

1 **Supplementary information for “Benchmarking RNA-seq with the Quartet**
2 **Reference Materials to establish Best Practices for Accurate Alternative Splicing**
3 **Detection”**

4 Content

5 1. Supplementary Methods.....4

- 6 1.1 Construction of the Quartet junction reference datasets.....4
- 7 1.2 Annotation of features for junctions, isoforms, and alternative splicing (AS) events.....5
- 8 1.3 Construction of a unified format for different AS event types.....6

9 2. Supplementary Figures.....10

- 10 Figure S1. Experimental and analytical design of this study.....10
- 11 Figure S2. Overview of bioinformatics design.12
- 12 Figure S3. Quality control of RNA-seq data.....13
- 13 Figure S4. Stepwise workflow for constructing junction reference datasets.14
- 14 Figure S5. Evaluation of junction-level detection performance of RNA-seq.15
- 15 Figure S6. Pairwise consistency in junction detection across short- and long-read RNA-seq.16
- 16
- 17 Figure S7. Read counts distribution of isoforms associated with FN, FP, and TP junctions...17
- 18 Figure S8. Coverage uniformity distribution of isoforms associated with FN, FP, and TP
- 19 junctions.....18
- 20 Figure S9. Length distribution of isoforms associated with FN, FP, and TP junctions.....19
- 21 Figure S10. Exon count distribution of isoforms associated with FN, FP, and TP junctions..20
- 22 Figure S11. GC content distribution of FN, FP, and TP junctions.21
- 23 Figure S12. Mappability distribution of FN, FP, and TP junctions.22
- 24 Figure S13. Mappability distribution of novel junctions flagged as FN and TP.23
- 25 Figure S14. Pairwise consistency in isoform quantification across laboratories.24
- 26 Figure S15. Impact of data quality on isoform quantification accuracy.....25
- 27 Figure S16. Comparison of three performance metrics.26
- 28 Figure S17. Number of DEIs detected by laboratories.27
- 29 Figure S18. PCCs of 18 isoform quantification pipelines.....28
- 30 Figure S19. Comparison of CV across 19 quantification pipelines.29
- 31 Figure S20. Comparison of SNR across 18 quantification pipelines.30
- 32 Figure S21. Impact of isoform quantification tools on DEI detection.31
- 33 Figure S22. Recall and precision of differential expression analysis pipelines.32
- 34 Figure S23. Performance of DEI detection tools combined with isoform quantification
- 35 pipelines.33
- 36 Figure S24. Impact of isoform complexity (K-value) on isoform quantification pipelines....34
- 37 Figure S25. Impact of isoform length on isoform quantification pipelines.....35
- 38 Figure S26. Impact of isoform number per gene on isoform quantification pipelines.....36
- 39 Figure S27. Impact of exon number per isoform on isoform quantification pipelines.....37
- 40 Figure S28. Impact of exon length on isoform quantification pipelines.38
- 41 Figure S29. Impact of GC content on isoform quantification pipelines.....39
- 42 Figure S30. Impact of isoform expression on isoform quantification pipelines.40
- 43 Figure S31. Impact of isoform features based on the built-in truth.....41
- 44 Figure S32. Pairwise consistency in AS event quantification across laboratories.42

45	Figure S33. Consistency of ASE detection across 42 laboratories.....	43
46	Figure S34. Characteristics of ASEs with different detection consistency.....	44
47	Figure S35. Number of DSEs detected by laboratories.....	45
48	Figure S36. Distribution of dPSI values detected by SUPPA2 combined with 18 isoform quantification pipelines.....	46
50	Figure S37. SNR for SUPPA2 combined with 18 isoform quantification pipelines.....	47
51	Figure S38. Impact of the number of adjacent AS events on AS detection pipelines.....	48
52	Figure S39. Impact of AS event-associated isoform number on AS detection pipelines.....	49
53	Figure S40. Impact of the GC content on AS detection pipelines.....	50
54	Figure S41. Impact of the mappability on AS detection pipelines.....	51
55	Figure S42. Impact of PSI levels on AS detection pipelines.....	52
56	Figure S43. Impact of AS event-associated isoform expression levels on AS detection pipelines.....	53
58	Figure S44. A unified format for different AS event types.....	54
59	Figure S45. The intersections of annotated and novel AS events by four AS event detection tools.....	55
61	Figure S46. The intersections of annotated AS events by four AS event detection tools.....	56
62	Figure S47. The intersections of annotated and novel AS events by four AS event detection tools.....	57
64	Figure S48. The intersections of annotated AS events by four AS event detection tools.....	58
65	Figure S49. The number of annotated and novel DSEs by four AS event detection tools.....	59
66	Figure S50. The number of annotated DSEs by four AS event detection tools.....	60
67	Figure S51. The intersections of annotated and novel DSEs by four AS event detection tools.....	61
69	Figure S52. The intersections of annotated DSEs by four AS event detection tools.....	62
70	Figure S53. The intersections of annotated and novel DSEs by four AS event detection tools.....	63
72	Figure S54. The intersections of annotated DSEs by four AS event detection tools.....	64
73	Figure S55. The dPSI consistency of annotated and novel AS events by four AS event detection tools.....	65
75	Figure S56. The dPSI consistency of annotated AS events by four AS event detection tools.....	66
76	Figure S57. The MCC of four AS event detection tools for DSE detection.....	67
77	Reference	68
78		

79 **1. Supplementary Methods**

80 **1.1 Construction of the Quartet junction reference datasets**

81 **Long-read sequencing data.** Seven high-quality batches of long-read RNA-seq
82 (lRNA-seq) data from the Quartet samples were used to construct junction reference
83 datasets. These included two batches of multiplexed arrays sequencing (MAS-seq) data
84 generated using the PacBio platform, as well as two batches of direct RNA (dRNA)
85 sequencing and three batches of PCR-cDNA sequencing generated using the Oxford
86 Nanopore Technologies (ONT). Detailed descriptions of these data are available in our
87 companion study¹.

88 **lRNA-seq data processing and mapping.** For PCR-cDNA ONT sequencing data,
89 primer sequences were first removed using Pychopper (v2.7.10)
90 (<https://github.com/epi2me-labs/pychopper>), followed by Poly(A) tail trimming using
91 the trim_isoseq_polyA script from the official PacBio GitHub repository. MAS-seq data
92 were preprocessed using the standard Iso-Seq toolkit (v4.3.0) workflow
93 (<https://isoseq.how/>), which includes primer removal with lima and Poly(A) tail
94 trimming with isoseq refine. dRNA ONT sequencing data were used without
95 preprocessing. All seven batches of long-read data were aligned to the GRCh38 genome
96 assembly using Minimap2 (v2.28)², with gene annotations from Ensembl release-109.
97 The ONT data was aligned with Minimap2 using the parameters -ax splice, -uf, and -
98 k14. For the PacBio data, the alignment was performed using the parameters -ax
99 splice:hq and -uf. Junctions were extracted from BAM files using the
100 sjFromSAMcollapseUandM_inclOverlaps.awk script provided by STAR (v2.7.10b)³.

101 **Construction of the junction reference datasets.** Building the reference datasets
102 mainly involved three main steps: identifying high-confidence junctions, selecting
103 candidate gene sets with sufficient coverage, and generating the final reference datasets
104 (Figure S4).

105 (i) First, junctions were defined as high-confidence positives if, in each sample, at least
106 two out of three replicates had ≥ 3 supporting reads, and this criterion was met in at least
107 5 out of the 7 lRNA-seq batches. Conversely, junctions with no supporting reads in

108 any of the 3 replicates across all 7 batches were defined as high-confidence negative
109 junctions. For novel junctions, only high-confidence positive junctions were retained.
110 (ii) Next, genes were selected as candidates for inclusion in the reference datasets if, in
111 each sample, at least 2 out of 3 replicates had ≥ 3 supporting reads, and this criterion
112 was met in at least 5 out of the 7 batches. Genes that did not meet this threshold and
113 genes containing any low-confidence junctions were excluded.
114 (iii) Finally, junctions from the candidate gene sets were retained for inclusion in the
115 final junction reference dataset. For annotated junctions, we ensured that all junctions
116 within each gene were known. For novel junctions, only a reliable set of high-
117 confidence positive junctions was included.

118

119 **1.2 Annotation of features for junctions, isoforms, and alternative splicing (AS)
120 events**

121 **Features of junctions.** The length and exon number of isoforms containing each
122 junction were directly extracted from the gene annotation (GTF). Read coverage for
123 each junction was obtained from STAR-derived junction files (SJ.out.tab). Isoform
124 coverage uniformity was calculated using the geneBody_coverage.py script from
125 RSeQC (v5.0.4)⁴. For each isoform, 100 positions were uniformly sampled, and the
126 coverage at each position was divided by the maximum sequencing depth. The
127 proportion of positions with relative coverage greater than 0.75 was used as the final
128 measure of isoform coverage uniformity. GC content was calculated by extracting the
129 50 bp upstream of the junction start and 50 bp downstream of the junction end, and
130 computing the proportion of guanine and cytosine bases within this sequence.
131 Mappability was assessed using GenMap (v1.3.0)⁵ to compute genome-wide
132 mappability scores for each base, and the mean mappability of the 50 bp flanking
133 sequence was used as the junction's mappability measure.

134 **Features of isoforms.** Isoform length, the number of isoforms per gene, the number of
135 exons per isoform, and exon length were directly extracted from the gene annotation
136 (GTF). Isoform GC content was calculated by concatenating all exon sequences of the

137 isoform and computing the proportion of guanine and cytosine bases. Isoform
138 expression levels were obtained directly as FPKM values from the respective
139 quantification pipelines. The K value, which measures the complexity of exon-isoform
140 structures for each gene, was defined in previous studies⁶. The calculation procedure
141 was conducted as follows:

142 First, for a given gene harboring I distinct transcript isoforms and E exons, we
143 constructed a binary matrix $A \in \{0,1\}^{I \times E}$. Each element a_{ie} of this matrix was assigned a
144 value of 1 if the i-th isoform included the e-th exon, and 0 otherwise. This matrix A thus
145 provides a complete and precise representation of the exon composition of all annotated
146 transcript isoforms for the gene.

147 Subsequently, the singular value decomposition (SVD) of the matrix A was performed.
148 From this decomposition, we obtained the singular values of A, denoted $\sigma_1, \sigma_2, \dots, \sigma_r$,
149 where r is the rank of A. The maximum singular value, $\sigma_{\max}(A)$, and the minimum non-
150 zero singular value, $\sigma_{\min}(A)$, were identified.

151 Finally, the K-value was calculated as the ratio of these two singular values:

152
$$K \text{ value} = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} \quad (1)$$

153 **Features of AS events.** The number of transcript isoforms associated with each AS
154 event was directly obtained from the .ioe files generated by the generateEvents
155 command in SUPPA2 (v2.4) ⁷. The number of neighboring AS events was calculated
156 by selecting 150 bp upstream and downstream of each event and counting all event
157 types present within these regions. For sequence-based features, we extracted the 75 bp
158 upstream and downstream of the relevant junctions for SE, A5, A3, MX, AF, and AL
159 events, while the entire intronic region was used for RI events. GC content was
160 computed as the proportion of guanine and cytosine bases in these event-specific
161 sequences, and mappability was assessed using GenMap as described above.

162

163 **1.3 Construction of a unified format for different AS event types**

164 **Skipping exon (SE).** For SE events, four coordinates were used, including the start and
165 end coordinates of the alternative exon, as well as the end coordinate of the upstream

166 exon and the start coordinate of the downstream exon. For the rMATS output, these
167 four coordinates correspond to “exonStart_0base”, “exonEnd”, “upstreamEE”, and
168 “downstreamES”, respectively. For the PSI-Sigma output, the coordinates of the
169 alternative exon correspond to the “Target Exon,” while the end coordinate of the
170 upstream exon and the start coordinate of the downstream exon are derived from the
171 “Event Region” coordinates. In MAJIQ output files, SE events are divided into four
172 local splicing variations (LSVs). For LSVs with junction names C1_C2 or C2_C1, the
173 “junction_coord” corresponds to the end coordinate of the upstream exon and the start
174 coordinate of the downstream exon. For LSVs with junction name C1_A, the
175 “spliced_with_coord” corresponds to the start and end coordinates of the alternative
176 exon. PSI represents the mean PSI value across all LSVs that support inclusion of the
177 alternative exon. For the SUPPA2 output, the four coordinates were directly extracted
178 from the “event_id” field.

179 **Alternative 5' splice site (A5SS).** For A5SS, three coordinates were used: the end
180 coordinate of the alternative exon, the splice site coordinate on the alternative exon, and
181 the start coordinate of the downstream exon. These three coordinates correspond to the
182 junctions supporting inclusion or exclusion of the 5' splice site. For the rMATS output,
183 using + strand genes as an example, the three coordinates are “longExonEnd”,
184 “shortEE”, and “flankingES”. For the PSI-Sigma output, using + strand genes as an
185 example, the three coordinates correspond to the end coordinate of the “Target Exon”,
186 and the start and end coordinates of the “Event Region”. For the MAJIQ output, an
187 A5SS event is represented by two LSVs. The “junction_coord” with junction name
188 *Proximal* corresponds to the splice site of the alternative exon and the start coordinate
189 of the downstream exon, while the “junction_coord” with junction name *Distal*
190 corresponds to the end coordinate of the alternative exon and the start coordinate of the
191 downstream exon. The LSV with junction name *Proximal* represents the isoform that
192 includes the alternative exon, and its PSI value is used. For the SUPPA2 output, the
193 three coordinates were directly extracted from the “event_id” field.

194 **Alternative 3' splice site (A3SS).** For A3SS, similar to A5SS, three coordinates are

195 used to define the event, essentially mirroring the strand orientation. The outputs from
196 the different tools are processed in the same way as described above.

197 **Alternative first exon (AF).** For AF, two alternative exons are involved, with the one
198 at the start of the isoform designated as alternative exon 1 and the other as alternative
199 exon 2. AF is defined using five coordinates: the start and end coordinates of alternative
200 exon 1, the start and end coordinates of alternative exon 2, and the start coordinate of
201 the downstream exon. rMATS and PSI-Sigma do not detect this event type. For the
202 MAJIQ output, AF events are also represented as two LSVs. The “spliced_with_coord”
203 of the junction with the name *Proximal* corresponds to the start and end coordinates of
204 alternative exon 1, while that of the junction with the name *Distal* corresponds to the
205 start and end coordinates of alternative exon 2. In addition, the start coordinate of the
206 downstream exon can be extracted from the “junction_coord”. The LSV with junction
207 name *Distal* represents the isoform that includes the alternative exon 1, and its PSI
208 value is used. For the SUPPA2 output, the five coordinates were directly extracted from
209 the “event_id” field.

210 **Alternative last exon (AL).** For AL, similar to AF, five coordinates are used to define
211 the event, essentially mirroring the strand orientation. The outputs from the different
212 tools are processed in the same way as described above.

213 **Mutually exclusive exon (MX).** MX events involve two alternative exons. If the strand
214 is +, the 5' exon is defined as alternative exon 1; if the strand is -, the 3' exon is defined
215 as alternative exon 1. MX is defined using six coordinates: the start and end coordinates
216 of alternative exon 1, the start and end coordinates of alternative exon 2, the end
217 coordinate of the upstream exon, and the start coordinate of the downstream exon. For
218 rMATS output, using the + strand as an example, the six coordinates correspond to
219 “1stExonStart_0base”, “1stExonEnd”, “2ndExonStart_0base”, “2ndExonEnd”,
220 “upstreamEE”, and “downstreamES”. For PSI-Sigma output, MX events are split into
221 two separate events corresponding to the two alternative exons. The coordinates of the
222 “Target Exon” represent the alternative exons, while the “Event Region” spans the end
223 coordinate of the upstream exon and the start coordinate of the downstream exon. The

224 PSI value of the first alternative exon is used for analysis. For MAJIQ output, MX
225 events are represented by four LSVs with junction names *C1_A1*, *C2_A2*, *C2_A1*, and
226 *C2_A2*. The “spliced_with_coord” of LSVs with “spliced_with” equal to *A1* and *A2*
227 correspond to the coordinates of alternative exon 1 and alternative exon 2, respectively.
228 Additionally, the “junction_coord” of *C1_A1* and *C2_A2* contains the end coordinate
229 of the upstream exon and the start coordinate of the downstream exon. The PSI value
230 of the LSV with “spliced_with” equal to *A1* is used as the PSI for the MX event. For
231 the SUPPA2 output, the six coordinates were directly extracted from the “event_id”
232 field.

233 **Retained intron (RI).** For RI events, four coordinates are used, including the
234 start and end of the retained intron, as well as the start coordinate of the upst
235 steam exon and the end coordinate of the downstream exon. For rMATS output,
236 these correspond to “upstreamES”, “upstreamEE”, “downstreamES”, and “down
237 streamEE”. For PSI-Sigma output, the “Target Exon” corresponds to the coordi
238 nates of the retained intron, while the “Event Region” includes the end coordi
239 nate of the upstream exon and the start coordinate of the downstream exon. For
240 MAJIQ output, RI events are represented by two LSVs, with “junction_name”
241 as *C1_C2_intron/C2_C1_intron* and *C1_C2_spliced/C2_C1_spliced*. The “juncti
242 on_coord” of *C1_C2_intron/C2_C1_intron* corresponds to the coordinates of the
243 retained intron, while “spliced_with_coord” and “reference_exon_coord” provide
244 the start coordinate of the upstream exon and the end coordinate of the downst
245 steam exon. The PSI of the RI event is taken from the LSV with “junction_na
246 me” *C1_C2_intron/C2_C1_intron*. For the SUPPA2 output, the four coordinates
247 were directly extracted from the “event_id” field.

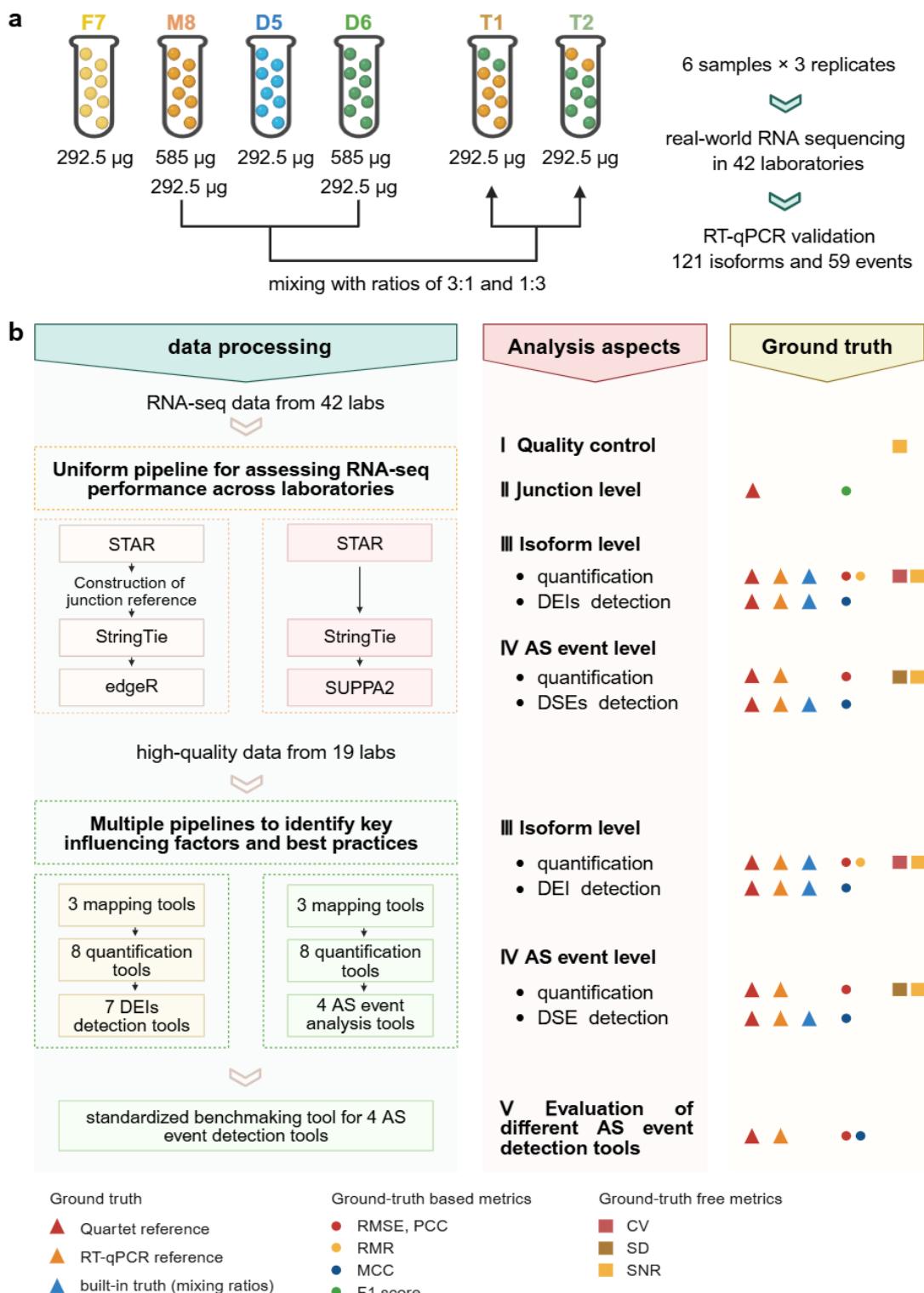
248

249

250

251 **2. Supplementary Figures**

252 **Figure S1. Experimental and analytical design of this study.**

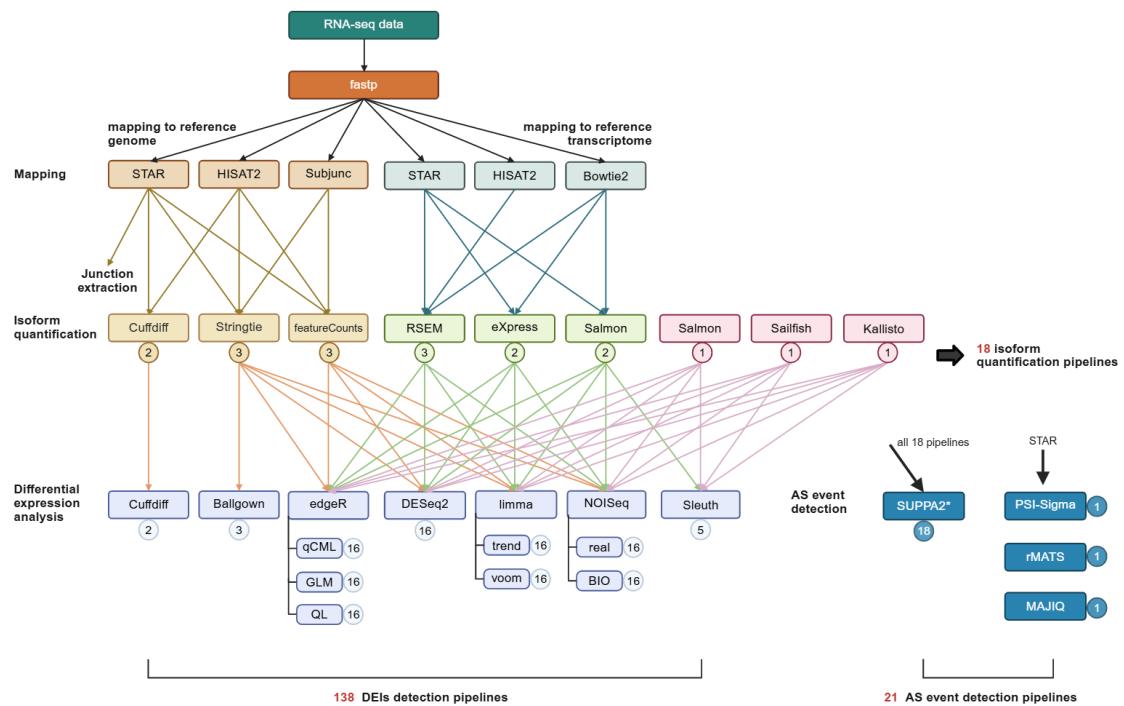


253

254 **a**, Preparation and mixing scheme of the Quartet RNA reference materials. **b**, Overview
255 of data analysis strategies and their objectives. A unified isoform and alternative

256 splicing (AS) event analysis pipeline was first applied to all RNA-seq data from 42
257 laboratories to evaluate overall detection performance. Based on signal-to-noise ratio
258 (SNR) and other quality metrics, high-quality laboratories were selected for further
259 analysis. Multiple combinations of mapping tools, isoform quantification tools,
260 differentially expressed isoforms (DEIs) detection methods, and AS event callers were
261 included to investigate the influence of experimental procedures, analysis pipelines, and
262 intrinsic features of isoforms and AS events. This allowed identification of best
263 practices for isoform and AS event analysis. Finally, a standardized benchmarking tool
264 was developed specifically for evaluating widely used AS event detection tools,
265 including SUPPA2 (v2.4), rMATS (v4.3.0)⁸, MAJIQ (v2.5.6)⁹, and PSI-Sigma (v2.3)¹⁰.
266 Throughout the study, three types of ground truth were employed, complemented by
267 both ground-truth based and ground-truth free performance metrics to comprehensively
268 assess RNA-seq performance in AS profiling. RMSE, root mean square error; PCC,
269 Pearson correlation coefficient; MCC, Matthews correlation coefficient; RMR,
270 recovery of mixing ratios; CV, coefficient of variation; SD, standard deviation. Created
271 in BioRender. Zhang, R. (2025) <https://BioRender.com/tgay6bs>.

272 **Figure S2. Overview of bioinformatics design.**

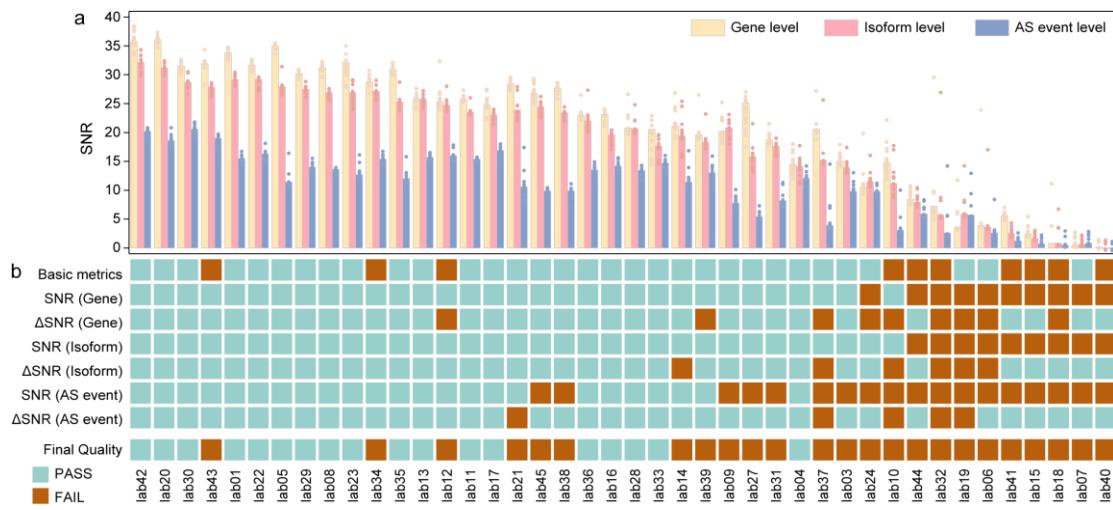


273

274 RNA-seq data analysis was performed using four mapping tools, eight isoform
 275 quantification tools, seven DEI detection tools, and four AS event detection tools,
 276 resulting in a total of 159 distinct analysis pipelines. Created in BioRender. Zhang, R.
 277 (2025) <https://BioRender.com/lstsz9>.

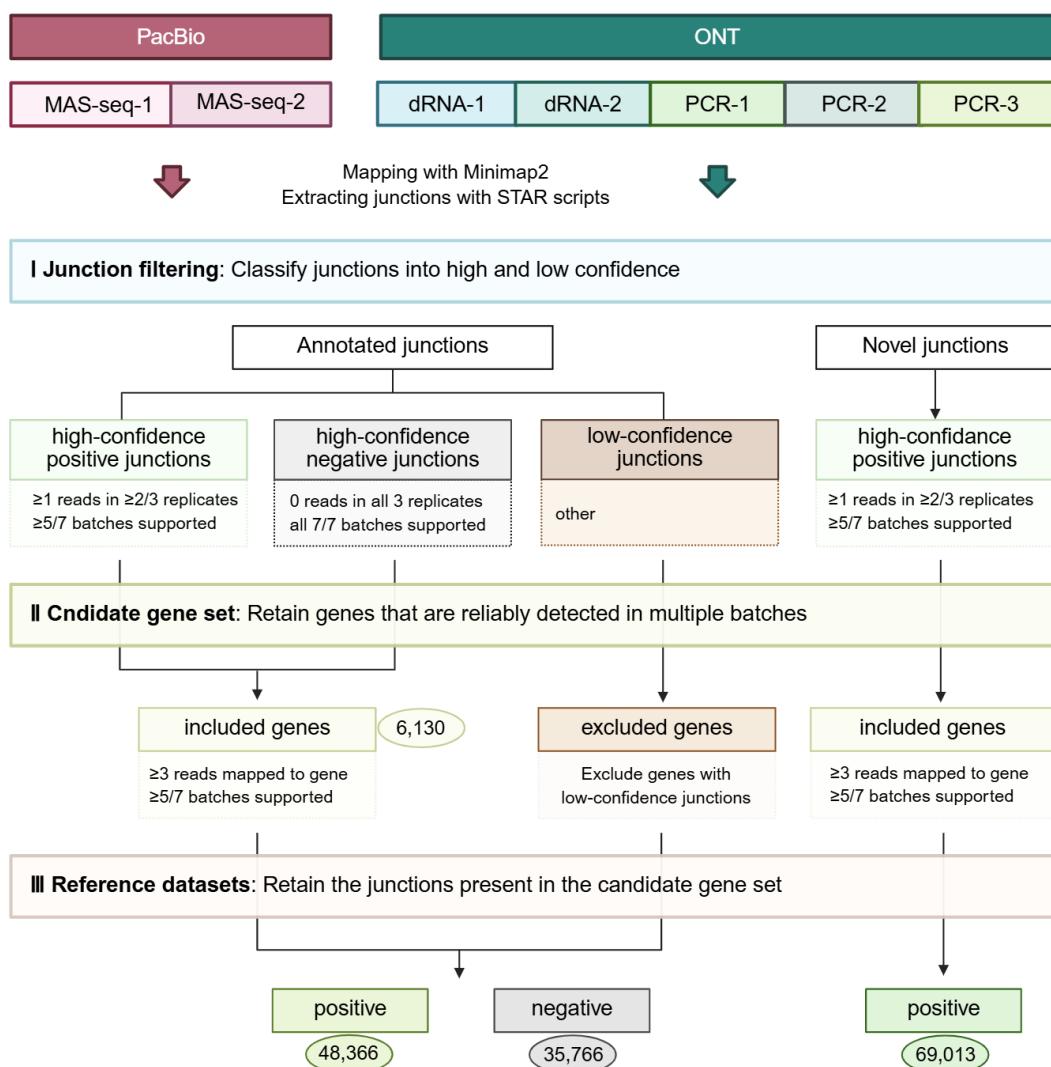
278 **Figure S3. Quality control of RNA-seq data.**

279



280 **a**, Signal-to-noise ratio (SNR) values to measure the quality of quantification data at
 281 gene, isoform, and AS event levels for 42 laboratories based on the Quartet and mixed
 282 samples (18 samples). Dots represented SNR values based on any 17 of the 18 samples.
 283 The red dashed line represents the SNR cutoff of 10. **b**, Multiple quality metrics were
 284 combined to assign quality tags for each laboratory. The basic metrics included the
 285 number of sequencing reads, base quality (Q30), mapping rate, gene body bias (5'-3'
 286 bias), the percentage of mapped reads in the intergenic region, duplicate rates, cross-
 287 contamination, and sample-identity check (Supplementary Data 1). Δ SNR represents
 288 the difference between the SNR computed from any subset of 17 out of 18 samples and
 289 that computed from all 18 samples, serving to identify low-quality outliers.

290 **Figure S4. Stepwise workflow for constructing junction reference datasets.**

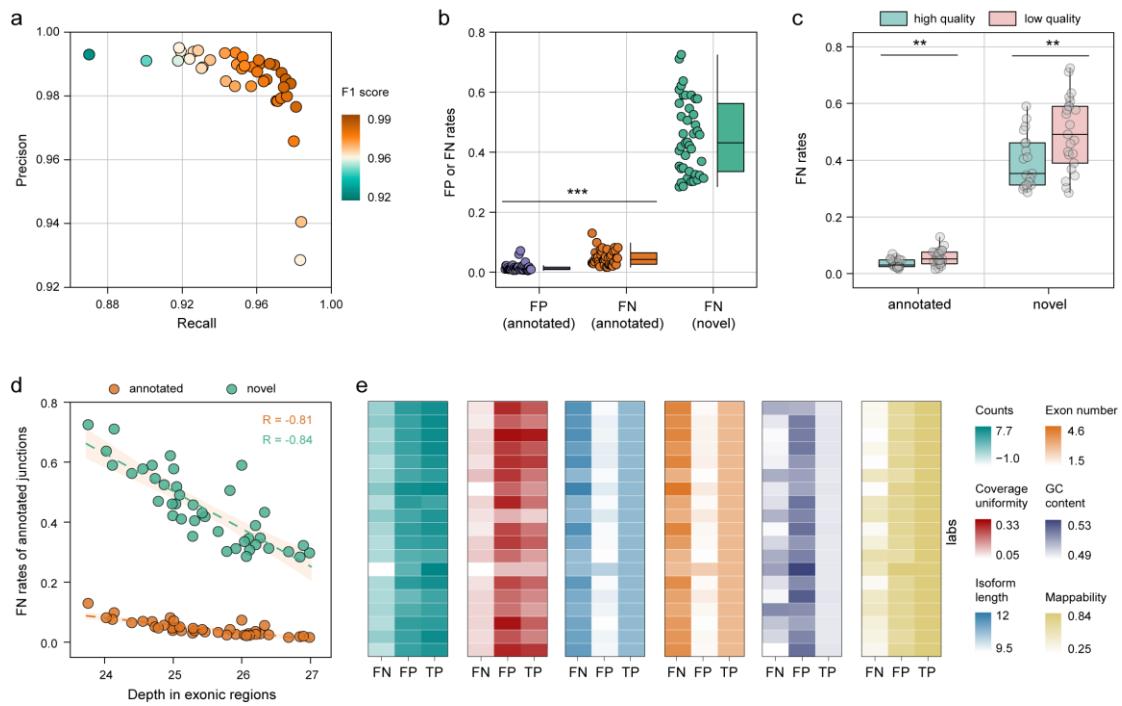


291

292 Seven batches of high-quality long-read sequencing data from our companion study
 293 were used to construct reference datasets (details in Supplementary Methods). First,
 294 positive and negative junctions supported by long-read sequencing data across multiple
 295 batches were selected. Second, a candidate gene set with sufficient read coverage was
 296 identified. Finally, positive and negative junctions located within the candidate gene set
 297 were further screened, resulting in a reference dataset in which all annotated positive
 298 and negative junctions within these genes were known, as well as an additional dataset
 299 containing novel junctions. Created in BioRender. Zhang, R. (2025)
 300 <https://BioRender.com/5f2hh7w>.

301

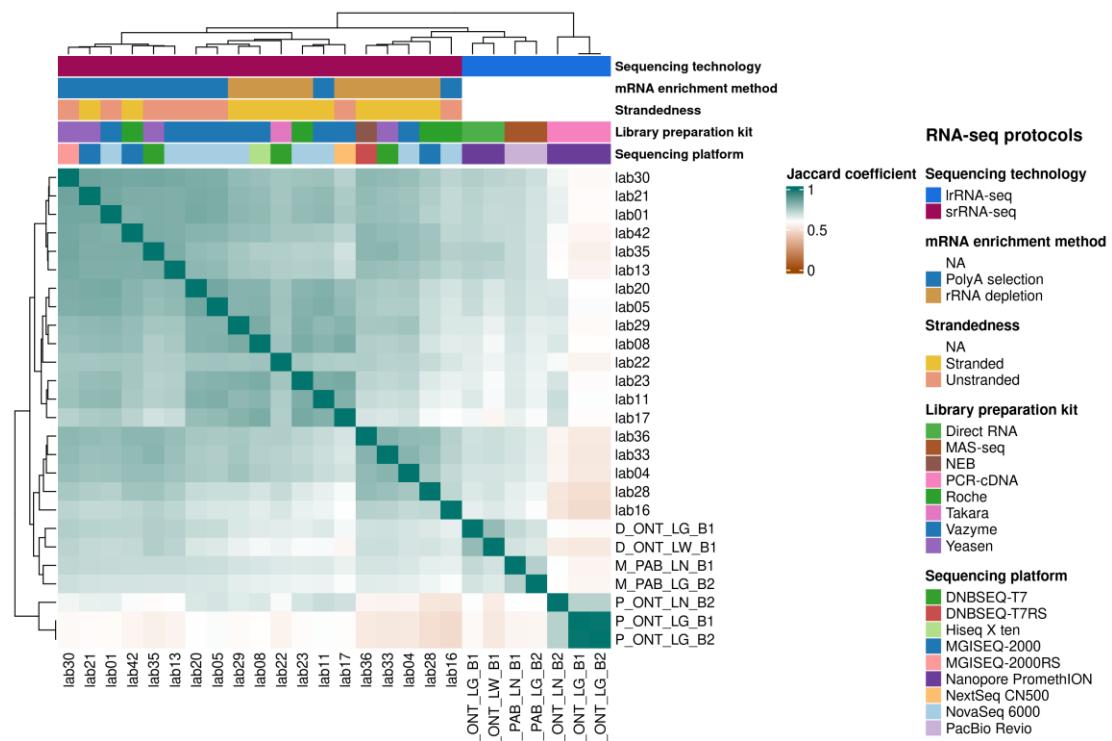
Figure S5. Evaluation of junction-level detection performance of RNA-seq.



302

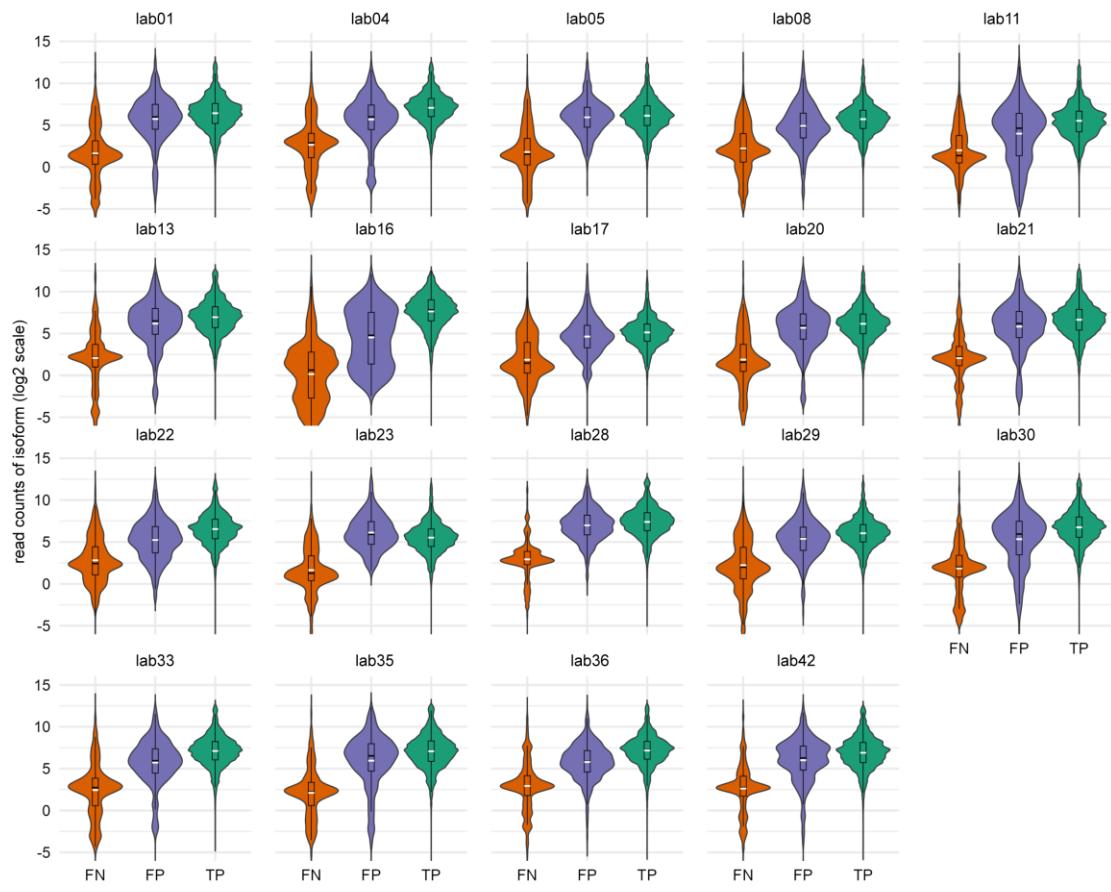
303 **a**, Recall, precision and F1 score of junction detection across 42 laboratories based on
 304 the junction reference datasets. **b**, Comparison of false positive (FP) and false negative
 305 (FN) rates of annotated and novel junctions across 42 laboratories. Significance testing
 306 among groups was conducted using paired t-tests. **c**, Comparison of FN rates of
 307 annotated and novel junctions between high-quality and low quality laboratories. Data
 308 are presented as median values (center lines) and the upper and lower quartiles (box
 309 limits). Significance testing among groups was conducted using Mann-Whitney U test.
 310 **d**, Correlation between FN rates of annotated and novel junctions and the sequencing
 311 depth in exonic regions. **e**, Comparison of several intrinsic features among true positive
 312 (TP), FP, and FN junctions, including read counts, coverage uniformity, isoform length
 313 and exon number of the associated isoforms, as well as GC content and mappability in
 314 the regions flanking the junctions. ** indicates a two-sided p -value < 0.01 , and ***
 315 indicates a two-sided p -value < 0.001 .

316 **Figure S6. Pairwise consistency in junction detection across short- and long-read**
 317 **RNA-seq.**



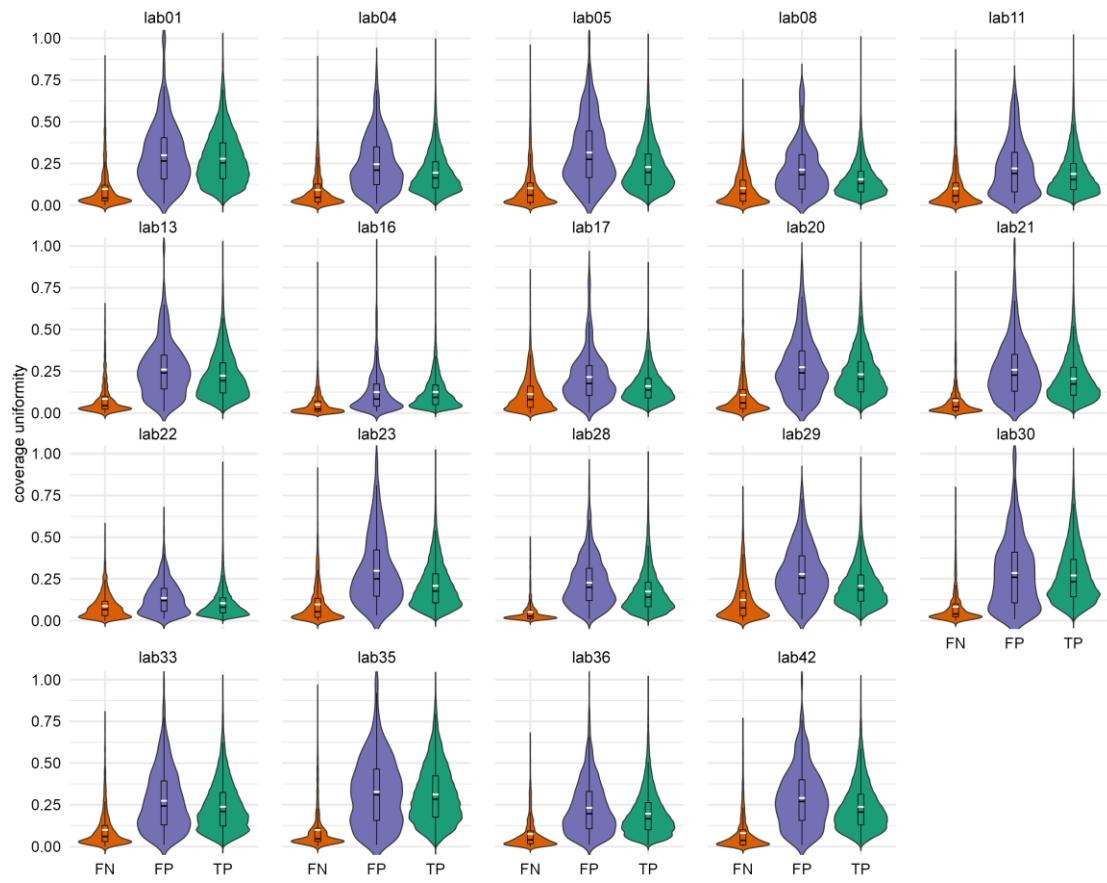
318
 319 Heatmap showing the Jaccard coefficients for junction detection results between each
 320 pair of 19 high-quality short-read RNA-seq batches from this study and 7 high-quality
 321 long-read RNA-seq batches from our companion study. Only junctions supported by
 322 ≥ 3 reads in at least two out of three replicates were included in the Jaccard coefficient
 323 calculation. Laboratories employing short-read RNA-seq clustered mainly according to
 324 the mRNA enrichment methods and strandedness, suggesting these are key factors
 325 influencing detection performance.

326 **Figure S7. Read counts distribution of isoforms associated with FN, FP, and TP**
327 **junctions.**



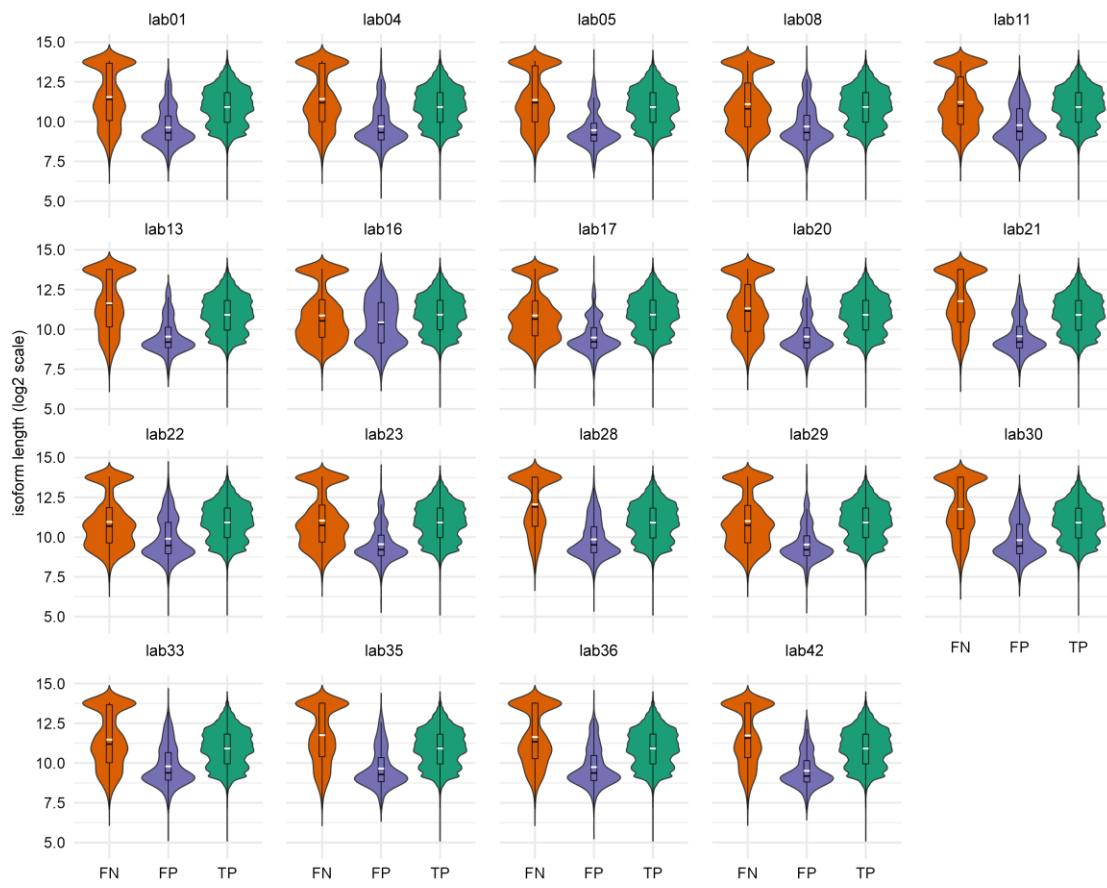
328
329 Violin plots illustrate the distribution of sequencing read counts for isoforms
330 corresponding to annotated junctions flagged as FN, FP, and TP in 19 high-quality
331 laboratories. Embedded boxplots indicate the mean (white line), median (black line),
332 and interquartile range (box limits) of the read count distribution.

333 **Figure S8. Coverage uniformity distribution of isoforms associated with FN, FP,**
334 **and TP junctions.**



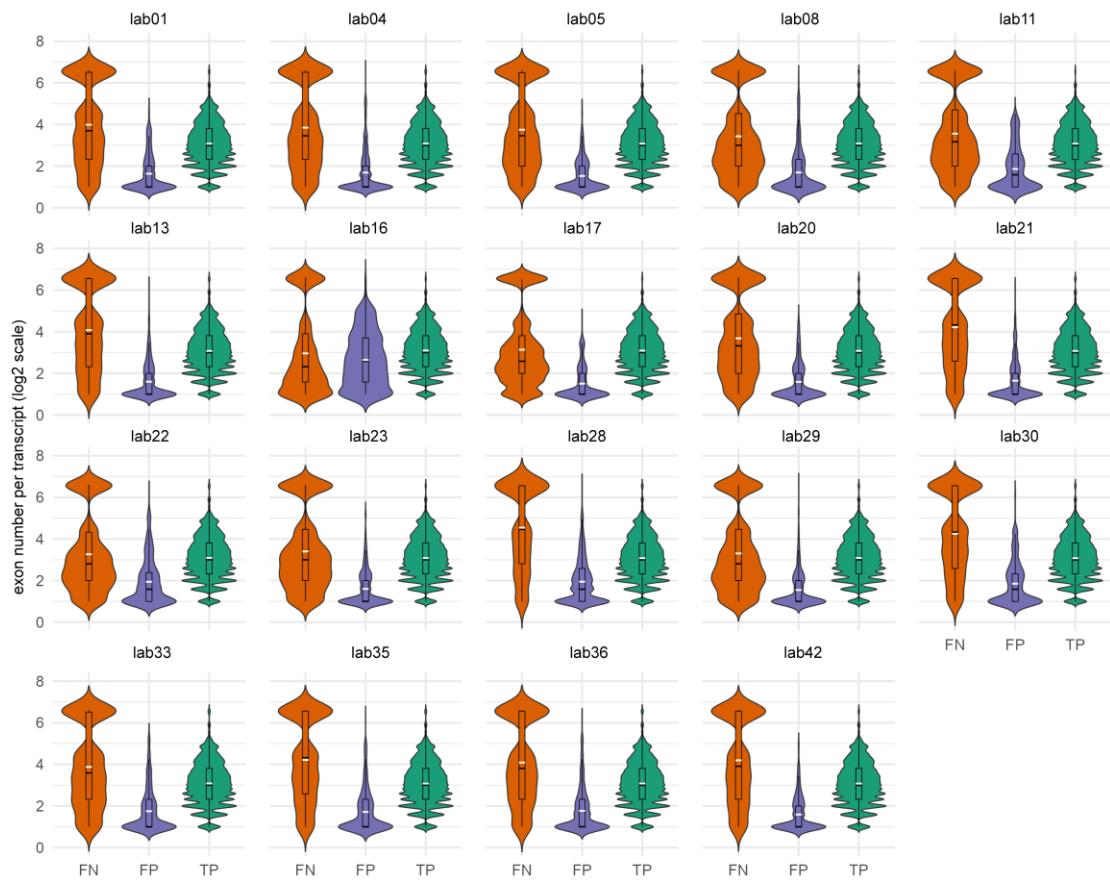
335
336 Violin plots illustrate the coverage uniformity distribution for isoforms corresponding
337 to annotated junctions flagged as FN, FP, and TP in 19 high-quality laboratories.
338 Coverage uniformity for each isoform was defined as the proportion of bases with
339 coverage depth greater than 75% of the maximum. The embedded boxplots depict the
340 mean (white line), median (black line), and the interquartile range (box limits) of the
341 coverage uniformity distribution.

342 **Figure S9. Length distribution of isoforms associated with FN, FP, and TP**
343 **junctions.**



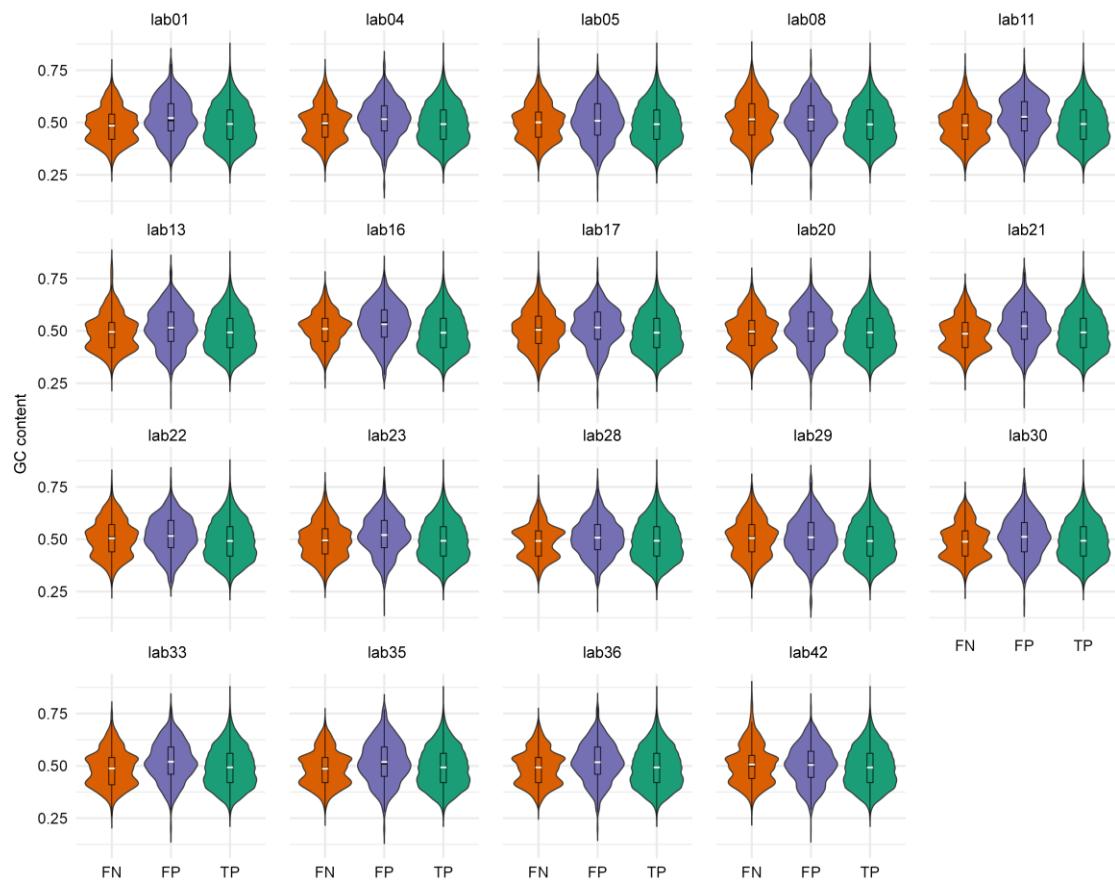
344
345 Violin plots illustrate the length distribution for isoforms corresponding to annotated
346 junctions flagged as FN, FP, and TP in 19 high-quality laboratories. The embedded
347 boxplots depict the mean (white line), median (black line), and the interquartile range
348 (box limits) of the length distribution.

349 **Figure S10. Exon count distribution of isoforms associated with FN, FP, and TP**
350 **junctions.**



351
352 Violin plots illustrate the exon number distribution of annotated junctions flagged as
353 FN, FP, and TP in 19 high-quality laboratories. The embedded boxplots depict the mean
354 (white line), median (black line), and the interquartile range (box limits) of the exon
355 number distribution.

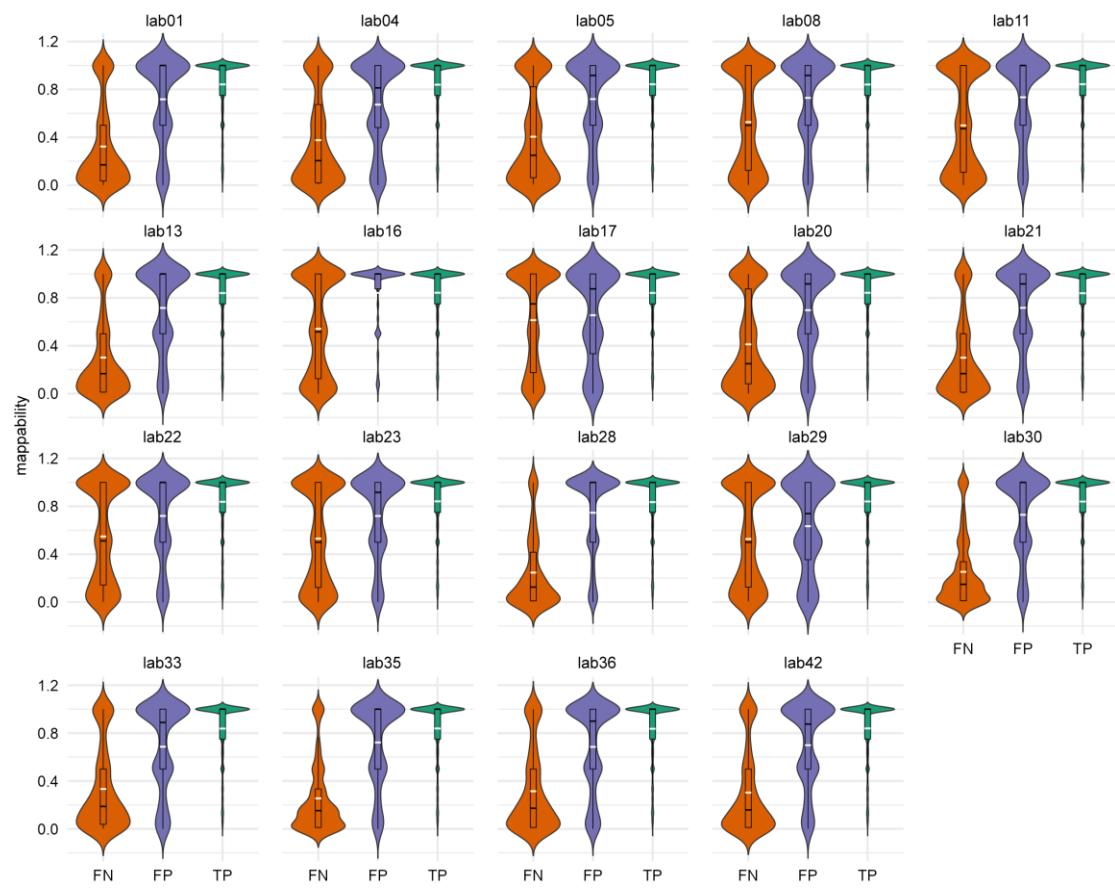
356 **Figure S11. GC content distribution of FN, FP, and TP junctions.**



357

358 Violin plots illustrate the GC content distribution of annotated junctions flagged as FN,
359 FP, and TP in 19 high-quality laboratories. GC content was calculated using a 100 bp
360 window flanking each junction, comprising 50 bp upstream of the start coordinate and
361 50 bp downstream of the end coordinate. The embedded boxplots show the mean (white
362 line), median (black line), and interquartile range (box limits) of the GC content.

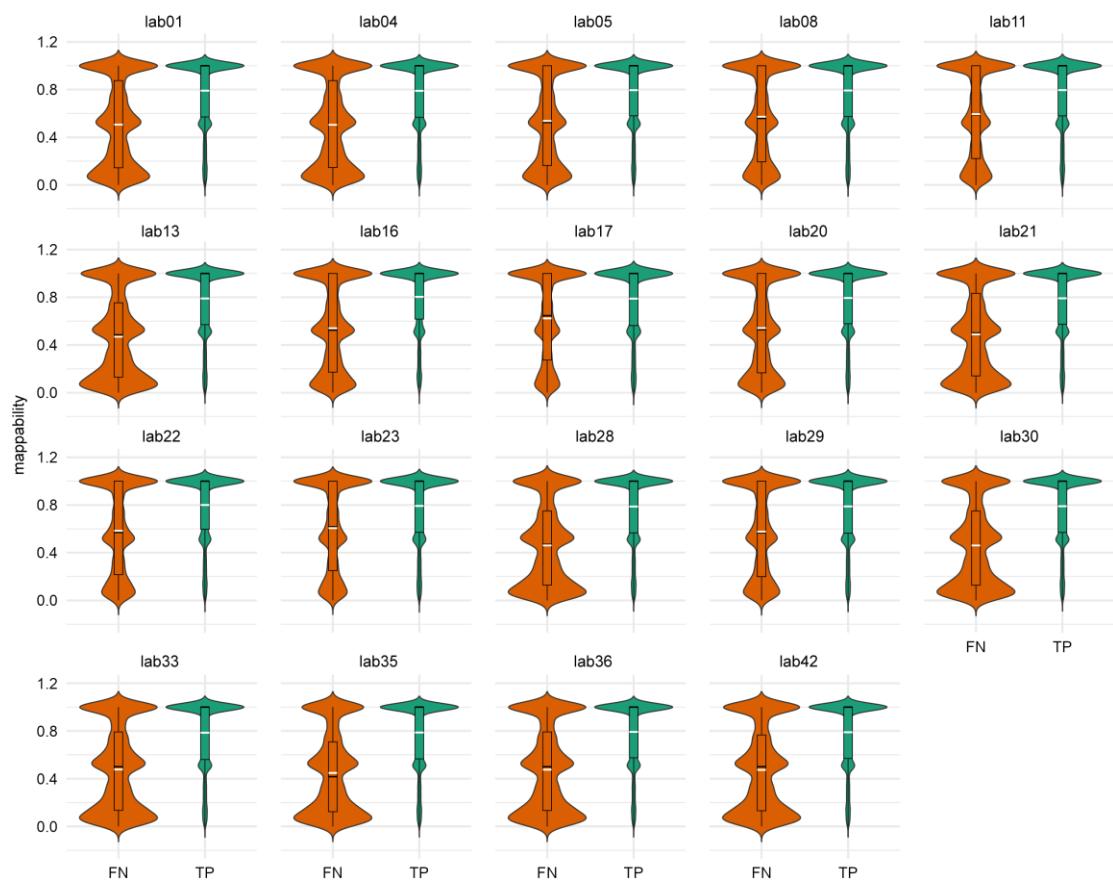
363

Figure S12. Mappability distribution of FN, FP, and TP junctions.

364

365 Violin plots illustrate the mappability distribution of annotated junctions flagged as FN,
 366 FP, and TP in 19 high-quality laboratories. Mappability was calculated using GenMap
 367 (v1.3.0)⁵ in a 50 bp window flanking each junction, comprising 25 bp upstream of the
 368 start coordinate and 25 bp downstream of the end coordinate. The embedded boxplots
 369 depict the mean (white line), median (black line), and the interquartile range (box limits)
 370 of the mappability distribution.

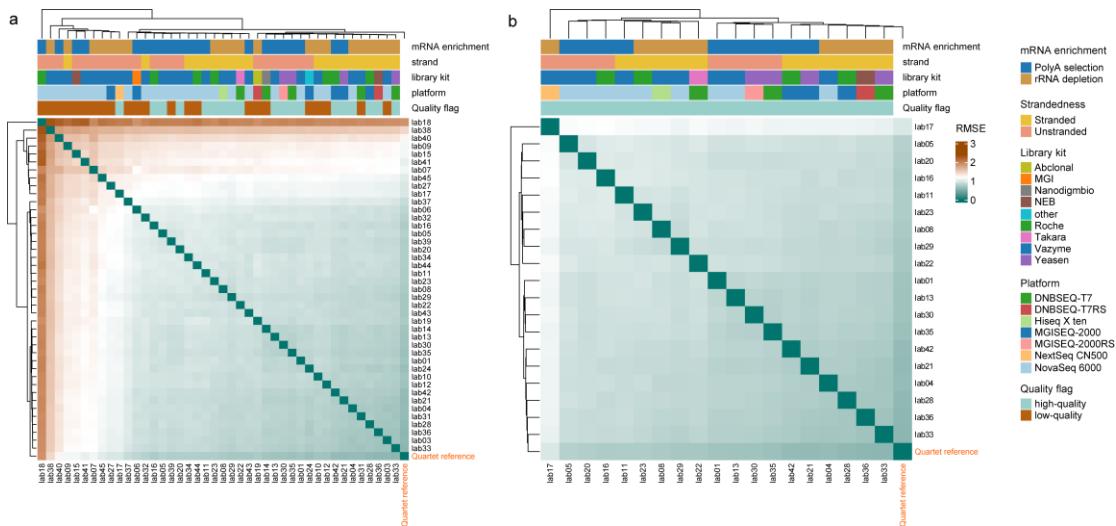
371 **Figure S13. Mappability distribution of novel junctions flagged as FN and TP.**



372

373 Violin plots illustrate the mappability distribution of novel junctions flagged as FN and
374 TP in 19 high-quality laboratories. Mappability was calculated using GenMap (v1.3.0)
375 ⁵ in a 50 bp window flanking each junction, comprising 25 bp upstream of the start
376 coordinate and 25 bp downstream of the end coordinate. The embedded boxplots depict
377 the mean (white line), median (black line), and the interquartile range (box limits) of
378 the mappability distribution.

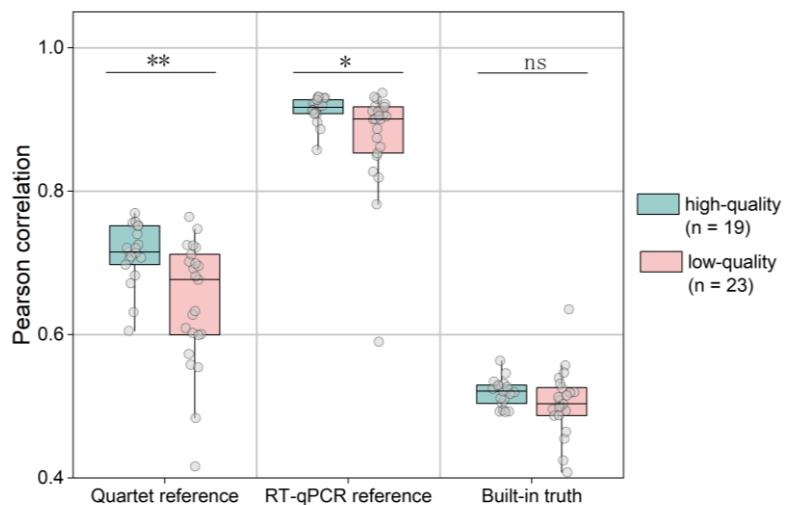
379 **Figure S14. Pairwise consistency in isoform quantification across laboratories.**



380

381 **a**, Heatmap showing the root mean square error (RMSE) of isoform-level log2FC
 382 values between each pair of the 42 participating laboratories. **b**, Heatmap showing
 383 RMSE values for log2FC comparisons among the 19 high-quality laboratories. Rows
 384 and columns are clustered based on hierarchical clustering of RMSE values.

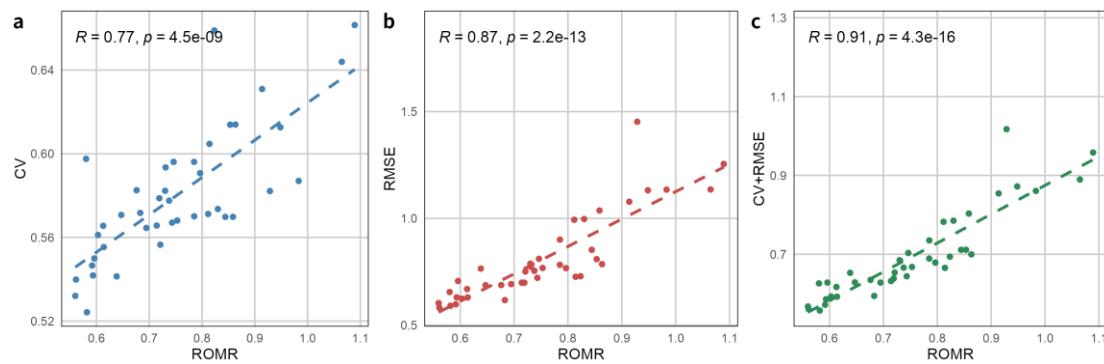
385 **Figure S15. Impact of data quality on isoform quantification accuracy.**



386

387 Isoform quantification accuracy was assessed using the Pearson correlation coefficient.
388 Statistical significance was evaluated using the Mann–Whitney U test. ** and * indicate
389 *p*-values < 0.001 and < 0.05, respectively; ns, not significant.

390 **Figure S16. Comparison of three performance metrics.**

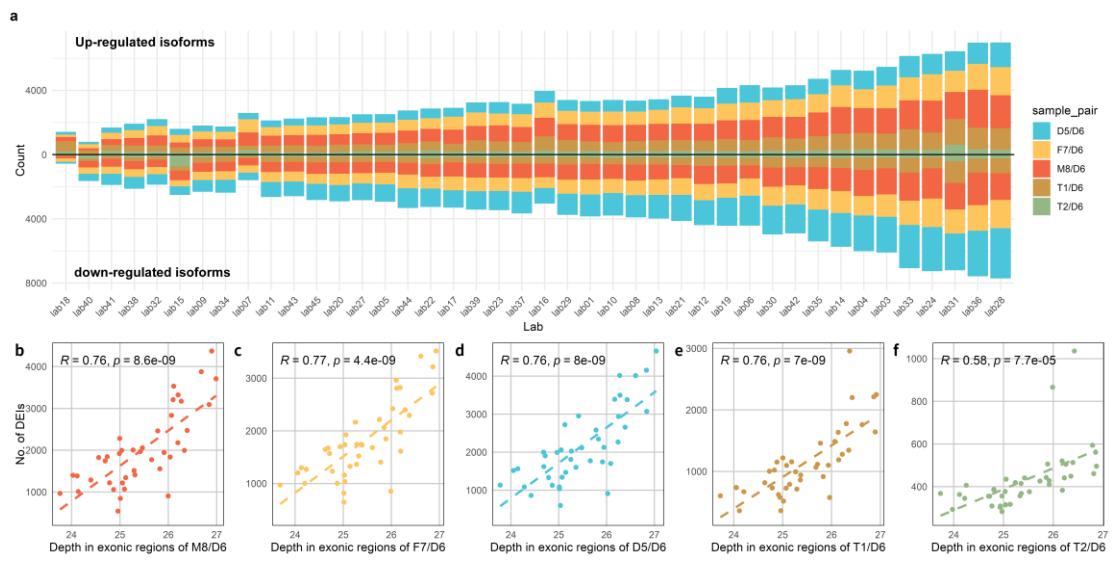


391

392 ROMR (Recovery of Mixing Ratios) measures how well RNA-seq expression data
393 reflect the known mixing ratios of mixed samples. Correlation of ROMR with (a)
394 coefficient of variation (CV), (b) root mean square error (RMSE), and (c) the average
395 of CV and RMSE. ROMR correlates most strongly with the combined accuracy and
396 reproducibility metric (average of CV and RMSE), suggesting its utility as an integrated
397 performance indicator. R represents the Pearson correlation coefficient.

398

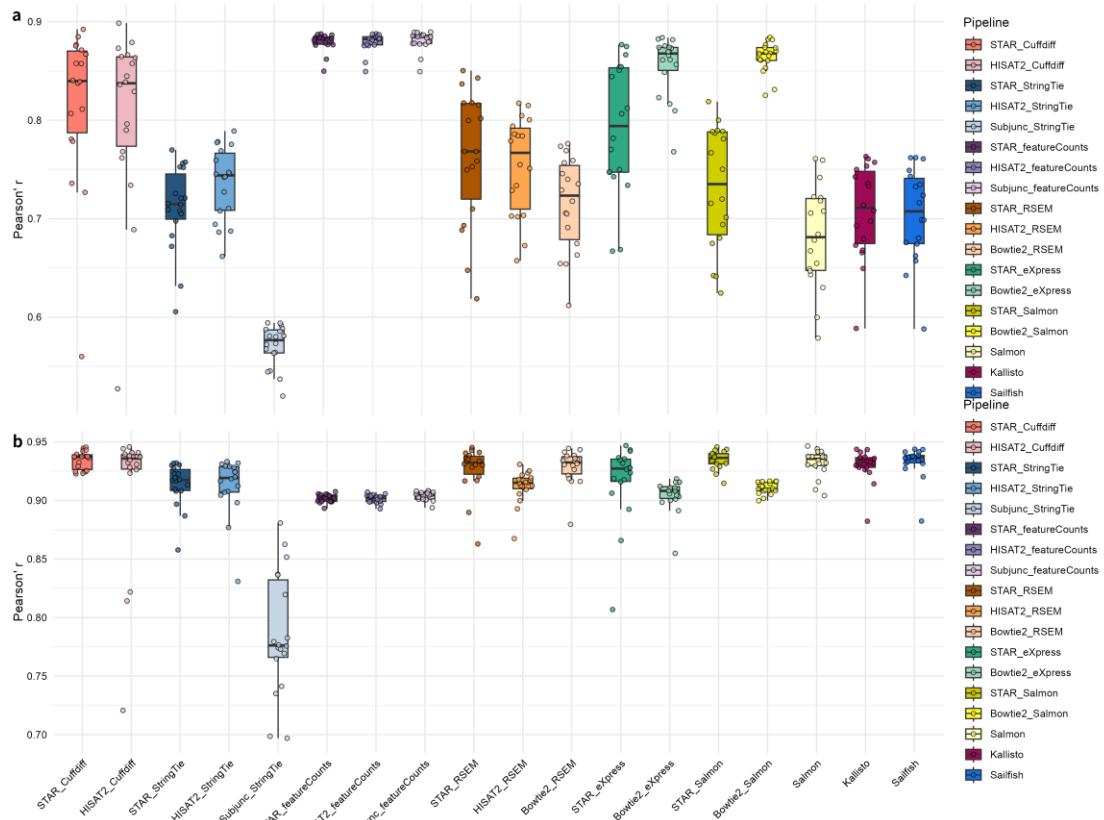
Figure S17. Number of DEIs detected by laboratories.



399

400 **a**, Total number of DEIs identified by each laboratory, including up-regulated and
 401 down-regulated isoforms. **b-f**, A positive correlation was observed between sequencing
 402 depth in exonic regions and the number of detected DEIs for the (b) M8 vs. D6, (c) F7
 403 vs. D6, (d) D5 vs. D6, (e) T1 vs. D6, (f) T2 vs. D6 sample pairs. *R* represents the Pearson
 404 correlation coefficient.

405 **Figure S18. PCCs of 18 isoform quantification pipelines.**



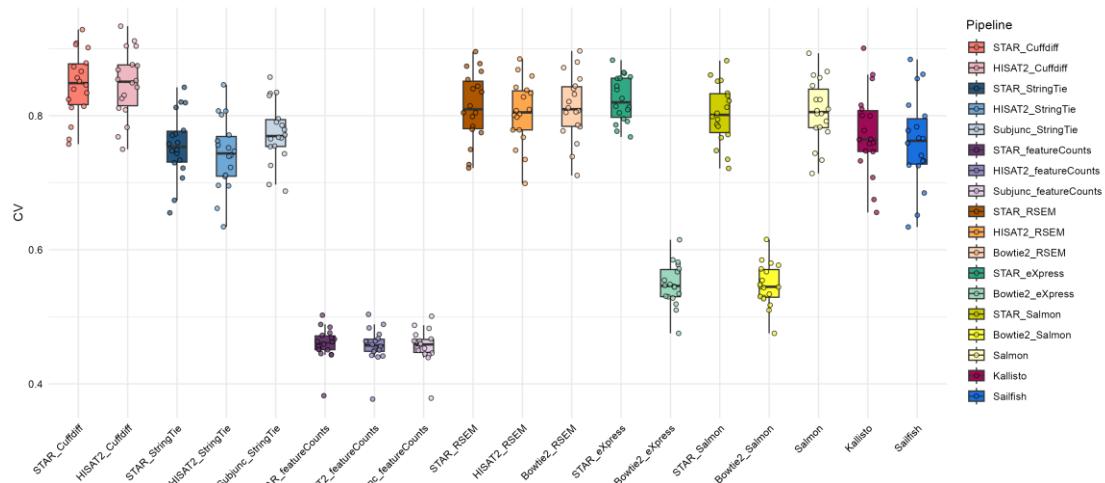
406

407 Boxplots show the distribution of Pearson correlation coefficients (PCCs) across 19
 408 high-quality datasets for each isoform quantification pipelines based on the (a) Quartet
 409 and (b) RT-qPCR reference datasets. Data are presented as median values (center lines)
 410 and the upper and lower quartiles (box limits).

411 **Figure S19. Comparison of CV across 19 quantification pipelines.**

412

413 Boxplots show the distribution of average coefficient of variation (CV) values of
 414 isoform expression levels among replicates in the Quartet samples across 19
 415 quantification pipelines. Individual points represent CV values from high-quality
 416 laboratories. Data are presented as median values (center lines) and the upper and lower
 417 quartiles (box limits).



418 **Figure S20. Comparison of SNR across 18 quantification pipelines.**

419

420 Boxplots display the distribution of SNR values for isoform expression data from the
 421 Quartet and mixed samples across 18 quantification pipelines. Individual points
 422 represent SNR values for high-quality laboratories. Data are presented as median values
 423 (center lines) and the upper and lower quartiles (box limits).

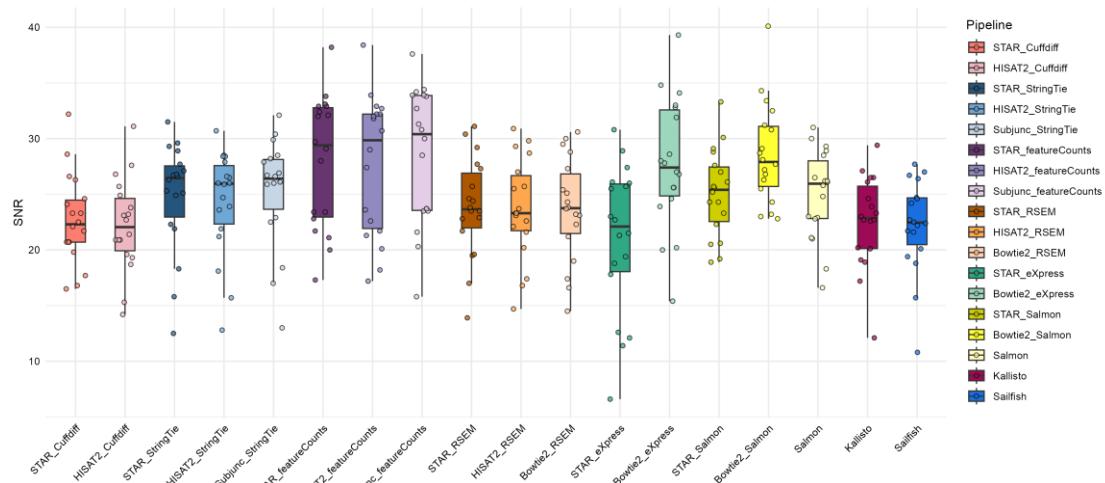
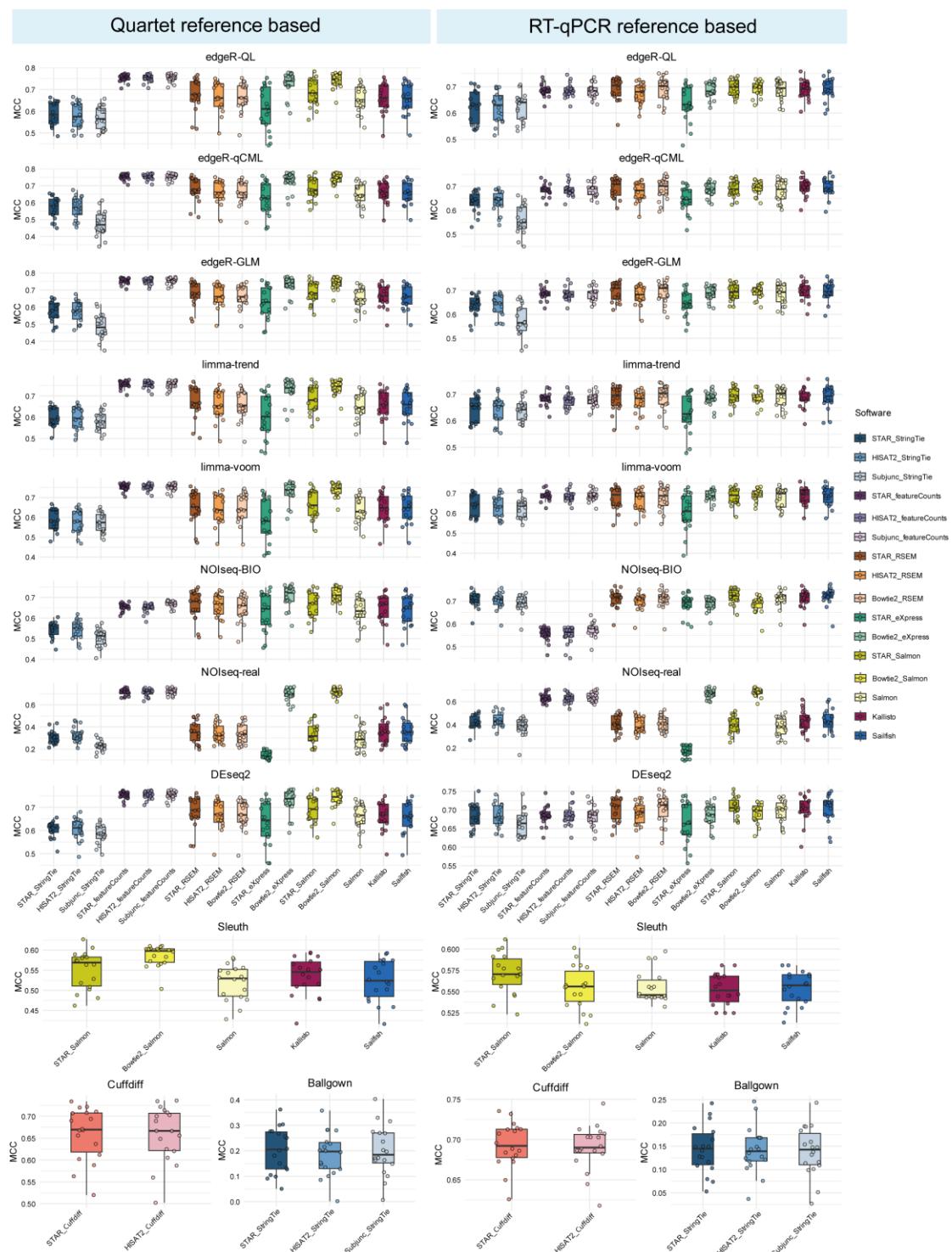
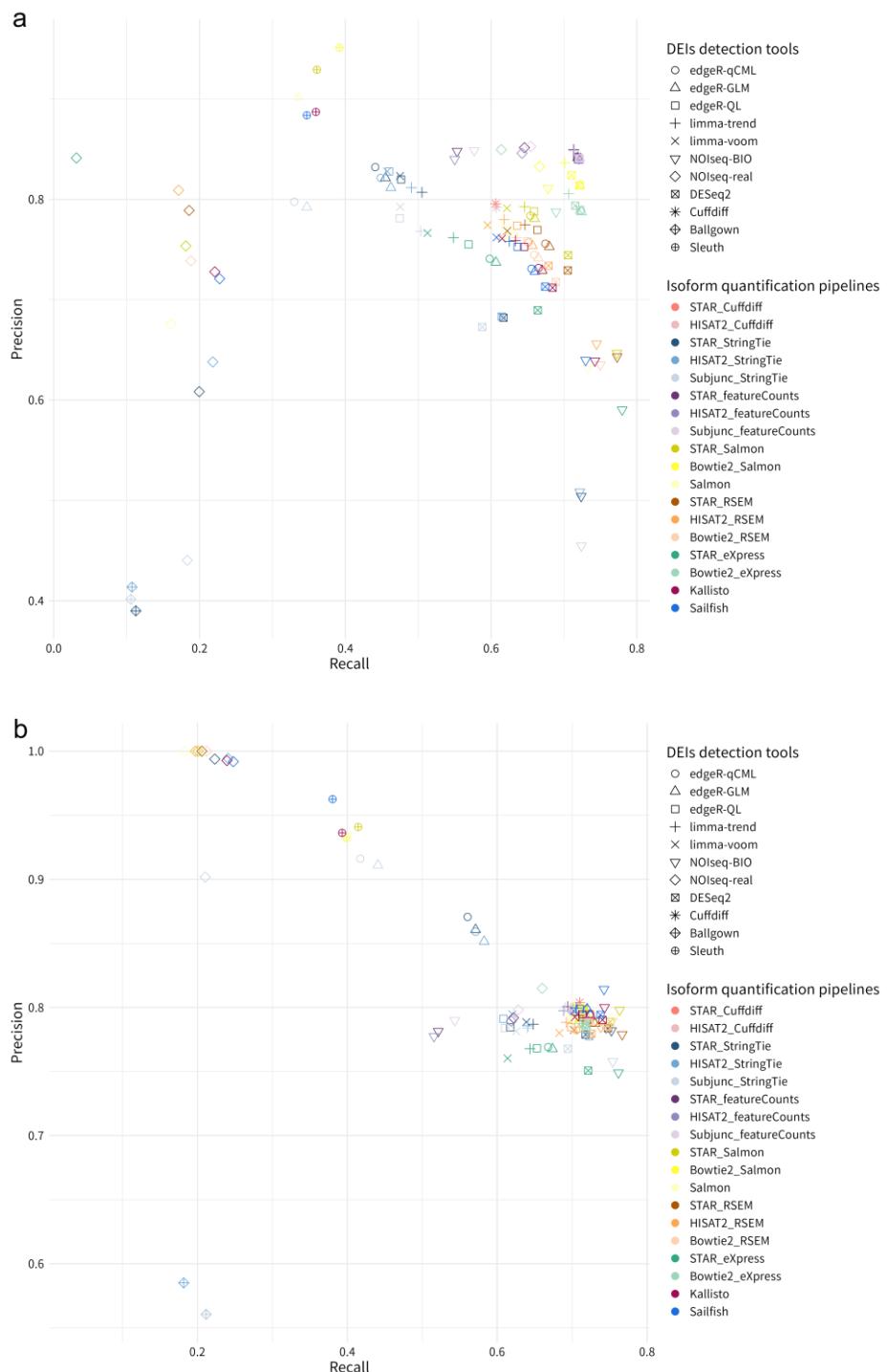


Figure S21. Impact of isoform quantification tools on DEI detection.

426 Boxplots show the distribution of Matthews correlation coefficients (MCC) across 19
 427 high-quality datasets for each differential expression analysis software, stratified by
 428 different isoform quantification tools. Data are presented as median values (center lines)
 429 and the upper and lower quartiles (box limits).

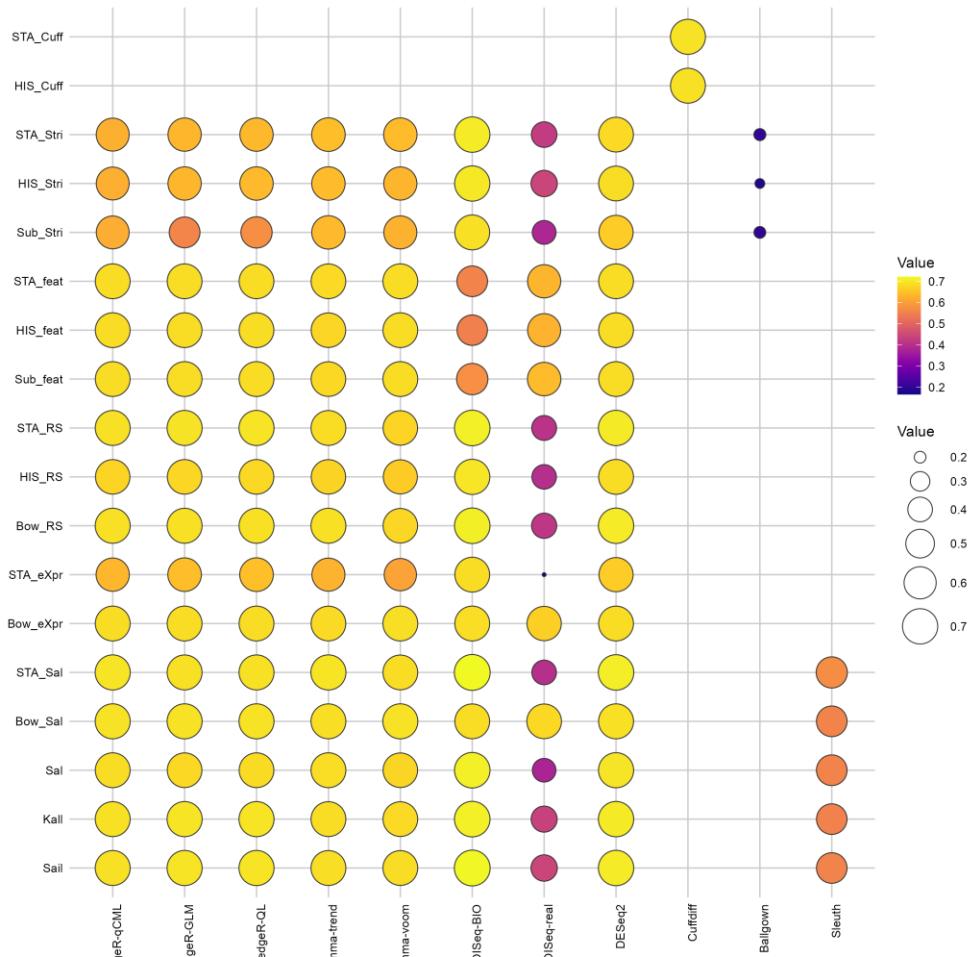
430 **Figure S22. Recall and precision of differential expression analysis pipelines.**



431

432 The scatter plot illustrates the recall and precision of 138 differential expression
 433 analysis pipelines based on the (a) Quartet and (b) RT-qPCR reference datasets. Each
 434 point represents the mean values calculated from the analysis results of 19 high-quality
 435 RNA-seq data. Different colors indicate distinct quantification pipelines, whereas
 436 different shapes denote different differential expression analysis tools.

437 **Figure S23. Performance of DEI detection tools combined with isoform**
 438 **quantification pipelines.**



439
 440 Bubble plots show the mean Matthews correlation coefficient (MCC) across 19 high-
 441 quality datasets for each combination. Both the size and the color intensity of the
 442 bubbles indicate the magnitude of MCC.

443 **Figure S24. Impact of isoform complexity (K-value) on isoform quantification**
444 **pipelines.**



445
446 Boxplots display the performance of 18 isoform quantification pipelines across 19 high-
447 quality datasets in relation to K-value of isoforms based on the (a) Quartet reference
448 datasets and (b) the recovery of mixing ratios against the built-in truth. Data are
449 presented as median values (center lines) and the upper and lower quartiles (box limits).

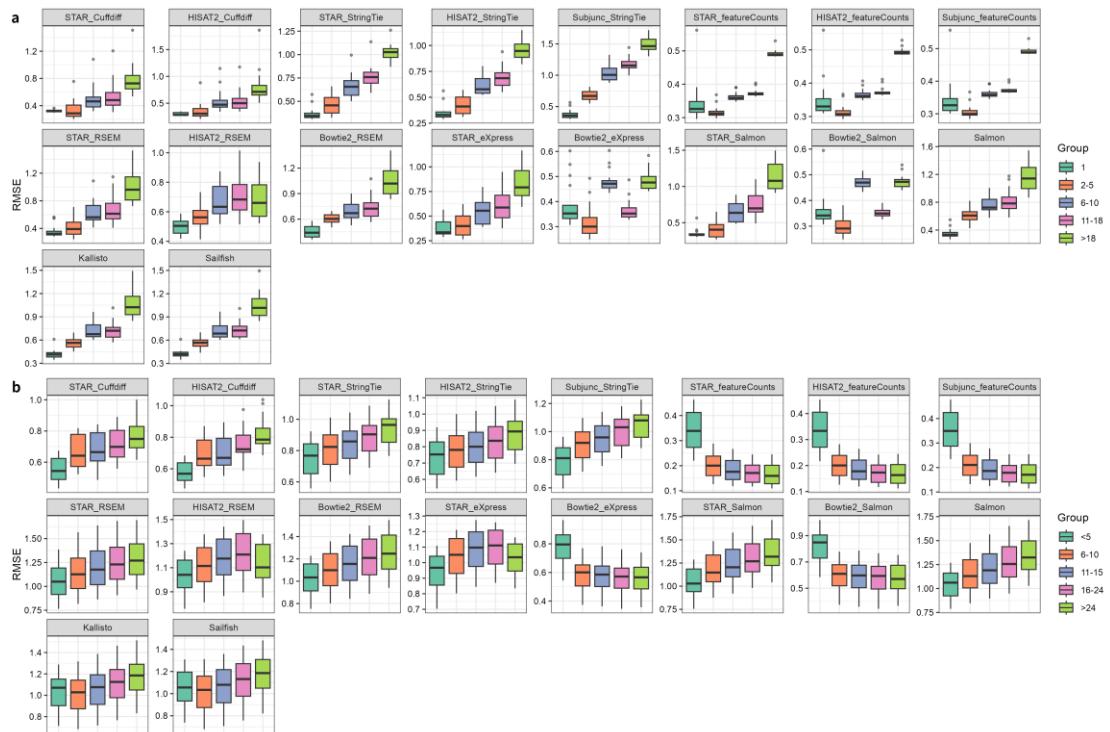
450

Figure S25. Impact of isoform length on isoform quantification pipelines.

451

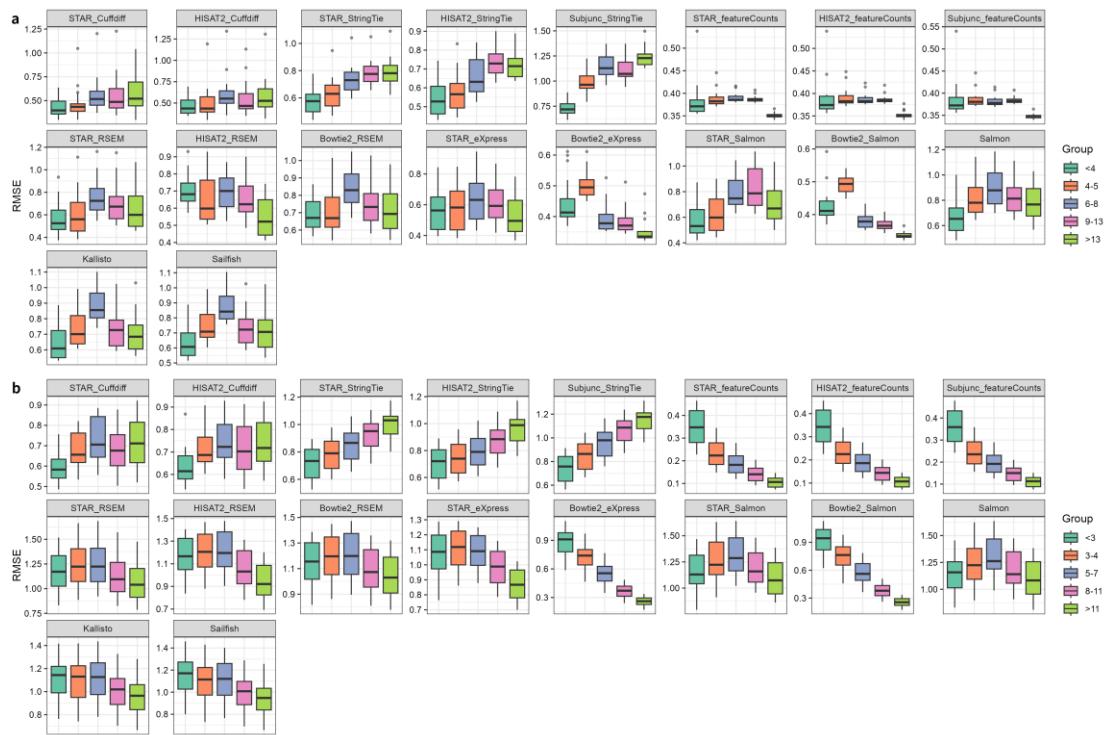
452 Boxplots display the performance of 18 isoform quantification pipelines across 19 high-
 453 quality datasets in relation to length of isoforms based on the (a) Quartet reference
 454 datasets and (b) the recovery of mixing ratios against the built-in truth. Data are
 455 presented as median values (center lines) and the upper and lower quartiles (box limits).

456 **Figure S26. Impact of isoform number per gene on isoform quantification**
457 **pipelines.**



458
459 Boxplots display the performance of 18 isoform quantification pipelines across 19 high-
460 quality datasets in relation to isoform number per gene based on the (a) Quartet
461 reference datasets and (b) the recovery of mixing ratios against the built-in truth. Data
462 are presented as median values (center lines) and the upper and lower quartiles (box
463 limits).

464 **Figure S27. Impact of exon number per isoform on isoform quantification**
465 **pipelines.**



466
467 Boxplots display the performance of 18 isoform quantification pipelines across 19 high-
468 quality datasets in relation to exon number per isoform based on the (a) Quartet
469 reference datasets and (b) the recovery of mixing ratios against the built-in truth. Data
470 are presented as median values (center lines) and the upper and lower quartiles (box
471 limits).

472

Figure S28. Impact of exon length on isoform quantification pipelines.

473

474 Boxplots display the performance of 18 isoform quantification pipelines across 19 high-
 475 quality datasets in relation to mean exon length of each isoform based on the (a) Quartet
 476 reference datasets and (b) the recovery of mixing ratios against the built-in truth. Data
 477 are presented as median values (center lines) and the upper and lower quartiles (box
 478 limits).

479

Figure S29. Impact of GC content on isoform quantification pipelines.



480

Boxplots display the performance of 18 isoform quantification pipelines across 19 high-quality datasets in relation to GC content based on the (a) Quartet reference datasets and (b) the recovery of mixing ratios against the built-in truth. Data are presented as median values (center lines) and the upper and lower quartiles (box limits).

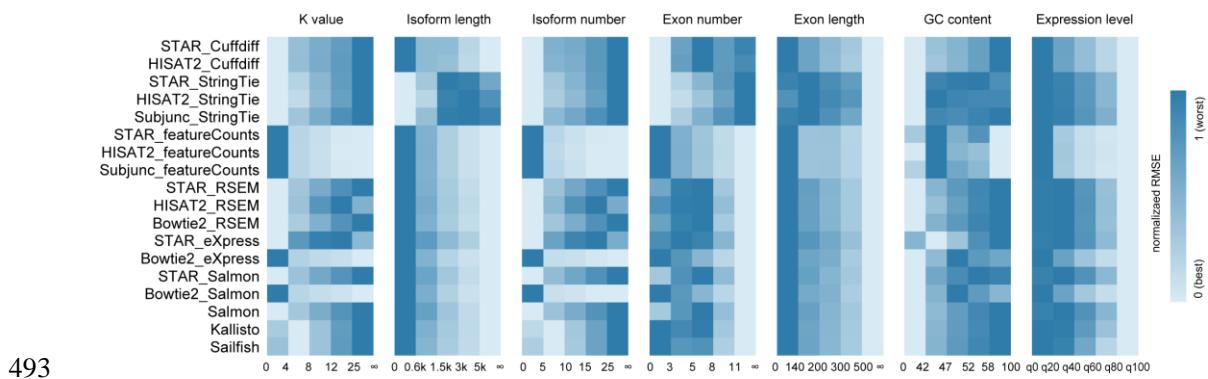
485

Figure S30. Impact of isoform expression on isoform quantification pipelines.

486

487 Boxplots display the performance of 18 isoform quantification pipelines across 19 high-
 488 quality datasets in relation to isoform expression level based on the (a) Quartet
 489 reference datasets and (b) the recovery of mixing ratios against the built-in truth. Data
 490 are presented as median values (center lines) and the upper and lower quartiles (box
 491 limits).

492 **Figure S31. Impact of isoform features based on the built-in truth.**

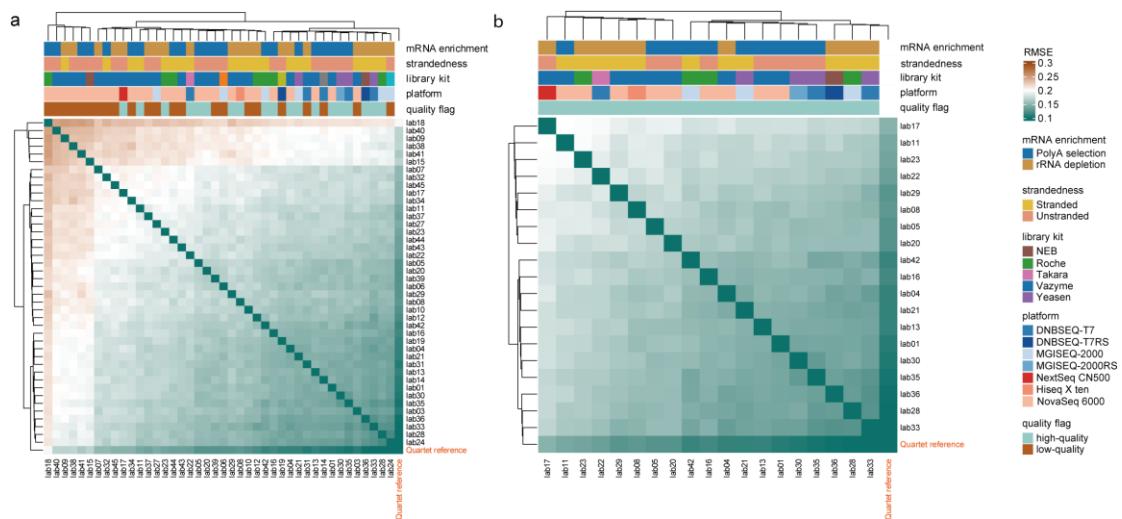


493

494 Evaluation of quantification pipelines with respect to multiple isoform features based
 495 on the built-in truth, including the isoform length, the number of isoforms per gene, the
 496 average length of exons per isoform, the number of exons per isoform, GC content, K-
 497 value (exon-isoform structural complexity), and expression level (FPKM).

498

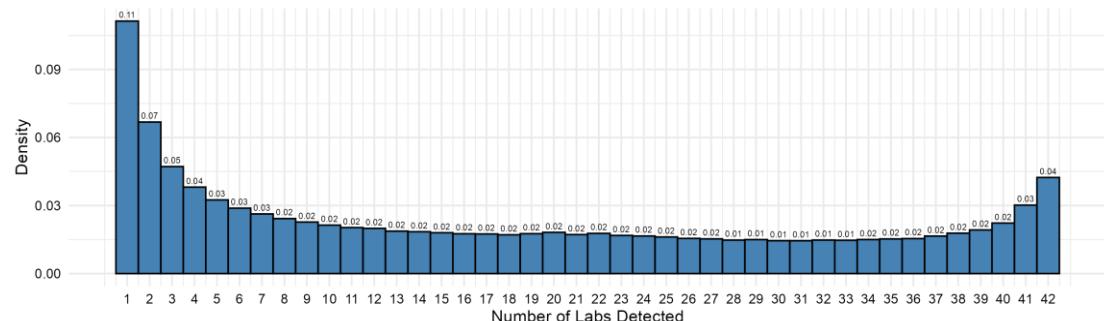
Figure S32. Pairwise consistency in AS event quantification across laboratories.



499

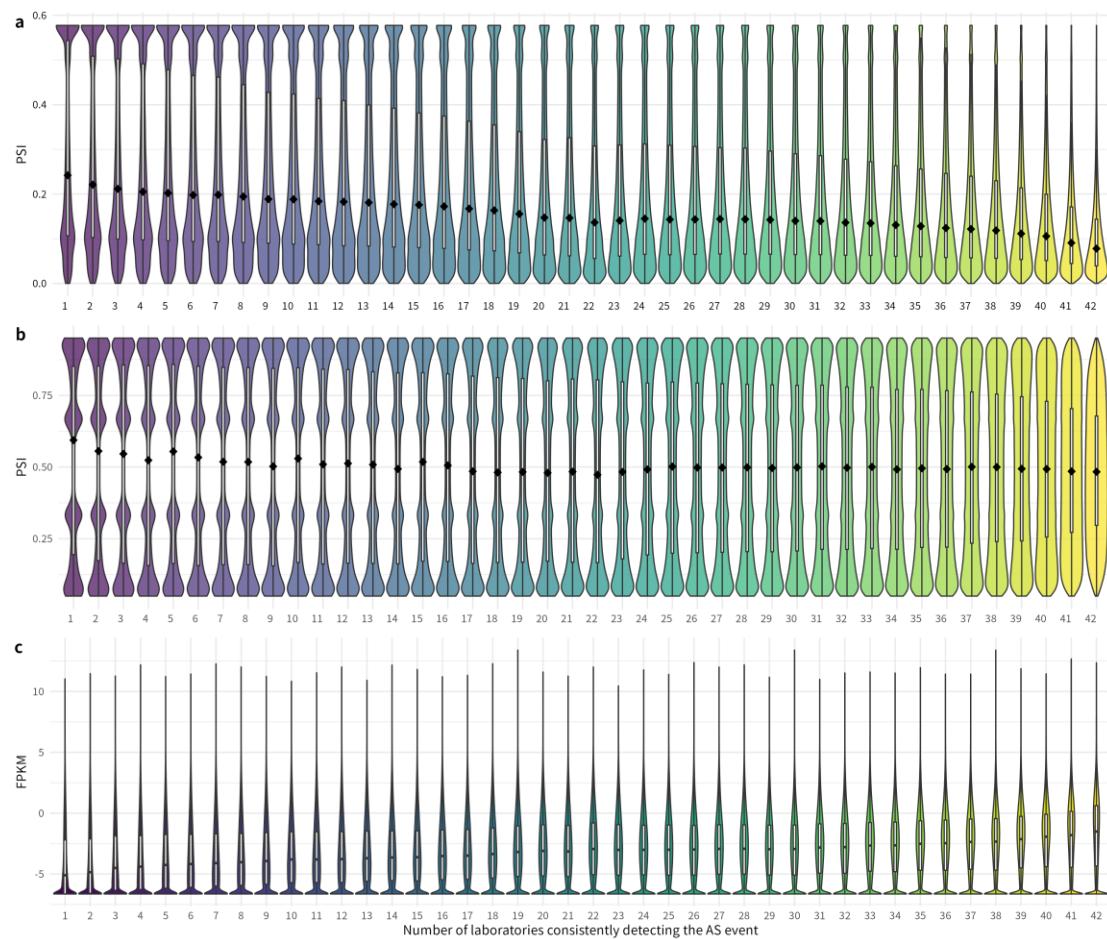
500 **a**, Heatmap showing the root mean square error (RMSE) of alternative splicing (AS)
501 event-level delta percent spliced-in (dPSI) values between each pair of the 42
502 participating laboratories. **b**, Heatmap showing RMSE values for dPSI comparisons
503 among the 19 high-quality laboratories. Rows and columns are clustered based on
504 hierarchical clustering of RMSE values.

505 **Figure S33. Consistency of ASE detection across 42 laboratories.**



506
507 The density plot illustrates the distribution of alternative splicing events (ASEs) with
508 respect to the number of laboratories that commonly detected them.

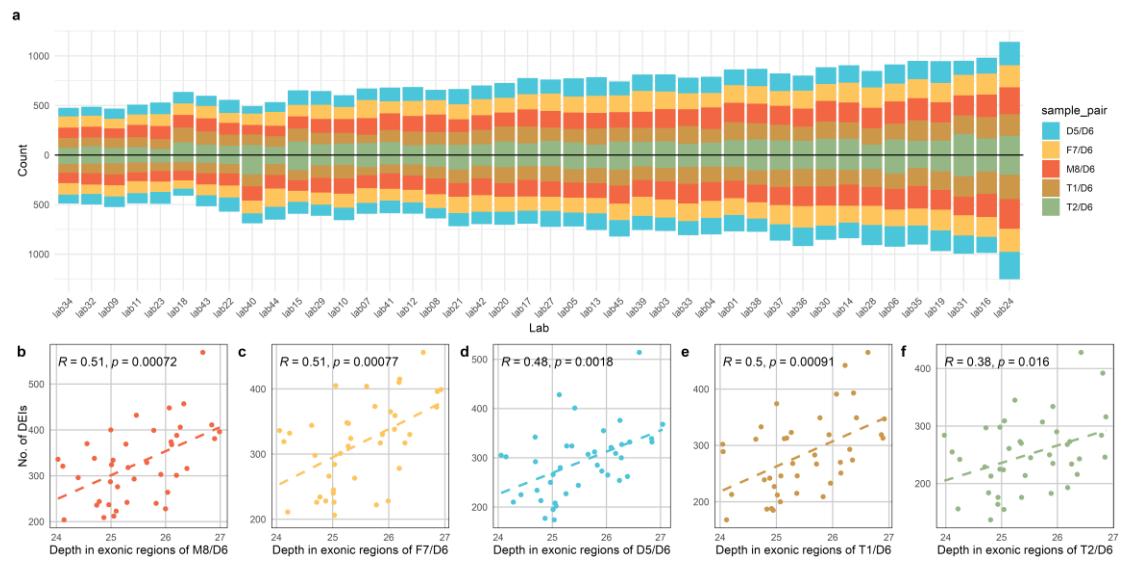
509

Figure S34. Characteristics of ASEs with different detection consistency.

510

511 **a**, Percent spliced-in (PSI) levels, **b**, PSI reproducibility (standard deviation), **c**, and
 512 expression levels of isoforms associated with ASEs, respectively, stratified by their
 513 detection consistency across 42 laboratories from low to high.

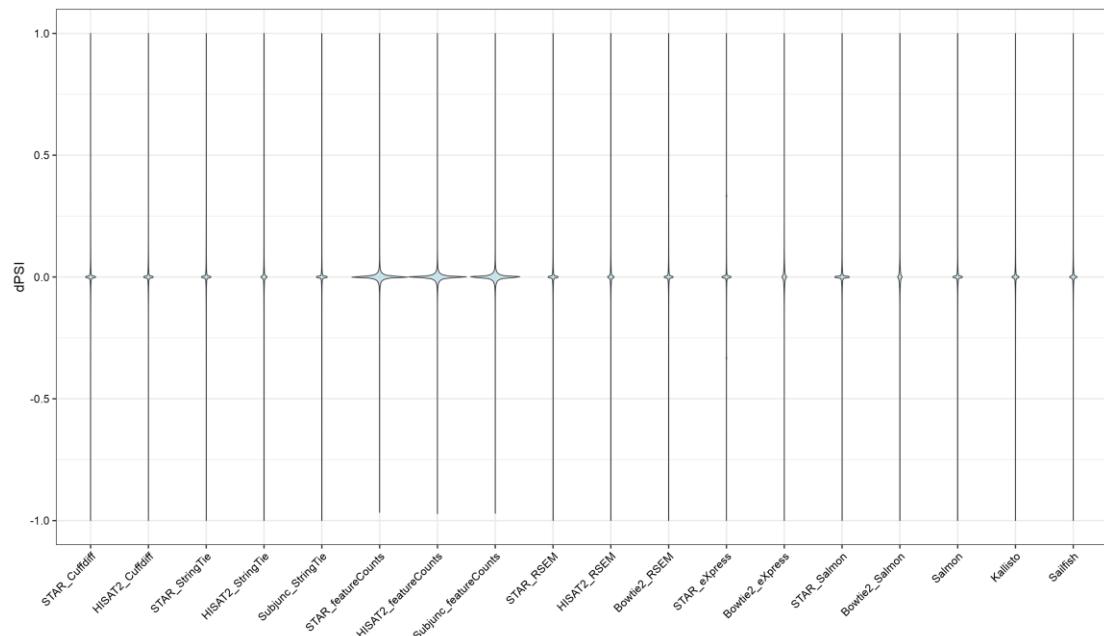
514 **Figure S35. Number of DSEs detected by laboratories.**



515

516 **a**, Total number of differential splicing events (DSEs) identified by each laboratory,
 517 including up-regulated and down-regulated events. **b–f**, A positive correlation was
 518 observed between sequencing depth in exonic regions and the number of detected DSEs
 519 for the (b) M8 vs. D6, (c) F7 vs. D6, (d) D5 vs. D6, (e) T1 vs. D6, and (f) T2 vs. D6
 520 sample pairs. R represents the Pearson correlation coefficient.

521 **Figure S36. Distribution of dPSI values detected by SUPPA2 combined with 18**
522 **isoform quantification pipelines.**

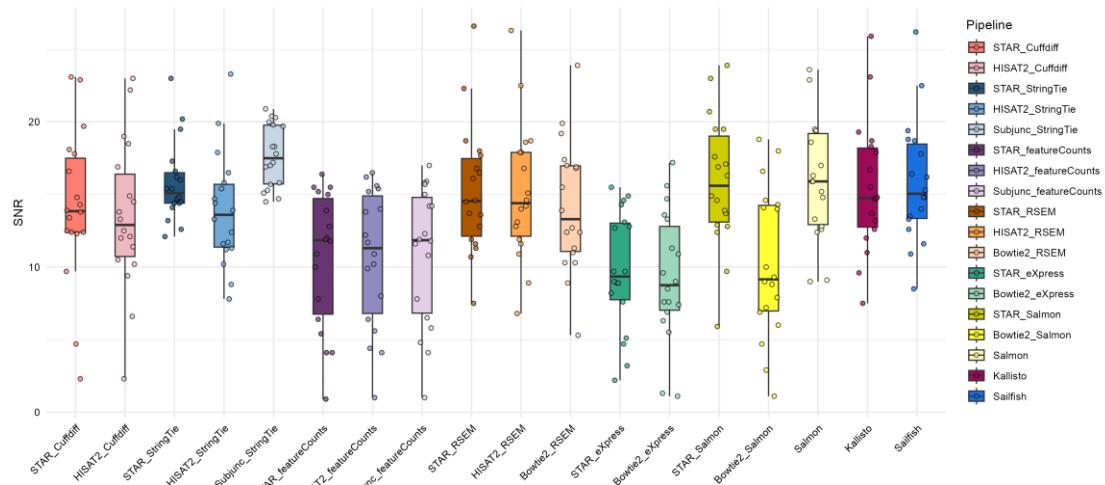


523
524 Ridgeline Plot show the delta percent spliced-in (dPSI) distributions analyzed by
525 SUPPA2 using TPM data from 18 different isoform quantification pipelines for all
526 detected alternative splicing (AS) events. The dPSI distribution of AS events included
527 in the Quartet reference datasets are showed in Fig.5d.

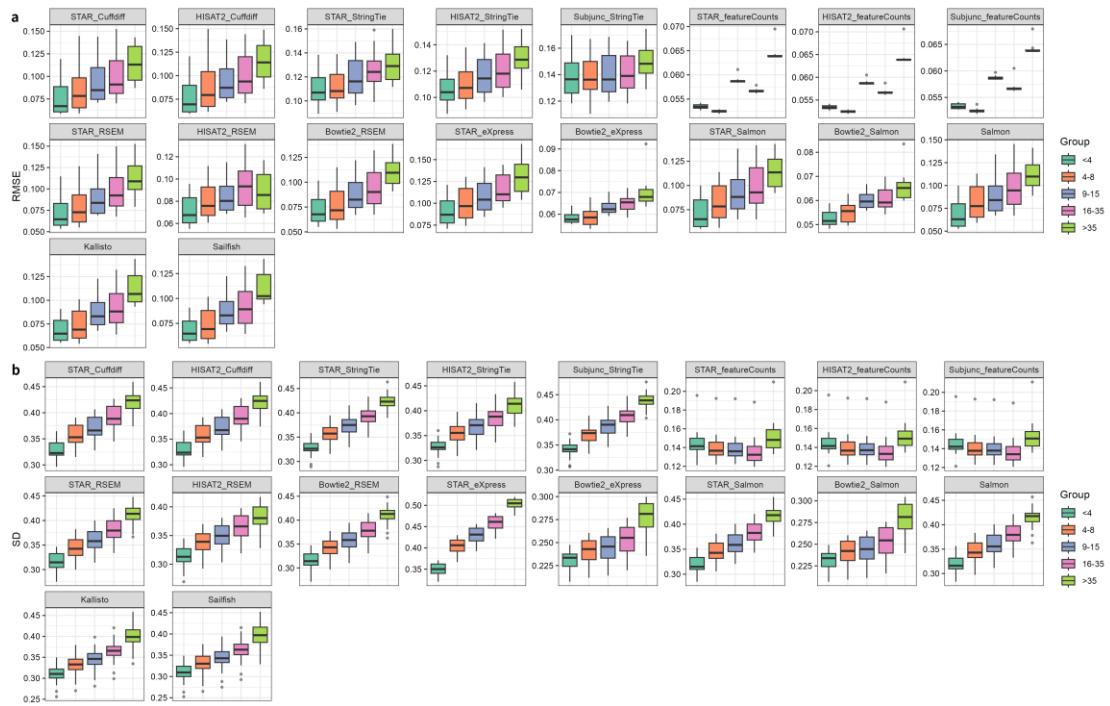
528 **Figure S37. SNR for SUPPA2 combined with 18 isoform quantification pipelines.**

529

530 Violin plots show the SNR distributions of quantification data (PSI) analyzed by
 531 SUPPA2 using TPM data from 18 different isoform quantification pipelines.



532 **Figure S38. Impact of the number of adjacent AS events on AS detection pipelines.**



533

534 Boxplots showing the (a) accuracy and (b) reproducibility of SUPPA2 combined with
 535 18 isoform quantification pipelines as a function of the number of AS events in adjacent
 536 regions. Reproducibility was assessed using the standard deviation (SD). Data are
 537 shown as median values (center lines) with upper and lower quartiles (box boundaries).

538 **Figure S39. Impact of AS event-associated isoform number on AS detection**
539 **pipelines.**



540
541 Boxplots showing the (a) accuracy and (b) reproducibility of SUPPA2 combined with
542 18 isoform quantification pipelines as a function of the isoform number associated with
543 AS events. Reproducibility was assessed using the standard deviation (SD). Data are
544 shown as median values (center lines) with upper and lower quartiles (box boundaries).

545

Figure S40. Impact of the GC content on AS detection pipelines.

546

547 Boxplots showing the (a) accuracy and (b) reproducibility of SUPPA2 combined with
 548 18 isoform quantification pipelines as a function of the GC content. The GC content of
 549 each AS event was calculated from two 75 bp regions flanking its splice junctions.
 550 Reproducibility was assessed using the standard deviation (SD). Data are shown as
 551 median values (center lines) with upper and lower quartiles (box boundaries).

552 **Figure S41. Impact of the mappability on AS detection pipelines.**



553

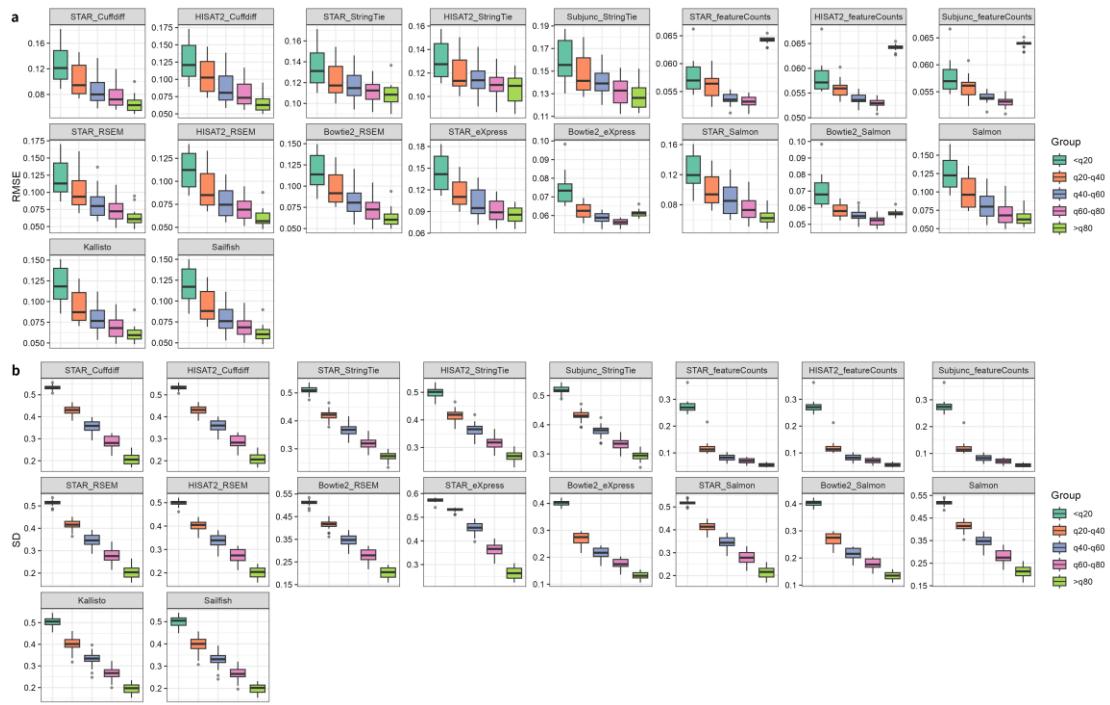
554 Boxplots showing the (a) accuracy and (b) reproducibility of SUPPA2 combined with
 555 18 isoform quantification pipelines as a function of the mappability. The mappability
 556 of each AS event was calculated from two 75 bp regions flanking its splice junctions.
 557 Reproducibility was assessed using the standard deviation (SD). Data are shown as
 558 median values (center lines) with upper and lower quartiles (box boundaries).

559 **Figure S42. Impact of PSI levels on AS detection pipelines.**



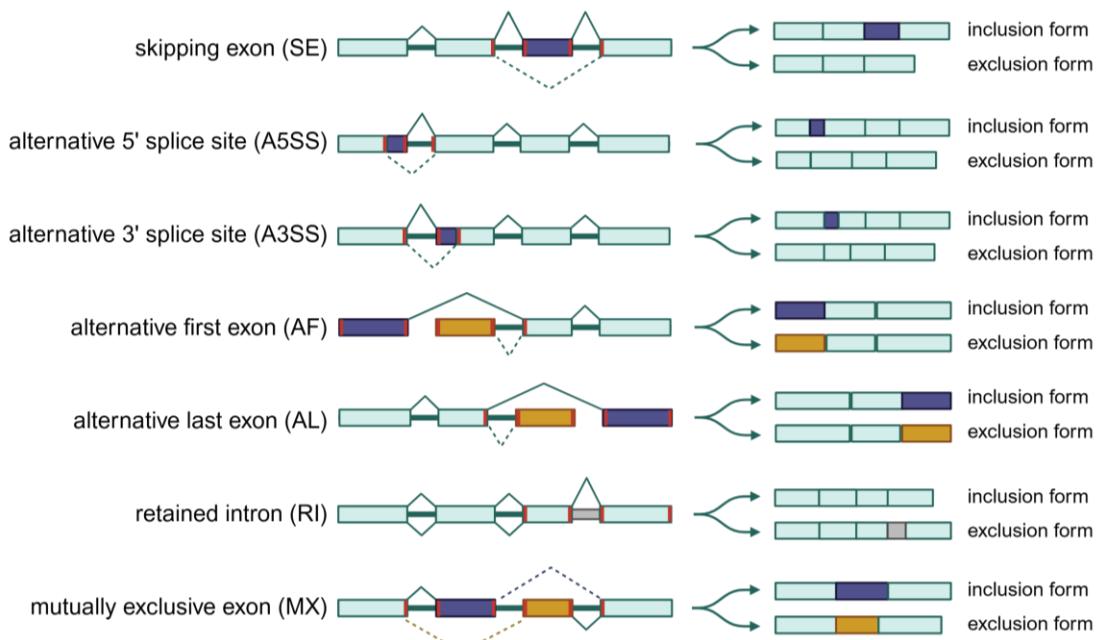
560
561 Boxplots showing the (a) accuracy and (b) reproducibility of SUPPA2 combined with
562 18 isoform quantification pipelines as a function of AS event PSI levels in 19 high-
563 quality RNA-seq datasets. Reproducibility was assessed using the standard deviation
564 (SD). Data are shown as median values (center lines) with upper and lower quartiles
565 (box boundaries).

566 **Figure S43. Impact of AS event-associated isoform expression levels on AS**
567 **detection pipelines.**



568 Boxplots showing the (a) accuracy and (b) reproducibility of SUPPA2 combined with
569 18 isoform quantification pipelines as a function of AS-related isoform expression
570 levels in 19 high-quality RNA-seq datasets. Isoform expression levels were defined as
571 the average FPKM of all isoforms related to each AS event. Reproducibility was
572 assessed using the standard deviation (SD). Data are presented as median values (center
573 lines) with upper and lower quartiles (box boundaries).

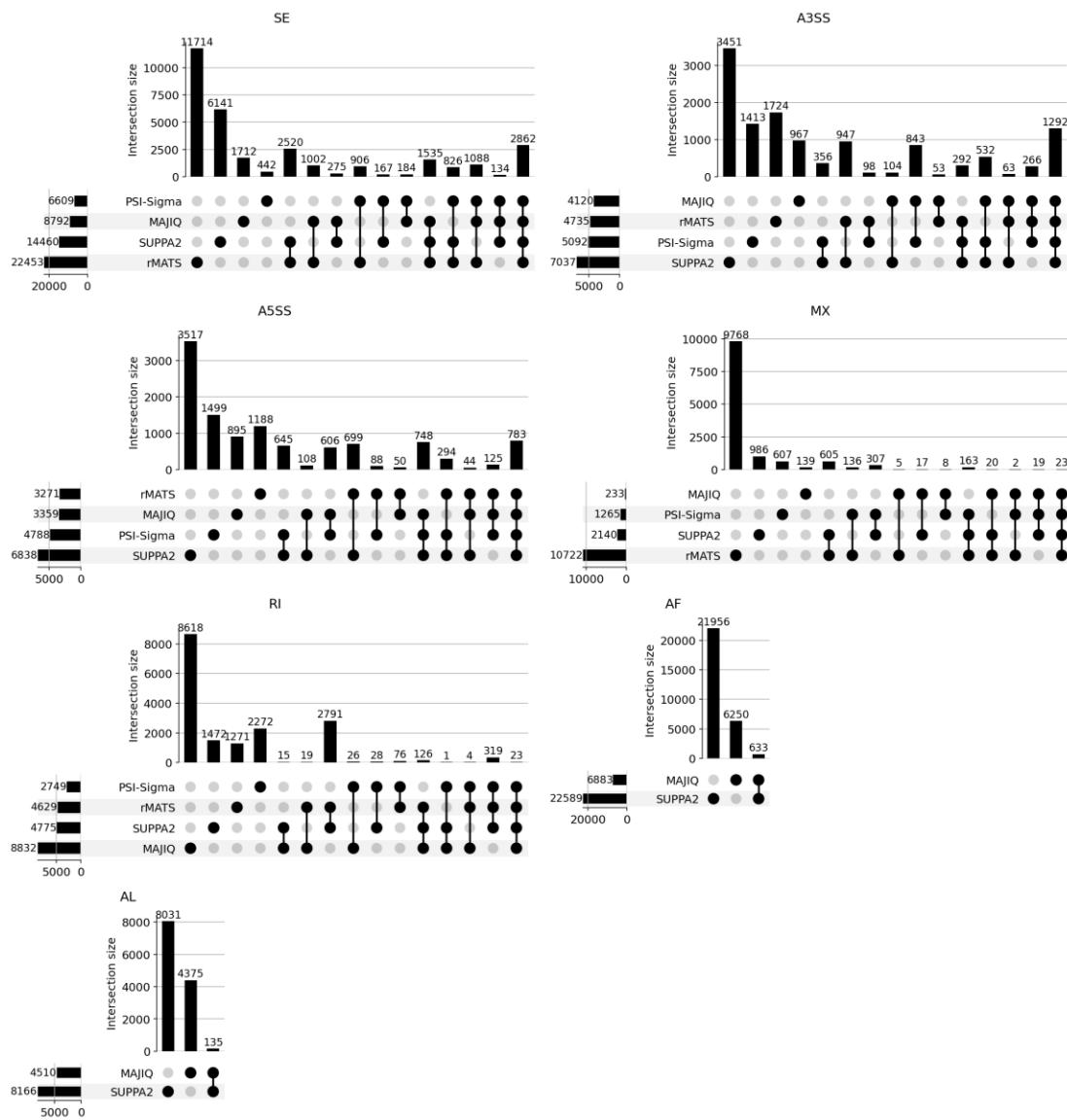
575 **Figure S44. A unified format for different AS event types.**



576

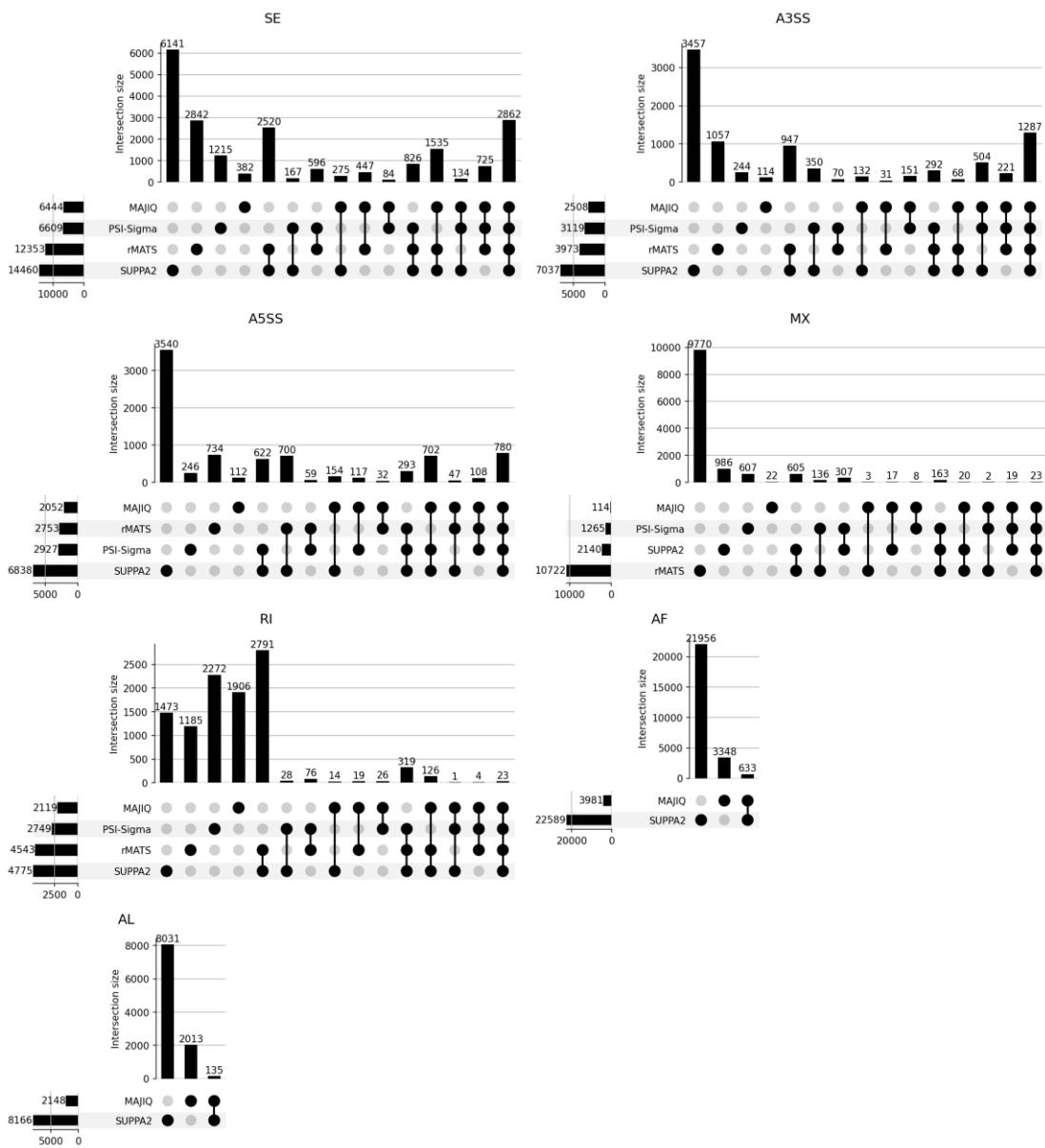
577 The coordinates of seven AS events defined in the unified format output are shown as
578 red lines in the corresponding splice graph. Created in BioRender. Zhang, R. (2025)
579 <https://BioRender.com/uvu00yo>.

580 **Figure S45. The intersections of annotated and novel AS events by four AS event**
 581 **detection tools.**



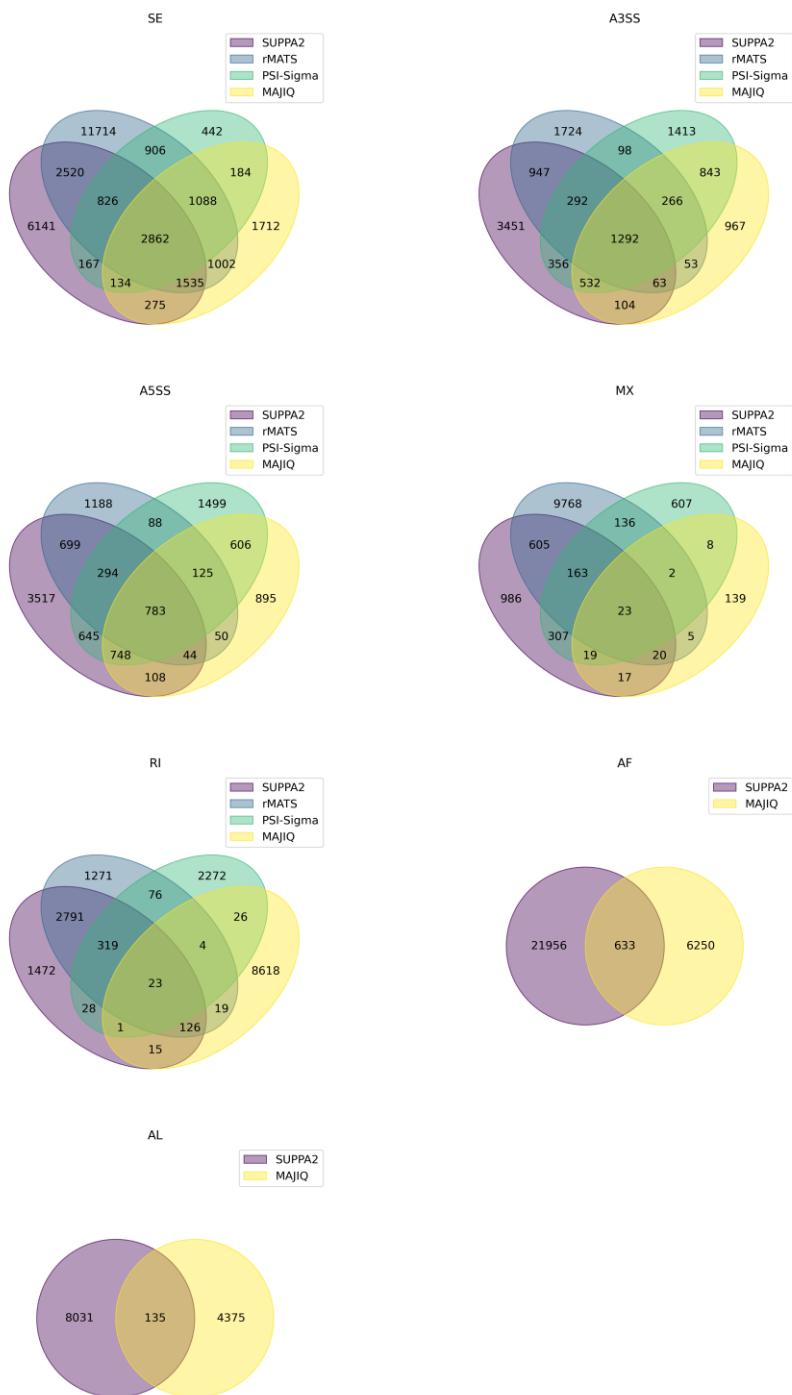
582
 583 The standardized benchmarking tool generates the Upset plots indicating intersections
 584 of alternative splicing (AS) event across MAJIQ, PSI-Sigma, rMATS, and SUPPA2,
 585 including both annotated and novel events. AS events with percent spliced-in (PSI)
 586 values between 0.05 and 0.95 were included. Data from the M8 sample of lab1 are
 587 shown as an example.

588 **Figure S46. The intersections of annotated AS events by four AS event detection
589 tools.**



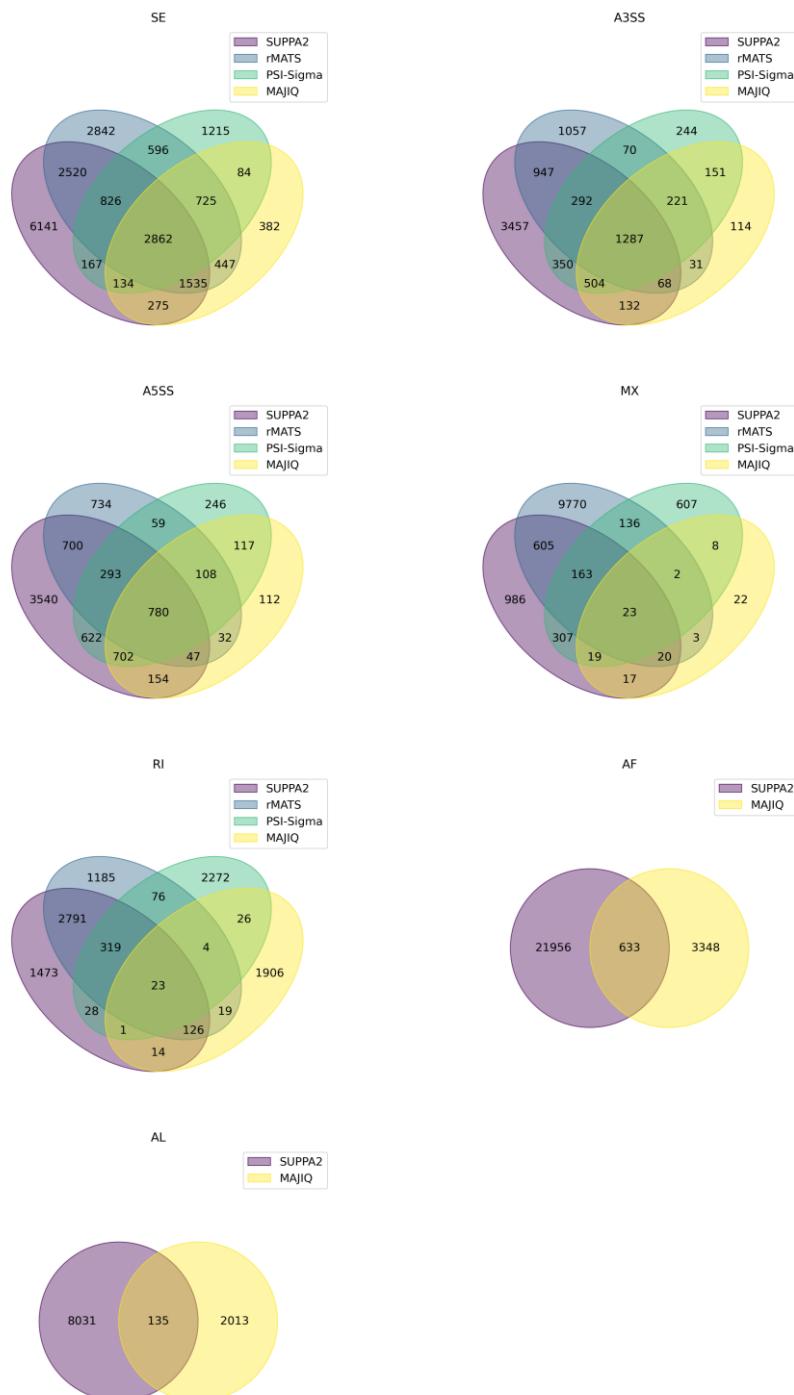
590
591 The standardized benchmarking tool generates the Upset plots indicating intersections
592 of annotated alternative splicing (AS) event across MAJIQ, PSI-Sigma, rMATS, and
593 SUPPA2. AS events with percent spliced-in (PSI) values between 0.05 and 0.95 were
594 included. Data from the M8 sample of lab1 are shown as an example.

595 **Figure S47. The intersections of annotated and novel AS events by four AS event**
 596 **detection tools.**



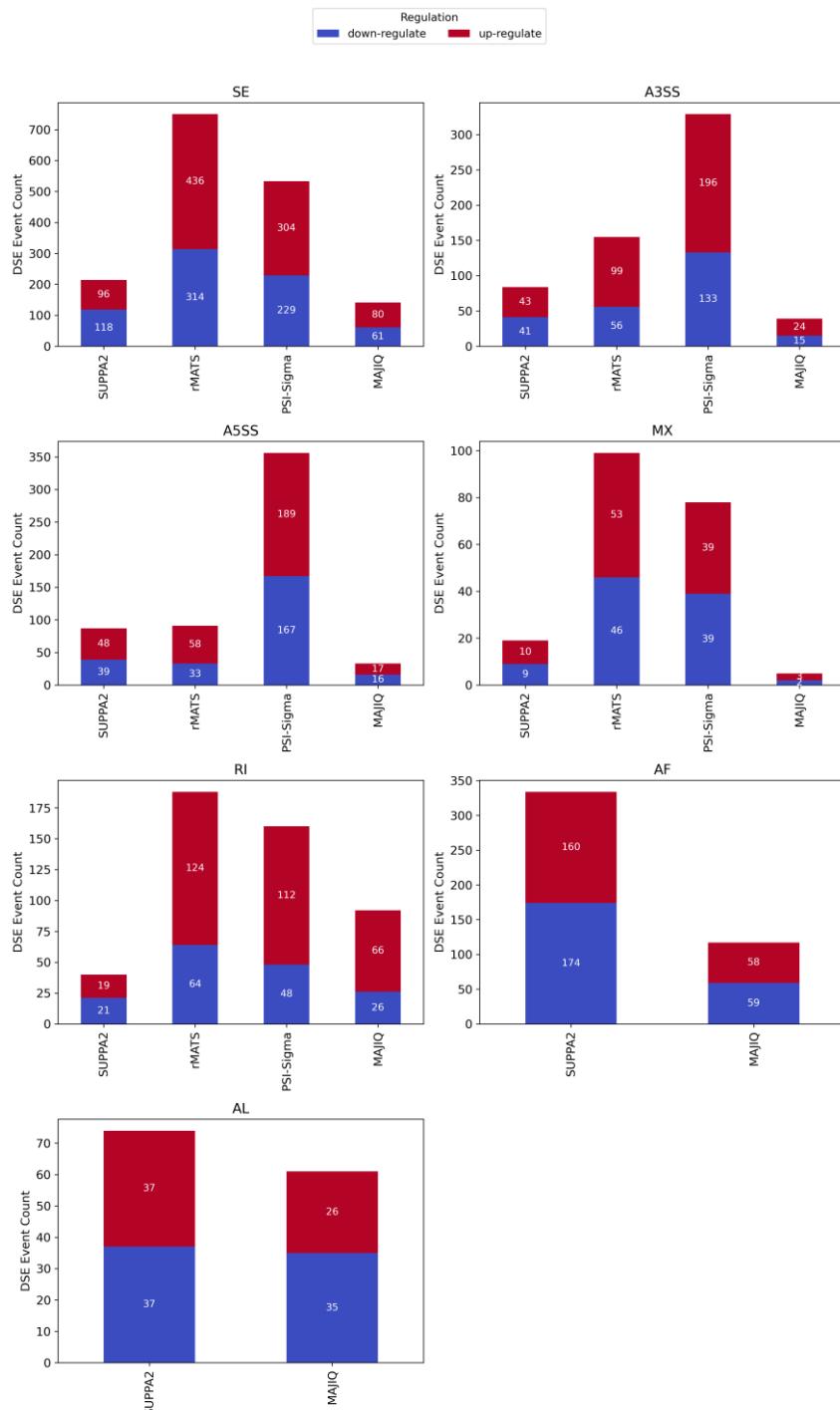
597
 598 The standardized benchmarking tool generates the Venn plots indicating intersections
 599 of alternative splicing (AS) event across MAJIQ, PSI-Sigma, rMATS, and SUPPA2,
 600 including both annotated and novel events. AS events with percent spliced-in (PSI)
 601 values between 0.05 and 0.95 were included. Data from the M8 sample of lab1 are
 602 shown as an example.

603 **Figure S48. The intersections of annotated AS events by four AS event detection
604 tools.**



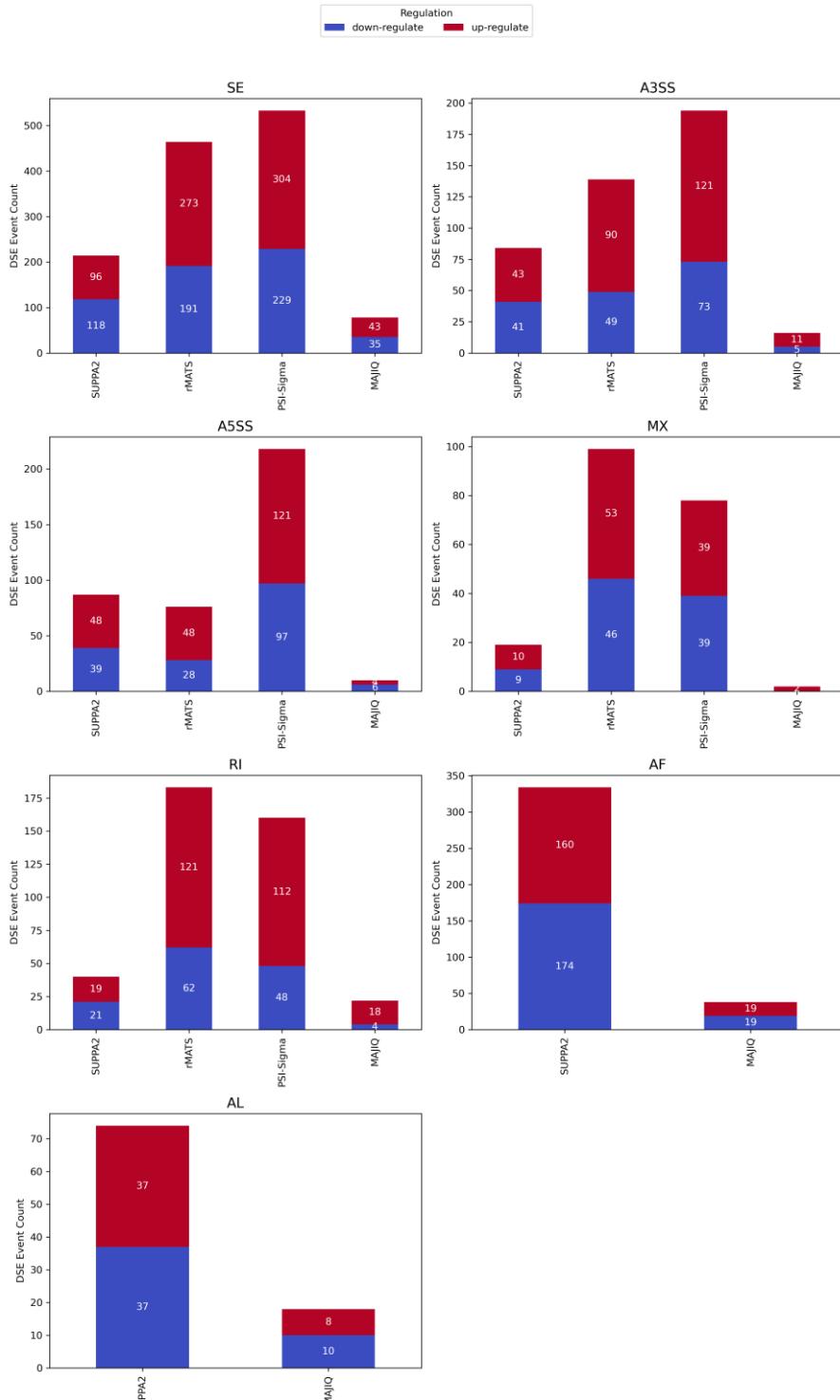
605
606 The standardized benchmarking tool generates the Venn plots indicating intersections
607 of annotated alternative splicing (AS) event across MAJIQ, PSI-Sigma, rMATS, and
608 SUPPA2. AS events with percent spliced-in (PSI) values between 0.05 and 0.95 were
609 included. Data from the M8 sample of lab1 are shown as an example.

610 **Figure S49. The number of annotated and novel DSEs by four AS event detection**
611 **tools.**



612
613 The standardized benchmarking tool generates the histogram indicating number of
614 differential splicing events (DSEs) detected by MAJIQ, PSI-Sigma, rMATS, and
615 SUPPA2, including both annotated and novel alternative splicing (AS) events. Data
616 from the M8/D6 sample pair of lab1 are shown as an example.

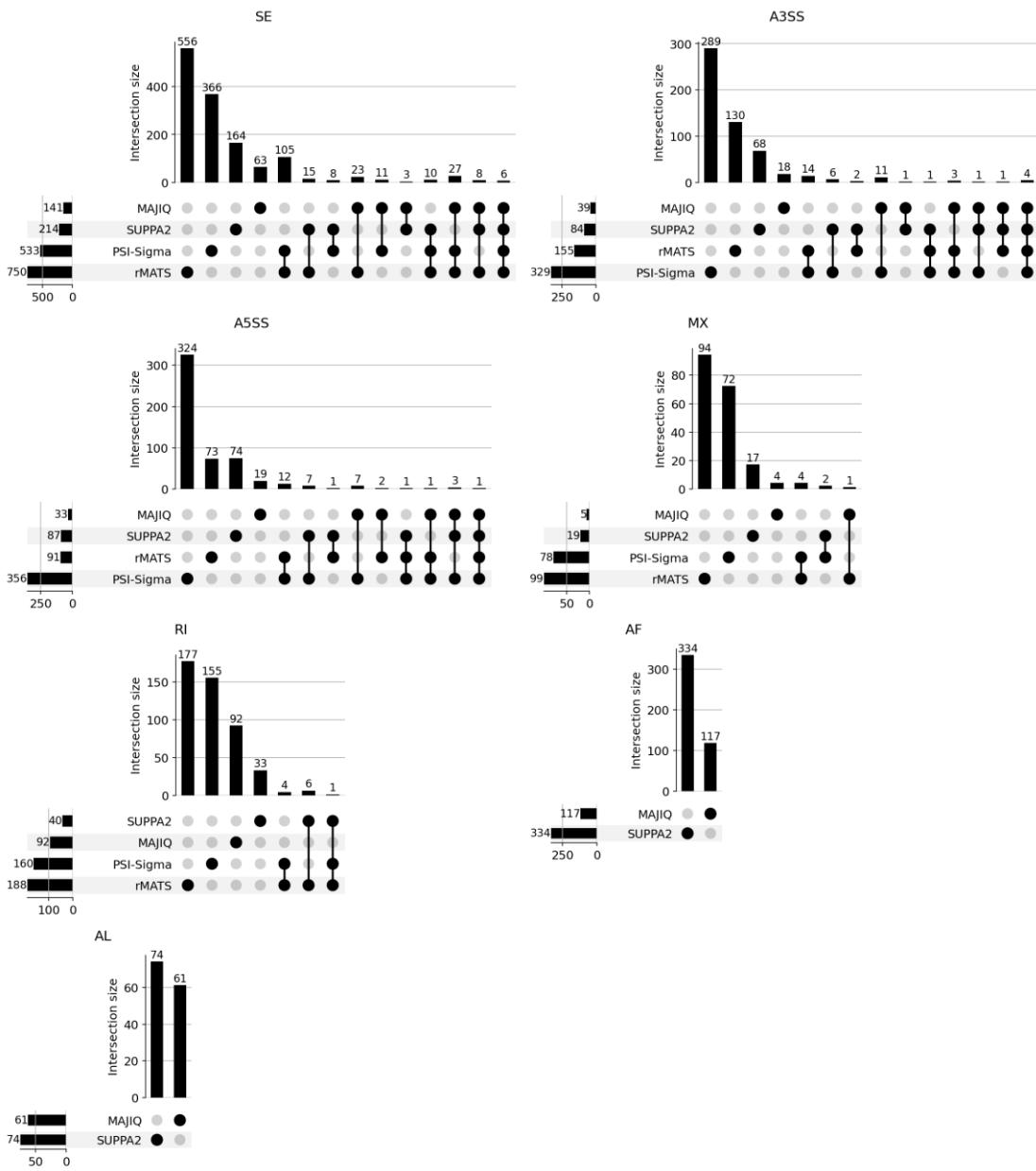
617 **Figure S50. The number of annotated DSEs by four AS event detection tools.**



618

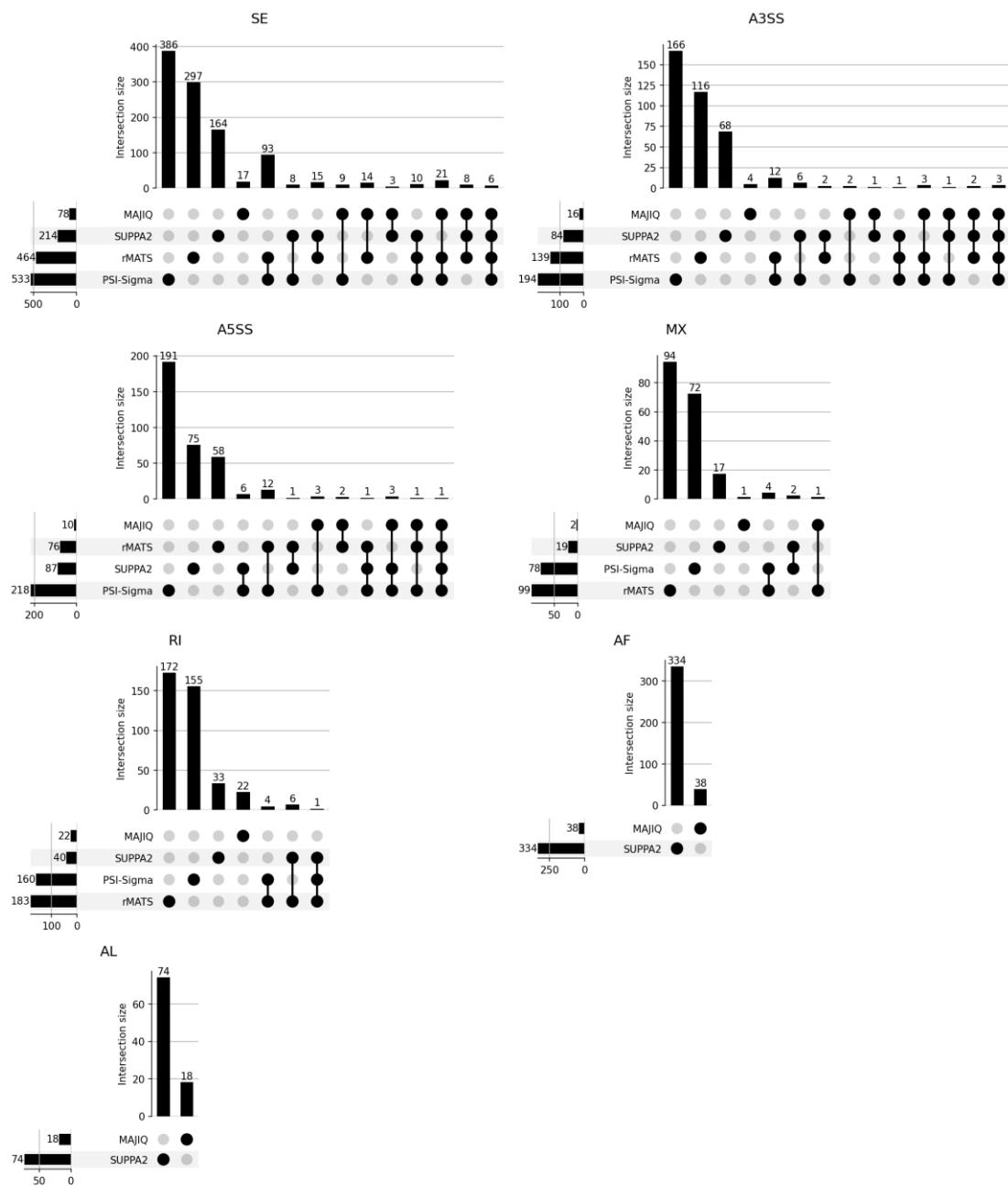
619 The standardized benchmarking tool generates the histogram indicating number of
 620 annotated differential splicing events (DSEs) detected by MAJIQ, PSI-Sigma, rMATS,
 621 and SUPPA2. Data from the M8/D6 sample pair of lab1 are shown as an example.

622 **Figure S51. The intersections of annotated and novel DSEs by four AS event**
623 **detection tools.**



624
625 The standardized benchmarking tool generates the Upset plots indicating intersections
626 of differential splicing events (DSEs) across MAJIQ, PSI-Sigma, rMATS, and SUPPA2,
627 including both annotated and novel alternative splicing (AS) events. Data from the
628 M8/D6 sample pair of lab1 are shown as an example.

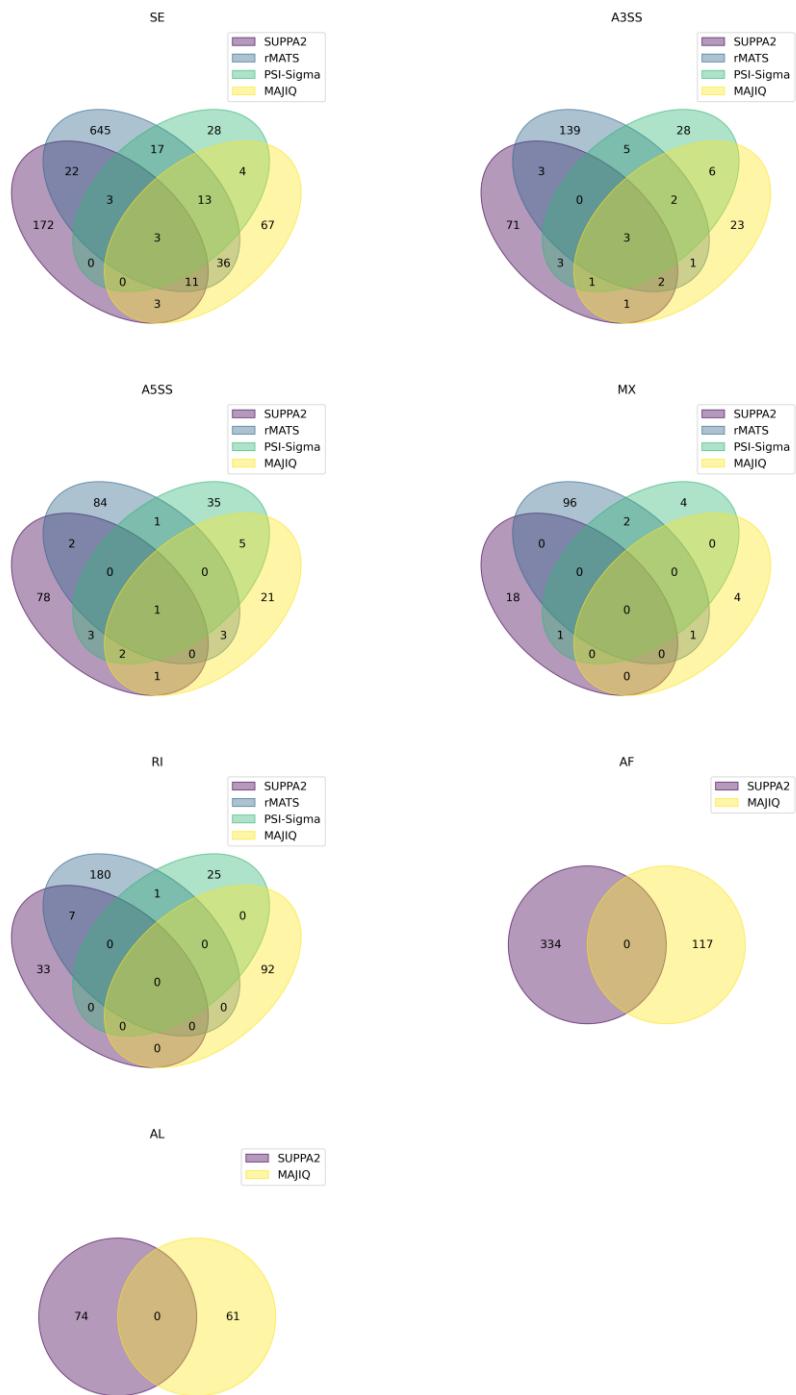
629 **Figure S52. The intersections of annotated DSEs by four AS event detection tools.**



630

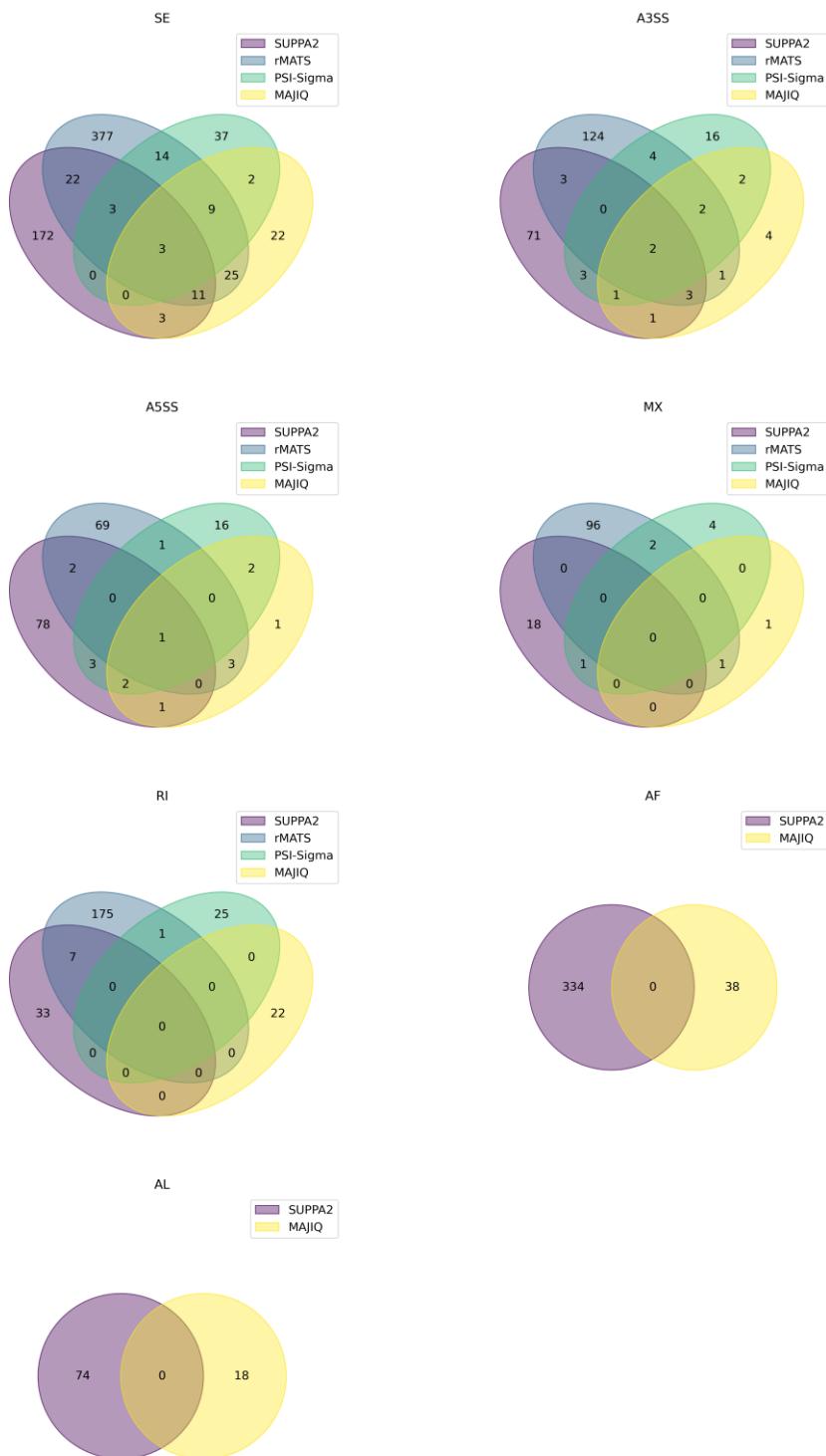
631 The standardized benchmarking tool generates the Upset plots indicating intersections
 632 of annotated differential splicing events (DSEs) across MAJIQ, PSI-Sigma, rMATS,
 633 and SUPPA2. Data from the M8/D6 sample pair of lab1 are shown as an example.

634 **Figure S53. The intersections of annotated and novel DSEs by four AS event**
635 **detection tools.**



636
637 The standardized benchmarking tool generates the Venn plots indicating intersections
638 of differential splicing events (DSEs) across MAJIQ, PSI-Sigma, rMATS, and SUPPA2,
639 including both annotated and novel alternative splicing (AS) events. Data from the
640 M8/D6 sample pair of lab1 are shown as an example.

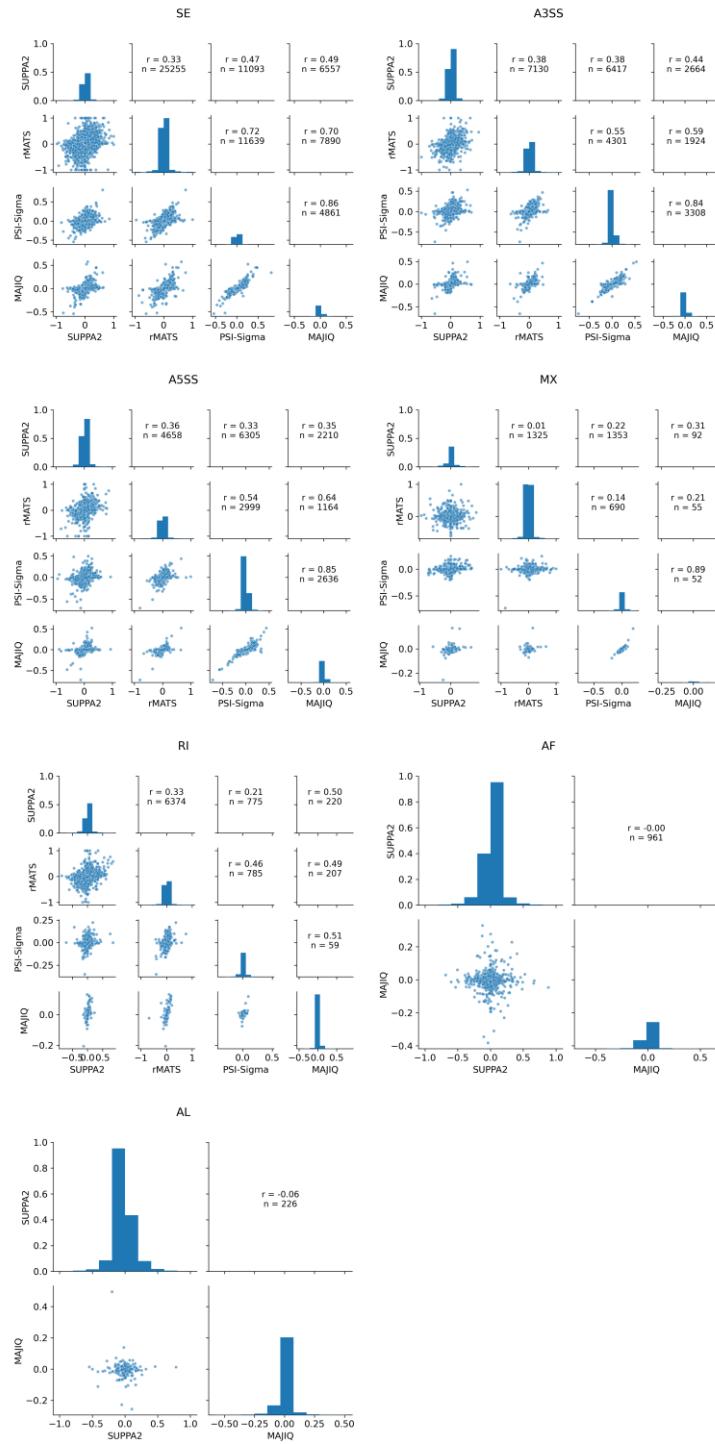
641 **Figure S54. The intersections of annotated DSEs by four AS event detection tools.**



642

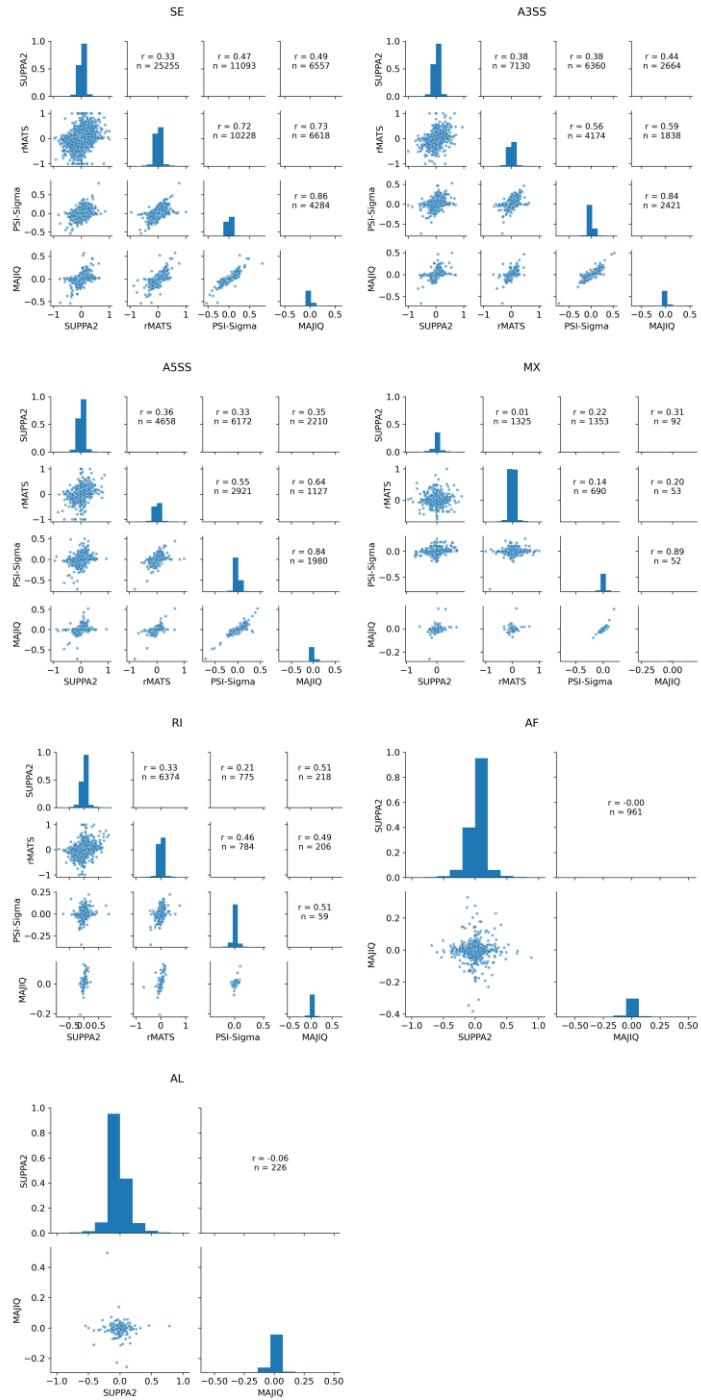
643 The standardized benchmarking tool generates the Venn plots indicating intersections
 644 of annotated differential splicing events (DSEs) across MAJIQ, PSI-Sigma, rMATS,
 645 and SUPPA2. Data from the M8/D6 sample pair of lab1 are shown as an example.

646 **Figure S55. The dPSI consistency of annotated and novel AS events by four AS**
647 **event detection tools.**



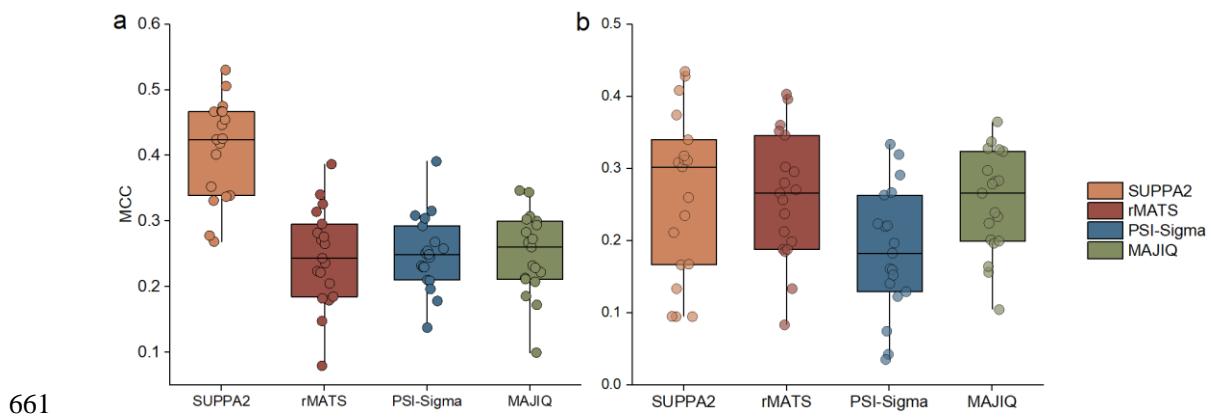
648
649 The standardized benchmarking tool generates the Scatter plots indicating the
650 consistency of delta percent spliced-in (dPSI) results across MAJIQ, PSI-Sigma,
651 rMATS, and SUPPA2, including both annotated and novel alternative splicing (AS)
652 events. Data from the M8/D6 sample pair of lab1 are shown as an example.

653 **Figure S56. The dPSI consistency of annotated AS events by four AS event**
 654 **detection tools.**



655
 656 The standardized benchmarking tool generates the Scatter plots indicating the
 657 consistency of delta percent spliced-in (dPSI) results for annotated alternative splicing
 658 (AS) events across MAJIQ, PSI-Sigma, rMATS, and SUPPA2. Data from the M8/D6
 659 sample pair of lab1 are shown as an example.

660 **Figure S57. The MCC of four AS event detection tools for DSE detection.**



661
662 The Matthews correlation coefficient (MCC) of 19 high-quality laboratories in
663 differential splicing event (DSE) detection against the (a) Quartet and (b) RT-qPCR
664 reference datasets by four AS event detection tools. Data are presented as median values
665 (center lines) with upper and lower quartiles (box boundaries).

666 **Reference**

667 1. Chen, Q. et al. A ratio-based framework using Quartet reference materials for integrating long-
668 and short-read RNA-seq. *bioRxiv*, 2025.2009.2015.676287 (2025).

669 2. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100
670 (2018).

671 3. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

672 4. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*
673 **28**, 2184-2185 (2012).

674 5. Pockrandt, C., Alzamel, M., Iliopoulos, C.S. & Reinert, K. GenMap: ultra-fast computation of
675 genome mappability. *Bioinformatics* **36**, 3687-3692 (2020).

676 6. Pardo-Palacios, F.J. et al. Systematic assessment of long-read RNA-seq methods for transcript
677 identification and quantification. *Nat Methods* **21**, 1349-1363 (2024).

678 7. Trincado, J.L. et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis
679 across multiple conditions. *Genome Biol* **19**, 40 (2018).

680 8. Wang, Y. et al. rMATS-turbo: an efficient and flexible computational tool for alternative splicing
681 analysis of large-scale RNA-seq data. *Nat Protoc* **19**, 1083-1104 (2024).

682 9. Vaquero-Garcia, J. et al. RNA splicing analysis using heterogeneous and large RNA-seq datasets.
683 *Nat Commun* **14**, 1230 (2023).

684 10. Lin, K.T. & Krainer, A.R. PSI-Sigma: a comprehensive splicing-detection method for short-
685 read and long-read RNA-seq analysis. *Bioinformatics* **35**, 5048-5054 (2019).

686