For the Supplement:

*incRNA* was run with the following datasets:

For all three organisms, we used expression data including polyA+ RNA sequencing, polyA- RNA sequencing, small RNA sequencing, histone modification data including dozens of modification types such as H3K4me3, H3K36me3,H3K27me3 etc.(See supplemental Table (raw data table) in detail).

After the whole machine learning procedure, we got predicted ncRNA bins on whole genome. Adjacent bins were merged together allowing 50 bp gaps to get ncRNA fragments. If ncRNA fragments overlapped with known exonic region on the same strand or overlapped with known\_ncRNA on the opposite strand, we filter them out, sinse these might be false positives. We also filter ncRNA fragments overlapped with known ncRNAs on the same strand, only keep novel ncRNAs. Then, the remained ncRNA fragments were separated into several types based on their genomic locations (Supplemental Table 1). If ncRNA fragments overlapped with pseudogene and transposable element more than one nucleotide, we classified them correspondingly. Antisense ncRNA fragment means more than 50% length of a ncRNA fragment overlapped with known coding transcripts on the opposite strand, intronic ncRNA fragment means a ncRNA fragment fully embeded in a coding gene’s intron on the same strand, the remainders were intergenic ncRNA fragments(Lu et al.,2011). We also divided known lncRNA into different subtypes exactly the same procedure as above(Supplemental Table 2).

**Supp. Table 1. ncRNA fragment genomic location**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Species | Data type | ncRNA type | Antisensea | Intronicb | Intergenicc | Overlap with TEd | Overlap with pseudogenee | | Ambiguousf | |
| *D.melanogaster* | All | ncRNA1 | 36 | 31 | 24 | 62 | | 5 | 0 |
| ncRNA2 | 746 | 515 | 256 | 20 | | 0 | 27(13) |
| Embryo | ncRNA1 | 11 | 23 | 7 | 2 | | 1 | - |
| ncRNA2 | 209 | 187 | 40 | 6 | | - | 10(3) |
| *C.elegans* | All | ncRNA1 | 735(162)g | 522(64) | 1,930 | 55 | | 63(12) | 1,352(671) |
| ncRNA2 | 623(3) | 387(9) | 1,525 | 16 | | 19 | 136(10) |
| Embryo | ncRNA1 | 683(148) | 482(60) | 1,717 | 44 | | 47(8) | 1,254(623) |
| ncRNA2 | 608(3) | 376(9) | 1,429 | 16 | | 19 | 135(10) |
| *H.sapiens* | All | ncRNA1 | 82,555 | 60,914 | 122,048 | 720,181 | | 7,692 | 3,419 |
| ncRNA2 | 1,372(2) | 943 | 8,118 | 48,551(72) | | 1,674 | 619(24) |
| H1\_ESC | ncRNA1 | 47,344 | 52,264 | 69,140 | 364,098 | | 4,988 | 2,130 |
| ncRNA2 | 328(2) | 325 | 1,958 | 26,527(59) | | 304 | 270(13) |

a, More than 50% of ncRNA fragments overlapped with known coding transcripts on the opposite strand.

b, ncRNA fragments were fully embeded in coding gene’s intron on the same strand.

c, ncRNA fragments were fully embeded in intergenic region.

d, More than one nucleotide ncRNA fragments overlapped with transposable element.

e, More than one nucleotide ncRNA fragments overlapped with pseudogene.

f, None of the above.

g. Numbers in the parentheses are those predicted ncRNA fragments already annotated in the database (i.e. wormbase, flybase and gencode) but not confirmed or validated. They were not included in the gold-standard training set.

**Supp. Table 2 Gold standard ncRNA**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | *D.melanogaster* | | | | *C.elegans* | | | | *H.sapiens* | | | |
| **ncRNA type** | **ncRNA subtype** | **Numbers**  **in Raw Annotation** | | **Number of Bins** | **Number of Nucleotides Covered** | **Numbers**  **in Raw Annotation** | | **Number of Bins** | **Number of Nucleotides Covered** | **Numbers**  **in Raw Annotation** | | **Number of Bins** | **Number of Nucleotides Covered** |
| ncRNA1 | rRNA | 96 | | 262 | 12,958 | 22 | | 228 | 11,414 | 531 | | 1,202 | 60,053 |
| tRNA | 292 | | 456 | 22,169 | 609 | | 909 | 45,577 | 625 | | 938 | 46,643 |
| snRNA | 47 | | 135 | 7,230 | 114 | | 286 | 14,311 | 1,944 | | 4,209 | 210,373 |
| snoRNA | 286 | | 691 | 33,994 | 139 | | 309 | 15,306 | 1,521 | | 3,325 | 168,125 |
| miRNA | 237 | | 422 | 22,291 | 222 | | 406 | 19,701 | 1,595 | | 2,668 | 129,201 |
| snlRNA | - | | - | - | 4 | | 7 | 348 | - | | - | - |
| scRNA | - | | - | - | 1 | | 3 | 105 | - | | - | - |
| 7SK\_RNA | - | | - | - | - | | - | - | 298 | | 1,767 | 88,751 |
| Y\_RNA | - | | - | - | - | | - | - | 809 | | 1,675 | 83,999 |
| ncRNA2 | Antisense | 94 | | 1,989 | 129,950 | 48 | | 737 | 49,679 | 72 | | 695 | 83,845 |
| Intronic | 17 | | 166 | 12,981 | 4 | | 53 | 3,663 | - | | - | - |
| Intergenic | 147 | | 2,256 | 199,547 | 136 | | 2,379 | 176,002 | 460 | | 3,935 | 565,365 |
| TE overlapped | 33 | | 492 | 42,091 | - | | - | - | 7,343 | | 121,010 | 133,333,108 |
| Ambiguous | 26 | | 1,400 | 37,794 | 23 | | 474 | 37,106 | 30 | | 577 | 47,243 |