DElite: a tool for integrated differential expression analysis

Davide Baldazzi*, Michele Doni, Beatrice Valenti,
Maria Elena Ciuffetti, Stefano Pezzella and Roberta Maestro

Unit of Oncogenetics and Functional Oncogenomics (CRO), Centro di Riferimento Oncologico di Aviano
(CRO Aviano) IRCCS, Aviano, Italy

One of the fundamental aspects of genomic research is the identification of differentially expressed (DE) genes between two conditions. In the past decade, numerous DE analysis tools have been developed, employing various normalization methods and statistical modelling approaches. In this article, we introduce *DElite*, an R package that leverages the capabilities of four state-of-the-art DE tools: edgeR, limma, DESeq2, and dearseq. *DElite* returns the outputs of the four tools with a single command line, thus providing a simplified way for non-expert users to perform DE analysis. Furthermore, *DElite* provides a statistically combined output of the four tools, and *in vitro* validations support the improved performance of these combination approaches for the detection of DE genes in small datasets. Finally, *DElite* offers comprehensive and well-documented plots and tables at each stage of the analysis, thus facilitating result interpretation. Although *DElite* has been designed with the intention of being accessible to users without extensive expertise in bioinformatics or statistics, the underlying code is open source and structured in such a way that it can be customized by advanced users to meet their specific requirements. *DElite* is freely available for download from https://gitlab.com/soc-fogg-cro-aviano/DElite.

KEYWORDS

bioinformatics, RNA-seq, differential expression analysis, combination methods, NGS, Software, R package, gene expression

1 Introduction

One of the main goals of transcriptome analysis is to identify significant differences in gene expression patterns between groups or conditions. Differentially expressed (DE) genes are identified based on the extent of variation in gene expression levels between two comparison classes and the statistical significance of this variation. In RNA-sequencing analysis, the expression level of a transcriptomic element is quantified as the number of sequenced fragments aligned to it. Nevertheless, the precise quantification of gene expression and the detection of DE genes are affected by several factors besides sequencing technology, such as gene length and nucleotide composition, sequencing depth, isoforms, overlapping transcripts and cohort size (Zhang et al., 2014). In this regard, the performance of the different tools in pinpointing DE genes in cohorts of small size (e.g. *in vitro* experiments where the comparison classes usually consist of few replicates per condition) is poorly defined. A number of DE analysis tools have been developed based on different mathematical and statistical approaches, either parametric or non-parametric, with the aim of minimizing the impact of these factors. As yet, there is no consensus on the most appropriate approach or algorithm that may yield the most reliable results (Kvam et al., 2012).

Ideally, the integration of different DE tools based on different statistics could help identify the most robust results. However, this can be challenging for users without a strong bioinformatics background. A number of user-friendly suites incorporating different DE tools are available, but these essentially generate descriptive and diagnostic plots rather than performing statistical integration of the results (Nguyen et al., 2024; Liu et al., 2023; Teichman et al., 2023; Sangket et al., 2022; Hunt et al., 2022; Chao et al., 2021; Helmy et al., 2021; López-Fernández et al., 2019; Jiménez-Jacinto et al., 2019; Ge et al., 2018; Li and Andrade, 2017; Varet et al., 2016). The statistical combination of the output of different DE tools by P-value combination methods besides increasing statistical power by combining the summary statistics can allow for the detection of patterns or relationships that may not be apparent through descriptive or diagnostic plots.

Here we present *DElite*, a package developed in the R environment that returns the output of four state-of-the-art DE tools, namely edgeR (Robinson et al., 2010), limma (Ritchie et al., 2015), DESeq2 (Love et al., 2014), and dearseq (Gauthier et al., 2020) with a single command line. To enhance detection capability, *DElite* also provides a combined output of the four tools. Six different statistical methods for combining *p*-values are implemented in *DElite*. Finally, *DElite* produces a report that includes detailed descriptions and explanations of each step, as well as tables and graphs of the different stages of the analysis, thus facilitating the interpretation of the results even for non-expert users. In this work, the different approaches of DE analysis and their integration were cross-examined on datasets of varying sizes. Additionally, *in vitro* validations were carried out to determine their performance in detecting DE genes in small datasets.

2 Materials and methods

2.1 *DElite* development

The *DElite* package was developed in R v.4.1.2 and also tested with R v.3.6.3 (R Core Team, 2022). *DElite* wraps in a single command line the serial execution of edgeR (Robinson et al., 2010), limma (Ritchie et al., 2015), DESeq2 (Love et al., 2014), and dearseq (Gauthier et al., 2020). Data import and DE analysis follow the developer guidelines of each tool[1] (Law et al., 2016). The minimum requirements for running *DElite* are the metadata table containing the comparison classes and the quantitative data in the form of a raw count matrix. The standard *DElite* workflow uses default values for filtering thresholds, but these parameters can be customised by the user.

To filter out low-expressed genes and generate filtered counts to be process in parallel by the four tools, *DElite* offers three alternative options: the rowSums function (total sum of the counts attributed to the gene in the entire dataset), the filterByExpr function from edgeR,

or a filter based on the gene variance parameter. Filtered counts are then normalized using the normalization method built-in in each tool and DE analysis is conducted. edgeR, limma, and dearseq compute normalization factors via the calcNormFactors function applying the TMM method. Differently, DESeq2 normalize gene counts via the "median of ratios" method. For each normalization strategy, *DElite* provides a series of plots demonstrating the effect of the filtering and normalization phases, and it also calculates the Cook's Distance (Cook, 1977) for each gene in each sample to identify potential outliers. Moreover, *DElite* generates a number of descriptive plots including MultiDimensional Scaling (MDS), Principal Component Analysis (PCA), volcano plots and heatmaps of DE genes.

Importantly, to improve detection power, *DElite* combines the results from the four tools into a unified output. Specifically, *DElite* re-processes the results of edgeR, limma, DESeq2, and dearseq, by computing the mean of the fold change values returned by each tool. In addition, it calculates a combined *p*-value. The user can select among six different *p*-value combination methods, namely Lancaster's (Lancaster, 1961), Fisher's (Mosteller and Fisher, 1948), Stouffer's (Stouffer et al., 1949), Wilkinson's (Wilkinson, 1951), Bonferroni-Holm's (Holm, 1979), Tippett's (Wishart, 1952). *DElite* also returns the intersection of the genes identified as DE by all four tools, attributing to them the least significant p-value (Max-P). This allows users to identify the most robust observations. The Max-P intersection value and the results from the Lancaster's combination method are provided by default. Based on recent works that suggest that the Wilcoxon rank-sum test better controls false positives rates when dealing with large datasets (Li et al., 2022), *DElite* provides also the results of this test. Adjusted p-values (padj) are then calculated with the Benjamini–Hochberg correction (Benjamini et al., 2001). Finally, the user can define fold-change and adjusted p-value (padj) thresholds to filter for differentially expressed genes. A comprehensive tutorial is included in the tool and provided as Supplementary File S1.

2.2 *DElite* assessment

DElite was tested on both synthetic and real-world RNA-sequencing datasets. Synthetic datasets were generated using the generateSyntheticData function from the compcode R package (v.1.30.0) (Soneson, 2014) as described by Soneson et al. (Soneson and Delorenzi, 2013). Three distinct cohorts of different size were generated: a small cohort consisting of three samples per condition (mimicking a common experimental scenario of *in vitro* experiments); a medium-size cohort with ten samples per condition; a large cohort made of 100 samples per condition. For the small and medium cohorts (both covering 12,500 genes), nine distinct types of datasets, approximating a negative binomial distribution, were systematically generated. For each type of dataset, ten independent replicates were produced. Each dataset type featured a different number, magnitude, and direction of genes expected to be scored as DE, and presence or absence of outliers to simulate real world data (Table 1). The large synthetic cohort (100 samples per condition in ten independent replicates) included 20,000 genes, 10% of which were set as DE genes. DE genes were unevenly distributed between the two comparison classes (40%

TABLE 1 This table lists the nine types of synthetic datasets employed to evaluate the performance of DE approaches implemented in *DElite*. NB, Negative Binomial distribution.

#	Generation method	# Of DE genes	Upregulated genes in either class [%]
1	NB Distribution	0	0
2	NB Distribution	1,250	50
3	NB Distribution	1,250	100
4	NB Distribution	4,000	50
5	NB Distribution	4,000	100
6	NB Distribution + Single Outlier	0	0
7	NB Distribution + Single Outlier	1,250	50
8	NB Distribution + Random Outliers	0	0
9	NB Distribution + Random Outliers	1,250	50

upregulated in one class and 60% in the other). Single and random outliers were introduced using the compcodeR function (Soneson, 2014). Synthetic counts (available upon request) were processed in *DElite*, with features filtered using the filterByExpr function with default parameters. Differential expression was determined using thresholds of padj ≤0.05 and an absolute log2 fold change $\left(|log2\,(FC)|\right)$ ≥ 1. For each tool and combination method, a confusion matrix was constructed and the performance was assessed using a range of metrics including sensitivity (true positive rate), specificity (true negative rate), F1-score (F1, the harmonic mean of precision and sensitivity), and Matthew's correlation coefficient (MCC, a measure describing the agreement between predictions and expectations).
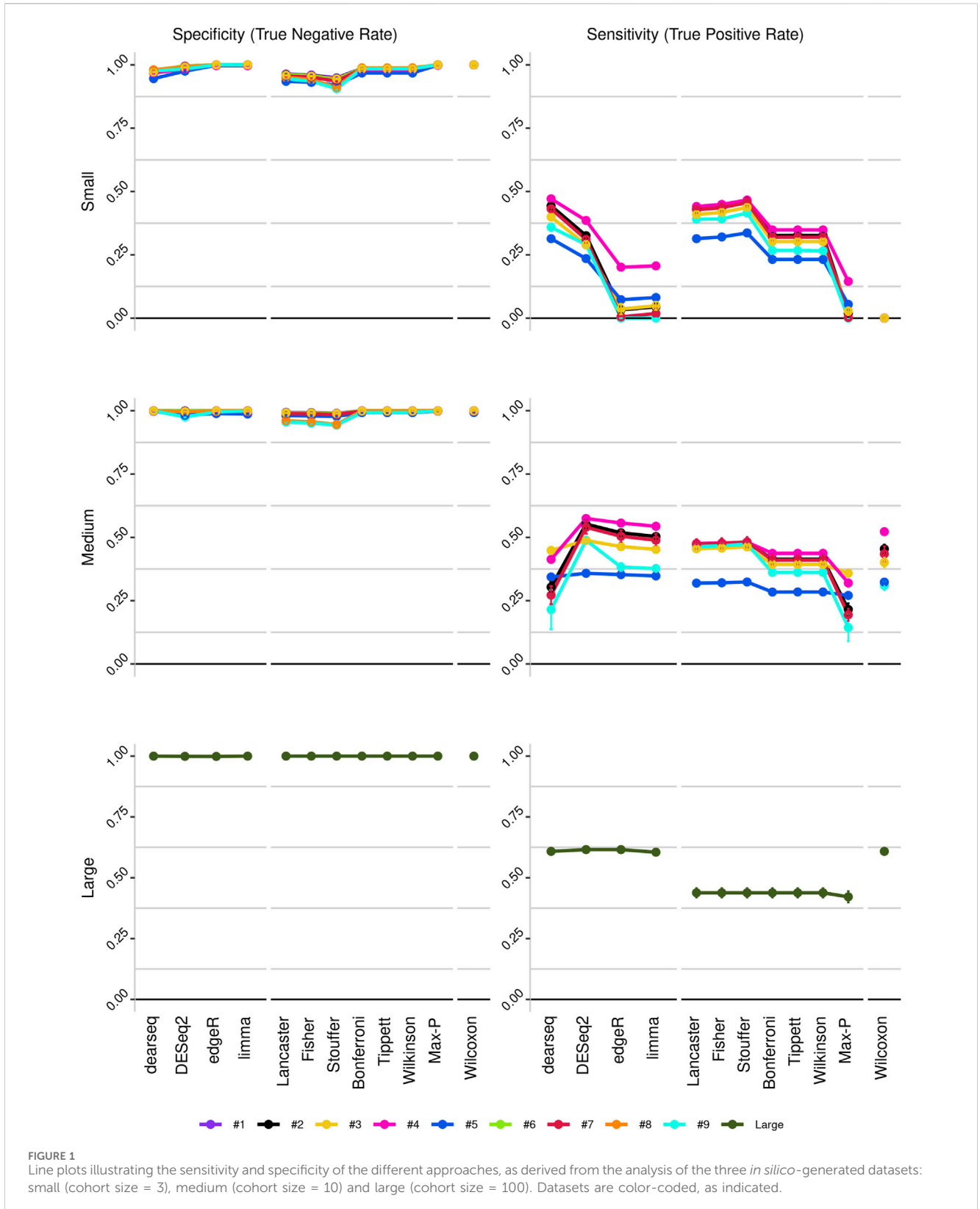
DElite was also run on real-world data. To this end, we used RNA-sequencing data of cell models of Extraskeletal Myxoid Chondrosarcoma (EMC) (project identifier PRJNA692081 at https://www.ncbi.nlm.nih.gov/sra). EMC is a rare tumor that may express two different fusion transcripts, either EWSR1-NR4A3 (EN) or TAF15-NR4A3 (TN). We recently reported that the expression of EN or TN correlates with a differential activation of axon guidance and semaphorin genes, in both human samples and cell lines (Brenca et al., 2021). On these grounds, we used *DElite* to compare the transcriptome of EN and TN cell lines, four biological replicates each (raw data and results are available in Supplementary File S2). A representative set of semaphorins (SEMA3F, 3G, 4C, 4F, 6D) was selected to validate DE analysis results by RT-qPCR (primers are listed in Supplementary File S3). Features with fewer than 10 total counts (rowsums ≥10) were filtered out. Differential expression was determined with thresholds of padj ≤0.05 and an absolute log2 fold change $\left(|log2\,(FC)|\right)$ ≥ 0.6. Total RNA extraction, reverse transcription, and RT-qPCR were as previously described (Brenca et al., 2021). The comparative Ct (ΔΔCt) method and the geometric average of two housekeeping genes (GAPDH and β-actin) were used to calculate relative gene expression.

3 Results

DElite is a novel R package that allows to perform DE analysis based on edgeR, limma, DESeq2, and dearseq tools using a single command line. *DElite* output is designed to be user-friendly and accessible even to users without a strong bioinformatics background. All intermediate and final *DElite* results, including plots and tabular files, are stored into a dedicated directory along with a final report. Besides listing all the steps executed, the report of *DElite* illustrates and describes in detail each step and plot of the analysis.

To improve detection capability, *DElite* also returns the intersection (Max-P) and a statistically combined output of the four algorithms. Six different *p*-value statistical combination methods (Lancaster's, Fisher's, Stouffer's, Wilkinson's, Bonferroni-Holm's, and Tippett's) plus the Wilcoxon rank-sum test are implemented in *DElite*. The performance of these combination methods was evaluated on synthetic datasets (Supplementary Table S1). For each tool and combination method, a confusion matrix was constructed and a comprehensive set of metrics was computed to compare the performance of the different approaches (Figure 1; Supplementary Figures S1–S4). Both individual and combined approaches tended to perform better in terms of sensitivity, F1 score, and MCC when datasets were characterized by a relatively even distribution of DE genes between the two classes (for instance, in Figure 1 compare sensitivity for dataset #4, in which of the DE genes, 50% are upregulated and 50% are downregulated in the test class, and sensitivity for dataset #5, in which all the DE genes are upregulated in one class). Individual tools and *DElite* combined results showed comparable specificity, irrespective of cohort size. In medium and large cohorts, individual tools, primarily DESeq2, edgeR and limma, showed overall superior performance in terms of sensitivity, F1 score, and MCC, whereas combination methods performed better than dearseq. We did not observe the claimed improvement (Li et al., 2022) of the Wilcoxon rank-sum test over DESeq2 and edgeR when dealing with large datasets. Noteworthy, combination approaches demonstrated a subtle improvement in sensitivity compared to single tools, especially edgeR and limma, in the analysis of small datasets (Figure 1). This suggests that when dealing with cohorts of limited size, as is often the case in in vitro experiments, a combination approach may be more effective for the identification of DE genes. To address this hypothesis, *DElite* was run on RNA-sequencing data of cell models mimicking the two biological variants, EN and TN, of a rare tumor (EMC) (Brenca

FIGURE 1
Line plots illustrating the sensitivity and specificity of the different approaches, as derived from the analysis of the three *in silico*-generated datasets: small (cohort size = 3), medium (cohort size = 10) and large (cohort size = 100). Datasets are color-coded, as indicated.

et al., 2021). The results of this analysis are reported in Table 2. We focused on semaphorin genes, which have been previously reported to have a major role in the different biology of EN and TN (Brenca et al., 2021). With the exception of dearseq, which failed to call any

of the investigated semaphorins as DE, limma, edgeR, DESeq2, and *DElite* combination approaches identified a variable number of DE semaphorins (Figure 2). To provide orthogonal validation of these results, a set of targets, representative of three different outcomes,

TABLE 2 DE analysis in TN vs. EN cell models (padj ≤0.05, |log (FC)| ≥ 0.6). The number of upregulated (UP) and downregulated (DOWN) genes are indicated.

		UP	DOWN
Individual tool	dearseq	0	0
	edgeR	460	426
	DESeq2	813	1,571
	Limma	501	388
Combination method	Lancaster	2,620	4,398
	Fisher	2,772	5,027
	Stouffer	3,080	6,617
	Tippett	589	698
	Bonferroni	588	696
	Wilkinson	589	696
	Max-P	0	0

FIGURE 2
The heatmap depicts the DE of 11 semaphorin genes in TN vs. EN cell models. Asterisks indicate the instances where differential expression was |log2(FC)| ≥ 0.6 and padj ≤0.05.
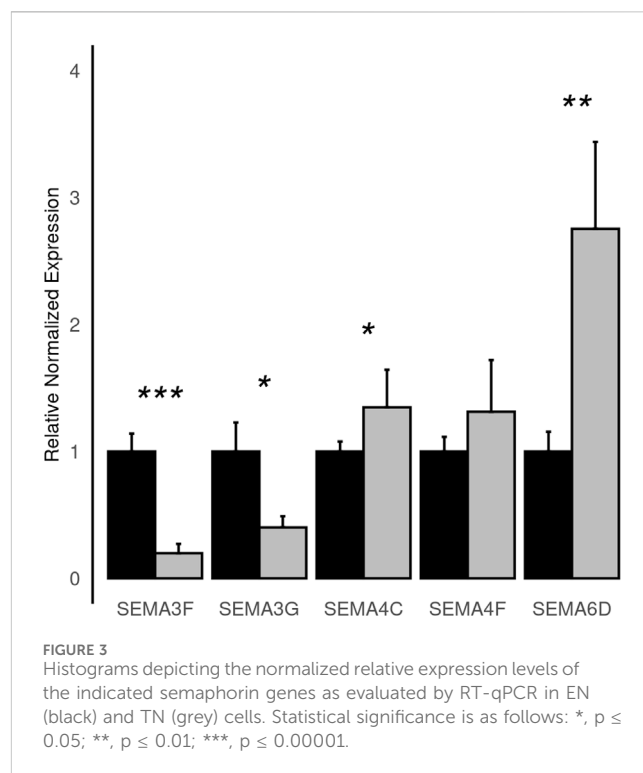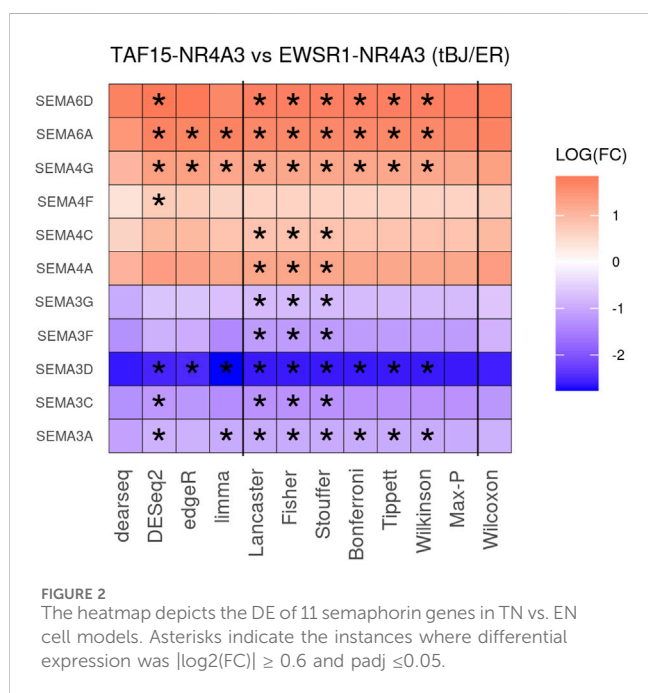

FIGURE 3
Histograms depicting the normalized relative expression levels of the indicated semaphorin genes as evaluated by RT-qPCR in EN (black) and TN (grey) cells. Statistical significance is as follows: *, p ≤ 0.05; **, p ≤ 0.01; ***, p ≤ 0.00001.

were evaluated by RT-qPCR. These included SEMA4F, which was detected as DE exclusively by DESeq2; SEMA6D, identified as DE by both DESeq2 and all *DElite* combination methods; SEMA3G, SEMA3F, and SEMA4C, which were exclusively identified as DE by *DElite* combination methods (Lancaster, Fisher, Stouffer). As illustrated in Figure 3, the differential expression of SEMA4F, detected exclusively by DESeq2, was not confirmed by RT-qPCR. In contrast, all the SEMAs identified by *DElite* combination methods, including the one detected also by DESeq2 (SEMA6D), were confirmed by RT-qPCR to be expressed in a statistically different manner by the two cell models. Taken together, these results support the notion that the integration of diverse DE algorithms, through the use of *p*-value combination methods, may increase the sensitivity in detecting DE genes, particularly when dealing with cohorts of small size.

4 Discussion

This study presents *DElite*, a R package for DE analysis that offers user-friendly functionalities, accompanied by a detailed report. The main advantage of *DElite* is that it enables the execution of DE analysis with four state-of-the-art tools (edgeR, limma, DEseq2, and dearseq) with just a single command line. Moreover, *DElite* presents the additional functionality of combining the results of the four algorithms in a statistically controlled manner, unlike other packages that offer the possibility of running different DE analysis tools, but in most cases only generate descriptive and diagnostic plots (Supplementary Table S2) (Nguyen et al., 2024; Liu et al., 2023; Teichman et al., 2023; Sangket et al., 2022; Hunt et al., 2022; Chao et al., 2021; Helmy et al., 2021; López-Fernández et al., 2019; Jiménez-Jacinto et al., 2019; Ge et al., 2018; Li and Andrade, 2017; Varet et al., 2016). To our knowledge, only ExpressAnalystR and RCPA provide a statistically combined output of the implemented tools. ExpressAnalystR relies on two *p*-value combination methods, whilst RCPA on six. Nevertheless, unlike *DElite*, which can be launched with a single command line, these tools require multiple command inputs, do not provide a final analysis report, and use only parametric approaches (DESeq2, edgeR, limma). As observed on both synthetic and real-world data, *DElite* statistical combination methods appear to improve sensitivity over individual tools, particularly when dealing with small datasets.

The current version of *DElite* is based on the four DE algorithms that represent the today's state-of-the-art for bulk RNA-sequencing data analysis. However, we are committed to further improving it by

integrating additional tools as well as pipelines for single-cell RNA-sequencing data analysis.

Data availability statement

The data used in the study were deposited in the SRA repository, accession number PRJNA692081 (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA692081).

Author contributions

DB: Conceptualization, Data curation, Formal Analysis, Investigation, Software, Writing–original draft. MD: Data curation, Formal Analysis, Investigation, Software, Writing–original draft. BV: Investigation, Validation, Writing–review and editing. MC: Investigation, Validation, Writing–review and editing. SP: Formal Analysis, Software, Writing–original draft. RM: Conceptualization, Funding acquisition, Supervision, Writing–review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2024.1440994/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Confusion matrices reporting true positives, false positives, false negatives, and true negatives computed from the results of the different DE tools and combination methods implemented in DElite applied to small **(A)** and medium **(B)** size datasets. A gradient color scale, from best (green) to worst (red), is used. Numbers within each cell represent the sum of the output of ten independent replicates.

SUPPLEMENTARY FIGURE S2
Line plots illustrating the false negative rate and false positive rate of the different approaches, as derived from the analysis of the three in silico-generated datasets: small (cohort size = 3), medium (cohort size = 10) and large (cohort size = 100). Datasets are color-coded, as indicated.

SUPPLEMENTARY FIGURE S3
Line plots illustrating the F1-score and Matthews's Correlation Coefficient of the different approaches, as derived from the analysis of the three in silico-generated datasets: small (cohort size = 3), medium (cohort size = 10) and large (cohort size = 100). Datasets are color-coded, as indicated.

SUPPLEMENTARY FIGURE S4
Line plots illustrating the accuracy and balanced accuracy of the different approaches, as derived from the analysis of the three in silico-generated datasets: small (cohort size = 3), medium (cohort size = 10) and large (cohort size = 100). Datasets are color-coded, as indicated.

References

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* 125 (1-2), 279–284. doi:10.1016/s0166-4328(01)00297-2

Brenca, M., Stacchiotti, S., Fassetta, K., Sbaraglia, M., Janjusevic, M., Racanelli, D., et al. (2021). NR4A3 fusion proteins trigger an axon guidance switch that marks the difference between EWSR1 and TAF15 translocated extraskeletal myxoid chondrosarcomas published correction appears in. *J. Pathol.* 254 (5), 606. doi:10.1002/path.5737

Chao, K. H., Hsiao, Y. W., Lee, Y. F., Lee, C. Y., Lai, L. C., Tsai, M. H., et al. (2021). RNASeqR: an R package for automated two-group RNA-seq analysis workflow. *IEEE/ACM Trans. Comput. Biol. Bioinform* 18 (5), 2023–2031. doi:10.1109/TCBB.2019.2956708

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics* 19 (1), 15–18. doi:10.2307/1268249

Gauthier, M., Agniel, D., Thiébaut, R., and Hejblum, B. P. (2020). dearseq: a variance component score test for RNA-seq differential analysis that effectively controls the false discovery rate. *Nar. Genom Bioinform* 2 (4), lqaa093. Published 2020 Nov 19. doi:10.1093/nargab/lqaa093

Ge, S. X., Son, E. W., and Yao, R. (2018). iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinforma.* 19 (1), 534. Published 2018 Dec 19. doi:10.1186/s12859-018-2486-6

Helmy, M., Agrawal, R., Ali, J., Soudy, M., Bui, T. T., and Selvarajoo, K. (2021). GeneCloudOmics: a data analytic cloud platform for high-throughput gene expression analysis. *Front. Bioinform* 1, 693836. Published 2021 Nov 25. doi:10.3389/fbinf.2021.693836

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statistics* 6 (2), 65–70.

Hunt, G. P., Grassi, L., Henkin, R., Smeraldi, F., Spargo, T. P., Kabiljo, R., et al. (2022). GEOexplorer: a webserver for gene expression analysis and visualisation. *Nucleic Acids Res.* 50 (W1), W367–W374. doi:10.1093/nar/gkac364

Jiménez-Jacinto, V., Sanchez-Flores, A., and Vega-Alvarado, L. (2019). Integrative differential expression analysis for multiple EXperiments (ideamex): a web server tool for integrated RNA-seq data analysis. *Front. Genet.* 10, 279. Published 2019 Mar 29. doi:10.3389/fgene.2019.00279

Kvam, V. M., Liu, P., and Si, Y. (2012). A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am. J. Bot.* 99 (2), 248–256. doi:10.3732/ajb.1100340

Lancaster, H. O. (1961). The combination of probabilities: an application of orthonormal functions. *Aust. J. Statistics* 3 (1), 20–33. doi:10.1111/j.1467-842X.1961.tb00058.x

Law, C. W., Alhamdoosh, M., Su, S., Smyth, G. K., and Ritchie, M. E. (2016). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res* 5, 1408. Published 2016 Jun 17. doi:10.12688/f1000research.9005.1

Li, Y., and Andrade, J. (2017). DEApp: an interactive web interface for differential expression analysis of next generation sequence data. *Source Code Biol. Med.* 12, 2. Published 2017 Feb 3. doi:10.1186/s13029-017-0063-4

Li, Y., Ge, X., Peng, F., Li, W., and Li, J. J. (2022). Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biol.* 23 (1), 79. Published 2022 Mar 15. doi:10.1186/s13059-022-02648-4

Liu, P., Ewald, J., Pang, Z., Legrand, E., Jeon, Y. S., Sangiovanni, J., et al. (2023). ExpressAnalyst: a unified platform for RNA-sequencing analysis in non-model species. *Nat. Commun.* 14 (1), 2995. Published 2023 May 24. doi:10.1038/s41467-023-38785-y

López-Fernández, H., Blanco-Míguez, A., Fdez-Riverola, F., Sánchez, B., and Lourenço, A. (2019). DEWE: a novel tool for executing differential expression RNA-Seq workflows in biomedical research. *Comput. Biol. Med.* 107, 197–205. doi:10.1016/j.compbiomed.2019.02.021

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Mosteller, F., and Fisher, R. A. (1948). Questions and answers. *Am. Statistician* 2 (5), 30–31. doi:10.2307/2681650

Nguyen, H., Nguyen, H., Maghsoudi, Z., Tran, B., Draghici, S., and Nguyen, T. (2024). RCPA: an open-source R package for data processing, differential analysis, consensus pathway analysis, and visualization. *Curr. Protoc.* 4 (5), e1036. doi:10.1002/cpz1.1036

R Core Team (2022). R: a language and environment for statistical computing. Available at: https://www.R-project.org/.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi:10.1093/bioinformatics/btp616

Sangket, U., Yodsawat, P., Nuanpirom, J., and Sathapondecha, P. (2022). bestDEG: a web-based application automatically combines various tools to precisely predict differentially expressed genes (DEGs) from RNA-Seq data. *PeerJ* 10, e14344. Published 2022 Nov 10. doi:10.7717/peerj.14344

Soneson, C. (2014). compcodeR--an R package for benchmarking differential expression methods for RNA-seq data. *Bioinformatics* 30 (17), 2517–2518. doi:10.1093/bioinformatics/btu324

Soneson, C., and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinforma.* 14, 91. Published 2013 Mar 9. doi:10.1186/1471-2105-14-91

Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A., and Williams, Jr (1949). The American soldier: adjustment during army life. *Stud. Soc. Psychol. World War II* 1, 599. doi:10.1001/jama.1949.02900490055028

Teichman, G., Cohen, D., Ganon, O., Dunsky, N., Shani, S., Gingold, H., et al. (2023). RNAlysis: analyze your RNA sequencing data without writing a single line of code. *BMC Biol.* 21 (1), 74. Published 2023 Apr 7. doi:10.1186/s12915-023-01574-6

Varet, H., Brillet-Guéguen, L., Coppée, J. Y., and Dillies, M. A. (2016). SARTools: a DESeq2-and EdgeR-based R pipeline for comprehensive differential analysis of RNA-seq data. *PLoS One* 11 (6), e0157022. Published 2016 Jun 9. doi:10.1371/journal.pone.0157022

Wilkinson, B. (1951). A statistical consideration in psychological research. *Psychol. Bull.* 48 (2), 156–158. doi:10.1037/h0059111

Wishart, J. (1952) The methods of statistics. By L.H.C. Tippett. Fourth revised edition. Pp 395. 38s. 1952. London: williams and norgate ltd.; New York: john wiley and sons inc. 37 301. (London : Williams and Norgate Ltd. doi:10.2307/3610082*Math. Gaz.*

Zhang, Z. H., Jhaveri, D. J., Marshall, V. M., Bauer, D. C., Edson, J., Narayanan, R. K., et al. (2014). A comparative study of techniques for differential expression analysis on RNA-Seq data. *PLoS One* 9 (8), e103207. Published 2014 Aug 13. doi:10.1371/journal.pone.0103207