than the false negative rate of the DSLE2 meta-analysis model. The false negative rate of the DSLR2, DL, EB, HO, RML, and SJ random-effects models increases with the increase of the number of studies.

**Matthews correlation coefficient**

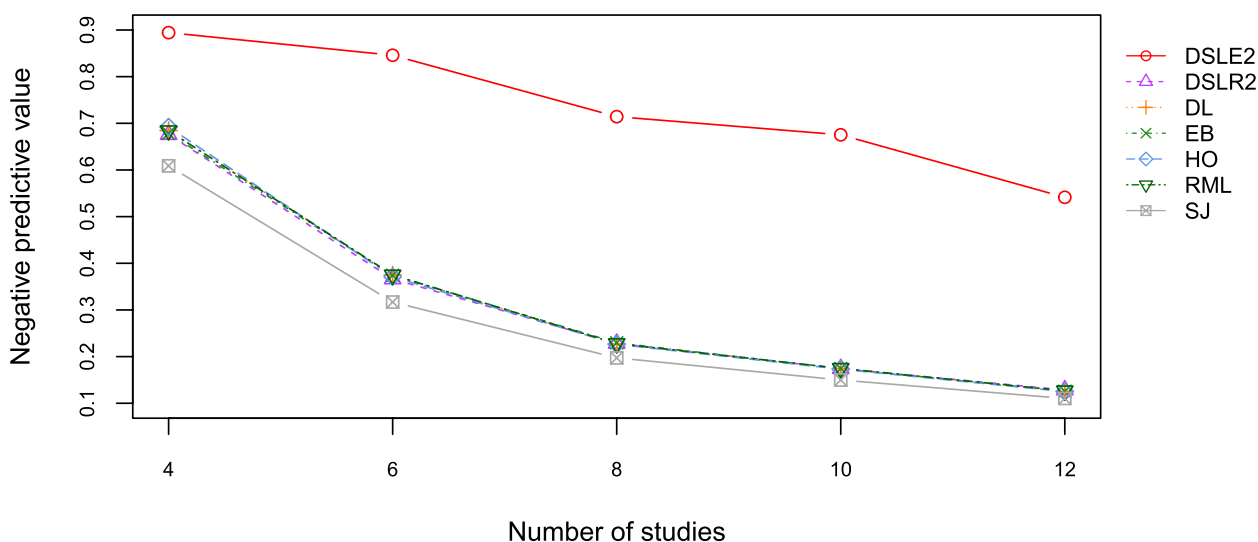
Matthews correlation coefficient is the correlation coefficient that describes the actual class and the predicted class. We compared the Matthews correlation coefficients of seven random-effects models under three hypothesis testing conditions (see Fig. 4, Figure S5 and

Figure S6). Under the first hypothesis, when the number of studies is greater than 4, the Matthews correlation coefficient of the DSLE2 meta-analysis model is the highest among the Matthews correlation coefficient of the seven random-effects models. The Matthews correlation coefficient of SJ meta-analysis model is the lowest among the Matthews correlation coefficients of the seven meta-analysis models. The Matthews correlation coefficients of DSLR2, DL, EB, HO, and RML random-effects models are close, which are lower than the Matthews correlation coefficient of the DSLE2 meta-analysis model and higher than the Matthews correlation coefficient of the SJ random-effects model. The Matthews correlation coefficient of DSLR2, DL, EB, HO, RML, SJ meta-analysis models decreases as the number of studies increases. Under the second hypothesis, the Matthews correlation coefficient of the DSLE2 random-effects model is the lowest among the Matthews correlation coefficients of the seven meta-analysis models; the Matthews correlation coefficient of the SJ random-effects model is the highest among the Matthews correlation coefficients of the seven random-effects models; the Matthews correlation coefficients of DSLR2, DL, EB, HO, and RML meta-analysis models are relatively close. The Matthews correlation coefficients of HO, RML, and SJ random-effects models increase with the increase of the number of studies. Under the third hypothesis, the Matthews correlation coefficient of the DSLE2 random effects meta-analysis model is the lowest among the Matthews correlation coefficients of the seven meta-analysis models, and the Matthews correlation coefficient of the DSLR2 meta-analysis model is the highest among the Matthews correlation coefficients of the seven random-effects meta-analysis models.



**Fig. 4** Plot of Matthews correlation coefficient curves for 7 random-effects models under the first hypothesis





**Fig. 5** Plot of negative predictive value curves for 7 random-effects models under the first hypothesis

**Negative predictive value**

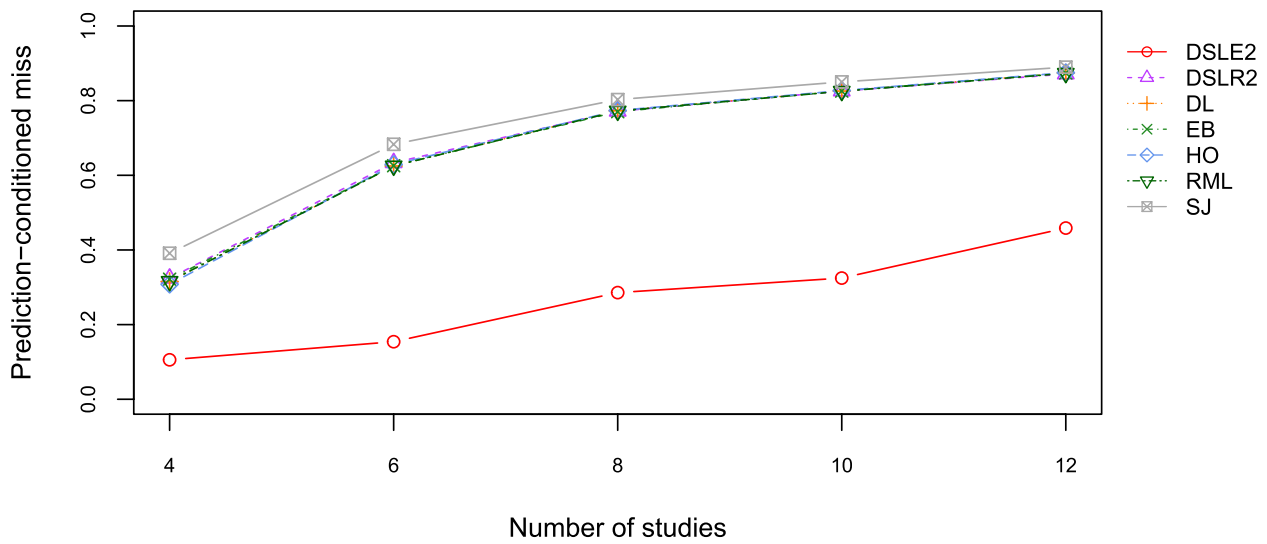
Negative predictive value refers to the proportion of the number of non-significantly differentially methylated sites correctly predicted by meta-analysis model to the number of all non-significantly differentially methylated sites predicted by meta-analysis model. We compared the negative predictive values of seven meta-analysis models (see Fig. 5, Figure S7 and Figure S8). Under the first hypothesis, the negative predictive values of the DSLE2 meta-analysis model are the highest among seven random-effects models; the negative predictive values of the SJ meta-analysis model are the lowest among seven random-effects models. The negative predictive values of the DSLR2, DL, EB, HO, and RML random-effects models are relatively close, and they are all lower than the negative predictive value of the DSLE2 random-effects model and higher than the negative predictive values of SJ random-effects model. Under the second hypothesis, the negative predictive values of the SJ meta-analysis model are the lowest among the seven random-effects models. The negative predictive values of the DL, EB, HO, and RML random-effects models are relatively close. Under the third hypothesis, the negative predictive values of the DSLE2 meta-analysis model are the highest among the seven random effects models; the negative predictive values of the SJ random-effects model are the lowest among the seven meta-analysis models. The negative predictive rates of DSLR2, DL, EB, HO, and RML meta-analysis models are relatively close.

**Prediction-conditioned miss**

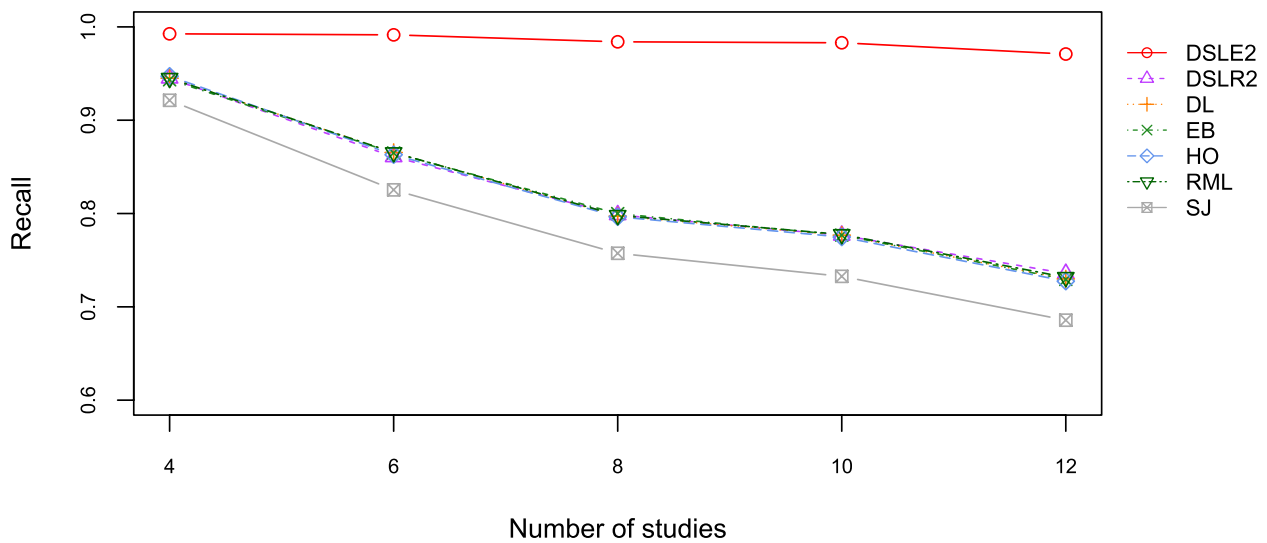
PCMiss (Prediction-Conditioned Miss) refers to the proportion of the number of sites incorrectly predicted as non-significantly differentially methylated to the number of all non-significantly differentially methylated sites predicted by the meta-analysis model. We compared the PCMiss values of seven meta-analysis models under three hypothesis (see Fig. 6, Figure S9 and Figure S10). Under the first hypothesis, the PCMiss values of the DSLE2 random-effects model are the lowest among the PCMiss values of the seven meta-analysis models; the PCMiss values of the SJ random-effects model are the highest among the seven meta-analysis models. The PCMiss values of the DSLR2, DL, EB, HO, and RML random-effects meta-analysis models are relatively close; the PCMiss values of the seven models all increase with the increase of the number of studies. Under the second hypothesis, the PCMiss values of the SJ model are the highest among the PCMiss values of the seven models. The PCMiss values of the DL, EB, HO, and RML random-effects models are close. Under the third hypothesis testing, the PCMiss values of the DSLE2 model are the lowest among the PCMiss values of seven random effects meta-analysis models; the PCMiss values of the SJ meta-analysis model are the highest among the seven random-effects models. The PCMiss values of the DSLR2, DL, EB, HO, and RML meta-analysis models are relatively close.

**Recall**

This paper measured the recall rates of seven random-effects models under three hypothesis (see Fig. 7, Figure



**Fig. 6** Plot of PCMiss-value curves for 7 random-effects models under the first hypothesis



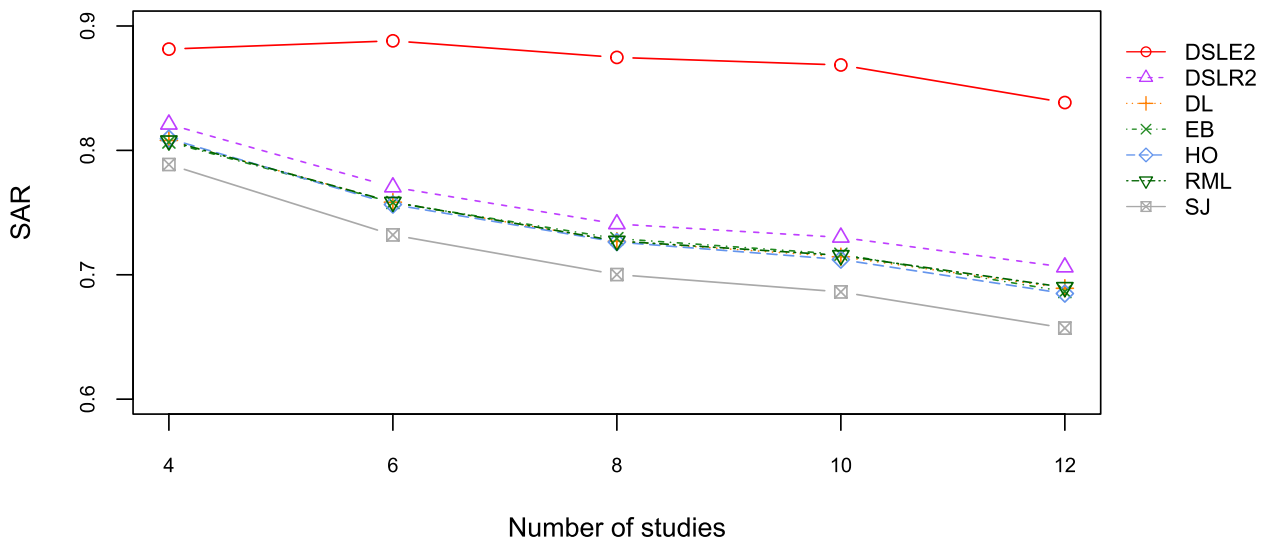
**Fig. 7** Plot of recall curves for 7 random-effects meta-analysis models under the first hypothesis

S11 and Figure S12). Under the first hypothesis, the recall rates of the DSLE2 meta-analysis model are the highest among seven random effects meta-analysis models; the recall rates of the SJ random-effects model are the lowest among seven meta-analysis models. The recall rates of the DL, EB, HO, and RML meta-analysis models are relatively close. The recall rates of the DSLR2, DL, EB, HO, RML, and SJ random-effects models decreases as the number of studies increases. Under the second hypothesis, the recall rates of the DSLE2 model are the highest among the seven meta-analysis models, followed by the recall rates of the DSLR2 model, and the recall rates of the SJ are the lowest among seven models. The recall

rates of the DL, EB, HO, and RML random-effects models are relatively close. Under the third hypothesis, the recall rates of the DSLE2 meta-analysis model are the highest among the seven random-effects models, and the recall rate of the SJ model is the lowest among the seven meta-analysis models. The recall rates of DSLR2, DL, EB, HO, and RML random-effects models are relatively close.

**SAR**

SAR combines accuracy, AUC (the area under the receiver operating characteristic curve) value, and root mean square error. The SAR value is more robust than a single indicator. We computed the SAR values of seven



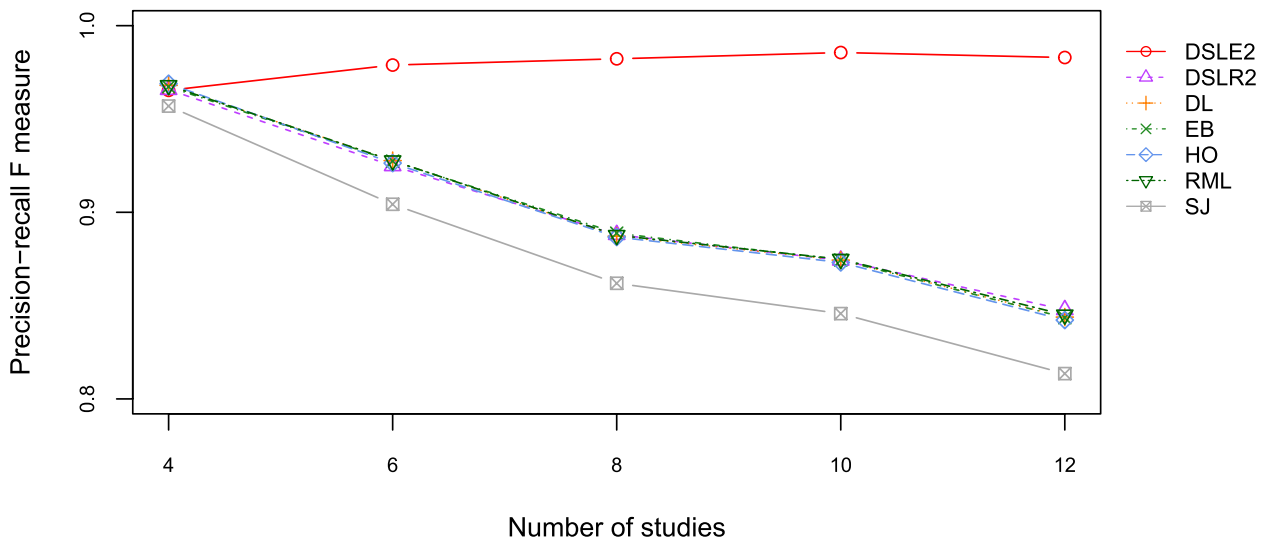
**Fig. 8** Plot of SAR-value curves for 7 random-effects meta-analysis models under the first hypothesis

meta-analysis models under three hypothesis (see Fig. 8, Figure S13 and Figure S14). Under the first hypothesis, the SAR values of the DSLE2 random-effects model are the highest among the SAR values of seven meta-analysis models, and the SAR values of SJ model are the lowest among the seven random-effects models. The SAR values of DL, EB, HO, and RML meta-analysis models are relatively close. Under the second hypothesis, the SAR values of DSLR2 random-effects model are the highest among the SAR values of seven meta-analysis models, and the SAR values of the DSLE2 model are lowest among the seven random-effects models. The SAR values of the DL, EB, HO, RML, and SJ random-effects meta-analysis

models are relatively close. The SAR values of the seven meta-analysis models all increase with the increase of the number of studies. Under the third hypothesis, the SAR values of the DSLR2 meta-analysis model are the highest among the SAR values of seven random-effect models. Moreover, the SAR values of the DL, EB, HO, and RML meta-analysis models are relatively close.

**Precision-recall F measure**

Precision and recall indicators sometimes conflict with each other. In this case, precision and recall indicators need to be considered comprehensively. The most common method is to calculate the F value

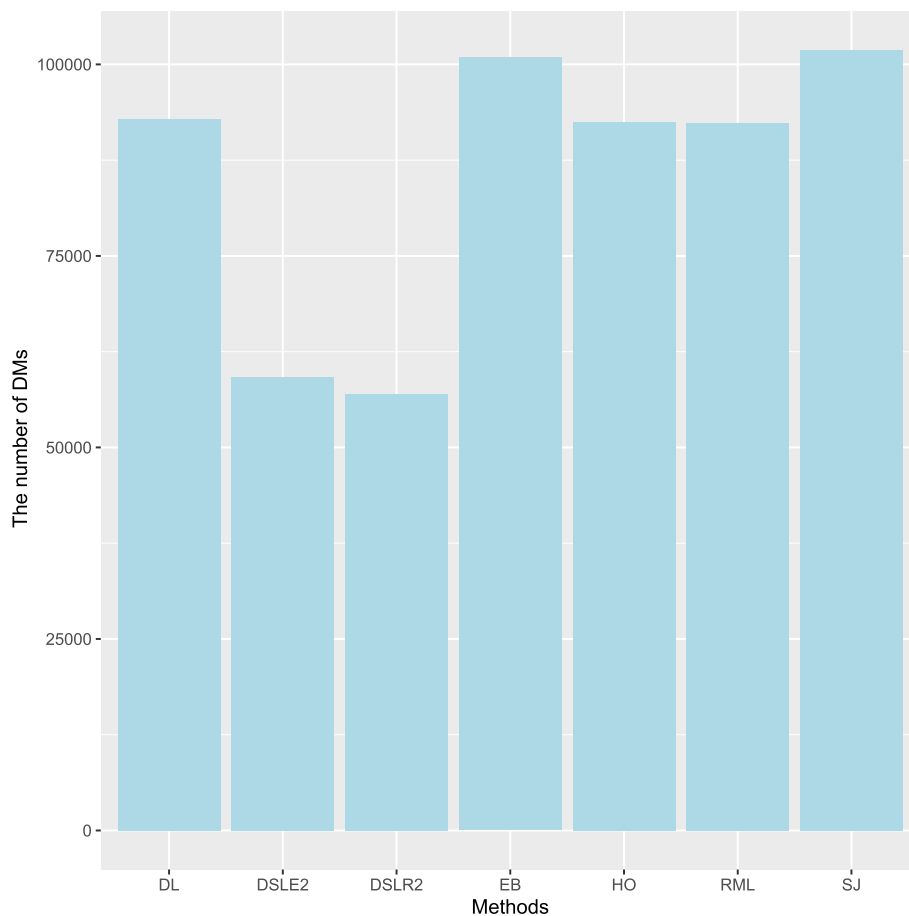


**Fig. 9** Plot of F-value curves for 7 random-effects meta-analysis models under the first hypothesis

of precision and recall. We computed the  $F$  values of 7 random-effects models under three hypothesis (see Fig. 9, Figure S15 and Figure S16). Under the first hypothesis, the  $F$  values of the DSLE2 random-effects model are the highest among the  $F$  values of seven random-effects models, and the  $F$  values of the SJ meta-analysis model are the lowest among the seven meta-analysis models. The  $F$  values of DSLR2, DL, EB, HO, and RML meta-analysis models are relatively close. Under the second hypothesis, the  $F$  values of the SJ meta-analysis model are the highest among the  $F$  values of seven random-effects models, and the  $F$  values of the DSLE2 random-effects model are the lowest among the seven meta-analysis models. The  $F$  values of DSLR2, DL, EB, HO, and RML random-effects models are relatively close. Under the third hypothesis, the  $F$  values of the DSLE2 random-effects model is the lowest among the  $F$  values of seven meta-analysis models.

#### Application of DSLE2 random-effects meta-analysis model to lung cancer methylation data

Lung cancer is the second most common cancer worldwide and the most common cancer among men. According to the World Cancer Statistics Center, more than 2.2 million cases of lung cancer occur every year [17]. We collected three sets of lung cancer methylation data in the GEO database: GSE63704, GSE83842, and GSE85845. The EPIC and 450 K methylation data of lung cancer in TCGA database were also collected. A total of 927 samples and 156,680 methylation sites were analyzed. The distribution of the number of significantly differential methylation sites determined by 7 random-effects models is shown in the Fig. 10. The number of significantly differential methylation sites determined by the DSLE2 random-effects model is more than the number of significantly differential methylation sites determined by DSLR2 and is less than the number of significantly differential methylation sites determined by DL, EB, HO, RML, and SJ meta-analysis models.



**Fig. 10** Histogram of the number distribution of significantly differential methylation sites identified by 7 random-effects meta-analysis models

The DSLE2 random-effects model identified 59,146 significantly differential methylation sites, of which 2,230 significantly differential methylation sites distributed in 1stExon of 3,039 genes, and 2,003 significantly differential methylation sites were distributed in 3'UTR of 1,660 genes, and 4899 significantly differentially methylated sites distributed in the 5'UTR of 2754 genes, and 21,029 significantly differentially methylated sites distributed in the Body of 7056 genes, and 7500 significantly differentially methylated sites distributed in the TSS1500 of 5113 genes, and 6210 significantly differentially methylated sites distributed in TSS200 of 3963 genes. Research shows that lung cancer is closely related to A2BP1 [8], AACS [18], DNAH10 [19], PINK1 [20], and other genes with significantly differential methylation sites. The significantly differentially methylated sites identified by the DSLE2 meta-analysis model may affect the expression of the corresponding genes.

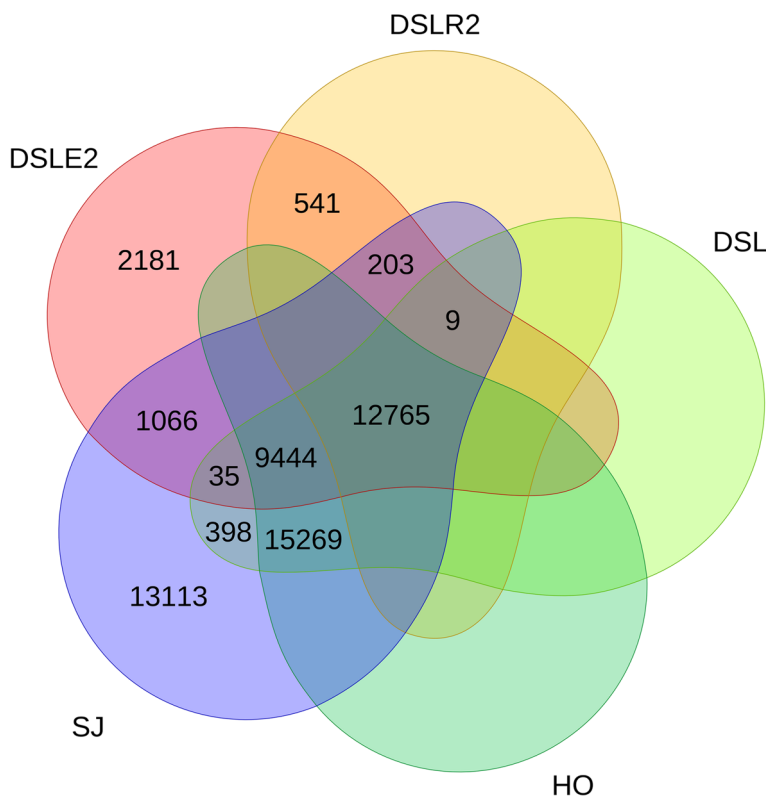
**Application of DSLE2 meta-analysis model to Parkinson’s methylation data**

Parkinson’s disease (PD) is a chronic neurodegenerative disease [21]. According to the World Health Organization (WHO), there were 2.5 million and 6.1 million cases of PD worldwide in 1990 and 2016 [22]. However,

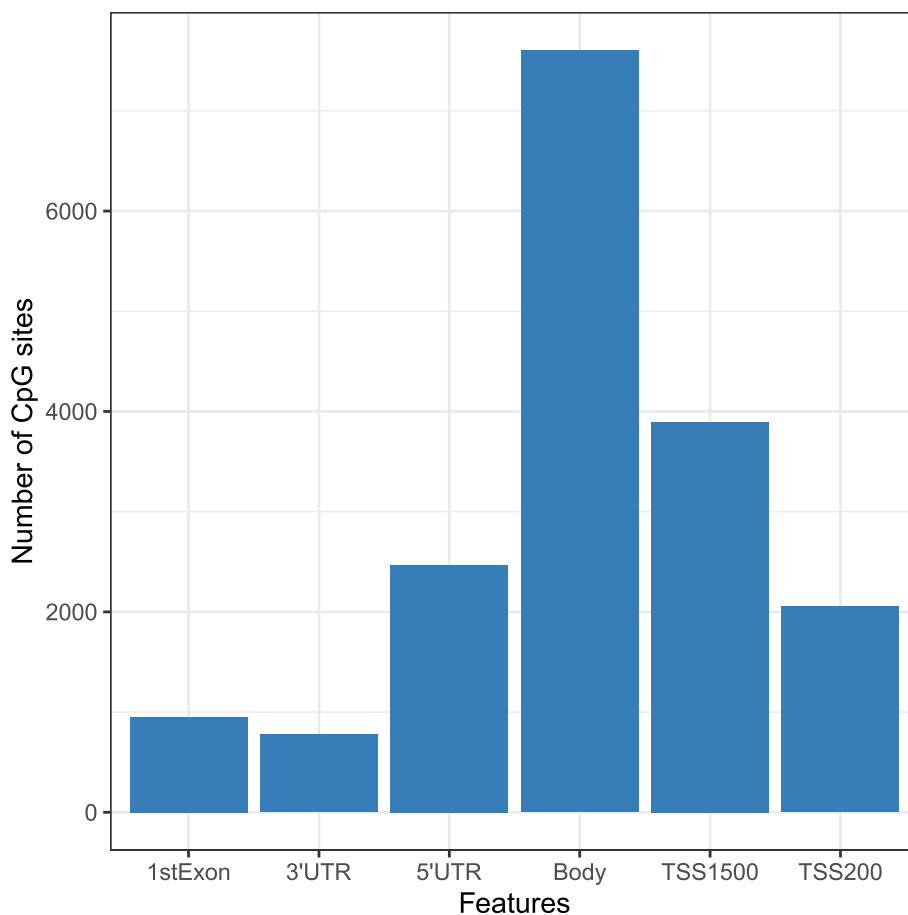
the number of PD patients increased significantly to 8.5 million, and it is estimated that by 2040, the number of PD patients worldwide may exceed 17 million [23]. In 2019, PD caused 5.8 million disabilities, an 81% increase since 2000, and 329,000 deaths, an increase of more than 100 percent since 2000 [23]. PD causes serious trouble to people’s life, but there is no clear cause and effective treatment. DSLE2 random-effects model was used to further study the causes of PD.

We collected five sets of PD methylation data in the GEO database: GSE72774, GSE72776, GSE111629, GSE145361 and GSE165081. A total of 3,080 samples and 161,261 methylation sites were analyzed. The DSLE2 model identified 26,244 significantly differential methylation sites (Fig. 11). Most of significantly differential methylation sites identified by DSLE2, SJ, HO, DSL, DSLR2 meta-analysis methods were same. DSLE2 model independently identified 2181 significantly differential methylation sites.

We further analyzed the location of significantly differential methylation sites on genes (Figs. 12 and 13). Most of the significant differential methylation sites were located on the gene body, followed by TSS1500 and 5'UTR (Fig. 12). Relatively few Significantly differential methylation sites were located at 3'UTR and the



**Fig. 11** Venn of significantly differential methylation sites identified by 5 random-effects meta-analysis models



**Fig. 12** Histogram of significantly differential methylation sites on genes. The horizontal axis refers features, and the vertical axis is the number of significantly differential methylation sites

first exon (Fig. 12). Most of the significantly differential methylation sites are uniquely located on a single gene and only a small number of genes contain multiple significantly differential methylation sites (Fig. 13). There are 51, 62 and 122 significantly differentially methylation sites for BCOR, HDAC4 and PRDM16, respectively. Research shows that PD is closely related to BCOR [24], HDAC4 [25], PRDM16 [26], and other genes with significantly differential methylation sites.

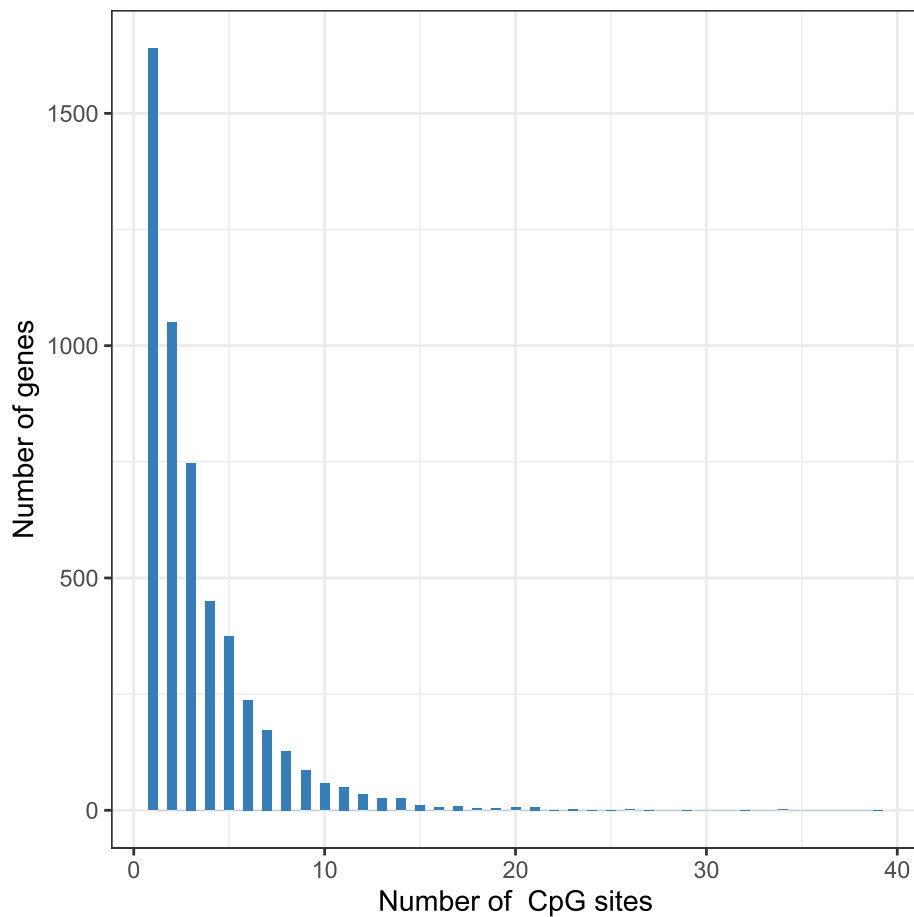
### Discussion and conclusion

For high-throughput methylation sequencing data, we proposed the DSLE2 random effects meta-analysis model based on the between-study variance estimator  $E_m$ . Under the alternative hypothesis that the effect sizes of all studies are not 0, the DSLE2 meta-analysis model performs better than the other 6 random-effects models in terms of accuracy, negative prediction rate, recall rate, Matthews correlation coefficient, PCMiss value, and SAR value, F value.

We applied the DSLE2 meta-analysis model to the lung cancer methylation datasets. 59,146 significantly differential methylation sites were identified, with 2,230, 2,230, 4899, 21,029, 7500, and 6210 significantly differentially methylated sites were located on 1stExon, 3'UTR, 5'UTR, Body, TSS1500, and TSS200 of 3039, 1660, 2754, 7056, 5113, and 3963 genes, respectively. Studies have shown that lung cancer is closely related to genes such as A2BP1, AACCS, and DNAH10 with significantly differential methylation sites.

We further applied the DSLE2 meta-analysis model to the PD methylation datasets. 26,244 significantly differential methylation sites, with 952, 776, 2466, 7604, 3897 and 2052 significantly differentially methylated sites were located on 1stExon, 3'UTR, 5'UTR, Body, TSS1500, and TSS200 of 703, 700, 1240, 3059, 2495 and 1267 genes, respectively. Research shows that PD is closely related to BCOR, HDAC4, PRDM16, and other genes with significantly differential methylation sites.

This work presents a new between-study variance estimator for meta-analysis model. First, the primary purpose



**Fig. 13** Histogram of significantly differential methylation sites on genes. The horizontal axis is the number of significantly differential methylation sites, and the vertical axis is the number of genes

of the between-study variance estimator is to measure the extent of variability in effect sizes across different studies that is attributable to true differences rather than within-study sampling error. Accurately estimating between-study variance allows us to distinguish between variability due to real differences in study effects and variability due to random error, providing a clearer understanding of the consistency of the effect across studies. Second, meta-analyses synthesize data from multiple studies to derive conclusions with greater statistical power. By accurately accounting for between-study variance, the proposed model enhances the validity of meta-analytic results, leading to more trustworthy conclusions. And it can improve the quality of evidence synthesis, promoting better decision-making in clinical practice, policy, and further research. Third, understanding heterogeneity is key to identifying which treatments work best for specific populations in medical research. A better estimation of between-study variance can help in understanding consistent treatment effects, aiding in the development of personalized therapeutic strategies. Moreover, efficiently allocating research resources requires an

understanding of where variability lies. By identifying true sources of heterogeneity, researchers and funding bodies can focus efforts on areas with the most significant impact, optimizing the use of limited resources.

In addition, after significant differential methylation sites identified by DSLE2 meta-analysis model, several software tools and platforms can be used for downstream analysis. These tools help in various aspects such as functional annotation, pathway enrichment analysis and so on. We can use Chip Analysis Methylation Pipeline (ChAMP) to annotate significantly differential methylated regions (DMRs), perform gene ontology (GO) analysis, and integrating with other epigenetic data. GREAT (Genomic Regions Enrichment of Annotations Tool) can be used to annotate and analyze the functional significance of sets of genomic regions, including DMRs, by associating them with nearby genes and functional terms. We can also use GSEA (Gene Set Enrichment Analysis) to do pathway enrichment analysis based on significant differential methylation sites identified by DSLE2.



**Abbreviations**

DSL E2	Two-step estimation starting with the DSL estimate and the $E_{m1}$ in the second step
RML	Restricted maximum likelihood estimate
DSL R2	Two-step estimation starting with the DSL estimate and the $R^2$ in the second step
DSL	DerSimonian and Laird estimate
SJ	Sidik and Jonkman estimate
GEO	Gene Expression Omnibus database
AUC	The area under the receiver operating characteristic curve
$M$	The methylated fluorescence signal intensity
$U$	The unmethylated fluorescence signal intensity
HM450K	Infinium Human Methylation 450 K BeadChip
TCGA	The Cancer Genome Atlas database
EB	Empirical Bayes estimate
HO	Hedges and Olkin estimate
PCMiss	Prediction-conditioned miss
WHO	World Health Organization
PD	Parkinson's disease
1stExon	The first exon of a gene
TSS1500	200–1,500 Bases upstream of the transcriptional start site
TSS200	0–200 Bases upstream of the transcriptional start site

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11316-3>.

Supplementary Material 1: Figure S1 to S16 are additional files which are plots of accuracy, false negative rates and other measures under the second and third hypothesis.

**Acknowledgements**

Not applicable.

**Authors' contributions**

S.J., N.W. and Y.Z. designed this project. N.W. conceived the idea, developed the model, and performed the analysis. Y.Z. and S.J. wrote the manuscript. F.Z. performed data analysis and revision of this paper. All authors read and approved the final manuscript.

**Funding**

This work was supported by the National Natural Science Foundation of China (Grant number: 62271173, 11971130, and 62172122), the Interdisciplinary Research Foundation of HIT (Grant number: IR2021109).

**Data availability**

All data used in this paper are available from the GEO database (<http://www.ncbi.nlm.nih.gov/geo>) and TCGA database (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>). The accession numbers of lung cancer methylation data used in this paper are GSE63704, GSE83842, and GSE85845. The EPIC and 450 K methylation data of lung cancer in TCGA database are collected from [https://portal.gdc.cancer.gov/analysis\\_page?app=Downloads](https://portal.gdc.cancer.gov/analysis_page?app=Downloads). The accession numbers of PD methylation data in the GEO database are GSE72774, GSE72776, GSE111629, GSE145361 and GSE165081.

**Declarations****Ethics approval and consent to participate**

Because the data in this meta-analysis are from published articles and the data are publicly available, it was not necessary to obtain ethics approval and consent to participate.

**Consent to publication**

Not applicable.

**Competing interests**

The authors declare no competing interests.

Received: 20 October 2023 Accepted: 3 February 2025

Published online: 05 March 2025

**References**

- Wang X-M, Zhang X-R, Li Z-H, Zhong W-F, Yang P, Mao C. A brief introduction of meta-analyses in clinical practice and research. *Gene Med.* 2021;23(5):3312–20.
- Airy GB. *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations.* London: Forgotten Books; 2012.
- Shannon H. A statistical note on karl pearson's 1904 meta-analysis. *J R Soc Med.* 2016;109(8):310–1.
- Hoshdel AK, Attia J, Carney SL. Basic concepts in meta-analysis: A primer for clinicians. *Clinical Practice.* 2006;60(10):1287–94.
- Zhou S, Shen C. Avoiding definitive conclusions in meta-analysis of heterogeneous studies with small sample sizes. *JAMA Otolaryngol Head Neck Surg.* 2022;148(11):1003–4.
- Hill HW. The origin and prevalence of typhoid fever in the district of Columbia. *Am J Public Hygiene.* 1910;20(2):430–3.
- Schmitz S, Lowenstein EJ. The unwavering doctor who unraveled a medical mystery. *Int J Womens Dermatol.* 2019;5(2):137–9.
- Tielbeek JJ, Uffelmann E, Williams BS, Colodro-Conde L. Uncovering the genetic architecture of broad antisocial behavior through a genome-wide association study meta-analysis. *Mol Psychiatry.* 2022;27(11):4453–63.
- Kienle GS, Kiene H. The powerful placebo effect: Fact or fiction. *J Clin Epidemiol.* 1997;50(12):1311–8.
- Glass GV. Primary, secondary, and meta-analysis of research. *Am Educ Res Assoc.* 1976;5(10):3–8.
- Paule RC, Mandel J. Consensus values and weighting factors. *J Res Natl Bur Stand.* 1982;87(5):377–85.
- Hedges LV. A random effects model for effect sizes. *Psychol Bull.* 1983;93(2):388–95.
- Malzahn U, Bohning D, Holling H. Nonparametric estimation of heterogeneity variance for the standardised difference used in meta-analysis. *Biometrika.* 2000;87(3):619–32.
- Dettori JR, Norvell DC, Chapman JR. Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. *Global Spine J.* 2022;12(7):1624–6.
- Scheidt S, Vavken P, Jacobs C, Koob S, Cucchi D. Systematic reviews and meta-analyses. *Zeitschrift fur Orthopadie und Unfallchirurgie.* 2019;157(4):392–9.
- Kanters S. Fixed and random effects models. *Methods Mol Biol.* 2022;2345:41–65.
- Sung H, Ferlay J, Siegel RL, Laversanne M. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–49.
- Wang A, Wan P, Hebert JR, Marchand LL. Atopic allergic conditions and prostate cancer risk and survival in the multiethnic cohort study. *Br J Cancer.* 2023;129(6):974–81.
- Li M, Lin A, Luo P, Shen W, Xiao D, Gou L. Dnah10 mutation correlates with cisplatin sensitivity and tumor mutation burden in small-cell lung cancer. *Aging.* 2020;12(2):1285–303.
- Wang M, Luan S, Fan X, Wang J, Huang J, Gao X, Han D. The emerging multifaceted role of pink1 in cancer biology. *Cancer Sci.* 2022;113(12):4037–47.
- Qi L-F-R, Liu Y, Liu S, Xiang L, Liu Z, Liu Q, Zhao J-Q, Xu X. Phyllyrin promotes autophagosome formation in a53t-asyn-induced parkinson's disease model via modulation of reep1. *Phytomedicine.* 2024;134:155952.
- GBD 2016 Parkinson's Disease Collaborators. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet Neurol.* 2018;17(11):939–53.
- Organization W.H. Launch of who's parkinson disease technical brief. World Health Organization; 2022.
- Yemni EA, Monies D, Alkhairallah T, Bohlega S, Abouelhoda M, Magrashi A, Mustafa A, AlAbdulaziz B, Alhamed M, Baz B, Goljan E, Albar R, Jabaan A, Faquih T. Integrated analysis of whole exome sequencing and copy number evaluation in parkinson's disease. *Sci Rep.* 2019;9(1):3344.

25. Li Y, Gu Z, Lin S, Chen L, Dzreyan V, Eid M, Demyanenko S, He B. Histone deacetylases as epigenetic targets for treating parkinson's disease. *Brain Sci.* 2022;12(5):672.
26. Guo Y, Ma J, Huang H, Xu J, Jiang C, Ye K, Chang N, Ge Q, Wang G, Zhao X. Defining specific cell states of mptp-induced parkinson's disease by single-nucleus rna sequencing. *Int J Mol Sci.* 2022;23(18):10774.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.