**Table S2 The details of 8 benchmark datasets constructed in our work**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Species** | **NTR** | **NEV** | **NTE** | **NGT\_MF** | **NGT\_BP** | **NGT\_CC** |
| Human | 12501 | 735 | 1470 | 3841 | 11674 | 1505 |
| Mouse | 8965 | 527 | 1054 | 2735 | 12035 | 1188 |
| Arabidopsis | 9862 | 580 | 1160 | 2245 | 4787 | 563 |
| Rat | 4599 | 270 | 540 | 2542 | 7917 | 954 |
| Fly | 4521 | 265 | 531 | 1783 | 6123 | 857 |
| Budding Yeast | 3492 | 205 | 410 | 2025 | 4525 | 899 |
| Fission Yeast | 2332 | 137 | 274 | 1426 | 3885 | 720 |
| Nematoda | 2682 | 157 | 315 | 1160 | 4042 | 539 |

*Note:* NTR/NEV/NTE: the number of genes in training/validation/test datasets; NGT\_MF: the total number of MF terms in training, validation and test datasets;NGT\_BP: the total number of BP terms in training, validation and test datasets;NGT\_CC: the total number of CC terms in training, validation and test datasets.