Taylor & Francis
Taylor & Francis Group

TECHNICAL PAPER

OPEN ACCESS | Check for updates

3D RNA-seq: a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists

Wenbin Guo [a,b], Nikoleta A Tzioutziou[a], Gordon Stephen[b], Iain Milne [b], Cristiane PG Calixto [a], Robbie Waugh[a,c], John W. S. Brown[a,c], and Runxuan Zhang [b]

aDivision of Plant Sciences, University of Dundee at the James Hutton Institute, Dundee, UK; bInformation and Computational Sciences, The James Hutton Institute, Dundee, UK; cCell and Molecular Sciences, The James Hutton Institute, Dundee, UK

ABSTRACT

RNA-sequencing (RNA-seq) analysis of gene expression and alternative splicing should be routine and robust but is often a bottleneck for biologists because of different and complex analysis programs and reliance on specialized bioinformatics skills. We have developed the '3D RNA-seq' App, an R shiny App and web-based pipeline for the comprehensive analysis of RNA-seq data from any organism. It represents an easy-to-use, flexible and powerful tool for analysis of both gene and transcript-level gene expression to identify differential gene/transcript expression, differential alternative splicing and differential transcript usage (3D) as well as isoform switching from RNA-seq data. 3D RNA-seq integrates state-of-the-art differential expression analysis tools and adopts best practice for RNA-seq analysis. The program is designed to be run by biologists with minimal bioinformatics experience (or by bioinformaticians) allowing lab scientists to analyse their RNA-seq data. It achieves this by operating through a user-friendly graphical interface which automates the data flow through the programs in the pipeline. The comprehensive analysis performed by 3D RNA-seq is extremely rapid and accurate, can handle complex experimental designs, allows user setting of statistical parameters, visualizes the results through graphics and tables, and generates publication quality figures such as heat-maps, expression profiles and GO enrichment plots. The utility of 3D RNA-seq is illustrated by analysis of data from a time-series of cold-treated Arabidopsis plants and from dexamethasone-treated male and female mouse cortex and hypothalamus data identifying dexamethasone-induced sex- and brain region-specific differential gene expression and alternative splicing.

Introduction

RNA-seq is generally considered the method of choice to analyse gene expression. The availability of programs such as *Salmon* [1] and *Kallisto* [2] to quantify transcript and gene-level expression accurately and rapidly allows transcript-level analyses to be both feasible and routine. However, the analysis is often a source of frustration for experimental biologists. The key issues are: 1) reliance on the skills of often overstretched bioinformaticians with the necessary skills to perform the analysis; 2) variation in expertise and knowledge of state-of-the-art methodologies among different bioinformaticians and the continued use of out-of-date and error prone methods/programs (such *Tophat* [3] and *Cufflinks* [4]) affecting the quality and reproducibility of analysis; 3) variation in suitability and accuracy of different RNA-seq analysis programs; 4) variation in level of bioinformatics support and access to bioinformaticians' time (e.g. early career scientists often have more limited access to sufficient bioinformaticians' time); 5) cost of organizational bioinformatics support; 6) length of time commonly needed (often many months) to generate the data analysis; 7) limitations of many RNA-seq differential expression analysis programs to handle complex experimental designs (e.g. time-course or developmental series data); and 8) despite the ever-increasing appreciation that alternative splicing (AS) is an important level of post-transcriptional regulation, most RNA-seq analyses still focus on the gene-level expression thereby ignoring important information that is in the data.

3D RNA-seq is an interactive tool for RNA-seq analysis that is implemented using the R Shiny App. It can be run locally or accessed through a web interface (https://3drnaseq.hutton.ac.uk/). It has been developed to carry out 1) differential expression (DE) analysis of genes and transcripts; 2) differential alternative splicing (DAS) and 3) differential transcript usages (DTU) for RNA-seq data, hence *3D RNA-seq.* Isoform switch (IS) analysis for pairwise [5] and time-series [6] are also incorporated into 3D RNA-seq for advanced AS analysis. The definitions of DE, DAS, DTU and IS are explained in Fig. 1A-C. The program integrates the state-of-the-art, highly rated differential expression analysis tool, *Limma* [7,8], and adopts current best practice for RNA-seq analysis (Fig. 2). *3D RNA-seq* can be used regardless of the type of sample or species under investigation, it can handle complex experimental designs and standardizes the analysis process. An easy-to-use graphical interface takes users

A
DE genes/transcripts
Gene/transcript-based expression change between conditions

B
DAS genes
Compare each transcript change to gene change; then sumarise p-values to gene-level with F-test or Simes method

C
DTU transcripts
Compare each transcript change to the average change of all the remaining transcripts
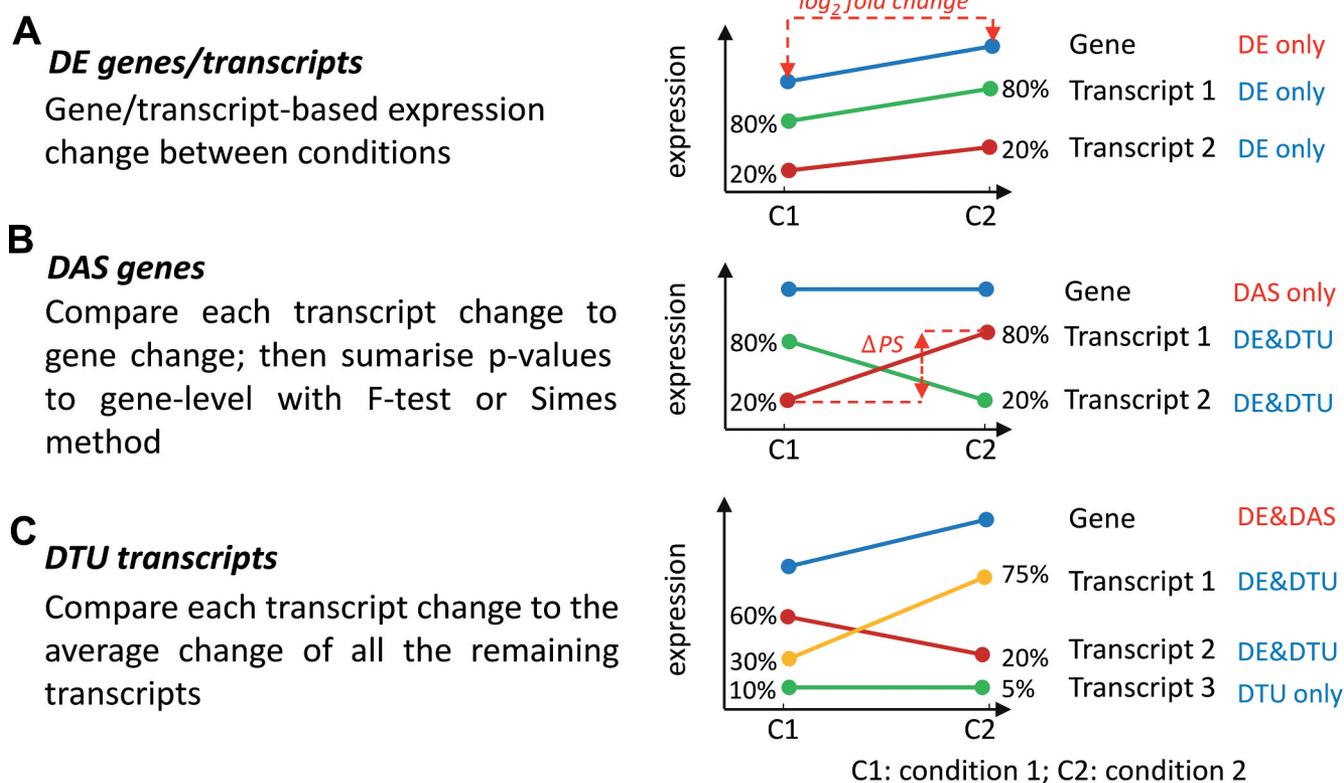
C1: condition 1; C2: condition 2

Figure 1. Definitions and criteria for identification of genes and transcripts with significant DE, DAS and DTU. The lines in the expression profile plots depict the average gene and transcript expression changes between two conditions C1 and C2. Total gene expression (blue) is the sum of the expression of all individual transcripts (red, green or yellow). The percentage values represent the expression ratios of transcripts to the gene. A) DE genes and transcripts are those whose abundance changes between conditions, as measured by changes in log_2 fold change (L₂FC). B) DAS genes must have more than one transcript and are determined by comparing the expression changes between individual transcripts to the gene level between conditions. The change in percentage spliced (ΔPS) is calculated as the percentage change in the abundance of a transcript compared to the total expression from gene. For a gene to be DAS, besides a pre-set p-value cut-off, at least one transcript must have a $\Delta PS \geq 0.1$. C) DTU transcripts are those transcripts which show different expression behaviour from the other transcripts of the same gene. They are determined by comparing the change in expression of each transcript to the average expression change of all the remaining transcripts of the gene. With these criteria, A) DE only genes are those where the gene and transcript expression levels change significantly but transcripts do not change their relative abundance within a gene. B) DAS only genes are those where the gene expression level does not change significantly but that of at least one transcript changes its relative abundance within a gene. C) DE+DAS genes show both significant gene-level expression changes and relative abundance changes of at least one transcript. Isoform switches [ISs] happen when a pair of transcripts reverse their relative abundance across different conditions or time points.

through the different steps of the analysis, visualizes the intermediate and final results through graphics and tables, and generates publication quality figures such as heat-maps, volcano plots, histograms, Venn diagrams and expression profiles. Finally, a comprehensive report is automatically generated at the end, which captures all the input info, each parameter used in every step of the analysis, all of the intermediate figures and tables generated as well as the methodologies used at each step. The full analysis report shows the key findings from the analysis and it provides a useful tool to diagnose the technical issues with the RNA-seq experiment and analysis. More importantly, it provides the ultimate guarantee of reproducible analyses.

There are six main advantages of *3D RNA-seq*: 1) accessibility, ease-of-use, extensive visualization and flexibility allows experimental biologists to control the analysis of their own data; 2) acceleration of RNA-seq analyses – in a typical analysis, RNA-seq data quality control, pre-processing and transcript quantification, the full differential expression/AS analysis and report generation using *3D RNA-seq* can be completed in less than three days; 3) ability to handle complex experimental designs; 4) transcript-level analysis for accurate differential expression and differential AS; 5) generation of

customized report with details of parameters and figures and tables; and 6) providing consistency and reproducibility in RNA-seq analysis to promote open science. We have demonstrated the applicability of 3D RNA-seq to different species by analysing RNA-seq data from complex experiments in Arabidopsis and mouse. The Arabidopsis analysis illustrates its application to time-series RNA-seq data while the mouse data show its ability to analyse data from experiments with multi-factorial designs. In the latter, comparison to the original differential gene expression analysis [9] and an advanced analysis of the same data with Sleuth [10,11], 3D RNA-seq provided far greater resolution in identifying dexamethasone-induced changes in both differential expression and differential alternative splicing.

Results

Application of 3D RNA-seq to RNA-seq analyses of expression/AS in plants

To demonstrate its utility, *3D RNA-seq* was applied to a subset of data from an RNA-seq time-series of Arabidopsis plants
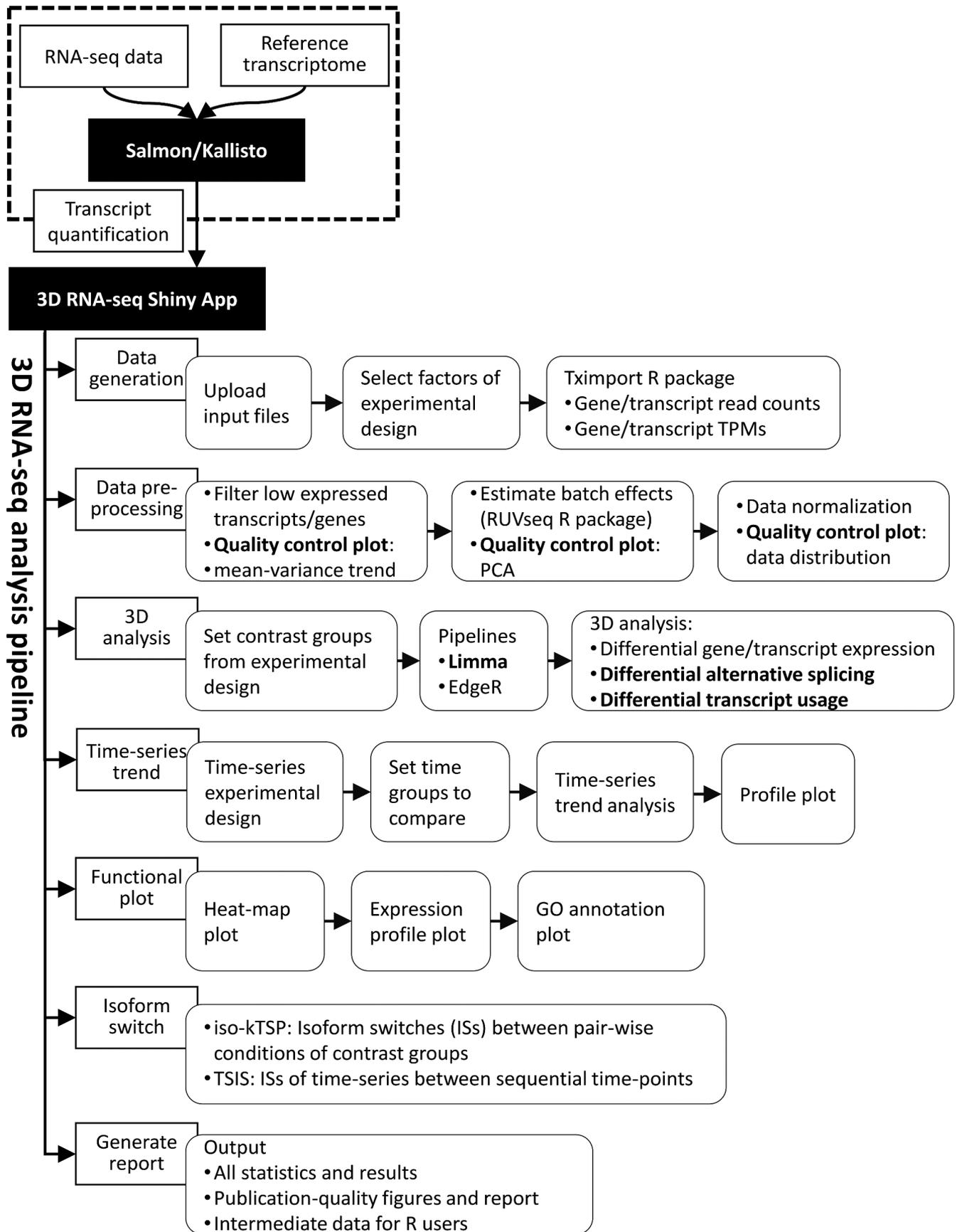
3D RNA-seq analysis pipeline

RNA-seq data

Reference transcriptome

Salmon/Kallisto

Transcript quantification

3D RNA-seq Shiny App

Data generation

Upload input files → Select factors of experimental design → Tximport R package
• Gene/transcript read counts
• Gene/transcript TPMs

Data pre-processing

• Filter low expressed transcripts/genes
• **Quality control plot**:
• mean-variance trend

→ • Estimate batch effects (RUVseq R package)
• **Quality control plot**: PCA

→ • Data normalization
• **Quality control plot**: data distribution

3D analysis

Set contrast groups from experimental design → Pipelines
• **Limma**
• EdgeR

→ 3D analysis:
• Differential gene/transcript expression
• **Differential alternative splicing**
• **Differential transcript usage**

Time-series trend

Time-series experimental design → Set time groups to compare → Time-series trend analysis → Profile plot

Functional plot

Heat-map plot → Expression profile plot → GO annotation plot

Isoform switch

• iso-kTSP: Isoform switches (ISs) between pair-wise conditions of contrast groups
• TSIS: ISs of time-series between sequential time-points

Generate report

Output
• All statistics and results
• Publication-quality figures and report
• Intermediate data for R users

Figure 2. 3D RNA-seq analysis pipeline.

exposed to cold stress (Supplementary Figure S1) [12,13]. Briefly, Arabidopsis Col-0 plants were grown at 20°C for 5 weeks, then the temperature was reduced to 4°C. Samples were harvested every 3 hours for the last day at 20°C, Day 1 at 4°C and Day 4 at 4°C, yielding 26 time-points in total (Supplementary Figure S1A). The data from six of these time-points was extracted to illustrate the utility of *3D RNA-seq*. The six time-points are 3 and 6 h into the dark period from the 20°C Day, Day 1 and Day 4 at 4°C (time points T2 and T3, T10 and T11, and T19 and T20, respectively, referred to here as time-points T2 and T3 from Day 0, Day 1 and Day 4 (Supplementary Figure S1A). The T2 and T3 time-points represent the equivalent time-point in the three days and thereby control for time-of-day expression variation. Transcript quantification was generated using *Salmon* [1] and AtRTD2-QUASI [14].

In the data pre-processing procedure, we removed the low expressed genes/transcripts using the mean-variance trend plots (Supplementary Figure S2) [7] and corrected batch effects using the RUVSeq method [15] in the App (Supplementary Figure S3). Expression data were normalized across samples to reduce sequencing biases. We also employed stringent filters to control the FDR of multiple testing.

Using *3D RNA-seq*, contrast groups were set up to compare the equivalent time-points before and after cold stress to control for time-of-day variation in expression due to photoperiodic and circadian changes [12]. For example, T10 (T2 in day 1 at 4°C) was compared to T2 in day 0 (20°C), T19 (T2 in day 4 at 4°C) was compared to T2 in day 0 (20°C) and so on (Supplementary Figure S1B). Other contrast groups can also be set up (Supplementary Figure S4). For example, the mean between T10 and T11 (T2 and T3 in day 1 at 4°C) can be compared to the mean of T2 and T3 in day 0 (20°C) etc. or the difference between T10 and T11 (T2 and T3 in day 1 at 4°C) can be compared to the difference between T2 and T3 in day 0 (20°C) etc. (Supplementary Figure S1C). We identified 5,023 DE genes and 2,346 DAS genes. These included 1,875 genes which were regulated only by AS (no significant differential expression at the gene level) and 4,185 DTU transcripts (Fig. 3A,B). In addition, 471 of the DAS genes also had significant abundance (DE) changes across the contrast groups (DE+DAS genes). The abundance of significant DTU transcripts can either change significantly (DE+DTU) or non-significantly (DTU-only) (Fig. 1C). Output histograms illustrate the variation in up- and down-regulation of DE genes and the number of significant isoform switches between the contrast groups (Fig. 3C,D). Volcano plots of fold changes in abundance ($\log_2$FC) against significance (FDR) highlight those DE genes and transcripts with the largest and most significant changes in the data (Fig. 3E). Heat maps of clustered DE genes and DTU transcripts are shown for the contrast groups of the six time-points in Fig. 4A and B. Functional annotation analysis was performed using RDAVIDWebService R package with significance threshold FDR < 0.05 [16]. The DE genes showed GO enrichment in cold-induced physiological and molecular events such as response to various stresses, transcription and altered ribosome production (Fig. 4C). Similarly, DAS genes were significantly associated with the spliceosome, RNA splicing and nucleotide/RNA/mRNA

binding terms reflecting cold-induced AS of splicing factors (Fig. 4D). The Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment analysis showed DE genes were significantly enriched with Starch and sucrose metabolism, Biosynthesis of secondary metabolites, Plant hormone signal transduction, Ribosome biogenesis in eukaryotes, Photosynthesis – antenna proteins, Plant–pathogen interactions and Glutathione metabolism pathways, while the DAS genes were related to Spliceosome and mRNA surveillance pathways (Supplementary Table S1). These results reflect the analysis of the complete time-series [12]. Experimental validation of expression, AS and isoform switches by high-resolution RT-qPCR and high-resolution RT-PCR have been described previously [12,13].

The *TSIS* analysis identified 1,688 significant switches between 1,144 pairs of alternatively spliced transcript isoforms in DAS genes with stringent cut-offs: probability >0.5, sum of average abundance differences >1 TPM and p-value <0.05 (Fig. 3D) [6]. These ISs were related to various AS events and approximately half of the isoforms were protein-coding transcripts with different functions according to the transcript annotations in AtRTD2 [14]. The switch frequency along the time-series showed that ISs occurred in the contrast groups comparing 4°C and 20°C time-points (Fig. 3D). Thus, a large number of genes and transcripts in Arabidopsis are sensitive to temperature reduction and likely contribute to acclimation and survival through AS regulation.

Application to RNA-seq analyses in animals

To illustrate the utility of the *3D RNA-seq* App in analysing multi-factor RNA-seq data from animals, we re-analysed RNA-seq data which studied the effects of dexamethasone treatment on cortical and hypothalamic neural progenitor/stem cells (NPSCs) from male and female mice [9]. The RNA-seq data consisted of male and female cortex and hypothalamus cell cultures treated or untreated with dexamethasone, each with three biological replicates (24 samples in total). The study identified differentially expressed genes common and unique to brain region, gender and dexamethasone treatment [9]. The same dataset has also been used to demonstrate the improved resolution of differentially expressed genes using *Sleuth* for transcript-level differential analysis and aggregation of transcript level *p*-values to give gene-level *p*-values [11]. Using these data, we performed two differential expression comparisons between: 1) the *Sleuth*/aggregated *p*-values method [10,11] and *3D RNA-seq* and 2) the results of the differential analysis in [9] and those generated by *3D RNA-seq* expression results.

To compare the results from *3D RNA-seq* and *Sleuth* directly, the *Kallisto* transcript quantifications were downloaded (see Availability of data and materials) and pre-processed in *3D RNA-seq*. This identified 43,836 expressed transcripts which had at least 3 samples with $\geq$ 1 CPM and 12,155 expressed genes which had at least one expressed transcript which were then used in the *Sleuth* and *3D RNA-seq* analyses. In addition, the *Sleuth* analysis in Yi *et al.* [11], only examined the effects of dexamethasone and did not distinguish brain region or gender effects and therefore the same contrast groups were set up in *3D RNA-seq* to match the
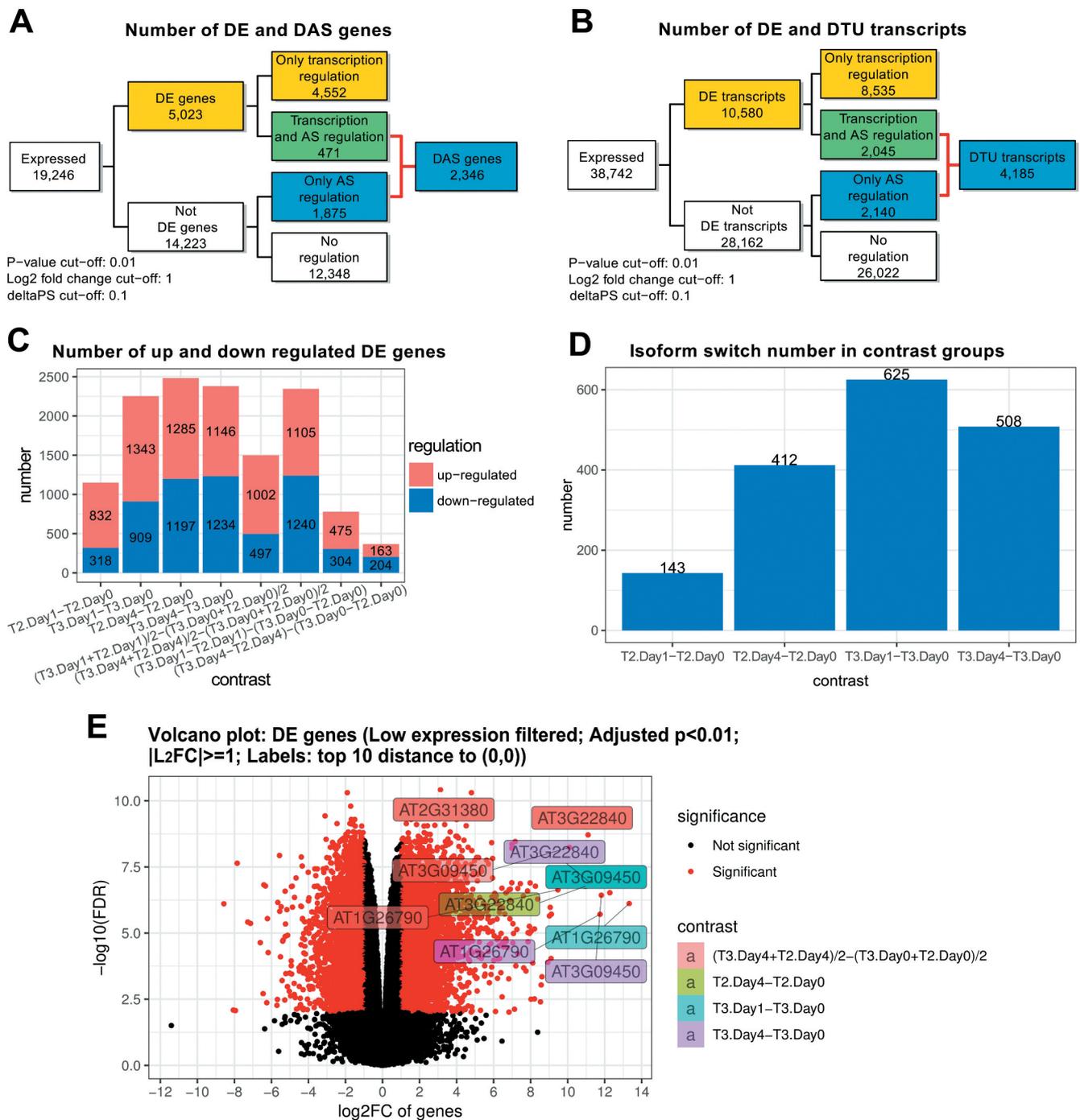
A Number of DE and DAS genes

B Number of DE and DTU transcripts

C Number of up and down regulated DE genes

D Isoform switch number in contrast groups

E Volcano plot: DE genes (Low expression filtered; Adjusted p<0.01; |L2FC|>=1; Labels: top 10 distance to (0,0))

Figure 3. Illustrations of visualization outputs from 3D RNA-seq. A) Summary figure of expressed genes and significant DE, DE+DAS and DAS genes from analysis of the Arabidopsis data; B) Summary figure of expressed transcripts and DE, DE+DTU and DTU transcripts; C) Number of significantly up- and down-regulated DE genes in different contrast groups, and D) Number of significant isoform switches in contrast groups. E) Volcano plot of significant DE genes. The top 10 genes with the smallest *p* values and biggest fold changes are highlighted and different colours refer to different contrast groups.

analysis in Yi *et al.* [11]. We then ran the *Sleuth*/aggregated *p*-values pipeline (see Availability of data and materials) and *3D RNA-seq* on the 43,836 expressed transcripts. Using *Sleuth*, we identified 3,237 DE genes (Fig. 5). GO enrichment analysis was performed using Fisher's exact test in topGO R package [17] in conjunction with genome annotation for mouse org.Mm.eg.db [18]. KEGG enrichment was queried against the KEGG database [19] by using Fisher's exact test. Significantly enriched GO terms

in categories of biological process (BP), cellular component (CC) and molecular function (MF) and KEGG pathways were determined with FDR < 0.05. Significant enrichment terms and pathways relevant to response to stress, immune system, inflammation, hormone response, splicing/spliceosome were extracted to illustrate effects of dexamethasone treatment on gene function (Fig. 6; Supplementary Table S2). The *3D RNA-seq* analysis used the same contrast groups and a BH adjusted
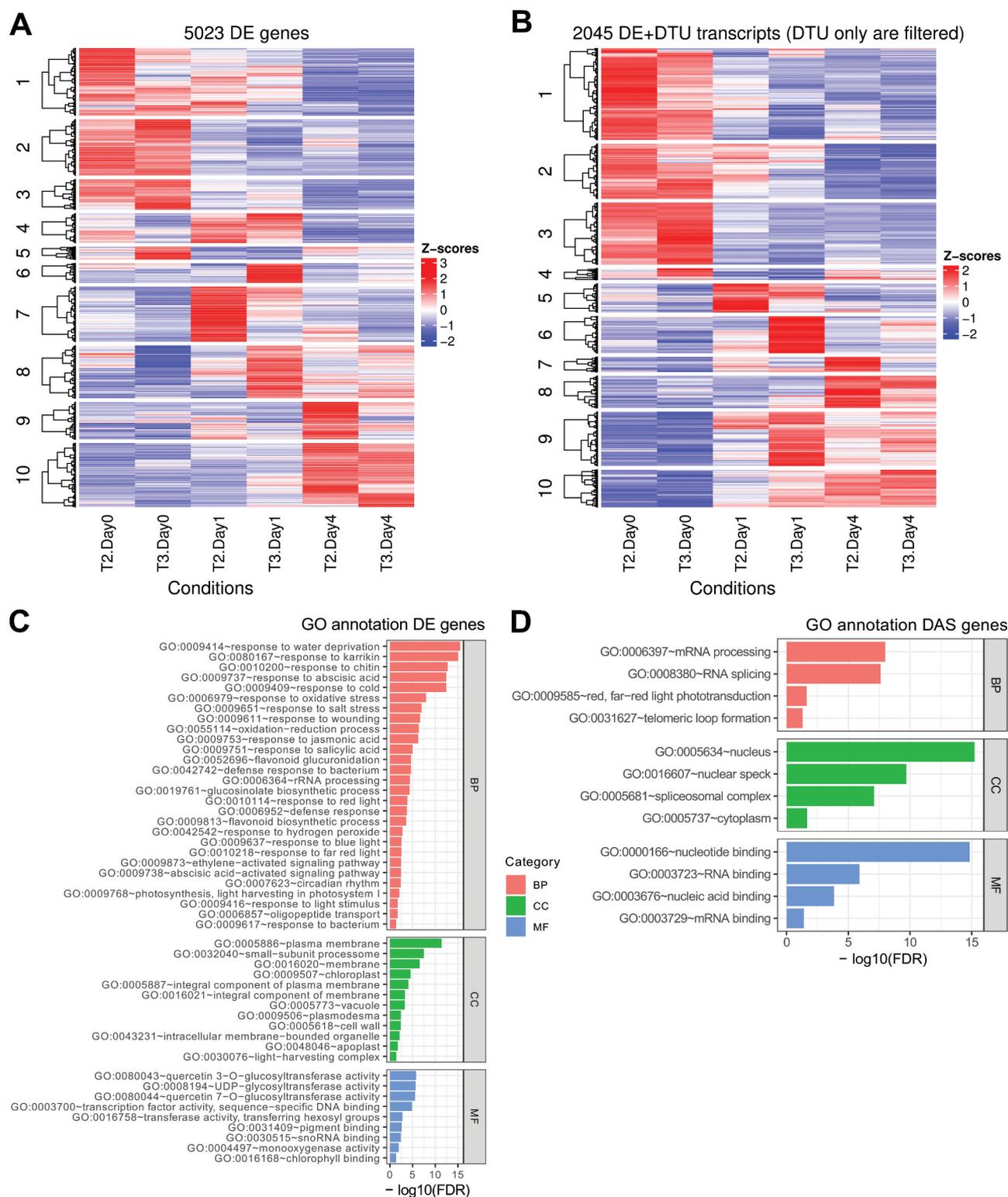
Figure 4. Visualization of clustered genes and transcripts and enriched GO terms. Heatmaps show the grouped expression profiles for A) DE genes and B) DTU transcripts across the samples. The top-enriched GO terms for C) DE and D) DAS genes are visualized with their associated FDRs.

p-value cut-off of <0.05 to identify significant changes [20]. Note that to maintain similar parameters for the direct comparison to *Sleuth*, the log$_2$ fold change and Δ percent spliced cut-offs were not applied in the *3D RNA-seq* analysis. The results of the *3D RNA-seq* pipeline showed a high degree of similarity with the *Sleuth* pipeline in terms of identified DE genes but in addition

resolved differentially alternatively spliced genes from DE genes and identified 1,649 genes with significant expression/AS changes. *3D RNA-seq* identified a total of 4,284 genes with differential expression and/or differential alternative splicing of which 3,700 and 896 were DE and DAS genes, respectively (with an overlap of 312 gene – DE+DAS genes) (Fig. 5). The *Sleuth* pipeline [11],
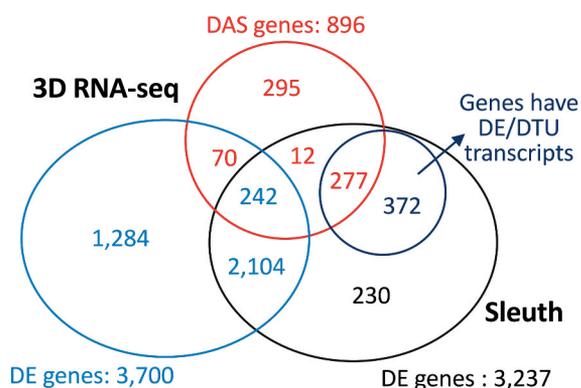
Figure 5. Comparison of the gene lists generated by 3D RNA-seq and Sleuth pipelines. The RNA-seq data on dexamethasone treatment of mice cells was taken from 9. Comparable parameters were applied when running 3D RNA-seq and Sleuth. The Venn diagram compares the DE genes from Sleuth to DE and DAS genes and DE and DTU transcripts from 3D RNA-seq.

which does not identify DAS genes, recovered 3,237 DE genes. Of the total (DE and DAS genes) and DE genes identified by *3D RNA-seq*, 2,635 and 2,346, respectively, were common to both analyses such that 81.4% and 72.5% of the total and DE genes identified by *Sleuth* were identified by *3D RNA-seq* (Fig. 5). The *3D RNA-seq* pipeline also identified 5,573 DE transcripts and 1,480 DTU transcripts with adjusted *p*-value <0.05. Interestingly, 531 of the DAS genes were among the DE genes defined by *Sleuth* – of these, 242 were DE+DAS. Of the 602 DE genes unique to the *Sleuth* analysis, 372 had significant DE/DTU transcripts in *3D RNA-seq* and did not carry over to significant DE or DAS genes (Fig. 5).

The resolution of DE and DAS genes by *3D RNA-seq* was also illustrated in the GO and KEGG pathway enrichment annotations. The DE genes from *Sleuth* had significant terms and pathways that related to response to stress, response to hormone and immune system as well as multiple splicing/spliceosome terms indicating alternative splicing regulation of gene expression (Fig. 6A; Supplementary Table S2A). The separation of DE and DAS genes by *3D RNA-seq* resolved the majority of stress, hormone and immune response terms to DE genes (transcriptional regulation) (Fig. 6B; Supplementary Table S2B) and the splicing enrichment terms (such as spliceosome and mRNA surveillance pathways) to the DAS genes (AS regulation) (Fig. 6C; Supplementary Table S2C). In addition, the 1,284 DE genes unique to the *3D RNA-seq* analysis (Fig. 5) were enriched for steroid hormone receptor activity (Fig. 6D) reflecting the nature of the chemical treatment.

The second comparison exploited the flexibility of *3D RNA-seq* to analyse the effects of multiple factors and detect sex-specific and brain region-specific DE and DAS genes and transcripts in the mouse data. In the original study, the effects of dexamethasone treatment on gene expression in NPSCs from different brain regions (cortical and hypothalamic) and sexes (male and female), were examined [9]. We therefore set up contrast groups (i.e. Female.Cortex.Dex vs Female.Cortex. Vehicle (untreated control), Male.Cortex.Dex vs Male.Cortex. Vehicle, Female.Hypothalmus.Dex vs Female.Hypothalmus. Vehicle and Male.Hypothalmus.Dex vs Male.Hypothalmus. Vehicle). The cut-offs were set as adjusted *p-value* <0.05, $L_2FC \geq$

1 and $\Delta PS \geq 0.1$. Across the four contrast groups, the 3D pipeline identified 930 DE genes and 509 DAS genes, of which 462 were only regulated by AS, and 2,121 DE transcripts and 628 DTU transcripts, of which 455 were AS regulated only. The results of individual contrast groups revealed that 1) both transcription and AS regulation were much more affected in cortex cells than in hypothalamic cells (Fig. 7A-C) and 2) more genes and transcripts were down-regulated in cortex cells while in hypothalamic cells up-regulation dominated the expression changes (Fig. 7A-C). The relative differences between brain regions at the DE gene level was described by Frahm *et al.* [9]. We also compared the DE genes from the 3D pipeline to those of [9] which used a *p-value* cut-off of <0.05 [note: *p*-values were not adjusted and no L_2FC cut-off was applied). There were 449 DE genes in common and 908 significant DE genes in [9] were filtered out in the 3D pipeline due to low L_2FC or insignificant-adjusted *p*-values (Fig. 7D]. Although stringent filters were used, the 3D analysis identified 481 unique DE genes. Finally, the isoform switch analysis identified 63 significant ISs of DAS genes in the contrast groups (Fig. 7E). Thus, compared to the Sleuth pipeline, *3D RNA-seq* provides both transcript- and gene-level analysis to resolve both genes and transcripts differentially expressed due to transcriptional regulation, gene and transcript expression changes due to AS regulation by identifying novel DAS genes, DTU transcripts and ISs. The re-analysis of the mouse data with *3D RNA-seq* demonstrates how novel information can be unlocked from publicly available RNA-seq data which has only been used to analyse differential gene-level expression. Finally, dexamethasone has recently gained prominence reducing mortality in hospitalized patients with COVID-19 that require mechanical ventilation, supplementary oxygen, or Extra Corporeal Membrane Oxygenation (ECMO) (https:// www.sps.nhs.uk/articles/dexamethasone-and-covid-19/). Our results raise the questions of whether dexamethasone treatment might affect expression or alternative splicing of endogenous genes in patients. 3D RNA-seq would be a valuable tool investigating such questions.

Discussion

3D RNA-seq is easy to use and designed for the maximum take-up by biologists

3D RNA-seq has the potential to revolutionize transcriptomics analysis by opening up RNA-seq analysis to wet lab scientists/ biologists with little or no bioinformatics experience. The *3D RNA-seq* web-based application requires no installation and no programming skills from users. The whole analysis can be accomplished entirely within a web browser. The pipeline consists of 24 single steps within seven tabbed panels. It integrates widgets that provide users with technical summaries of the research behind each step so that they can set the parameters appropriate for their studies. Recommended parameters are also given in the manual and are used as defaults. Tutorial and demonstration videos are available on YouTube (see Availability of data and materials). More importantly, *3D RNA-seq* integrates interactive visualization at every step of the analysis so that users can visualize the changes caused by modification of the parameters as they move through the
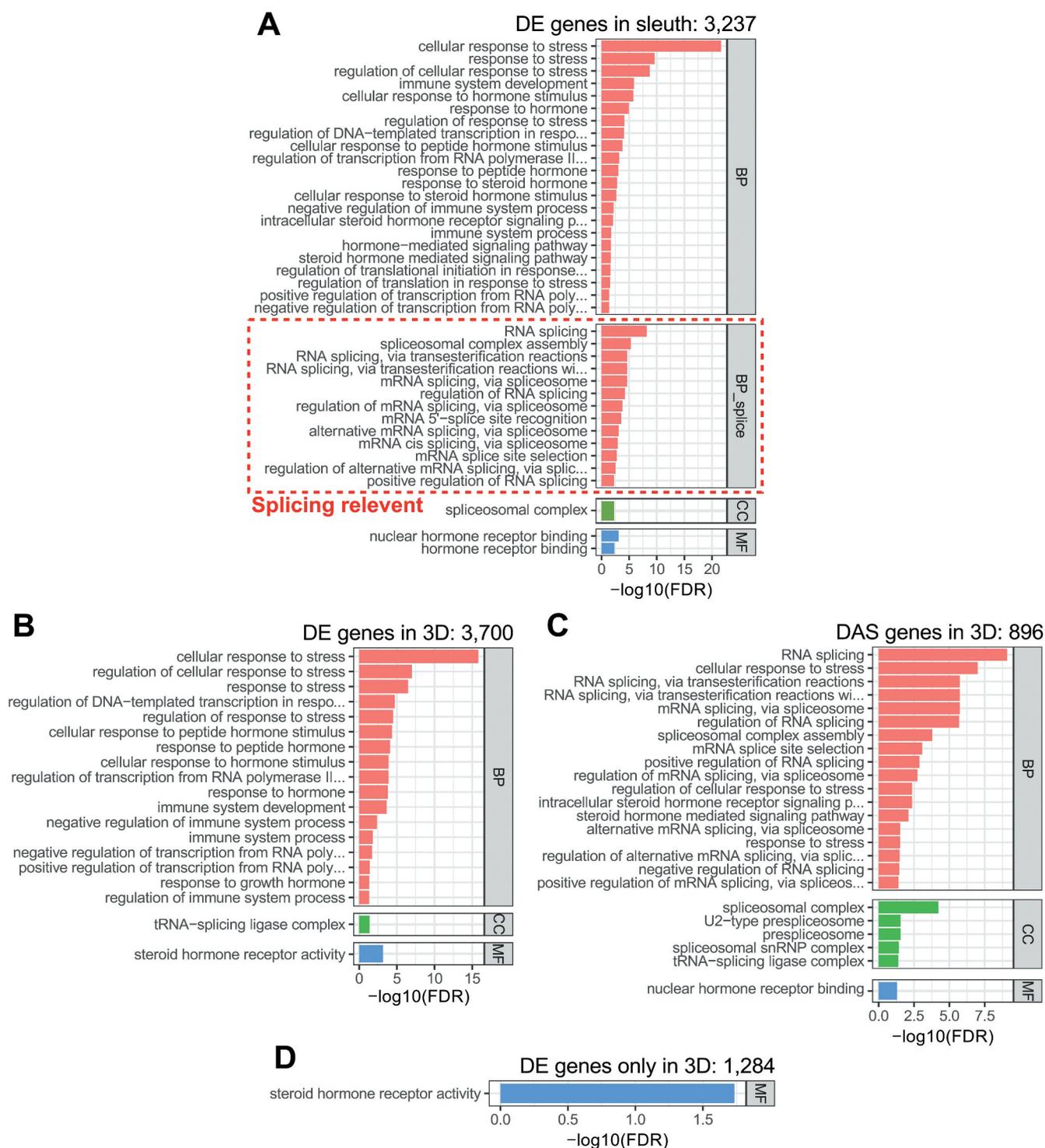
Figure 6. Top-enriched GO terms identified from 3D RNA-seq and Sleuth. The Fisher's exact test and topGO R package were used to generate significant enrichment gene ontology (GO) terms with FDR < 0.05. Significant terms relevant to response to stress, immune system, inflammation, hormone response, splicing/spliceosome are shown in the figure. A) Significantly enriched GO terms of DE genes from Sleuth; B) Significantly enriched GO terms of DE genes from 3D RNA-seq; C) Significantly enriched GO terms of DAS genes from 3D RNA-seq; and D) Significantly enriched GO terms of novel DE genes unique to 3D RNA-seq. Splicing/spliceosome-related GO terms are enriched in the DE genes in *Sleuth* (red dashed box in A) but are found in GO terms associated with DAS genes in 3D RNA-seq (C). BP: Biological process; BP_splice: Biological process with terms of splice, splicing, spliceosome and spliceosomal; CC: Cellular Component; MF: Molecular Function.

pipeline. Interactive visualization not only helps the user to understand the method behind each step better, but it also enables users to explore the optimum settings for the analysis, and ultimately offers reassurance and robustness in the results.

3D RNA-seq also allows the analysis of RNA-seq data with simple and complex experimental designs, such as time-series and developmental series data etc. The comparisons (contrast groups) can be set up between any pairs or groups of samples to cater for different investigations, a feature that is important for the
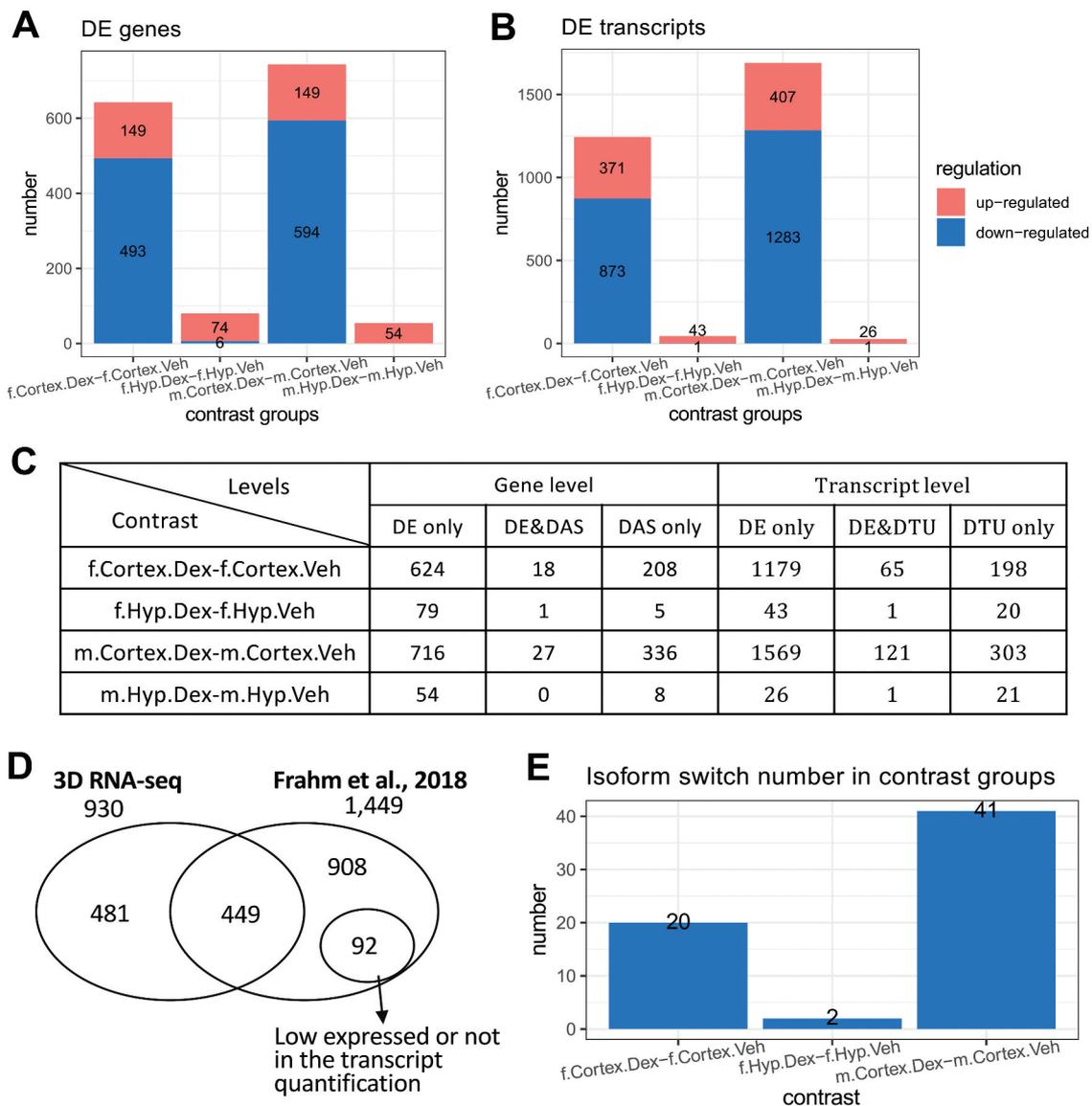
Figure 7. Sex-specific and tissue-specific expression analysis using 3D RNA-seq on the mouse data. Contrast groups were designed to investigate Dex-induced expression and alternative splicing changes between male and female and cortex and hypothalamus brain regions. Significant DE gene/transcript lists were generated by BH adjusted p-value < 0.05, $L_2FC \geq 1$ and $\Delta PS \geq 0.1$. A) Up- and down-regulated DE genes and B) DE transcripts. C) Summary of statistical analysis results from 3D RNA-seq in each contrast group. D) Venn diagram comparing the DE genes in the 3D RNA-seq analysis to the results in Frahm *et al.* [9], in which the significant DE genes were determined by p-value < 0.05 (multiple testing adjustment and L_2FC cut-off were not applied). 92 genes had low expression and were not included in the transcriptome quantification in 3D RNA-seq analysis. E) the number of Isoform switches in different contrast groups with the following cut-offs: probability of switch $\geq$ 0.5, difference of average TPMs at different conditions $\geq$ 1 TPM and adjusted p-value of the TPM difference < 0.05.

universal applicability of *3D RNA-seq* for RNA-seq analysis. The *3D RNA-seq* App can be used to analyse RNA-seq data from any species for which transcript quantifications using *Salmon/Kallisto* can be generated. We use *Salmon* or *Kallisto* because they are fastest and most accurate quantification methods for RNA-seq analysis. The accuracy of the quantifications using *Salmon* and *Kallisto* will depend on the quality and comprehensiveness of the reference transcriptome used to generate the transcript quantifications [14,21,22]. Multiple factors such as missing transcripts, transcript redundancy, transcript fragments and mis-annotations at the 5′ and 3′ ends of isoforms can drastically affect accuracy of transcript quantification [14,22–24]. For many species, comprehensive, optimized transcriptomes do not exist and using partial,

incomplete transcriptomes and/or which have not been filtered to remove such erroneous isoform information will also produce an inaccurate transcript and AS quantification. As transcriptomes improve for many species, transcript quantification and results from *3D RNA-seq* will also improve.

The speed and accuracy of the *3D RNA-seq* analysis have the potential to revolutionize gene expression research programmes. Generation of transcript quantifications and running *3D RNA-seq* is performed in less than three days. It reduces an analytical step which previously required skilled bioinformaticians and often took many weeks or even months to complete. This is significant not only in terms of satisfaction in result generation for the individual scientists/biologists but also for strategic

planning of research programmes where now multiple, consecutive RNA-seq experiments can be planned within the period of a funding proposal. More importantly, the rapid turn-over of the analysis provided by *3D RNA-seq* creates a level playing field for research groups with very different access to bioinformatics expertise or resources thereby supporting, in particular, smaller groups and early-career scientists. *3D RNA-seq* will also aid research projects in developing countries with limited access to computational infrastructures and bioinformatics expertise. Although the quantification of transcripts is not included in 3D RNA-seq, we have provided a tutorial document and YouTube video on performing such tasks using publicly available Galaxy servers (https://usegalaxy.eu/) (see Availability of data and materials). The transcript quantification output from Galaxy can be used as input to 3D RNA-seq directly.

3D RNA-seq facilitates speedy publication of RNA-seq analysis and improves the transparency and reproducibility of the analysis

The fast turn-over of the analysis provided by *3D RNA-Seq* will present a significant advantage for speedy publication of the results. In *3D RNA-seq*, multiple types of figures/summaries are generated and saved including commonly used visualizations such as heatmaps, GO enrichment plots, expression profile plots, volcano plots etc. Users can easily generate and save new figures for each new set of parameters. The format, resolution and size of the figure can be customized and previewed in *3D RNA-seq* before saving. In addition, the significant DE/DAS/DTU/IS gene and transcript lists are saved for further interpretation. The final technical report generated in the final step of *3D RNA-seq* records each step in the analysis with the parameters used and integrates all the saved figures. The report is comprehensive, accurate and reproducible including all the information required for 'Material and Methods' sections for the RNA-seq analysis as well as figures for the Results and Supplementary Materials. Furthermore, some publications have very poor technical descriptions of the methods and parameters used for RNA-seq analyses. In future, submission of such a report along with manuscripts to scientific journals would facilitate transparency and allow reviewers to identify issues with the analysis and how it relates to other published work.

3D RNA-seq provides enhanced alternative splicing analysis at the transcript level

Over the last 20 years, genome-wide expression analyses have relied on microarrays and more recently, RNA-seq analyses mostly at the gene level. RNA-seq allows gene expression to be analysed at the level of both genes and transcripts which in turn provides a means to study post-transcriptional processes such as alternative splicing. AS regulation plays key roles in gene expression and novel genes with altered gene expression at the AS level have been found in most of the RNA-seq studies. Despite the importance of AS and the potential to include AS analysis routinely in RNA-seq, AS is still largely ignored. For example, 4,065 publications from the Web of Science Core Collection (2008–2019) (https://wok.mimas.ac.uk/) were retrieved with the term 'RNA-seq differential gene expression', but only ca. 289 included 'differential alternative splicing'. *3D RNA-seq* provides by far the most detailed differential expression analysis on the transcript and alternative splicing level. In particular, it allows the identification of both DAS genes and DTU transcripts which are differentiated from DE genes/transcripts and a range of measurements, visualizations and statistical tests provide a comprehensive analysis of alternative splicing alongside differential expression. Additionally, *3D RNA-seq* has also integrated methods that detect isoform switches which can play pivotal roles in re-programming of gene expression through switching of functionally different transcript isoforms between, for example, normal and tumour tissues to provide signatures for cancer diagnostics and prognostics [5]. With the enhanced AS analysis in *3D RNA-seq*, key and novel genes under AS regulation which underpin important biological processes can be identified and provide new targets for medical intervention or crop improvement.

3D RNA-seq unlocks the discovery potential for RNA-seq data

Finally, the speed of *3D RNA-seq* now makes it feasible for biologists to re-analyse existing or publicly available RNA-seq data to give improved differential expression analysis and novel AS information as demonstrated here for the mouse data. Datasets from different labs can now be quickly compared using the same parameters and without performing new experiments. In addition, with new transcriptome releases, it is feasible to re-analyse datasets and update results. *3D RNA-seq* provides a reproducible platform to standardize analyses and utilize new and existing data for improved resolution and interpretation. 3D RNA-seq can be used to analyse expression and AS of non-coding RNAs as long as they are annotated in the reference transcriptome [13]. In the future, we aim to adapt the pipeline for third-generation direct RNA sequencing as well as small RNA analysis.

Materials and methods

We have developed 3D RNA-seq App, an R package which provides a web-based shiny App for flexible and powerful differential expression and alternative splicing analysis of RNA-seq data. The program is designed for ease-of-use and can be run by biologists with minimal bioinformatics experience or by bioinformaticians with little exposure to RNA-seq analysis.

The whole analysis from transcript quantification to data pre-processing, differential gene and transcript expression, differential alternative splicing, differential transcript usage and isoform switching analysis can be performed in less than three days. The analysis is optimized with quality control plots and generates a report containing customizable publication-quality figures and data tables to aid interpretation of results.

3D RNA-seq enables routine analysis of differential alternative splicing and isoform switching and can handle complex experimental designs. The speed and accuracy of the analysis have the potential to revolutionize what is achievable with RNA-seq technology.

Input files for 3D RNA-seq App

Three simple input files are required for *3D RNA-seq* analysis (Supplementary Figure S5). 1) A meta-data table in 'csv' (comma delimited) format containing the information of experimental design, including conditions/treatments, biological replicates, sequencing replicates (same biological sample divided and sequenced on different sequencing lanes) and quantification file names (Supplementary Figure S5A). 2) Transcript quantification outputs generated using *Salmon* [1] or *Kallisto* [2] (Supplementary Figure S5B). Transcript quantification is the first difficulty for most biologists who lack programming skills. A detailed user manual with demo video of quantification by using *Salmon/Kallisto* on the Galaxy interface [25] has been provided (see Availability of data and materials). 3) A transcript-gene association file consisting of a 'csv' spreadsheet with the first column listing transcript names and second column listing gene IDs. This relates transcript names to gene IDs which allows expression analysis at both gene and transcript level and the AS analysis within each gene (Supplementary Figure S5C). It is also possible to extract this info from other files when that information is available (Supplementary Figure S5D and E). All the inputs can be easily uploaded by selecting corresponding files from a local computer.

Data pre-processing with optimal quality control

First, *Salmon* and *Kallisto* outputs ('quant.sf' from *Salmon* and 'abundance.tsv' from *Kallisto*) for each sample, will be imported into *3D RNA-seq* and read counts will be generated using the tximport R package [26], which takes into account the transcript length and library size. Sequencing replicates, if present, will be merged to increase sequencing depth. Then, the data is pre-processed in various steps to reduce noise (e.g. removal of low expressed genes and transcripts) and technical variance (e.g. batch effects). In each step, interactive visualization plots are produced to show the immediate impact of the parameter change and facilitate the optimization of parameters for pre-processing. 1) Removal of low expressed transcripts. RNA-seq read counts follow approximately the negative binomial distribution. The variance of $\log_2$ transformed read counts decreases monotonically with the increase of mean (Supplementary Figure S2) [7,27,28]. However, low expressed transcripts do not follow such a distribution and thus should be removed. On the basis that an expressed transcript should have a minimum count per million (CPM), n, in at least m samples, the cut-off can be adjusted by the user and the effect on mean-variance trends visualized immediately. 2) Similarities and differences between samples are visualized by Principal Component Analysis (PCA) plots (Supplementary Figure S3). PCA is a method to project data variance of thousands of variables (transcripts/genes) to a few principal components (PC) that capture the majority of the data variability. PCA plots can reveal major variances in an experiment in an unbiased manner which allows technical effects (e.g. batch effects among biological replicates caused by different experimental conditions) to be identified. 4) Correction of technical effects if required. Batch effects, if detected, can be corrected by different methods such as the RUVSeq R package, which estimates a co-efficient for batch effect term which can be incorporated into the design matrix with the main factors in linear regression models for 3D analysis (Supplementary Figure S3B) [15]. 5) Normalization of read counts across the libraries. Read counts can be normalized using Trimmed Mean of M-values (TMM), Relative Log Expression (RLE) or upper-quartile method [29]. Read count distributions before and after normalization can be visualized in the plots.

Principle of 3D analysis: identification of DE, DAS and DTU genes and transcripts

RNA-seq experimental designs often involve a single, or multiple experimental conditions that may affect gene expression. The *3D RNA-seq* App provides a flexible way to set up contrast groups where users can select any samples or groups of samples of interest for comparison. Thus, it allows analyses of both simple pair-wise comparisons and complex experimental designs such as time-series, developmental series and multiple conditions (Supplementary Figure S4). For each contrast group, different statistics are defined for robust DE/DAS/DTU (3D) predictions (Fig. 1A-C). 1) For differential expression, $\log_2$ fold change (L_2FC) is the difference of $\log_2$-CPM values in contrast groups; 2) for differential alternative splicing, Δ percent spliced (ΔPS) is the difference of *PS* values which are defined as the ratios of transcript average abundances divided by the average gene abundances; and 3) *p*-values of multiple comparisons are adjusted to control the false discovery rate (FDR) [30, 31].

Stringency of the analysis can be modified by changing the cut-off settings. A gene/transcript is identified as DE in a contrast group if L_2FC of expression is greater than or equal to an established cut-off value (e.g. $L_2FC \geq 1$) and with adjusted *p*-value less than a cut-off (e.g. *p*< 0.01 or 0.05) (Fig. 1A). At the AS level, gene expression is compared to transcript expression in the contrast groups [8]. To identify DAS genes, the expression of each transcript is compared to the weighted average expression of all the transcripts for the same gene (weight on transcript expression variance). The *p*-value of each test is converted to gene-wise *p*-value by using the F-test or Simes method [8] (Fig. 1B). To identify DTU transcripts, each transcript is compared to the weighted average of all the other transcripts of the same gene (Fig. 1C). A gene is DAS in a contrast group if the adjusted *p*-value is less than an established cut-off value and at least one transcript of the gene has a ΔPS greater than an established cut-off value (e.g. ΔPS >0.1) (Fig. 1B). A transcript is DTU if the adjusted *p*-value and absolute values of ΔPS is greater than the selected cut-off values, respectively (Fig. 1C). *3D RNA-seq* thereby identifies significant DE genes and transcripts, DAS genes and DTU transcripts and these results

are saved in summary figures (see Fig. 3A,B), and tables of identified transcript/gene lists in csv files.

3D RNA-seq is flexible and can work with all experimental designs

Within 3D RNA-seq, the contrast groups (comparisons) can be set-up in multiple ways: 1) between any two samples with biological replicates not just limited to the comparison to control samples only; 2) between groups of samples, where the samples can be grouped at a higher level (e.g. group samples by gender); 3) contrast groups can be set up to test the interactions of different factors (e.g. the interactions between gender and treatment. Do different genders respond to the treatment differently?); 4) Besides discrete comparisons of samples and groups, 3D RNA-seq also allows the comparison of trend changes over time in each parallel condition group by using the spline method [31] to study dynamic activities and changes of transcriptome over time and over several conditions simultaneously. For example, the identification of the dynamics of stress-responsive genes and AS events in plants [12,13,32,33], timing mechanisms of circadian clocks [34, 35, 36] and drug responses [37,38]. Adjusted p-values are used to report the genes and transcripts that show significant differences between samples/groups/series at both levels of transcription and AS (DE, DAS and DTU).

Transcript isoform switch analysis

Transcript isoform switches (ISs) are a prominent type of DAS within a gene where a pair of alternatively spliced isoforms reverse the order of their relative expression abundances in response to stimulus [6]. In the *3D RNA-seq* analysis pipeline, the iso-kTSP method is introduced to study the ISs between pair-wise conditions in the user-defined contrast groups [5] while the Time-Series Isoform Switch (TSIS) method is used to identify the ISs between any consecutive time-points in time-series or developmental data [6].

Outputs and visualization

3D RNA-seq allows users to save results of significant DE/DAS/DTU gene/transcript lists, intermediate data of the whole analysis and publication-quality plots to a local folder by clicking 'action' buttons at the various steps in the App. Four folders are created to save 3D analysis outputs, 'result', 'figure', 'report' and 'data'. The gene and transcript expression in read counts and TPMs, testing statistics and analysis results of 3D analysis will be saved as '.csv' files in the 'result' folder. All the figures of data pre-processing, 3D analysis and downstream visualization are saved as 'png' and 'pdf' formats with user specified width, height and resolution in the 'figure' folder. The intermediate variables generated during 3D analysis are saved as '.RData' R objects in the 'data' folder for R users to carry out further analysis if required. In the last step of 3D analysis, a report in three formats, 'html', 'pdf' and 'word', will be generated in the 'report' folder using *R Markdown* [39]. The report includes sections of 'Methods', 'Results', 'Supplementary figures', 'Supplementary material' and 'References', in which all the parameters selected by the users for each step are recorded. The publication-quality figures and reports provide all information required for publications and it also provides sufficient information to reproduce the same analysis and results.

Genes and transcripts that are significantly changed between contrast groups can be visualized within the App in Venn diagrams, histograms and volcano plots (e.g. the number of up- and down-regulated DE genes and transcripts, isoform switches etc.) (see Fig. 3C-E). Heat-maps are used to visualize co-expression clusters of significant 3D genes and transcripts and investigate their expression pattern changes across conditions (Fig. 4A,B). Individual gene and transcript profiles and ΔPS plots can be generated to give users intuitive visualization of those with significant abundance and AS changes, thereby allowing a convenient and detailed investigation of individual genes/transcripts and providing a means of selecting candidate genes and transcripts for experimental validation and future research. The lists of genes and transcripts with significant changes in DE/DAS/DTU can be downloaded and exported out of 3D RNA-seq for gene ontology (GO) enrichment analysis using appropriate web databases or GO analysis programs. The significantly enriched annotation results can be re-imported into the *3D RNA-seq* App to generate GO annotation plots (Supplementary Figure S6 and see Fig. 4C,D). To generate publication-quality figures, users can set widths, heights, resolution and colours for histograms, plots, heat-maps etc. to be saved. The analysis reports contain full details of the custom-configured parameters, methods and results tables and figures from the entire 3D analysis. The analysis report is a detailed document that guarantees the reproducibility of the analysis and results.

3D RNA-seq adopts the best practice and integrates the state-of-art methods for DE and DAS analysis

3D RNA-seq adopts best practice and integrates state-of-art methods for RNA-seq data pre-processing: 1) Tximport to convert TPM values to read counts while taking transcript length and sequencing depth into account [26]; 2) mean-variance trend plots to filter low expressed genes/transcripts to improve fit to statistical models and remove discrepancies due to read count distribution assumptions [7]; 3) PCA plots to visualize sample variances and identify technical effects; 4) RUVSeq to estimate batch effects [15]; 5) a number of read count normalization methods to correct the variances and reduce the false positives for highly abundant transcript outliers [29].

Limma voom [7] was chosen as the engine for the DE analysis (DE, DAS and DTU) for five reasons. Firstly, from different studies, *Limma* is consistently one of the best performing methods for RNA-seq differential expression analysis and has a good control of FDR [10,40,41]. Secondly, it allows both the DE and DAS analysis within the same framework. More importantly, the DE analysis and DAS analysis differentiates genes that are affected by transcriptional regulation (e.g. by transcriptional factors) and alternative splicing regulation (e.g. by splicing factors) as shown in Fig. 3–7. Thirdly, *Limma* employs a linear model that runs very quickly and compared with methods requiring bootstrapping data, such as Sleuth [10], the running time and memory required is significantly reduced. Fourthly, *Limma* has good compatibility with other bioinformatics

R packages for a range of other functions, such as estimation of batch effects and normalization of sequencing biases. In addition to batch effects that occur systematically for all the samples in a certain biological replicate, RNA-seq data may have variations in sample quality due to, for example, degradation or contamination of specific samples. These problematic samples are often shown as outliers in the PCA plot. The *Limma-voomWithQualityWeights* function has been incorporated in the 3D RNA-seq App to balance the outliners [7,42]. Lastly, but most importantly, *Limma* allows flexible experimental designs, where any pairs or groups of samples can be compared in contrast to most of the current DE tools which offer limited choices for comparisons.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Funding

Data availability statement

The *3D RNA-seq* web interface is available at https://3drnaseq.hutton.ac. uk. The R package version (ThreeDRNAseq) is available on Github at https://github.com/wyguo/ThreeDRNAseq.

Manuals for both versions and transcript quantification on Galaxy interface can be accessed from https://github.com/wyguo/ThreeDRNAseq/tree/master/vignettes/user_manuals

Tutorial and demo video can be viewed from https://www.youtube.com/watch?v=rqeXECX1-T4

The *Kallisto* transcript quantifications from the dexamethasone treatment on mice were downloaded from:

https://figshare.com/articles/kallisto_quantifications_of_Frahm_et_al_2017/6203012.

The Sleuth/aggregated *p*-values pipeline is at:

https://pachterlab.github.io/sleuth_walkthroughs/pval_agg/analysis.html.

Authors' contributions

RZ and JB supervised all aspects of the work; WG, RZ, JB designed the 3D analysis pipeline and WG coded the whole 3D RNA-seq App; NT and CC collected and prepared the Arabidopsis cold response RNA-seq data; GS and IM set up and maintained the web service; WG, RZ JB and RW drafted the manuscript; All authors engaged in discussions and testing of the 3D RNA-seq App; All authors were involved in revision and improvement of the final manuscripts; RZ, JB and RW interpreted the main findings.

Availability of the 3D RNA-seq App

The *3D RNA-seq* App is available as a web service and an R package. 1) A publicly accessible version of *3D RNA-seq* is running at the James Hutton Institute and can be viewed at: https://3drnaseq.hutton.ac.uk/. The user can upload input files and carry out the entire analysis within a browser. The intermediate data and results of 3D analysis can be downloaded during the final step. This web interface allows biologists to use *3D RNA-seq* without installing it. The *3D RNA-seq* manual and a YouTube tutorial (https://www.youtube.com/watch?v=rqeXECX1-T4) to demonstrate the usage of *3D* RNA-seq and upstream quantification step using *Salmon*can be found at the same web address; 2) The R package *ThreeDRNAseq* can be installed locally and users can run the *3D RNA-seq* App by typing a command 'run3DApp()' through RStudio on a local PC. This version is suitable for people with a good knowledge of R and who prefer to run the analysis locally. The user manuals for both using 3D RNA-seq App and command lines of ThreeDRNAseq R package for 3D analysis is provided at:

https://github.com/wyguo/ThreeDRNAseq/tree/master/vignettes/user_manuals

ORCID

Wenbin Guo ⬤ http://orcid.org/0000-0002-1829-6044
Iain Milne ⬤ http://orcid.org/0000-0002-4126-0859
Cristiane PG Calixto ⬤ http://orcid.org/0000-0002-9935-9325
Runxuan Zhang ⬤ http://orcid.org/0000-0001-7558-765X

References

[1] Patro R, Duggal G, Love MI, et al. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417–419.

[2] Bray NL, Pimentel H, Melsted P, et al. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525–527.

[3] Trapnell C, Pachter l, Salzburg SL.TopHat: discovering splice junctions with RNA-seq Bioinformatics. 2009;25(9):1105–1111.

[4] Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–515.

[5] Sebestyén E, Zawisza M, Eyras E. Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. Nucleic Acids Res. 2015;43(3):1345–1356.

[6] Guo W, Calixto CPG, Brown JWS, et al. TSIS: an R package to infer alternative splicing isoform switches for time-series data. Bioinformatics. 2017;33(20):3308–3310.

[7] Law CW, Chen Y, Shi W, et al. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29.

[8] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

[9] Frahm KA, Waldman JK, Luthra S, et al. A comparison of the sexually dimorphic dexamethasone transcriptome in mouse cerebral cortical and hypothalamic embryonic neural stem cells. Mol Cell Endocrinol. 2018;471:42–50.

[10] Pimentel H, Bray NL, Puente S, et al. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14(7):687–690.

[11] Yi L, Pimentel H, Bray NL, et al. Gene-level differential analysis at transcript-level resolution. Genome Biol. 2018;19(1):53.

[12] Calixto CPG, Guo W, James AB, et al. Rapid and dynamic alternative splicing impacts the arabidopsis cold response transcriptome. Plant Cell. 2018;30(7):1424–1444.

[13] Calixto CPG, Tzioutziou NA, James AB, et al. Cold-dependent expression and alternative splicing of arabidopsis long non-coding RNAs. Front Plant Sci. 2019;10:235.

[14] Zhang R, Calixto CPG, Marquez Y, et al. A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. Nucleic Acids Res. 2017;45(9):5061–5073.

[15] Risso D, Ngai J, Speed TP, et al. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol. 2014;32(9):896–902.

[16] Fresno C, Fernández EA. RDAVIDWebService: A versatile R interface to DAVID. Bioinformatics. 2013;29(21):2810–2811.

[17] Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R package version 2.38.1. 2019. Available at: https://bioconductor.org/packages/release/bioc/html/topGO.html

[18] Carlson M. org.Mm.eg.db: genome wide annotation for Mouse. R package version 3.4.0. Bioconductor;2019. DOI:10.18129/B9.bioc.org.Mm.eg.db

[19] KEGG PATHWAY Database. n.d. [cited 2020 Jan 6]. Available from: https://www.genome.jp/kegg/pathway.html

[20] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B (Methodological). 1995;57(1):289–300.

[21] Brown JWS, Calixto CPG, Zhang R. High-quality reference transcript datasets hold the key to transcript-specific RNA-sequencing analysis in plants. New Phytologist. 2017;213(2):525–530.

[22] Zhang R, Calixto CPG, Tzioutziou NA, et al. AtRTD - a comprehensive reference transcript dataset resource for accurate quantification of transcript-specific expression in Arabidopsis thaliana. New Phytol. 2015;208(1):96–101.

[23] Alamancos GP, Pagès A, Trincado JL, et al. Leveraging transcript quantification for fast computation of alternative splicing profiles. Rna. 2015;21(9):1521–1531.

[24] Soneson C, Love MI, Patro R, et al. A junction coverage compatibility score to quantify the reliability of transcript abundance estimates and annotation catalogs. Life Sci Alliance. 2019;2(1):e201800175.

[25] Galaxy | Europe. n.d. [cited 2020 Jan 6]. Available from: https://usegalaxy.eu/

[26] Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2016;4:1521.

[27] Benaroya H, Mi Han S. Probability models in engineering and science. Mech Eng. 2005;48(4):740. .

[28] McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40(10):4288–4297.

[29] Bullard JH, Purdom E, Hansen KD, et al. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010;11(1):94.

[30] Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 2001;29(4):1165–1188.

[31] Hastie TJ, Chambers JM. Generalized additive models. Chapter 7 of Statistical Models in Seds. In: Monographs on statistics and applied probability. New York: Wadsworth & Brooks/Cole; 1992.

[32] Bahieldin A, Atef A, Sabir JSM, et al. RNA-Seq analysis of the wild barley (H. spontaneum) leaf transcriptome under salt stress. Comptes Rendus - Biologies. 2015;338(5):285–297.

[33] Ma L, Coulter JA, Liu L, et al. Transcriptome analysis reveals key cold-stress-responsive genes in winter rapeseed (Brassica rapa L. Int J Mol Sci. 2019;20(5):1071.

[34] Blair EJ, Bonnot T, Hummel M, et al. Contribution of time of day and the circadian clock to the heat stress responsive transcriptome in Arabidopsis. Sci Rep. 2019;9(1). DOI:10.1038/s41598-019-41234-w

[35] Kim JA, Shim D, Kumari S, et al. Transcriptome analysis of diurnal gene expression in Chinese cabbage. Genes (Basel). 2019;10(2):130.

[36] Oh S, Song S, Dasgupta N, et al. The analytical landscape of static and temporal dynamics in transcriptome data. Front Genet. 2014;5(FEB). DOI:10.3389/fgene.2014.00035

[37] Jardim-Perassi BV, Alexandre PA, Sonehara NM, et al. RNA-Seq transcriptome analysis shows anti-tumor actions of melatonin in a breast cancer xenograft model. Sci Rep. 2019;9(1). DOI:10.1038/s41598-018-37413-w

[38] Kang W, Jia Z, Tang D, et al. Time-course transcriptome analysis for drug repositioning in fusobacterium nucleatum-infected human gingival fibroblasts. Front Cell Dev Biol. 2019;7. DOI:10.3389/fcell.2019.00204

[39] Baumer B, Udwin D. R Markdown. Wiley Interdiscip Rev Comput Stat. 2015;7(3):167–177.

[40] Rapaport F, Khanin R, Liang Y, et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol. 2013;14(9):R95.

[41] Tang M, Sun J, Shimizu K, et al. Evaluation of methods for differential expression analysis on multi-group RNA-seq count data. BMC Bioinformatics. 2015;16(1):360.

[42] Liu R, Holik AZ, Su S, et al. Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. Nucleic Acids Res. 2015;43(15):e97-e97.