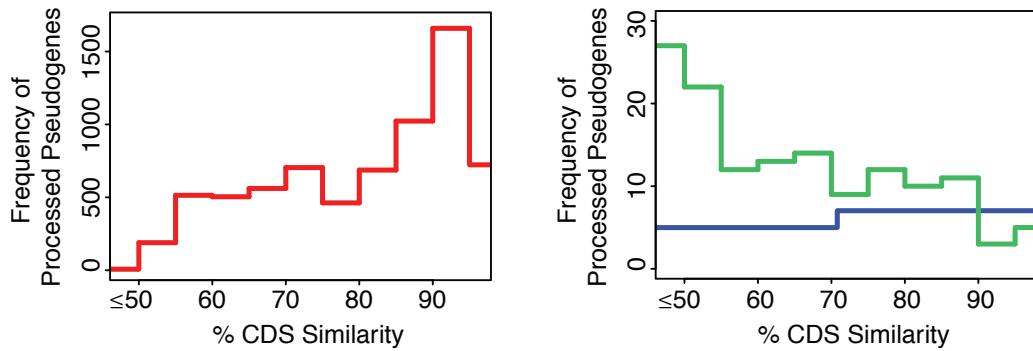


A

| Organism | Total Pseudogenes | Biotype Distribution | | ENCODE Functional Genomics Data | Completed Manual Annotation |
|-----------|-------------------|----------------------|------------|---------------------------------|-----------------------------|
| | | Processed | Duplicated | | |
| Human | 12,358 | 8908 | 2266 | ✓ | ✓ |
| Worm | 911 | 159 | 566 | ✓ | ✓ |
| Fly | 145 | 16 | 109 | ✓ | ✓ |
| Zebrafish | 229 | 21 | 177 | ✓ | ✓ |
| Macaque | 11,136 | 6570 | 1725 | X | X |
| Mouse | 13,169 | 7811 | 1827 | ✓ | X |

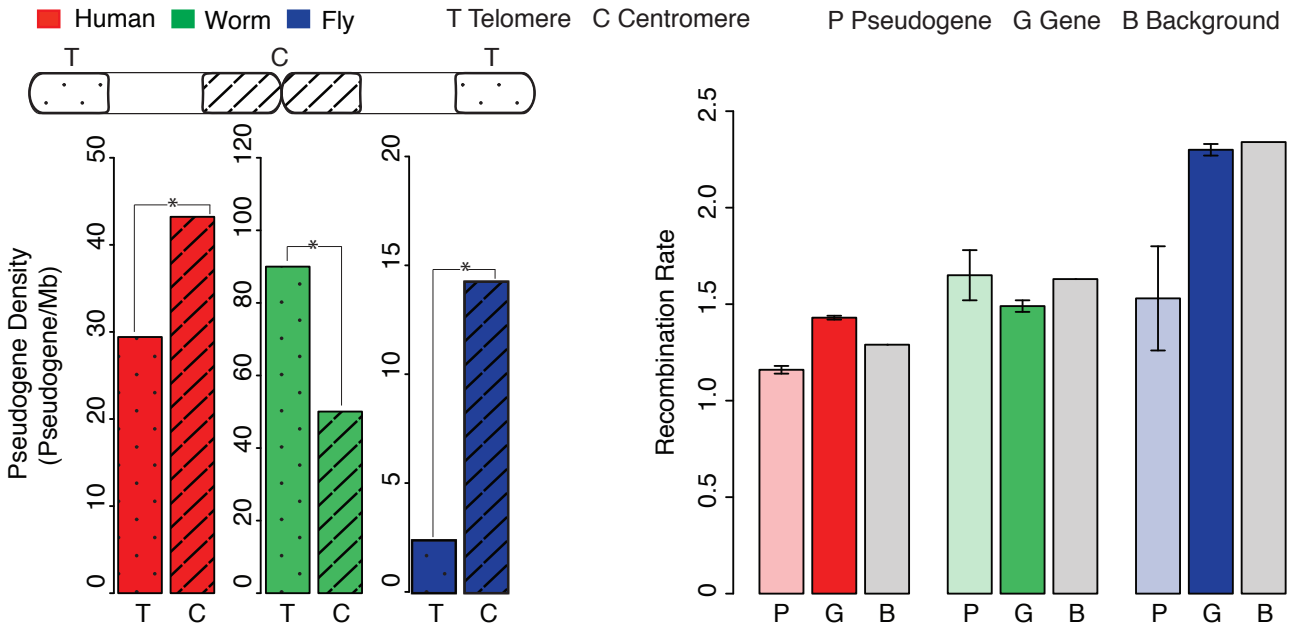
B



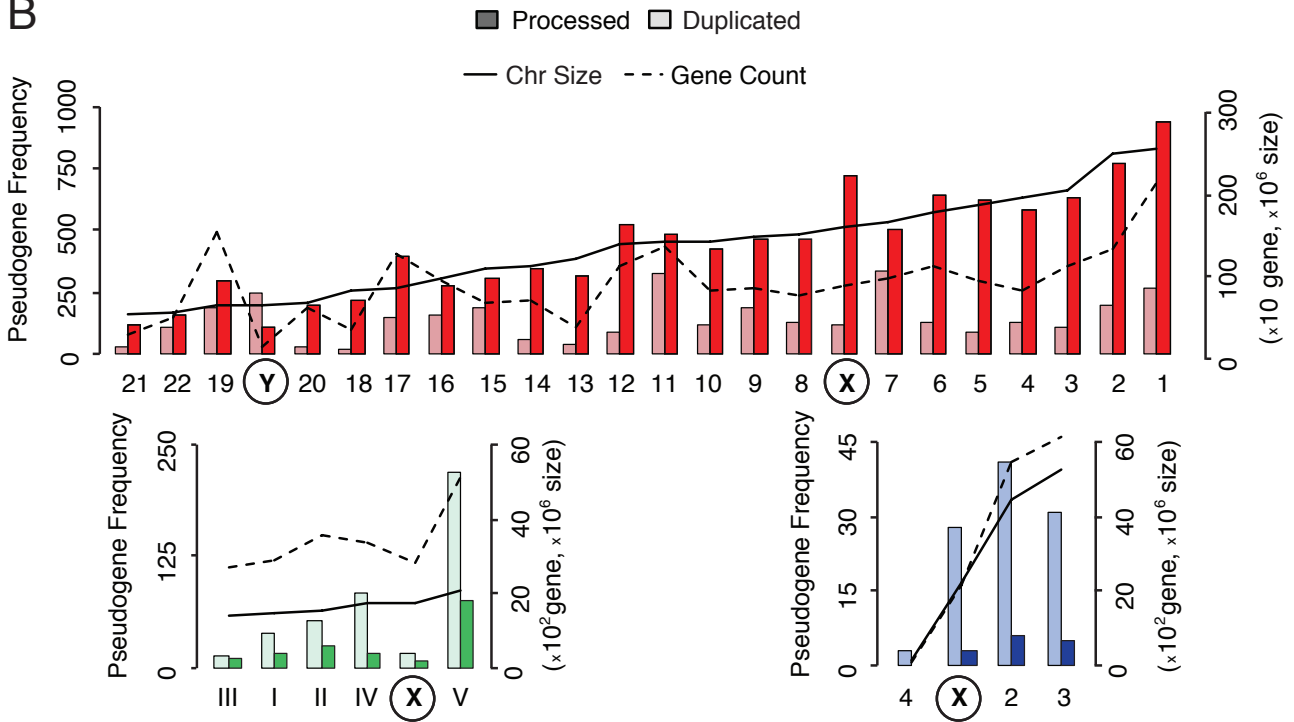
C

| Organism | Defect / Pseudogene x MB | | |
|----------|--------------------------|-------------|------|
| | Insertion | Deletion | Stop |
| Human | 4.4 | 4.9 | 2.4 |
| Worm | 25.8 | 7.45 | 2.5 |
| Fly | 7.9 | 12.7 | 1.1 |

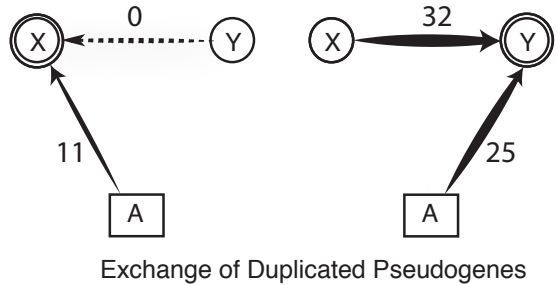
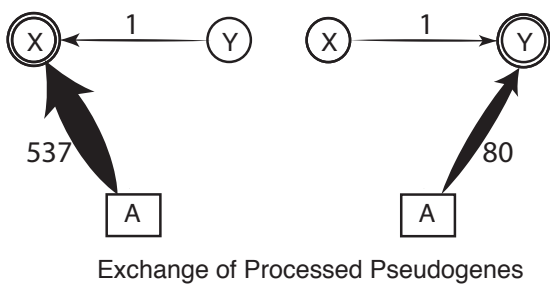
A



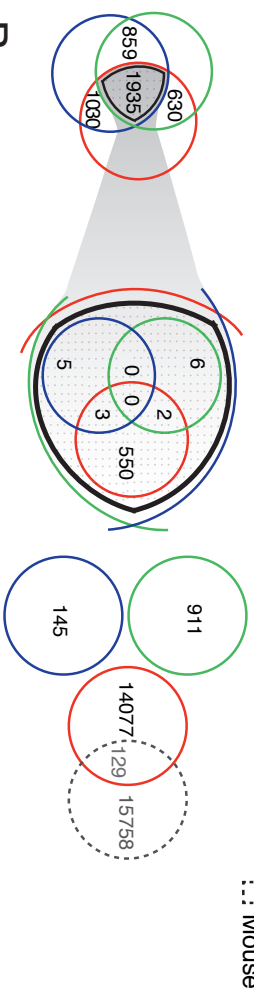
B



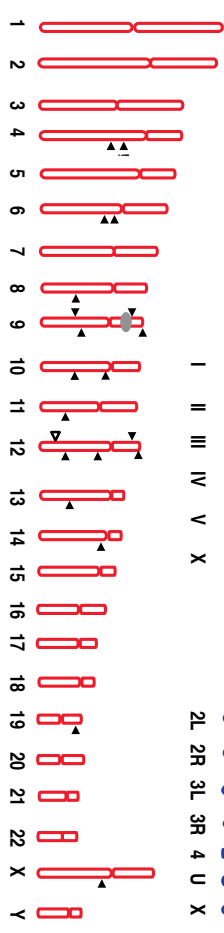
C



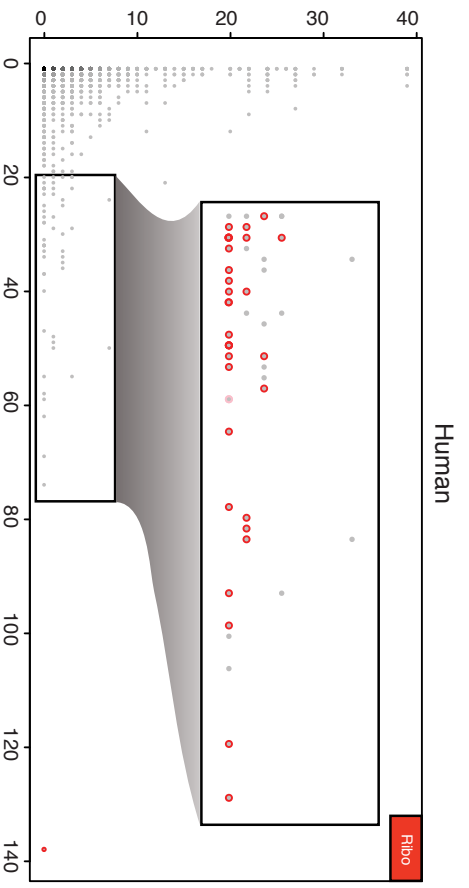
A Orthologous Genes
Parent Genes amongst 1935 1-1-1 orthologs
Orthologous Pseudogenes



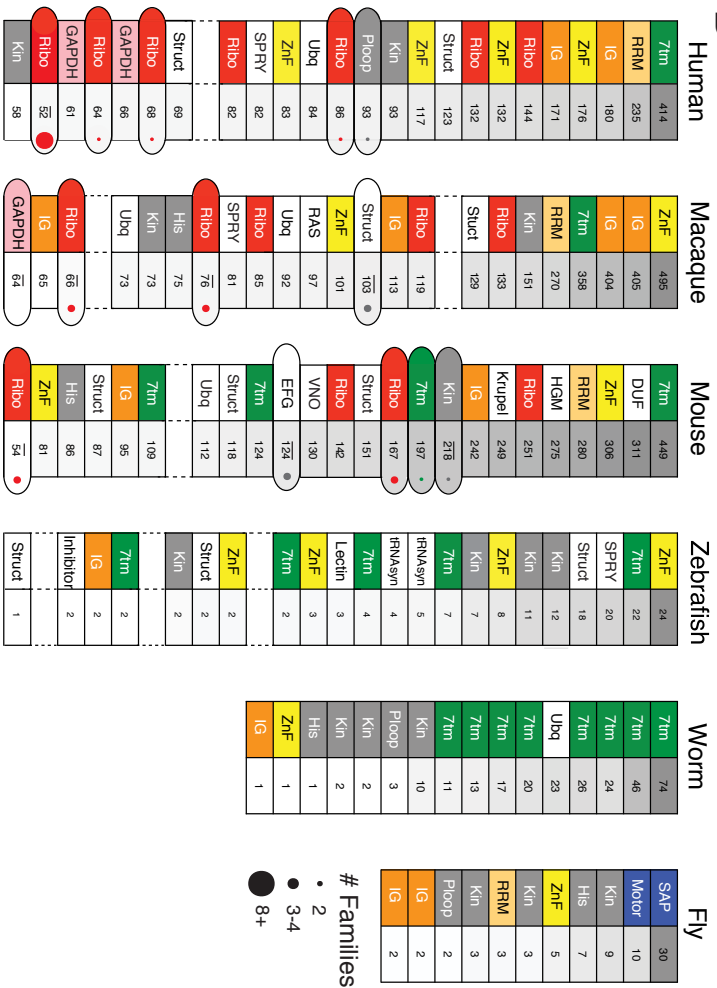
B Gene
● Processed
▲ Duplicated
▼ Transcribed

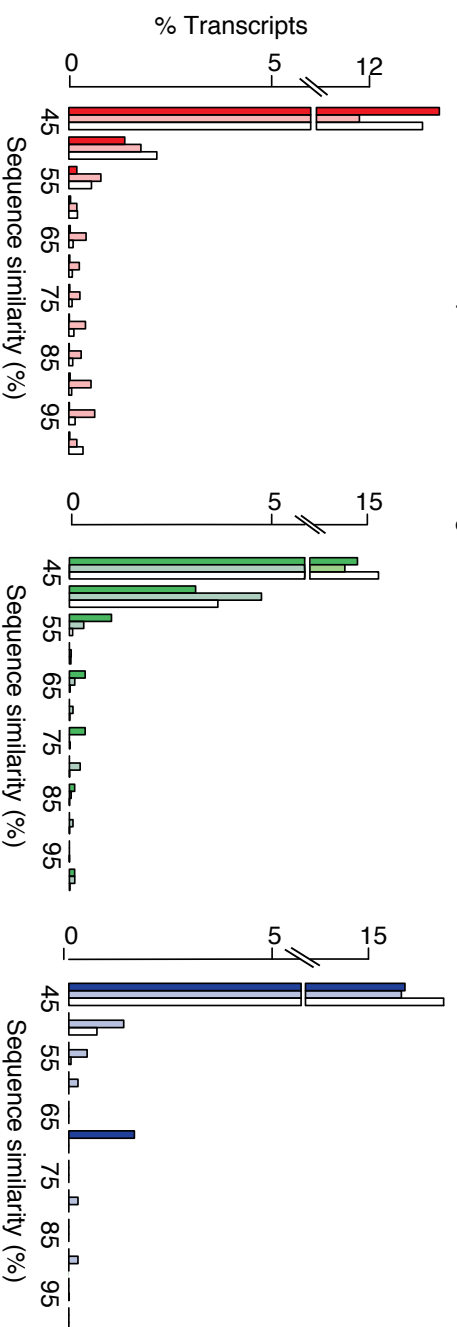
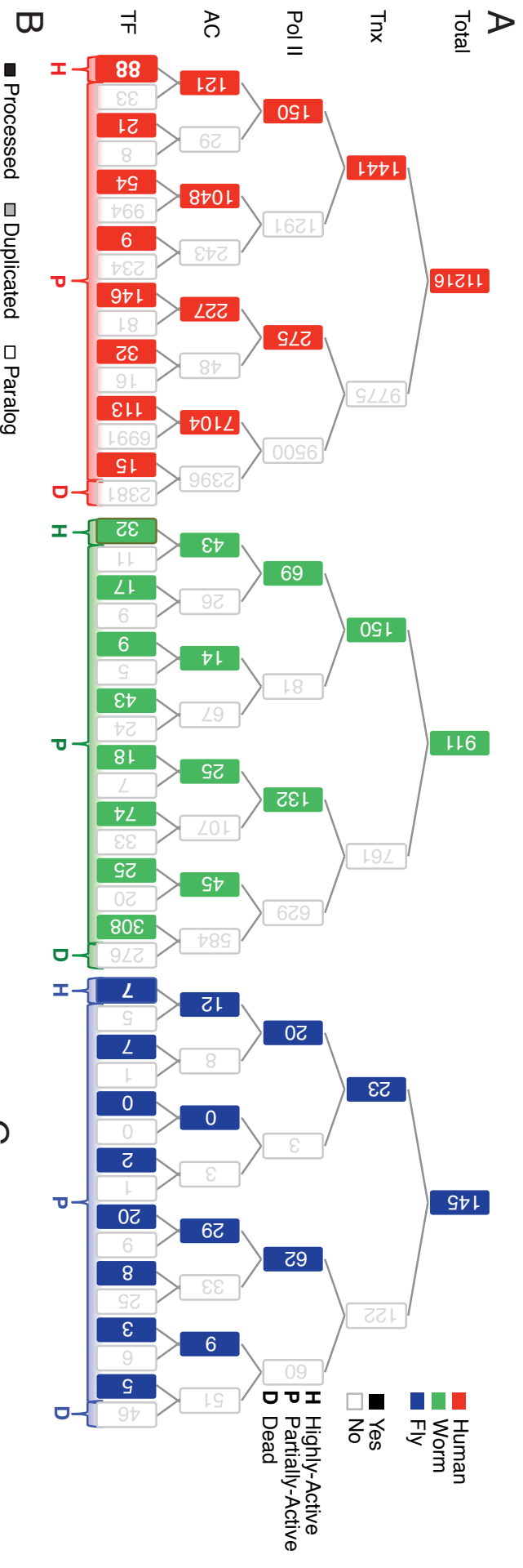


C Number of Paralogs per Gene



D





D

| Translation Candidates | Biotype | Coexp Coef | pVal | % Similarity CDS / UTR | Defect | Trnx | Pol II | AC | TF |
|---|------------|------------|--------|------------------------|-----------|------|--------|----|----|
| SLIT-ROBO Rho GTPase activating protein 2B pseudogene | Duplicated | 0.80 | 5.9E-7 | 58 / 50 | ins | ✓ | - | - | ✓ |
| PRKY-004, Y-linked protein kinase pseudogene | Duplicated | -0.14 | 0.42 | 96 / 51 | ins / del | ✓ | ✓ | - | ✓ |
| Fer-1-like 4 (C. elegans), pseudogene | Unitary | -0.38 | 0.03 | 62 / 32 | ins / del | ✓ | - | - | ✓ |

C

| Pseudogene | Parent | | Paralog | |
|------------|--------|-----|---------|-----|
| | - | + | - | + |
| + | 11 | 131 | 31 | 336 |
| - | 37 | 253 | 17 | 48 |
| + | 5 | 6 | 7 | 12 |
| - | 20 | 13 | 18 | 7 |
| + | 0 | 4 | 3 | 3 |
| - | 4 | 1 | 1 | 2 |