Computers in Biology and Medicine 107 (2019) 197–205

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

DEWE: A novel tool for executing differential expression RNA-Seq workflows in biomedical research

Hugo López-Fernández[a,b,c,d,e], Aitor Blanco-Míguez[a,b,f], Florentino Fdez-Riverola[a,b,c], Borja Sánchez[f], Anália Lourenço[a,b,c,g,*]

[a] ESEI: Escuela Superior de Ingeniería Informática, University of Vigo, Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004, Ourense, Spain
[b] CINBIO - Centro de Investigaciones Biomédicas, University of Vigo, Campus Universitario Lagoas-Marcosende, 36310, Vigo, Spain
[c] SING Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Hospital Álvaro Cunqueiro, 36312, Vigo, Spain
[d] Universidade do Porto, Rua Alfredo Allen, 208, 4200-135, Porto, Portugal
[e] Instituto de Biologia Molecular e Celular (IBMC), Rúa Alfredo Allen, 208, 4200-135, Porto, Portugal
[f] Department of Microbiology and Biochemistry of Dairy Products, Instituto de Productos Lácteos de Asturias (IPLA), Consejo Superior de Investigaciones Científicas (CSIC), Paseo Río Linares s/n, 33300, Villaviciosa, Asturias, Spain
[g] CEB - Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal

ARTICLE INFO

Keywords:
Differential expression
RNA-Seq
Open-source software
Workflow management
Translational application

ABSTRACT

Background: Transcriptomics profiling aims to identify and quantify all transcripts present within a cell type or tissue at a particular state, and thus provide information on the genes expressed in specific experimental settings, differentiation or disease conditions. RNA-Seq technology is becoming the standard approach for such studies, but available analysis tools are often hard to install, configure and use by users without advanced bioinformatics skills.

Methods: Within reason, DEWE aims to make RNA-Seq analysis as easy for non-proficient users as for experienced bioinformaticians. DEWE supports two well-established and widely used differential expression analysis workflows: using Bowtie2 or HISAT2 for sequence alignment; and, both applying StringTie for quantification, and Ballgown and edgeR for differential expression analysis. Also, it enables the tailored execution of individual tools as well as helps with the management and visualisation of differential expression results.

Results: DEWE provides a user-friendly interface designed to reduce the learning curve of less knowledgeable users while enabling analysis customisation and software extension by advanced users. Docker technology helps overcome installation and configuration hurdles. In addition, DEWE produces high quality and publication-ready outputs in the form of tab-delimited files and figures, as well as helps researchers with further analyses, such as pathway enrichment analysis.

Conclusions: The abilities of DEWE are exemplified here by practical application to a comparative analysis of monocytes and monocyte-derived dendritic cells, a study of clinical relevance. DEWE installers and documentation are freely available at https://www.sing-group.org/dewe.

1. Introduction

Transcriptomics profiling aims to identify and quantify all transcripts present within a cell type or tissue at a particular state, and thus provides information on which genes are being expressed in precise experimental settings, differentiation or disease conditions. Such profiling is essential to understand how changes in gene expression relate to functional changes in the organism, as well as to provide insights into transcriptional regulation, signalling pathways and gene network organisation [1]. Traditional transcriptomic approaches were based on microarrays cDNA-DNA hybridisation, but high-throughput sequencing of mRNA (also called RNA-Seq) offers many advantages over hybridisation-based studies. Deep sequencing allows the identification and quantification of eventually all mRNA in the samples of the experiment with potentially high accuracy. Accuracy depends on the sequencing depth of a cell type at a specific condition, including small RNAs and other non-coding RNAs, such as micro-RNAs. The increase of sequencing coverage in new platforms and the introduction of depletion

* Corresponding author. ESEI: Escuela Superior de Ingeniería Informática, University of Vigo, Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004, Ourense, Spain.
E-mail address: analia@uvigo.es (A. Lourenço).

https://doi.org/10.1016/j.compbiomed.2019.02.021
Received 22 November 2018; Received in revised form 21 February 2019; Accepted 21 February 2019
0010-4825/ © 2019 Elsevier Ltd. All rights reserved.

techniques have enabled dual RNA-Seq, i.e. to perform simultaneous transcriptomic studies in interacting organisms. For instance, it is now possible to characterise host-pathogen interactions in a single experiment [2]. Moreover, RNA-Seq can identify *de novo* transcripts as it is not dependent on previous probe design and synthesis [3].

The many advantages of RNA-Seq are partly possible due to the generation of an enormous number of raw sequencing reads, typically tens of millions for a standard experiment, which capture even low abundant transcripts. Consequently, the analysis of RNA-Seq data requires software specifically designed to handle huge amounts of data.

Over recent years, a number of data analysis methods and software tools were developed to support the different tasks generally included in RNA-Seq data analysis [4]. Typically, the main stages of a differential expression (DE) workflow include: (*i*) trimming reads and clipping adapters (e.g. using FastQC [5] or Trimmomatic [6]); (*ii*) reading alignments (e.g. using Bowtie2 [7] or HISAT2 [8]); (*iii*) transcript assembly and quantification (e.g. with StringTie [9], Cufflinks [10] or iReckon [11]); and, (*iv*) the DE analysis itself (e.g. supported by Ballgown [12], edgeR [13], DESeq [14], baySeq [15], or Cuffdiff [10]).

Existing software varies greatly in terms of the stages of analysis covered. Notably, some software combines several of the previous tools in order to implement complete workflows [16–18]. Moreover, since the installation, the configuration and the use of these tools are not always trivial, a variety of interfaces exists to help non-proficient end-users [19]. For example, easyRNASeq [20], Next*presso* [21], Galaxy for RNA-Seq [22], RNASeqGUI [23], RobiNA [24], RSeqFlow [25], and SePIA [26].

Despite these efforts, RNA-Seq interfaces are still affected by several technical difficulties [19]. Therefore, this work presents DEWE (Differential Expression Workflow Executor), a new RNA-Seq DE analysis tool that enables the execution of complete workflows by non-proficient users as well as analysis customisation by experienced bioinformaticians. DEWE runs inside a Docker container to expedite installation and configuration in the main operating systems, i.e. Windows, Mac OS X and Linux [27–29]. Likewise, DEWE interface was designed to minimise the software learning curve. Ultimately, the aim of DEWE is to allow less experienced users (in particular, biomedical and health researchers) to use known analysis workflows as a black box, while enabling more advanced users to customise existing workflows, or even build their own pipelines, according to particular needs and interests.

2. Materials and Methods

2.1. DEWE differential expression analysis workflows

DEWE offers built-in, easy-to-configure and well-consolidated workflows to conduct differential expression analyses as well as enables the execution of individual analysis steps. In particular, DEWE workflows entail the following steps (Fig. 1): (i) the creation of a reference index for the genome of interest, (ii) the alignment of reads to the reference index, (iii) transcript assembly and quantification, and (iv) the differential expression analysis itself. Noteworthy, DEWE workflows do not include quality control or pathway enrichment as integrated steps, i.e. these should be manually executed before and after the analysis, respectively. However, it is highly recommended to perform quality control steps over the raw sequence reads [19]. Additionally, DEWE integrates the IGV viewer to support the interactive exploration of the expression files [27]. Supplementary Material S1 collects the third-party tools currently included in DEWE.

2.1.1. Quality control of the samples

Quality control of raw reads includes the analysis of sequence quality, Guanine-Cytosine (GC) content, and the presence of adaptors, among others. To enable such quality controls, DEWE integrates the FastQC [5] and Trimmomatic tools [6].

FastQC provides a modular set of analyses, which gives a basic idea of whether the input data have any problems that could affect downstream analysis. DEWE allows the analysis of multiple raw reads (in FASTQ format) at the same time, generating individual quality reports. Trimmomatic filters raw reads by discarding low-quality reads, trim adaptor sequences and poor-quality bases. DEWE supports the trimming of both single- and paired-end raw sequence reads in FASTQ format. Section 4.1 of the user's manual provides technical details on these steps.

2.1.2. Creation of the reference index

Initially, reads are mapped against a reference genome in order to identify the corresponding genomic positions. For this purpose, DEWE integrates Bowtie2 [7] and HISAT2 [8] tools. Prior to the alignment, a reference index must be created or imported. DEWE accepts reference genomes in FASTA format without any size limitation. Section 5.2.1 of the user's manual provides technical details on this step.

2.1.3. Alignment to the reference sequence

The alignment to the reference sequence allows the collection of subsets of reads and the quantification of the transcripts represented by these reads. As stated before, DEWE integrates Bowtie2 [7] and HISAT2 [8] tools for this purpose. Moreover, DEWE enables the analysis of both single- and paired-end raw sequence reads in FASTQ format without any size limitation.

The output alignment files produced in the Sequence Alignment Map (SAM) format are converted into the Binary Alignment Map (BAM) format using SAMtools [29].

Technical details on this step can be found in sections 4.2-4.3 and 5.3-5.4 of the user's manual.

2.1.4. Transcript assembly and quantification

DEWE uses the StringTie [9] tool to conduct transcript discovery and abundance estimation. StringTie assembles the transcripts from the RNA-Seq reads aligned to the reference index and performs their quantification. It follows a netflow algorithm, i.e. the assembly and quantification of highly expressed transcripts are executed simultaneously, removing the corresponding reads, and repeating the process until all the reads are used. So, DEWE generates one model per sample using a user-uploaded GTF file; then, all these models are merged into a single GTF file, which is used to assemble the transcripts from the input alignments. The transcripts are quantified in terms of Fragments Per Kilobase Million (FPKM) and Transcripts Per Million (TPM). The FPKM quantification is used by the Ballgown tool to perform differential expression analysis.

Additionally, HTSeq [30] is executed to produce raw counts using the alignment files in BAM format. Then, the raw counts are normalised in the Trimmed Mean of M-values (TMM). While the FPKM/TPM normalisation, produced by StringTie, tends to perform poorly when samples have heterogeneous transcript distributions, TMM normalisation is able to ignore highly variable and/or highly expressed features, and thus improves performance [4]. The TMM quantification is used by the edgeR tool to perform differential expression analysis.

Details on how data should be normalised can be found in sections 4.2 and 4.3 of DEWE user's manual.

2.1.5. Differential expression analysis

The last step of the workflow is the differential expression analysis, i.e. the exploratory analysis of the obtained data and corresponding statistical modelling. For this step, DEWE integrates two R packages, i.e. Ballgown [12] and edgeR [13].

Ballgown offers functions to organise, visualise, and analyse the expression measurements of transcriptome assembly. It includes functions for interactive exploration of the transcriptome assembly, visualisation of transcript structures and feature-specific abundances for each locus, and *post hoc* annotation of assembled features to annotated features. In turn, edgeR enables the analysis of RNA-Seq expression

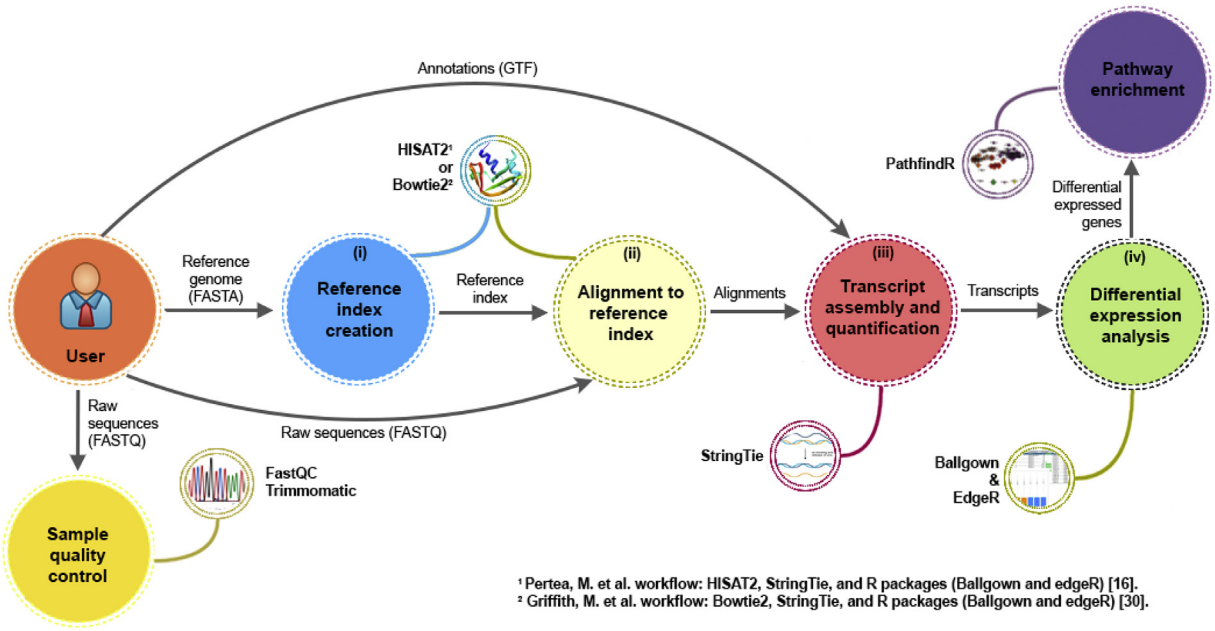

Fig. 1. Steps comprised of a typical differential expression analysis workflow. The central flow represents the main steps of a DEWE workflow. Quality control of the samples and pathway enrichment are included as additional/optional steps. Currently, DEWE implements two workflows, i.e. the Pertea, M. et al. workflow [16], and the Griffith, M et al. workflow [28].

profiles with biological replication. It implements a range of statistical methods based on the negative binomial distributions, including empirical Bayes estimation, exact tests, generalised linear models and quasi-likelihood tests.

DEWE generates a summary report of these two analyses, notably of the overlapping genes, and allows interactive visualisation of the main outputs (see section 3.2 for more details). Technical details on this step are collected in sections 4.2-4.3, and 5.7 of the user's manual.

2.1.6. Pathway enrichment

As an additional step, DEWE enables the discovery of enriched pathways via the R library pathfindR [31]. Using a GSEA approach, pathfindR performs pathway enrichment using active subnetworks, i.e. groups of interconnected genes in a protein-protein interaction network (PIN) containing most of the significant genes. First, the enriched pathways are detected from the PINs, and then the pathways are grouped by user-selected hierarchical cluster analysis. As input, pathfindR uses the significant differentially expressed genes tables (i.e. q-value < 0.05) from Ballgown and edgeR analyses. Section 5.9 of the user's manual provides the corresponding technical details.

2.2. DEWE implementation

The DEWE software v1.2 is implemented in Java 8 using AIBench [32], which is a framework for the rapid development of scientific applications, with several successful biomedical developments [33–36]. The source code of DEWE is divided into four modules: (*i*) the *api*, which contains the Application Programming Interface (API) definition; (*ii*) the *core*, which incorporates the default API implementation; (*iii*) the *gui*, which includes several reusable GUI components; and, (*iv*) the *aibench*, which implements the final GUI application. Thanks to this structure, each part of DEWE, i.e. business logic, GUI components and end-user application, is conveniently isolated.

The first two modules, i.e. the *api* and the *core*, provide the basis for the functionalities offered in DEWE and, most notably, support the integration of third-party applications. A Java interface was developed to run each software using the so-called binaries executor, which acts as a gateway between DEWE and those external applications. The *gui* module, which was created in Java Swing using the freely available extensions SwingX and GC4S [37] (https://www.sing-group.org/gc4s/), provides GUI components to display results and collect user inputs. The JSparklines library was used for enhanced data table visualisation [38]. Finally, the *aibench* module defines the end-user application by creating an AIBench deployment that relies on the other modules.

Since some of the used third-party software (shown in Supplementary Material S1) are only available for Linux-based systems, the Docker technology was used to simplify the installation and setup. Notably, a Docker container, which includes DEWE along with all its

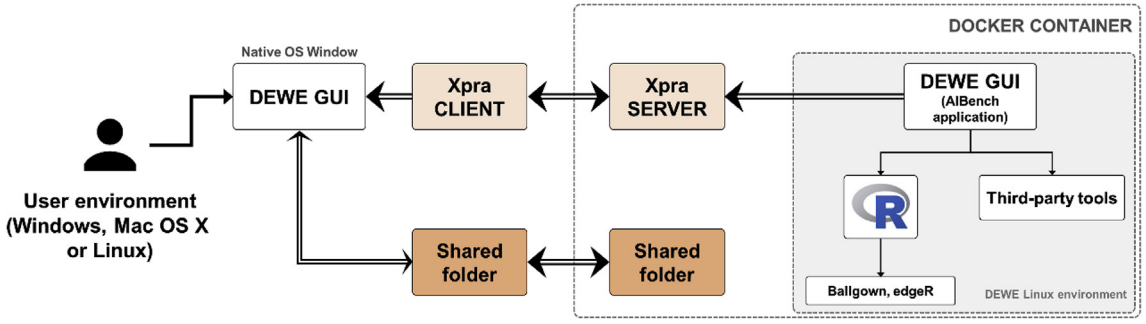

Fig. 2. DEWE deployment architecture.

dependencies, was created. As illustrated in Fig. 2, when this container is executed, the DEWE GUI can be accessed using an Xpra remote display client (https://www.xpra.org/). Since Xpra clients are available for the main operating systems (i.e. Windows, Mac OS X and Linux), DEWE becomes a multiplatform application and, more importantly, issues related with installing and configuring a variety of third-party dependencies are eliminated.

In addition to the Docker-based installers, it is available a VirtualBox (https://www.virtualbox.org/) machine, with all dependencies installed and configured.

The DEWE software is released under a GNU GPL 3.0 License (http://www.gnu.org/copyleft/gpl.html). The software, as well as all documentation and training tutorials, are publicly available at https://www.sing-group.org/dewe. The source code is available at https://github.com/sing-group/dewe.

2.3. Case study dataset: monocyte-derived dendritic cells

The translational application of DEWE to practical biomedical studies is illustrated through the comparative characterisation of the differential expression of monocyte cells and monocyte-derived dendritic cells (moDCs). This dataset comprises reads obtained from human monocyte and moDCs samples and is freely available for download at https://www.sing-group.org/dewe/downloads.html.

These RNA-Seq samples were aligned and annotated against the *Homo sapiens* reference genome. The Pertea, M. et al. workflow [16] was applied to compare the expression of monocytes (i.e. the control condition) and moDCs (i.e. the treatment condition). The introduction and the description of the isolation, differentiation and sequencing methods are exposed in Supplementary Material S2. The results of the DE analysis are presented in the Discussion section and are further disclosed in Supplementary Material S3.

3. Results

The motivation of DEWE is to equip users less proficient in bioinformatics with the means to execute differential expression analyses while enabling GUI-supported advanced customisation if desired. Therefore, among DEWE's main contributions, it is relevant to notice the out-of-the-box use of well-established and varied analysis tools, including the customised execution of individual tools as well as complete workflows, and the user-friendly management and visualisation of a large number of differential expression results generated.

The following subsections describe DEWE contributions in some detail.

3.1. Workflow execution

The sequence of steps to execute a DEWE built-in workflow is depicted in Fig. 3 and can be described as follows:

I. Selection of the workflow to be executed from the *Workflow catalogue*. Currently, DEWE implements two workflows: the Pertea, M. et al. workflow [16], and the Griffith, M et al. workflow [28] (Fig. 1).
II. Introduction or creation of the reference genome index. To build a new index, the reference genome must be provided in FASTA format.
III. Configuration of the workflow, namely: (A) the reference genome index; (B) the name of the experimental conditions; (C) the samples to analyse (in FASTQ or compressed FASTQ format); (D) the annotation file (in GTF format); (E) the working directory where all results will be saved; and, (F) the configuration parameters for the analysis tools. DEWE provides an option to fill all the required information about the samples automatically, and thus minimise the manual effort, and possible errors in data introduction.
IV. Final checking of configuration setup, after which the workflow is launched. As additional support, DEWE shows all commands and steps in the log window as well as displays the execution progress in the progress bar.
V. When the analysis is completed, the results are automatically displayed in the graphical interface.

Fig. 4 collects the list of results generated in the working directory. After a workflow execution, DEWE automatically displays all the generated data tables and enables the generation of additional outputs, in the form of plots and new data tables (refer to the next subsection, *Workflow results*, for more details).

3.2. Workflow results

The main results of a built-in workflow are the outputs provided by the Ballgown and the edgeR packages. At first, DEWE presents a common set of results (Fig. 5):

- The significant differentially expressed genes (q-value < 0.05) between the two conditions (Fig. 5A). A different threshold can be used on user demand.
- The distribution of the p-values obtained for fold-change values (Fig. 5B).
- The distribution of the differential expression fold-change values (Fig. 5C).
- The distribution of the differential expression of the p-values of the genes (Fig. 5D).

Additionally, DEWE makes available results provided only by the Ballgown package, namely:

- The distribution of the FPKM values across the samples (Fig. 6A).
- The distribution of the differential expression p-values of transcripts (Fig. 6B).
- The FPKMs correlation between the two conditions (Fig. 6C and D).
- The heatmap of the FPKM values for the statistically significant genes (q-value < 0.05) (Fig. 6E).
- The principal component analysis of the global variance of the experiment groups (Fig. 6F).

Additionally, the user may generate a plot of the FPKM distribution for a given transcript (Fig. 6G) and a plot describing the structure and expression levels of the distinct isoforms of a particular transcript gene in a given sample (Fig. 6H). The overlap that exists between the predictions of differentially expressed genes outputted by Ballgown and edgeR is described in a data table and a Venn diagram.

Considering the variety of plots of possible interest in DE analyses, and their usual inclusion in scientific articles, DEWE enables figure customisation. By default, the plots illustrated in Figs. 4 and 5 are created in grayscale with a fixed size (1000×1000) and format (JPEG), but the user may customise the format, size and colour.

Detailed information on the results provided for a workflow can be found in section 6 of the user's manual.

3.3. Performance

As a basic reference of the performance to be expected in the practical use of DEWE, Supplementary Material S4 reports the execution times achieved by each workflow over the moDCs case study, described in the Materials and Methods section, and other three example datasets, described in the Supplementary Material S5. These executions were performed in four different environments: (*i*) the DEWE Virtual Machine with three different configurations; (*ii*) Linux using the Docker image, which is equivalent to run it as a native Linux application; (*iii*) Windows using the Docker image; and, (*iv*) Mac OS X using the Docker

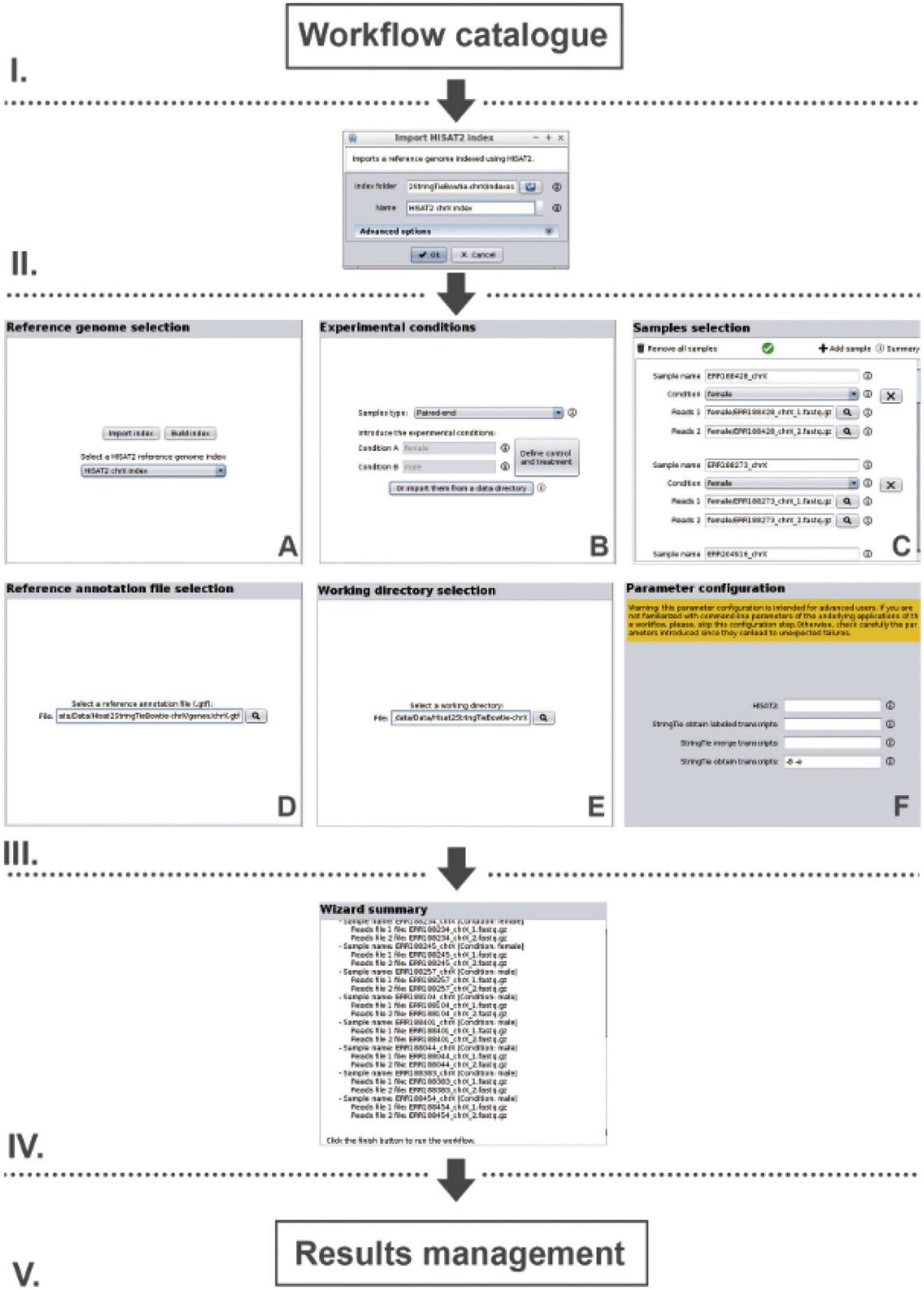

Fig. 3. Steps to execute a DE workflow in DEWE. I. selection of the workflow from the Workflow catalogue. II. build or import of the reference genome index. III. inputs introduction: (A) reference genome, (B) experimental conditions, (C) samples to analyse, (D) annotation file, (E) working directory, and (F) tool parameters. IV. checking the configuration summary. V. visualisation of the results and generation of additional analysis.

image. For each environment, the total execution time is reported along with the execution time of each main analysis step. Due to the size of the *Homo sapiens* HG38 and the moDCs datasets, a minimum of 8 GM of RAM is required to reproduce these analyses.

4. Discussion

The first comparison of DEWE's design premises with those of similar purpose tools enabled the identification of key requirements in

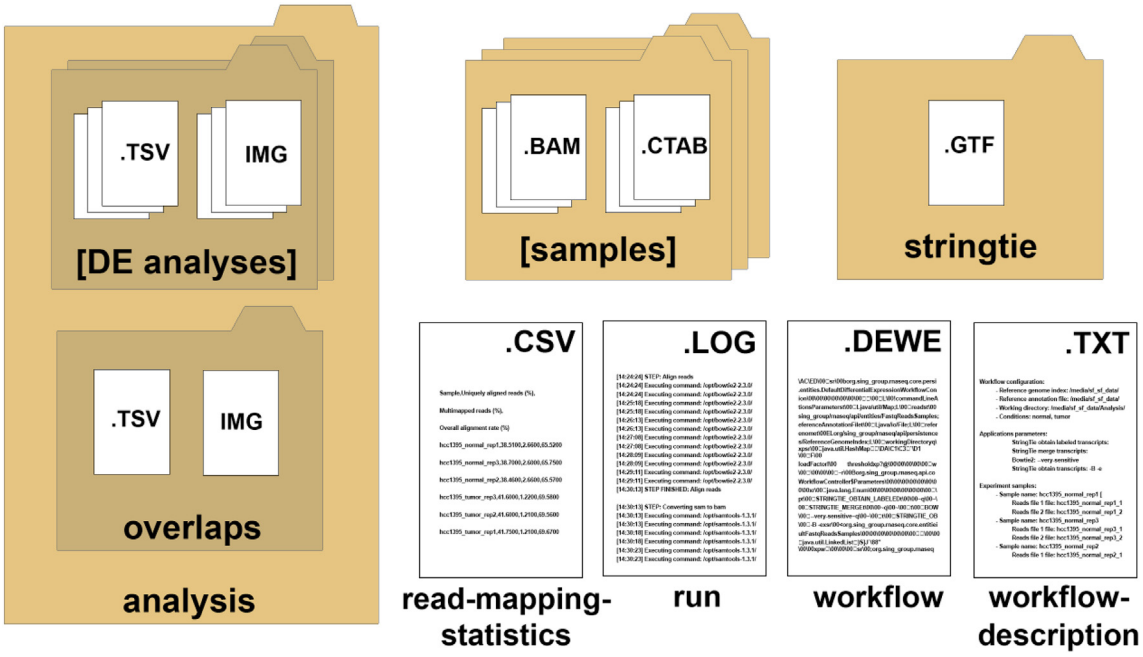

Fig. 4. Results generated after running a DE analysis in DEWE. The *analysis* folder contains the differential expression results; the *samples* folders contain the alignment and transcription files of each sample; the *stringtie* folder contains the merged annotations; the *read-mapping-statistics.csv* file contains the statistical results of the sample alignment; *run.*log file reports all steps and commands executed; the *workflow.dewe* file keeps the workflow configuration and can be imported for further executions; and, the *workflow-description.txt* file contains the summary of the selected inputs and the configuration of the analysis in a human-readable format.

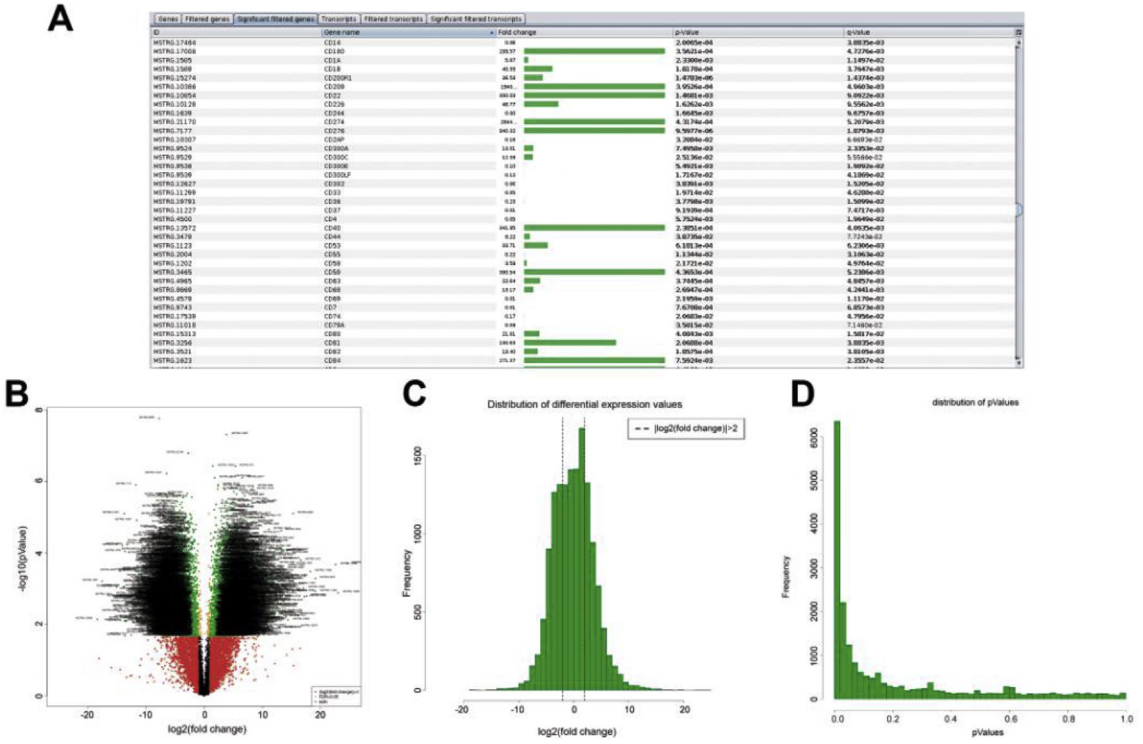

Fig. 5. DE results generated by Ballgown for the moDCs case study. (A) Visualisation of a differentially expressed genes; (B) Volcano plot of p-values against fold changes; (C) Overall distribution of differential expression fold changes for genes; (D) Overall distribution of differential expression.

terms of software installation, configuration, usability and documentation. The software analysed were ArrayExpressHTS [39], easyRNASeq [20], Galaxy [22], PRADA [40], RNASeqGUI [23], and RobiNA [24].

Most of the DE tools are platform dependent, except for RobiNA. To overcome/minimise installation issues, DEWE provides all-in-one installers, i.e. the automatic download and installation of the required components, for Linux, Windows and Mac OS platforms. Similar to RNASeqGUI and RobiNA, DEWE'S interface aims to guide users throughout the analysis. Other tools, such as PRADA, ArrayExpressHTS or easyRNASeq, perform the analysis via command line, which can be somewhat challenging to less proficient users. Regarding

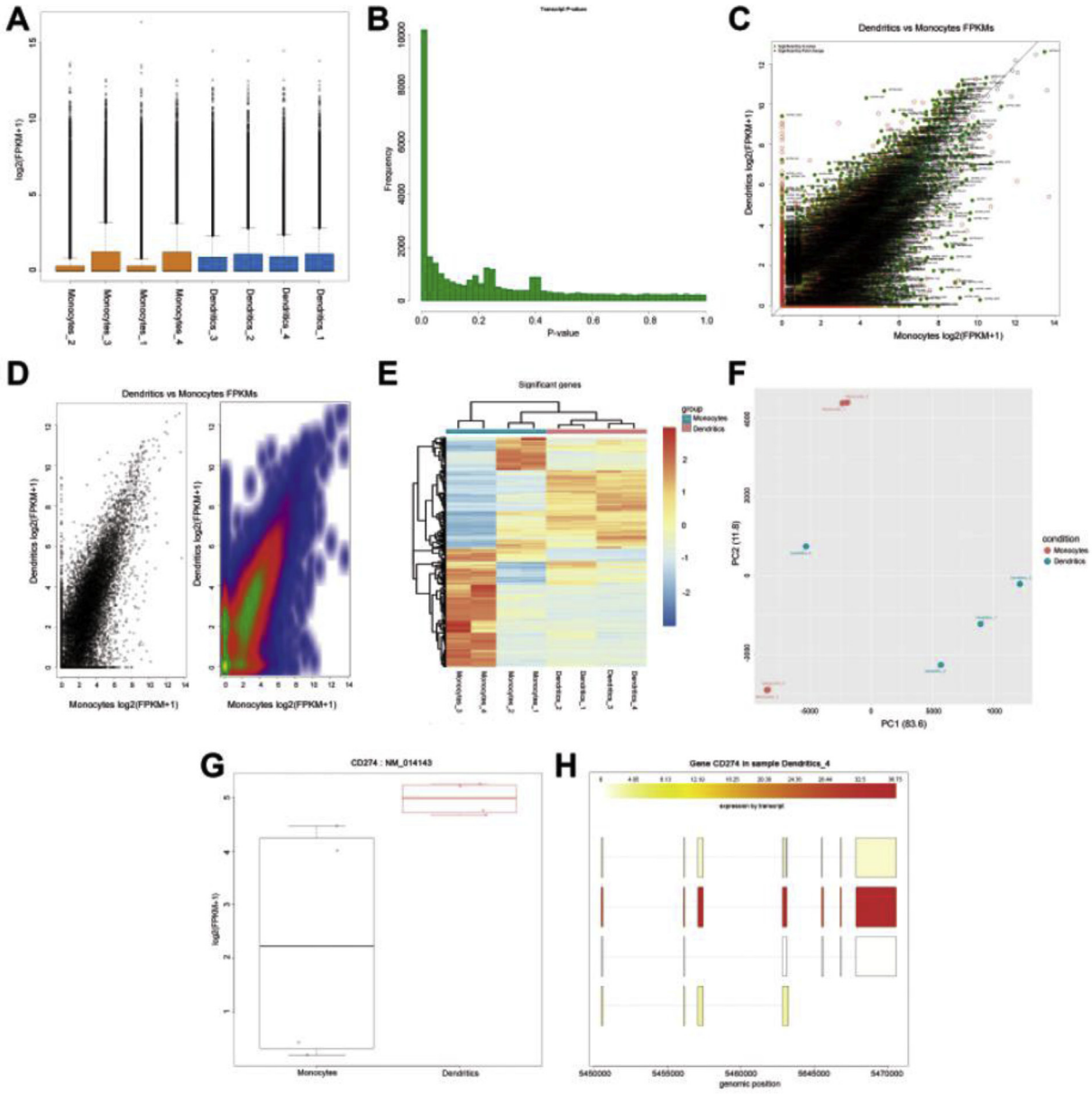

Fig. 6. Additional results generated by Ballgown for the moDCs case study: (A) distribution of FPKM values across the samples; (B) distribution of differential expression p-values for transcripts; (C) FPKM values correlation between the two conditions; (D) FPKM values correlation as density plot; (E) Heatmap of the FPKM values for the significant genes across the samples; (F) principal component analysis plot; (G) FPKM distribution for a transcript; (H) structure and expression levels of the distinct isoforms of the gene CD274 in specific sample.

documentation, DEWE follows the same approach of Galaxy and RNASeqGUI, i.e. provides a web user manual with detailed installation and usage explanations using practical examples. Other tools, such as PRADA or RobiNA, provide somewhat more limited documentation (e.g. missing installation instructions, and fewer examples).

Other key software aspects such as completeness (i.e. whether the tool included all major DE analysis steps), organism's availability, report generation and reproducibility were further inspected. Like most of the similar tools, DEWE can potentially analyse data from any eukaryote organism for which a complete genome is available and implements the main steps of DE analysis, including quality control and reads filtering. Exceptions made for PRADA, which does not support differential analysis and only processes human data, and ArrayExpressHTS that does not support differential analysis. Regarding report generation, DEWE provides an assortment of usual output results, including different data plots as well as run and debug logs. In contrast, PRADA does not allow for generating graphs, and ArrayExpressHTS only provides reports for raw and aligned data. Lastly, DEWE experiments can be easily reproduced using the two files generated upon workflow configuration: one in plain text that allows the user to know the workflow configuration and one in DEWE format for automatic reproduction of the workflow.

The practical use of DEWE was validated using a case study of biomedical interest that compared the gene expression of monocytes and monocyte-derived dendritic cells (moDCs). The Pertea, M. et al. workflow [16] was applied to 4 samples of monocytes and moDCs (Supplementary Material S2). The obtained results (detailed in Supplementary Material S3) revealed that genes encoding transcription regulatory factors, such as FOXO1, RUNX3 and C/EBPo, were highly expressed in monocytes. Among the top up-regulated genes were C1QC, TREM2, HAMP, APOE, PPARG, CEP55, SYK and APOC1, whereas genes such as SYN1 and SERPINA1 were among the top under-expressed genes. As described in previous studies, these expression changes are related to inflammation (interacting with TNF, IFN-γ, and IL-6

cytokines) and lipid metabolism [41]. Another gene of interest was CD14, which is known to be positively expressed in monocyte cells and negatively expressed in moDCs [41]. In the present analysis, this gene obtained a log2-fold change value of −4.06, which agrees with previous studies.

Also, the RNA-Seq results showed that both HLA-DRB1 and HLA-DRB6 were over-expressed in moDCs. Previous studies demonstrated that mature moDCs expressed high levels of MHC class II, a protein complex responsible for the presentation of antigens to T cells, and costimulatory molecules, which mediated in T cell activation [42]. Interestingly, in the present analysis, the HLA-DRA1 gene was found under-expressed, which was unexpected as the MHC class II is a heterodimer composed of an alpha chain encoded by HLA-DRA gene and a beta chain encoded by one of the HLA-DRB genes. This apparent discordance in regulation might be due to inter-individual variability [43]. However, and although some costimulatory molecules necessary in the moDC-T cell interaction were overexpressed (i.e. CD40: log2-fold change value of 8.42, CD80: log2-fold change value of 4.45, PD-L1: log2-fold change value of 11.47, and PD-L2: log2-fold change value of 8.47), other genes mandatory in this interaction (e.g. CD86, ICOSL and OX40L) did not show increased expression (i.e. ICOSL decreased: log2-fold change value of −6.27). All these genes are mandatory for T cell activation [44]. In this sense, the maturation marker CD83 was not overexpressed in moDCs, so RNA-Seq results supported the idea that moDCs generated *in vitro* are immature DCs with a limited ability to activate T cells.

5. Conclusion

DEWE is a new RNA-Seq analysis tool specifically designed to allow users less proficient in bioinformatics to conduct differential expression analyses on their own, whereas enabling analysis customisation and software extension by more advanced users. DEWE offers out-of-the-box, easy-to-configure, and well-established analysis tools, including individual DE steps as well as complete workflows. DEWE's interface enables the user-friendly management of differential expression results, including the preparation of high-quality and publication-ready plots and data tables.

The present paper describes the overall software architecture as well as its main functionalities, linking these descriptions to the corresponding, and more detailed, sections of the user manual. The translational application of DEWE to biomedical problems was illustrated with a case study of clinical relevance: the comparative characterisation of the differential expression of human monocyte cells and monocyte-derived dendritic cells.

DEWE is open to further extension, in particular to new types of analysis and workflows, such as the Cufflinks-based protocol. DEWE (https://www.sing-group.org/dewe) is freely distributed under the GPLv3 license. A comprehensive user manual is available at https://www.sing-group.org/dewe/manual.html.

6. Summary

Transcriptomics profiling aims to identify and quantify all transcripts present within a cell type or tissue at a particular state, and thus provide information on the genes expressed in specific experimental settings, differentiation or disease conditions. RNA-Seq technology is becoming the standard approach for such studies, requiring specific analysis software that facilitate the work of laboratory scientists. Available tools are often hard to install, configure and use by users without advanced bioinformatics skills. Therefore, this paper presents DEWE, an alternative, open source software that aims to reduce the learning curve of less knowledgeable users (in particular, biomedical and health researchers), while enabling analysis customisation and software extension by advanced users. Its two key assets are a user-friendly interface, which enables the customised execution of individual tools as well as of complete, well-established workflows, and broad management and visualisation of differential expression results. Currently, DEWE supports two well-established and widely used differential expression analysis workflows: one combines Bowtie2 and StringTie whereas the other applies HISAT2 and StringTie; and, both use the Ballgown and edgeR packages in the final stages of the analysis. DEWE also enables the tailored execution of individual tools as well as helps with the management and visualisation of differential expression results. It produces high quality and publication ready outputs, e.g. tab-delimited files and figures, as well as aids in further analysis, such as pathway enrichment analysis. Thanks to the Docker container, DEWE can be easily installed in Windows, Mac OS X and Linux operating systems, requiring minimal configuration effort. The practical and translational application of DEWE was illustrated with a comparative analysis between monocytes and monocyte-derived dendritic cells. DEWE installers and documentation are freely available at https://www.sing-group.org/dewe.

Authors' contribution

ABM, AL, BS, and HLF conceived and designed the tool. ABM and HLF built and tested the tool. ABM, AL, BS, FFR and HLF drafted the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of interest

Borja Sánchez is on the scientific board and is a co-founder of Microviable Therapeutics SL. The other authors do not have competing interests.

Acknowledgment

Authors are thankful to Noé Vázquez for his guidance on how to setup Xpra in the Docker image. SING group thanks CITI (Centro de Investigación, Transferencia e Innovación) from the University of Vigo for hosting its IT infrastructure. This work was supported by the Spanish "Programa Estatal de Investigación, Desarrollo e Inovación Orientada a los Retos de la Sociedad" (grant AGL2013-44039R); the Asociación Española Contra el Cancer ("Obtención de péptidos bioactivos contra el Cáncer Colo-Rectal a partir de secuencias genéticas de microbiomas intestinales", grant PS-2016); the Consellería de Educación, Universidades e Formación Profesional (Xunta de Galicia) under the scope of the strategic funding of ED431C2018/55-GRC Competitive Reference Group; the Portuguese Foundation for Science and Technology under the scope of the strategic funding of UID/BIO/04469/2019 unit; and the Asturias Regional Plan I + D + i for research groups (FYCYT-IDI/2018/000236). H. López-Fernández is supported by a post-doctoral fellowship from Xunta de Galicia (ED481B 2016/068-0).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2019.02.021.

References

[1] I.M. Dykes, C. Emanueli, Transcriptional and post-transcriptional gene regulation by long non-coding RNA, genomics, Proteomics, Bioinf. 15 (2017) 177–186, https://doi.org/10.1016/j.gpb.2016.12.005.

[2] A.J. Westermann, L. Barquist, J. Vogel, Resolving host–pathogen interactions by dual RNA-seq, PLoS Pathog. 13 (2017) e1006033, https://doi.org/10.1371/journal.ppat.1006033.

[3] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (2010) 511–515, https://doi.org/10.1038/nbt.1621.

[4] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M.W. Szcześniak, D.J. Gaffney, L.L. Elo, X. Zhang, A. Mortazavi, A survey of best practices for RNA-seq data analysis, Genome Biol. 17 (2016) 13, https://doi.org/10.1186/s13059-016-0881-8.

[5] S. Andrews, FastQC: a quality control tool for high throughput sequence data, (n.d.). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/(Accessed 28 November, 2017).

[6] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120, https://doi.org/10.1093/bioinformatics/btu170.

[7] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359, https://doi.org/10.1038/nmeth.1923.

[8] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (2015) 357–360, https://doi.org/10.1038/nmeth.3317.

[9] M. Pertea, G.M. Pertea, C.M. Antonescu, T.-C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat. Biotechnol. 33 (2015) 290–295, https://doi.org/10.1038/nbt.3122.

[10] C. Trapnell, D.G. Hendrickson, M. Sauvageau, L. Goff, J.L. Rinn, L. Pachter, Differential analysis of gene regulation at transcript resolution with RNA-seq, Nat. Biotechnol. 31 (2012) 46–53, https://doi.org/10.1038/nbt.2450.

[11] A.M. Mezlini, E.J.M. Smith, M. Fiume, O. Buske, G.L. Savich, S. Shah, S. Aparicio, D.Y. Chiang, A. Goldenberg, M. Brudno, iReckon: simultaneous isoform discovery and abundance estimation from RNA-seq data, Genome Res. 23 (2013) 519–529, https://doi.org/10.1101/gr.142232.112.

[12] A.C. Frazee, G. Pertea, A.E. Jaffe, B. Langmead, S.L. Salzberg, J.T. Leek, Ballgown bridges the gap between transcriptome assembly and expression analysis, Nat. Biotechnol. 33 (2015) 243–246, https://doi.org/10.1038/nbt.3172.

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

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

[15] T.J. Hardcastle, Generalized empirical Bayesian methods for discovery of differential data in high-throughput biology, Bioinformatics 32 (2016) 195–202, https://doi.org/10.1093/bioinformatics/btv569.

[16] M. Pertea, D. Kim, G.M. Pertea, J.T. Leek, S.L. Salzberg, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown, Nat. Protoc. 11 (2016) 1650–1667, https://doi.org/10.1038/nprot.2016.095.

[17] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, Nat. Protoc. 7 (2012) 562–578, https://doi.org/10.1038/nprot.2012.016.

[18] J. Costa-Silva, D. Domingues, F.M. Lopes, RNA-Seq differential expression analysis: an extended review and a software tool, PLoS One 12 (2017) e0190152, https://doi.org/10.1371/journal.pone.0190152.

[19] A. Poplawski, F. Marini, M. Hess, T. Zeller, J. Mazur, H. Binder, Systematically evaluating interfaces for RNA-seq analysis from a life scientist perspective, Briefings Bioinf. 17 (2016) 213–223, https://doi.org/10.1093/bib/bbv036.

[20] N. Delhomme, I. Padioleau, E.E. Furlong, L.M. Steinmetz, easyRNASeq: a bioconductor package for processing RNA-Seq data, Bioinformatics 28 (2012) 2532–2533, https://doi.org/10.1093/bioinformatics/bts477.

[21] O. Grana, M. Rubio-Camarillo, F. Fdez-Riverola, D.G. Pisano, D. Glez-Pena, Nextpresso: next generation sequencing expression analysis pipeline, Curr. Bioinform. 12 (2017), https://doi.org/10.2174/1574893612666170810153850.

[22] E. Afgan, D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Grüning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, J. Goecks, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update, Nucleic Acids Res. 44 (2016) W3–W10, https://doi.org/10.1093/nar/gkw343.

[23] F. Russo, D. Righelli, C. Angelini, Advancements in RNASeqGUI towards a reproducible analysis of RNA-seq experiments, BioMed Res. Int. 2016 (2016) 1–11, https://doi.org/10.1155/2016/7972351.

[24] M. Lohse, A.M. Bolger, A. Nagel, A.R. Fernie, J.E. Lunn, M. Stitt, B. Usadel, RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics, Nucleic Acids Res. 40 (2012) W622–W627, https://doi.org/10.1093/nar/gks540.

[25] Y. Wang, G. Mehta, R. Mayani, J. Lu, T. Souaiaia, Y. Chen, A. Clark, H.J. Yoon, L. Wan, O.V. Evgrafov, J.A. Knowles, E. Deelman, T. Chen, RseqFlow: workflows for RNA-Seq data analysis, Bioinformatics 27 (2011) 2598–2600, https://doi.org/10.1093/bioinformatics/btr441.

[26] K. Icay, P. Chen, A. Cervera, V. Rantanen, R. Lehtonen, S. Hautaniemi, SePIA: RNA and small RNA sequence processing, integration, and analysis, BioData Min. 9 (2016) 20, https://doi.org/10.1186/s13040-016-0099-z.

[27] H. Thorvaldsdottir, J.T. Robinson, J.P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, Briefings Bioinf. 14 (2013) 178–192, https://doi.org/10.1093/bib/bbs017.

[28] M. Griffith, J.R. Walker, N.C. Spies, B.J. Ainscough, O.L. Griffith, Informatics for RNA sequencing: a web resource for analysis on the cloud, PLoS Comput. Biol. 11 (2015) e1004393, https://doi.org/10.1371/journal.pcbi.1004393.

[29] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome project data processing subgroup, the sequence alignment/map format and SAMtools, Bioinformatics vol. 25, (2009), https://doi.org/10.1093/bioinformatics/btp352 2078–9.

[30] S. Anders, P.T. Pyl, W. Huber, HTSeq–a Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2015) 166–169, https://doi.org/10.1093/bioinformatics/btu638.

[31] E. Ulgen, O. Ozisik, O.U. Sezerman, pathfindR, An R package for pathway enrichment analysis utilizing active subnetworks, Preprint (2018), https://doi.org/10.1101/272450.

[32] D. Glez-Peña, M. Reboiro-Jato, P. Maia, M. Rocha, F. Díaz, F. Fdez-Riverola, AIBench: a rapid application development framework for translational research in biomedicine, Comput. Methods Progr. Biomed. 98 (2010) 191–203, https://doi.org/10.1016/j.cmpb.2009.12.003.

[33] H. López-Fernández, M. Reboiro-Jato, D. Glez-Peña, F. Aparicio, D. Gachet, M. Buenaga, F. Fdez-Riverola, BioAnnote: a software platform for annotating biomedical documents with application in medical learning environments, Comput. Methods Progr. Biomed. 111 (2013) 139–147, https://doi.org/10.1016/j.cmpb.2013.03.007.

[34] G. Pérez-Rodríguez, D. Glez-Peña, N.F. Azevedo, M.O. Pereira, F. Fdez-Riverola, A. Lourenço, Enabling systematic, harmonised and large-scale biofilms data computation: the biofilms experiment workbench, Comput. Methods Progr. Biomed. 118 (2015) 309–321, https://doi.org/10.1016/j.cmpb.2014.12.005.

[35] H. López-Fernández, M. Reboiro-Jato, D. Glez-Peña, J.R. Méndez Reboredo, H.M. Santos, R.J. Carreira, J.L. Capelo-Martínez, F. Fdez-Riverola, Rapid development of Proteomic applications with the AIBench framework, J. Integr. Bioinform. 8 (2011) 171, https://doi.org/10.2390/biecoll-jib-2011-171.

[36] H. López-Fernández, J.E. Araújo, S. Jorge, D. Glez-Peña, M. Reboiro-Jato, H.M. Santos, F. Fdez-Riverola, J.L. Capelo, S2P: a software tool to quickly carry out reproducible biomedical research projects involving 2D-gel and MALDI-TOF MS protein data, Comput. Methods Progr. Biomed. 155 (2018) 1–9, https://doi.org/10.1016/j.cmpb.2017.11.024.

[37] H. López-Fernández, M. Reboiro-Jato, D. Glez-Peña, R. Laza, R. Pavón, F. Fdez-Riverola, GC4S: a bioinformatics-oriented Java software library of reusable graphical user interface components, PLoS One 13 (2018) e0204474, https://doi.org/10.1371/journal.pone.0204474.

[38] H. Barsnes, M. Vaudel, L. Martens, JSparklines: making tabular proteomics data come alive, Proteomics 15 (2015) 1428–1431, https://doi.org/10.1002/pmic.201400356.

[39] A. Goncalves, A. Tikhonov, A. Brazma, M. Kapushesky, A pipeline for RNA-seq data processing and quality assessment, Bioinformatics 27 (2011) 867–869, https://doi.org/10.1093/bioinformatics/btr012.

[40] W. Torres-García, S. Zheng, A. Sivachenko, R. Vegesna, Q. Wang, R. Yao, M.F. Berger, J.N. Weinstein, G. Getz, R.G.W. Verhaak, PRADA: pipeline for RNA sequencing data analysis, Bioinformatics 30 (2014) 2224–2226, https://doi.org/10.1093/bioinformatics/btu169.

[41] C. Dong, G. Zhao, M. Zhong, Y. Yue, L. Wu, S. Xiong, RNA sequencing and transcriptomal analysis of human monocyte to macrophage differentiation, Gene 519 (2013) 279–287, https://doi.org/10.1016/j.gene.2013.02.015.

[42] S.J. Sung, Monocyte-derived dendritic cells as antigen-presenting cells in T-cell proliferation and cytokine production, (2008), pp. 97–106, https://doi.org/10.1007/978-1-59745-366-0_9.

[43] M. Kitcharoensakkul, L.B. Bacharier, H. Yin-Declue, J.S. Boomer, D. Burgdorf, B. Wilson, K. Schechtman, M. Castro, Temporal biological variability in dendritic cells and regulatory T cells in peripheral blood of healthy adults, J. Immunol. Methods 431 (2016) 63–65, https://doi.org/10.1016/j.jim.2016.02.006.

[44] M. Hubo, B. Trinschek, F. Kryczanowsky, A. Tuettenberg, K. Steinbrink, H. Jonuleit, Costimulatory molecules on immunogenic versus tolerogenic human dendritic cells, Front. Immunol. 4 (2013), https://doi.org/10.3389/fimmu.2013.00082.