MASTERARBEIT | MASTER'S THESIS

Titel | Title

The role of alternative splicing in Hawaiian bobtail squid (*Euprymna scolopes*) gene regulation

verfasst von | submitted by

Pt. Georgios Efthymiadis

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 220
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Joint-Masterstudium Evolutionary Systems Biology
Betreut von | Supervisor:	Assoz. Prof. Dr. Oleg Simakov

Acknowledgements

The following master's thesis was conducted from October 2023 to September 2024. I would like to express my deep gratitude to Oleg Simakov for the trust over the past year and the invaluable experience. Being part of his group, equipped me with essential scientific skills and theoretical knowledge that will shape my future endeavors.

I owe the biggest thanks to my supervisor, Thea Rogers, whose guidance and unwavering assistance made this project possible. Thanks to her everyday support not only I enhanced my technical expertise but most importantly I expanded my critical thinking and problem-solving abilities. I am very grateful for her role in shaping my academic growth, and I wish her continued success as a scientist. I truly hope to collaborate with both Oleg and Thea and other members of this group, as colleagues in the future.

To my parents, Kosmas and Ioanna, I extend my deepest thanks for their selfless love and continuous support, which have made this journey possible. Lastly, I dedicate this effort to my beloved partner, Anna-Maria, whose constant encouragement and love have been a cornerstone of this achievement, as she is the unsung hero of the entire process.

Kurzfassung

Die Rolle des alternativen Spleißens bei der Genregulation des Hawaiianischen Bobtail-Tintenfischs (*Euprymna scolopes*).

Alternatives Spleißen (AS) ist ein kritischer zellulärer Mechanismus, der die transkriptomische Vielfalt erhöht, indem er es einzelnen Genen ermöglicht, mehrere reife mRNA-Isoformen zu produzieren, die möglicherweise zu funktionell unterschiedlichen Proteinen führen. Dieser Prozess wird durch verschiedene genomische Bedingungen reguliert, die gewebespezifische Spleißmuster bestimmen.

Trotz zahlreicher Studien, die die Rolle von AS in der Genregulation bei verschiedenen Metazoen-Spezies beleuchten, bleibt seine Rolle in Tintenfischen vollständig unerforscht. Coleoide Tintenfische, bekannt für ihre genomische Komplexität und hochentwickelten Nervensysteme, zeigen einzigartige Verhaltens- und morphologische Anpassungen. Unter diesen hochentwickelten Tieren dient eine Tintenfischart als aufstrebendes Genomik-Modell. Der Hawaii-Zwergtintenfisch (*Euprymna scolopes*), dessen evolutionäre Geschichte durch Reorganisationsereignisse geprägt ist und dessen Genominhalt mit einer Vielzahl regulatorischer Elemente angereichert ist, stellt einen vielversprechenden Kandidaten dar, um zu untersuchen, wie Spleißentscheidungen beeinflusst werden.

Durch den Einsatz modernster Next-Generation-Sequencing-(NGS)-Technologien bieten aktuelle Publikationen Einblicke in verschiedene regulatorische genomische Merkmale, einschließlich konservierter Gencluster und dreidimensionaler chromosomaler Konfigurationen, und positionieren *E. scolopes* als Schlüsselmodell für das Studium der Genregulation und Evolution von Tintenfischen. Die Untersuchung der Rolle des alternativen Spleißens in *E. scolopes* wird nicht nur unser Verständnis seiner genomischen Architektur erweitern, sondern auch entscheidende Details über die Auswirkungen von AS auf die transkriptomische Vielfalt in verschiedenen Geweben liefern. Dieses Vorhaben wird erheblich zur umfassenderen Erforschung der Tintenfischgenomik sowie zu den komplexen regulatorischen Mechanismen des alternativen Spleißens beitragen.

Schlüsselwörter: Alternatives Spleißen, *Euprymna scolopes*, Genomik, Genregulation

Abstract

The role of alternative splicing in Hawaiian bobtail squid (*Euprymna scolopes)* gene regulation.

Alternative splicing (AS) is a critical cellular mechanism that enhances transcriptomic diversity by enabling individual genes to produce multiple mature mRNA isoforms, potentially resulting in functionally diverse proteins. This process is regulated by various genomic conditions that dictate tissue-specific splicing patterns.

Despite numerous studies elucidating the role of AS in gene regulation across different metazoan species, its role in cephalopods remains completely unexplored. Coleoid cephalopods, known for their genomic complexity and highly developed nervous systems, exhibit unique behavioral and morphological adaptations. Among these highly sophisticated animals, a squid species serves as an emerging genomics model. The Hawaiian bobtail squid (*Euprymna scolopes*), with an evolutionary history characterized by reorganization events and a genomic content enriched with a vast repertoire of regulatory elements, consists of a promising candidate to explore how splicing decisions are influenced.

Utilizing state-of-the-art next-generation sequencing (NGS) technologies, recent publications provide insights into various regulatory genomic traits, including conserved gene clusters and three-dimensional chromosomal configurations, positioning *E. scolopes* as a key model for studying cephalopod gene regulation and evolution. Investigating the role of alternative splicing in *E. scolopes* will not only enhance our understanding of its genomic architecture but also provide pivotal details on the impact of AS on transcriptomic diversity across various tissues. This endeavor will contribute significantly to the broader comprehension of cephalopod genomics as well as to the complex regulatory mechanisms underlying alternative splicing.

Keywords: Alternative splicing, *Euprymna scolopes*, genomics, gene regulation

Contents

List of Tables

List of Figures

2. Introduction

2.1 Cephalopods cephalopod evolution

Cephalopods consist of a unique class of marine invertebrates, that emerged around 530 million years ago (Mya), during the late Cambrian period (Lemche HM 1957). The divergence between Coleoidea (octopuses, cuttlefish, and squid) and Nautiloidea occurred approximately 350 to 480 Mya (Bergman 2013; Kröger, 2011). Coleoid cephalopods are known to possess the largest and most elaborate invertebrate nervous systems (Ritschard et al., 2019), with high neuronal complexity as well as cognitive abilities (Young, 1965), that likely contribute to their dynamic behavioral repertoire, such as rapid changes in skin texture and color for camouflage. Moreover, large expansions in gene families, like C2H2 zinc fingers and protocadherins, predominantly expressed in nervous tissues have been recorded (Albertin et al., 2015; Garrett et al 2012). These gene families are hypothesized to facilitate neuronal diversity in both vertebrates and cephalopods (Albertin et al., 2015; Styfhals et al., 2022).

Apart from their advanced nervous system, coleoids cephalopods excel in the interplay between gene regulation and evolution, where both changes in transcription level due to gene regulation as well as non-synonymous changed in protein-coding regions comprise significant evolutionary drivers (Ritschard et al., 2019). Furthermore, cephalopods employ epigenetic mechanisms such as DNA methylation and histone modification (Macchi et al 2022) that can regulate transcript diversity in a similar way observed in vertebrate nervous tissues (Liscovitch-Brauer et al., 2017; Garrett & Rosenthal, 2012).

Additionally, species-specific symbiotic organs, like in the case of Euprymna genus, offer promising models for exploring the evolution of organismal novelties (Albertin et al., 2020; Ritschard et al., 2019). Given these specialized attributes, coleoid cephalopods can be considered focal points for extensive research across various disciplines, including behavioral biology, developmental biology, and neurobiology (O'Brien et al., 2018; Hanlon et al., 2018; Wollesen et al., 2015; Koenig et al., 2016).

2.2 *Euprymna scolopes* as emerging genomics model

This project focuses on the Hawaiian bobtail squid (Family: Sepiolidae), *Euprymna scolopes*, a model organism known for its symbiotic relationships with bacterial consortia in two organs, namely the light organ likely evolved through the subfunctionalisation of genes expressed in the eye, and the accessory nidamental gland, enriched with novel, species-specific orphan genes (Belcaid et al., 2019). The *E. scolopes* genome is notably large (approximately 5.1 Gb) and is characterized by a high repetitive content (Belcaid et al., 2019; Schmidbaur et al., 2022). This repetitive nature is largely due to transposable elements (TEs), particularly from the long interspersed nuclear elements (LINEs) family, which constitute 48.2% to 50.5% of the identifiable repetitive elements (Heckman et al., 2021). This particularly large number of TEs in Euprymna genus, is also contributing to the significant disparity in genome sizes between Euprymna species and closely related taxa like *Rondeletiola* species.

Initially, the substantial size and complexity of the *E. scolopes* genome had been attributed to a speculated whole-genome duplication event (Hallinan and Lindberg, 2011; Yoshida et al., 2011). However, after the emergence of new genomic data, recent research has identified a large-scale genomic reorganization in *E. scolopes*, suggesting its potential role in shaping cephalopod complexity (Belcaid et al., 2019). Changes in the genomic content may provide the template for new gene linkages, that can directly impact gene expression, as different regulatory blocks arise or change position (Harewood et al 2014). Furthermore, additional findings propose that the genome's size and complexity may also stem from high variability in intronic sequences (Ritschard et al., 2019).

2.3 Alternative splicing and transcriptomic diversity

Numerous studies on AS and isoform quantification have provided valuable insights into the transcriptional enrichment of eukaryotic organisms (Pan et al., 2008) as well as of various methods for assessing the diversity of transcriptome (Jones et al., 2024).

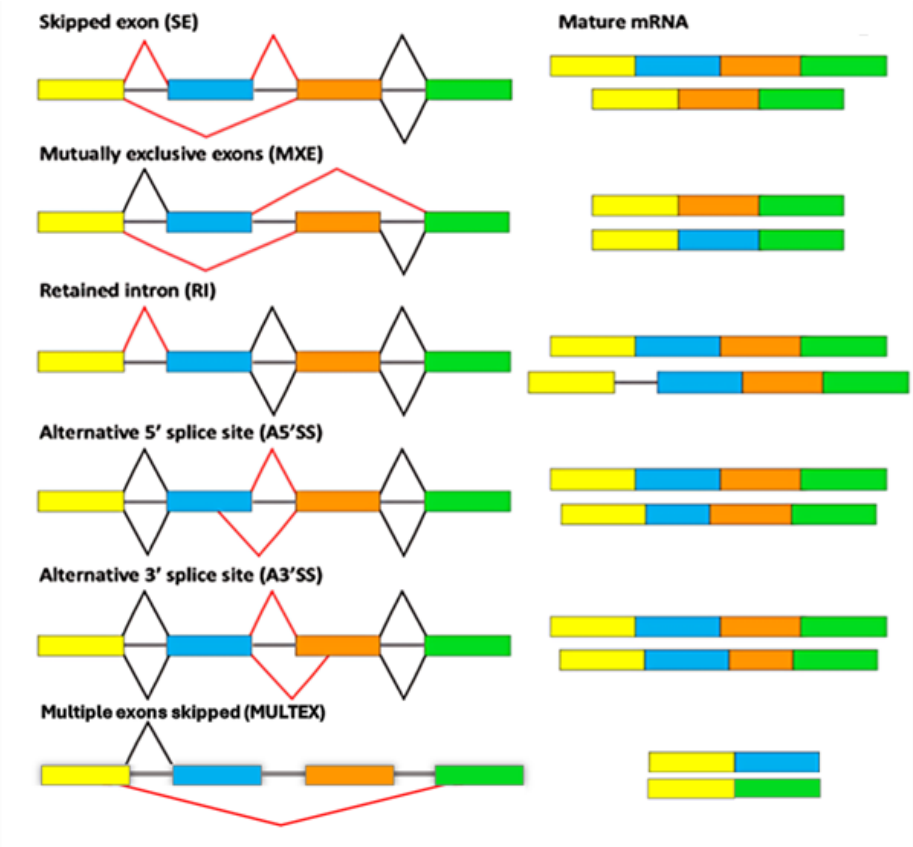
Almost 98% of the eukaryotic genome consists of non-coding regions that do not encode protein sequences (Mattick et al., 2001), meaning that only 2% consists of protein coding genes, able to produce functional

molecules, raising the reasonable question of how such genomic content generates vast phenotypic complexity. One explanation lies in the ability of a gene transcript to be spliced into different combinations of exons, leading to the production of diverse isoforms (Fig. 1a,b) with variable functions. Although not all isoforms might lead to functional proteins in a specific environmental context, this cellular process acting on precursor mRNA enhances the transcriptomic and proteomic diversity of multicellular organisms. Moreover, recent comparative studies across different metazoan species suggest that AS complexity increases with organismal complexity, correlating with the number of unique cell types (Barbosa et al., 2012; Chen et al., 2014).

Data derived from different studies have shown that AS occurs frequently in neural tissues leading to the generation of a high isoform variety, characterizing their dynamic functionality and additionally contributing to their development and function (Marasco et al., 2023; Baralle et al., 2017). AS also contributes to intraspecific biodiversity by producing different isoforms between sexes and morphs of the same species (Rogers et al., 2021; Cordellier et al 2020; Grantham et al 2018).

Given the significance of the nervous system in development and behavior, it is reasonable to speculate that it could consist of a target for natural selection to act upon. Conversely, if AS is important in phenotypic complexity, it is feasible that is involved in the evolution of unique structures in the nervous systems, like in the case of coleoids. This can lead to an extensive transcriptome and a potentially more functionally diverse proteome, emphasizing its pivotal role in complex organisms like *E. scolopes*.

Figure 1. *A schematic diagram delineating the six studied modes of AS. The exons are depicted by different colored boxes, whereas introns are represented by grey lines. The connections between exons signify segments excised from the final transcript. Of note, the red lines emphasize AS events (adapted from Rogers et al 2021, supplementary material).*



2.4 Phenotypic plasticity and alternative splicing

One of the distinctive traits that cephalopods are renowned for, is their ability to swiftly modulate body coloration by mimicking their surroundings (Cloney et al., 1968). The influence of transcriptomic diversity from AS could potentially be one of the major mechanisms that contribute to this observed rapid phenotypic plasticity. Together with other reported genomic features, such us differential gene expression and RNA editing (Rosenthal & Eisenberg 2023; Albertin et al., 2022), the hypothesis that multiple protein isoforms resulting from AS can aid these mechanisms provides a plausible explanation for their adaptive capabilities.

When a gene is undergoing AS, a dominant isoform typically emerges alongside several others transcribed at lower levels, suggesting their functional neutrality in specific environmental contexts (Ahi & Singh, 2022). This splicing variation that may contain deleterious effects, accumulates, but due to purifying selection, remains unexpressed (Xing & Lee, 2007, Rogers et al 2021). However, the unutilized isoforms may harbor concealed phenotypic effects (Wright & Laue 2021). When external pressures disrupt these

hidden effects, these transcript variants may confer advantages upon release, leading to the selection of novel functional isoforms. Thus, with the use of this hidden genetic variation, AS could potentially contribute to adaptive evolution by influencing ecological adaptation over time.

2.5 Detection of alternative splicing

Genomic analysis has significantly changed since the arrival of next-generation sequencing (NGS) techniques. In the context of AS, the focus is directed to the identification of splice junctions, which facilitate the transition from precursor to mature mRNA. The canonical splicing sites in eukaryotic mRNA processing typically involve a conserved sequence motif, with the donor sequence GU at the 5' end of the upstream exon and the splice acceptor sequence AG at the 3' end of the downstream exon (Zhang et al., 2021 ; Dou et al 2006). Thus, the quantification of AS events relies on the approach of analyzing sequencing reads that comprise splice junctions (Ding et al., 2017; Au et al., 2010). A common method of detecting AS isoforms is to localize reads where the sequence transitions from the GU motif of the 5' exon to the AG motif of the 3' exon differ from those in the reference genome.

Additionally, to investigate AS events in a robust way is by using short-read data alongside a reference genome of good quality. This method quantifies differences in splicing levels based on read coverage, where high read depth reliably detects variations (e.g.in the case where sample A has half the reads mapping to a splice junction compared to sample B, this indicates that the isoform using that splice junction is expressed half as frequently compared to sample A).

Moreover, recent advancements in the field of sequencing, particularly with long-read data, have significantly contributed to studies focusing on AS, by allowing the sequencing of entire isoforms (Wright et al., 2022). In essence, short-read data are indispensable for achieving high sequence depth and accurately identifying AS junctions while long-read data can additionally offer the ability to sequence entire isoforms (Weirather et al., 2017).

2.6 Influence of genomic features on alternative splicing

Recent surveys have uncovered evidence supporting large, conserved gene orders (termed as synteny) across phyla boundaries, even in early branching lineages like Porifera, suggesting the presence of a fundamental genetic framework , similar regulatory constraints and potential functional significance dating back to the metazoan ancestor (Simakov et al 2022; Robert et al 2022; Simakov and Kawashima 2017).

In *Euprymna scolopes*, Schmidbauer et al. 2022 categorized two types of syntenic regions (termed as microsyntenies): MACIs (microsyntenies associated with cephalopod innovations), that emerged post genome rearrangement and are specific to cephalopods, likely responsible for the unique functions that only evolved in these lineages, and metazoan microsyntenies, orthologous gene sets shared with distantly related species, implying the significance of evolutionary conserved functions across different phyla.

The evolutionary trajectory of coleoids has been shaped by large-scale genomic reshuffling, including the loss of the ancient bilaterian microsyntenies (Albertin et al., 2015), leading to the disruption of conserved gene linkages in the genome of *E. scolopes* and other coleoid species, such as *Octopus bimaculoides*, giving rise to the cephalopod-specific syntenic regions (Belcaid et al., 2019; Bosch 2019). Given that all genes are subject to AS, understanding how AS is influenced by these conserved microsyntenic regions may elucidate whether evolution primarily affects genes unique to cephalopods or is more pronounced in universally distributed orthologs shared across metazoans.

Furthermore, the three-dimensional organization of the squid genome is characterized by topologically associating domains or TADs (Schmidbaur et al 2022; Rouressol et al 2023). By bringing close transcription start sites and enhancer regions, TADs can facilitate transcriptional coordination, that could potentially affect the splicing of precursor mRNA within or near these domains (Dixon et al., 2012; Pope et al., 2014; Symmons et al., 2016).

The spatial distribution of AS genes within TADs could reveal patterns indicative of regulatory influences on splicing patterns. Components like the conserved zinc finger protein CTCF (CCCTC-binding factor),

contributing to TAD formation, may act as regulators of AS for genes near TAD boundaries, while other elements within TADs could influence splicing variation for genes located deeper within these domains (Alharbi et al., 2021; Tian et al 2020). Understanding how AS genes are organized within TADs can provide insights into how genomic architecture influences splicing patterns, or even if differential splicing across various tissues is impacted by divergent of TAD structures.

2.7 Impact of this study

Despite the high genomic complexity and extensive mRNA production (Albertin et al., 2022) of coleoid cephalopods, little to no information exist about the impact of AS in their gene regulation. To address this key question, we focus on the Hawaiian bobtail squid, *Euprymna scolopes*, an emerging model organism for genomic studies, due to the intriguing attributes of its genome that provide a template to investigate how a complex mechanism like AS is coordinated. Expectations from this analysis include first insights on the splicing profile of a squid species as well as differential splicing patterns across tissues, deciphering the distribution of AS modes, and understanding the impact of genomic organization on AS. The research is motivated by hypotheses concerning the role of AS in gene regulation, including tissue specific AS profiles, its association in microsyntenic regions, and how it is impacted from the three-dimensional genome architecture. The findings are expected to provide a foundation for future relevant studies on other coleoid species, studies dedicated to the functionality of AS isoforms in squids, and endeavors focusing on how genome regulatory blocks influence splicing patterns.

The outcomes of this project align with existing literature by providing further insights into the crucial role of AS in enhancing functional diversity, particularly in the complex tissues of metazoan species. Furthermore, additional findings highlight evidence of the complex regulation of AS in cephalopods, influenced by three-dimensional genome organization and group-specific gene clusters, contributing to ongoing research on how evolutionary pressures shape transcriptomic flexibility in highly sophisticated organisms.

3. Materials and methods

In this study, a range of bioinformatics tools were employed. Each step of the analysis is detailed in the respective section of the constructed pipeline. Data visualization and statistical analysis were conducted in the R environment (v. 4.2.1), while Python3 (v. 3.11.7) was employed for the manipulation of large datasets. Some scripts were optimized using LLMs approaches, to enhance efficiency via code debugging.

3.1 Sample collection and library preparation

Detailed descriptions of specimen collection, preparation, RNA extraction, and the generation of long and short read data are available in two publications. Rouressol et al. 2023 cover the sequencing and mapping of short read data, while Rogers et al. 2024 provide details about the generation of long read data. These studies collectively offer a comprehensive RNA-seq dataset from a wide range of tissues and samples, divided into two distinct batches.

In this study, a total of 17 male adult samples from both batches, covering six different tissues were selected. The exclusive focus on male samples was due to the availability of long read data solely for this gender. The tissues of interest included the central brain and left optic lobe (referred to as brain tissues), the hectocotylus (referred to as a hybrid tissue), and the skin, mantle, and left gills (referred to as somatic tissues). For each tissue, three samples were selected, except for the left gills, where only two male samples were available. A complete description of the sample details can be found in the supplementary material, Table 5.

3.2 Quality control

Initially, the FastQC (v.0.12.1) (Andrew S., 2010) package was utilized with default parameters to generate quality control reports for both R1 and R2 reads (forward and reverse, respectively) of all samples. After a comprehensive review of the reports, raw RNA-seq data that required treatment, underwent trimming of adaptor content using Trim Galore (v.0.6.10) (Krueger et al 2023). For all samples, no instances of low-quality reads were observed. The parameters specified for Trim Galore included the Illumina adaptor (--illumina) and paired-end reads (--paired). De novo, FastQC reports were generated for the curated data.

3.3 Indexing and mapping reads on the reference genome

Subsequently, the reference genome (BioProject number PRJNA6616848, Schmidbaur et al. 2022) was indexed, with the help of the GTF file containing the gene indices (Rogers et al 2024) and sample reads were mapped using the STAR alignment tool (v. 2.7.11a) (Dobin et al 2013). Genome indexing utilized the -genomeGenerate run mode, with the -sjdbOverhang parameter to define splice junction characteristics. BAM files were generated for each sample, sorted by genomic coordinates (--SortedByCoordinate). Quality assessment of BAM files was conducted using Samtools flagstat (v.1.19) (Li et al 2009), while Samtools index was employed to generate indexes for each BAM file to facilitate subsequent analyses.

3.4 Gene expression analysis

Using featureCounts tool of Subread package (v. 2.0.6) (Liao et al 2014) on BAM files generated by STAR, genes expressed in each tissue were identified. An expression threshold based on the log2 fold change was established, with genes exhibiting a fold change above 1 being considered as expressed (Love et al 2014). Genes failing to surpass the predetermined expression level, were deemed to represent captured noise and were filtered out from downstream analysis.

In addition to the default parameters of featureCounts, we specified several additional others, that included -g to designate gene IDs as the attribute type, -t set to exons to focus counting on reads mapping exclusively to exonic regions associated with gene expression, -M to include counts from multi-mapping reads, and -p for paired-end read data processing.

By default, featureCounts output raw counts i.e. the number of reads that are assigned to each feature (e.g., genes, exons, transcripts), which aren't characterized by a particular unit. After normalization, using DESeq2 (v. 1.38.3) (Love et al 2014), the counts are represented in CPM units (counts per million). The normalization process scales the counts to represent the expression of each feature in CPM, accounting for differences in sequencing depth across samples.

3.5 Alternative splicing event detection

The identification of AS events was conducted using Spladder (v. 3.0.5) (Kahles et al., 2016). The Spladder build mode, employed with default settings and the --ignore-mismatches parameter to account for imperfectly mapped reads, processed BAM files to construct a custom benchmark of detected AS events. This mode was run separately for each of the six tissues to assess tissue specific AS events and collectively for all tissues, enabling a comprehensive AS event analysis.

The filtration procedure in build mode is implemented at the program's initiation during the construction of splice graphs, by using criteria like number of mismatches, to remove low-quality reads or number of alignments to confirm intron edges. This integrated filtering along with the additional step of gene expression analysis, ensured a robust approach by including reliable splicing events.

An important metric in alternative splicing (AS) studies is the Percent Spliced In (PSI) index, which quantifies the proportion of transcripts that include a specific exon or splice site compared to all possible transcripts that could include it. Specifically, for each detected AS event, the corresponding PSI value represents the ratio of normalized read counts indicating the inclusion of a transcript over the total normalized reads for that event. For instance, for a hypothetical exon skipping event, a PSI value of 0.6 would mean that the exon is included in approximately 60% of the transcripts in the sample. This metric is crucial for comparing different samples, as shifts in PSI values indicate changes in splicing patterns between samples or for a particular AS event. The PSI value ranges from 0 to 1 denoting that one of the two isoforms is expressed while a middle value of 0.5 indicating the equal expression of both isoforms. PSI can be calculated using the following formula:

$$\boldsymbol{PSI = \frac{Number\ of\ trasncripts\ that\ support\ isoform}{Total\ number\ of\ transcripts}}$$

PSI values used for the construction of heatmaps, visualizing the matrix of AS events (Figure 2b). In instances of NaN (Not a Number) values, due to inadequate coverage or very low expression levels, these values were substituted with 0 to prevent computational errors. Moreover, in cases where the total sum of

PSI values for the selected subset of samples was 0 or the PSI value of every sample was 1, these AS events were also removed, since there are no differences across the selected subset of samples. Additionally the pvclust package (v. 2.2-0) (Suzuki et al 2006) in R was utilized, to generate cladograms for each AS mode, employing the bootstrap method for assessing clustering profiles of samples.

Spladder test mode was used to perform pairwise comparisons between all tissues to identify differential splicing events, utilizing default parameters. Differential splicing was determined based on the delta Percent Spliced In (dPSI) value, which quantifies the difference in the percentage of transcripts including a specific exon between two different tissues. The results of these analyses are presented in Table 4 which details the differential splicing events for each tissue pairwise comparison, while Table 5 displays the number of unique differentially spliced genes in a matrix format.

3.6 Venn diagrams

Venn diagrams was used to quantify the overlap of AS genes between brain tissues and other tissue types (Figure 3A). Additionally, a comprehensive Venn diagram (Figure 3B) incorporating all six target tissues was created in order to identify conserved AS genes shared across all tissues and to determine the unique number of tissue specific AS genes. The generation of the Venn diagrams was carried out using the R package ggVennDiagram (version 1.5.2) (Gao, C.-H et al 2024).

To further analyze the overlaps, we conducted a Super Exact Test using the R package SuperExactTest (v. 1.1.0) (Wang M. et al 2015). This approach defined the extent of shared AS genes beyond visual representation, enabling the statistical validation of the overlaps observed in the Venn diagrams, ensuring that they did not occur due to random chance. The results are depicted at the Supplementary Table 1.

3.7 Protein annotation analysis

Protein annotation enrichment analysis was conducted to identify enriched functional categories across tissues, aiming to the determination of the predicted proteins generated by genes with alternative isoforms.

The purpose of this step was to determine if AS genes are enriched in specific functions compared to the entire genome.

Protein annotation from Rogers et al. 2024 was employed to assist in defining the functions of the corresponding genes. The annotation data were processed in R to create tables correlating gene IDs with their specific functions. Contingency tables were then constructed for pairwise tissue comparisons, identifying differences in gene function expression. To infer for statistical significance Fisher's exact test was utilized, with p-values adjusted for multiple comparisons using FDR correction. Only p-adjusted values < 0.05 were considered significant. The results are summarized in Table 6.

Additionally, genomic background comparisons were conducted for all six tissues. The genomic background which represents all the expressed genes in the *E. scolopes* genome, was used in order to identify tissue-specific enrichments of unique AS genes compared to the total number of expressed genes.

3.8 Microsynteny association analysis

Data of identified Microsynteny IDs and their corresponding conserved gene orders (Schmidbaur et al., 2022) were utilized, to quantify the number of AS genes associated with two types of microsyntenies in each tissue. Furthermore, the count of non-syntenic AS genes for each tissue was determined by subtracting the combined number of AS genes in MACIs and metazoan microsyntenies from the total number of AS genes observed in that tissue.

The percentage of AS genes within each of the three categories (MACIs, metazoan, non-syntenic) was calculated to determine their respective proportions. For non-syntenic genes, the percentage was computed by dividing the number of non-syntenic AS genes by the total number of AS genes in that tissue.

A custom Python script was used to assess matching cases of AS genes ids, with the provided information on microsyntenic gene orders. Subsequently, Fisher's exact test was employed to assess significant enrichment of AS genes in MACIs or Metazoan syntenies for each tissue. This test, chosen for its suitability

with small sample sizes (less than 1000 values), evaluated the association between AS gene presence in these synteny categories.

3.9 Interacting gene pairs

In order to investigate if interacting gene may display coordinating splicing, data on the interaction frequencies of *Euprymna scolopes* gene pairs including corresponding orthologous gene pairs from the closely related species *Octopus bimaculoides* were analyzed (Rogers et al in prep) . The gene interaction pairs were categorized into four groups (see Figure 6). The analysis focused exclusively on exon skipping due to the straightforward detection of changes in exon inclusion levels using RNA-seq data. PSI values from all 17 samples across six different tissues were utilized.

The interacting gene pairs were required to have both genes undergoing AS and detected by SplAdder. Aiming to determine the relationship between their splicing patterns, Pearson correlation analysis of PSI values for each gene pair was conducted. The calculated mean and median correlation values provided an overview of the four categories (Table 9). A high correlation may suggest potential co-splicing, (i.e. the splicing pattern of one gene is related to its interacting pair), while a low or no correlation may signify independent splicing patterns of the gene pair. Mean correlation provides the average correlation coefficient for gene pairs within each interaction category, while median correlation offers a view to the central tendency unaffected by outliers. Finally, the results of the probability density function for the four interaction categories were visualized with the use of a kernel density plot.

In addition, the non-parametric Wilcoxon signed-rank test was employed, to infer for significant differences for each gene pair on their PSI values across different interaction categories. The mean p-value was calculated to indicate any statistical significance or not, of splicing level differences between the two genes in each pairwise comparison of the different categories (Supplementary Table 4).

3.10 Localization of AS genes at TADs

Micro-C data from Rogers et al in prep. were used for the creation of Hi-C contact maps. Reads were previously mapped to the reference genomes (Belcaid et al., 2019; Albertin et al., 2022) using Hi-CPro v. 3.1.0 (Servant et al., 2015). Insulation scores were calculated using FAN-C (v. 0.9.28) (Kruse et al 2020) to estimate the likelihood of boundary locations forming Topologically Associated Domains (TADs). To determine the positioning of AS genes within or at the boundaries of TADs, a subset of 70 random individual AS genes was selected (10 AS genes for each tissue and 10 of the brain tissues overlaps). Except for randomness, additional criteria of the selected AS genes subsets, were that the candidate genes must belong to a gene interacting pair that both genes are AS, and their average interaction frequency is above 10 (KR normalized). Hi-C contact maps, focused on a resolution of 25 kb, were visualized with Juicebox (v. 1.9.8) (Robinson et al 2018).

3.11 Isoform classification

To effectively distinguish AS events from transcriptional noise, an approach diverging from conventional methodologies was implemented. While conventional processes involve the use of more than one tools to identify AS events and assess their overlap, in this study, SQANTI3 (v. 3-5.2) (Pardo-Palacios et al., 2024) was integrated into the pipeline alongside Spladder (following gene expression filtering) for sanity verification. SQANTI3 takes input, long-read data in order to classify isoforms as genuine or artifacts. The available of long-read data for four of the six target tissues (left optic lobe, hectocotylus, central brain, and skin), facilitated the implementation of this approach

SQANTI3 employs a two-way validation approach and distinguishes artifacts (or transcriptional noise) from genuine novel isoforms through two main steps: isoform classification and filtering. Its filtering mechanism prioritizes features such as adenine content at the 3' end, consecutive adenines, and retention of transcription start sites (TSS). Moreover, SQANTI3 excludes isoforms with minimal short-read coverage and single-exon structures as potential artifacts.

Initially, the sqanti3_qc.py function with default parameters was employed, which utilizes long-read data in collapsed GFF format (Rogers et al. 2024), along with the GTF file and reference genome of *Euprymna scolopes* (Schmidbaur et al. 2022), to generate the initial classification file. Subsequently, we applied the sqanti3_filter.py rules function with default parameters to the initial output, to assess the isoforms characterization.

The data were then categorized into artifacts and true isoforms and overlapped with results from Spladder. Our criterion was that if SQANTI3 identified at least one true isoform for each gene that Spladder characterized as alternatively spliced, then it was considered a true alternatively spliced gene.

4. Results

4.1 Tissue-specific analysis on alternative splicing

Reported paired read mapping rates ranged from 99.97% to 100%. Table 1 presents the number of expressed genes, the identified AS genes, the proportion of AS genes relative to the total expressed genes, and the derived AS events for each tissue. These results reflect the overall AS gene pool per tissue, not tissue specific AS genes, indicating that common gene IDs may appear across different tissues (refer to Figure 3 for the number of unique AS genes per tissue).

	Central brain	Left optic lobe	Hectocotylus	Skin	Left gills	Mantle
Expressed genes	16771	18682	17556	17315	15147	15857
AS genes	2515	2665	2294	1718	1286	1882
Proportion	14.99%	14.26%	13.06%	9.92%	8.49%	11.86%
AS events	8189	8205	7385	4765	2998	6161

Table 1. *Number of total expressed genes in each tissue studied, along with the total alternatively spliced (AS) genes per tissue and the corresponding proportion of these AS genes relative to the expressed genes. The fourth row illustrates the number of AS events per tissue generated by the AS genes, considering that one gene may undergo multiple modes of splicing or be spliced more than once for the same mode, resulting in a higher count of AS events compared to AS genes.*

The left optic lobe exhibits the highest number of expressed genes, while the left gills have the lowest. This trend is consistent for AS genes, with the left optic lobe having the highest number and the left gills having significantly fewer, less than half as many. However, the central brain shows the highest proportion of AS genes relative to the total expressed genes. Hectocotylus, which is characterized as hybrid tissue as a part of this is tissue is neurated, based on the total number of AS genes detected, correlates more to brain tissues rather than somatic tissues.

Concerning the proportion of AS events a similar pattern is observed, as brain tissues exhibit significantly higher measure, followed by the hybrid tissue hectocotylus, and then the three somatic tissues. Notably, the

discrepancy in the number of total AS events and AS genes arises because a single gene can produce multiple transcripts, each capable of being spliced in various ways and involving multiple AS modes (Figure 1). These results demonstrate that tissues influenced by the presence of nerves are more prone to AS compared to less innervated tissues. Another intriguing finding revealed that the number of genes undergoing AS using the 3' alternative acceptor and 5' alternative donor site are relatively enriched in the cellular splicing machinery of *E. scolopes*.

Different types of AS events across tissues were quantified, Tables in 2 and 3 detail the total number of AS genes and events by the splicing modes depicted in Figure 1, for each tissue. Exon skipping mode was the commonest AS mode, followed by intron retention.

Tissue	3' SS	5' SS	ES	IR	MES	MUTEX	Total AS genes
Central brain	728	652	1315	970	233	70	2515
Hectocotylus	663	584	1094	875	188	55	2294
Left optic lobe	728	661	1439	1022	268	88	2665
Skin	475	445	801	591	115	51	1718
Left gills	361	291	608	322	78	35	1286
Mantle	573	487	930	681	133	58	1882

Table 2. *Total number and classification of AS genes for each of the six AS modes across each tissue (3'SS = alternative 3' acceptor site, 5'SS = alternative 5' donor site, ES = exon skipping, IR = intron retention, MES = multiple exons skipped, MUTEX = mutually exclusive exons).*

Tissue	3' SS	5' SS	ES	IR	MES	MUTEX	Total AS events
Central brain	1505	1074	2694	2436	333	147	8189
Hectocotylus	1694	1055	2183	2035	312	106	7385
Left optic lobe	1492	1048	2958	2170	376	161	8205
Skin	1045	722	1494	1213	190	101	4765
Left gills	918	417	960	566	99	38	2998
Mantle	1342	914	1824	1680	262	139	6161

Table 3. *Total number and classification of AS events for each of the six AS modes across each tissue (3'SS = alternative 3' acceptor site, 5'SS = alternative 5' donor site, ES = exon skipping, IR = intron retention, MES = multiple exons skipped, MUTEX = mutually exclusive exons).*

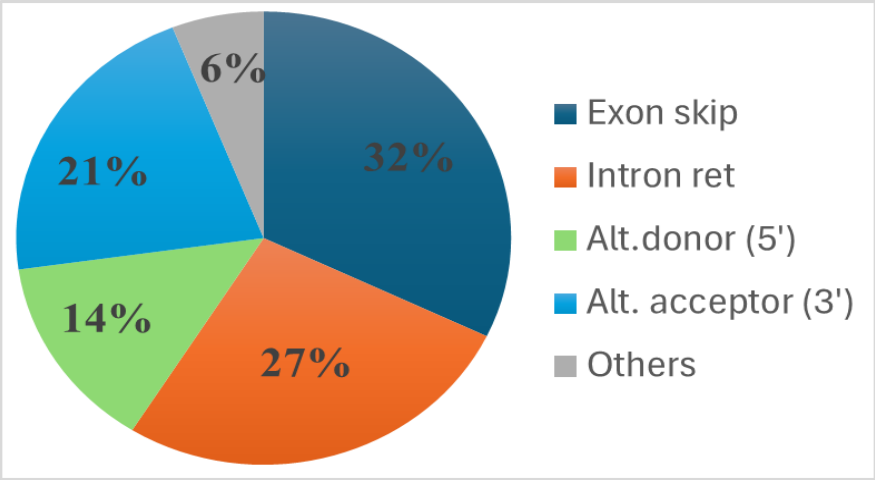


Figure 2. *Illustration of E.scolopes splicing profile, constructed from six different tissues, highlighting the proportions of each AS mode. The category "Others" includes the sum of the modes mutually exclusive exons and multiple exons skip.*

4.2 Tissue-specific splicing patterns and clustering analysis revealed by PSI values

Employing percent spliced in (PSI) values across the six AS modes, tissue-specific splicing patterns were assessed. Results were visualized with six heatmaps (Figure 2a) and an additional cladogram showing sample clustering and branching points. Robustness of the clustering was confirmed using Bootstrap method, with derived associated values indicating the clustering confidence.

The samples of brain tissues (central brain, left optic lobe) consistently clustered together across the different AS modes, indicating high similarity in their broad splicing patterns. In the case of the somatic tissues, skin and left gills samples also exhibited consistent clustering, suggesting uniform splicing patterns. Even though two of the three mantle samples continuously clustered together, one deviated frequently. Notably, hectocotylus samples showed variable clustering, occasionally aligning with brain tissues, while other times with somatic tissues, highlighting this tissue's distinct splicing patterns.

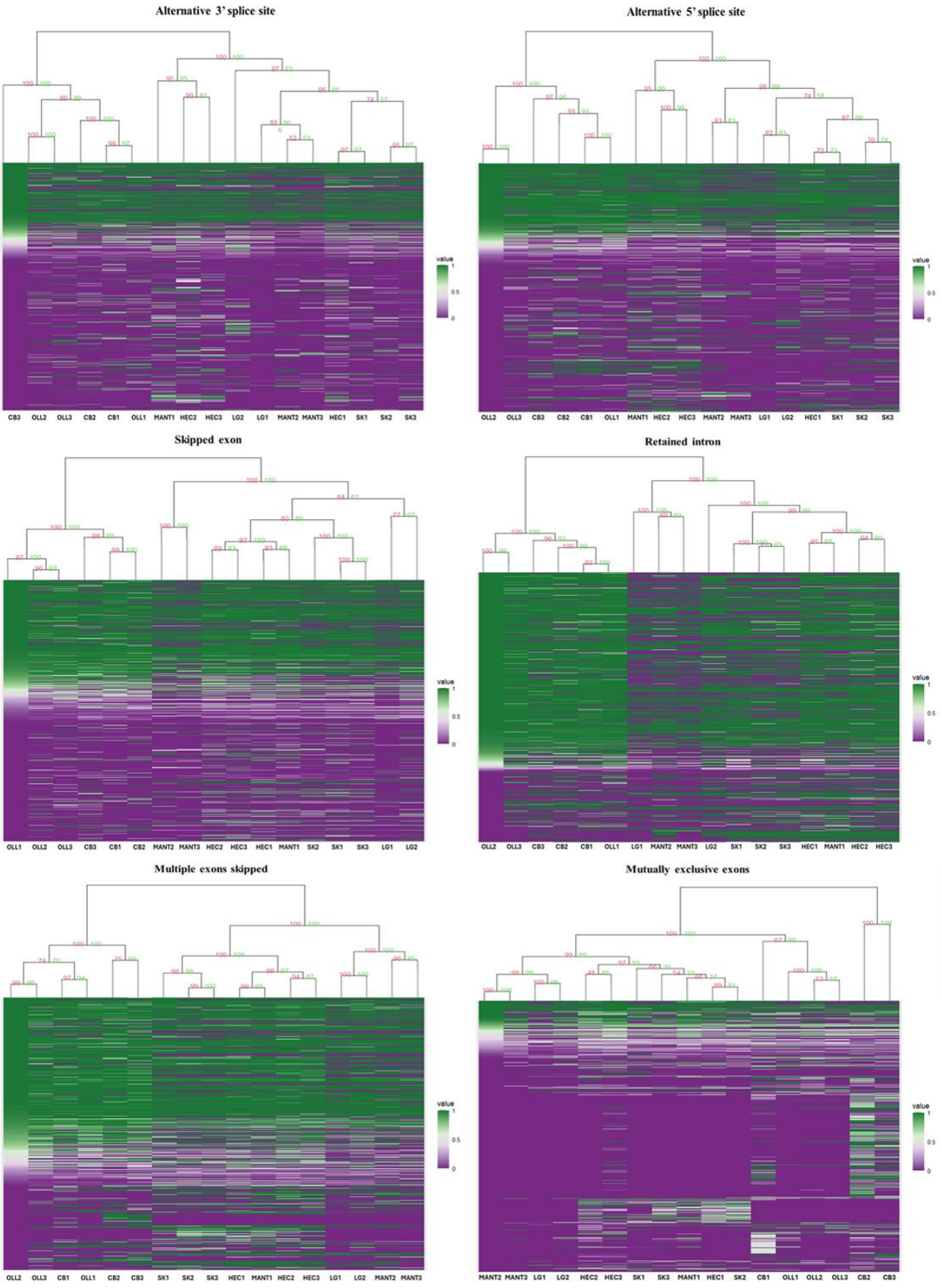

Alternative 3' splice site
CB3 OLL2 OLL3 CB2 CB1 OLL1 MANT1 HEC2 HEC3 LG2 LG1 MANT2 MANT3 HEC1 SK1 SK2 SK3
value
1
0.5
0
Alternative 5' splice site
OLL2 OLL3 CB3 CB2 CB1 OLL1 MANT1 HEC2 HEC3 MANT2 MANT3 LG1 LG2 HEC1 SK1 SK2 SK3
value
1
0.5
0
Skipped exon
OLL1 OLL2 OLL3 CB3 CB1 CB2 MANT2 MANT3 HEC2 HEC3 HEC1 MANT1 SK2 SK1 SK3 LG1 LG2
value
1
0.5
0
Retained intron
OLL2 OLL3 CB3 CB2 CB1 OLL1 LG1 MANT2 MANT3 LG2 SK1 SK2 SK3 HEC1 MANT1 HEC2 HEC3
value
1
0.5
0
Multiple exons skipped
OLL2 OLL3 CB1 OLL1 CB2 CB3 SK1 SK2 SK3 HEC1 MANT1 HEC2 HEC3 LG1 LG2 MANT2 MANT3
value
1
0.5
0
Mutually exclusive exons
MANT2 MANT3 LG1 LG2 HEC2 HEC3 SK1 SK3 MANT1 HEC1 SK2 CB1 OLL1 OLL2 OLL3 CB2 CB3
value
1
0.5
0

Figure 3. *Visual representations of the percentage of the four six distinct patterns of alternative splicing are depicted in heatmaps. Percent spliced-in values (PSI) indicate the proportion of alternative isoforms at a splice site. A PSI value of 1 or 0 signifies exclusive expression of one isoform, while 0.5 denotes equal expression of both isoforms. The red and green numbers indicate the bootstrap probability values, highlighting the same observation out of 100 distinct samplings (CB = central brain, OLL = left optic lobe, HEC =Hectocotylus, SK = skin, MANT = mantle, LG = left gills).*

4.3 Differential splicing analysis

The differential splicing analysis across the six tissues was conducted in order to assess patterns of individual splicing. Therefore, despite that two tissues might be grouped together based on PSI values considering they share similarities in broad splicing patterns, in differential splicing analysis, splicing events might vary between them, because more localized differences are detected.

The dPSI (delta percent spliced in) metric proved invaluable to quantify these local splicing changes, as it measures changes in PSI values between two groups, providing a quantitative assessment of local splicing variations. Only results with adjusted p-values on log scale (p-adj < 0.05) were considered for robustness and reduced bias.

Table 4 summarizes differential splicing findings, detailing different AS events across tissue comparisons. Exon skipping (ES) mode showed the highest numbers, except in three instances (central brain and left optic lobe, hectocotylus and left gills, hectocotylus and skin) where intron retention (IR) was more prevalent. Additionally, Table 5 presents a matrix highlighting the number of unique differentially spliced genes. Several key insights can emerge from this analysis.

Particularly, the brain tissues (central brain and left optic lobe) exhibited the highest number of differentially spliced events and genes with skin. Their comparison with left gills and mantle showed additional large disparity in the number of events and genes, underscoring a significant splicing pattern divergence between brain and somatic tissues.

Surprisingly and despite its hybrid nature, hectocotylus showed a substantial number of differentially spliced genes and events with both central brain and left optic lobe. In contrast, it revealed minimal differences with the somatic tissues, with the lowest being compared to mantle.

Lastly, pairwise comparisons among somatic tissues revealed fewer differentially spliced genes and events, indicating more conserved splicing profiles in these tissues. Overall, these findings underscore a complex landscape of AS across the six tissues.

Tissue comparisons	3´ SS	5´ SS	ES	IR	MES	MUTEX	Total number
Central brain & Left gills	6	14	92	34	44	25	**215**
Central brain & Mantle	11	7	54	8	63	40	**183**
Central brain & Hectocotylus	8	17	150	33	72	44	**324**
Central brain & Skin	0	21	180	40	88	41	**370**
Central brain & Left optic lobe	2	8	27	39	20	10	**106**
Hectocotylus & Left gills	1	4	4	5	11	13	**38**
Hectocotylus & Mantle	1	0	6	2	0	1	**10**
Hectocotylus & Skin	3	0	4	12	14	16	**49**
Left optic lobe & Hectocotylus	25	28	195	56	88	38	**430**
Left optic lobe & Left gills	3	16	97	21	51	29	**217**
Left optic lobe & Mantle	10	11	93	29	69	43	**255**
Left optic lobe & Skin	17	25	219	56	102	51	**470**
Skin & Left gills	0	1	8	3	20	2	**34**
Skin & Mantle	0	1	5	2	12	0	**20**
Mantle & Left gills	0	0	5	0	8	11	**24**

Table 4. *Enumeration of the differentially alternatively spliced events identified in pairwise comparisons between different tissues. The identification of significantly alternatively spliced events was based on adjusted p-values on log scale (p-adj < 0.05) obtained from each pairwise tissue comparison. These stringent criteria ensure the robustness of the results, highlighting AS event differences between the tissues analyzed.*

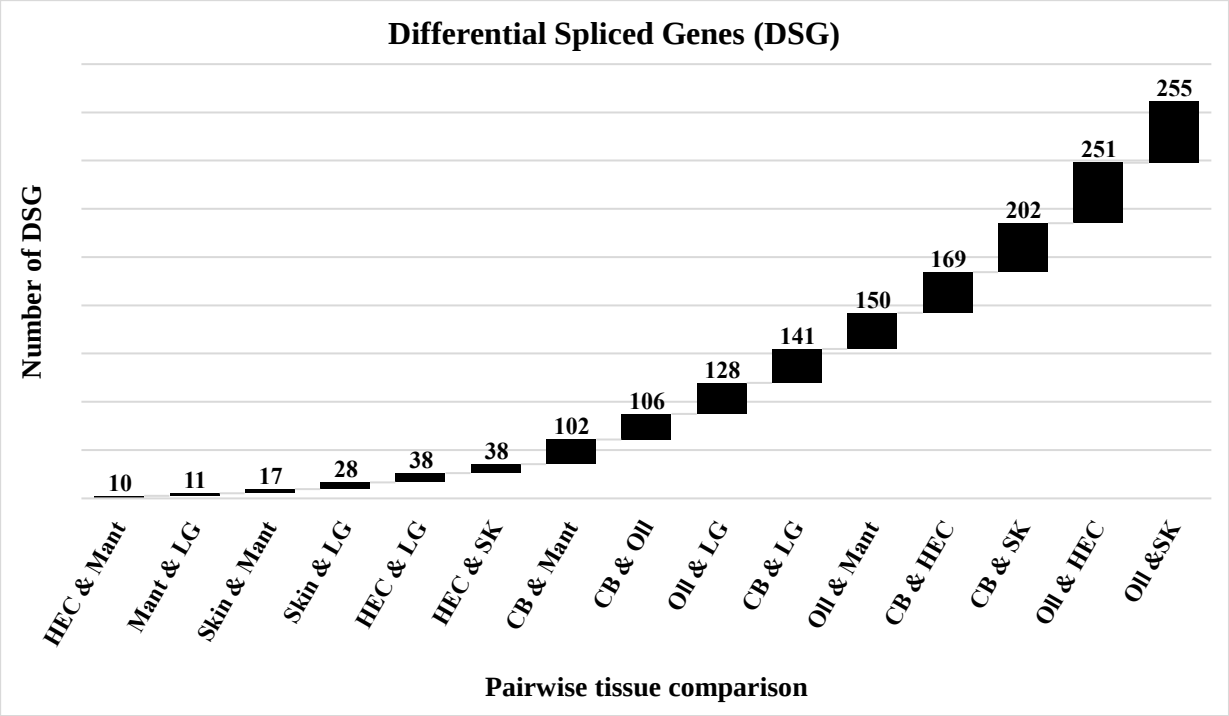


Figure 4. *Waterfall plot visualizing the count of the differentially spliced genes across all possible pairwise tissue comparisons (p-adj on log scale < 0.05). The numbers are ranked from the smallest difference (HEC & Mant) to the highest (Oll & SK). Abbreviations: CB = central brain, Oll = left optic lobe, HEC = hectocotylus, LG = left gills, SK = skin and Mant = mantle.*

4.4 Overlapping AS genes across different tissues

Venn diagrams were used to assess the overlap of shared AS genes across the brain, hybrid and somatic tissues. A significant number of common AS genes exhibited between brain tissues and hectocotylus (Figure 3A), indicating the highest similarity among all comparisons. In contrast, brain tissues compared with left gills showed the fewest shared AS genes (Figure 3B), while comparisons with mantle and skin (Figures 3C and 3D, respectively) revealed nearly identical numbers of shared genes. These findings underscore the dissimilar AS profile of brain tissues compared to somatic and propose a closer association with hectocotylus.

Subsequently, a comprehensive Venn diagram was constructed, to illustrate the shared sub-groups of AS genes across the six target tissues, identifying a conserved group of 696 AS genes. Additionally, unique AS gene counts were determined for each tissue: 440 for left optic lobe, 328 for central brain, 205 for hectocotylus, 105 for skin, 104 for mantle, and 48 for left gills. These results are analogous with the overall number of AS per tissue depicted in Table 1, proposing that the enrichment of AS genes in brain tissues might be indicative of higher complexity and regulation of AS in these regions. Furthermore, the comprehensive Venn diagram revealed two intriguing findings: 311 AS genes shared by the central brain and left optic lobe, and 186 AS genes shared across all tissues except the left gills.

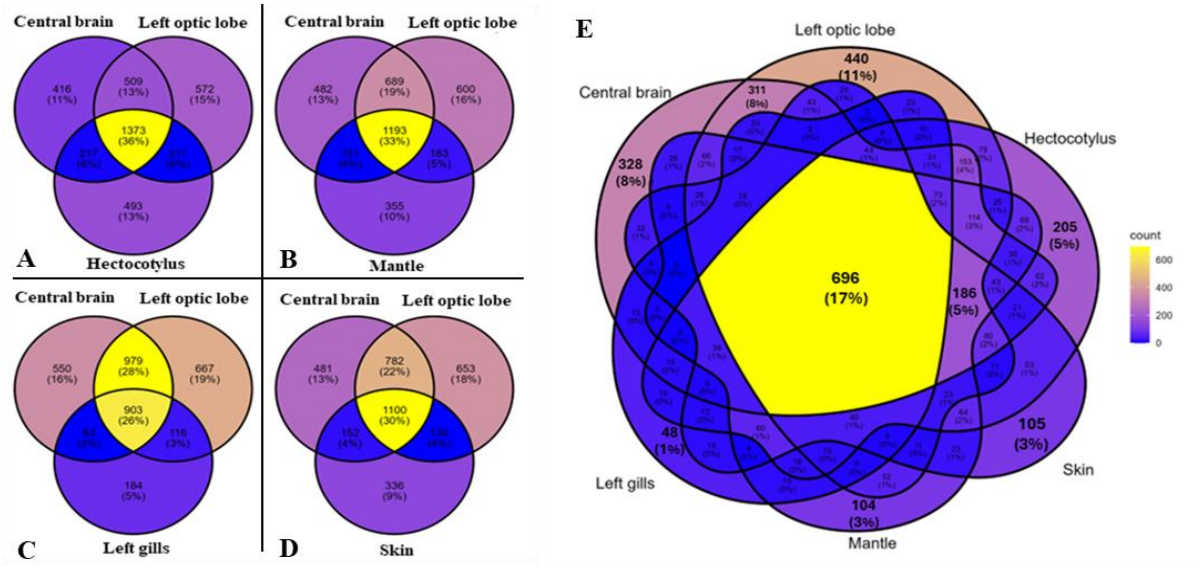


Figure 5. *Venn diagrams illustrate the distribution of AS genes across multiple comparisons, with a color gradient indicating gene counts: blue for the fewest and yellow for the most genes. Panels A-D compare common AS genes between two brain tissues (central brain and left optic lobe) and the rest of the tissues. Panel E shows the distribution of AS genes across all six tissues. The peripheral regions (highlighted with bold letters) represent unique AS genes specific to each tissue, while the central yellow area highlights the core set of AS genes shared by all six tissues. The findings, evaluated using the Super Exact Test, are detailed in Supplementary Table 1.*

4.5 Protein annotation analysis

Protein annotation proved instrumental for elucidating the functional implications of identified AS gene sets across tissues. Tissue - pairwise functional enrichment analysis was conducted for each tissue's unique AS gene set. For all analyses, Fisher's exact test, adjusted for multiple comparisons was utilized to identify significant enrichments (adjusted p-value < 0.05).

The conserved set of 696 AS genes did not exhibit significant enrichment in any terms and background comparisons showed also no significant findings. On the other hand, pairwise comparisons of functional enrichment between brain tissues and the rest of the tissues are summarized in Table 6. Brain tissues and hectocotylus showed similar predicted functions of AS genes, with no significant differences noted. In contrast, comparisons with somatic tissues revealed significant differences, particularly with the left optic lobe showing more enriched functions compared to the central brain.

Enriched in specific protein domains, common to brain tissues compared to somatic tissues included WD40/YVTN repeat-like-containing domain superfamily, WD40-repeat-containing domain superfamily, Zinc finger, and P-loop containing nucleoside triphosphate hydrolase.

Unlike somatic tissues, hectocotylus showed no significant differences in functional enrichment compared to brain tissues, underscoring its hybrid nature at functional level too. Somatic tissues did not exhibit significant enrichment in any pairwise comparisons, highlighting functional distinctions between brain and non-brain tissues in terms of AS genes.

	Hectocotylus	**Left Gills**	**Mantle**	**Skin**
Central Brain	0	4	2	4
Left Optic Lobe	0	14	11	13

Table 5. *Number of significantly enriched protein annotations identified using Fisher's exact test between the two brain tissues (central brain, left optic lobe) and the four non-brain tissues (hectocotylus, left gills, mantle, and skin), after the adjustment of p-value. All protein annotations are enriched in both brain tissues.*

4.6 Microsyntenies

The aim of this section was to provide insights into the relationship between AS and conserved gene clusters. The focal point constituted whether AS predominantly affects sets of genes contributing to cephalopod-specific traits or is more prevalent in the universally shared ortholog gene sets among metazoans. In *E. scolopes*, MACIs comprise of 2.297 genes, whereas Metazoan microsyntenies include 1.486 genes. Table 7 shows the proportion of AS genes within MACIs and Metazoan syntenies relative to these totals. The non-syntenic genes comprise the total number of AS genes that do not appear within either of the two microsyntenies. Their proportion is calculated relative to the total number of AS identified per tissue. Total AS genes are summarized in Table 1, and tissue-specific unique AS genes are depicted in Figure 3E.

MSD	Central brain	Left optic lobe	Hectocotylus	Skin	Left gills	Mantle
MACIs	492	509	425	338	262	367
% AS genes	21.42%	22.16%	18.50%	14.71%	11.41%	15.98%
Metazoan	194	211	213	148	113	158
% AS genes	13.06%	14.20%	14.33%	9.96%	7.60%	10.63%
Non-syntenic	1829	1945	1656	1232	911	1357
% AS genes	72.72%	72.98%	72.19%	71.71%	70.84%	72.10%

Table 6. *Number of syntenic AS genes per tissues found in the two confirmed microsyntenic types. The number of non-syntenic AS genes per tissue was also evaluated by subtraction of the total syntenic genes form the total AS genes for each tissue. The proportion of each category was evaluated.*

This analysis showed that the highest proportion of AS genes is observed in MACIs, which comprise syntenic gene clusters specific to cephalopods linked with their evolutionary innovations. To identify potential statistical significance of the differences in AS gene counts between MACIs and Metazoan syntenies Fisher's exact test was assessed. This statistical test evaluates the association between categorical variables, particularly suitable for small sample sizes datasets.

Tissue	p-value	odds ratio
Central brain	4.354e-11	1.815
Left optic lobe	7.09e-10	1.720
Hectocotylus	0.000847	1.356
Skin	1.804e-05	1.559
Mantle	2.854e-06	1.598
Left gills	0.000118	1.564
Core set	0.000839	1.689

Table 7. *Statistical analysis of AS genes distribution between MACIs and Metazoan syntenies across various tissues. For each tissue, the p-value and odds ratio from Fisher's exact test are presented, indicating the likelihood and strength of the association between AS genes and syntenic regions. For clarification, the results of the core set of genes are visualized under the section of total genes.*

The results from Fisher's exact tests revealed significant differences in the number of AS genes between MACIs and Metazoan syntenies, across all six target tissues. In addition, the odds ratios, proved the association between categorical variables, by consistently exceed 1 across all tissues, indicating a higher likelihood of finding AS genes in MACIs compared to Metazoan syntenies.

4.7 Spatial organization influences alternative splicing

To investigate the spatial localization of AS genes, we selected a subset of AS genes and examined their positions within the genome, specifically assessing their occurrence within or at the borders of Topologically Associated Domains (TADs). Out of the 70 genes analyzed in different chromosomes, 62 of them (approximately 88.5%), were located near TAD borders. Figure 6 illustrates three examples of these genes: g25295, g2678, and g3842. For a more detailed look of the selected genes, the genome assembly of *E. scolopes* can be accessed at https://www.cephalopodresearch.org/ceph_gdatab/. The remainder of the contact maps (excluding the 10 of the brain tissue overlap) are included in the Supplementary Material in Figures 1-6.

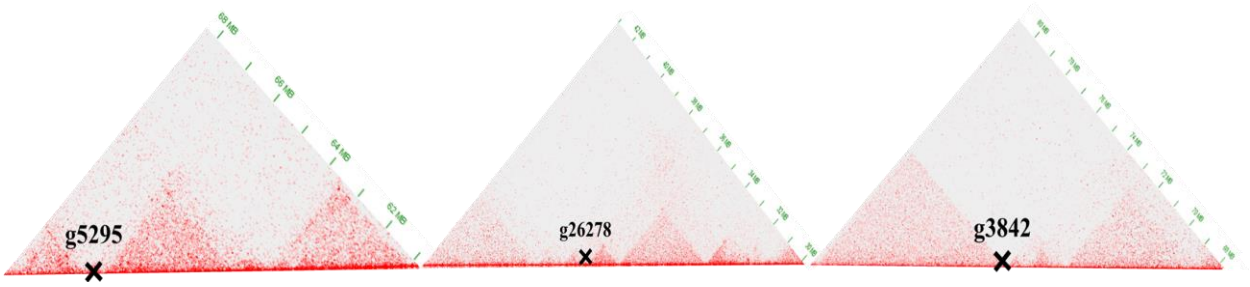


Figure 6. *Example of three triangle contact maps, exhibiting the spatial location of three AS genes. All three contact maps have a resolution of 25kb. The black arrow signifies the actual position of the AS gene.*

To evaluate gene locations relative to TAD structures, insulation scores for each gene were assessed. Higher insulation scores suggest that a gene is located within a TAD, while lower scores indicate that a gene may be closer to a TAD boundary. Subsequently, the mean and median insulation scores for each AS gene total were compared with those of expressed and non-AS gene totals within the same tissue. Detailed results for each score per tissue category and window size are provided in Supplementary Tables 2-3. Additionally, the mean and median insulation scores were calculated for all genes in the genome, without specific tissue inference.

The analysis was conducted using 500 and 1000 kb window sizes to prove that even in largely different sizes, the same trend was maintained. Across the entire genome, mean and median insulation scores were -0.37 and -0.17 for 500 kb window size, and -0.43 and -0.25 for 1000 kb window size. The corresponding mean and median scores, for the expressed (non-AS) genes, were -0.56 and -0.45 for 500 kb, and -0.69 and -0.61 for 1000 kb. In contrast, AS genes exhibited the lowest insulation score values, with mean and median scores of -0.71 and -0.62 for 500 kb, and -0.84 and -0.71 for 1000 kb. The results of the insulation score distribution for both AS and expressed genes at 500kb and 1000kb window sizes are highlighted in the density plots of Figure 5. Moreover, the standard deviations of the AS and expressed genes were 0.93 and 0.95 for 500 kb, and 0.94 and 0.93 for 1000 kb, respectively. These results highlight the lower insulation scores of AS genes, compared to both the expressed genes and the genome average, an observation consistent with their proximity to TAD boundaries.

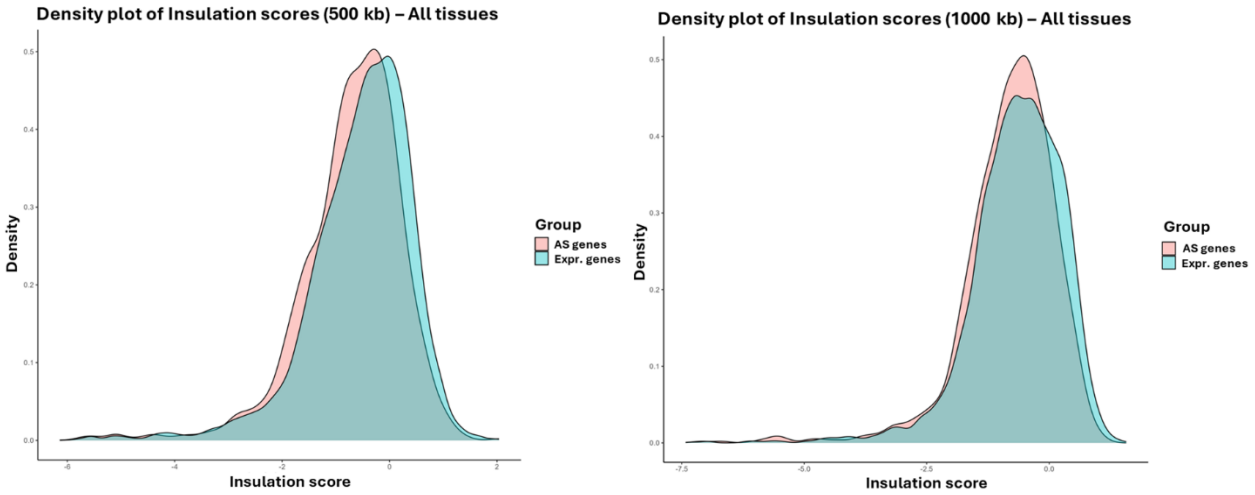


Figure 7. *Density plots illustrate the insulation scores for window sizes of 500 kb and 1000 kb, comparing the distributions between AS genes and Expressed (non-AS) genes.*

4.8 Gene interaction pairs show no correlation across sample PSI values

Previous endeavors (Seitan et al., 2013; Gorkin et al., 2014; Baguette A. 2020), have found that genes interacting in the three-dimensional space, have their co-expression regulated. Motivated from that, this section aimed to explore whether gene pairs with high interaction frequencies exhibit in addition coordinated splicing patterns, specifically in terms of their PSI values.

Generated data (Rogers et al in prep) providing information over the interaction frequencies of gene pairs in *E. scolopes* genome and their corresponding orthologs from the closely related species *O. bimaculoides* were utilized, to tackle this hypothesis. Gene pairs were grouped into four categories based on interaction frequencies: 1) "Interacting in both species" (interaction frequency > 10 in both), 2) "Interacting in *E. scolopes* only" (interaction frequency over 10 only in *E. scolopes*), 3) "Interacting in *O. bimaculoides* only" (interaction frequency over 10 only in *O. bimaculoides*), and "Not interacting in either species" (no significant interaction frequency in both).

Figure 6 highlights the results, showing only a few co-splicing events (green peak), for the "Interacting in *E. scolopes* only" category. This small subset of gene pairs with relatively high correlation values, averaging

~0.75, while the overall correlations were mostly between -0.5 and 0.5, indicating no association of the PSI values.

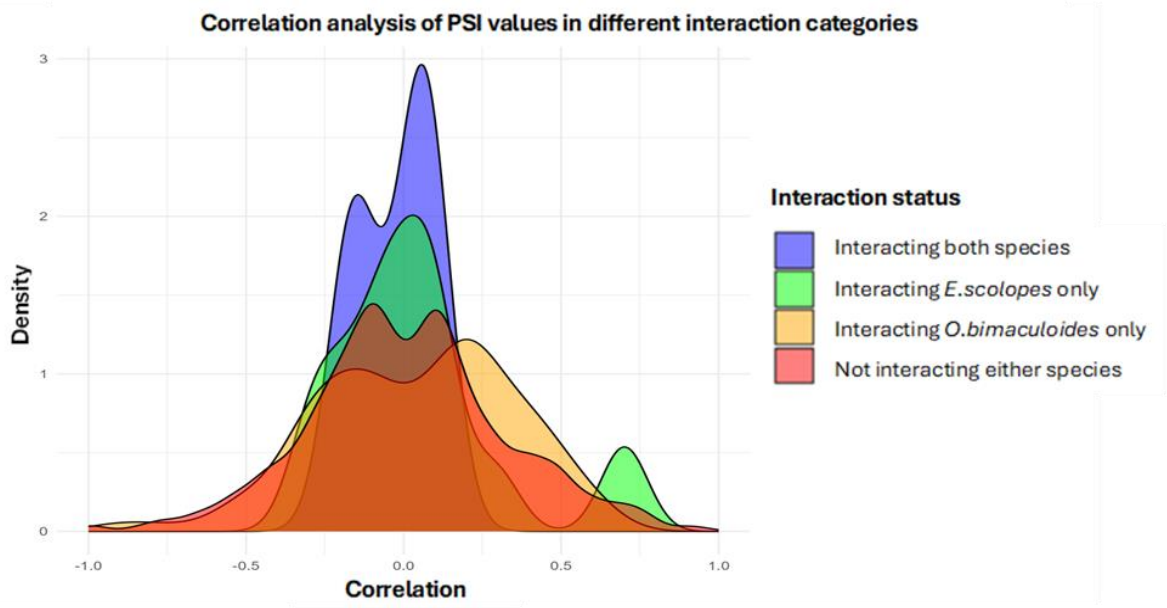


Figure 8. *PSI correlation analysis of all AS genes across all six tissues, categorized by the four different gene interaction types.*

Table 9 showcases the results of the correlation analysis, displaying the mean and median correlation values, which ranged from -0.03 to 0.04 and -0.02 to 0.08 respectively. These numbers propose no positive or negative correlation. As the data didn't follow a normal distribution (based on Shapiro-Wilk normality test), the non-parametric Wilcoxon test was employed to infer for any statistically significant differences across the four categories (Supplementary Table 4).

Despite the general uncorrelation, a small peak of AS genes interacting in *E.scolopes* showcased relatively high correlation of about 0.75. Moreover, one case of perfect positive correlation was observed between the gene pair g9258 and g8243, which displayed perfectly correlated PSI values across all samples (correlation = 1). The predicted functions for these genes are immunoglobulin production (g9258) and Zinc finger C2H2-type domains (g8243).

Interaction status	Mean correlation	Median correlation
Interacting both species	-0.03	0.03
Interacting *E.scolopes* only	0.04	-0.02
Interacting *O.bimaculoides* only	0.04	0.08
Not interacting either species	0.01	0.00

Table 8. *Correlation analysis of the PSI values of the gene interaction pairs for each category.*

4.9 Isoform classification

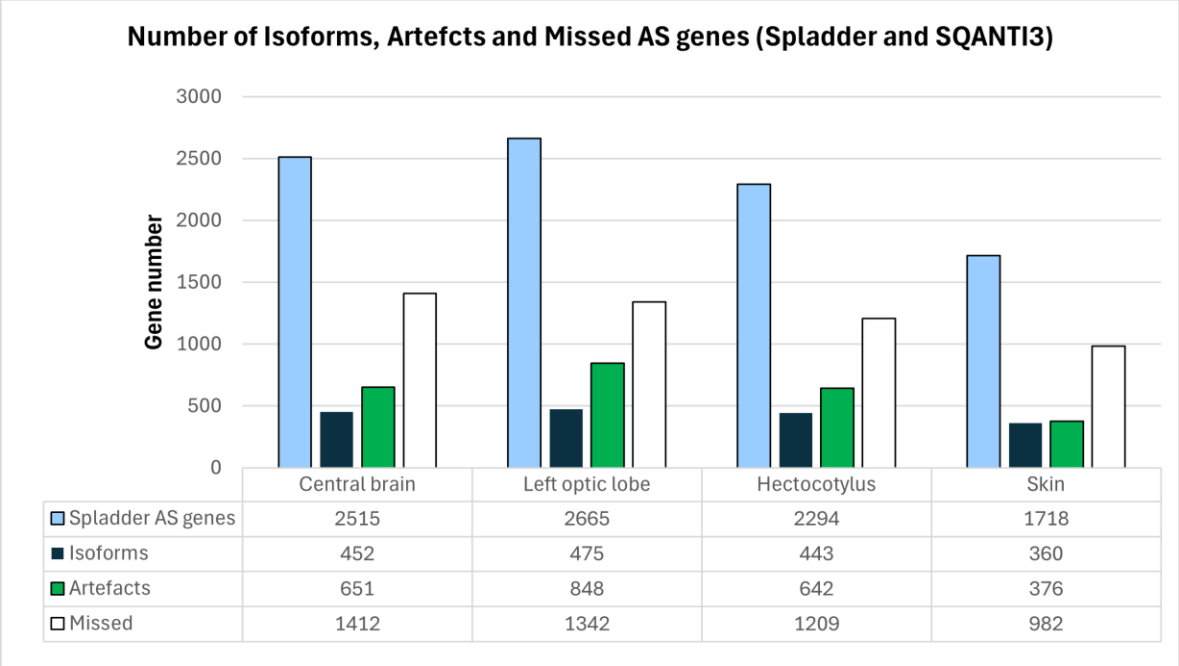
Skepticism often surrounds AS studies due to concerns about reliability and potential transcriptional noise. Numerous publications address this perceived unreliability by acknowledging the uncertainty surrounding current methods and emphasizing the need to reduce biases and ensure robust, well-grounded results (Hiller et al., 2006; Jones et al., 2024; Wan et al., 2018). To mitigate these challenges, predictions from several bioinformatics tools (Denti et al., 2020; Hesham et al., 2020; Ding et al., 2017) are used to comprehensively assess AS events. while avoiding overly stringent criteria that could lead to false negatives (Type II errors).

In this section, a novel bioinformatics strategy is proposed, by combining Spladder which utilizes short-read RNA-seq data along with SQANTI3, which accepts long-reads as an input and employs machine learning to categorize isoforms based on genome alignment. This combination aimed to enhance AS analysis precision by categorizing isoforms into specific structural categories. This filtering approach defines a gene as truly AS based on two criteria: first, it must be expressed and identified as alternatively spliced by Spladder (after the additional filtering using featureCounts), and second, that at least one isoform for that gene must be predicted by SQANTI3. Conversely, if SQANTI3 determines that all isoforms of an alternative gene are artifacts, then it is not considered as truly AS. The results are illustrated at Figure 8.

The total AS genes per tissue (see Table 1) , before their overlap with SQANTI3's prediction, are shown in the light blue bar. The Isoforms category (dark blue bar) represents AS genes that were validated with at least one isoform by SQANTI3, while the category of Artifacts (green bar) includes genes identified by

Spladder as undergoing alternative splicing but categorized as artifacts by SQANTI3 prediction. Artifacts exceeded in number the isoforms in most tissues except in the case of skin, where numbers are nearly equal.

Lastly, the total of genes evaluated as alternatively spliced from Spladder, but missed from SQANTI3 prediction, are depicted in the white bar. In summary, SQANTI3's stringent filtering seems to not consider a significant number of AS genes identified by Spladder, potentially missing genuine AS events.



	Central brain	Left optic lobe	Hectocotylus	Skin
Spladder AS genes	2515	2665	2294	1718
Isoforms	452	475	443	360
Artefacts	651	848	642	376
Missed	1412	1342	1209	982

Figure 9. *Visually representation and enumeration of the results obtained from the combination of Spladder and SQANTI3 analyses. The first row illustrates the total count of AS genes per tissue, identified by Spladder (post featureCounts filtering). The subsequent three rows present: 1) the total count of genes identified as, AS by both tools, indicating genes with isoforms; 2) genes labeled as, AS spliced by Spladder but lacking isoforms according to SQANTI3; and 3) genes identified as, AS by Spladder but not captured by SQANTI3, thus not categorized as isoforms or artefacts.*

5. Discussion

5.1 Alternative splicing is more prevalent in neurated tissues

Numerous projects focused on AS face criticism stating that many identified events may represent transcriptional noise rather than genuine splicing. Researchers argue that a significant portion of detected events could be erroneous (Tress et al 2017) or represent splice errors (Bénitière et al 2024), resulting in non-functional protein products. At the bioinformatics level, distinguishing true AS events from noise (Wan et al 2018; Busch et al 2015) involves rigorous filtering based on criteria such as read coverage, or junction-spanning reads, to identify consistent patterns. In addition, since splicing requires gene expression, an additional filtering step needs to be implemented to exclude genes near the expression threshold that may not be reliably expressed.

In line with the above, the analysis of the *E. scolopes* splicing profile provided reliable results. Tissues were categorized into three groups: brain (central brain, left optic lobe), hybrid (hectocotylus), and somatic (left gills, mantle, skin). Brain tissues exhibited the highest total and unique number of AS genes as well as derived events, followed by the hectocotylus, with somatic tissues showing the lowest number, particularly in the left gills.

This project further marks the dynamic regulation of AS in neurally enriched tissues, a common event across different metazoans, reinforcing the role of transcriptional variation in the regulation of neural tissues (Raj & Blencowe, 2015; Furlanis & Scheiffele, 2018), as well as their need for functional diversity and regulatory plasticity (Baralle & Giudice, 2017; Lipscombe, 2005). Furthermore, previous studies (Hao Su et al. 2018) highlight that the prevalence of AS in neural tissues enhances synaptic transmission, while the high number of AS genes in hectocotylus exhibits a decentralized control of nervous system, with various neuronal processing occurring in this organ.

5.2 Two modes occur in relatively greater proportions in *E.scolopes* unique splicing profile

The constructed splicing profile of *E. scolopes* (Figure 2) indicates a distinct mode distribution that doesn't align with any known animal models (Sammeth et al., 2008). Different AS modes and their proportions target certain transcript processes, and their variations are attributed to genomic architecture, splicing machinery, or evolutionary pressures (Tammer et al 2022).

Exon skipping emerged as the predominant mode, consistent with patterns observed in other metazoans (Wright et al 2022), underscoring its important role in cephalopods as well. However, compared to other species, elevated proportions of 5' alternative donor and 3' alternative acceptor sites were observed too. Notably, in certain tissues such as the left gills, genes utilizing the 3' alternative acceptor site outnumbered those employing intron retention, which is typically commoner in metazoan gene regulation (Middleton et al., 2017; Schmitz et al., 2017).

These modes provide isoforms with varied N-terminal and C-terminal sequences, potentially affecting localization, stability, and interactions (Wachutka et al., 2019; Akerman et al., 2007), while studies on splicing kinetics (Baralle et al., 2017; Marasco et al., 2023) suggest they also impact differential processing rates facilitated by configurations that enhance efficiency of pre-mRNA's structural context. This fine-tuning of transcripts diverges from exon skipping, which often aims to alter isoform functions (Stucki et al 2013; Black 2003).

Thus, the squid's preference for less common AS modes may reflect a strategy to modulate splicing efficiency, allowing rapid adaptation to environmental changes without significant isoform alterations. As this is the first analysis shedding light on the splicing mechanisms of a coleoid cephalopod, future studies on related species could yield additional insights into the evolutionary significance behind cephalopod mode diversity.

5.3 Uncovering broad splicing patterns and localized differences

The seventeen samples were clustered based on their broad PSI values across all six AS modes. Brain tissue samples exhibited strong homogeneous clustering, indicating coordinated splicing regulation. The variable clustering of hectocotylus samples aligning with both brain and somatic tissues, highlighted its unique characteristics. All samples of skin and left gills, as well as two of the mantle tissue, consistently clustered together, whereas one mantle sample deviated from the others. These findings underscored tissue-specific splicing profiles and reflect different implications for biological functions and adaptations.

Even though two tissues may cluster based on similar PSI values, localized differences can lead to variability in splicing events, suggesting potential functional consequences. Considering this, the use of differential splicing analysis revealed intriguing insights. While the overall splicing patterns of brain tissues were similar, 106 differentially spliced genes between the central brain and left optic lobe were identified. This variability indicates distinct regulatory mechanisms at tissue level, highlighting the dynamic regulation of AS in brain tissues. In contrast, somatic tissues showed little to no differential splicing events, while hectocotylus exhibited minimal differences with somatic tissues but significant differences with brain tissues, suggesting specific factors influencing AS events.

5.4 Shared groups of AS genes may be indicative of common splicing regulation in gene groups

The comprehensive Venn diagram in Figure 5E showcases three different cases of identified AS sets. Initially, a core set of 696 AS genes shared by all tissues was detected. Even though protein annotation analysis results didn't reveal any specific functional enrichment, this set could possibly serve housekeeping genes or playing important role in multiple tissues (Zhu et al., 2008; Eisenberg et al., 2013). Transcriptional enrichment of broadly expressed genes could potentially reflect a balance between basic cellular processes and the flexibility required for tissue-specific functions.

Furthermore, a smaller set of 186 AS genes shared among five tissues (excluding left gills) may be proof of specific similarities in splicing behavior, reflecting the transcriptional versality and importance of this

gene set. Finally, an additional intriguing finding included a set of 311 AS genes shared only between the brain tissues, likely representing a core gene set crucial for basic brain function or may even be indicative of an ancestral synapomorphy (Mavrodiev et al., 2020; Venkatesh et al., 2001).

5.5 Hectocotylus unique splicing profile is attributes to its hybrid nature

The high number of AS genes observed in hectocotylus may couple with its role as a sex-specific organ, aligning with previous studies that identified similar organs as centers of AS, essential for sex-specific gene regulation and adaptation. Specifically, previous efforts (Rogers et al., 2021; Cordellier et al 2020) have shown that AS plays a crucial role in generating transcripts, exclusively expressed in such organs, providing sex-specific advantages.

Moreover, the high neuron count of hectocotylus could potentially contribute to its elevated number of AS genes and derived events, comparable to brain tissues, as well as to its substantial shared number of AS genes with them (Figure 5A). Protein enrichment analysis showed that AS genes of hectocotylus have similar biological functions with the brain tissues, highlighting possible functional consistency and/or reflecting potential convergent evolutionary pressures for functional versatility.

However, despite previous similarities, differential splicing analysis revealed distinct splicing patterns of hectocotylus compared to brain tissues, indicating event variability. These localized differences may be due to the specialized reproductive demands of male individuals, or even due to the organ's rapid regeneration capabilities compared to the other arms (Bello et al., 2020; Imperadore et al., 2018).

Thus, further analysis is needed to not only shed light to the functionality of hectocotylus' isoforms but to also decipher the structure of a regular arm's splicing profile, to clarify whether these unique splicing patterns are indeed specific to hectocotylus.

5.6 Protein annotation analysis

The production of multiple protein isoforms from AS can be beneficial for adapting to various environmental needs without requiring new genes to evolve. This flexibility is particularly advantageous

for complex tissues such as the nervous system, which demands a high degree of functional versatility (Barbosa-Morais et al., 2012). Protein annotation analysis revealed significant functional enrichments (Table 5) of genes undergoing AS in brain tissues, with key domains associated with protein interactions, DNA binding, and energy transfer (Licatalosi & Darnell, 2010) being more prevalent in neurated tissues compared to somatic tissues, further distinguishing their crucial tissue-specific roles.

Many studies on other model organisms have identified AS events that lead to functional proteins or have the potential to be translated (Weatheritt et al 2016; Marquez et al 2015). To better understand the functionality of the detected alternative isoforms in *E. scolopes*, further research is needed. A key next step will involve assessing in silico whether the open reading frame (ORF) of isoforms is preserved, which can help determine the likelihood of functional protein production, while wet lab approaches, such as RT-PCR for event validating or Western blotting for protein production detection (Anczuków & Krainer 2016; Vandenbroucke et al 2001), can be also employed to evaluate biological relevance of alternative isoforms.

5.7 Alternative splicing is enriched in cephalopod specific conserved regions

Since all multi-exon genes have the potential to undergo splicing, this process is largely regulated by specific cis-regulatory elements (Baralle & Giudice, 2017; Contreras et al., 2013), conserved splicing modulators and concordant motifs (Hsiao et al. 2016) that coordinate AS within certain genomic regions.

Analysis of microsyntenies in *E. scolopes* reveals significant enrichment of AS genes in cephalopod-specific microsyntenic clusters (MACIs) compared to conserved metazoan syntenies, suggesting a crucial role of isoform diversity in regions associated with cephalopod-specific evolution and innovation. Although further research is needed to define the precise synteny clusters where these AS-rich regions appear, first insights highlight an evolutionary drive towards transcriptomic complexity in MACIs, showcasing the importance of AS in the cephalopod genomic architecture.

Future efforts could involve identification of specific genomic traits in gene clusters by the generation of additional Hi-C data to map cis-regulatory elements and investigate if the localization of these regions e.g.

on chromosomal surface, govern splicing decisions. Another approach would involve exploring whether evolutionary pressures, such as positive selection, might drive cephalopod-specific innovations.

5.8 TADs borders act as alternative splicing hotspots

The spatial organization at TAD boundaries appears to be critical in AS modulation. Results from this study align with previous findings (Naftelberg, 2015; McArthur & Capra, 2021; Alharbi et al., 2021), characterizing TAD boundaries as alternative splicing hotspots, enriched with regulatory elements such us CTCF molecules, that significantly influence the splicing behavior of precursor mRNAs. Among the 70 AS genes manually observed, the vast majority were located near TAD boundaries (Figure 6 Supplementary Figures 1-6), while AS genes exhibited lower mean and median insulation score than expressed genes across all tissues (Figure 7), supporting the notion that these regions have indeed a crucial role in AS regulation.

Elements situated at TAD boundaries, like the CTCF molecules, are known to regulate AS in many ways. Notably, they are known to interact with the elongation speed of RNA polymerase II (Lang et al 2014; Fong et al 2014), modulating the inclusion or exclusion of exons. Moreover, their role in transcriptional regulation is closely related to post-splicing regulation (Phillips & Corces, 2009). CTCF binding at DNA sites within alternative exons can create a transcriptional elongation roadblock, favoring exon fate (Dujardin, 2013), while proximal CTCF binding sites influence pre-mRNA splicing decisions by mediating alternative exon or even intron inclusion (Velasco et al., 2017; Nanavaty et al., 2020; Lopez et al., 2020). In addition, histone factors like HP1α also affect AS regulation, by binding downstream of cassette exons, along with the co-localization of CTCF (Agirre et al., 2015). These findings demonstrate the reciprocal connection between spatial genome organization and AS regulation, by highlighting the role of TAD boundaries as regulatory hubs that influence splicing patterns.

Further investigation is needed to elucidate the exact forces driving these regulatory processes, as the reorganization events of the *E. scolopes* genome (Albertin & Simakov, 2020) may have altered the landscape of cis-regulatory elements (Rogers & Simakov, 2023), resulting in novel regulatory forces. For

example, recent studies (Bhat et al., 2024) claim that genes located near nuclear speckles have greater access to spliceosome concentrations leading to enhanced splicing levels, something that doesn't necessary couple with the presence of CTCF molecules. Consequently, exploring new regulatory regions is essential, as changes in TAD boundaries can affect not only the coordination of AS but also gene expression.

5.9 Gene interactions reveal different levels of splicing complexity

Correlation analysis of different interaction categories (Figure 8, Table 9) revealed no significant PSI correlation between highly interacting AS gene pairs in *E. scolopes*. The low mean and median correlation values indicate a clear lack of significant positive or negative PSI correlation, suggesting that coordinated splicing does not occur.

On the contrary, these findings imply that AS operates independently to the gene interactions within a TAD that typically facilitate co-expression, by enabling differential regulations, distinct from the mechanisms governed by shared cis-regulatory elements. In other words, spatial proximity does not necessarily lead to coordinated splicing patterns. In any case, it's been suggested that AS may be more evolutionary flexible compared to gene expression, by altering the isoform balance of constrained genes while maintaining the expression of essential isoforms (Wright et al 2022). Thus, AS may provide an alternate route to gene expression, by supplying the organisms with a manifold of patterns through transcriptional regulation, ultimately impacting phenotypic complexity.

5.10 Perfect PSI correlation of gene pairs might reveal complex functional interactions

Among the interacting gene pairs studied, only a small portion of interacting AS genes exhibited high correlation, with mean values around 0,75. In this confined group, one interacting gene pair exhibited a perfect correlation across the PSI values of all samples, indicating that their ASE are highly similar.

This pair includes genes involved in immunoglobulin production (g9258) and zinc finger C2H2-type domain (g8243). The observed interaction may propose a potential regulatory mechanism that links immune response to transcriptional regulation, or even indicate that the protein products of these genes work

together. Another explanation of the observed coordinated splicing patterns is that these genes may be subjected to similar transcriptional or post-transcriptional regulations.

However, further research is necessary to clarify the underlying mechanisms influencing this coordinated splicing behavior. Investigating their expression levels, their derived transcript variants in different conditions, as well as identifying cis-regulatory elements that might be involved in controlling their AS patterns, can provide deeper understanding.

5.11 Alternative isoform classification

Developing a comprehensive analysis framework and robust bioinformatics toolkit that minimizes biases from transcriptional noise in AS projects is essential. While SQANTI3 excels in isoform classification, introducing detailed classification categories through rigorous machine learning methods and filtering processes, it failed to capture a wide range of AS events across tissues, as either artifacts or isoforms. This suggests that tools like SplAdder, which specializes in detecting AS events and SQANTI3, measure AS events differently. Despite both tools quantifying similar numbers of AS events per tissue, indicating that incomplete gene annotation is not the only issue (even though a better gene annotation could only act in a positive way), SQANTI3's stringent algorithm may lead to false negatives. This discrepancy underscores the importance of carefully combining bioinformatics tools and the need for specialized tools that utilize long-read data to accurately support AS studies.

Closing remarks

The present analysis of AS in six tissues of *E. scolopes* revealed splicing patterns that underscore their corresponding transcriptional complexity. The prevalence of AS across different tissue types, particularly in the brain tissues and hectocotylus, highlights its role in enhancing transcriptomic diversity and potentially facilitating adaptive responses to environmental challenges or sex-specific demands, through the generation of distinct isoforms. This diversity is governed by various complex mechanisms, involving regulatory elements, splicing modulators, and chromatin architecture, which influence AS patterns and possibly contribute to the evolutionary innovation observed in cephalopods.

Additionally, the distinct evolutionary trajectory of *E. scolopes* may correspond to its unique splicing landscape. For example, the detected enrichments of AS genes in cephalopod-specific microsyntenic regions suggests an evolutionary pressure toward transcriptomic diversity, possibly driven by the demands of a changing environment and the need for functional versatility. This adaptive mechanism not only highlights the dynamic nature of gene regulation in *E. scolopes* but also implies that their evolutionary history has shaped distinct splicing patterns that could potentially diverge from those of ancestral mollusks.

Future research will involve the integration of computational tools along with experimental validations, as well as the addition of more tissues in the analysis, will be proved crucial for a deeper understanding of AS dynamics in *E. scolopes*. Addressing these complexities will further illuminate how AS shapes transcriptomic landscapes and impacts biological versatility of cephalopods, contributing to our understanding of this cellular mechanism even genome wide.

Bibliography

1. Agirre, E., Bellora, N., Alló, M., Pagès, A., Bertucci, P., Kornblihtt, A. R., & Eyras, E. (2015). A chromatin code for alternative splicing involving a putative association between CTCF and HP1α proteins. BMC Biology, 13(1), 31. https://doi.org/10.1186/s12915-015-0141-5
2. Akerman, M., & Mandel-Gutfreund, Y. (2007). Does distance matter? Variations in alternative 3′ splicing regulation. Nucleic Acids Research, 35(16), 5487–5498. https://doi.org/10.1093/nar/gkm603
3. Albertin, C. B., & Simakov, O. (2020). Cephalopod Biology: At the Intersection Between Genomic and Organismal Novelties. Annual Review of Animal Biosciences, 8(1), 71–90. https://doi.org/10.1146/annurev-animal-021419-083609
4. Albertin, C. B., Simakov, O., Mitros, T., Wang, Z. Y., Pungor, J. R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C. W., & Rokhsar, D. S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature, 524(7564), 220–224. https://doi.org/10.1038/nature14668
5. Alexis M. Bergman. (2013). Phylogeography of Sepioteuthis lessoniana (the bigfin reef squid) and Uroteuthis duvauceli (the Indian squid).
6. Alharbi, A. B., Schmitz, U., Bailey, C. G., & Rasko, J. E. J. (2021). CTCF as a regulator of alternative splicing: new tricks for an old player. Nucleic Acids Research, 49(14), 7825–7838. https://doi.org/10.1093/nar/gkab520
7. Au, K. F., Jiang, H., Lin, L., Xing, Y., & Wong, W. H. (2010). Detection of splice junctions from paired-end RNA-seq data by SpliceMap. Nucleic Acids Research, 38(14), 4570–4578. https://doi.org/10.1093/nar/gkq211
8. Bandel, K., & Von Boletzky, S. (2016). Features of development and functional morphology required in the reconstruction of early coleoid cephalopods. https://www.researchgate.net/publication/284392203
9. Baralle, F. E., & Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. In Nature Reviews Molecular Cell Biology (Vol. 18, Issue 7, pp. 437–451). Nature Publishing Group. https://doi.org/10.1038/nrm.2017.27

10. Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., Slobodeniuc, V., Kutter, C., Watt, S., Çolak, R., Kim, T. H., Misquitta-Ali, C. M., Wilson, M. D., Kim, P. M., Odom, D. T., Frey, B. J., & Blencowe, B. J. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science, 338(6114), 1587–1593. https://doi.org/10.1126/science.1230612
11. Belcaid, M., Casaburi, G., McAnulty, S. J., Schmidbaur, H., Suria, A. M., Moriano-Gutierrez, S., Sabrina Pankey, M., Oakley, T. H., Kremer, N., Koch, E. J., Collins, A. J., Nguyen, H., Lek, S., Goncharenko-Foster, I., Minx, P., Sodergren, E., Weinstock, G., Rokhsar, D. S., McFall-Ngai, M., … Nyholm, S. V. (2019). Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. Proceedings of the National Academy of Sciences of the United States of America, 116(8), 3030–3035. https://doi.org/10.1073/pnas.1817322116
12. Macchi, F., Edsinger, E. & Sadler, K.C. Epigenetic machinery is functionally conserved in cephalopods. BMC Biol 20, 202 (2022). https://doi.org/10.1186/s12915-022-01404-1
13. Bello, G. (2020). Evolution of the hectocotylus in sepiolinae (Cephalopoda: Sepiolidae) and description of four new genera. European Journal of Taxonomy, 2020(655). https://doi.org/10.5852/ejt.2020.655
14. Bénitière, F., Necsulea, A., & Duret, L. (2024). Random genetic drift sets an upper limit on mRNA splicing accuracy in metazoans. ELife, 13. https://doi.org/10.7554/eLife.93629.3
15. Schmidbaur, H., Kawaguchi, A., Clarence, T. et al. Emergence of novel cephalopod gene regulation and expression through large-scale genome reorganization. Nat Commun 13, 2172 (2022). https://doi.org/10.1038/s41467-022-29694-7
16. Blekhman, R., Marioni, J. C., Zumbo, P., Stephens, M., & Gilad, Y. (2010). Sex-specific and lineage-specific alternative splicing in primates. Genome Research, 20(2), 180–189. https://doi.org/10.1101/gr.099226.109
17. Boletzky, S. V. (2003). Biology of Early Life Stages in Cephalopod Molluscs.
18. Bosch, T. C. G. (2019). Squid genomes in a bacterial world. Proceedings of the National Academy of Sciences, 116(8), 2799–2801. https://doi.org/10.1073/pnas.1822166116

19. Chen, Y., Akin, O., Nern, A., Tsui, C. Y. K., Pecot, M. Y., & Zipursky, S. L. (2014). Cell-type-Specific Labeling of Synapses In Vivo through Synaptic Tagging with Recombination. Neuron, 81(2), 280–293. https://doi.org/10.1016/j.neuron.2013.12.021
20. Chénais, B., Caruso, A., Hiard, S., & Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. Gene, 509(1), 7–15. https://doi.org/10.1016/j.gene.2012.07.042
21. Cloney, R. A., & Florey, E. (1968). Ultrastructure of cephalopod chromatophore organs. Zeitschrift Fur Zellforschung Und Mikroskopische Anatomie, 89(2), 250–280. https://doi.org/10.1007/BF00347297
22. Cocquet, J., Chong, A., Zhang, G., & Veitia, R. A. (2006). Reverse transcriptase template switching and false alternative transcripts. Genomics, 88(1), 127–131. https://doi.org/10.1016/j.ygeno.2005.12.013
23. Contreras, L. M., Huang, T., Piazza, C. L., Smith, D., Qu, G., Gelderman, G., Potratz, J. P., Russell, R., & Belfort, M. (2013). Group II intron-ribosome association protects intron RNA from degradation. RNA, 19(11), 1497–1509. https://doi.org/10.1261/rna.039073.113
24. Crick, F. (1970). Central Dogma of Molecular Biology. In NATURE (Vol. 227).
25. Mary E Grantham, Jennifer A Brisson, Extensive Differential Splicing Underlies Phenotypically Plastic Aphid Morphs, Molecular Biology and Evolution, Volume 35, Issue 8, August 2018, Pages 1934–1946, https://doi.org/10.1093/molbev/msy095
26. Ding, L., Rath, E., & Bai, Y. (2017). Comparison of Alternative Splicing Junction Detection Tools Using RNASeq Data. Current Genomics, 18(3), 268–277. https://doi.org/10.2174/1389202918666170215125048
27. Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature, 485(7398), 376–380. https://doi.org/10.1038/nature11082
28. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

29. Dujardin, G., Lafaille, C., Petrillo, E., Buggiano, V., Gómez Acuña, L. I., Fiszbein, A., Godoy Herz, M. A., Nieto Moreno, N., Muñoz, M. J., Alló, M., Schor, I. E., & Kornblihtt, A. R. (2013). Transcriptional elongation and alternative splicing. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1829(1), 134–140. https://doi.org/10.1016/j.bbagrm.2012.08.005
30. Eisenberg, E., & Levanon, E. Y. (2013). Human housekeeping genes, revisited. Trends in Genetics, 29(10), 569–574. https://doi.org/10.1016/j.tig.2013.05.010
31. Fu, X.-D., & Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nature Reviews Genetics, 15(10), 689–701. https://doi.org/10.1038/nrg3778
32. Garrett, A. M., Schreiner, D., Lobas, M. A., & Weiner, J. A. (2012). γ-Protocadherins Control Cortical Dendrite Arborization by Regulating the Activity of a FAK/PKC/MARCKS Signaling Pathway. Neuron, 74(2), 269–276. https://doi.org/10.1016/j.neuron.2012.01.028
33. Garrett, S., & Rosenthal, J. J. C. (2012). RNA editing underlies temperature adaptation in K+ channels from polar octopuses. Science, 335(6070), 848–851. https://doi.org/10.1126/science.1212795
34. Gelfman, S., Burstein, D., Penn, O., Savchenko, A., Amit, M., Schwartz, S., Pupko, T., & Ast, G. (2012). Changes in exon–intron structure during vertebrate evolution affect the splicing pattern of exons. Genome Research, 22(1), 35–50. https://doi.org/10.1101/gr.119834.110
35. Graveley, B. R. (2005). Mutually Exclusive Splicing of the Insect Dscam Pre-mRNA Directed by Competing Intronic RNA Secondary Structures. Cell, 123(1), 65–73. https://doi.org/10.1016/j.cell.2005.07.028
36. Hallinan, N. M., & Lindberg, D. R. (2011). Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. Genome Biology and Evolution, 3(1), 1150–1163. https://doi.org/10.1093/gbe/evr087
37. Hanlon, R. T. , and J. B. (2018). Cephalopod behaviour. Messenger.
38. Cordellier, M., Schneider, J.M., Uhl, G. et al. Sex differences in spiders: from phenotype to genomics. Dev Genes Evol 230, 155–172 (2020). https://doi.org/10.1007/s00427-020-00657-6

39. Harewood, L., & Fraser, P. (2014). The impact of chromosomal rearrangements on regulation of gene expression. Human Molecular Genetics, 23(R1), R76–R82. https://doi.org/10.1093/hmg/ddu278
40. Tian, G.G., Li, J. & Wu, J. Alternative splicing signatures in preimplantation embryo development. Cell Biosci 10, 33 (2020). https://doi.org/10.1186/s13578-020-00399-y
41. Heath-Heckman, E., & Nishiguchi, M. K. (2021). Leveraging Short-Read Sequencing to Explore the Genomics of Sepiolid Squid. Integrative and Comparative Biology, 61(5), 1753–1761. https://doi.org/10.1093/icb/icab152
42. Henning Mourier Lemche, W. J. S. (n.d.). The anatomy of Neopilina galatheae Lemche, 1957 (Mollusca Tryblidiacea).
43. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
44. https://github.com/FelixKrueger/TrimGalore
45. House, A. E., & Lynch, K. W. (2008). Regulation of Alternative Splicing: More than Just the ABCs. Journal of Biological Chemistry, 283(3), 1217–1221. https://doi.org/10.1074/jbc.R700031200
46. Hsiao, Y. H. E., Bahn, J. H., Lin, X., Chan, T. M., Wang, R., & Xiao, X. (2016). Alternative splicing modulated by genetic variants demonstrates accelerated evolution regulated by highly conserved proteins. Genome Research, 26(4), 440–450. https://doi.org/10.1101/gr.193359.115
47. Imperadore, P., & Fiorito, G. (2018). Cephalopod tissue regeneration: Consolidating over a century of knowledge. In Frontiers in Physiology (Vol. 9, Issue MAY). Frontiers Media S.A. https://doi.org/10.3389/fphys.2018.00593
48. Jin, H. J., & Kaplan, D. L. (2003). Mechanism of silk processing in insects and spiders. Nature, 424(6952), 1057–1061. https://doi.org/10.1038/nature01809
49. John Zachary Young. (1965). The central nervous system of Nautilus. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 249(754), 1–25. https://doi.org/10.1098/rstb.1965.0006

50. Jones, E. F., Haldar, A., Oza, V. H., & Lasseigne, B. N. (2024). Quantifying transcriptome diversity: a review. In Briefings in Functional Genomics (Vol. 23, Issue 2, pp. 83–94). Oxford University Press. https://doi.org/10.1093/bfgp/elad019

51. Kahles, A., Ong, C. S., Zhong, Y., & Rätsch, G. (2016). SplAdder: Identification, quantification and testing of alternative splicing events from RNA-Seq data. Bioinformatics, 32(12), 1840–1847. https://doi.org/10.1093/bioinformatics/btw076

52. Kerwin, A. H., & Nyholm, S. V. (2017). Symbiotic bacteria associated with a bobtail squid reproductive system are detectable in the environment, and stable in the host and developing eggs. Environmental Microbiology, 19(4), 1463–1475. https://doi.org/10.1111/1462-2920.13665

53. Kim, E., Magen, A., & Ast, G. (2007). Different levels of alternative splicing among eukaryotes. Nucleic Acids Research, 35(1), 125–131. https://doi.org/10.1093/nar/gkl924

54. Koenig, K. M., Sun, P., Meyer, E., & Gross, J. M. (2016). Eye development and photoreceptor differentiation in the cephalopod *Doryteuthis pealeii*. Development (Cambridge), 143(17), 3168–3181. https://doi.org/10.1242/dev.134254

55. Kröger, B., & Kr6ger, B. (2005). This content downloaded from 185.44.79.99 on Sun. Source: Paleobiology, 31(2), 253–268. https://doi.org/10.1666/03079.s1

56. Kröger, B., Vinther, J., & Fuchs, D. (2011). Cephalopod origin and evolution: A congruent picture emerging from fossils, development and molecules. BioEssays, 33(8), 602–613. https://doi.org/10.1002/bies.201100001

57. Kruse, K., Hug, C. B., & Vaquerizas, J. M. (2020). FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. Genome Biology, 21(1), 303. https://doi.org/10.1186/s13059-020-02215-9

58. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

59. Li, X. J., Friedman, M., & Li, S. (2007). Interacting proteins as genetic modifiers of Huntington disease. In Trends in Genetics (Vol. 23, Issue 11, pp. 531–533). https://doi.org/10.1016/j.tig.2007.07.007

60. Liao, Y., Smyth, G. K., & Shi, W. (2014a). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics, 30(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

61. Liao, Y., Smyth, G. K., & Shi, W. (2014b). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics, 30(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

62. Licatalosi, D. D., & Darnell, R. B. (2010). RNA processing and its regulation: Global insights into biological networks. In Nature Reviews Genetics (Vol. 11, Issue 1, pp. 75–87). https://doi.org/10.1038/nrg2673

63. Liscovitch-Brauer, N., Alon, S., Porath, H. T., Elstein, B., Unger, R., Ziv, T., Admon, A., Levanon, E. Y., Rosenthal, J. J. C., & Eisenberg, E. (2017). Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. Cell, 169(2), 191-202.e11. https://doi.org/10.1016/j.cell.2017.03.025

64. López Soto, E. J., & Lipscombe, D. (2020). Cell-specific exon methylation and CTCF binding in neurons regulate calcium ion channel splicing and function. ELife, 9. https://doi.org/10.7554/eLife.54879

65. Minghui Wang, Yongzhong Zhao, and Bin Zhang (2015) Efficient Test and Visualization of Multi-Set Intersections. *Scientific Reports* 5: 16923.

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

67. Marasco, L. E., & Kornblihtt, A. R. (2023). The physiology of alternative splicing. Nature Reviews Molecular Cell Biology, 24(4), 242–254. https://doi.org/10.1038/s41580-022-00545-z

68. Mattick, J. S., & Gagen, M. J. (2001). The Evolution of Controlled Multitasked Gene Networks: The Role of Introns and Other Noncoding RNAs in the Development of Complex Organisms. Molecular Biology and Evolution, 18(9), 1611–1630. https://doi.org/10.1093/oxfordjournals.molbev.a003951

69. Mavrodiev, E. V. (2020). Synapomorphies Behind Shared Derived Characters: Examples from the Great Apes' Genomic Data. Acta Biotheoretica, 68(3), 357–365. https://doi.org/10.1007/s10441-019-09368-6

70. Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J. J.-L., Bomane, A., Cosson, B., Eyras, E., Rasko, J. E. J., & Ritchie, W. (2017). IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biology, 18(1), 51. https://doi.org/10.1186/s13059-017-1184-4

71. Naftelberg, S., Schor, I. E., Ast, G., & Kornblihtt, A. R. (2015). Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. Annual Review of Biochemistry, 84(1), 165–198. https://doi.org/10.1146/annurev-biochem-060614-034242

72. Nagai, K., Muto, Y., Krummel, D. A. P., Kambach, C., Ignjatovic, T., Walke, S., & Kuglstatter, A. (2001). Structure and Assembly of the Spliceosomal snRNPs. Biochemical Society Transactions, 29(2), 15–26. https://doi.org/10.1042/bst0290015

73. Nam, D. K., Lee, S., Zhou, G., Cao, X., Wang, C., Clark, T., Chen, J., Rowley, J. D., & Wang, S. M. (2002). Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. Proceedings of the National Academy of Sciences, 99(9), 6152–6156. https://doi.org/10.1073/pnas.092140899

74. Nanavaty, V., Abrash, E. W., Hong, C., Park, S., Fink, E. E., Li, Z., Sweet, T. J., Bhasin, J. M., Singuri, S., Lee, B. H., Hwang, T. H., & Ting, A. H. (2020). DNA Methylation Regulates Alternative Polyadenylation via CTCF and the Cohesin Complex. Molecular Cell, 78(4), 752-764.e6. https://doi.org/10.1016/j.molcel.2020.03.024

75. O'Brien, C. E., Roumbedakis, K., & Winkelmann, I. E. (2018). The current state of cephalopod science and perspectives on the most critical challenges ahead from three early-career researchers. In Frontiers in Physiology (Vol. 9, Issue JUN). Frontiers Media S.A. https://doi.org/10.3389/fphys.2018.00700

76. PACBio. (2023). TECH NOTE Increase scRNA-seq of low-expression genes and isoforms with depletion | Increase scRNA-seq coverage of low-expression genes and isoforms with depletion

Combining PacBio ® MAS-Seq and Jumpcode Single Cell RNA Boost Kit improves isoform detection sensitivity Highlights.

77. Pan, Y. F., Wansa, K. D. S. A., Liu, M. H., Zhao, B., Hong, S. Z., Tan, P. Y., Lim, K. S., Bourque, G., Liu, E. T., & Cheung, E. (2008). Regulation of Estrogen Receptor-mediated Long Range Transcription via Evolutionarily Conserved Distal Response Elements. Journal of Biological Chemistry, 283(47), 32977–32988. https://doi.org/10.1074/jbc.M802024200
78. Pardo-Palacios, F. J., Arzalluz-Luque, A., Kondratova, L., Salguero, P., Mestre-Tomás, J., Amorín, R., Estevan-Morió, E., Liu, T., Nanni, A., McIntyre, L., Tseng, E., & Conesa, A. (2024). SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms. Nature Methods, 21(5), 793–797. https://doi.org/10.1038/s41592-024-02229-2
79. Phillips, J. E., & Corces, V. G. (2009). CTCF: Master Weaver of the Genome. Cell, 137(7), 1194–1211. https://doi.org/10.1016/j.cell.2009.06.001
80. Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K., Thurman, R. E., Cheng, Y., Gülsoy, G., Dennis, J. H., Snyder, M. P., Stamatoyannopoulos, J. A., Taylor, J., Hardison, R. C., Kahveci, T., … Gilbert, D. M. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature, 515(7527), 402–405. https://doi.org/10.1038/nature13986
81. Raj, B., & Blencowe, B. J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. In Neuron (Vol. 87, Issue 1, pp. 14–27). Cell Press. https://doi.org/10.1016/j.neuron.2015.05.004
82. Regan, J. A., Green, M., Kao, C., Powers, E., Neff, J., Strickler, J. H., McKinney, M. S., Shah, S. H., & Zhang, T. (2021). Clonal hematopoiesis association with cardiac function and mortality in patients with solid tumors. Journal of Clinical Oncology, 39(15_suppl), 10586–10586. https://doi.org/10.1200/JCO.2021.39.15_suppl.10586
83. Ritschard, E. A., Whitelaw, B., Albertin, C. B., Cooke, I. R., Strugnell, J. M., & Simakov, O. (2019). Coupled Genomic Evolutionary Histories as Signatures of Organismal Innovations in Cephalopods:

Co-evolutionary Signatures Across Levels of Genome Organization May Shed Light on Functional Linkage and Origin of Cephalopod Novelties. BioEssays, 41(12). https://doi.org/10.1002/bies.201900073

84. Robinson, J. T., Turner, D., Durand, N. C., Thorvaldsdóttir, H., Mesirov, J. P., & Aiden, E. L. (2018). Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. Cell Systems, 6(2), 256-258.e1. https://doi.org/10.1016/j.cels.2018.01.001

85. Gao, C.-H., Chen, C., Akyol, T., Dușa, A., Yu, G., Cao, B., and Cai, P. (2024). ggVennDiagram: intuitive Venn diagram software extended. iMeta 3, 69. doi: 10.1002/imt2.177.

86. Rogers, T. F., Palmer, D. H., & Wright, A. E. (2021). Sex-Specific Selection Drives the Evolution of Alternative Splicing in Birds. Molecular Biology and Evolution, 38(2), 519–530. https://doi.org/10.1093/molbev/msaa242

87. Rogers, T. F., Yalçın, G., Briseno, J., Vijayan, N., Nyholm, S. V., & Simakov, O. (2024). Gene modelling and annotation for the Hawaiian bobtail squid, *Euprymna scolopes*. Scientific Data, 11(1). https://doi.org/10.1038/s41597-023-02903-8

88. Thea F Rogers, Oleg Simakov, Emerging questions on the mechanisms and dynamics of 3D genome evolution in spiralians, *Briefings in Functional Genomics*, Volume 22, Issue 6, November 2023, Pages 533–542, https://doi.org/10.1093/bfgp/elad043

89. Rosenthal, J. J. C., & Eisenberg, E. (2023). Extensive Recoding of the Neural Proteome in Cephalopods by RNA Editing. Annual Review of Animal Biosciences, 11(1), 57–75. https://doi.org/10.1146/annurev-animal-060322-114534

90. Rouressol, L., Briseno, J., Vijayan, N., Chen, G. Y., Ritschard, E. A., Sanchez, G., Nyholm, S. V., McFall-Ngai, M. J., & Simakov, O. (2023). Emergence of novel genomic regulatory regions associated with light-organ development in the bobtail squid. IScience, 26(7). https://doi.org/10.1016/j.isci.2023.107091

91. Ruiz-Velasco, M., Kumar, M., Lai, M. C., Bhat, P., Solis-Pinson, A. B., Reyes, A., Kleinsorg, S., Noh, K.-M., Gibson, T. J., & Zaugg, J. B. (2017). CTCF-Mediated Chromatin Loops between Promoter and

Gene Body Regulate Alternative Splicing across Individuals. Cell Systems, 5(6), 628-637.e6. https://doi.org/10.1016/j.cels.2017.10.018

92. Sammeth, M., Foissac, S., & Guigó, R. (2008). A General Definition and Nomenclature for Alternative Splicing Events. PLoS Computational Biology, 4(8), e1000147. https://doi.org/10.1371/journal.pcbi.1000147
93. Schmitz, U., Pinello, N., Jia, F., Alasmari, S., Ritchie, W., Keightley, M.-C., Shini, S., Lieschke, G. J., Wong, J. J.-L., & Rasko, J. E. J. (2017). Intron retention enhances gene regulatory complexity in vertebrates. Genome Biology, 18(1), 216. https://doi.org/10.1186/s13059-017-1339-3
94. Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biology, 16(1), 259. https://doi.org/10.1186/s13059-015-0831-x
95. Shigeno, S., Andrews, P. L. R., Ponte, G., & Fiorito, G. (2018). Cephalopod brains: An overview of current knowledge to facilitate comparison with vertebrates. In Frontiers in Physiology (Vol. 9, Issue JUL). Frontiers Media S.A. https://doi.org/10.3389/fphys.2018.00952
96. Shigeno, S., Sasaki, T., Moritaki, T., Kasugai, T., Vecchione, M., & Agata, K. (2008). Evolution of the cephalopod head complex by assembly of multiple molluscan body parts: Evidence from Nautilus embryonic development. Journal of Morphology, 269(1), 1–17. https://doi.org/10.1002/jmor.10564
97. Shomrat, T., Turchetti-Maia, A. L., Stern-Mentch, N., Basil, J. A., & Hochner, B. (2015). The vertical lobe of cephalopods: an attractive brain structure for understanding the evolution of advanced learning and memory systems. In Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology (Vol. 201, Issue 9, pp. 947–956). Springer Verlag. https://doi.org/10.1007/s00359-015-1023-6
98. Simakov, O., & Kawashima, T. (2017). Independent evolution of genomic characters during major metazoan transitions. Developmental Biology, 427(2), 179–192. https://doi.org/10.1016/j.ydbio.2016.11.012

99. Singh, P., & Ahi, E. P. (2022). The importance of alternative splicing in adaptive evolution. Molecular Ecology, 31(7), 1928–1938. https://doi.org/10.1111/mec.16377

100. Stucki, M., Suormala, T., Fowler, B., Valle, D., & Baumgartner, M. R. (2009). Cryptic Exon Activation by Disruption of Exon Splice Enhancer. Journal of Biological Chemistry, 284(42), 28953–28957. https://doi.org/10.1074/jbc.M109.050674

101. Styfhals, R., Seuntjens, E., Simakov, O., Sanges, R., & Fiorito, G. (2019). In silico Identification and Expression of Protocadherin Gene Family in Octopus vulgaris. In Frontiers in Physiology (Vol. 10, Issue JAN). Frontiers Media S.A. https://doi.org/10.3389/fphys.2018.01905

102. Robert, N.S.M., Sarigol, F., Zimmermann, B. et al. Emergence of distinct syntenic density regimes is associated with early metazoan genomic transitions. BMC Genomics 23, 143 (2022). https://doi.org/10.1186/s12864-022-08304-2

103. Styfhals, R., Zolotarov, G., Hulselmans, G., Spanier, K. I., Poovathingal, S., Elagoz, A. M., De Winter, S., Deryckere, A., Rajewsky, N., Ponte, G., Fiorito, G., Aerts, S., & Seuntjens, E. (2022). Cell type diversity in a developing octopus brain. Nature Communications, 13(1), 7392. https://doi.org/10.1038/s41467-022-35198-1

104. Su, C. H., Dhananjaya, D., & Tarn, W. Y. (2018). Alternative splicing in neurogenesis and brain development. In Frontiers in Molecular Biosciences (Vol. 5, Issue FEB). Frontiers Media S.A. https://doi.org/10.3389/fmolb.2018.00012

105. Oleg Simakov et al. ,Deeply conserved synteny and the evolution of metazoan chromosomes. Sci. Adv.8, eabi5884 (2022). DOI:10.1126/sciadv.abi5884

106. Suria, A. M., Tan, K. C., Kerwin, A. H., Gitzel, L., Abini-Agbomson, L., Bertenshaw, J. M., Sewell, J., Nyholm, S. V., & Balunas, M. J. (2020). Hawaiian Bobtail Squid Symbionts Inhibit Marine Bacteria via Production of Specialized Metabolites, Including New Bromoalterochromides BAC-D/D′. MSphere, 5(4). https://doi.org/10.1128/mSphere.00166-20

107. Suzuki, R., & Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics, 22(12), 1540–1542. https://doi.org/10.1093/bioinformatics/btl117

108. Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W., & Spitz, F. (2016). The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. Developmental Cell, 39(5), 529–543. https://doi.org/10.1016/j.devcel.2016.10.015

109. Tammer, L., Hameiri, O., Keydar, I., Roy, V. R., Ashkenazy-Titelman, A., Custódio, N., Sason, I., Shayevitch, R., Rodríguez-Vaello, V., Rino, J., Lev Maor, G., Leader, Y., Khair, D., Aiden, E. L., Elkon, R., Irimia, M., Sharan, R., Shav-Tal, Y., Carmo-Fonseca, M., & Ast, G. (2022). Gene architecture directs splicing outcome in separate nuclear spatial regions. Molecular Cell, 82(5), 1021-1034.e8. https://doi.org/10.1016/j.molcel.2022.02.001

110. Thomas, P. D., Hill, D. P., Mi, H., Osumi-Sutherland, D., Van Auken, K., Carbon, S., Balhoff, J. P., Albou, L.-P., Good, B., Gaudet, P., Lewis, S. E., & Mungall, C. J. (2019). Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. Nature Genetics, 51(10), 1429–1433. https://doi.org/10.1038/s41588-019-0500-1

111. Titus, M. B., Chang, A. W., & Olesnicky, E. C. (2021). Exploring the Diverse Functional and Regulatory Consequences of Alternative Splicing in Development and Disease. In Frontiers in Genetics (Vol. 12). Frontiers Media S.A. https://doi.org/10.3389/fgene.2021.775395

112. Tress, M. L., Abascal, F., & Valencia, A. (2017). Alternative Splicing May Not Be the Key to Proteome Complexity. In Trends in Biochemical Sciences (Vol. 42, Issue 2, pp. 98–110). Elsevier Ltd. https://doi.org/10.1016/j.tibs.2016.08.008

113. Venkatesh, B., Erdmann, M. V., & Brenner, S. (2001). Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. Proceedings of the National Academy of Sciences, 98(20), 11382–11387. https://doi.org/10.1073/pnas.201415598

114. Wachutka, L., Caizzi, L., Gagneur, J., & Cramer, P. (2019). Global donor and acceptor splicing site kinetics in human cells. ELife, 8. https://doi.org/10.7554/eLife.45056

115. Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature, 456(7221), 470–476. https://doi.org/10.1038/nature07509

116. Wang, Z. Y., & Ragsdale, C. W. (2017). Cadherin genes and evolutionary novelties in the octopus. In Seminars in Cell and Developmental Biology (Vol. 69, pp. 151–157). Elsevier Ltd. https://doi.org/10.1016/j.semcdb.2017.06.007

117. Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D., & Au, K. F. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. F1000Research, 6, 100. https://doi.org/10.12688/f1000research.10571.2

118. Wessler, S. R. (2006). Transposable elements and the evolution of eukaryotic genomes. Proceedings of the National Academy of Sciences, 103(47), 17600–17601. https://doi.org/10.1073/pnas.0607612103

119. Wollesen, T., Rodríguez Monje, S. V., McDougall, C., Degnan, B. M., & Wanninger, A. (2015). The ParaHox gene Gsx patterns the apical organ and central nervous system but not the foregut in scaphopod and cephalopod mollusks. EvoDevo, 6(1). https://doi.org/10.1186/s13227-015-0037-z

120. Wright, A. H., & Laue, C. L. (2021). Evolvability and complexity properties of the digital circuit genotype-phenotype map. Proceedings of the Genetic and Evolutionary Computation Conference, 840–848. https://doi.org/10.1145/3449639.3459393

121. Wright, C. J., Smith, C. W. J., & Jiggins, C. D. (2022). Alternative splicing as a source of phenotypic diversity. Nature Reviews Genetics, 23(11), 697–710. https://doi.org/10.1038/s41576-022-00514-4

122. Xing, Y., & Lee, C. (2007). Relating Alternative Splicing to Proteome Complexity and Genome Evolution (pp. 36–49). https://doi.org/10.1007/978-0-387-77374-2_3

123. Yoshida, M. aki, Ishikura, Y., Moritaki, T., Shoguchi, E., Shimizu, K. K., Sese, J., & Ogura, A. (2011). Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. Gene, 483(1–2), 63–71. https://doi.org/10.1016/j.gene.2011.05.027

124. Zhu, J., He, F., Hu, S., & Yu, J. (2008). On the nature of human housekeeping genes. Trends in Genetics, 24(10), 481–484. https://doi.org/10.1016/j.tig.2008.08.004

125. Wan, Y., Larson, D.R. Splicing heterogeneity: separating signal from noise. *Genome Biol* 19, 86 (2018). https://doi.org/10.1186/s13059-018-1467-4

126. Busch A & Hertel K: Splicing predictions reliably classify different types of alternative splicing. *RNA* 21: 813-823 (2015). doi:10.1261/rna.048769.114

127. McArthur E, Capra JA. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. Am J Hum Genet. 2021 Feb 4;108(2):269-283. doi: 10.1016/j.ajhg.2021.01.001. PMID: 33545030; PMCID: PMC7895846.

128. Olga Anczuków & Adrian R. Krainer: Splicing-factor alterations in cancers. *RNA* 22: 1285-1301 (2016). doi: 10.1261/rna.057919.116

129. Ina I. Vandenbroucke, Jo Vandesompele, Anne De Paepe, Ludwine Messiaen, Quantification of splice variants using real-time PCR, *Nucleic Acids Research*, Volume 29, Issue 13, 1 July 2001, Page e68, https://doi.org/10.1093/nar/29.13.e68

130. Zhang, Y., Qian, J., Gu, C. et al. Alternative splicing and cancer: a systematic review. Sig Transduct Target Ther 6, 78 (2021). https://doi.org/10.1038/s41392-021-00486-7

131. Yimeng Dou, Kristi L. Fox-Walsh, Pierre F. Baldi, and Klemens J. Hertel. Genomic splice-site analysis reveals frequent alternative splicing close to the dominant splice site. October 19, 2006, doi: 10.1261/rna.151106

132. Albertin, C.B., Medina-Ruiz, S., Mitros, T. et al. Genome and transcriptome mechanisms driving cephalopod evolution. Nat Commun 13, 2427 (2022). https://doi.org/10.1038/s41467-022-29748-w

133. Vlad C. Seitan, Andre J. Faure, Ye Zhan, Rachel Patton McCord, Bryan R. Lajoie, Elizabeth Ing-Simmons, Boris Lenhard, Luca Giorgetti, Edith Heard, Amanda G. Fisher, Paul Flicek, Job Dekker and Matthias Merkenschlager. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res. 2013 23: 2066-2077 originally published online September 3, 2013. doi:10.1101/gr.161620.113

134. Gorkin DU, Leung D, Ren B. The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell. 2014 Jun 5;14(6):762-75. doi: 10.1016/j.stem.2014.05.017. PMID: 24905166; PMCID: PMC4107214.

135. Baguette, Audrey. Characterization of the interplay between the genome architecture and gene co-expression. McGill University (Canada), 2020.

136. Hiller, Michael, Klaus Huse, Karol Szafranski, Niels Jahn, Jochen Hampe, Stefan Schreiber, Rolf Backofen, and Matthias Platzer. "Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing." The American Journal of Human Genetics 78, no. 2 (2006): 291-302.

137. Hesham, Dina, and Shahenda El-Naggar. "Transcriptomic analysis revealed an emerging role of alternative splicing in embryonal tumor with multilayered rosettes." Genes 11, no. 9 (2020): 1108.

138. Denti, Luca. "Algorithms for analyzing genetic variability from Next-Generation Sequencing data." (2020).

139. Furlanis, Elisabetta, and Peter Scheiffele. "Regulation of neuronal differentiation, function, and plasticity by alternative splicing." Annual review of cell and developmental biology 34, no. 1 (2018): 451-469.

140. Lipscombe, Diane. "Neuronal proteins custom designed by alternative splicing." Current opinion in neurobiology 15, no. 3 (2005): 358-363.

141. Su, Chun-Hao, and Woan-Yuh Tarn. "Alternative splicing in neurogenesis and brain development." Frontiers in molecular biosciences 5 (2018): 12.

142. Black, Douglas L. "Mechanisms of alternative pre-messenger RNA splicing." Annual review of biochemistry 72, no. 1 (2003): 291-336.

143. Weatheritt, Robert J., Timothy Sterne-Weiler, and Benjamin J. Blencowe. "The ribosome-engaged landscape of alternative splicing." Nature structural & molecular biology 23, no. 12 (2016): 1117-1123.

144. Marquez, Yamile, Markus Höpfler, Zahra Ayatollahi, Andrea Barta, and Maria Kalyna. "Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity." Genome research 25, no. 7 (2015): 995-1007.

145. Lang, F., Li, X., Vladimirova, O. et al. CTCF interacts with the lytic HSV-1 genome to promote viral transcription. Sci Rep 7, 39861 (2017). https://doi.org/10.1038/srep39861

146. Nova Fong, Hyunmin Kim, Yu Zhou, Xiong Ji, Jinsong Qiu, Tassa Saldi, Katrina Diener, Ken Jones, Xiang-Dong Fu, and David L. Bentley. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. 2014, 28: doi:10.1101/gad.252106.114

147. López Soto, Eduardo Javier, and Diane Lipscombe. "Cell-specific exon methylation and CTCF binding in neurons regulate calcium ion channel splicing and function." Elife 9 (2020): e54879.

148. Bhat, Prashant, Amy Chow, Benjamin Emert, Olivia Ettlin, Sofia A. Quinodoz, Mackenzie Strehle, Yodai Takei et al. "Genome organization around nuclear speckles drives mRNA splicing efficiency." Nature (2024): 1-9.