Neural microexons contain lengthened sequence and extended RNA structure between the branchpoint and splice site motif

Authors: Alexandra Randazza[1,2], *Kathryn E. Howe[1,2], *John R. McCoy[1,2], Abigail Hatfield[1], Tara Doucet-O'Hare[1,2], +Lela Lackey[1,2]

* equal contributions

+ corresponding author

Affiliations: [1]Clemson University Institute for Human Genetics, [2]Department of Genetics and Biochemistry

Running title: Sequence and structure of neural microexons

Key words: microexons, splicing, RNA structure, neural development, chick embryo

Abstract:

Microexons are extremely short exons that are highly conserved in vertebrates and are essential for normal brain development. Their small size poses a challenge for canonical exon-definition splice site recognition, which typically relies on spliceosomal interactions across standard length exons. Here, we determine the sequence and RNA structural features of neural microexons in humans in comparison with chickens. We find that both human and chicken neural microexons share intronic features that extend the functional distance between the branchpoint and the 3' splice site, including distal positioning and lengthening of polypyrimidine tracts. Using experimental RNA structure-probing on a subset of neural microexons, we show that the human branchpoint-to-splice site region is unusually accessible and relatively unpaired compared to other exon classes. However, we find little support for shared RNA secondary structures between orthologous human and chicken precursor RNAs despite high sequence conservation. Using the chick embryo as a developmental model, we demonstrate that a subset of neural microexons undergo dynamic, stage-specific regulation during brain development that correlates with expression of known microexon regulators, *SRRM4* and *NOVA1*. Our data suggest that neural microexon splicing fluctuates during brain development and may rely on an extended polypyrimidine tract and structural accessibility for splicing efficiency, rather than conserved secondary structures.

Introduction:

The brain exhibits some of the highest levels of alternative splicing in the human body, producing a diverse array of RNA isoforms, including transcripts containing microexons (Porter et al. 2018). Microexons are short exons (3-27 nucleotides (nts) in length) that are highly conserved in vertebrates (Irimia et al. 2014; Li et al. 2015; Torres-Mendez et al. 2019). Microexons are usually part of the coding sequence and retain open reading frames, frequently resulting in small amino acid insertions in proteins. These insertions can alter the protein's

41 binding affinities and partners (Irimia et al. 2014; Mackensen and Irimia 2025). Specific
42 microexons have been linked to neurological disorders and development (Mackensen and Irimia
43 2025), including a microexon in cytoplasmic polyadenylation element binding protein 4 (*CPEB4*)
44 that broadly regulates translational efficiency and is associated with autism (Parras et al. 2018;
45 Olla et al. 2023; Garcia-Cabau et al. 2025), and a microexon in roundabout guidance receptor 1
46 (*ROBO1*) that influences neural migration during brain development (Johnson et al. 2019).
47 Changes in microexon splicing can be caused by misregulation of the splicing factor
48 serine/arginine repetitive matrix 4 (SRRM4) protein and have been associated with a variety of
49 neurological disorders, such as autism and schizophrenia (Irimia et al. 2014; Quesnel-Vallieres
50 et al. 2016; Gandal et al. 2018). Defining the mechanisms that regulate microexon splicing are
51 essential for understanding early brain development and neurodevelopmental disorders.
52
53 Splicing depends on both sequence and structural elements to guide splice site selection and
54 promote efficiency (Lee and Rio 2015; Shenasa and Hertel 2019; Herbert et al. 2023). For
55 example, hairpin formation at the 5' splice site of the microtubule associated protein tau (*MAPT*)
56 precursor RNA can inhibit spliceosomal recognition and is essential for appropriate splice site
57 recognition and balance of mature mRNA isoforms (Kumar et al. 2022). CryoEM structures of
58 early spliceosome assembly on hairpin containing precursor RNA in yeast reveal that hairpins
59 can also promote splicing (Li et al. 2019). Our previous work studying the secondary structure
60 of precursor RNAs found a key role for RNA flexibility in splicing fidelity, with mis-spliced RNAs
61 much more flexible than robustly spliced RNAs (Herbert et al. 2025). However, the structure of
62 most precursor RNAs, including precursor RNAs containing microexons, remains unknown.
63 Furthermore, how structural and sequence elements interact to influence splice site selection is
64 unclear.
65

66 Mechanistically, most vertebrates rely on the exon-definition model of splicing. This model
67 presupposes that most vertebrate exons are 120-150 nts in length, making them both more
68 consistent and significantly shorter than most vertebrate introns at 1,100-1,700 nts (Gazave et
69 al. 2007; Gelfman et al. 2012). In the exon-definition model, three splice sites are identified,
70 starting with the original 5' splice site and the downstream 5' splice site. This downstream 5'
71 splice site assists with identification of the 3' splice site, before the spliceosome remodels into
72 an intron-spanning complex for splicing with the original 5' splice site and the newly identified 3'
73 splice site (De Conti et al. 2013; Li et al. 2019; Zhang et al. 2024). Exon definition is assisted by
74 members of the family of serine-arginine rich RNA binding proteins (RBPs), including serine and
75 arginine rich splicing factor 1 (SRSF1), that can bind exon enhancer elements or be recruited by
76 the U1 small nuclear ribonucleoprotein (snRNP) component of the spliceosome (Jobbins et al.
77 2022). In addition to sequence elements, RNA structure can alter the distance between splice
78 sites or regulatory elements (Saha et al. 2021; Carranza et al. 2022; Saha and Ghosh 2022),
79 potentially influencing intron or exon definition splicing. Given their small size, microexons are
80 typically thought to fall outside the bounds of conventional exon-definition splicing (Ustianenko
81 et al. 2017).
82
83 Here we analyze the properties of neural-specific microexons in humans and in chickens.
84 Neural developmental milestones in the chicken embryo are well characterized, providing a unique
85 opportunity to explore microexon conservation in a developmental context (Williams et al. 2022).
86 We find that both human and chicken neural microexons have extended intronic polypyrimidine
87 tracts that push the predicted branch-point farther into the intron. In addition, our experimentally-
88 based structure models show that human microexons have less base-pairing between the branch-
89 point and 3' splice site than the same region in standard-size exons. Our results suggest that
90 despite their small size, microexons may be capable of using exon-definition for splice site

91 selection. Using the chicken model system for analysis of neurological development, we show that
92 conserved microexons have dynamic temporal regulation during embryonic brain development that
93 correlates with expression of known microexon regulators, neuro-oncological ventral antigen 1
94 *(NOVA1)* and *SRRM4*.
95
96 **Results:**
97 **Neural microexons are highly conserved**
98 We classified microexons into two groups using the VastDB splicing database (Tapial et al.
99 2017): those with altered splicing in neural tissues and those without, termed "neural
100 microexons" and "non-neural microexons", respectively. We found that the intronic regions of
101 neural microexons are significantly more conserved in vertebrates than the intronic regions of
102 non-neural microexons or moderately sized exons (100-200 nts, midexons, Figure 1A, Supp
103 Figure S1). Intronic conservation is specific to the microexon itself and not the precursor RNA in
104 general as this intronic conservation is not apparent at intronic regions upstream or downstream
105 of microexons (Supp Figure S2A). Neural microexon splicing can be identified and separately
106 clustered from non-neural microexons in humans, chickens, and mice (Figure 1B, Supp Figure
107 S3). Microexons with neural regulation are frequently shared between species, with 35.7% of
108 human neural microexons also displaying neural regulation in chicken and mouse microexons
109 (Supp Figure S2B). This contrasts with midexons where shared neural regulation is not common
110 between species with only 2.1% of human neural midexons also displaying neural regulation in
111 chicken and mouse (Supp Figure S2B). Our results support prior studies showing that neural
112 microexons and their proximate intronic regions are highly conserved in vertebrates (Irimia et al.
113 2014; Torres-Mendez et al. 2019).
114
115 We analyzed the overall length of the introns flanking neural microexons in contrast to other
116 classes of exon. We found that introns flanking neural microexons are significantly shorter than

117 introns flanking midexons in humans (Figure 2A and B) as observed in other studies (Li et al.
118 2015; Ustianenko et al. 2017). However, this was not true for chicken microexons. Overall,
119 chicken introns are much shorter than human introns, and introns flanking chicken neural
120 microexons are longer than introns flanking other classes of exon (Figure 2A-D). To further
121 investigate the difference in human and chicken intron lengths, we analyzed repetitive element
122 content and tandem repeats in the introns (Smit 1996-2010; Jurka 2000). We found that both
123 upstream and downstream introns flanking human neural microexons have significantly less
124 nucleotides from repetitive sequences than introns flanking midexons (Figure 2E and F, Supp
125 Figure S4A and B). In contrast, there was no significant difference in repeat content in introns
126 flanking chicken neural microexons compared to introns flanking other classes of exons (Figure
127 2E and F). This implies that introns around human microexons could be protected from repeat
128 increases, suggesting a functional role for introns flanking microexons.
129
130 **Neural microexons do not retain specific secondary structure elements across species**
131 To better understand the mechanisms underlying splicing regulation of neural microexons we
132 selected 13 neural microexons conserved in humans and chickens for structural studies. These
133 13 microexons were selected based on their expression levels, percent spliced-in (PSI) values,
134 and neurological disorder association (Table 1). Consistent with high vertebrate conservation at
135 microexons, the sequence similarity between humans and chickens is higher around
136 microexons than around midexons (Table 1; Supp Figure S5). We expected RNA structures to
137 be highly similar between humans and chickens, as RNA structure can maintain even more
138 information about functional conservation than sequence alone (Bernhart et al. 2008; Eddy
139 2014). Regions with conserved RNA structure can be identified through covariation analysis,
140 which detects nucleotide variation that preserves base pairing and thus maintains functional
141 RNA structures (Bernhart et al. 2008). We performed covariation and conservation analysis
142 using alignments at our 13 select microexons from 9 model species and 100 vertebrates. We

143 found evidence for covariation at 11 microexons using model species and at 6 microexons using
144 100 vertebrates (Table 1). For example, conservation and covariation analysis of a microexon in
145 *ArfGAP with GTPase domain, ankyrin repeat and PH domain 1* (*AGAP1*) suggests the presence
146 of a stem-loop encompassing the microexon (Figure 3A).
147
148 To test predicted covariant/conserved structures and compare the structural similarity of human
149 and chicken microexon precursor RNAs, we developed experimentally-based *in vitro* structural
150 models of all 13 microexons and local intronic regions. We performed both selective 2' hydroxyl
151 acylation by primer extension (SHAPE) and mutational profiling (SHAPE-MaP) and dimethyl
152 sulfide (DMS) mutational profiling (DMS-MaP) chemical probing and found high reproducibility
153 between replicates (~0.9). We also found reasonable correlation between SHAPE and DMS
154 data, (~0.4), as these methods depend on different chemical reagents (Supp Tables S1 and S2)
155 (Siegfried et al. 2014; Zubradt et al. 2017). To confirm that *in vitro* structure data was
156 representative of the RNA structure in the cell, nuclei from a medulloblastoma cell line were
157 isolated and probed with both 5NIA and DMS. *In vitro* and cellular SHAPE reactivities were
158 highly correlated (mean ~0.7) (Supp Figures S6-9, Supp Table S3). *In vitro* and cellular DMS-
159 MaP experiments were less correlated (~0.5), but likely less accurate overall as DMS only
160 provides highly accurate reactivity at adenines and cytosines, which were limited in our cellular
161 targets (Supp Table S3). We found that all SHAPE-MaP replicates as well as SHAPE-MaP
162 derived *in vitro* and in cellular comparisons were highly similar.
163
164 In *AGAP1*, we found that the nucleotides involved in the predicted covariation are maintained as
165 base-pairs in a stem-loop downstream of the microexon in representative structures for both
166 human and chicken (Figure 3B). However, predicted covarying nucleotides were not
167 consistently found in the secondary structure predictions in the other 10 microexons containing
168 potential covariation (Supp Figures S10-21). In addition, base-pairing predicted by

169 covariation/conservation did not successfully predict low SHAPE or DMS reactivity overall (Supp
170 Figure S22). We do not find evidence for covariation as a mechanism to maintain conserved
171 RNA structures around neural microexons. In addition, despite the high sequence similarity
172 between humans and chickens, such as in *AGAP1* (0.79), the overall structural correlation
173 based on the experimental reactivity data is relatively low (0.39) (Figure 3C, Supp Table S4),
174 although still significantly higher than randomized reactivities (Figure 3D). Despite the relatively
175 low overall structural correlation, the structural similarity we do see is dependent on sequence
176 similarity (Figure 3C). Supporting this, despite relatively high sequence similarity scores (>0.5),
177 the microexons in ER membrane protein complex subunit 1 (EMC1) and dynactin subunit 1
178 (DCTN1) could not be distinguished from background as structurally similar between humans
179 and chickens (Figure 3C). Our data do not suggest that microexons or their surrounding intronic
180 sequence use covariation or maintain high levels of RNA structural similarity.
181
182 We observed that human and chicken sequences frequently contain small insertions or
183 deletions, which could contribute to low structure similarity scores; therefore, we focused on
184 regions of contiguous high similarity scores to identify whether small local regions of structure
185 could be conserved between humans and chickens. We identified blocks of structural similarity
186 and found that these blocks typically include seeds of sequence similarity of at least 0.5 (Figure
187 3F; Supp Figures S10-21). *AGAP1* contains extensive structural blocks encompassing the
188 microexon and region of predicted covariation/conservation (Figure 3F). This suggests that
189 small regions of structure in the flanking introns and microexons are conserved between
190 humans and chickens. However, these structural blocks are not located in a consistent pattern
191 within the 13 tested microexons and can be upstream, downstream or in the microexon (Figure
192 3F, Supp Figures S10-21). Our data suggest that neural microexons do not have high levels of
193 structural conservation but have maintained blocks of structural similarity that primarily correlate
194 with high sequence similarity. Due to their varied localization and length within our selected

195 microexons, these regions of structural similarity likely contribute in unique ways to the
196 regulation of each individual precursor RNA.
197
198 **The branchpoint-to-splice site region in neural microexons is lengthened by an extended**
199 **polypyrimidine tract and structurally accessibly nucleotides**
200 Due to their small size, microexons have been found to require robust splicing sequence signals
201 to promote exon inclusion (Li et al. 2015). Splicing sequence signals can be compared to 5' and
202 3' splice site consensus sequences and quantified for strength (MAXENT score, (Yeo et al.
203 2004)). We found that the 3' splice sites of neural microexons were significantly weaker than the
204 3' splice sites of neural midexons in humans (Figure 4A) and in chickens (Supp Figure S23A),
205 as was also observed previously in (Irimia et al. 2014). As expected, human midexons and non-
206 neural microexons have higher polypyrimidine frequencies between 5 to 15 nts from the 3'
207 splice site, but we found that neural microexons have the highest polypyrimidine frequency
208 between 15 to 30 nts from the GU (Figure 4C). While the weaker 3' splice site around
209 microexons may be partially offset by the significantly stronger 5' splice site in microexons
210 compared to midexons (Figure 4B, Supp Figure S23B), the 3' splice site score may also not fully
211 capture the unusual properties of the polypyrimidine tract flanking neural microexons. MAXENT
212 strength is based on the presence of a strong polypyrimidine tract within 18 nts of the GU splice
213 site dinucleotide (Yeo et al. 2004). A similar trend also occurs in chicken neural microexons
214 (Supp Figure S23C). This unusual position of the polypyrimidine tract is not apparent at the
215 downstream constitutive 3' splice site for neural microexons, midexons, or non-neural
216 microexons (Figure 4D, Supp Figure S23D). The unusual position of the polypyrimidine tract in
217 microexons may not reflect their contribution to splice site selection (Yeo et al. 2004).
218
219 Based on the unusual properties of polypyrimidine tracts in neural microexons (Figure 4C), we
220 analyzed the predicted location of the branchpoint using SVM-BPfinder (Corvelo et al. 2010).

221 We found, in both humans and in chickens, that the branchpoint was predicted to be
222 significantly farther from the 3' splice site around neural microexons, (48 vs 43 nts in humans,
223 50 vs 43 nts in chicken, Figure 5A, Supp Figure S24A). We observed that human and chicken
224 neural microexons have significantly longer polypyrimidine tracts in unusually close proximity to
225 the predicted branchpoint (Figure 5B and C, Supp Figure S24B and C). In addition, the region
226 between the branchpoint and the 3' splice site has significantly higher accessibility in human
227 neural microexons compared to human neural midexons (Figure 5D). The high reactivity in the
228 branchpoint-to-splice site region indicates that this region is accessible and likely extended. In
229 contrast, we did not find any other patterns of consistent accessibility in our experimental
230 reactivity data at core elements, including the branchpoint, 3' splice site and 5' splice site (Supp
231 Figure S25). This suggests that neural microexons use a distinct and shared splicing
232 mechanism that lengthens the distance between the branchpoint and splice site through both
233 sequence and structural elements, potentially allowing for exon-definition to assist in splice site
234 identification.
235
236 **The chick embryo model reproduces regulated neural microexon splicing during brain**
237 **development**
238 We employed a chick embryo model system to analyze microexon splicing during embryonic
239 development. We dissected embryonic brain and heart tissue at five time-points (Hamburger-
240 Hamilton Stages 11, 15, 27, 31 and 35), prepared sequencing libraries and performed short
241 read sequencing. We identified microexon splicing specific to brain tissue for 12 of 13 selected
242 neural microexons. The transition between early stages HH11-15 and later stages HH27-35 is
243 especially dramatic, with nearly half of microexons sharply increasing in inclusion at this
244 developmental stage in the brain (Figure 6A; *FRY microtubule binding protein (FRY), AGAP1,*
245 *protein tyrosine phosphatase receptor type K (PTPRK), CPEB4, amyloid beta precursor protein*
246 *binding family B member 2 (APBB2)* and *DCTN1*). The inclusion of the *ROBO1* microexon

247 decreases at the transition between HH15 and HH27, consistent with its known biological role in
248 axon guidance (Johnson et al. 2019). The inclusion of microexons in dedicator of cytokinesis 7
249 (*DOCK7)* and myocyte enhancer factor 2A (*MEF2A)* remain relatively stable during early
250 development in chicken brain tissues (Figure 6A). Splicing of microexons in our select 13 genes
251 is not strongly affected by sex (Supp Figure S26). In addition, transcript levels for the 13
252 selected microexon genes are similar between males and females in both brain and heart tissue
253 (Supp Figure S27). The level of transcription is not correlated with microexon splicing in either
254 brain or heart (Supp Figure S28). We find strong stage-dependent splicing regulation of
255 microexons during chick development similar to expected stage-specific splicing regulation of
256 human microexons during brain development based on limited evidence (Gonatopoulos-
257 Pournatzis and Blencowe 2020).
258
259 At the transition point between HH15 and HH27, where we identified dynamic microexon
260 regulation (Figure 6A), we also see dramatic increases in expression of the RNA binding
261 proteins *NOVA1* and *SRRM4* in the brain at the RNA level, followed by a slow decrease from
262 HH27 to HH31 and HH35 (Figure 6B). NOVA1 and SRRM4 are classic microexon regulatory
263 factors that influence human microexon splicing (Mackensen and Irimia 2025). *SRRM4* is
264 expressed 50-fold higher in brain and *NOVA1* is expressed 6-fold higher in brain than in heart
265 tissue (Figure 6B and C; Supp Figure S29). *RNA binding fox-1 homolog 1 (RBFOX1)* and *RNA*
266 *binding fox-1 homolog 3 (RBFOX3)* may be relevant for neural specific splicing as they also
267 have higher expression in brain than in heart tissue and increase at the HH27 transition. Other
268 splicing factors, *polypyrimidine tract binding protein 1 (PTBP1), serine and arginine rich splicing*
269 *factor 11 (SRSF11)* and *RNA binding protein with serine rich domain 1 (RNPS1)*, may not be
270 related to neural microexon specific regulation as they have similar RNA expression in both
271 brain tissues and heart tissue (Supp Figure S29).
272

273 **Discussion**

274 Microexons present a conceptual challenge for canonical models of splice site recognition due
275 to their extremely small size. The exon-definition model, which relies on coordinated recognition
276 of splice sites across an exon, is incompatible with microexons because they lack sufficient
277 sequence length to physically accommodate cross-exon interactions (Li et al. 2019). In this
278 study, we identify factors that contribute to lengthening of the crucial branchpoint-to-splice site
279 region including 1) distal placement of the polypyrimidine tract farther into the intron (Figure 4C)
280 and 2) extension of the polypyrimidine tract sequence (Figure 5A and B). In addition, we find
281 that 3) the branchpoint-to-splice site region is highly accessible and likely unpaired in neural
282 microexons (Figure 5D). As exon-definition is the predominant model in vertebrates with long
283 introns (De Conti et al. 2013), the sequence and structural lengthening of neural microexons
284 may allow them to take advantage of this splicing mechanism to aid in robust splicing despite
285 their small size.

286

287 Due to shared sequence content, we were surprised to find little evidence that neural
288 microexons rely on conserved RNA secondary structures maintained through covariation.
289 Although covariation analysis predicted conserved base-pairing around many of the selected
290 microexons, these predictions were not supported by experimental SHAPE reactivity data (Supp
291 Figure S22). Moreover, structural similarity between orthologous human and chicken precursor
292 RNAs was relatively low, even in cases of high sequence similarity (Figure 3C). The lack of
293 strong structural or accessibility conservation suggests that, for neural microexons, large
294 regions of specific structure are not strictly required for function across species. We find
295 evidence for small local conserved structures at unique positions in each neural microexon we
296 analyzed. These local structures, in blocks of conserved sequence, may be adequate for
297 functional regulation. In addition, other structural properties, such as decreased pairing within
298 the extended branchpoint-to-splice site region, may be critical instead of specific structural

299 motifs. Future studies will be important for directly testing the mechanistic role of the extended
300 branchpoint-to-splice site region in splicing efficiency.
301
302 We are interested in functional roles for neural microexons; however, many of these genes are
303 difficult to study as they are involved in embryonic brain development (Irimia et al. 2014;
304 Gonatopoulos-Pournatzis and Blencowe 2020). The chick embryo is a well-established
305 developmental model system (Williams et al. 2022). We harvested brain and heart tissue from
306 five embryonic developmental stages. We find that the majority of our selected neural
307 microexons undergo dynamic, stage-specific regulation during brain development. The timing of
308 increased microexon inclusion coincides with elevated expression of known microexon
309 regulators, including SRRM4 and NOVA1, supporting a conserved regulatory program across
310 vertebrates. Microexon splicing is regulated at the splicing level during brain development and
311 not with transcriptional changes to the microexon gene as changes in microexon
312 inclusion/exclusion are not correlated with gene expression. The chick embryo is an accessible
313 developmental system that allows for perturbation of regulation, including splicing, and analysis
314 of downstream impact on morphology and function. Future studies on the role of broad changes
315 to microexon inclusion, such as through SRRM4 knock-down, and targeted changes to specific
316 microexons, such as microexon exclusion within *ROBO1*, are exciting potential future
317 experiments with this model system.
318
319 Limitations of this study should be acknowledged. Branchpoint positions were inferred
320 computationally using a program based on mammalian splicing (Corvelo et al. 2010). Although
321 human and chicken ribosomal RNA sequences are highly similar, and consistent trends were
322 observed across both species, branchpoint prediction may be less accurate for chickens. In
323 addition, our *in vitro* structural assays were performed on RNA fragments of the microexon and
324 surrounding intronic sequences. We find high correlation between our *in vitro* results and the

325 microexon RNA structure derived from cellular RNA, suggesting that the *in vitro* RNA is
326 representative of *in vivo* RNA. However, these experiments may not fully capture the influence
327 of co-transcriptional folding, RNA-binding proteins, or spliceosomal components.
328
329 In summary, our data confirm high conservation of introns surrounding microexons (Figure 1A).
330 We find that neural microexons do not share specific local structures between orthologous
331 regions in human and chicken but have unique regions of similar structural accessibility that
332 correlate with high sequence similarity (Figure 3). However, neural microexons do have similar
333 polypyrimidine tract positioning in both humans and chickens (Figure 4C, Supp Figure S23C),
334 and human neural microexons have a significantly more accessible branchpoint-to-splice site
335 region (Figure 5D). These features suggest a mechanism by which neural microexons may
336 overcome size constraints to achieve efficient and regulated splicing during neural development.
337 This work provides a foundation for future mechanistic studies of microexon splicing regulation.
338
339 **Materials and Methods**
340 **Splice site selection**
341 Scripts for analyses and plots are available on GitHub
342 (https://github.com/randazzal/Microexons_Hsa_Gga.git). VastDB tables “EVENT
343 INFORMATION” and “MAIN PSI TABLE” datasets for human (hg38), chicken (galGal4), and
344 mouse (mm10) were downloaded from VastDB (Tapial et al. 2017). For each species, the
345 VastDB events were filtered by size with microexons > 0 and <= 27 nts and midexons > 100 and
346 <= 200 nts. Exon skipping events were selected. To ensure selection of mostly alternative
347 exons, events with a percent spliced in (PSI) <= 5% and >= 95% in 4 or more samples were
348 removed. Events with 10 or more reads in at least half the samples and 20 or more reads in at
349 least a fifth of the samples were considered high confidence and used for further analysis. The
350 samples from VastDB were divided into neural and non-neural categories. Welch’s t-test was

351 used to identify exons that had significantly higher or lower PSI values in neural samples. After
352 adjustment with Benjamini–Hochberg, events with p-values < 0.005 were classified as “neural”
353 while those not meeting this threshold were classified as “non-neural”
354 (event_filtering_criteria.R). UCSC LiftOver (Hinrichs et al. 2006) was used to update the galGal4
355 coordinates to galGal6.
356
357 **Exon and intron characterization**
358 Conservation was measured using phastCons scores (Siepel et al. 2005) from UCSC Table
359 Browser (assembly: hg38, group: Comparitive Genomics, track: UCSC 100 Vertebrates, table:
360 Cons 100 Verts (phastCons100way) (Karolchik et al. 2004)
361 (Figure_1/PhastCons_A3ss_sample.R). Splice site strength was calculated with MaxEntScan
362 (Yeo et al. 2004) (Figure_4/splice_site_strength_human.R). Repeat content in upstream and
363 downstream introns was generated with bedtools intersect (bedtools2 v2.29.2) (Quinlan and
364 Hall 2010) and repeat annotations for hg38 and galGal6 from UCSC Table Browser (track:
365 RepeatMasker, table: rmsk) (Smit 1996-2010). Repeat content was also calculated with Tandem
366 Repeat Finder (TRF) (Benson 1999) using the recommended parameters
367 (Figure_2/repetitive_sequence_plots.R). Branchpoints were identified using SVM-BPfinder
368 (Corvelo et al. 2010) with “-s Hsap” and “-l 100” for both human and chicken due to the absence
369 of a model trained for chicken sequences. The branchpoint with the highest “svm_scr” per
370 intronic sequence was used for analysis (Figure_5/BPfinder_plots_human.R). All analyses
371 including sequence content (Figure_4/polypyrimidine_probabilities_human.R), and intron length
372 (Figure_2/human_intron_length_plots.R) were conducted in R v4.1.2.
373
374 **Human and chicken sequence alignment**
375 The VastDB table “EVENT CONSERVATION” was used to identify the chicken orthologs of
376 selected human microexons and midexons. Sequences 200 nts upstream from the 3’ splice site

377 and 200 nts downstream from the 5' splice site were used for alignment. The global alignment
378 score between orthologous flanking introns was calculated using PairwiseAligner from the
379 Bio.Align package (https://biopython.org/docs/1.76/api/Bio.Align.html;
380 Supplemental_Figures/alignment_score_seq_ident.ipynb) with default parameters. In addition,
381 the unaligned sequence identity was calculated as "matches" divided by 200. Bootstrapping was
382 performed in R v4.1.2.
383
384 **Conservation/covariation analysis**
385 Multiple alignment files of the 450 nts in each direction from the 3' splice site of a subset of
386 neural microexons were downloaded from the multiz100way track (track: UCSC 100
387 vertebrates, table: Multiz Align (multiz100way) (Blanchette et al. 2004). The MAF files were
388 converted to fasta files with Galaxy MAF to FASTA tool (Blankenberg et al. 2011). Specific
389 species were kept and realigned using Clustal Omega v1.2.4 (Sievers et al. 2011). Sequences
390 on the human negative strand required an in-house python script to generate the reverse
391 complement of all the sequences in the alignment (Figure_3/reverse_complement.ipynb). The
392 ViennaRNA (Lorenz et al. 2011) package RNAalifold (Bernhart et al. 2008) was utilized to
393 identify shared RNA structures between multiple species around the microexons using the
394 following command: RNAalifold -p –maxBPspan 66 –sci –aln –color <alignment_fasta>.
395
396 **Chemical structure probing**
397 Gene fragments for microexons and midexons of interest were designed and synthesized by
398 Integrated DNA Technologies (IDT). Most human microexon and all midexon sequences contain
399 450 nucleotides (nts) flanking the 3' splice site in both directions. Due to sequence constraints
400 additional human microexons (*CPEB4, DCTN1, ASAP2, MEF2A, DOCK7*) and all chicken
401 microexon sequences contain 200 nts flanking the 3' splice site in both directions. A T7 promoter
402 followed by a stabilizing hairpin sequence was on the 5' end and a universal sequence was on

403 the 3' end of each gene fragment. The HiScribe T7 High Yield RNA synthesis kit (New England
404 Biolabs, NEB) was used for *in vitro* RNA synthesis using the gene fragments acting as a
405 template following the standard protocol. The synthesized RNA was treated with DNase I for 20
406 minutes at 37°C and cleaned up with RNA XP beads (Beckman Coulter). The RNA was
407 incubated in folding buffer (400 mM bicine, 200 mM NaCl, 20 mM MgCL2), RNase inhibitor, and
408 5ul of 25 mM 5NIA or DMSO at 37°C for 5 minutes for SHAPE-MaP experiments (Siegfried et al.
409 2014). DMS-MaP experiments (Zubradt et al. 2017) used 5ul of 10% DMS in ethanol or ethanol
410 alone at 37°C for 5 minutes followed by incubation with 20% beta-mercaptoethanol for 5
411 minutes on ice. Treatment with the chemical probe was followed immediately by RNA XP bead
412 cleanup and error-prone reverse transcription (RT) using SuperScript™ II (Invitrogen,
413 ThermoFisher Scientific) with modified buffer (10X: 500 mM Tris pH 8, 750 mM KCl). 6nM
414 manganese, RNase inhibitor, and reverse primer 5'-TGTTGGAGTCACTCGACTCCGGT-3' were
415 added to the RT reaction. The RT product was cleaned up with RNA XP beads and used as the
416 template for second strand synthesis (Invitrogen Second Strand cDNA Synthesis Kit;
417 ThermoFisher Scientific). Libraries were prepared using the Nextera DNA Library Preparation
418 Kit (Illumina). The paired-end libraries were sequenced at 2 x 150 on an Illumina NovaSeq.
419 Experiments were performed in duplicate.
420
421 **Structure probing analysis**
422 Structure probing analysis was conducted with ShapeMapper v2.2.0 (Busan and Weeks 2018)
423 using default parameters and the "--dms" flag for DMS probing samples. The DMSO/EtOH
424 reads were designated as "untreated" and the 5NIA/DMS reads as "modified", respectively. The
425 Pearson correlation between replicates was calculated (Supp Table S1), and due to highly
426 similar reactivity profiles, the replicates were merged. Normalized reactivity profiles were limited
427 between 0 and 4 and utilized for downstream analysis. Representative 2D structures for human

428 and chicken sequences were generated with RNAstructure v6.5 (Reuter and Mathews 2010)
429 incorporating reactivity data and a max distance of 66 base pairs. Regions bound by primers for
430 library prep were set as single-stranded.
431
432 **Sliding Reactivity similarity**
433 Human and chicken sequence alignments from PairwiseAligner from the Bio.Align package with
434 default parameters were used to align normalized reactivity values. Spearman correlations
435 between human and chicken SHAPE reactivities were calculated for both the full sequences
436 (Supp Table S4; Figure_3/spearman_reactivity_correlation.R) and centered 15-nt sliding
437 windows (step size = 1). For each window, sequence identity was calculated as the number of
438 matching nucleotides divided by the window size (15 nts). Positions containing missing values
439 were included in the window, but windows with less than 8 valid reactivity values were skipped.
440 To determine significance, both human and chicken reactivity values were shuffled, and the
441 correlation calculations were repeated 1000 times for both the full sequences and centered 15-
442 nt sliding windows (Figure_3/observed_struct_correlation_vs_shuffled.R). A null distribution of
443 correlation run lengths were computed for each microexon with a run counting as 2 or more
444 neighboring windows having Spearman correlations > 0.2. The number of runs at any given
445 length were compiled across the 1000 iterations. The length at which < 5% of runs were longer
446 was set as the threshold for significance. Runs longer than this threshold in the unshuffled data
447 were interpreted to include regions of highly similar reactivity profiles
448 (Figure_3/identify_struct_similar_regions_agap1.R;
449 Figure_3/sequence_identity_struct_overlay_agap1.R). All correlations and plots were done in R
450 v4.1.2.
451
452 **Chicken brain and heart sample collection**

453 Tetra-H chicken (*Gallus gallus domesticus*) eggs were incubated at ~37.5 °C, monitored by
454 thermometer, for 42 hours, 56 hours, 5, 7 and 9 days at a ~50% humidity to reach Hamburger-
455 Hamilton (HH) stages 11, 15, 27, 31 and 35, respectively (Hamburger and Hamilton 1992).
456 Initially, fertilized eggs obtained from the Clemson Morgan Poultry Center (Clemson, SC) were
457 stored at 15 °C for up to one week, and before incubation eggs were left at room temperature
458 for 24 hours. Eggs were incubated in a Georgia Quail Farm (GQF) "Digital Sportsman" Model
459 1502 Incubator (Georgia Quail Farm, Cat. No. 1502) in humid conditions and were rotated
460 mechanically every hour to facilitate embryo development. To collect embryos at HH stages 11
461 and 15, eggs were opened and the contents transferred to a Petri dish. Embryos were isolated
462 from the yolk using a cloverleaf filter paper technique, and developmental stage was confirmed
463 based on established Hamburger–Hamilton morphological criteria prior to tissue processing
464 (Hamburger and Hamilton 1992; Chapman et al. 2001). The filter paper and embryo were
465 transferred to a Petri dish with isotonic saline. The vitelline membrane was surgically removed
466 to fully expose the embryo, and the heart was isolated and excised by sharp transection at its
467 superior inflow and inferior outflow connections. The brain was subsequently dissected, and
468 neural tissue encompassing the hindbrain and all more rostral regions were collected. After
469 excision, both the brain and heart tissues were immediately transferred to TRIzol reagent
470 (Invitrogen, Thermo Fisher Scientific), mechanically dissociated with pipetting, and flash frozen
471 on dry ice. Embryos at HH stage 27 and later were harvested by opening the eggshell and
472 transferring the embryo to a Petri dish containing isotonic saline. The vitelline membrane was
473 carefully removed to expose the embryo. The heart was isolated and excised via a supra-aortic
474 transection and immediately submerged in 500ul TRIzol reagent. For brain tissue collection, the
475 ocular structures were removed along with surrounding non-neural tissues to fully expose the
476 brain. Following gross dissection and cleaning of the neural tissue with saline, the brain was
477 transferred to 500ul TRIzol reagent, mechanically dissociated using sterile scissors, and flash
478 frozen. All tissue samples were stored at −80°C until RNA extraction.

479

480 **RNA extraction and library preparation**

481 The samples stored in TRIzol were homogenized with 2.4mm metal bulk beads (Fisher
482 Scientific) at 5 m/s for 60 seconds, repeating until tissue was fully homogenized. An additional
483 500ul TRIzol was added, and RNA extraction was performed following the manufacturer's
484 suggested protocol. After RNA isolation, samples were treated with DNase (Turbo DNA Free Kit,
485 Invitrogen) according to the manufacturer's instructions. Following standard ribosomal RNA
486 depletion (Qiagen QIAseq FastSelect -rRNA HMR Kit or NEBNext rRNA Depletion Kit v2), cDNA
487 was synthesized using the SuperScript™ IV First-Strand Synthesis System (Invitrogen,
488 ThermoFisher Scientific) and Second Strand cDNA Synthesis Kit (Invitrogen, ThermoFisher
489 Scientific) both according to the standard protocol. Libraries were prepared using the Nextera
490 DNA Library Preparation Kit (Illumina). The paired-end libraries were sequenced at 2 x 150 on
491 the Illumina NovaSeq. Samples were collected in triplicate per sex and per tissue.

492

493 **RNA sequencing analysis**

494 Read quality was checked with FastQC v0.11.9 (https://github.com/s-andrews/FastQC). Reads
495 were then mapped to the galGal6 genome assembly with STAR v2.7.10a (Dobin et al. 2013)
496 (Figure_6/star.sh) followed by filtering for uniquely mapped reads with Samtools v1.10 (Li et al.
497 2009) (Figure_6/mapq_filt.sh). Aligned files were sorted and indexed with Samtools and gene
498 expression was quantified with featureCounts from Subread v1.6.4 (Liao et al. 2014) followed by
499 analysis with edgeR (Robinson et al. 2010) (Figure_6/edgeR_analysis_plots.R). Splicing
500 analysis was performed with rMATs v4.1.1 (Wang et al. 2024) (Figure_6/rmats_psi_heatmap.R)
501 comparing splicing patterns between male and female derived tissues and between consecutive
502 developmental stages for each tissue. Skipped exon events <= 27nts in length were included for
503 further analysis with a special focus on the selected 13 microexons.

Acknowledgments

The authors want to acknowledge Dr. Susan Chapman for her valuable mentorship, training, and equipment sharing which facilitated the chicken embryo work in this manuscript. In addition, the authors wish to acknowledge the Institute for Human Genetics Molecular Core and the Institute for Human Genetics Bioinformatics Core, which supported the generation of the sequencing libraries and analysis presented in this work.

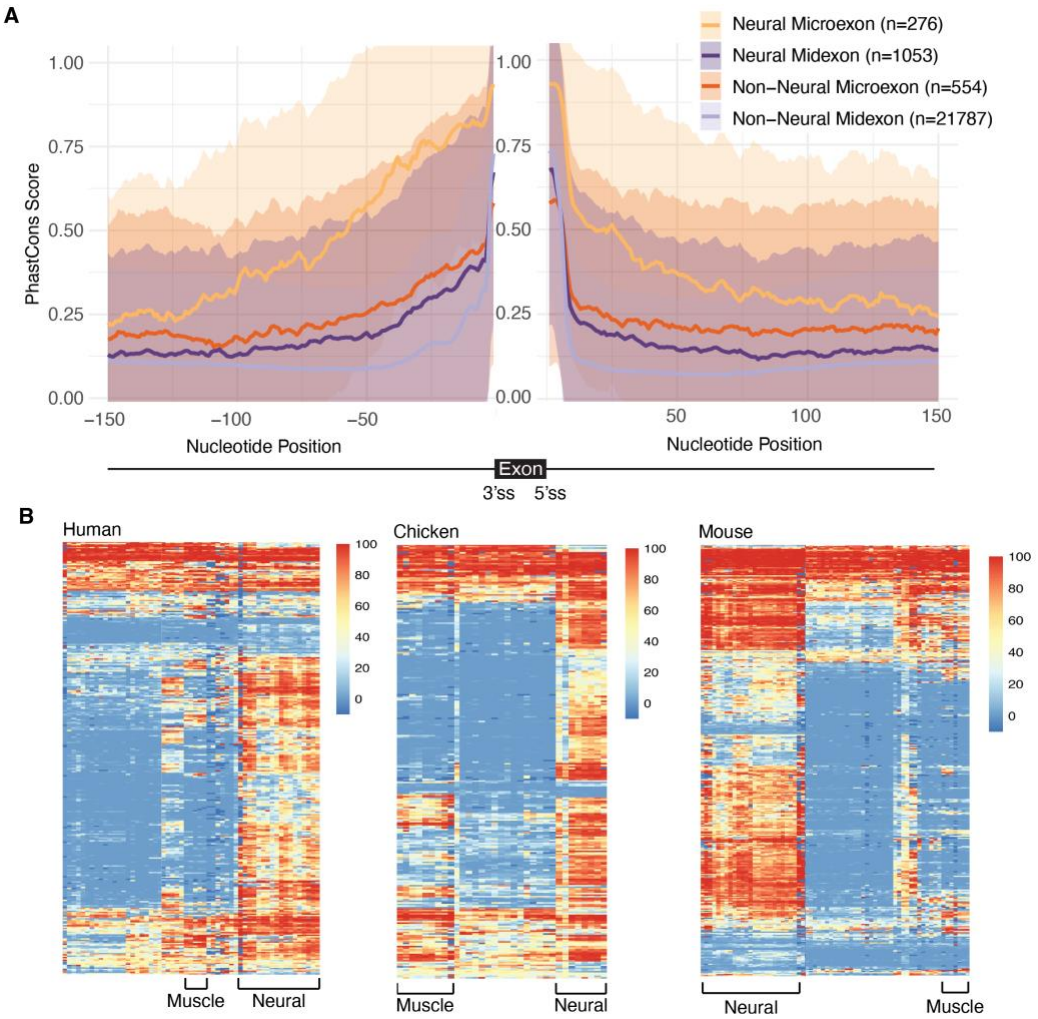
Data availability statement

Scripts used for data analysis and plotting are available at https://github.com/randazzal/Microexons_Hsa_Gga.git. Raw sequencing reads from embryonic chicken heart and brain tissues will be deposited to the gene expression omnibus. Other data available upon request.

Funding

This work was supported by the National Institutes of Health (NIH) and National Institute of General Medical Sciences grant (R35 GM142851) to Lela Lackey. Experiments used resources in the Institute for Human Genetics Molecular Core and Bioinformatics Core, which are supported by a COBRE in Human Genetics 1P20GM139769 awarded to Dr. Trudy F. C. Mackay and Dr. Robert R. H. Anholt from the National Institute of General Medical Sciences of the National Institutes of Health.

526 **Figures**



527

528 **Figure 1. Neural microexons are flanked by highly conserved intronic sequences and**

529 **spliced in a tissue-specific manner.** (A) Introns flanking neural microexons are more

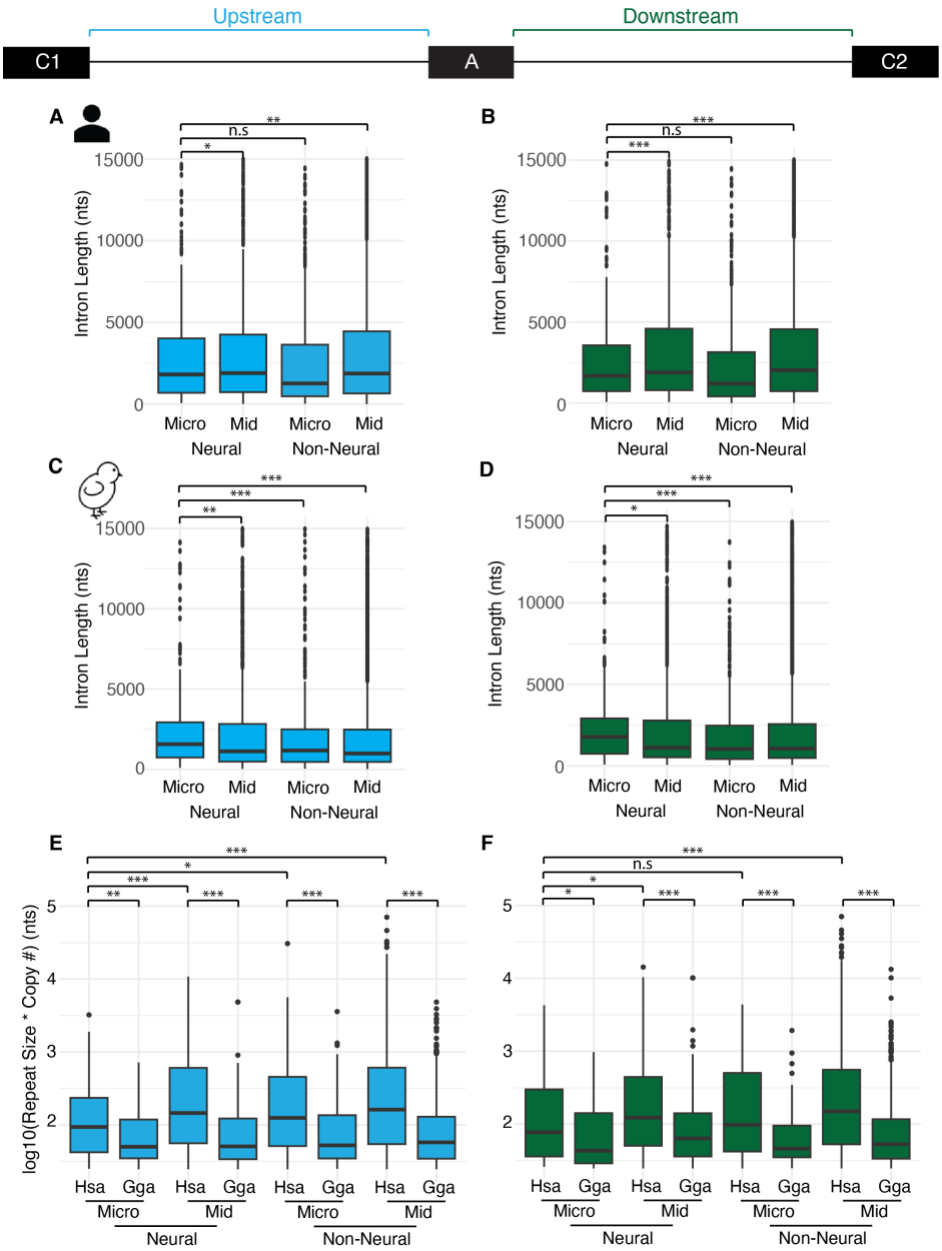
530 conserved than introns flanking other exons, as shown by high mean PhastCons scores across

531 +/- 150 nucleotides of intronic sequence. Shaded regions show +/- 1 standard deviation. (B)

532 Heatmap of microexon percent spliced in (PSI) values for a subset of human, chicken, and

533 mouse tissues from VastDB. Dark blue indicates missing values. 3'ss = 3' splice site, 5'ss = 5'

534 splice site.



535

536 **Figure 2. Short flanking intron length in human neural microexons is maintained despite**

537 **expanded repetitive elements.** (A) Human microexons have short upstream introns (blue) and

538 (B) short downstream introns (green). The upstream constitutive exon 1 (C1), alternative exon

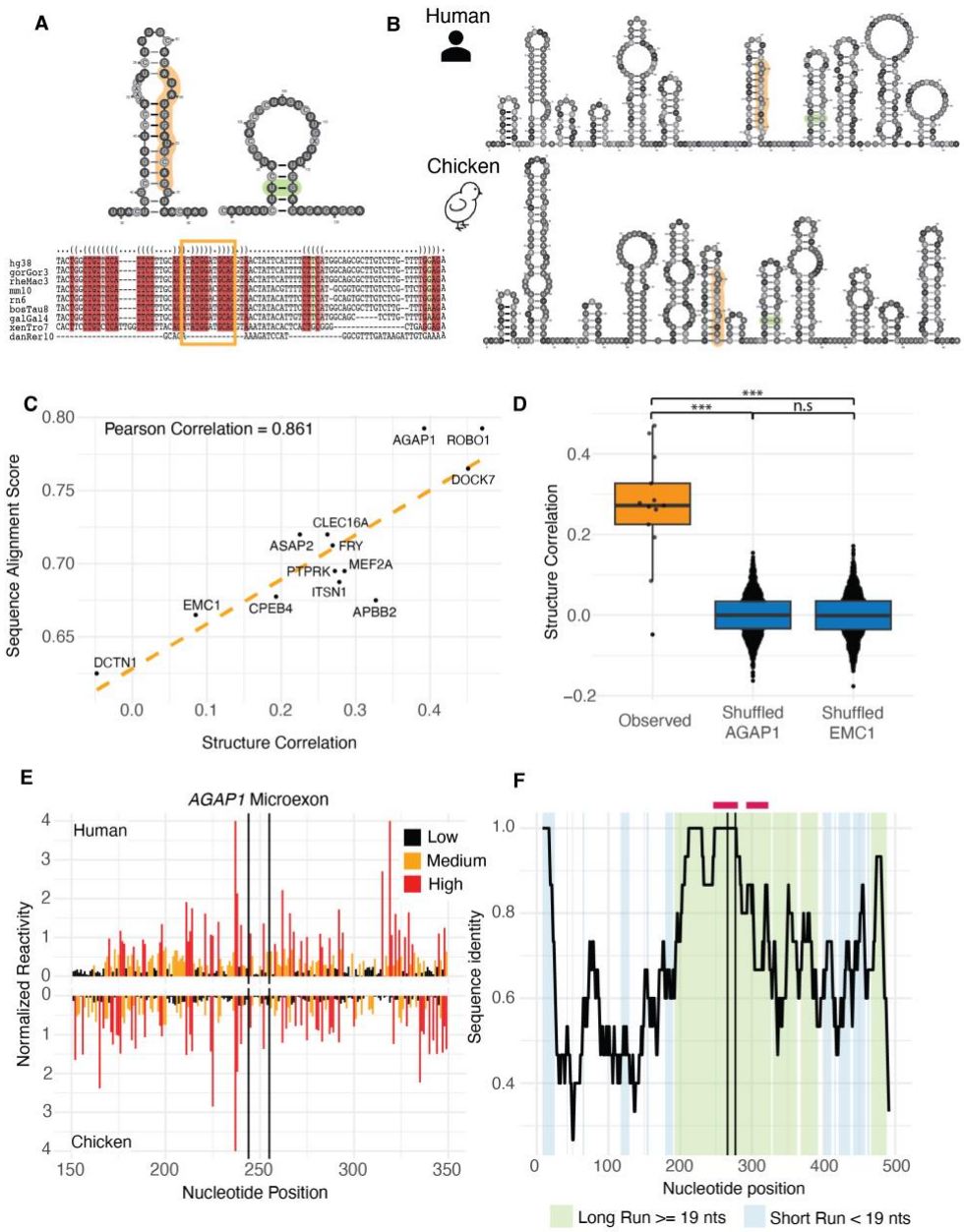
539 (A), downstream constitutive exon 2 (C1) and intronic regions are not to scale. (C) Although

540 overall smaller than human introns, chicken neural microexon upstream introns and (D)

541 downstream introns are comparatively longer than other exon classes. (y-axis display limited to

542 15,000). (E) Microexons in humans (Hsa) have less repetitive sequence than other classes of

543 human exons at both upstream and (F) downstream introns. Microexons in chickens (Gga) have
544 less repetitive sequence than in human and do not vary by exon class. (Wilcoxon signed-rank
545 test; * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.005, n.s = p-value > 0.05).
546



547

548 **Figure 3. Secondary structure depends on sequence conservation in *AGAP1* and other**
549 **microexons.** (A) Predicted *AGAP1* structure and sequence alignment based on covariation
550 (green) and conservation (red) within common model organisms. Microexon (orange) and
551 covariate positions (green) highlighted. (B) Representative structures of human (top) and
552 chicken (bottom) *AGAP1*. (C) Structural similarity between human and chicken microexons is
553 highly correlated with sequence similarity (Spearman). (D) Structural similarity between human
554 and chicken is significant as compared to background (Spearman correlation, background

555 =1000 iterations of shuffled reactivity values, Wilcoxon signed-rank test; *= p-value < 0.05, ** =
556 p-value < 0.01, *** = p-value < 0.005, n.s = p-value > 0.05). (E) Regions of similarity and
557 difference can be identified in reactivity profiles of human (top) and chicken (bottom) *AGAP1*
558 with (F) similar regions of different sizes highlighted in green and blue (>= 19 nts or < 19 nts
559 respectively). Splice sites indicated by black vertical lines. Covariation predicted structures
560 indicated by red boxes.
561

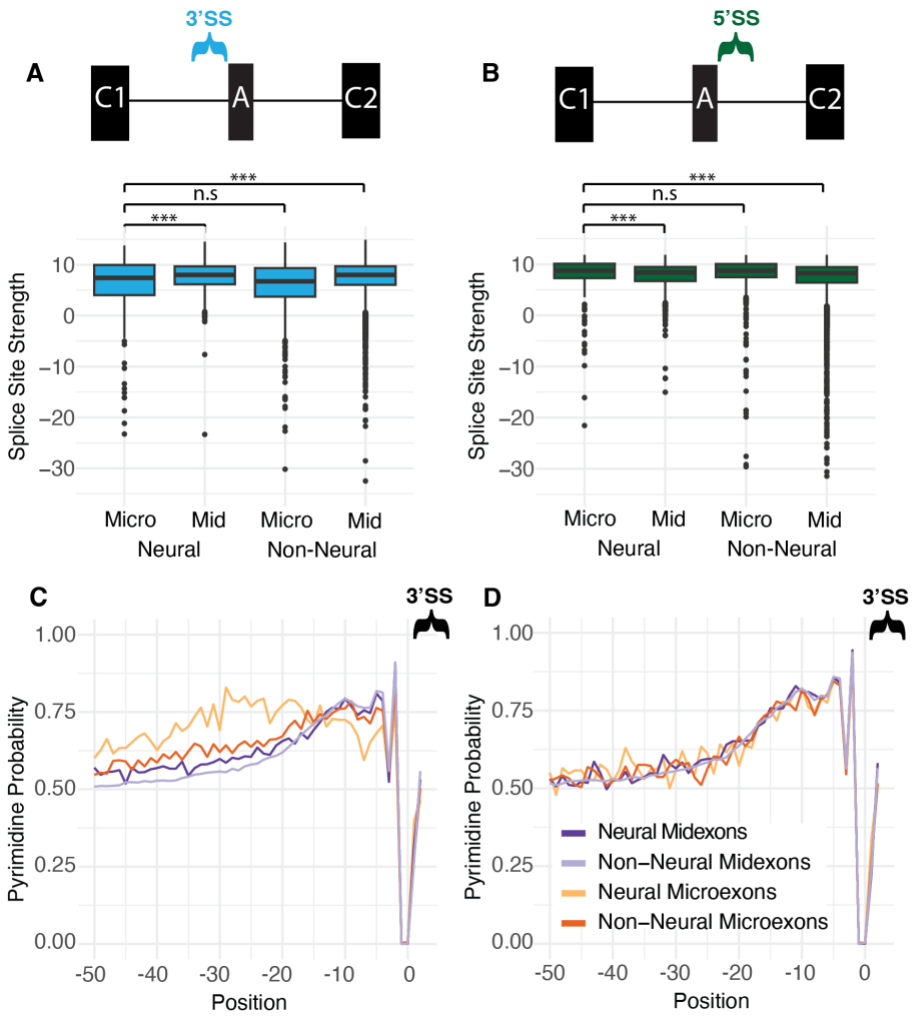


Figure 4. Human neural microexons have weak 3' splice sites and a shifted polypyrimidine tract. (A) Human neural microexons have weak 3' splice site strength scores. (B) Human neural microexons have strong 5' splice site strength scores. (C) The polypyrimidine tract is shifted farther into the intron in human neural microexons as shown by the shift in probability of cytosine/uridine per nucleotide upstream. (D) Altered polypyrimidine tracts are not evident in human neural microexons or other classes at the 3'ss of exons downstream. (Wilcoxon signed-rank test; * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.005).

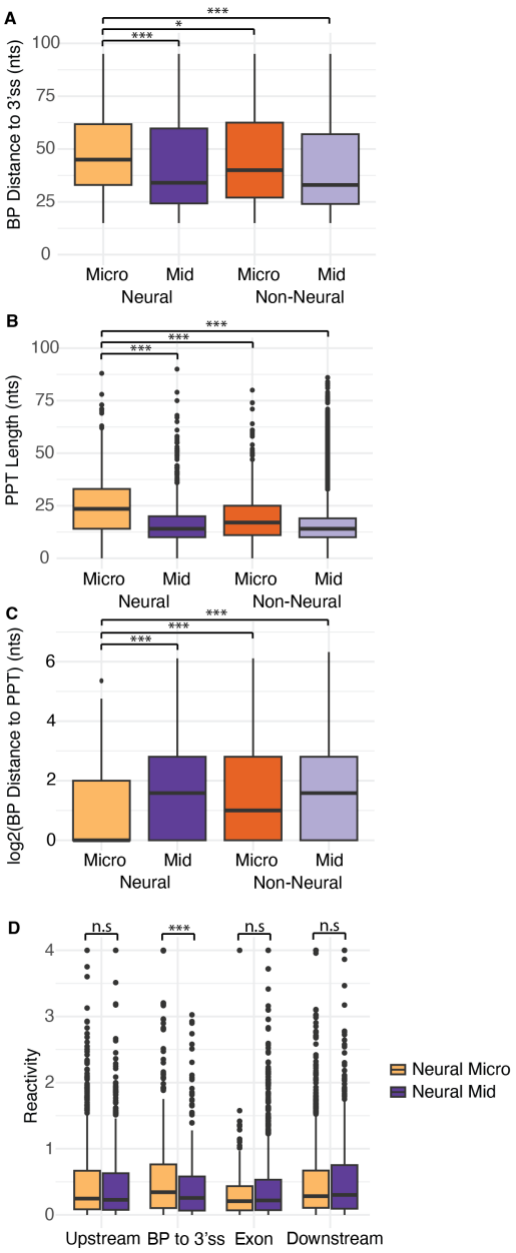
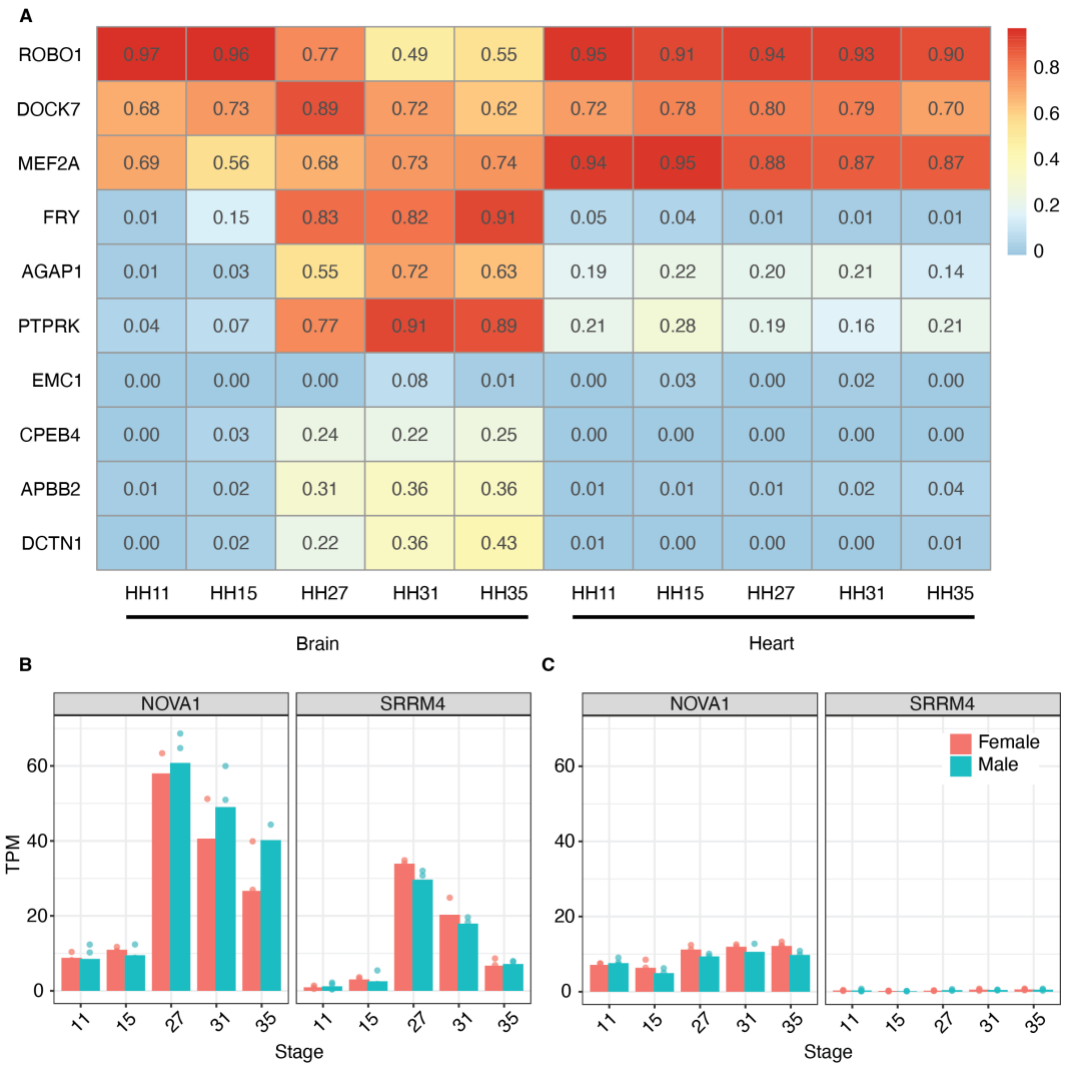


Figure 5. Human neural microexons have an extended branchpoint-to-splice site region through unique sequence and RNA structure elements. (A) Human neural microexons have a longer distance between the top-scoring branch-point (BP) and the 3' splice site compared to other classes of exon. (B) Likewise, human neural microexons have longer polypyrimidine tracts (PPT). (C) Human neural microexon polypyrimidine tracts are very close to the top-scoring branchpoint. Measurements were generated with SVM-BPfinder. (D) The branchpoint-to-splice site region of human neural microexons is significantly more accessible than human neural

581 midexons using experimental reactivity data in select regions. Regions upstream of the
582 branchpoint, within the exon and downstream of the 3' splice site have similar accessibility.
583 (neural microexons n = 13, neural midexons n = 7, Wilcoxon signed-rank test; *= p-value < 0.05,
584 ** = p-value < 0.01, *** = p-value < 0.005, n.s = p-value > 0.05).
585



586

587 **Figure 6. The majority of select neural microexons are spliced in a tissue-specific and**

588 **temporally regulated manner during chick embryonic development.** (A) Chicken neural

589 microexons have dynamic splicing during HH11, HH15, HH27, HH31, and HH35 developmental

590 stages as well as differences between brain and heart tissue. Includes both male and female

591 chickens. Samples collected in triplicate. (B) Expression of the RNA binding proteins *NOVA1*

592 and *SRRM4* show a pattern of upregulation between HH15 and HH27 that corresponds to the

593 dynamic splicing changes of many neural microexons between HH15 and HH27. (C) This

594 pattern of *NOVA1* and *SRRM4* upregulation is not evident in heart tissue.

595 **Tables**

596 **Table 1. Selected conserved neural microexons and their properties (Tapial et al. 2017).**

597 **Gene function summarized from (Irimia et al. 2014; Gonatopoulos-Pournatzis and**

598 **Blencowe 2020; NCBI 2025)**

Gene	*Human Expression (RPKM)*	*Human Average PSI*	*Gene Function*	*Microexon Function*	*Human to Chicken Sequence Alignment*	*Covariation (9 model / 100 vertebrates)*
AGAP1	15.8	13.1%	Endolysosomal trafficking	unknown	79.3%	yes/yes
APBB2	10.4	14.8%	Amyloidogenesis	unknown	67.5%	yes/yes
ASAP2	14.7	8.3%	Cellular migration	unknown	72.0%	yes/no
CLEC16A	6.4	9.7%	Autophagy and mitophagy regulation	unknown	72.0%	no/no
CPEB4	15.9	12.4%	Polyadenylation regulator	CPEB4 condensate interactions (Olla et al. 2023; Garcia-Cabau et al. 2025)	67.8%	yes/no
DCTN1	76.1	11.9%	Axonal transport	unknown	62.5%	no/no
DOCK7	8.3	19.8%	Neuronal polarization	unknown	76.5%	yes/yes
EMC1	11.4	13.9%	Membrane protein biogenesis	unknown	66.5%	yes/yes
FRY	10.3	21.3%	Bipolar spindle organization	unknown	71.3%	yes/no
ITSN1	9.8	6.4%	Signaling	Tissue specific protein interactions (Dergai et al. 2010)	68.8%	yes/no
MEF2A	17.8	13.6%	Neuronal differentiation and survival	unknown	69.5%	yes/yes
PTPRK	11.7	11.2%	Synaptic transmission	unknown	69.5%	yes/yes
ROBO1	27.7	88.5%	Axon guidance	Midline crossing (Johnson et al. 2019)	79.3%	yes/no

599

References

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**: 573-580.

Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. 2008. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9**: 474.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED et al. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* **14**: 708-715.

Blankenberg D, Taylor J, Nekrutenko A, Galaxy T. 2011. Making whole genome multiple alignments usable for biologists. *Bioinformatics* **27**: 2426-2428.

Busan S, Weeks KM. 2018. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA* **24**: 143-148.

Carranza F, Shenasa H, Hertel KJ. 2022. Splice site proximity influences alternative exon definition. *RNA Biol* **19**: 829-840.

Chapman SC, Collignon J, Schoenwolf GC, Lumsden A. 2001. Improved method for chick whole-embryo culture using a filter paper carrier. *Dev Dyn* **220**: 284-289.

Corvelo A, Hallegger M, Smith CW, Eyras E. 2010. Genome-wide association between branch point properties and alternative splicing. *PLoS Comput Biol* **6**: e1001016.

De Conti L, Baralle M, Buratti E. 2013. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip Rev RNA* **4**: 49-60.

Dergai M, Tsyba L, Dergai O, Zlatskii I, Skrypkina I, Kovalenko V, Rynditch A. 2010. Microexon-based regulation of ITSN1 and Src SH3 domains specificity relies on introduction of charged amino acids into the interaction interface. *Biochem Biophys Res Commun* **399**: 307-312.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.

Eddy SR. 2014. Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. *Annu Rev Biophys* **43**: 433-456.

Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, Won H, van Bakel H, Varghese M, Wang Y et al. 2018. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**.

Garcia-Cabau C, Bartomeu A, Tesei G, Cheung KC, Pose-Utrilla J, Pico S, Balaceanu A, Duran-Arque B, Fernandez-Alfara M, Martin J et al. 2025. Mis-splicing of a neuronal microexon promotes CPEB4 aggregation in ASD. *Nature* **637**: 496-503.

Gazave E, Marques-Bonet T, Fernando O, Charlesworth B, Navarro A. 2007. Patterns and rates of intron divergence between humans and chimpanzees. *Genome Biol* **8**: R21.

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 Res* **22**: 35-50.

Gonatopoulos-Pournatzis T, Blencowe BJ. 2020. Microexons: at the nexus of nervous system development, behaviour and autism spectrum disorder. *Curr Opin Genet Dev* **65**: 22-33.

Hamburger V, Hamilton HL. 1992. A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn* **195**: 231-272.

Herbert A, Hatfield A, Lackey L. 2023. How does precursor RNA structure influence RNA processing and gene expression? *Biosci Rep* **43**.

Herbert A, Hatfield A, Randazza A, Miyamoto V, Palmer K, Lackey L. 2025. Precursor RNA structural patterns at SF3B1 mutation sensitive cryptic 3' splice sites. *RNA Biol* **22**: 1-15.

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F et al. 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**: D590-598.

650 Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, Quesnel-
651 Vallieres M, Tapial J, Raj B, O'Hanlon D et al. 2014. A highly conserved program of
652 neuronal microexons is misregulated in autistic brains. *Cell* **159**: 1511-1523.
653 Jobbins AM, Campagne S, Weinmeister R, Lucas CM, Gosliga AR, Clery A, Chen L, Eperon LP,
654 Hodson MJ, Hudson AJ et al. 2022. Exon-independent recruitment of SRSF1 is
655 mediated by U1 snRNP stem-loop 3. *EMBO J* **41**: e107640.
656 Johnson V, Junge HJ, Chen Z. 2019. Temporal regulation of axonal repulsion by alternative
657 splicing of a conserved microexon in mammalian Robo1 and Robo2. *Elife* **8**.
658 Jurka J. 2000. Repbase update: a database and an electronic journal of repetitive elements.
659 *Trends Genet* **16**: 418-420.
660 Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The
661 UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**: D493-496.
662 Kumar J, Lackey L, Waldern JM, Dey A, Mustoe AM, Weeks KM, Mathews DH, Laederach A.
663 2022. Quantitative prediction of variant effects on alternative splicing in MAPT using
664 endogenous pre-messenger RNA structure probing. *Elife* **11**.
665 Lee Y, Rio DC. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev*
666 *Biochem* **84**: 291-323.
667 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
668 Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and
669 SAMtools. *Bioinformatics* **25**: 2078-2079.
670 Li X, Liu S, Zhang L, Issaian A, Hill RC, Espinosa S, Shi S, Cui Y, Kappel K, Das R et al. 2019. A
671 unified mechanism for intron and exon definition and back-splicing. *Nature* **573**: 375-380.
672 Li YI, Sanchez-Pulido L, Haerty W, Ponting CP. 2015. RBFOX and PTBP1 proteins regulate the
673 alternative splicing of micro-exons in human brain transcripts. *Genome Res* **25**: 1-13.
674 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for
675 assigning sequence reads to genomic features. *Bioinformatics* **30**: 923-930.
676 Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL.
677 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26.
678 Mackensen T, Irimia M. 2025. From Tiny Exons to Big Insights: The Expanding Field of
679 Microexons. *Annu Rev Genomics Hum Genet* **26**: 77-102.
680 NCBI NCfBI. 2025. NCBI Gene. Vol 2025.
681 Olla I, Pardinas AF, Parras A, Hernandez IH, Santos-Galindo M, Pico S, Callado LF, Elorza A,
682 Rodriguez-Lopez C, Fernandez-Miranda G et al. 2023. Pathogenic Mis-splicing of
683 CPEB4 in Schizophrenia. *Biol Psychiatry* **94**: 341-351.
684 Parras A, Anta H, Santos-Galindo M, Swarup V, Elorza A, Nieto-Gonzalez JL, Pico S,
685 Hernandez IH, Diaz-Hernandez JI, Belloc E et al. 2018. Autism-like phenotype and risk
686 gene mRNA deadenylation by CPEB4 mis-splicing. *Nature* **560**: 441-446.
687 Porter RS, Jaamour F, Iwase S. 2018. Neuron-specific alternative splicing of transcriptional
688 machineries: Implications for neurodevelopmental disorders. *Mol Cell Neurosci* **87**: 35-
689 45.
690 Quesnel-Vallieres M, Dargaei Z, Irimia M, Gonatopoulos-Pournatzis T, Ip JY, Wu M, Sterne-
691 Weiler T, Nakagawa S, Woodin MA, Blencowe BJ et al. 2016. Misregulation of an
692 Activity-Dependent Splicing Network as a Common Mechanism Underlying Autism
693 Spectrum Disorders. *Mol Cell* **64**: 1023-1034.
694 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
695 *Bioinformatics* **26**: 841-842.
696 Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction
697 and analysis. *BMC Bioinformatics* **11**: 129.
698 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential
699 expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

700 Saha K, Fernandez MM, Biswas T, Joseph S, Ghosh G. 2021. Discovery of a pre-mRNA
701 structural scaffold as a contributor to the mammalian splicing code. *Nucleic Acids Res*
702 **49**: 7103-7121.
703 Saha K, Ghosh G. 2022. Cooperative engagement and subsequent selective displacement of
704 SR proteins define the pre-mRNA 3D structural scaffold for early spliceosome assembly.
705 *Nucleic Acids Res* **50**: 8262-8278.
706 Shenasa H, Hertel KJ. 2019. Combinatorial regulation of alternative splicing. *Biochim Biophys*
707 *Acta Gene Regul Mech* **1862**: 194392.
708 Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM. 2014. RNA motif discovery by SHAPE
709 and mutational profiling (SHAPE-MaP). *Nat Methods* **11**: 959-965.
710 Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J,
711 Hillier LW, Richards S et al. 2005. Evolutionarily conserved elements in vertebrate,
712 insect, worm, and yeast genomes. *Genome Res* **15**: 1034-1050.
713 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M,
714 Soding J et al. 2011. Fast, scalable generation of high-quality protein multiple sequence
715 alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.
716 Smit AH, R; Green, P 1996-2010. RepeatMasker Open-3.0.
717 Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, Quesnel-
718 Vallieres M, Permanyer J, Sodaei R, Marquez Y et al. 2017. An atlas of alternative
719 splicing profiles and functional associations reveals new regulatory programs and genes
720 that simultaneously express multiple major isoforms. *Genome Res* **27**: 1759-1768.
721 Torres-Mendez A, Bonnal S, Marquez Y, Roth J, Iglesias M, Permanyer J, Almudi I, O'Hanlon D,
722 Guitart T, Soller M et al. 2019. A novel protein domain in an ancestral splicing factor
723 drove the evolution of neural microexons. *Nat Ecol Evol* **3**: 691-701.
724 Ustianenko D, Weyn-Vanhentenryck SM, Zhang C. 2017. Microexons: discovery, regulation, and
725 function. *Wiley Interdiscip Rev RNA* **8**.
726 Wang Y, Xie Z, Kutschera E, Adams JI, Kadash-Edmondson KE, Xing Y. 2024. rMATS-turbo: an
727 efficient and flexible computational tool for alternative splicing analysis of large-scale
728 RNA-seq data. *Nat Protoc* **19**: 1083-1104.
729 Williams RM, Lukoseviciute M, Sauka-Spengler T, Bronner ME. 2022. Single-cell atlas of early
730 chick development reveals gradual segregation of neural crest lineage from the neural
731 plate border during neurulation. *Elife* **11**.
732 Yeo G, Holste D, Kreiman G, Burge CB. 2004. Variation in alternative splicing across human
733 tissues. *Genome Biol* **5**: R74.
734 Zhang Z, Kumar V, Dybkov O, Will CL, Zhong J, Ludwig SEJ, Urlaub H, Kastner B, Stark H,
735 Luhrmann R. 2024. Structural insights into the cross-exon to cross-intron spliceosome
736 switch. *Nature* **630**: 1012-1019.
737 Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2017. DMS-MaPseq
738 for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* **14**: 75-82.

739