**Additional file 5.** The statistics of all datasets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Source | lncRNA | PCT\* | Total |
| Human training | Gencode V30 | 23,000 | 23,000 | 46,000 |
| Human testing | Gencode V30 | 6,000 | 6,000 | 12,000 |
| Human gene-wise training | Gencode V30 | 12,000 | 12,000 | 24,000 |
| Human non-gene-wise training | Gencode V30 | 12,000 | 12,000 | 24,000 |
| Human gene-wise testing | Gencode V30 | 3,855 | 6,909 | 10,764 |
| Mouse | Gencode M20 | 17,885 | 57,992 | 75,877 |
| Arabidopsis | RefSeq v94 | 4,081 | 47,135 | 51,216 |
| Bos taurus | RefSeq v94 | 6,822 | 49,547 | 56,369 |
| C.elegans | RefSeq v94 | 1,665 | 26,849 | 28,514 |
| Chicken | RefSeq v94 | 5,681 | 34,130 | 39,811 |
| Chimpanzee | RefSeq v94 | 10,709 | 59,186 | 69,895 |
| Frog | RefSeq v94 | 2,793 | 32,794 | 35,587 |
| Fruit fly | RefSeq v94 | 2,853 | 26,433 | 29,286 |
| Gorilla | RefSeq v94 | 4,457 | 38,615 | 43,072 |
| Pig | RefSeq v94 | 7,841 | 42,664 | 50,505 |
| Zebrafish | RefSeq v94 | 5,935 | 43,129 | 49,064 |

PCT\*: Protein Coding Transcript