

DATABASE

Open Access



Cortexa: a comprehensive resource for studying gene expression and alternative splicing in the murine brain

Stephan Weißbach^{1,2}, Jonas Milkovits¹, Stefan Pastore^{2,3}, Martin Heine¹, Susanne Gerber^{2*} and Hristo Todorov^{2*}

*Correspondence:
sugerber@uni-mainz.de; hristo.todorov@uni-mainz.de

¹ Institute of Developmental Biology and Neurobiology (iDN), Johannes Gutenberg University Mainz, 55128 Mainz, Germany

² Institute of Human Genetics, University Medical Center, Johannes Gutenberg University Mainz, 55131 Mainz, Germany

³ Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg University Mainz, 55128 Mainz, Germany

Abstract

Background: Gene expression and alternative splicing are strictly regulated processes that shape brain development and determine the cellular identity of differentiated neural cell populations. Despite the availability of multiple valuable datasets, many functional implications, especially those related to alternative splicing, remain poorly understood. Moreover, neuroscientists working primarily experimentally often lack the bioinformatics expertise required to process alternative splicing data and produce meaningful and interpretable results. Notably, re-analyzing publicly available datasets and integrating them with in-house data can provide substantial novel insights. However, such analyses necessitate developing harmonized data handling and processing pipelines which in turn require considerable computational resources and in-depth bioinformatics expertise.

Results: Here, we present Cortexa—a comprehensive web portal that incorporates RNA-sequencing datasets from the mouse cerebral cortex (longitudinal or cell-specific) and the hippocampus. Cortexa facilitates understandable visualization of the expression and alternative splicing patterns of individual genes. Our platform provides SplicePCA—a tool that allows users to integrate their alternative splicing dataset and compare it to cell-specific or developmental neocortical splicing patterns. All standardized gene expression and alternative splicing datasets can be downloaded for further in-depth downstream analysis without the need for extensive preprocessing.

Conclusions: Cortexa provides a robust and readily available resource for unraveling the complexity of gene expression and alternative splicing regulatory processes in the mouse brain. The data portal is available at <https://cortexa-rna.com/>

Keywords: Alternative splicing, Gene expression, Cerebral cortex, Hippocampus, Embryonic brain development

Background

Transcriptional regulation plays a crucial role in the developing and adult mammalian brain [1–3]. Major changes during development can be observed particularly in gene expression [4, 5] and alternative splicing [6–9] patterns.



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Gene expression is the best-understood, and most extensively studied among these processes. It has been a major focus in the study of brain development [10], cell identity [11, 12], and disease-related alterations [13–15]. However, a common issue with gene expression analysis when comparing results from heterogeneous sources is the occurrence of batch effects that are unrelated to biological factors, due to differences in sequencing technology, experimental handling, and bioinformatic processing [16–20].

Alternative splicing (AS) of precursor mRNA is a fundamental process that enables the generation of various transcripts and subsequent proteins from the same gene. Therefore, it significantly increases the available transcriptomic variability [21]. AS is particularly important in the central nervous system, playing a vital role during cortical development [6, 21] and in the determination and maintenance of neuronal cell identity [22]. Moreover, AS is an important regulator of gene expression since it can initiate the degradation of mRNA by introducing premature stop codons which leads to nonsense-mediated decay [23].

Although several high-quality datasets focusing on AS in the murine brain are freely available, interpreting the different types of splicing events for individual genes remains challenging. Moreover, harnessing the potential of multiple datasets requires a harmonized processing strategy to ensure the comparability of results. A limitation of existing data portals, namely, the single cell atlas of the Allen Brain Institute [24], Neuron Subtype Transcriptome [25], or Brain RNA-Seq [26], is that they are focused on gene expression and do not offer the analysis of custom data in the context of cell-specific or developmental changes in alternative splicing.

Here, we introduce Cortexa—a novel data portal for accessing a variety of high-quality neocortical and hippocampal transcriptomic datasets, analyzed for gene expression and alternative splicing. Batch effects between different studies have been minimized using a standardized analysis pipeline (Figure 1a). We offer easily interpretable summaries and visualization of results that allow a broad range of scientists to explore the expression and AS patterns of individual genes. Additionally, we developed SplicePCA—a tool that performs a principal component analysis of splicing events for a selected gene set and enables the investigation of splicing patterns related to developmental changes or variations across cell types [6, 27–29]. All standardized datasets included in Cortexa are publicly available for download, allowing users to integrate them into their research easily.

Construction and content

Datasets

We analyzed publicly available *in vivo* paired-end RNA sequencing data of the mouse cerebral cortex and hippocampus as well as *in vivo* data from neural progenitor cells (NPCs) and neurons (Fig. 1c) with a minimum read length of 100 bp. We downloaded the sequencing data from NCBI SRA or GEO respectively. Specifically, we used SRP055008 [6], GSE133291 [22], and GSE96950 [9]. Further datasets can easily be integrated, refer to <https://cortexa-rna.com/datasets>.

RNA-seq analysis

We used a standardized RNA-seq pipeline (Fig. 1a) to analyze the transcriptomic data for gene expression and alternative splicing. In brief, we trimmed the reads for

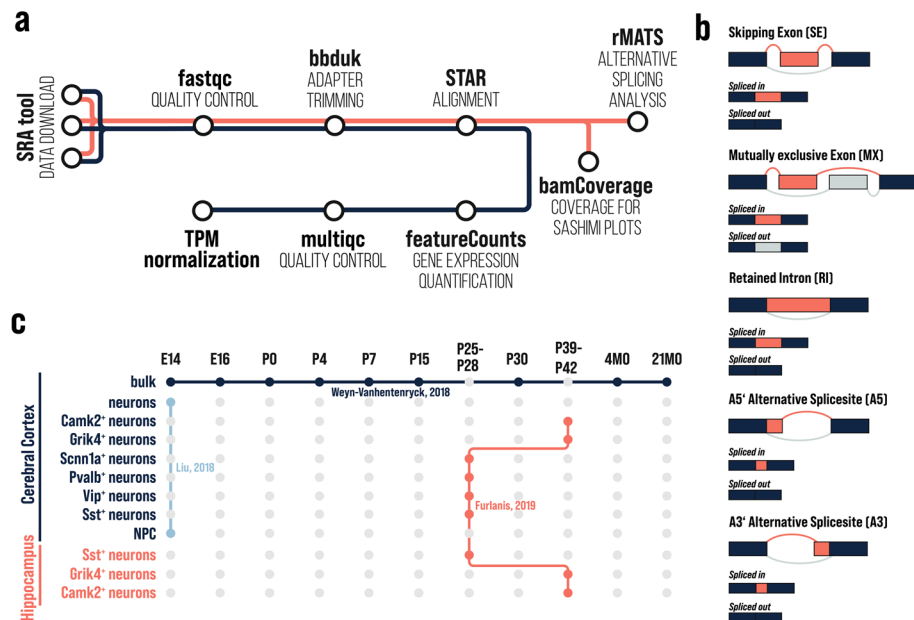


Fig. 1 **a** Standardized RNA-seq pipeline used for gene expression (blue) and alternative splicing (orange). **b** Visualization of possible alternative splicing events and the splicing outcome referred to as spliced-in and spliced-out. **c** Publicly available datasets integrated into Cortexa: SRP055008 [6], GSE133291 [22], and GSE96950 [9]

adapter sequences with BBDuk (version 39.01) [30]. The trimmed reads were aligned to the reference genome mm39 (released 19.10.2022) downloaded from Gencode using STAR (version 2.7.10b) [31] and indexed using samtools (version 1.18) [32]. FeatureCounts, provided by SubRead (version 2.0.6) [33], was used to quantify the expression of each respective gene. All gene expression counts were normalized to transcripts per million (TPM).

We utilized rMATS turbo (version 4.1.2) [34] with default settings to detect AS events. We analyzed the data for five alternative splicing events: cassette exon (skipping exon), mutually exclusive exons, intron retention, alternative A5' splice site, and alternative A3' splice site (Fig. 1b), as defined in rMATS [34]. The coverage for sashimi plots was analyzed in 10 bp steps with bamCoverage (version 3.5.2) from deepTools2 [35] converted to wig with Encode bigWigToWig.

Webapp

The web application was built with Next.js frontend framework, utilizing SQLite database for backend data storage. TypeScript enhances code maintainability and type safety, while Tailwind CSS streamlined styling. Prisma serves as the ORM tool for efficient database management. The application follows the REST principles for communication between frontend and backend components, optimizing interoperability and scalability. The website is hosted on servers of the Johannes Gutenberg University, Mainz, Germany. A detailed tutorial on how to interpret alternative splicing events presented in Cortexa is provided at <https://cortexa-rna.com/tutorial>.

Visualization of genes

Gene expression was normalized to transcripts per million (TPM) and represented as a barplot for each dataset. Alternative splicing events (Fig. 1b) are visualized as sashimi plots.

SplicePCA

SplicePCA performs principal component analysis (PCA) on averaged percentage spliced-in (PSI) values (Fig. 2a). Initially, the user can either select a subset of genes or perform SplicePCA on all genes. All events that have missing values are removed from the subsequent analysis. Next, PSI values are averaged within their respective group (e.g. E14.5 from the developmental data set [6]). These averaged PSI values are the input for the PCA, and the resulting values are plotted and available for download. Moreover, SplicePCA allows users to integrate their in-house analyzed output files (rMATS). We recommend processing the files as described in Sect. 3.2 and in the tutorial available at https://github.com/s-weissbach/cortexa_SplicePCA_example/.

Example usage of SplicePCA

To demonstrate the use of SplicePCA, we obtained *Nova2*- knock-out (KO) and wild-type (WT) data from NCBI GEO with the accession number GSE103314 [36]. We performed quality control, trimming, alignment, and alternative splicing analysis as described in Sect. 3.2 RNA-seq analysis.

Subsequently, the cassette exons from the rMATS output file were uploaded to <https://cortexa-rna.com/pca> and analyzed in the context of developmental [6] and NPC/neuron-specific [9] alternative splicing events. Next, the results from SplicePCA were downloaded and plotted using matplotlib (version 3.9.0) [37].

Utility and discussion

Alternative splicing is a prevalent regulatory mechanism in the brain that plays an important role during development and in specifying and maintaining neural cell types [6, 7, 9, 22, 28, 36, 38]. However, *Mus musculus* has ~22,000 protein-coding genes [39] of which almost all multi-exon genes undergo alternative splicing [40]. Functional implications of these alternative splicing events remain in many cases elusive. Cortexa is an easy-to-use web tool to access alternative splicing events for genes of interest in a developmental and neuronal cell-type-specific context to formulate and investigate research hypotheses.

Additionally, principal component analysis has proven to be a powerful tool to summarize alterations in the alternative splicing landscape [6, 27–29]. Representation of samples in two-dimensional space allows investigating the similarity or divergence of global splicing patterns between different experimental conditions. In our analysis of developmental alternative splicing, we observed a characteristic bell-shaped trajectory across diverse iterations, which aligns with findings reported in the literature [6]. However, the interpretation of principal components can be challenging in terms of associating them with biologically meaningful factors. By design, principal components capture the direction of maximal variance in the original data [41] which

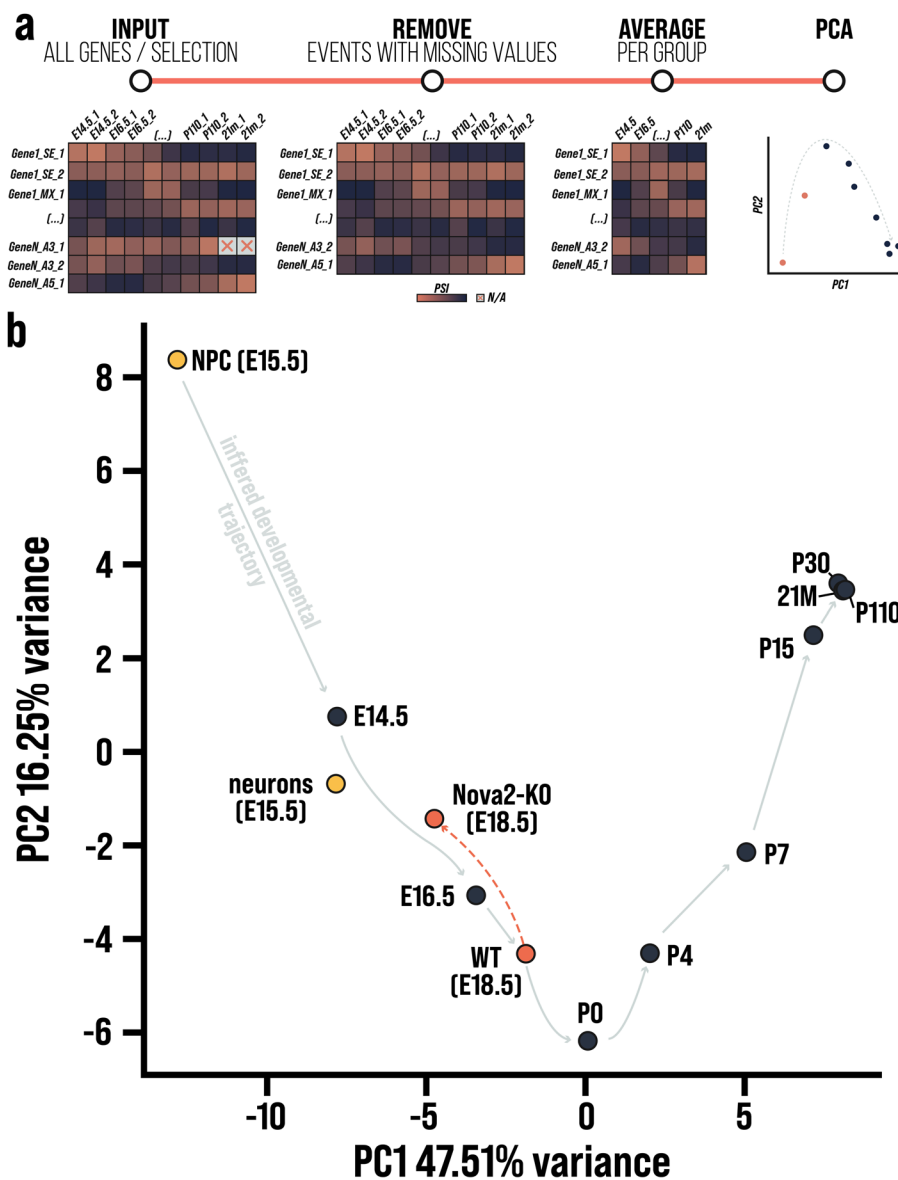


Fig. 2 **a** Schematic representation of SplicePCA. SplicePCA takes percentage spliced-in (PSI) values for all genes or a selected subset of genes as input. Alternative splicing events with missing values are removed from the dataset. The remaining PSI values are averaged over individual replicates per experimental group. Finally, PCA is performed on the averaged data, resulting in a representation of splicing patterns across groups in two-dimensional space. **b** Cassette exon data of cortical samples of WT (E18.5) and Nova2-KO (E18.5), analyzed with SplicePCA and compared to splicing changes across normal development of the mouse cortex. PCA of alternative splicing data across cortical development forms a characteristic bell-shaped trajectory (indicated by gray arrows) with P0 as its inflection point. The conditional knock-out of Nova2 resulted in a substantial deviation from the inferred normal splicing trajectory. Nova2-KO samples (E18.5) were associated with a less mature splicing pattern than E16.5 wild-type samples

does not necessarily reflect experimental or biological factors. Despite these limitations, PCA remains a valuable exploratory tool, and SplicePCA offers a user-friendly method for investigating alternative splicing in the context of development or different cell types.

Using SplicePCA, researchers can select splice events for specific genes, integrate their data, and interpret results in the context of cortical development and specific neuronal cell types. To showcase the usefulness of this approach, we re-analyzed cortical *Nova2*-KO and WT samples at embryonic day E18.5 [36] and used the SplicePCA tool. NOVA2 belongs to the class of RNA-binding proteins, governing alternative splicing during cortical development and in mature neurons [42]. Specifically, NOVA2 is required to regulate neuronal migration through splicing *Dab1* which is part of the Reelin pathway [2]. By using SplicePCA, we revealed a striking effect of *Nova2*-KO during cortical development. E18.5 knockout samples were located between E14.5 and E16.5 wild-type samples on the inferred developmental splicing trajectory, indicating a less mature splicing pattern (Fig. 2b). Thus, NOVA2 splicing activity contributes significantly to the splicing changes between E14.5 to E18.5, as reported previously [2, 36, 43]. These results support the relevance of SplicePCA which combines available datasets with new datasets. Cortexa thus allows the use of publicly available data without extensive re-analysis, which would otherwise require significant computational resources.

Conclusions

In summary, Cortexa gives access to high-quality, publicly available transcriptomic data to a broad range of scientists without the need to gain expertise in the computational aspects of gene expression and alternative splicing analysis. For experienced users, SplicePCA offers a powerful tool to summarize alterations in the alternative splicing landscapes upon experimental manipulation and compare results to the normal splicing trajectory during cerebral cortical development or in distinct neuronal cell types. Ultimately, Cortexa offers an easy-to-use extensive platform for gaining novel insights into gene expression and alternative splicing regulatory processes of the mouse brain.

Acknowledgements

Not applicable.

Author contributions

S.W. and H.T. conceived the idea and conceptualized the work. H.T., S.G., and M.H. supervised the project. S.W. and S.P. performed the bioinformatic analysis. J.M. and S.W. implemented and tested the website. S.W. created visualizations. S.W. and H.T. wrote the manuscript. M.H. and S.G. acquired funding and edited the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work has been supported by the Emergent Algorithmic Intelligence Centre of the Johannes Gutenberg University Mainz, funded by the Carl-Zeiss Foundation.

Availability of data and materials

Publicly available sequencing data were downloaded from NCBI SRA or GEO respectively. Specifically, we used SRP0550086, GSE13329122, and GSE969509. The *Nova2*-KO data that were used to test SplicePCA and a tutorial are available on GitHub at https://github.com/s-weissbach/cortexa_SplicePCA_example/. All processed datasets used in this study are available on Zenodo at <https://zenodo.org/records/13170518>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 6 June 2024 Accepted: 28 August 2024

Published online: 05 September 2024

References

1. Lennox AL, Mao H, Silver DL. RNA on the brain: emerging layers of post-transcriptional regulation in cerebral cortex development. *WIREs Dev Biol.* 2018;7: e290.
2. Vuong CK, Black DL, Zheng S. The neurogenetics of alternative splicing. *Nat Rev Neurosci.* 2016;17:265–81.
3. He X, Rosenfeld MG. Mechanisms of complex transcriptional regulation: implications for brain development. *Neuron.* 1991;7:183–96.
4. Loo L, et al. Single-cell transcriptomic analysis of mouse neocortical development. *Nat Commun.* 2019;10:134.
5. Telley L, et al. Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science.* 2016;351:1443–6.
6. Weyn-Vanhentenryck SM, et al. Precise temporal regulation of alternative splicing during neural development. *Nat Commun.* 2018;9:2189.
7. Zhang X, et al. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell.* 2016;166:1147–1162.e15.
8. Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol.* 2017;18:437–51.
9. Liu J, Geng A, Wu X, Lin R-J, Lu Q. Alternative RNA splicing associated with mammalian neuronal differentiation. *Cereb Cortex.* 2018;28:2810–6.
10. Kang HJ, et al. Spatio-temporal transcriptome of the human brain. *Nature.* 2011;478:483–9.
11. Scholpp S, et al. Her6 regulates the neurogenetic gradient and neuronal identity in the thalamus. *Proc Natl Acad Sci.* 2009;106:19895–900.
12. Bedogni F, et al. Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proc Natl Acad Sci.* 2010;107:13129–34.
13. Fromer M, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci.* 2016;19:1442–53.
14. Zhang B, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell.* 2013;153:707–20.
15. Voineagu I, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature.* 2011;474:380–4.
16. Goh WWB, Wang W, Wong L. Why batch effects matter in omics data, and how to avoid them. *Trends Biotechnol.* 2017;35:498–507.
17. Zheng W, Chung LM, Zhao H. Bias detection and correction in RNA-Sequencing data. *BMC Bioinform.* 2011;12:290.
18. Weißbach S, et al. Reliability of genomic variants across different next-generation sequencing platforms and bioinformatic processing pipelines. *BMC Genom.* 2021;22:1–15.
19. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform.* 2020;2:lqaa078.
20. O'Rawe J, et al. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Med.* 2013;5:28.
21. Raj B, Blencowe BJ. Alternative splicing in the mammalian nervous system: Recent insights into mechanisms and functional roles. *Neuron.* 2015;87:14–27.
22. Furlanis E, Traunmüller L, Fucile G, Scheiffele P. Landscape of ribosome-engaged transcript isoforms reveals extensive neuronal-cell-class-specific alternative splicing programs. *Nat Neurosci.* 2019;22:1709–17.
23. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci.* 2003;100:189–92.
24. Yao Z, et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature.* 2023;624:317–32.
25. Huntley MA, et al. Genome-wide analysis of differential gene expression and splicing in excitatory neurons and interneuron subtypes. *J Neurosci.* 2020;40:958–73.
26. Zhang Y, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci.* 2014;34:11929–47.
27. Feng H, et al. Complexity and graded regulation of neuronal cell-type-specific alternative splicing revealed by single-cell RNA sequencing. *Proc Natl Acad Sci.* 2021;118: e2013056118.
28. Jacko M, et al. Rbfox splicing factors promote neuronal maturation and axon initial segment assembly. *Neuron.* 2018;97:853–868.e6.
29. Martí-Gómez C, et al. Functional impact and regulation of alternative splicing in mouse heart development and disease. *J Cardiovasc Transl Res.* 2022;15:1239–55.
30. BBMap. SourceForge <https://sourceforge.net/projects/bbmap/> (2022).
31. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
32. Danecek P, et al. Twelve years of SAMtools and BCFtools. *GigaScience.* 2021;10:giab008.
33. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30:923–30.
34. Shen S. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data | PNAS. *Proc Natl Acad Sci.* 2014. <https://doi.org/10.1073/pnas.1419161111>.
35. Ramírez F, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44:W160–5.

36. Saito Y, et al. Differential NOVA2-mediated splicing in excitatory and inhibitory neurons regulates cortical development and cerebellar function: neuron. *Neuron*. 2019;101:707-720.E5.
37. Ari N, Ustazhanov M. Matplotlib: a 2D graphics environment. *IEEE J Mag*. 2014.
38. Heck J, et al. More than a pore: how voltage-gated calcium channels act on different levels of neuronal communication regulation. *Channels*. 2021;15:322–38.
39. Breschi A, Gingeras TR, Guigó R. Comparative transcriptomics in human and mouse. *Nat Rev Genet*. 2017;18:425–40.
40. Jiang W, Chen L. Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Comput Struct Biotechnol J*. 2021;19:183–95.
41. Todorov H, Fournier D, Gerber S. Principal components analysis: theory and application to gene expression data analysis. *Genom Comput Biol*. 2018;4: 100041.
42. Meldolesi J. Alternative splicing by NOVA factors: from gene expression to cell physiology and pathology. *Int J Mol Sci*. 2020;21:3941.
43. Yano M, Hayakawa-Yano Y, Mele A, Darnell RB. Nova2 regulates neuronal migration through an RNA switch in disabled-1 signaling. *Neuron*. 2010;66:848–58.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.