Computational and Structural Biotechnology Journal 20 (2022) 3783–3795

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

MSPJ: Discovering potential biomarkers in small gene expression datasets *via* ensemble learning

HuaChun Yin [a,b,c,1], JingXin Tao [b,1], Yuyang Peng [a,1], Ying Xiong [c], Bo Li [b,*], Song Li [a,d,*], Hui Yang [a,d,*]

[a] Department of Neurosurgery, Xinqiao Hospital, The Army Medical University, Chongqing 400037, China
[b] College of Life Sciences, Chongqing Normal University, Chongqing 401331, China
[c] Department of Neurobiology, Chongqing Key Laboratory of Neurobiology, The Army Medical University, Chongqing 400038, China
[d] Guangyang Bay Laboratory, Chongqing Institute for Brain and Intelligence, Chongqing, China

ARTICLE INFO

Article history:
Received 14 March 2022
Received in revised form 10 July 2022
Accepted 11 July 2022
Available online 14 July 2022

Keywords:
Small sample size
Random sampling
Feature selection
Differentially expressed genes
Machine learning

ABSTRACT

In transcriptomics, differentially expressed genes (DEGs) provide fine-grained phenotypic resolution for comparisons between groups and insights into molecular mechanisms underlying the pathogenesis of complex diseases or phenotypes. The robust detection of DEGs from large datasets is well-established. However, owing to various limitations (e.g., the low availability of samples for some diseases or limited research funding), small sample size is frequently used in experiments. Therefore, methods to screen reliable and stable features are urgently needed for analyses with limited sample size. In this study, MSPJ, a new machine learning approach for identifying DEGs was proposed to mitigate the reduced power and improve the stability of DEG identification in small gene expression datasets. This ensemble learning-based method consists of three algorithms: an improved multiple random sampling with *meta*-analysis, SVM-RFE (support vector machines-recursive feature elimination), and permutation test. MSPJ was compared with ten classical methods by 94 simulated datasets and large-scale benchmarking with 165 real datasets. The results showed that, among these methods MSPJ had the best performance in most small gene expression datasets, especially those with sample size below 30. In summary, the MSPJ method enables effective feature selection for robust DEG identification in small transcriptome datasets and is expected to expand research on the molecular mechanisms underlying complex diseases or phenotypes.

1. Introduction

Transcriptomics technology has become ubiquitous in systems biology, and identifying the differentially expressed genes (DEGs) is a critical step in analyses of these high-throughput data. Analyses of DEGs provide key insights into the mechanisms underlying diseases and a basis for the discovery of diagnostic biomarkers [1–3]. Meanwhile, the DEG analysis was applied to several various fields, such as ecology, evolution and population genetics [4–6]. However, identifying DEGs from high-dimensional datasets is a challenging task. In the context of transcriptomic data analyses, many candidate biomarkers are unstable and yield false positive results in clinical applications [7]. The instability of derived biomarkers has many explanations, of which a core reason may be that the sample size of datasets is insufficient [8,9]. Large-scale datasets are considered to have significant statistical power for DEG identification [10,11], whereas small sample size poses significant challenges for datasets analysis, including the "curse of dimensionality" and the overfitting of training datasets [12]. Unfortunately, large sample size is not typically feasible in broad laboratory biological experiments owing to the scarcity of specimens or the prohibitive costs of datasets preparation.

To mitigate the reduced power and improve the stability of biomarkers in gene expression datasets with small sample sizes, many methods for DEG identification have been proposed (such as, based on noise distribution [13,14] and optimized machine learning algorithms and prediction models [15,16]). Apart from

Abbreviations: DEGs, differentially expressed genes; SVM-RFE, support vector machines-recursive feature elimination; MSPJ, the Joint method of Meta-analysis, SVM-RFE, and Permutation test; SMDs, standardized mean differences; GEO, Gene Expression Omnibus; SAM, significance analysis of microarrays; SNR, signal noise ratio; mRMR, minimum-redundancy-maximum-relevance; GA, genetic algorithm; RF, random forest; GO, gene ontology; ROC, receiver operating characteristic; AUC, area under the ROC curve (AUC); FDR, false positive rate.

* Corresponding authors at: College of Life Sciences, Chongqing Normal University, Chongqing 401331, China (B. Li); Guangyang Bay Laboratory, Chongqing Institute for Brain and Intelligence, Chongqing, China (S. Li and H. Yang).

E-mail addresses: libcell@cqnu.edu.cn (B. Li), dlisong3@163.com (S. Li), huiyangxinqiao@163.com (H. Yang).

[1] These authors contribute equally to this work.

https://doi.org/10.1016/j.csbj.2022.07.022
2001-0370/

strategies to improve statistical robustness in studies with small sample sizes, multiple random sampling is recognized as a good strategy, in which samples are randomly extracted with replacement from the original dataset to construct independent sub-datasets [17]. The use of multiple sub-datasets can effectively adjust the skewed distributions of errors caused by a small sample size. This method of combining multiple datasets can simulate large-scale datasets and provide sufficient statistical power [18]. To improve the consistency of biomarkers derived from multiple datasets, several popular feature selection methods (such as *meta*-analysis, permutation test and SVM-RFE) were usually used in biomedicine filed [19–21]. Meta-analysis is a statistical technique for aggregating the results of various independent but related studies to identify DEGs [19,22–24]. SVM-RFE is a method for feature selection by iteratively training an SVM classifier with the current set of features and removing the least important feature indicated by the SVM [25]. In cases with very small sample sizes, multiple sub-datasets are generated by random sampling, and the permutation-derived test statistics for each gene in sub-datasets are combined to determine DEGs [26,27].

In this study, an improved ensemble learning-based method, MSPJ (the Joint method of Meta-analysis, SVM-RFE, and Permutation test) was proposed to discover DEGs adapting to transcriptome dataset with small sample size, so that improve the reproduction of DEGs by integrating results from multiple sub-datasets. With the aim of exploring stable molecular signatures by utilizing small samples to reconstruct large datasets and sub-datasets. MSPJ is expected to provide opportunities to identify reliable, stable biomarkers, irrespective of sample size, in a cost-effective manner.

2. Materials and methods

2.1. Multiple random sampling procedure

Resampling, a simple but powerful statistical technique, is used to obtain multiple subsets of data by random sampling with replacement. In this study, a multiple random sampling strategy was used to draw a sample from the original dataset without replacement to create subsets. Each subset had independent samples, and sizes were not identical. The random sampling method was as follows:

Original dataset:$\{x_{E1}, x_{E2}, x_{E3}, \cdots, x_{EM}, x_{C1}, x_{C2}, x_{C3}, \cdots, x_{CN}\}$;
Sample size: $S = M + N$
Take e from $\{x_{E1}, x_{E2}, x_{E3}, \cdots, x_E\}$ and c from $\{x_{C1}, x_{C2}, x_{C3}, \cdots, x_C\}$ where $M > e > 3$ & $N > c > 3$
Sampling K times,
Subset:$\{y_{e1}, y_{e2}, y_{e3}, \cdots, y_{em}, y_{c1}, y_{c2}, y_{c3}, \cdots, y_{en}\}$;
Sample size:.$s = m + n$

Where x_E and x_C represent the E-th sample of experimental group and C-th sample of the control group in the original dataset individually. S stands for the original sample size, while M and N represent original sample size of the experimental and control group, respectively. Similarly, y_e and y_c represent the e-th sample of the experimental group and c-th sample of the control group in the sub-datasets. K denotes the number of random sampling sets. s represents the subset sample size, m represents the experimental group of subset sample size, n represents the control group of subset sample size.

2.2. Application of the MSPJ approach to detect differentially expressed genes

To identify stable candidate biomarkers, the MSPJ method was built by integrating three algorithms (*meta*-analysis, SVM-RFE, and permutation test) for DEG detection. Using this joint method, shared DEGs estimated by the three algorithms were defined as robust biomarkers. An overview of the whole process of MSPJ is provided in Fig. 1. MSPJ procedure contains two steps. Firstly, three sets of DEGs were identified by three methods (*meta*-analysis based on multiple subsets, SVM-RFE based on nested 5-fold cross-validation, and permutation test in original dataset) individually. Secondly, the overlap of DEGs from three methods was obtained, and the share DEGs in overlapping area were considered as the robust potential biomarkers.

2.2.1. Meta-analysis for gene selection

The *meta*-analysis strategy was used in the MSPJ method for the initial screening of DEGs by combining the results from multiple sub-datasets [28]. For continuous outcomes according to the following formula:

1) The standard deviation was computed for gene i in every sub-datasets:

$$SD_k = \sqrt{\frac{(n_e - 1) \times sd_e^2 + (n_c - 1) \times sd_c^2}{(n_e + n_c - 2)}},\ \mathrm{k} = 1, 2, 3, \cdots, \mathrm{N};$$

2) The standardized mean differences (SMDs) were calculated in experimental and control groups of sub-datasets: $SMD_{\mathrm{k}} = \frac{\bar{ME}_{\mathrm{k}} - \bar{MC}_{\mathrm{k}}}{SD_{\mathrm{k}}}$;

3) The weight vectors $\omega_{\mathrm{k}} = \frac{n_e + n_c}{\mathrm{N}_e + \mathrm{N}_c}$;

4) The combining effect sizes were calculated: $SMD_i = \sum_{k=1}^{N} SMD_k \times \omega_k$

Where n_e and n_c are number of observations in experimental and control groups, and the sd_e and sd_c represent the standard deviation in experimental and control groups. The k is the N-th sub-datasets, N is the number of sub-datasets, ω is the weight vectors. And $\bar{ME}_i$ and $\bar{MC}_i$ denote the mean values in the two groups. A gene was considered up-regulated when SMD > 0.5 with 95% confidence interval did not cross the null hypothesis. In contrast, a gene was defined as a down-regulated gene if SMD < 0.5 with 95% confidence intervals did not exhibit an invalid line cross.

2.2.2. SVM-RFE algorithm for gene ranking

As one of the most widely used backward elimination algorithms, SVM-RFE works by iteratively removing the "worst" gene until the predefined size of the final gene subset is reached [29]. The feature ranking score as the ranking criterion was calculated by the coefficients of the weight vector in each dataset according to the following computing steps:

Inputs:
Training dataset $\{X_i, Y_i\}_{i=1}^{N}$
Output:
Feature ranked list R.
Initialize:
Subset of surviving features,$S = [1, 2, \cdots n]$
Feature ranked list R, $R = []$
While S is not empty, do:
1). Restrict the features of X_i to the remaining S
2). Train SVM to get weight vectors $\omega = \sum_i \alpha_i Y_i X_i$, i=1,2,3, . . .,N
3). the ranking score of i-th feature is defined as
$\mathbb{C}_i = (\omega_i)^2$, i = 1,2,3, . . .,N
4). Find the feature with the smallest ranking criteria
$f = argmin(\mathbb{C}_i)$
5). Add feature index f to R, $R = \{f\} \cup R$
6). Eliminate the feature index f from S, $S = S$

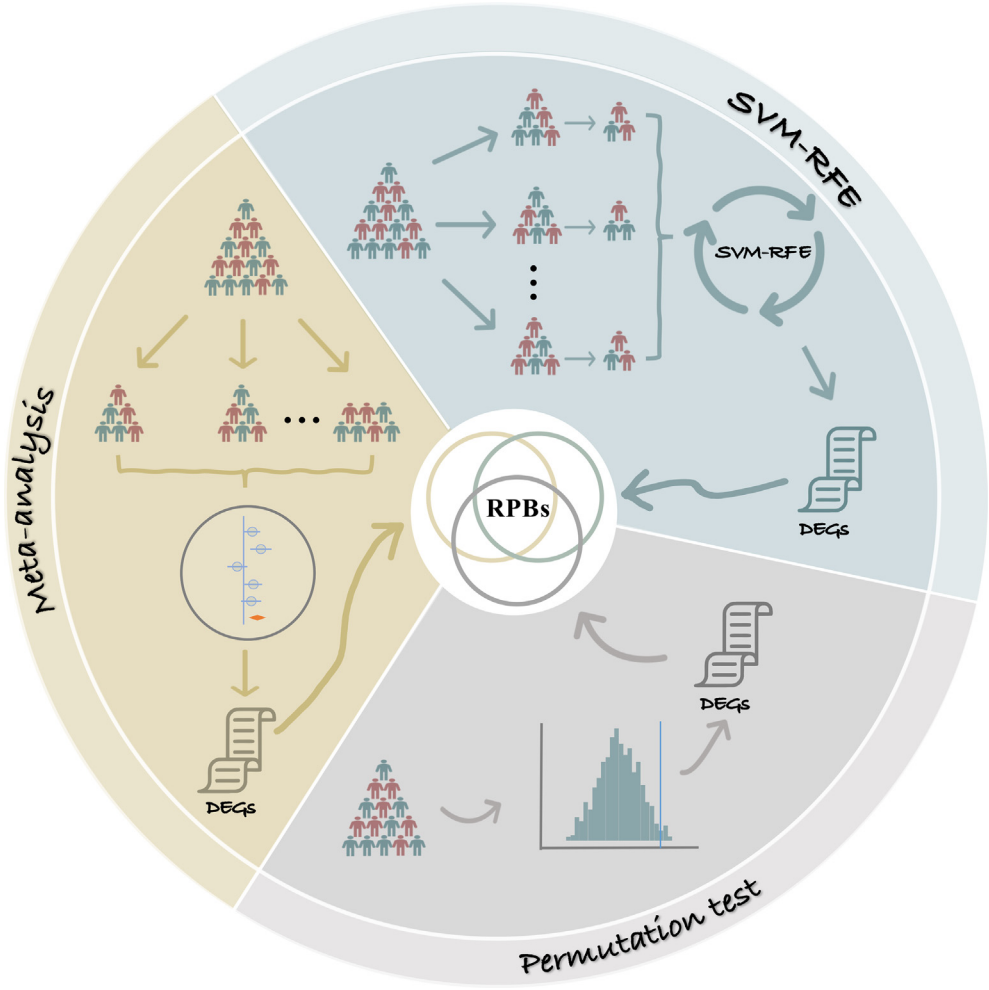

Fig. 1. A working mechanism of the MSPJ. The RPBs represents robust potential biomarkers.

Where N is the number of training samples, X_i is the i-th training sample, Y_i encodes the class label of X_i, S is the surviving feature size, n is the n-th feature index, α is the SVM classifier, ω is the weight vectors, C is the ranking criteria and f is the smallest ranking criterion.

2.2.3. Permutation test for the detection of DEGs

The permutation test is increasingly used to construct sampling distributions without replacement, especially for data sets with small sample sizes and ambiguous data distributions. Generally, DEGs were identified by a permutation test by the following basic steps:

The experimental group: A $\{x_{E1}, x_{E2}, x_{E3}, \cdots, x_{EM}\}$;
The control group: B $\{x_{C1}, x_{C2}, x_{C3}, \cdots, x_{CN}\}$;

(1) compute the difference in mean expression values between experimental and control groups in the original dataset:$T_0 = \bar{A} - \bar{B}$
(2) randomly sampling the mixed data [A, B] into subset A' and B'.
(3) compute the difference in mean expression values between experimental and control groups in the permuted dataset:$T = \bar{A}' - \bar{B}'$
(4) repeat steps (2) and (3) K times and evaluate all permutations, and $K \leq (\mathrm{M} + \mathrm{N})!$
(5) return the p-value as the p-value of DEGs when T exceeds T_0:

$$P = \frac{1}{K}\sum_{i=1}^{K}(T > T_0)$$

here, T_0 is the statistic for the observation dataset, and T is the statistic for the permutation subset. $\bar{A}$ and $\bar{B}$ are the mean expression values for the experimental and control groups. $\bar{A}'$ and $\bar{B}'$ are the mean expression values in experimental and control groups in the permutation dataset. The M represents the sample sizes of A, and N represents the sample sizes of B. The K represents the permutation rounds.

2.3. Datasets used to assess the performance of MSPJ

To determine the validity of the MSPJ method for DGE detection, 259 gene expression datasets (containing 94 simulated and 165 real datasets) were used with different sizes generated by DNA microarray or RNA-seq platforms.

2.3.1. Simulated datasets

To test the stability of feature selection methods, several simulated gene expression datasets (including microarray and RNA-seq data) were utilized. Microarray datasets ($\log_2$-ratios) were simu-

lated using the madsim package in R [30] and RNA-seq datasets were generated using the SPsimSeq package [31].

2.3.2. Real datasets

A number of real datasets with various sample sizes were applied to further compare MSPJ with established methods. These gene expression datasets were mainly collected from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). The detailed information for all gene expression datasets is provided in Table A.1.

2.3.3. Data preprocessing

Before DEG detection, raw *.cel files for the microarray datasets were read for RMA normalization and the generation of gene expression data matrices [32]. Then, the genes with low expression values in all samples (threshold < 0.1) were removed in subsequent analyses of each dataset [33]. To reduce sample error resulting from random sampling, the original datasets and subsets were further normalized by the upper quantile method using the preprocessCore package [34]. All computations were implemented in the R environment (v4.0.0).

2.4. Comparison with classical methods for gene selection

Ten methods were used for a comparative analysis with the newly developed MSPJ method, including five DEG identification methods (limma [35], significance analysis of microarrays (SAM) [36], T-test, Wilcoxon's test, multtest [37]) and five gene rank-based methods (RankProd [38], signal noise ratio (SNR), minimum-redundancy-maximum-relevance (mRMR) [39], genetic algorithm (GA) [40], random forest (RF) [41]).

Those ten methods have been widely used, and can be categorized into three main types: filter, wrapper, and embedded approach. Among of them, Limma, SAM, T-test, Wilcoxon's test, multtest, mRMR, RankProd and SNR were the well-known filter approach algorithms [42]. GA was applied as a wrapper feature selection method, and RF was known as an embedded method [43,44]. The gene expression matrices were used to identify DEGs by each of these 11 methods independently. The DEG identification methods provided directly a subset of genes, and the feature ranking methods provided the ranking. Genes were considered as DEGs when the adjusted p-value was below 0.05 or the arbitrary rank was in the top 30%. A gene set enrichment analysis of gene ontology (GO) terms was implemented using clusterProfiler with an adjusted p-value cutoff of 0.05 [45]. The Jaccard index was used to evaluated consensus DEGs and the related GO terms identified by different methods.

The Jaccard index was defined as follows: $J(A,B) = |\frac{A \cap B}{A \cup B}|$, where A and B represent the DEGs or related GO terms identified by two methods, respectively.

2.5. Application of a classification algorithm for method assessment

Prediction modeling plays an important role in algorithm assessment by classifying experimental and control cases. In this study, the SVM algorithm was implemented using the R package e1071, and selected to assess the stability of biomarkers from MSPJ and conventional methods [46]. The top ten biomarkers were selected to construct the SVM classifier. The nested 5-fold cross-validation strategy was applied to partitioning the training-test datasets and calculating the evaluating indexes. The classification accuracy was used to assess the classification performance of gene features from training datasets. The area under the receiver operating characteristic (ROC) curve (AUC), specificity, sensitivity, and accuracy were used to evaluate the performance of the MSPJ and conditional methods using the pROC package [47].

3. Results

3.1. The type i error control in simulated datasets

The type I error rate was calculated as the proportion of simulations where non-DEGs tested positively, *i.e.*, the false positive rate (FDR). The sample size of all simulated datasets was limited to a maximum of 30 samples, and all non-DEGs were false positive. To better compare feature selection methods based on hypothesis testing, the DEGs were identified using a nominal p-value cutoff of 0.05. As for gene rank-based methods, the DEGs were identified by an arbitrary rank (the preset DEGs number). In most cases, keeping the FDR < 10% is a conservative approach for a well-calibrated test. Under baseline conditions for 20 microarray simulation datasets, only three methods (MSPJ, SAM, and multtest) performed quite well and even in maintained FDR < 5% (Fig. 2A). The RF algorithm performed reasonably well but with the FDR of > 10%. However, limma, T-test, RankProd, and Wilcoxon's test called around half non-DEGs. The GA, mRMR and SNR methods had a totally different gene list in simulation datasets.

As for the simulated RNA-seq datasets, all methods maintained the FDR of > 5% (Fig. 2B). For each of SNR, mRMR, MSPJ, SAM, and Wilcoxon's test, the median FDR value is simultaneously less than < 30% in most datasets. Of which, SAM has a high upper-quantile, and limma, multtest, T-test, and RankProd showed an FDR of approximately 50%. The FDR values of GA and RF were much higher than those of other methods. It was noted that MSPJ method applied to small datasets showed better performance with respect to type I error control than those of 11 classical methods both in microarray and RNA-seq datasets.

3.2. The time consuming and memory usage in simulation datasets

The complexity algorithms may result in high computational cost during the massive iterations of feature identification, especially as the search strategies drift towards more exhaustive [48]. Hence, the time consuming and memory usage of 11 methods were evaluated using the simulation datasets. The detailed score of time consuming and memory usage were normalized across the different methods. For each method, the mean value of time consuming or memory usage obtained from three repeated experiments. The time and memory scores were shifted and scaled to $\sigma = 1$ and $\mu = 0$, and then applied the unit probability density function of a normal distribution on these values to get the scores = 1-scores and back into [0,1] range [49].

In the Fig. 2C, all methods have consistency in the time consuming and memory usage both in microarray and RNA-seq technologies. Most of filter methods spend less memory and time than the wrapper and embedded methods in different samples and gene size. In terms of time consuming, limma, SNR, SAM and T-test were rarely influenced by sample and gene size. The time consuming of Wilcoxon's test, mRMR, multtest and RF increased with sample and gene size. The RankProd, MSPJ and GA spend more time consuming on both microarray and RNA-seq technologies. As for the memory usage, multtest, limma, mRMR, SNR, Wilcoxon's test and T-test achieved low memory usage. The memory usage of wrapper method (GA and MSPJ) was less than filter methods (SAM and RankProrid) and embedded method RF.

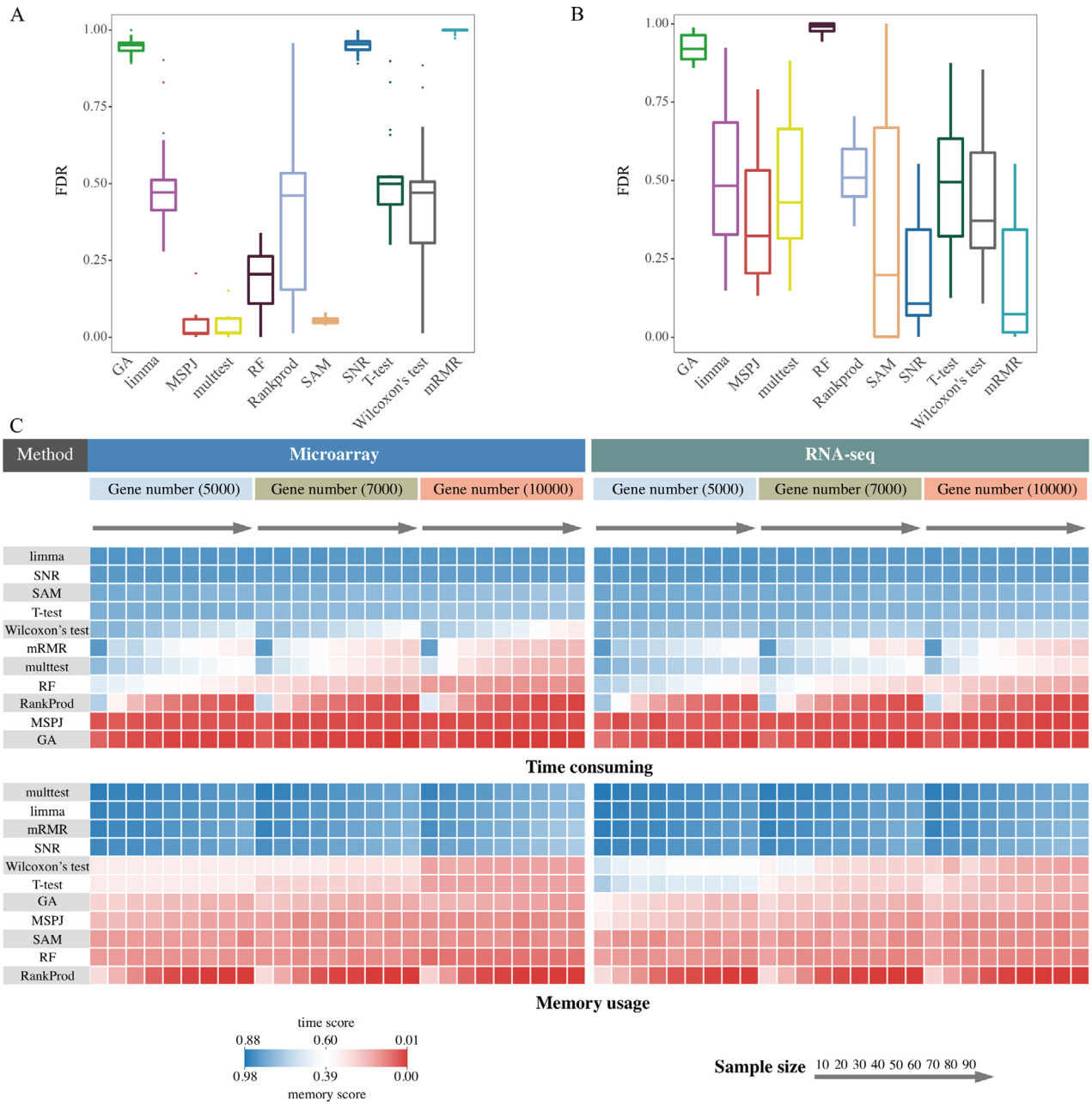

Fig. 2. (A) Type I error control for 11 methods applied to small simulated microarray datasets. (B) Type I error control for 11 methods applied to small simulated RNA-seq datasets. For each method, the box plot represents the values obtained from 20 experiments. All samples in simulated datasets were<30 and contained 6000–20000 genes. (C) The rank of time consuming and memory usage for 11 methods applied to simulated microarray and RNA-seq datasets.

3.3. Comparative analysis of DEG detection from real datasets

To compare different methods for identifying DEGs, three datasets with known properties were utilized (Table 1). The Benjamini and Hochberg procedure was applied to statistical tests with no adjustment, like the MSPJ, T-test, and Wilcoxon's test. As an adjusted p-value of < 0.01 was the criterion for DEG detection methods. The top 30% of genes were considered statistically significant for the gene rank-based methods. The identification of unique and common DEGs, gene set enrichment analyses, and AUC analyses for different sample sizes are important analysis types for a comprehensive evaluation.

Large-scale datasets (163 datasets with different sample sizes) were utilized to assess the robustness of gene selection methods. Detailed sample information was provided in Table A.1.

3.3.1. Assessment of the application of eleven methods to small datasets

The 11 individual gene selection methods were applied to two real datasets to identify DEGs. The GA, RF, and mRMR methods yielded the most unique DEGs; however, the intersection of DEGs obtained by other methods in microarray data was small (as shown in Fig. 3A). The MSPJ method generated the highest number of genes compared with the DEGs identified methods. The MSPJ, limma, SAM, RankProd, T-test, and Wilcoxon's test generated more shared DEGs. For RNA-seq datasets, the number of DEGs identified by different methods were similar to those of the microarray datasets (Fig. 3B). The GA and RF methods detected more total and unique DEGs and generated similar results. Apart from DEG counts, the Jaccard scores were used to evaluate DEGs identified using different methods. Jaccard scores closing to 0.5 (between 0.4 and 0.6)

Table 1
Summary of real datasets.

Technique	Accession number	Sample size	Gene numer	Size per class	Organism	Ref.
Microarray	GSE16515	20	12,937	10:10	Human	[50]
RNA-seq	PMID: 21179090	18	9,300	12:6	Fly	[51]
Microarray	GSE10072	107	12,937	58:49	Human	[52]

indicated that similar unique and common DEGs were detected by two methods. Jaccard scores for comparisons between MSPJ and limma, SAM, T-test, and Wilcoxon's test were all within 0.40–0.60 for the microarray datasets (Fig. 3C). The Jaccard scores for comparisons between MSPJ and limma, SAM, T-test, and Wilcoxon's test fell in 0.42 ∼ 0.55 for RNA-seq datasets (Fig. 3D). Overall, the DEGs from MSPJ, limma, SAM, T-test and Wilcoxon's test methods were very similar; however, the Jaccard scores for comparisons between MSPJ and other four methods were closer to 0.5 than those of comparisons in other methods.

The Jaccard scores of GO terms were consistent with those of the DEGs enriched by different methods. MSPJ-derived DEGs generated more unique and common biological functions in comparisons with ones of SAM, T-test, and Wilcoxon's test (Jaccard score: 0.55–0.60) for microarray datasets, as shown in Fig. 3E. For RNA-seq datasets, there existed similar results according to biological functions (Jaccard scores of 0.55 ∼ 0.58) (Fig. 3F). The AUC values revealed that MSPJ, limma, RankProd, T-test and Wilcoxon's test outperformed other methods in DEG discovery for all datasets (Fig. 3G and H).

3.3.2. Good performance of MSPJ revealed by testing on large-scale datasets

We next compared various methods for analyses of large-scale datasets, 123 microarray datasets and 40 RNA-seq datasets were tested. After identifying DEGs using diverse methods, the Jaccard score was computed to evaluate consensus with respect to DEGs and GO enrichment. Overall, the similarity between the methods was not significant for DEG identification both in microarray and RNA-seq datasets.

In microarray datasets, mRMR, SNR, MSPJ, RF, RankProd and GA had higher Jaccard scores than those of other methods for all sample sizes (Fig. 4A). But around 10 to 40 samples, SNR, mRMR, MSPJ, and RF had higher Jaccard scores than those of the other methods. Each method retained its independence for DEG detection on different sample sizes, except RankProd had high Jaccard scores around 20 to 40 samples. For GO terms, most methods showed similar trends in Jaccard scores, while T-test, GA, and limma showed a decrease as the sample size increased (Fig. 4B). Of note, RF, SAM, RankProd, and MSPJ showed high Jaccard scores for sample sizes between 20 and 30. According to Jaccard scores for DEGs and functional enrichment, only RF and MSPJ stayed on top of the momentum under 30 sample sizes. Consistency among the detected DEGs does not fully reflect the advantages and disadvantages of different methods; accordingly, we evaluated the discriminant ability of the model based on the top 10 DEGs between experimental and control groups in the DNA microarray datasets. The top 10 DEGs from five DEG identification methods were ranked by the p-value. Several indices should be considered for the classification and prediction of biomarkers, such as AUC, specificity, sensitivity, and accuracy. For these four indices, MSPJ and SNR showed good performance with small sample sizes (Fig. 4C–F). Then, RankProd, RF, and GA showed similar trends with respect to sample size, behind MSPJ and SNR.

In RNA-seq datasets, SNR, limma, MSPJ, and T-test had higher Jaccard scores than those of the other methods for around 12 to 60 samples (Fig. 5A). For GO terms, MSPJ, and RankProd showed higher Jaccard scores than those of the other methods with samples size between 12 and 50 (Fig. 5B). According to Jaccard scores for DEGs and functional enrichment, only MSPJ kept the higher Jaccard for around 12 to 50 samples. Similarly, according to four indices (AUC, specificity, sensitivity, and accuracy), MSPJ, limma, T-test and SNR showed good performance with sample size between 12 and 70 (Fig. 5C–F).

3.3.3. Robustness of MSPJ for DEG identification in small datasets

To explore the stability of the performance of MSPJ for feature retrieval in small size datasets, the impact of sample size was investigated by computing overlap in DEGs and GO terms between subsets and the original dataset for 11 methods. Six sub-datasets with 10%, 20%, 30%, 40%, 50%, and 60% of samples from each of two groups were generated by random sampling, and the random sampling was repeated ten times. The average Jaccard index for overlap in the DEGs or GO terms between subsets and the original dataset was evaluated for 11 methods and various small sample sizes. Except for the case where Jaccard value = 1, a high Jaccard score (i. e., the positive value close to 1) indicates that the method is more robust regardless of changes in sample size. AUC, specificity, sensitivity, and accuracy for different sample sizes were also evaluated.

One example of DNA microarray dataset GSE10072 containing 107 samples, DEG identification methods were sensitive to changes in sample size (Fig. 6A). The number of DEGs detected by most methods increased with the sampling rate, other than GA, SNR, mRMR and RF. Among all methods, MSPJ showed the least variation in DEG numbers across sample rates. For most methods, except for multtest, SNR, GA, mRMR, and RF, the Jaccard scores of DEGs and GO terms between subsets and the original dataset increased as the sample size increased (Fig, 6B and C). In the assessment of the sensitivity for sample size, mRMR showed complete concordance with DEG detection (Jaccard scores = 1), shown in Fig. 6B-E. Therefore, mRMR will not be compared with other methods in terms of the consistency of DEGs and GO terms. MSPJ clearly showed the highest Jaccard score of DEGs for sub-datasets in the 10–20% range. Up to 30%, MSPJ was only behind the T-test. For GO terms, MSPJ also stably ranked high in the 10–30% range. We also considered the retention of the top 100, 200, and 300 DEGs by different methods with regard to sampling rate between the sub-dataset and original dataset. For both DEGs and GO enrichment terms, RankProd consistently showed the highest Jaccard score for all sampling rates and all top-ranked genes (Fig. 6D and E, Fig. A.1). MSPJ showed a moderate Jaccard score for DEG detection, and the Jaccard scores for MSPJ were among the top five for the evaluation of GO terms for all top ranked genes and 10–30% sampling rates.

Most methods consistently showed high AUC values and high specificity in the 10–60% range, except for mRMR, multtest, GA, SNR, and Wilcoxon's test (Fig. 6F and G). For sensitivity, GA, RankProd, mRMR, and SNR showed decrease as the sampling rate increased, while other methods maintained a high sensitivity across all sample rates (Fig. 6H). Moreover, the biomarkers identified by GA, mRMR, and SNR had a lower accuracy for the prediction of control and experimental groups than that of other methods, irrespective of the sampling rate (Fig. 6I). Overall, MSPJ, limma, RF, and T-test have superior performance based on four indices for the assessment of classification and prediction.

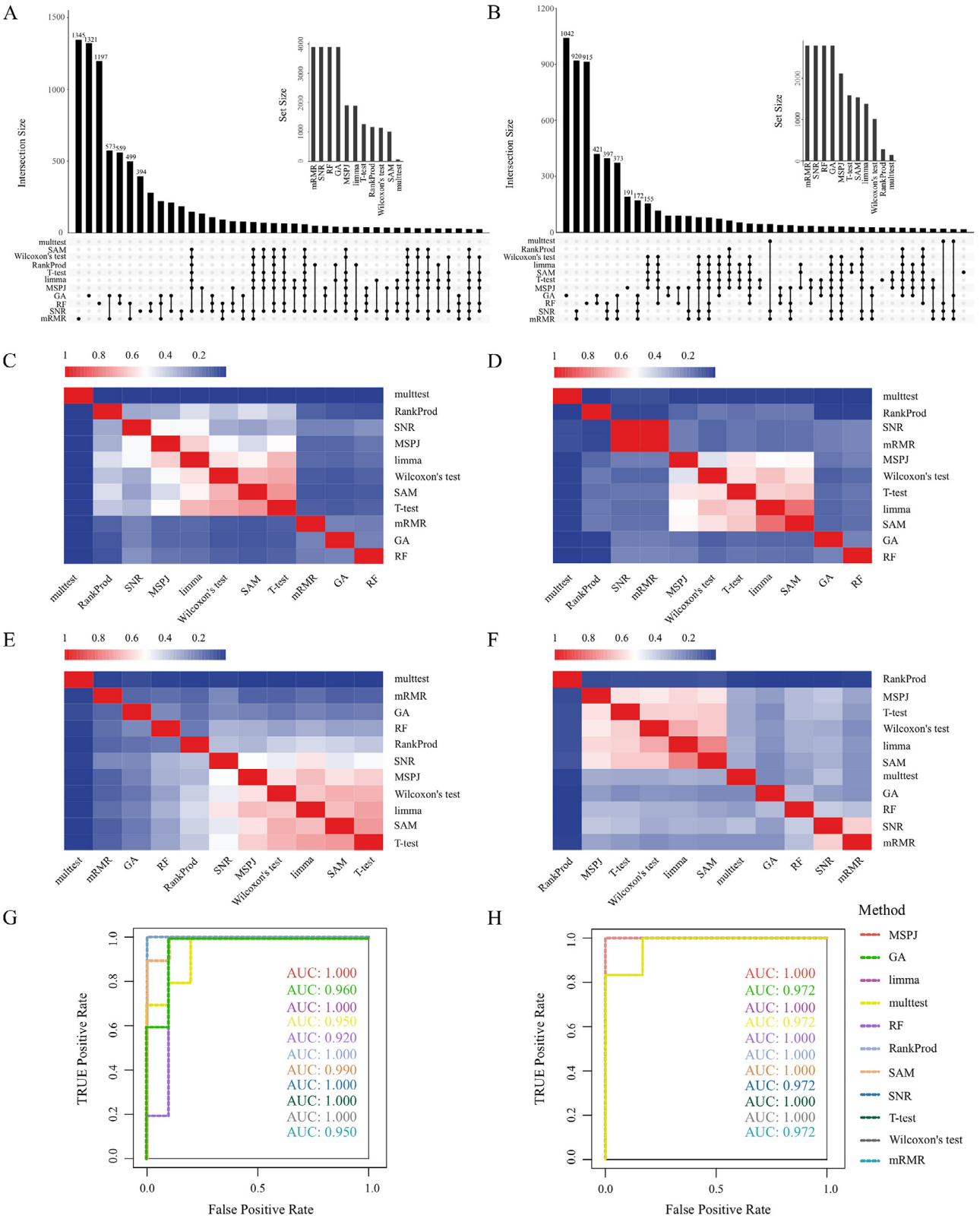

Fig. 3. Comparison of eleven methods with small sample sizes. (A) UpSet plot of DEGs obtained by 11 methods from a microarray dataset. (B) UpSet plot of DEGs from 11 methods from the RNA-seq dataset. (C) The Jaccard scores of DEGs from the microarray dataset. (D) The Jaccard scores of DGEs from the RNA-seq dataset. (E) The Jaccard scores of GO terms from the microarray dataset. (F) The Jaccard scores of GO terms from the RNA-seq dataset. (G) The AUC values of 11 methods for the microarray dataset. (H) The AUC values of 11 methods for the RNA-seq dataset.

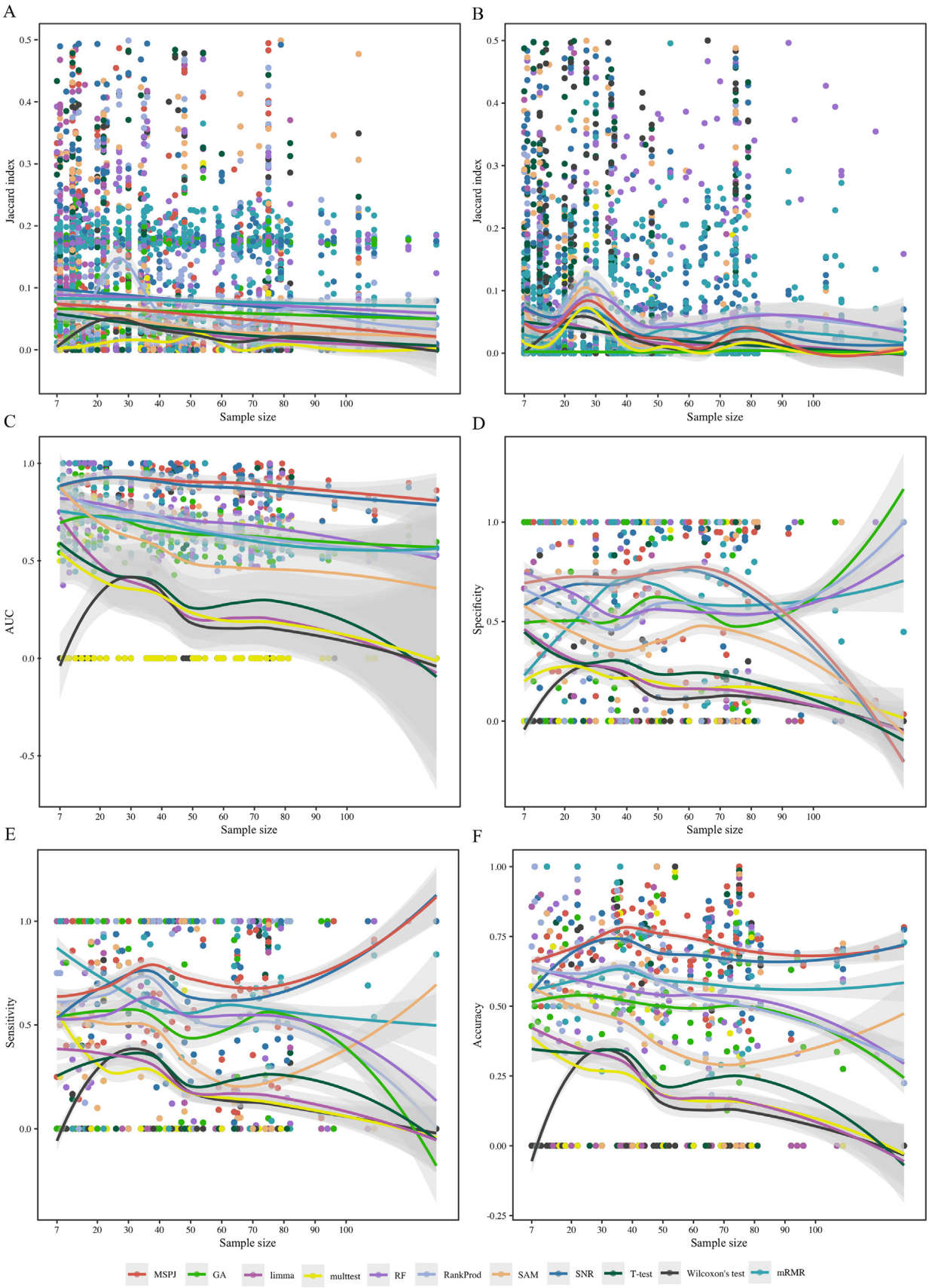

Fig. 4. Application of different methods to large-scale microarray datasets. The value of the smoothing parameter (loess) for a curve fitting was chosen for<1,000 observations, and the generalized additive model was used for>1,000 observations. (A) The similarity of DEG analysis methods for small datasets. (B) Similarity of enriched GO terms for DEGs identified using different methods in small datasets. When Jaccard score > 0.5, Jaccard index = 1 - Jaccard score, else Jaccard index = Jaccard score. Detailed information is provided in Table A.2. (C)∼(F) showed the AUC, specificity, sensitivity and accuracy of the top ten DEGs identified using different methods, respectively. The detailed values are reported in Table A.3.

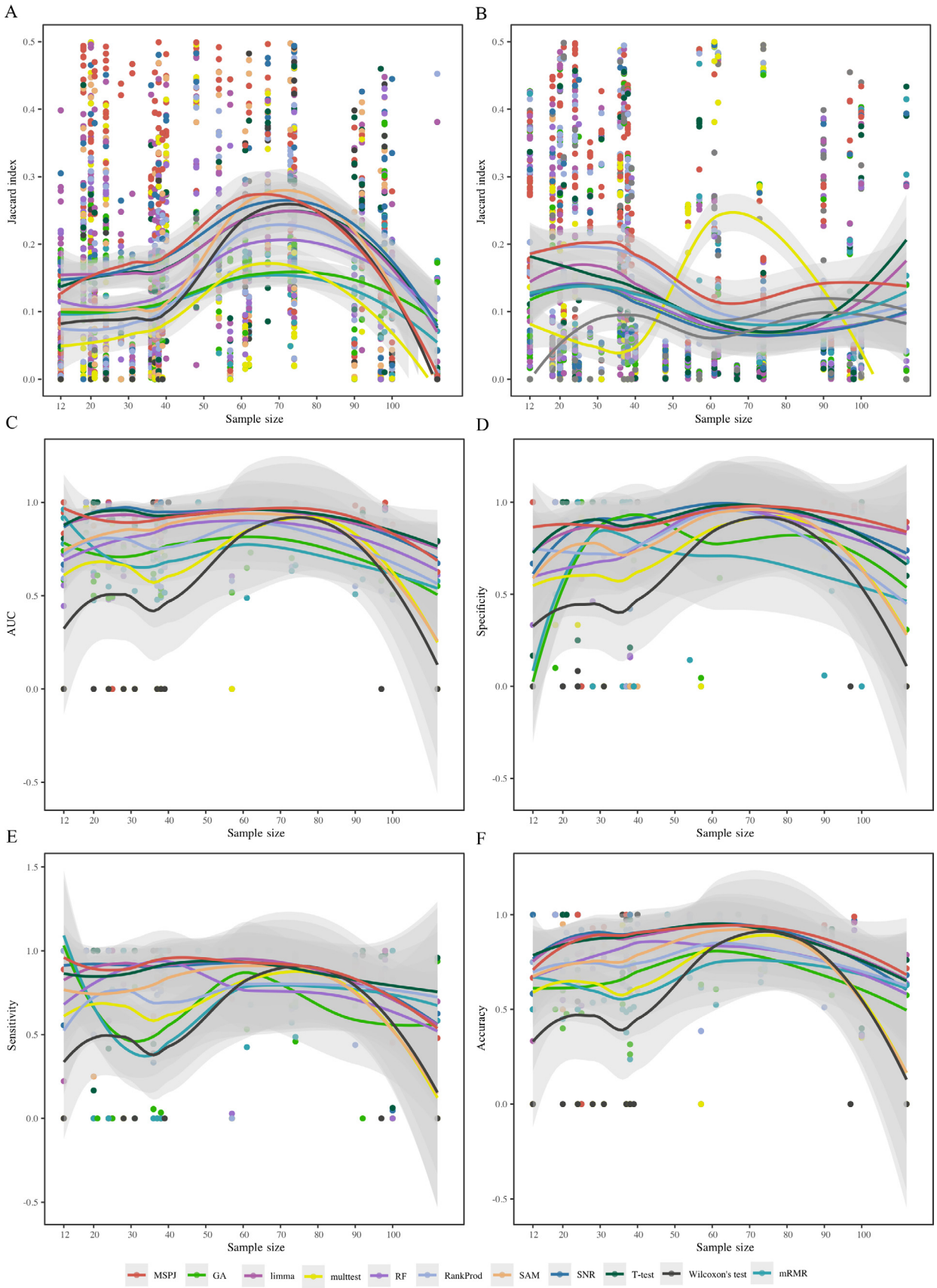

Fig. 5. Application of different methods to large-scale RNA-seq datasets. The value of the smoothing parameter (loess) for a curve fitting was chosen for<1,000 observations, and the generalized additive model was used for>1,000 observations. (A) Similarity of DEG analysis methods for small datasets. (B) Similarity of enriched GO terms for DEGs identified using different methods in small datasets. When Jaccard score > 0.5, Jaccard index = 1 - Jaccard score, else Jaccard index = Jaccard score. Detailed information is provided in Table A.2. (C)∼(F) showed the AUC, specificity, sensitivity and accuracy of the top ten DEGs identified using different methods, respectively. The detailed values are reported in Table A.3.

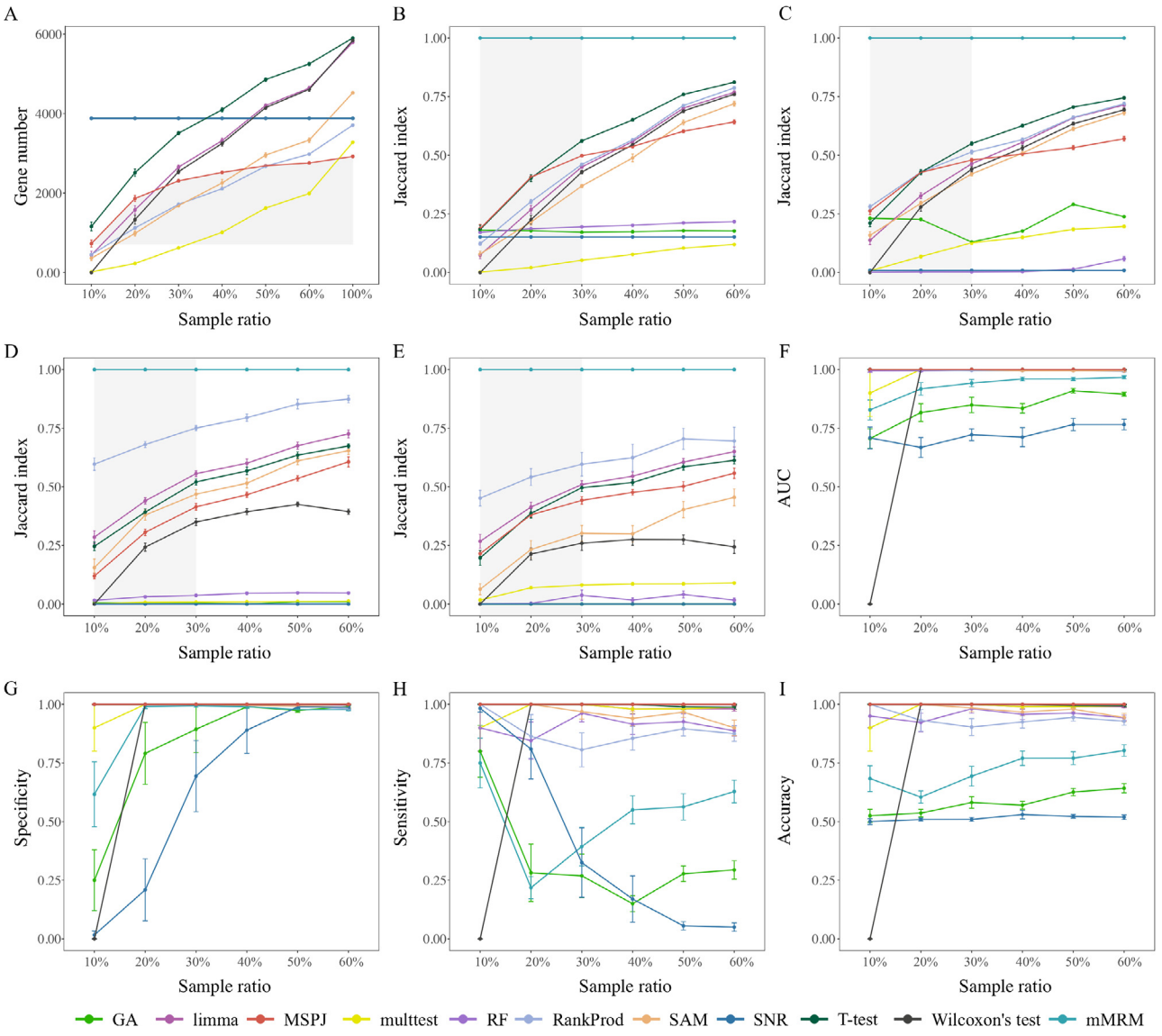

Fig. 6. Robust DEG detection for different sampling rates in large datasets. The entire process was repeated 10 times for random sampling, and each repetition employed a different random seed. (A) The number of discovered DEGs for different sampling rates. (B) The Jaccard scores of DEG counts for the comparison between the subset and original microarray dataset. (C) The Jaccard scores of GO terms for the comparison between the subset and microarray original dataset. (D) The Jaccard scores of the top 100 DEGs between the subset and microarray original dataset. (E) The Jaccard scores of GO terms enriched from the top 100 DEGs between the subset and microarray original dataset. (F)∼(I) The AUC, specificity, sensitivity and accuracy of sub-sampling microarray datasets, respectively.

4. Discussion

Owing to high experimental costs or low availability of samples, many gene expression datasets have small sample sizes. DEG analysis methods often show poor performance when the sample size is small [53]. The scDEA method proposed by Li *et al.* [54] was developed for differential expression analyses of scRNA-seq data. Further, the scDEA did not provide a significant advantage over other approaches when the sample size is small. To resolve these issues, an improved gene selection approach named MSPJ was developed in this study by integrating *meta*-analysis, SVM-RFE, and permutation test frameworks. To our knowledge, this is the first differential expression analysis method specifically targeting small gene expression datasets.

In this study, we compared various methods with default parameters, as implemented in widely used packages. Ten representative methods were used for a comparative analysis with MSPJ. Among these, limma is often used for gene discovery by differential expression analyses of microarray and high-throughput PCR data, and SAM is a frequently used nonparametric method for analyses of microarray dataset. The T-test, multtest, Wilcoxon's test, mRMR and RankProd (rank product method) methods are well-established statistical methodologies for feature selection [38]. The RankProd method is expected to be applicable to small datasets [55]. As a supervised machine learning approach, RF has gained substantial popularity for feature selection [56]. The GA approach, as an unsupervised search method, is often used to select a set of features to discriminate between groups, especially for classification in cases with small sample sizes [57]. Thus, RF and GA were used as representative supervised and unsupervised strategies for our comparative analysis. SNR is a signal–noise-ratio based feature selection method for ranking genes [58]. Our comparison of these distinct approaches for DEG identification provided a general and important reference for research in the field.

MSPJ was obviously superior to other methods with respect to type I error control using simulated microarray datasets. Using simulated RNA-seq datasets, MSPJ ranked highly in terms of type

I error control, only behind SNR and mRMR. The high noise level of RNA-seq datasets is an issue for the accurate detection of DEGs [59]. Hence, it is reasonable that the SNR method is superior to the MSPJ method for analyses of simulated RNA-seq datasets. Nevertheless, SNR and mRMR had a quite loose on type I error control in microarray datasets. Perez's research indicated that mRMR algorithm is not suitable for high domain feature problems [60], and this may be the reason why mRMR kept unstable in high microarray and RNA-seq dimensional data, in terms of type I error control. MSPJ had the robust type I error control both on microarray and RNA-seq datasets. In terms of time and memory consumption, MSPJ did not perform well. Because it was based on *meta*-analysis for 40 sub-datasets and classifier model development provided by machine learning, and the proposed method taken slightly high computational cost for feature selection. However, it was worth to consider exchanging more time consuming and memory usage for robust feature detection based on small samples.

Using real datasets for a large-scale microarray and RNA-seq datasets, SNR and MSPJ had an outperformance in terms of similarity of gene detection for <30 samples. In terms of gene enrichment of functional entries, RankProd and MSPJ had good performance under 30 samples. However, MSPJ and SNR were comparable and superior to the other methods in terms of feature gene classification and prediction both in microarray and RNA-seq datasets. In brief, the overall results for 165 bulk datasets revealed that MSPJ showed good performance for DEG detection under small sample sizes.

In this study, the DEGs identified by individual methods were also assessed with different sampling rates. In terms of DEG counts, the rank-based methods were not included in the evaluation. For the 10–30% sampling rate, the T-test, MSPJ, limma, and Wilcoxon's test detected the most DEGs. Moreover, we found that the number of DEGs identified using MSPJ was least influenced by sample size, while the number of DEGs obtained by other methods was more sensitive to sample size. Furthermore, comparing the gene identification and GO terms of sub-datasets and the biomarker classification of sub-datasets from different methods, changes in sample size clearly had the least impact on the mRMR, T-test and MSPJ, in the range of 10–30%. The more detailed analysis revealed that RankProd was demonstrated the most robust DEG detection. Although MSPJ was not the most stable with respect to gene ranking, and it still outperformed many methods across different sampling rates. T-test method outperformed than others in several aspects, however, it had some serious limitations in actual scenario, such as the absence of fold change values of genes between controls and cases, and assumptions regarding the distribution of datasets (normal distribution), and so on [61]. The distribution assumption also applied to SAM and limma (normal or Poisson distribution) [35]. MSPJ had no restrictions with respect to the distribution of the data and therefore could be applied to various types of omics datasets, such as proteomics, metabolomics, and single cell RNA-seq datasets. Moreover, the MSPJ method could be used to visualize expression levels of each gene with estimated significance measurements (Fig. A.2).

Overall, our results support the performance of MSPJ for DEG identification in datasets with small sample sizes, especially those with<30 samples. Although the datasets considered in our study were large and included datasets from various scientific fields, it is not, strictly speaking, representative of a "population of datasets" and hence the results may not be generalizable. Accordingly, the method does not necessarily perform better on all datasets and its performance is not limited to the sample sizes evaluated. Depending on the research requirements, our method can be an additional option for gene discovery.

5. Conclusion

Comprehensive analyses were conducted to evaluate the performance of the newly developed method, MSPJ, for analyses of transcriptome datasets with different sample sizes, including simulated and real datasets. Our systematic large-scale comparative analysis using 165 real datasets revealed that MSPJ showed good average prediction performance for biomarkers, with high rates of common and unique of DEGs and gene function identification. The robustness to sample size enables effective DEG detection by MSPJ. The MSPJ method described here is effective under limited sample sizes for gene expression datasets and thus provides stable scores. Our method can be easily applied to high-throughput transcriptional datasets of any size from microarray or RNA-seq experiments. It is even theoretical possible to apply the method to other omics data types due to the free distribution (e.g., metabolomics, proteomics, and single-cell RNA-seq).

6. Availability of data and code

All artificial datasets used to evaluate the performance of gene selection methods were deposited in the Zenodo repository (https://doi.org/10.5281/zenodo.6320499). The real datasets were collected from GEO and pre-processed according to the pipeline described in this manuscript.

All R source code for MSPJ, dataset analyses and benchmarking in this study was released on GitHub (https://github.com/libcell/MSPJ).

Funding

This work was supported by grants from the Nursery Project of Army Medical University (No. 2019R054), Natural Science Foundation of Chongqing, China (Grant No. CSTC2019JCYJ-MSXMX0527), Open Fund of Yunnan Key Laboratory of Plant Reproductive Adaptation and Evolutionary Ecology, Yunnan University, Chongqing Technology Innovation and Application Development Special key Project (cstc2019jscx-dxwtBX0010), and Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN202100538).

Authorship statement

Hui Yang and Bo Li conceived the study and provided the relevant bioinformatics technologies; HuaChun Yin and Yuyang Peng performed the datasets analysis. HuaChun Yin, Bo Li and JingXin Tao wrote the R code. Ying Xiong and Song Li supervised the project. Song Li, Bo Li, and Huachun Yin wrote the manuscript.

CRediT authorship contribution statement

HuaChun Yin: Software, Methodology, Validation, Formal analysis, Investigation, Data curation, Resources, Writing – original draft, Writing – review & editing, Visualization. **JingXin Tao:** Software, Methodology. **Yuyang Peng:** Validation. **Ying Xiong:** Supervision. **Bo Li:** Conceptualization, Software, Methodology, Resources, Writing – original draft, Writing – review & editing. **Song Li:** Writing – original draft, Writing – review & editing, Supervision. **Hui Yang:** Conceptualization, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.07.022.

References

[1] Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat Rev Genet 2016;17(5):257–71.
[2] Zhao W, Beers DR, Hooten KG, Sieglaff DH, Zhang A, Kalyana-Sundaram S, et al. Characterization of gene expression phenotype in amyotrophic lateral sclerosis monocytes. JAMA Neurol 2017;74(6):677–85.
[3] Ye X, Zhang N, Jin Y, Xu B, Guo C, Wang X, et al. Dramatically changed immune-related molecules as early diagnostic biomarkers of non-small cell lung cancer. FEBS J 2020;287(4):783–99.
[4] Ansai S, Mochida K, Fujimoto S, Mokodongan DF, Sumarto BKA, Masengi KWA, et al. Genome editing reveals fitness effects of a gene for sexual dichromatism in Sulawesian fishes. Nat Commun 2021;12(1):1350.
[5] Avila-Magana V, Kamel B, DeSalvo M, Gomez-Campo K, Enriquez S, Kitano H, et al. Elucidating gene expression adaptation of phylogenetically divergent coral holobionts under heat stress. Nat Commun 2021;12(1):5731.
[6] Lin J, Zhang W, Zhang X, Ma X, Zhang S, Chen S, et al. Signatures of selection in recently domesticated macadamia. Nat Commun 2022;13(1):242.
[7] Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013;14(5):365–76.
[8] Tang ZQ, Han LY, Lin HH, Cui J, Jia J, Low BC, et al. Derivation of stable microarray cancer-differentiating signatures using consensus scoring of multiple random sampling and gene-ranking consistency evaluation. Cancer Res 2007;67(20):9996–10003.
[9] Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. Nat Commun 2017;8:15180.
[10] Blanco N, Harris AD, Magder LS, Jernigan JA, Reddy SC, O'Hagan J, et al. Sample size estimates for cluster-randomized trials in hospital infection control and antimicrobial stewardship. JAMA Netw Open 2019;2(10):e1912644.
[11] Verbruggen F, Aron AR, Band GP, Beste C, Bissett PG, Brockett AT, et al. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. Elife 2019;8.
[12] Fu WJ, Carroll RJ, Wang S. Estimating misclassification error with small samples via bootstrap cross-validation. Bioinformatics 2005;21(9):1979–86.
[13] Zeisel A, Amir A, Kostler WJ, Domany E. Intensity dependent estimation of noise in microarrays improves detection of differentially expressed genes. BMC Bioinf 2010;11:400.
[14] Ye P, Ye W, Ye C, Li S, Ye L, Ji G, et al. scHinter: imputing dropout events for single-cell RNA-seq data with limited sample size. Bioinformatics 2020;36(3):789–97.
[15] Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 2006;22(22):2825–7.
[16] van de Wiel MA, Neerincx M, Buffart TE, Sie D, Verheul HM. ShrinkBayes: a versatile R-package for analysis of count-based sequencing data in complex study designs. BMC Bioinf 2014;15:116.
[17] Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 2005;365(9458):488–92.
[18] Sweeney TE, Haynes WA, Vallania F, Ioannidis JP, Khatri P. Methods to increase reproducibility in differential gene expression via meta-analysis. Nucleic Acids Res 2017;45(1):e1.
[19] Panagiotou OA, Willer CJ, Hirschhorn JN, Ioannidis JP. The power of meta-analysis in genome-wide association studies. Annu Rev Genomics Hum Genet 2013;14:441–65.
[20] Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. Neuroimage 2014;92:381–97.
[21] Yang Q, Li B, Tang J, Cui X, Wang Y, Li X, et al. Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data. Brief Bioinform 2020;21(3):1058–68.
[22] Ren S, Zhang Z, Xu C, Guo L, Lu R, Sun Y, et al. Distribution of IgG galactosylation as a promising biomarker for cancer screening in multiple cancer types. Cell Res 2016;26(8):963–6.
[23] Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;36(5):411–20.
[24] Winkelbeiner S, Leucht S, Kane JM, Homan P. Evaluation of Differences in Individual Treatment Response in Schizophrenia Spectrum Disorders: A Meta-analysis. JAMA Psychiatry 2019;76(10):1063–73.
[25] Ding Y, Wilkins D. Improving the performance of SVM-RFE to select genes in microarray data. BMC Bioinf 2006;7(Suppl 2):S12.
[26] Tsai CA, Chen YJ, Chen JJ. Testing for differentially expressed genes with microarray data. Nucleic Acids Res 2003;31(9):e52.
[27] Yang H, Churchill G. Estimating p-values in small microarray experiments. Bioinformatics 2007;23(1):38–43.
[28] Schwarzer G. meta: An R Package for Meta-Analysis. R News 2007.
[29] Tang Y, Zhang YQ, Huang Z. Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis. IEEE/ACM Trans Comput Biol Bioinform 2007;4(3):365–81.
[30] Dembele D. A flexible microarray data simulation model. Microarrays (Basel) 2013;2(2):115–30.
[31] Assefa AT, Vandesompele J, Thas O. SPsimSeq: semi-parametric simulation of bulk and single-cell RNA-sequencing data. Bioinformatics 2020;36(10):3276–8.
[32] Gautier L, Cope L, Bolstad BM, Irizarry RA. affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004;20(3):307–15.
[33] Chavez A, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, et al. Comparison of Cas9 activators in multiple species. Nat Methods 2016;13(7):563–7.
[34] Bolstad B. preprocessCore: a collection of pre-processing functions. R Package Version 2013.
[35] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43(7):e47.
[36] Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98(9):5116–21.
[37] Katherine S, Pollard SD, van der Laan MJ. Multiple testing procedures: R multtest package and applications to genomics. Bioinformatics and computational biology solutions using R and bioconductor, 2005.
[38] Del Carratore F, Jankevics A, Eisinga R, Heskes T, Hong F, Breitling R. RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets. Bioinformatics 2017;33(17):2774–5.
[39] Castillo-Secilla D, Galvez JM, Carrillo-Perez F, Verona-Almeida M, Redondo-Sanchez D, Ortuno FM, et al. KnowSeq R-Bioc package: The automatic smart gene expression tool for retrieving relevant biological knowledge. Comput Biol Med 2021;133:104387.
[40] F Aragón Royón AJV, A Arauzo Azofra: FSinR: an exhaustive package for feature selection. 2020, arXiv:2002.10330v1.
[41] Breiman L. Random forests. Machine Learn 2001;45:5–32.
[42] Bommert A, Welchowski T, Schmid M, Rahnenfuhrer J. Benchmark of filter methods for feature selection in high-dimensional gene expression survival data. Brief Bioinform 2022;23(1).
[43] Shaban WM, Rabie AH, Saleh AI, Abo-Elsoud MA. A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. Knowl Based Syst 2020;205:106270.
[44] Bader-El-Den M, Teitei E, Perry T. Biased random forest for dealing with the class imbalance problem. IEEE Trans Neural Netw Learn Syst 2019;30(7):2163–72.
[45] Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (N Y) 2021;2(3):100141.
[46] David Meyer ED, Hornik Kurt, Weingessel Andreas, Leisch Friedrich, Chang Chih-Chung, Lin Chih-Chen. Package 'e1071'. R J 2019.
[47] Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics 2005;21(20):3940–1.
[48] Chimienti A, Ferraris C, Pau D. A complexity-bounded motion estimation algorithm. IEEE Trans Image Process 2002;11(4):387–92.
[49] Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol 2019;37(5):547–54.
[50] Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, Lingle W, et al. FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. Cancer Cell 2009;16(3):259–66.
[51] Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al. The developmental transcriptome of Drosophila melanogaster. Nature 2011;471(7339):473–9.
[52] Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PLoS ONE 2008;3(2):e1651.
[53] Wang T, Li B, Nelson CE, Nabavi S. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. BMC Bioinf 2019;20(1):40.
[54] Li HS, Ou-Yang L, Zhu Y, Yan H, Zhang XF. scDEA: differential expression analysis in single-cell RNA-sequencing data via ensemble learning. Brief Bioinform 2022;23(1).
[55] Servant N, Gravier E, Gestraud P, Laurent C, Paccard C, Biton A, et al. EMA - A R package for Easy Microarray data analysis. BMC Res Notes 2010;3:277.
[56] Couronne R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinf 2018;19(1):270.

[57] Garcia-Diaz P, Sanchez-Berriel I, Martinez-Rojas JA, Diez-Pascual AM. Unsupervised feature selection algorithm for multiclass cancer classification of gene expression RNA-Seq data. Genomics 2020;112(2):1916–25.
[58] Xiaoxu H. Nonnegative principal component analysis for cancer molecular pattern discovery. IEEE/ACM Trans Comput Biol Bioinform 2009;7(3):537–49.
[59] Maddirevula S, Kuwahara H, Ewida N, Shamseldin HE, Patel N, Alzahrani F, et al. Analysis of transcript-deleterious variants in Mendelian disorders: implications for RNA-based diagnostics. Genome Biol 2020;21(1):145.
[60] Perez NP, Guevara Lopez MA, Silva A, Ramos I. Improving the Mann-Whitney statistical test for feature selection: an approach in breast cancer diagnosis on mammography. Artif Intell Med 2015;63(1):19–31.
[61] Zhang J, Liu L, Zhao K, Guo C, Li S. SABR for operable stage I non-small-cell lung cancer: comparison to surgery. Lancet Oncol 2021;22(12):e536.