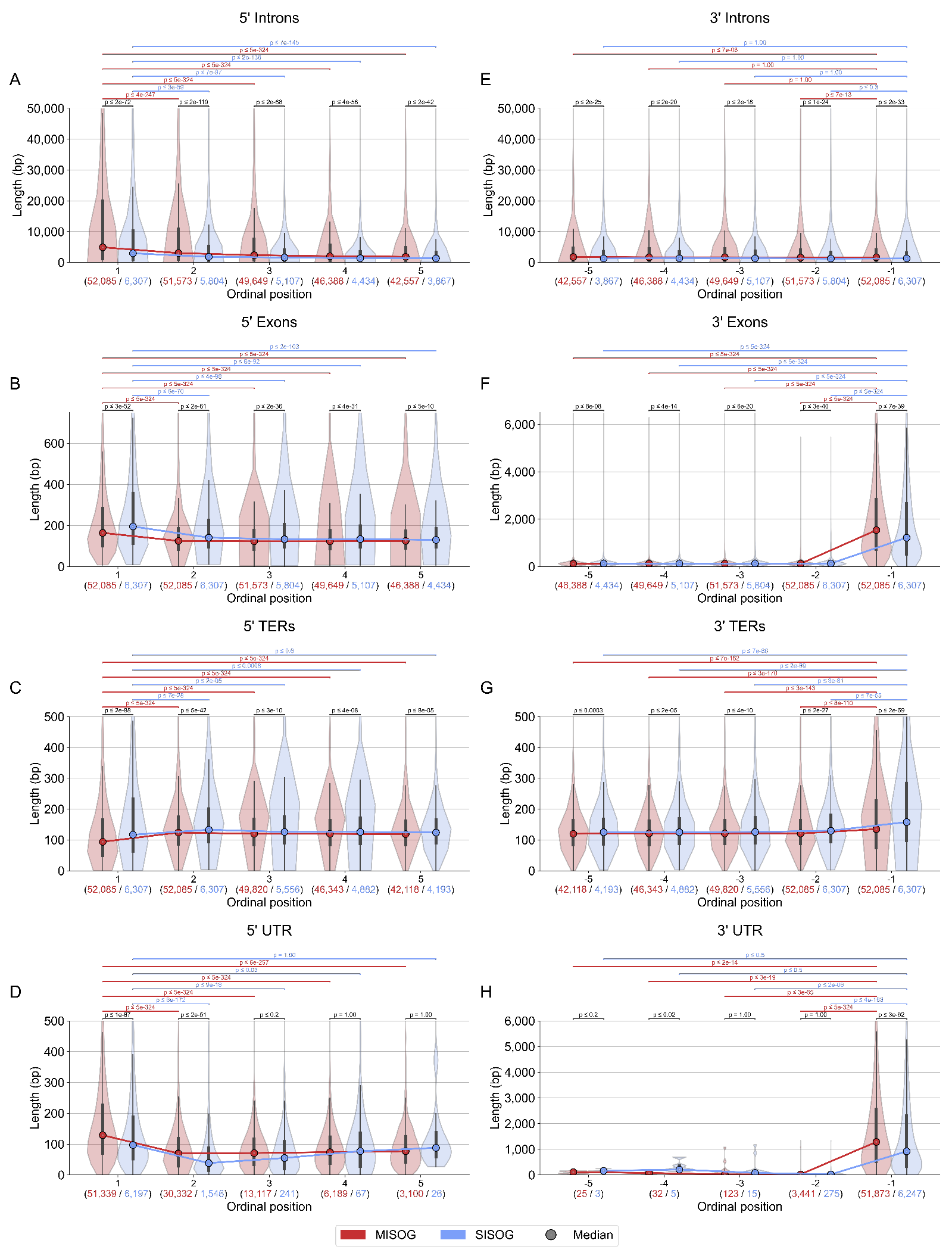
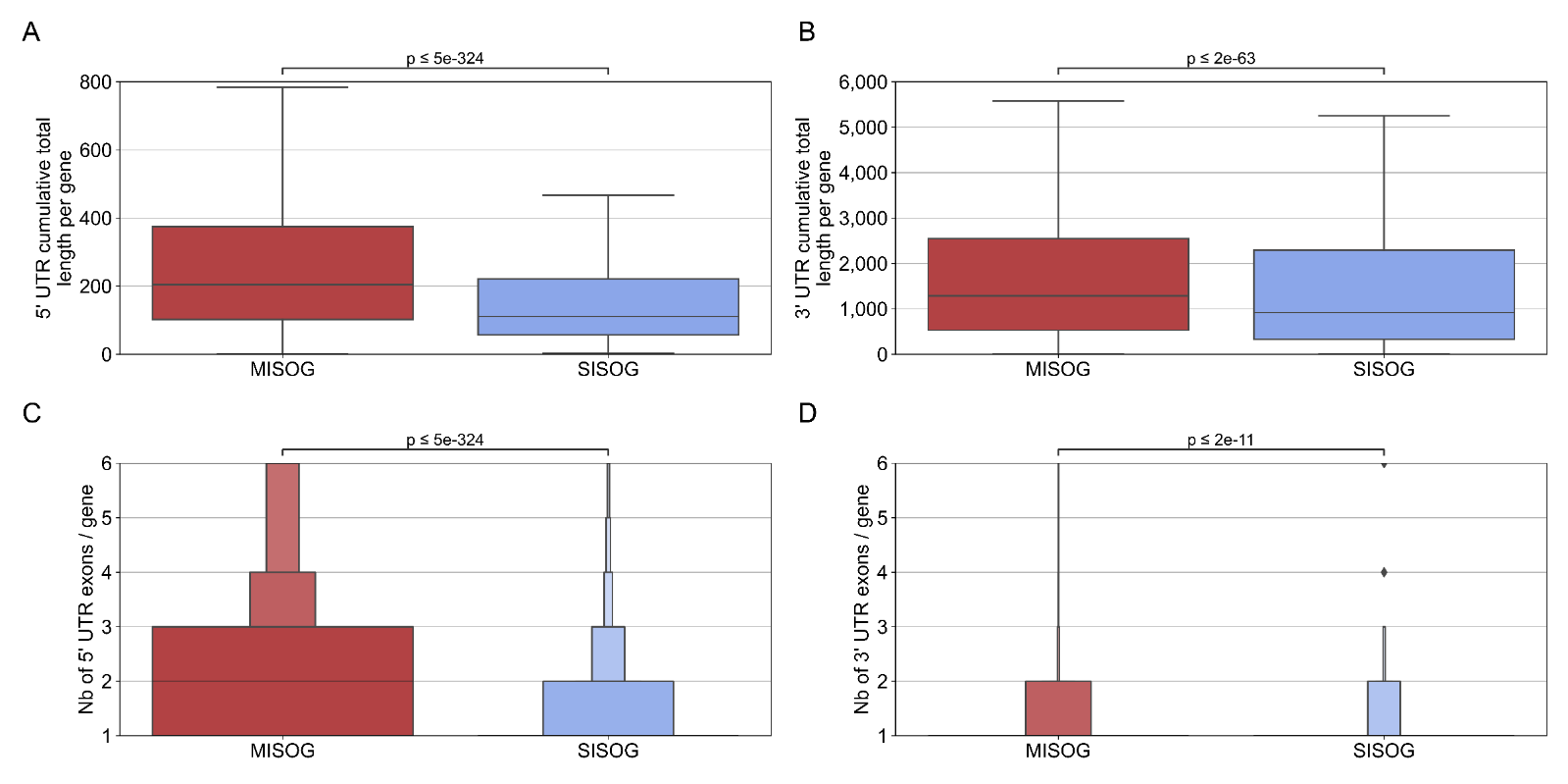
**All supplementary Figures are available in original quality at** [**https://github.com/weber8thomas/MISOG\_SISOG**](https://github.com/weber8thomas/MISOG_SISOG)**.**

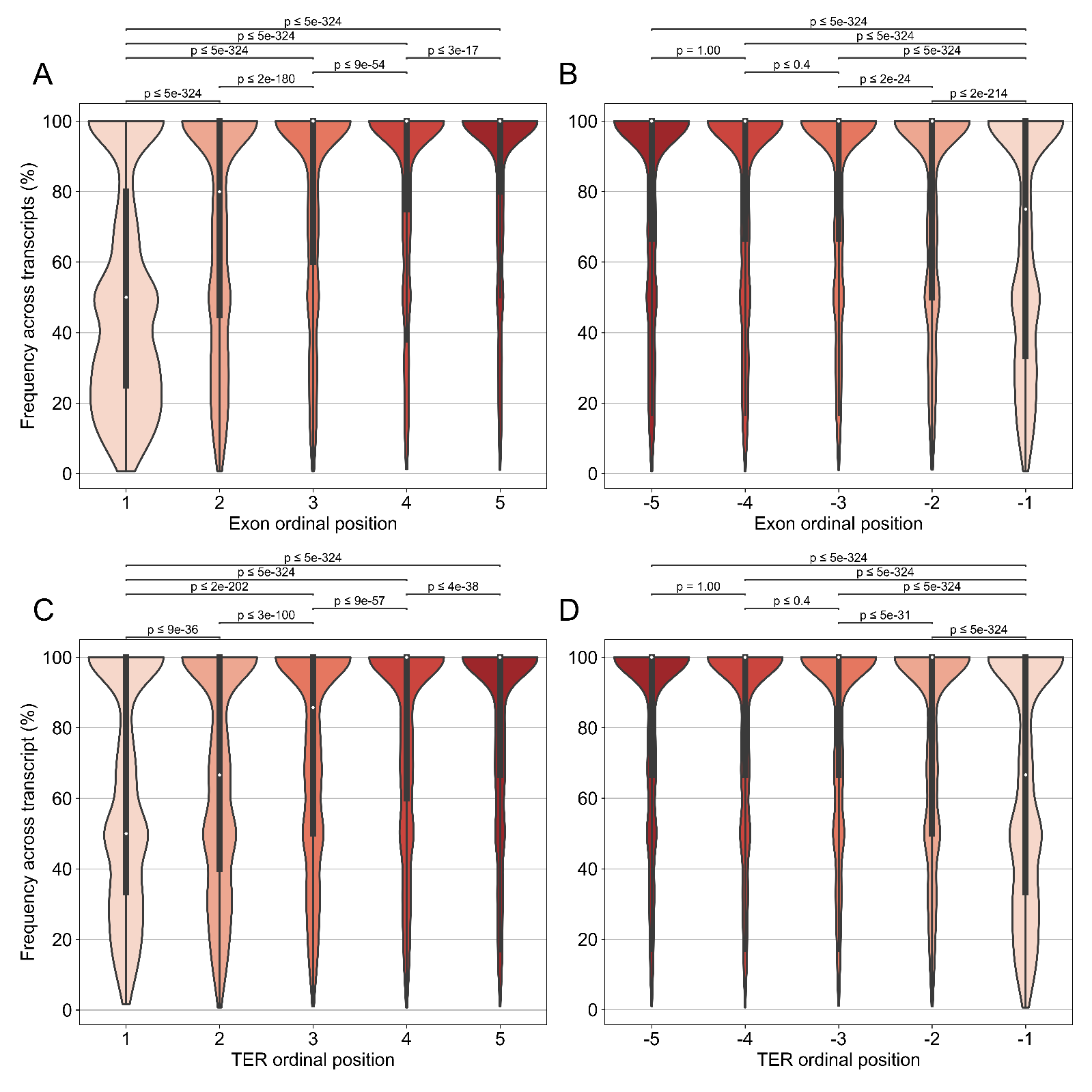


**Figure S1 – Length distribution of gene elements (Intron, Exon, TER and UTR exons) according to ordinal positions.**

All plots represent the length distribution of gene elements according to ordinal position (as illustrated in Fig. S4F). Plots A-D correspond to the first five 5’ elements in the genes (1 up to 5), plots E-H correspond to the last five 3’ elements (-5 to -1). The p-values obtained through Mann-Whitney U tests are displayed above the violin plots. Black p-values correspond to comparisons between MISOG and SISOG while colored p-values represent comparisons between first elements and following ones (MISOG: red, SISOG: blue). Values under each plot represent the numbers of observed elements at the ordinal position for MISOG (red) and SISOG (blue).

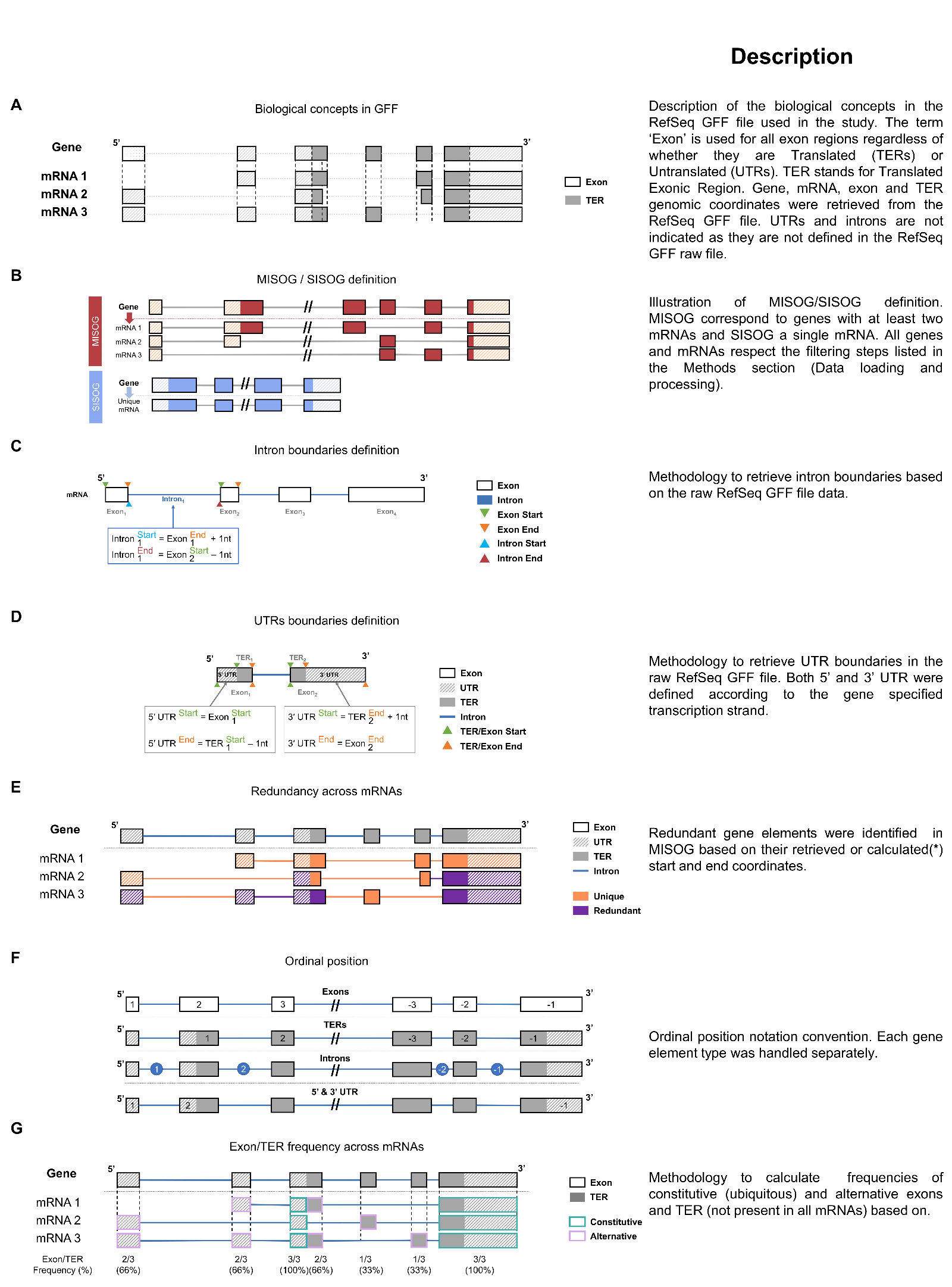
**Figure S2 – Cumulative length and number of 5’ and 3’ UTR exons by gene**

Cumulative total lengths of 5’ (A) and 3’ (B) UTR regions per gene. Number of 5’ and 3’ UTR exons per gene are displayed in (C, D). The p-values obtained through Mann-Whitney U tests are displayed above the boxplots (A, B) and boxenplots (C, D).



**Figure S3 – Percentage of alternative exons and TERs according to their ordinal positions.**

All plots show calculated frequencies of alternative exons (A, B) and alternative TERs (C, D) as illustrated in Fig. S4G. Plots (A, C) correspond to the first five 5’ elements in the genes (1 up to 5), while plots (B,D) correspond to the last five 3’ elements (-5 to -1). The p-values obtained through Mann-Whitney U tests are displayed above the violin plots.



**Figure S4 – Material and Methods schemas**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **MISOG** | **SISOG** | **Ratio MISOG/SISOG** |
| **Number of genes** | **10,995** | **6,307** | **1.74** |
| **Number of mRNAs** | **52,983** | **6,307** | 8.40 |

**Table S1 – Number of genes and mRNAs in MISOG and SISOG**

Numbers correspond to total number of HCPG MISOG and SISOG after excluding 1,983 single-TER genes.

Numbers in bold correspond to cited values in the main text.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Median** | **Lower quartile** | **Upper quartile** | **Mean** | **Std** |
| **Number of mRNAs per MISOG** | **3** | 2 | 5 | 5 | 6 |

**Table S2 – Number of mRNAs by gene in MISOG**

Numbers in bold correspond to cited values in the main text.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **MISOG** | **SISOG** | **Ratio MISOG/SISOG** |
| **Nb of Exons** | **165,585** | **65,079** | 2.5 |
| **Nb of TERs** | **150,276** | **62,728** | 2.4 |
| **Nb of Introns** | **164,826** | **58,738** | 2.8 |

**Table S3 – Count of Exons, TERs and Introns for MISOG and SISOG**

Presented values correspond to elements with unique coordinates (Start and End) for a gene (illustrated in Fig. S4E).

Numbers in bold correspond to cited values in the main text.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Median (bp)** | **Lower quartile (bp)** | **Upper quartile (bp)** | **Mean (bp)** | **Std (bp)** | **Coverage (bp)** | **p-value (MWU)** |
| **Gene length** | **MISOG** | **36,988** | 15,527 | 90,370 | 84,864 | 149,983 | **933,084,569** | **4.425e-164** |
| **SISOG** | **20,617** | 8,294 | 51,898 | 51,048 | 105,313 | **321,962,756** |
| **Ratio MISOG/SISOG** | **1.79** | 1.87 | 1.74 | 1.66 | 1.42 | **2.90** | / |
| **mRNA length** | **MISOG** | **3,341** | 2,143 | 5,058 | 4,000 | 2,917 | 211,946,987 | **1.231e-90** |
| **SISOG** | **2,688** | 1,587 | 4,627 | 3,533 | 2,895 | 22,283,817 |
| **Ratio MISOG/SISOG** | **1.24** | 1.35 | 1.09 | 1.13 | 1.01 | 9.51 | / |

**Table S4 – Length of genes and mRNA for MISOG and SISOG**

Gene lengths were computed based on the Start and End coordinates (respectively corresponding to TSS and TTS) present in the raw RefSeq GFF file. The mRNA lengths were computed as the sum of all exon lengths associated with a specific mRNA. Numbers in bold correspond to cited values in the main text.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Number** | | | | | | | | |
|  | **Exons** | | | **TERs** | | | **Introns** | | |
| **MISOG** | **SISOG** | **Ratio MISOG/SISOG** | **MISOG** | **SISOG** | **Ratio MISOG/SISOG** | **MISOG** | **SISOG** | **Ratio MISOG/SISOG** |
| **Median** | **12** | **7** | 1.71 | **11** | **7** | 1.57 | 12 | 6 | 2.00 |
| **Lower quartile** | 8 | 4 | 2.00 | 7 | 4 | 1.75 | 7 | 3 | 2.33 |
| **Upper quartile** | 19 | 12 | 1.58 | 17 | 12 | 1.42 | 19 | 11 | 1.73 |
| **Mean** | 15 | 10 | 1.46 | 14 | 10 | 1.37 | 15 | 9 | 1.61 |
| **Std** | 11 | 11 | 1.01 | 11 | 11 | 1.00 | 12 | 11 | 1.08 |
| **p-value (MWU)** | 0.000e+00 | | / | 1.830e-307 | | / | 0.000e+00 | | / |

**Table S5 – Number of exons, TERs and introns**

Presented values correspond to elements with unique coordinates (Start and End) for a gene (illustrated in Fig. S4E).

Numbers in bold correspond to cited values in the main text.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Length** | | | | | | | | |
|  | **Exons** | | | **TERs** | | | **Introns** | | |
| **MISOG** | **SISOG** | **Ratio MISOG/SISOG** | **MISOG** | **SISOG** | **Ratio MISOG/SISOG** | **MISOG** | **SISOG** | **Ratio MISOG/SISOG** |
| **Median (bp)** | **134** | **136** | 0.99 | **119** | **126** | 0.94 | **1,963** | **1,336** | 1.47 |
| **Lower quartile (bp)** | 93 | 96 | 0.97 | 81 | 88 | 0.92 | 638 | 447 | 1.43 |
| **Upper quartile (bp)** | 203 | 200 | 1.02 | 166 | 173 | 0.96 | 5,838 | 3,511 | 1.66 |
| **Mean (bp)** | 345 | 342 | 1.01 | 160 | 173 | 0.93 | 8,495 | 4,789 | 1.77 |
| **Std (bp)** | 888 | 880 | 1.01 | 254 | 316 | 0.80 | 26,797 | 16,488 | 1.63 |
| **p-value (MWU)** | 4.004e-08 | | / | 6.935e-140 | | / | 0.000e+00 | | / |

**Table S6 –Length of exons, TERs, and introns**

Presented values correspond to elements with unique coordinates (Start and End) for a gene (illustrated in Fig. S4E). Length was computed using Start and End coordinates. Numbers in bold correspond to cited values in the main text.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **5’ part** | | | | | | |
|  |  | **Ordinal position** | **Median (bp)** | **Lower quartile (bp)** | **Upper quartile (bp)** | **Mean (bp)** | **Std (bp)** | **Count** |
| **Introns** | **MISOG** | **1** | **4,901** | 1,181 | **19,992** | 19,865 | 45,012 | 52,085 |
| **2** | **3,032** | 966 | 10,760 | 14,986 | 40,643 | 51,573 |
| **3** | 2,347 | 739 | 7,477 | 10,266 | 28,653 | 49,649 |
| **4** | 2,012 | 666 | 5,627 | 7,495 | 21,758 | 46,388 |
| **5** | **1,847** | 626 | 4,765 | 6,174 | 18,414 | 42,557 |
| **SISOG** | **1** | **2,953** | 920 | **10,314** | 12,316 | 32,666 | 6,307 |
| **2** | **1,823** | 571 | 5,193 | 7,457 | 23,747 | 5,804 |
| **3** | 1,589 | 544 | 4,125 | 5,490 | 16,434 | 5,107 |
| **4** | 1,356 | 476 | 3,490 | 4,241 | 12,611 | 4,434 |
| **5** | **1,341** | 408 | 3,225 | 4,051 | 14,206 | 3,867 |
| **Exons** | **MISOG** | **1** | **164** | 100 | 283 | 237 | 251 | 52,085 |
| **2** | **125** | 85 | 184 | 194 | 332 | 52,085 |
| **3** | 123 | 85 | 177 | 224 | 530 | 51,573 |
| **4** | 123 | 87 | 175 | 271 | 730 | 49,649 |
| **5** | **125** | 90 | 174 | 302 | 849 | 46,388 |
| **SISOG** | **1** | **195** | 112 | 357 | 299 | 355 | 6,307 |
| **2** | **141** | 95 | 225 | 333 | 758 | 6,307 |
| **3** | 133 | 95 | 205 | 354 | 927 | 5,804 |
| **4** | 134 | 95 | 198 | 407 | 1,110 | 5,107 |
| **5** | **130** | 95 | 186 | 377 | 936 | 4,434 |
| **TERs** | **MISOG** | **1** | **94** | 49 | 165 | 161 | 274 | 52,085 |
| **2** | **124** | 86 | 174 | 172 | 234 | 52,085 |
| **3** | 120 | 85 | 167 | 165 | 272 | 49,820 |
| **4** | 120 | 85 | 164 | 165 | 250 | 46,343 |
| **5** | **119** | 85 | 162 | 151 | 199 | 42,118 |
| **SISOG** | **1** | **117** | 64 | 233 | 217 | 368 | 6,307 |
| **2** | **133** | 94 | 201 | 223 | 437 | 6,307 |
| **3** | 126 | 90 | 175 | 182 | 419 | 5,556 |
| **4** | 126 | 89 | 171 | 196 | 456 | 4,882 |
| **5** | **124** | 90 | 165 | 157 | 230 | 4,193 |
| **5' UTR exons** | **MISOG** | **1** | **129** | 70 | 227 | 182 | 185 | 51,339 |
| **2** | **70** | 29 | 119 | 94 | 120 | 30,332 |
| **3** | 71 | 33 | 116 | 92 | 123 | 13,117 |
| **4** | 74 | 38 | 123 | 95 | 120 | 6,189 |
| **5** | **77** | 42 | 125 | 93 | 82 | 3,100 |
| **SISOG** | **1** | **97** | 52 | 188 | 151 | 171 | **6,197** |
| **2** | **39** | 15 | 88 | 71 | 109 | **1,546** |
| **3** | 55 | 20 | 108 | 88 | 132 | **241** |
| **4** | 77 | 28 | 135 | 127 | 266 | **67** |
| **5** | **88** | 59 | 137 | 136 | 154 | **26** |

**Table S7 – Length of first 5 exons, TERs, introns, and 5' UTR exons**

Ordinal positions (from the 5' mRNA end) were separately identified for each biological concept (introns, exons, TER, UTR). Numbers in bold correspond to cited values in the main text.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **3’ part** | | | | | | |
|  |  | **Ordinal position** | **Median (bp)** | **Lower quartile(bp)** | **Upper quartile (bp)** | **Mean (bp)** | **Std (bp)** | **Count** |
| **Introns** | **MISOG** | **-5** | 1,768 | 596 | 4,703 | 6,774 | 20,159 | 42,118 |
| **-4** | 1,665 | 566 | 4,465 | 6,099 | 20,288 | 46,343 |
| **-3** | 1,669 | 553 | 4,415 | 5,890 | 19,154 | 49,820 |
| **-2** | 1,567 | 500 | 4,080 | 5,319 | 16,048 | 52,085 |
| **-1** | 1,636 | 605 | 4,163 | 5,089 | 16,115 | 52,085 |
| **SISOG** | **-5** | 1,285 | 436 | 3,579 | 4,695 | 15,763 | 4,193 |
| **-4** | 1,329 | 424 | 3,424 | 4,746 | 14,583 | 4,882 |
| **-3** | 1,329 | 450 | 3,614 | 4,478 | 13,653 | 5,556 |
| **-2** | 1,239 | 389 | 3,266 | 4,094 | 14,445 | 6,307 |
| **-1** | 1,296 | 493 | 3,123 | 4,198 | 19,203 | 6,307 |
| **Exons** | **MISOG** | **-5** | 123 | 88 | 168 | 153 | 176 | 46,388 |
| **-4** | 124 | 88 | 167 | 154 | 179 | 49,649 |
| **-3** | 123 | 91 | 166 | 158 | 212 | 51,573 |
| **-2** | 122 | 89 | 165 | 156 | 201 | 52,085 |
| **-1** | 1,537 | 721 | 2,840 | 2,189 | 2,284 | 52,085 |
| **SISOG** | **-5** | 130 | 92 | 174 | 166 | 244 | 4,434 |
| **-4** | 129 | 94 | 176 | 175 | 262 | 5,107 |
| **-3** | 130 | 95 | 180 | 180 | 248 | 5,804 |
| **-2** | 132 | 96 | 183 | 198 | 288 | 6,307 |
| **-1** | 1,221 | 529 | 2,654 | 1,935 | 2,103 | 6,307 |
| **TERs** | **MISOG** | **-5** | 120 | 85 | 163 | 143 | 161 | 25 |
| **-4** | 121 | 85 | 162 | 145 | 172 | 32 |
| **-3** | 121 | 88 | 163 | 149 | 189 | 123 |
| **-2** | 121 | 89 | 164 | 153 | 198 | 3,441 |
| **-1** | 135 | 75 | 227 | 247 | 439 | 51,873 |
| **SISOG** | **-5** | 125 | 87 | 167 | 154 | 241 | 3 |
| **-4** | 125 | 89 | 169 | 160 | 249 | 5 |
| **-3** | 126 | 90 | 173 | 160 | 217 | 15 |
| **-2** | 130 | 94 | 180 | 184 | 257 | 275 |
| **-1** | 158 | 98 | 284 | 303 | 551 | 6,247 |
| **3' UTR exons** | **MISOG** | **-5** | 111 | 62 | 145 | 100 | 54 | 42,557 |
| **-4** | 77 | 55 | 96 | 82 | 50 | 46,388 |
| **-3** | 21 | 12 | 107 | 130 | 249 | 49,649 |
| **-2** | 22 | 10 | 34 | 39 | 121 | 51,573 |
| **-1** | **1,285** | 527 | 2,548 | 1,950 | 2,225 | 52,085 |
| **SISOG** | **-5** | 161 | 161 | 208 | 192 | 54 | 3,867 |
| **-4** | 210 | 210 | 362 | 324 | 225 | 4,434 |
| **-3** | 79 | 48 | 106 | 195 | 340 | 5,107 |
| **-2** | 21 | 9 | 36 | 33 | 86 | 5,804 |
| **-1** | **918** | 324 | 2,297 | 1,649 | 2,011 | 6,307 |

**Table S8 – Length of last 5 exons, TERs, introns, and 3' UTR exons**

Ordinal positions (from the 3' mRNA end) were separately identified for each biological concept (introns, exons, TER, UTR). Numbers in bold correspond to cited values in the main text.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Median (bp)** | **Lower quartile (bp)** | **Upper quartile (bp)** | **Mean (bp)** | **Std (bp)** |
| **Distance TSS ↔ START** | **MISOG** | **1,577** | 154 | 15,054 | 19,125 | 54,100 |
| **SISOG** | **143** | 59 | 639 | 3,575 | 15,476 |
| **Ratio MISOG / SISOG** | **11.0** | 2.6 | 23.6 | 5.3 | 3.5 |
| **Distance STOP ↔ TTS** | **MISOG** | 1,388 | 565 | 2,794 | 2,248 | 4,002 |
| **SISOG** | 979 | 339 | 2,394 | 1,825 | 3,350 |
| **Ratio MISOG / SISOG** | 1.4 | 1.7 | 1.2 | 1.2 | 1.2 |

**Table S9 – Statistics for distance separating first/last transcript nucleotide from first/last coding nucleotide in 5'/3'**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Ordinal position** | **Median (%)** | **Lower quartile (%)** | **Upper quartile (%)** | **Mean**  **(%)** | **Std**  **(%)** |
| **5'** | **TER** | **1** | **50.00** | 33.33 | 100.00 | 61.94 | 32.29 |
| **2** | **66.67** | 40.00 | 100.00 | 67.32 | 31.09 |
| **3** | **85.71** | 50.00 | 100.00 | 74.37 | 29.23 |
| **4** | 100.00 | 60.00 | 100.00 | 79.01 | 27.72 |
| **5** | 100.00 | 66.67 | 100.00 | 82.53 | 26.09 |
| **Exon** | **1** | **50.00** | 25.00 | 80.00 | 51.05 | 32.22 |
| **2** | **80.00** | 45.00 | 100.00 | 70.44 | 32.33 |
| **3** | 100.00 | 60.00 | 100.00 | 79.53 | 28.79 |
| **4** | 100.00 | 75.00 | 100.00 | 83.92 | 26.16 |
| **5** | 100.00 | 80.00 | 100.00 | 86.20 | 24.41 |
| **3'** | **TER** | **-5** | 100.00 | 66.67 | 100.00 | 83.63 | 25.67 |
| **-4** | 100.00 | 66.67 | 100.00 | 83.34 | 25.85 |
| **-3** | 100.00 | 66.67 | 100.00 | 82.83 | 25.99 |
| **-2** | 100.00 | 50.00 | 100.00 | 79.18 | 27.95 |
| **-1** | **66.67** | 33.33 | 100.00 | 65.09 | 33.39 |
| **Exon** | **-5** | 100.00 | 66.67 | 100.00 | 81.63 | 27.75 |
| **-4** | 100.00 | 66.67 | 100.00 | 81.78 | 27.54 |
| **-3** | 100.00 | 66.67 | 100.00 | 82.58 | 26.68 |
| **-2** | 100.00 | 50.00 | 100.00 | 79.36 | 28.34 |
| **-1** | **75.00** | 33.33 | 100.00 | 67.92 | 32.96 |

**Table S10 – Frequencies of alternative exons and alternative TERs across mRNAs according to ordinal position**

Numbers in bold correspond to cited values in the main text.