

Gene numbers were based on the annotations in the Ensembl database (build 49) and also in literature. At present, it still remains a great challenge to obtain an accurate estimation of gene number in a genome, but we suspect that the actual gene numbers in these genomes are likely in a smaller range than the range 20,000 – 30,000 in table S1.

**Table S1** Number of genes in mammalian genomes

| Species    | Ensembl     |                 |             |             |           |               | Estimates in literature                                      |
|------------|-------------|-----------------|-------------|-------------|-----------|---------------|--|
|            | Known genes | Projected genes | Novel genes | Pseudogenes | RNA genes | <b>Total</b>  |  |
| Human      | 22,658      | N/A             | 82          | 2,081       | 4,421     | <b>29,242</b> | 20,000 – 25,000 [1] <sup>a</sup><br>20,500 [2]<br>19,700 [3] |
| Chimpanzee | 2,577       | 16,858          | 1,108       | 872         | 4,240     | <b>25,655</b> | 16,045 [4] <sup>b</sup>                                      |
| Macaque    | 874         | 17,340          | 3,691       | 1,762       | 4,003     | <b>27,670</b> | ~20,000 – 25,000 [5] <sup>c</sup>                            |
| Mouse      | 22,010      | N/A             | 1,483       | 1,190       | 3,014     | <b>27,697</b> | ~30,000 [6] <sup>d</sup>                                     |
| Rat        | 17,738      | 1,919           | 2,846       | 1,756       | 2,924     | <b>27,183</b> | 20,973 [7] <sup>e</sup>                                      |
| Dog        | 1,461       | 14,198          | 3,646       | 1,742       | 2,448     | <b>23,495</b> | 19,300 [8] <sup>f</sup>                                      |
| Cow        | 16,938      | 1,781           | 3,027       | 1,264       | 2,496     | <b>25,506</b> | N/A  |
| Horse      | 746         | 13,601          | 6,390       | 4,443       | N/A       | <b>25,180</b> | N/A  |
| Opossum    | 119         | 13,935          | 5,417       | 722         | 1,251     | <b>21,444</b> | 18,000 – 20,000 [9] <sup>g</sup>                             |
| Platypus   | 54          | 10,866          | 7,031       | 547         | 3,074     | <b>21,572</b> | N/A  |

<sup>a</sup>Protein-coding genes.

<sup>b</sup>Based on the alignment to human RefSeq transcripts.

<sup>c</sup>Predicted by Ensembl gene models, Gnomon (NCBI) and Nscan.

<sup>d</sup>Protein-coding genes.

<sup>e</sup>Predicted by the Ensembl gene prediction pipeline.

<sup>f</sup>Based on evidence-based method.

<sup>g</sup>Protein-coding genes.

## References

1. **Finishing the euchromatic sequence of the human genome.** *Nature* 2004, **431**:931-945.
2. Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, Kellis M, Lindblad-Toh K, Lander ES: **Distinguishing protein-coding and noncoding genes in the human genome.** *Proc Natl Acad Sci USA* 2007, **104**:19428-19433.
3. Goodstadt L, Ponting CP: **Phylogenetic reconstruction of orthology, paralogy, and**

- conserved synteny for dog and human.** *PLoS Comput Biol* 2006, **2**:e133.
4. **Initial sequence of the chimpanzee genome and comparison with the human genome.** *Nature* 2005, **437**:69-87.
  5. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, *et al.*: **Evolutionary and biomedical insights from the rhesus macaque genome.** *Science* 2007, **316**:222-234.
  6. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, *et al.*: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.
  7. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, *et al.*: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution.** *Nature* 2004, **428**:493-521.
  8. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas III EJ, Zody MC, *et al.*: **Genome sequence, comparative analysis and haplotype structure of the domestic dog.** *Nature* 2005, **438**:803-819.
  9. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, *et al.*: **Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences.** *Nature* 2007, **447**:167-177.