Fast and accurate differential splicing analysis across multiple conditions with replicates

Juan C. Entizne[1,*], Juan L. Trincado[1,*], Gerald Hysenaj[2], Babita Singh[1], Miha Skalic[1], David J. Elliott[2], Eduardo Eyras[1,3]

[1]Pompeu Fabra University, E08003, Barcelona, Spain

[2]Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle NE1 3BZ, UK.

[3]ICREA, E08010, Barcelona, Spain

* equal contribution

corresponding author: eduardo.eyras@upf.edu

Multiple approaches have been proposed to study differential splicing from RNA sequencing (RNA-seq) data[1], including the analysis of transcript isoforms[2,3], clusters of splice-junctions[4,5], alternative splicing events[6–8] and exonic regions[9]. However, many challenges remain unsolved, including the limitation in speed, the computing capacity and storage requirements, the constraints in the number of reads needed to achieve sufficient accuracy, and the lack of robust methods to account for variability between replicates and for analyses across multiple conditions. We present here a significant extension of SUPPA[8] to enable streamlined analysis of differential splicing across multiple conditions, taking into account biological variability. We show that SUPPA differential splicing analysis achieves high accuracy using extensive experimental and simulated data compared to other methods; and shows higher accuracy at low sequencing depth, with short read lengths, and using replicas with unbalanced depth, which has important implications for the cost-effective use of RNA-seq data for splicing analysis. We also validate the analysis of multiple conditions with SUPPA by studying differential splicing during iPS-cell to neuron differentiation and during erythroblast differentiation, providing support for the applicability of SUPPA for the robust analysis of differential splicing beyond binary comparisons.

Statistical significance in differential splicing is usually assessed using read counts. However, alternative splicing produces a change in the relative abundance of transcript isoforms from the same gene. This can be readily described in terms of differences in the transcript relative abundances estimated from RNA-seq and has shown a good agreement with experimental measurements[8,10,11]. We showed before that transcriptome quantification can be exploited for the fast estimate of event inclusion values with high accuracy[8]. We have now performed a significant extension of SUPPA to exploit the variability between replicates to determine the uncertainty of the estimate of the change in inclusion level, or ΔPSI, in terms of the abundance of the transcripts describing the event (see Methods). To evaluate the dynamic range of SUPPA we used it to analyze RNA-seq data after the knockdown of TRA2A and TRA2B splicing regulators compared to controls[12] (Fig. 1a). In agreement with the expected higher variability at low read count, there is greater uncertainty for the observed ΔPSI between conditions at low transcript expression. SUPPA estimates the significance considering the distribution between replicates for all events with similar transcript abundance (Fig. 1b), resulting in a lower bound in |ΔPSI| values for significant events that vary with the expression of the transcripts describing the event.

We compared SUPPA results with three other methods that calculate differential splicing using multiple replicates per condition: rMATS[7] and MAJIQ[5], which describe changes in terms of ΔPSI, and DEXSeq[9], which uses fold-changes. Importantly, we found that SUPPA is much faster than the other methods, devoting 24 seconds to PSI quantification and about 32 minutes and 47 seconds for differential splicing analysis on the same datasets (Fig. 1c). Since SUPPA performs the significance test directly on the ΔPSI values without needing to go back to the read data, it hence provides unmatched speed for differential splicing analysis. Comparing the results obtained with each method (Supp. Fig. 1), rMATS and DEXSeq detect many significant events with small inclusion changes that are not distinguishable from the variability between biological replicates, whereas MAJIQ does not detect as many significant events at low and high gene expression. In contrast, as SUPPA exploits the between-replicate variability to test for significance, it avoids the use of arbitrary global |ΔPSI| thresholds and detects significant events across a wide range of gene expression values (Supp. Fig. 1).

To test the accuracy of SUPPA, we simulated 317 differentially spliced cassette events with |ΔPSI|>0.2 between two conditions with 3 replicates each, and considered 317 different cassette events without change as a negative set (see Methods) (Supp. Table S1). We

simulated genome-wide RNA-seq reads using RSEM[13] at different sequencing depths: 120, 60, 25, 10 and 5 millions of paired-end 100nt reads per sample, and for different read-lengths: 100, 75, 50 and 25nt with 25M paired-end reads. Despite the differences in the numbers and length of the reads (Supp. Table S2), the benchmarking gene sets show similar distributions of gene expression at all depths and read lengths (Supp. Fig. 2). We then calculated differentially spliced events with SUPPA, rMATS, MAJIQ and DEXSeq, and evaluated the number of simulated cassette events that each method was able to measure, as well as the number of true positives and false positives. SUPPA was able to measure more events than the other methods at low sequencing depths (Fig. 1d and Supp. Fig. 2) and at short read lengths (Fig. 1e), and recovered more true positives for the same simulations (Figs. 1f and 1g). This result was the same regardless of whether the true positive was considered only based on the significance test, or imposing in addition the cutoff $|\Delta PSI|>0.2$ for the predictions (Supp. Table S3). Additionally, when considering an unbalanced configuration where one replicate had 120M reads and the other two replicates had 10M reads, SUPPA showed higher number of true positives (Fig. 1h). SUPPA also showed high correlation values between predicted and simulated ΔPSI values (Supp. Figs. 2c-2e) (Supp. Table S3) in the same simulated conditions. In contrast, the other methods had reduced performance at low depth and short read length, and showed a reduction of true positives for unbalanced sequencing depths (Fig. 1h), probably owing to them relying on having sufficient and balanced junction and/or exonic reads. SUPPA can also measure a higher number of events from the negative set (Supp. Figs. 2f and 2g) while maintaining a low proportion of false positives (Supp. Table S3).

To further evaluate the accuracy of SUPPA we used 83 experimentally validated events that change upon TRA2A and TRA2B knockdown compared to control cells[12]. SUPPA had similar accuracy to rMATS and MAJIQ in the estimation of ΔPSI values (Fig. 2a) (Supp. Tables S4-S5). Furthermore, using three other independent RT-PCR datasets published previously[10,14,15] SUPPA showed higher or similar accuracy compared to rMATS and MAJIQ (Supp. Figs 3a-3c) (Supp. Tables S6-S11). In particular, SUPPA had the highest accuracy for the data from Pimentel et al.[10], which has replicas with unbalanced depths. Finally, using 44 RT-PCR negative cassette events that did not show significant change upon the double knockdown of TRA2A and TRA2B, SUPPA had false positive rate similar to the other methods, either before (Fig. 2b) or after applying a $|\Delta PSI|>0.1$ cut-off (Supp. Fig. 3d) (Supp. Tables S12-S13).

We further performed a direct comparison of SUPPA, rMATS, MAJIQ and DEXSeq. Since exon cassette events are the most frequent form of splicing variation (48% of events from the human Ensembl annotation are of cassette type) we decided to match the identified cassette exons across all four methods (see Methods). We were able to identify 7116 cassette exons unambiguously defined by all methods (Supp. Fig. 4a) (Supp. Table S14). From the 793 events detected as significant by at least one method, only 32 were identified as significant by all methods (Supp. Fig 4b); and events detected by more methods tended to have higher ΔPSI values (Supp. Fig. 4c) and covered a smaller range of gene expression (Supp. Fig. 4d). Despite this low overlap, significant events from each method independently showed enrichment of TRA2B CLIP tags[12] (Fisher's exact test p-value < 0.01) and TRA2 binding motifs (z-score > 3). Hence, each method provides a reasonable description of the splicing patterns independently. To seek further support for this point, we selected for experimental validation 14 events that were predicted as significant only by SUPPA and further showed overlap with TRA2 motifs and TRA2B CLIP tags[12] (Supp. Table S15). However, from these 14 events, 7 could not be measured by RT-PCR due to low expression. From the other 7 cases tested, 5 changed significantly and all cases except one changed in the direction predicted by SUPPA (Fig. 2c). In particular, we validated a new event in *EML4* (Fig. 2d), a gene involved in cancer through a fusion with *ALK* that is not present in MDA-MB-231 cells[16]. We conclude that SUPPA is able to detect overlapping splicing changes common to other methods, but additionally can detect experimentally reproducible events that are not identified by any other method.

SUPPA is also able to analyze multiple conditions by computing the pairwise differential splicing between conditions, and detect groups of events with similar splicing patterns across conditions using density-based clustering. Clustering requires two parameters: the minimum number of events in a cluster (N), which can be interpreted as the expected size of the regulatory modules; and the maximum distance to cluster two events (D), which approximates the maximum average PSI difference at which two events are grouped together. To evaluate the ability of SUPPA to cluster events, we analyzed a 4-day time-course of differentiation of human iPS cells into bipolar neurons[17], which had not been analyzed yet for alternative splicing. In agreement with the observed enrichment of microexons (length < 28nt) in differentiated neurons[18,19], SUPPA identified 2314 regulated cassette events enriched in microexons compared to non-regulated cassette events (Fisher's exact test p-value < 2.2e-16, odds-ratio= 4.33961), with the majority of them (78%) more included in differentiated cells (ΔPSI > 0.1 between the first and fourth day). After evaluating various clustering parameters (Supp. Fig. 5a), using the parameters

D=0.08 and N=11 yielded 3 well-differentiated clusters (silhouette score = 0.63) (Figs. 3a-c). Clusters 0 and 1 increase inclusion at day 2 and 1, respectively, whereas cluster 2 decreases inclusion at day 1. Interestingly, clusters 0 and 1 are enriched in microexons with respect to unclustered regulated cassette events (Fisher's exact test p-value = 6.379e-05, odds-ratio = 4.708), whereas cluster 2 includes none. To further validate the findings with SUPPA, we performed a motif enrichment analysis in regulated events compared to non-regulated events. Notably, although the 2314 regulated events did not show any enrichment in motifs for RNA binding proteins (RBPs), those events in clusters were enriched in, among others, CELF, RBFOX and SRRM4 motifs. Consistent with this, *SRRM4* and members of the *CELF* and *RBFOX* families also show upregulation at the initial steps of iPS cell differentiation into neurons (Fig. 3d-f) (Supp. Fig. 5) (Supp. Table 16). These results agree with the described role of *CELF*, *RBFOX* and *SRRM4* genes in neuronal differentiation[19–22]. Additionally, they suggest for the first time a role of *ESRP1, ESRP2, RBMS3* and other factors in differential splicing during neuronal differentiation.

We also used SUPPA to analyze differential splicing across 5 steps of erythroblast differentiation[10]. After evaluating different parameters for clustering (Supp. Fig. 6), N=30 and D=0.07 yielded two homogeneous and well-differentiated clusters, one with high PSI values that decrease at the last step, and a second cluster with low inclusion with an increase at the last step (Supp. Fig. 6). In agreement with previous results[23], the cluster of events that increase inclusion at the late differentiation stage is enriched in intron retention events, as compared with the other cluster (Fisher's exact test p-value = 0.0009826, odds-ratio = 7.056), or by comparing with events not included in clusters (Fisher's exact test p-value = 0.01, odds-ratio = 1.769) (Supp. Fig. 6). We conclude that SUPPA provides a powerful approach to analyze splicing across multiple conditions, validated not only by intrinsic measures of clustering consistency, but also by recovering known biological results and new features.

In summary, our extensive evaluations here indicate that the new version of SUPPA provides a broadly applicable solution to current challenges in the analysis of differential splicing from RNA sequencing data across multiple conditions that will make it attractive to many potential users. SUPPA is faster than other methods, and maintains a high accuracy, especially at low sequencing depth and for short read-length. This provides the opportunity to study splicing in projects with limited budget, and to reuse available sequencing datasets with lower depth than usually required by other methods. Additionally, the low computing and storage requirements of

SUPPA makes possible to perform fast differential splicing processing and clustering analysis on a laptop. Thus, coupled with fast methods for transcript quantification[24–26], SUPPA facilitates the study of alternative splicing across multiple conditions without the need of large computational resources. The simplicity and modular architecture of SUPPA also makes it a very convenient tool in multiple contexts, as PSI values from other methods can be used in SUPPA for differential splicing analysis or for clustering across conditions.

The observed discrepancies between methods can be explained by various factors: the different ways in which a splicing change is represented, how changes in splicing inclusion levels are tested, and how biological variability affects these estimates. Most methods have the power to identify small significant ΔPSI values, but a fraction of these may be indistinguishable from the expected variability across replicates. Statistical tests that can separate significant splicing changes from the variability between-replicates thus provide an advantage to identify significant changes at low or high expression. By exploiting the biological variability without going back to the read data, SUPPA achieves unmatched speed and provides significance without the need of arbitrary ΔPSI thresholds.

Although SUPPA relies on genome annotation to define events, poorly annotated genomes can be improved and extended before analysis by SUPPA. In fact, recent analyses have shown that improved annotations lead to significantly better PSI estimates from RNA-seq compared to high-resolution RT-PCR measurements[27]. Current trends predict an increase in the number of efforts to improve the transcriptomes in multiple species and conditions. In this direction, SUPPA could play a key role for the systematic and rapid genome-wide analysis of splicing following annotation and sample updates. In conclusion, the speed, modularity and accuracy of SUPPA enable cost-effective use of RNA-seq for the robust and streamlined analysis of differential splicing across multiple biological conditions.

Acknowledgements

This work was supported by the MINECO and FEDER (BIO2014-52566-R) and AGAUR (SGR2014-1121) and Breast Cancer Now (2014NovPR355). GH is a BBSRC-funded PhD student. The authors thank C. Calixto, J. Brown, R. Zhang, M. Irimia and N. Barbosa-Morais for insightful discussions and to J. Vaquero-Garcia and Y. Barash for sharing the RT-PCR data from MAJIQ's paper.

Author Contributions

J.C.E, M.S. and EE designed and implemented the method and algorithms, J.C.E., J.L.T., and E.E. devised the analyses and J.C.E, J.L.T, and M.S. carried the benchmarking analyses. G.H and D.J.E. produced the datasets related to the double knockdown of TRA2A and TRA2B and performed the validation experiments. B.S. carried out the motif enrichment analyses. J.C.E, J.L.T. and EE wrote the manuscript with essential input from the other authors.

Methods

SUPPA Differential splicing

SUPPA considers two event distributions: one for the ΔPSI values between replicates, calculated as the difference of PSI values, as a function of the average abundance in the replicates of transcripts describing the event:

$$E_{rep} = \frac{1}{|R_c|} \sum_{r \in R_c} \log_{10} \left(\sum_{a} TPM_{a,r} \right)$$

where $r=1,..,|R_c|$ runs over the replicates in each condition $c=1,2$, and a indicates the two or more transcripts describing the event. A second distribution consists of the ΔPSI values between conditions, calculated as the difference of the means in the two conditions, calculated as the difference of the means in the two conditions, as a function of the average abundance of transcripts describing the event across both conditions:

$$E_{cond} = \frac{1}{2} \sum_{c=1,2} \frac{1}{R_c} \sum_{r \in R_c} \log_{10} (\sum_{a} TPM_{a,r,c})$$

Given the ΔPSI and E_{cond} values for an event, the significance is calculated from the comparison with the ΔPSI distribution between replicates for events with E_{rep} values similar to the observed E_{cond}. These are obtained by selecting the closest value E^*_{rep} from all points i from the between-replicate distribution:

$$E^*_{rep} = \min_i \{ | E_{i,rep} - E_{cond} | \}$$

using binary search and selecting a fixed number of events (1000 by default) around the E^*_{rep} value. The selected events define an empirical cumulative density function (ECDF) over |ΔPSI| from which a p-value is calculated:

$$p = (1 - ECDF(|\Delta PSI|)/2$$

Here we implicitly assume that the background distribution is symmetric. SUPPA includes an option to correct for multiple testing across all events from the same gene, as they cannot be considered to be entirely independent of each other, using the Benjamini-Hochberg method, where the false discovery rate (FDR) cut-off can be given as input.

SUPPA clustering analysis

SUPPA implements the density-based clustering algorithm DBSCAN[28] using the vectors of mean PSI values, and uses two parameters: the minimum number of events in a cluster (N), and the maximum distance to consider two events as cluster partners (D). Clustering is only performed with events that change significantly in at least one pair of adjacent conditions and three different distance metrics can be used: Euclidean, Manhattan and Cosine. Cluster qualities are reported using the silhouette score[29], which indicates how well the events are assigned to clusters; and the root mean square standard deviation (RMSSTD), which measures the homogeneity of each cluster. Additionally, the number and percentage of events in clusters is also reported. Motif enrichment analysis was performed as before[30]. Further details on the motif enrichment and analysis of differential expression are provided as Supplementary Material.

Simulated datasets

We used the quantification of the RefSeq transcripts for the 3 control samples from Best et al.[12] (GSE59335) as theoretical abundances, and considered 700 genes with only two isoforms containing any type of alternative splicing event (exon cassette, mutually exclusive, alternative 5'/3' splice-site, intron retention) and with absolute difference of relative abundance greater than 0.2:

$$\frac{|TPM_1 - TPM_2|}{TPM_1 + TPM_2} > 0.2$$

We simulated differential splicing by exchanging their theoretical TPM values in a second condition, keeping the same theoretical abundances for all other transcripts. For the benchmarking analysis we used only the 317 (45%) cases with cassette events, as these were the most abundant type and the easiest one to match across methods. Additionally, we considered a negative set of 317 cassette events in 317 genes with two isoforms sampled from the entire range of values of the difference of relative abundance. We used RSEM[13] to simulate

sequencing reads for the 2 conditions, 3 replicates each, at various depths: 120, 60, 25, 10 and 5 millions of 100nt paired-end reads per sample, and at various read lengths: 100nt, 75nt, 50nt and 25nt, at depth 25M paired-end reads (Supp. Tables S1-S3). Further details are given as Supplementary Material.

Experimental datasets

We analyzed RNA-seq data for the double knockdown of TRA2A and TRA2B in MDA-MB-231 cells and controls with 3 replicates per condition[12] (GSE59335), and used 83 RT-PCR validated events for comparison (Supp. Tables S4-S5). Additionally, we analysed RNA-seq from 5 steps of differentiating human erythroblasts[10] (GSE53635), with 3 replicates per condition. We further used 16 RT-PCR validated events between the initial (proerythroblast) and last (orthochromatic erythroblasts) steps[10,23] (Supp. Tables S6-S7). We also analyzed data from Cerebellum and Liver mouse tissues covering 8 different time points from 2 full circadian cycles[14] (GSE54651), and performed a comparison with the RT-PCR validated events used in Vaquero-Garcia *et al.*[5], comparing samples CT28, CT40 and CT52 in Cerebellum with the same circadian time points in Liver (Supp. Tables S8-S9). We also analyzed data for stimulated and unstimulated Jurkat T-cells[15] (SRP059357) for the comparison with RT-PCR validated events used in Vaquero-Garcia *et al.*[5] (Supp. Tables S10-S11). From the 54 RT-PCR validated events, we only used the 30 events that had experimental value $|\Delta PSI|>0.05$. We further used RNA-seq samples from a 4-day time-course, 3 replicates each, for the differentiation of human iPS cells into bipolar neurons[17] (GSE60548). Original data was for days 0,1,3,4 after initiation of differentiation, which we relabeled as days 0,1,2,3 for simplicity.

RNA-seq reads were used to quantify transcripts from Ensembl (version 75) and RefSeq (hg19/mm10) with Salmon[25]. Reads were also mapped to the human (hg19) or mouse (mm10) genomes using TopHat[31]. All methods other than SUPPA were used with these mappings. Cassette events from SUPPA and rMATS were matched to the RT-PCR validated events in each dataset, considering only those cases where the middle exon matched exactly the validated exons and confirming the flanking exons with the RT-PCR primers when available. Ambiguous matches were discarded from the comparison. For MAJIQ we selected the inclusion junction compatible with the validated event that had the largest posterior probability for $|\Delta PSI|>0.1$. For DEXSeq we considered whole exons that matched exactly with the middle exon of the experimentally validated cassette event. For the comparison with RT-PCR positives and negatives SUPPA was used with the RefSeq annotation. For all other analyses SUPPA was used with the Ensembl annotation. To select a set of cassette events common to all four methods, we selected the events measured by both SUPPA and rMATS such that the middle

exon matched exactly a DEXSeq exonic region and did not appear in more than one event from SUPPA or rMATS. From this set, we selected those for which any of the two inclusion junctions defined by SUPPA and rMATS was present in MAJIQ, and selected the junction with the largest posterior probability for |ΔPSI|>0.1. Further details are provided as Supplementary Material.

Time performance

Running time was measured using the Unix time command. For SUPPA running time was measured independently of the transcript quantification step. Similarly, for all other methods the running time did not include the read-mapping step. Time was measured independently for PSI calculation and for differential splicing analysis. All methods were run on a Unix machine with 12Gb of RAM and 8 Intel Xeon 2GHz CPU cores.

Experimental validation

Details on the experimental validation are given as Supplementary Material.

Software and datasets

SUPPA software for differential splicing analysis is available at
https://github.com/comprna/SUPPA
Commands and datasets used in this work are available at
https://github.com/comprna/SUPPA_supplementary_data

References

1. Alamancos, G. P., Agirre, E. & Eyras, E. *Methods to study splicing from high-throughput RNA sequencing data*. *Methods Mol. Biol.* **1126,** (2014).
2. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31,** 46–53 (2013).
3. Sebestyén, E., Zawisza, M. & Eyras, E. Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Res.* **43,** (2015).
4. Hu, Y. *et al.* DiffSplice: The genome-wide detection of differential splicing events with RNA-seq. *Nucleic Acids Res.* **41,** (2013).

5. Vaquero-Garcia, J. *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* **5,** e11752 (2016).
6. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7,** 1009–15 (2010).
7. Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E5593–601 (2014).
8. Alamancos, G. P., Pages, A., Trincado, J. L., Bellora, N. & Eyras, E. Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA* **21,** 1521–1531 (2015).
9. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22,** 2008–17 (2012).
10. Pimentel, H. *et al.* A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res.* **42,** 4031–42 (2014).
11. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338,** 1593–9 (2012).
12. Best, A. *et al.* Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons. *Nat. Commun.* **5,** 4760 (2014).
13. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
14. Zhang, R., Lahens, N. F., Ballance, H. I., Hughes, M. E. & Hogenesch, J. B. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 16219–24 (2014).
15. Cole, B. S. *et al.* Global analysis of physical and functional RNA targets of hnRNP L reveals distinct sequence and epigenetic features of repressed and enhanced exons. *RNA* **21,** 2053–66 (2015).
16. Lin, E. *et al.* Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. *Mol. Cancer Res.* **7,** 1466–76 (2009).
17. Busskamp, V. *et al.* Rapid neurogenesis through transcriptional activation in human stem cells. *Mol. Syst. Biol.* **10,** 760 (2014).
18. Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159,** 1511–23 (2014).
19. Li, Y. I., Sanchez-Pulido, L., Haerty, W. & Ponting, C. P. RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.*

25, 1–13 (2015).

20. Raj, B. *et al.* A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol. Cell* **56,** (2014).
21. Norris, A. D. *et al.* A pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. *Mol. Cell* **54,** 946–59 (2014).
22. Kim, K. K., Nam, J., Mukouyama, Y.-S. & Kawamoto, S. Rbfox3-regulated alternative splicing of Numb promotes neuronal differentiation during development. *J. Cell Biol.* **200,** 443–58 (2013).
23. Pimentel, H. *et al.* A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res.* **44,** 838–51 (2016).
24. Patro, R., Mount, S. M. & Kingsford, C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat. Biotechnol.* **32,** 462–4 (2014).
25. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. *Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference*. (2015). doi:10.1101/021592
26. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34,** 525–7 (2016).
27. Brown, J. W. S., Calixto, C. P. G. & Zhang, R. High-quality reference transcript datasets hold the key to transcript-specific RNA-sequencing analysis in plants. *New Phytol.* (2016). doi:10.1111/nph.14208
28. Ester, M., Kriegel, H. P., Sander, J. & Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. in *Proc. 2nd Int. Conf. Knowl. Discov. Data Min.* 226–231 (1996). doi:10.1.1.71.1980
29. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20,** 53–65 (1987).
30. Sebestyén, E. *et al.* Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* **26,** (2016).
31. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).

Figures

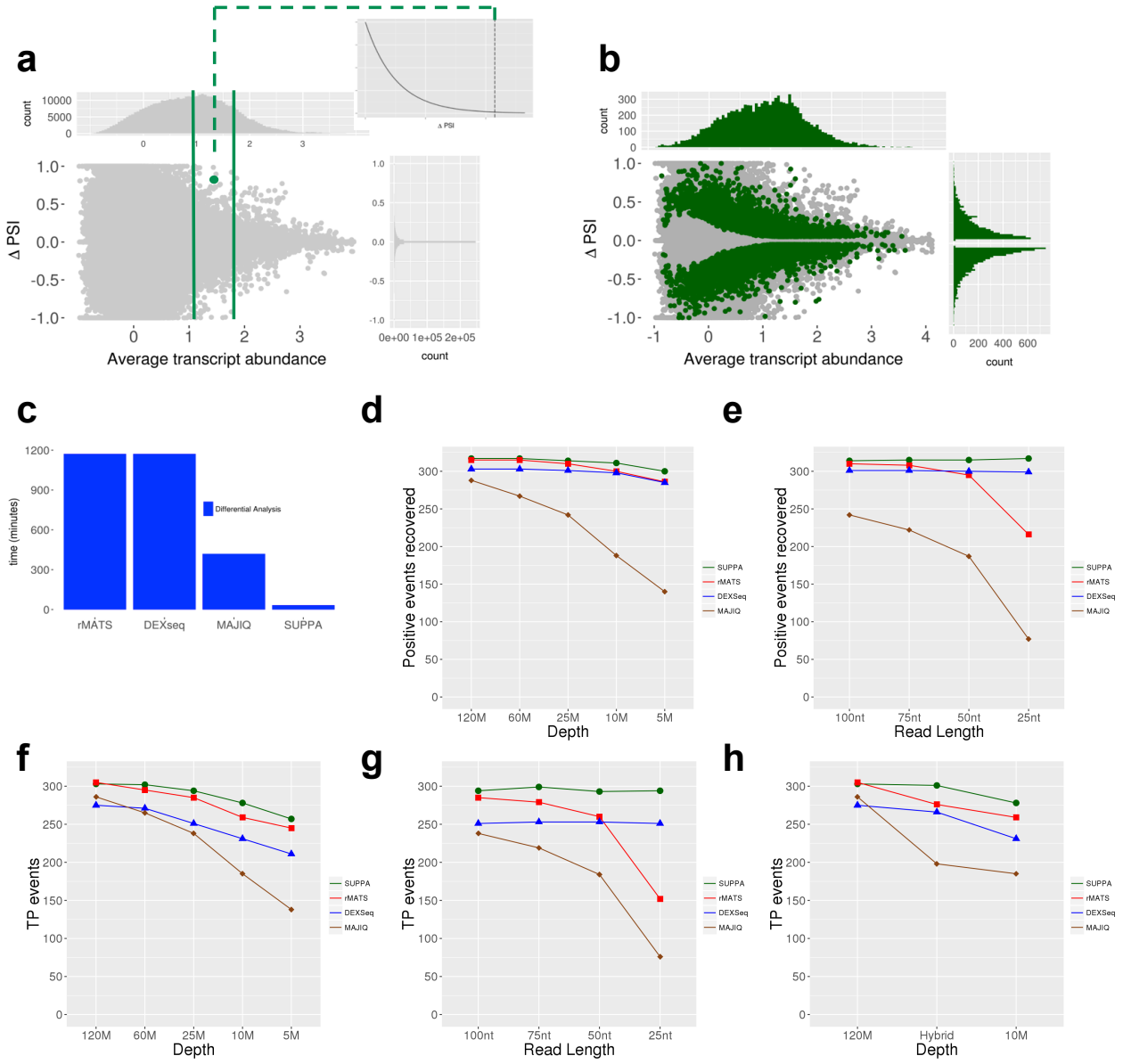

Figure 1. Overview of SUPPA differential splicing and benchmarking analysis**.** (**a**) The central panel displays the ΔPSI values between replicates (*y*-axis) as a function of the average transcript abundance (*x*-axis), using data from (Best et al. 2015) (Methods). The attached panels display the ΔPSI values along the x-axis (top panel) and along the y-axis (right panel). The green dot represents an example of ΔPSI observed between conditions and the top-right panel shows the between-replicate ΔPSI distribution against which the observed ΔPS is compared. **(b)** The central panel displays the ΔPSI values (y-axis) between conditions (green)

or between replicates (gray) as a function of the average transcript abundance (x-axis). Only events with p-value < 0.05 according to SUPPA are plotted in green. The attached panels display the distribution of the significant ΔPSI values along the x-axis (top panel) and along the y-axis (right panel). **(c)** Time performance of SUPPA compared to rMATS, MAJIQ and DEXSeq in the differential splicing analysis between two conditions, with 3 replicates each[12]. Time (*y*-axis) is given in minutes and in each case does not include the read mapping, transcript quantification steps or the calculation of PSI values. **(d)** Number of events measured by each method (*y*-axis) from the 317 positive simulated cassette events at different sequencing depths (*x*-axis). **(e)** As in (d) but for different read lengths (*x*-axis) at fixed depth (25M). **(f)** Number of true positive (TP) cases predicted as by each method (*y*-axis) at different sequencing depths (*x*-axis), calculated as the number of events with a statistically significant test according to each method: corrected p-value<0.05 for SUPPA, rMATS and DEXSEQ; and posterior(|ΔPSI|>0.1)>0.95 for MAJIQ. **(g)** As in (f) for different read lengths (*x*-axis) at fixed depth (25M). **(h)** Comparison of the true positives (TP) between the simulations with three replicas of 120M and 10M paired-end reads with an unbalanced simulation with one replica with 120M and the other two replicates with 10M paired-end reads (hybrid).

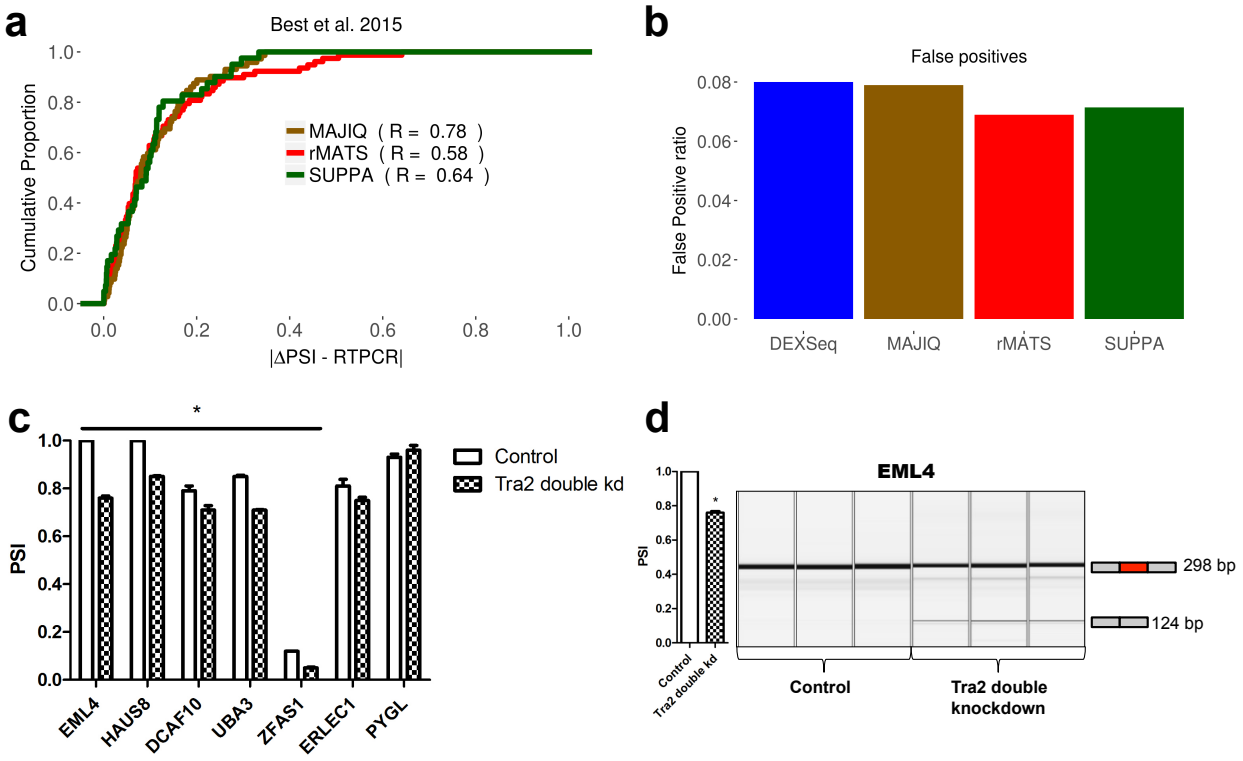

Figure 2. Experimental validation of differentially splicing predictions by SUPPA. **(a)** Comparison of predicted and experimentally validated ΔPSI values for 83 cassette events differentially spliced between the double knockdown of TRA2A and TRA2B and control in MDA-MB-231 cells. We show the cumulative proportion of cases (*y*-axis) according to the absolute difference between the predicted and the experimental value (|ΔPSI – RTPCR|). Additionally, we give for each method the Pearson correlation R between predicted and experimental values. **(b)** False positive rate (FPR) was calculated as the proportion of the measured RT-PCR negative events that were found significant: SUPPA (corrected p-value < 0.05), rMATS (FDR<0.05), MAJIQ (posterior > 0.95), DEXSeq (corrected p-value < 0.05). Out of the 44 negative events considered, each method measured 28 (SUPPA), 29 (rMATS), 38 (MAJIQ) and 25 (DEXSeq) events. We also show the FPR when an additional |ΔPSI|>0.1 cut-off is used. **(c)** Experimental validation by RT-PCR of a subset of novel events predicted as significant by SUPPA but not by rMATS, MAJIQ or DEXSeq between the double knockdown of TRA2A and TRA2B and controls. **(d)** Experimental validation of a new skipping event in EML4 upon knockdown of TRA2A and TRA2B (3 biological replicates shown in each case).

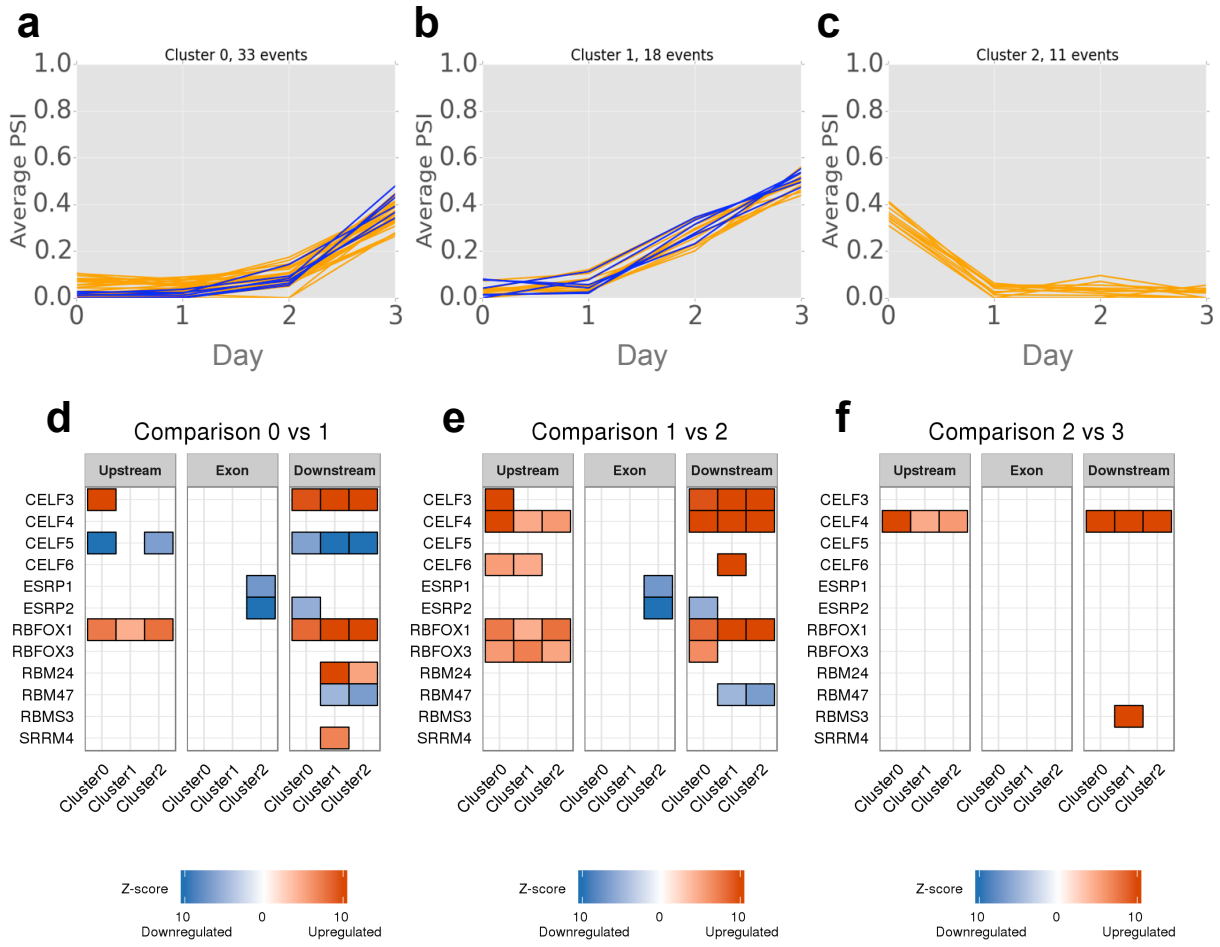

Figure 3. Prediction and clustering of differentially spliced events across neuronal differentiation. Density-based clustering performed on the 2314 cassette-exons events that change splicing significantly in at least one comparison between adjacent steps across 4 differentiation steps (days 0,1,2,3 after initiation of differentiation). Plots **(a-c)** show the average PSI per day of the events in the three clusters obtained using D=0.08 and N=11. Microexons are plotted in blue over the rest of the events in orange. **(d-f)** Motif enrichment associated to each of the 3 clusters in (a-c) in the regions upstream (200nt), exon, and downstream (200nt). Only enriched motifs associated to splicing factors that are differentially expressed are shown in each comparison between differentiation stages (days after differentiation 0,1,2,3). Splicing factors that are upregulated or downregulated are indicated in red and blue, respectively. The color intensity indicates the z-score of the motif enrichment in that particular region.