

- Y. Kapustin, O. Khovayko, D. Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, J. Ostell, V. Miller, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, R. L. Tatusov, T. A. Tatusova, L. Wagner and E. Yaschenko, *Nucleic Acids Res.* **35**, D5 (2007).
8. C. von Mering, L. J. Jensen, M. Kuhn, S. Chaffron, T. Doerks, B. Krüger, B. Snel and P. Bork, *Nucleic Acids Res.* **35**, D358 (2007).
 9. The UniProt Consortium, *Nucleic Acids Res.* **35**, D193 (2007).
 10. C. J. Date. An introduction to database systems. Addison-Wesley (1999).
 11. R. Durbin, S. Eddy, A. Krogh and G. Mitchison. Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge University Press (1998).
 12. H. Mi, N. Guo, A. Kejariwal and P. D. Thomas, *Nucleic Acids Res.* **35**, D247 (2007).
 13. R. D. Finn, J. Mistry, B. Schuster-Böckler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. L. Sonnhammer and A. Bateman, *Nucleic Acids Res.* **34**, D247 (2006).
 14. D. Wilson, M. Madera, C. Vogel, C. Chothia and J. Gough, *Nucleic Acids Res.* **35**, D308 (2007).
 15. <http://hgdownload.cse.ucsc.edu/goldenPath/>.
 16. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock, *Nat. Genet.* **25**, 25 (2000).
 17. E. A. Bruford, M. J. Lush, M. W. Wright, T. P. Sneddon, S. Povey and E. Birney, *Nucleic Acids Res.* **36**, D445 (2008).
 18. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.* **28**, 235 (2000).
 19. A. Andreeva, D. Howorth, J. Chandonia, S. E. Brenner, T. J. P. Hubbard, C. Chothia and A. G. Murzin, *Nucleic Acids Res.* **36**, D419 (2008).
 20. <http://scientific.thomsonreuters.com/pharma/geneseq/>.
 21. J. E. Stajich, *Methods Mol. Biol.* **406**, 535 (2007).
 22. E. Vivier, E. Tomasello, M. Baratin, T. Walzer and S. Ugolini, *Nat. Immunol.* **9**, 503 (2008).
 23. E. Kratz, P. M. Eimon, K. Mukhyala, H. Stern, J. Zha, A. Strasser, R. Hart and A. Ashkenazi, *Cell Death Differ.* **13**, 1631 (2006).
 24. J. Liu, Y. Zhang, X. Lei and Z. Zhang, *Genome Biol.* **9**, R69 (2008).
 25. R. D. Dowell, R. M. Jokerst, A. Day, S. R. Eddy and L. Stein, *BMC Bioinformatics* **2**, 7 (2001).