1.5 Bibliography

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol 215, 403-410.

Altschul, S.F., and Koonin, E.V. (1998). Iterated profile searches with PSI-BLAST-- a tool for discovery in protein databases. Trends Biochem Sci 23, 444-447.

Andreeva, A., Howorth, D., Chandonia, J.M., Brenner, S.E., Hubbard, T.J., Chothia, C., and Murzin, A.G. (2008). Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 36, D419-425.

Arguello, J.R., Chen, Y., Yang, S., Wang, W., and Long, M. (2006). Origination of an X-linked testes chimeric gene by illegitimate recombination in Drosophila. PLoS Genet 2, e77.

Attwood, T.K., Bradley, P., Flower, D.R., Gaulton, A., Maudling, N., Mitchell, A.L., Moulton, G., Nordle, A., Paine, K., Taylor, P., et al. (2003). PRINTS and its automatic supplement, prePRINTS. Nucleic Acids Res 31, 400-402.

Babushok, D.V., Ostertag, E.M., and Kazazian, H.H., Jr. (2007). Current topics in genome evolution: molecular mechanisms of new gene formation. Cell Mol Life Sci 64, 542-554.

Bartlett, G.J., Borkakoti, N., and Thornton, J.M. (2003). Catalysing new reactions during evolution: economy of residues and mechanism. J Mol Biol 331, 829-860.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002). The Pfam protein families database. Nucleic Acids Res 30, 276-280.

Berglund, A.C., Sjolund, E., Ostlund, G., and Sonnhammer, E.L. (2008). InParanoid 6: eukaryotic ortholog clusters with inparalogs. Nucleic Acids Res 36, D263-266.

Betran, E., Thornton, K., and Long, M. (2002). Retroposed new genes out of the X in Drosophila. Genome Res 12, 1854-1859.

Bjorklund, A.K., Ekman, D., and Elofsson, A. (2006). Expansion of protein domain repeats. PLoS Comput Biol 2, e114.

Bru, C., Courcelle, E., Carrere, S., Beausse, Y., Dalmar, S., and Kahn, D. (2005). The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res 33, D212-215.

Buljan, M., and Bateman, A. (2009). The evolution of protein domain families. Biochem Soc Trans 37, 751-755.

Buljan, M., Frankish, A., and Bateman, A. (2010). Quantifying the mechanisms of domain gain in animal proteins. Genome Biol 11, R74.

Burki, F., and Kaessmann, H. (2004). Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. Nat Genet 36, 1061-1063.

Caceres, J.F., and Kornblihtt, A.R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet 18, 186-193.

Cheng, Y., LeGall, T., Oldfield, C.J., Mueller, J.P., Van, Y.Y., Romero, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K. (2006). Rational drug design via intrinsically disordered protein. Trends Biotechnol 24, 435-442.

Chothia, C., and Gough, J. (2009). Genomic and structural aspects of protein evolution. Biochem J 419, 15-28.

Chothia, C., Gough, J., Vogel, C., and Teichmann, S.A. (2003). Evolution of the protein repertoire. Science 300, 1701-1703.

Ciccarelli, F.D., von Mering, C., Suyama, M., Harrington, E.D., Izaurralde, E., and Bork, P. (2005). Complex genomic rearrangements lead to novel primate gene function. Genome Res 15, 343-351.

Cohen, P. (2001). The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. Eur J Biochem 268, 5001-5010.

Coin, L. (2008). Protein Domains: New methods for detection and evolutionary analysis. In Wellcome Trust Sanger Institute (Cambridge, University of Cambridge).

Conant, G.C., and Wagner, A. (2005). The rarity of gene shuffling in conserved genes. Genome Biol 6, R50.

Deininger, P.L., Moran, J.V., Batzer, M.A., and Kazazian, H.H., Jr. (2003). Mobile elements and mammalian genome evolution. Curr Opin Genet Dev 13, 651-658.

Denning, D.P., Patel, S.S., Uversky, V., Fink, A.L., and Rexach, M. (2003). Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. Proc Natl Acad Sci U S A 100, 2450-2455.

Dosztanyi, Z., Chen, J., Dunker, A.K., Simon, I., and Tompa, P. (2006). Disorder and sequence repeats in hub proteins and their implications for network evolution. J Proteome Res 5, 2985-2995.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol 347, 827-839.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., et al. (2001). Intrinsically disordered protein. J Mol Graph Model 19, 26-59.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6, 197-208.

Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics 14, 755-763.

Ekman, D., Bjorklund, A.K., and Elofsson, A. (2007). Quantification of the elevated rate of domain rearrangements in metazoa. J Mol Biol 372, 1337-1348.

Ekman, D., Bjorklund, A.K., Frey-Skott, J., and Elofsson, A. (2005). Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. J Mol Biol 348, 231-243.

Emes, R.D., Goodstadt, L., Winter, E.E., and Ponting, C.P. (2003). Comparison of the genomes of human and mouse lays the foundation of genome zoology. Hum Mol Genet 12, 701-709.

Feng, D.F., Cho, G., and Doolittle, R.F. (1997). Determining divergence times with a protein clock: update and reevaluation. Proc Natl Acad Sci U S A 94, 13028-13033.

Fernandez-Fuentes, N., Dybas, J.M., and Fiser, A. (2010). Structural characteristics of novel protein folds. PLoS Comput Biol 6, e1000750.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., et al. (2010). The Pfam protein families database. Nucleic Acids Res 38, D211-222.

Fitch, W.M. (1970). Distinguishing homologous from analogous proteins. Syst Zool 19, 99-113.

Flicek, P., Aken, B.L., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., et al. (2010). Ensembl's 10th year. Nucleic Acids Res 38, D557-562.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531-1545.

Fuxreiter, M., Tompa, P., and Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. Bioinformatics 23, 950-956.

Gahmberg, C.G., and Tolvanen, M. (1996). Why mammalian cell surface proteins are glycoproteins. Trends Biochem Sci 21, 308-311.

Garcia-Blanco, M.A., Baraniak, A.P., and Lasda, E.L. (2004). Alternative splicing in disease and therapy. Nat Biotechnol 22, 535-546.

Garner, E., Cannon, P., Romero, P., Obradovic, Z., and Dunker, A.K. (1998). Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization. Genome Inform Ser Workshop Genome Inform 9, 201-213.

Goldstein, R.A. (2008). The structure of protein evolution and the evolution of protein structure. Curr Opin Struct Biol 18, 170-177.

Greene, L.H., Lewis, T.E., Addou, S., Cuff, A., Dallman, T., Dibley, M., Redfern, O., Pearl, F., Nambudiry, R., Reid, A., et al. (2007). The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. Nucleic Acids Res 35, D291-297.

Gribskov, M., McLachlan, A.D., and Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci U S A 84, 4355-4358.

Gsponer, J., and Babu, M.M. (2009). The rules of disorder or why disorder rules. Prog Biophys Mol Biol 99, 94-103.

Gsponer, J., Futschik, M.E., Teichmann, S.A., and Babu, M.M. (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science 322, 1365-1368.

Gu, X., Wang, Y., and Gu, J. (2002). Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat Genet 31, 205-209.

H Lodish, A.B., S L Zipursky, P Matsudaira, D Baltimore, and J Darnell (2000). Molecular Cell Biology, 4th edition edn (New York, W. H. Freeman).

Haber, J.E. (2000). Partners and pathwaysrepairing a double-strand break. Trends Genet 16, 259-264.

Haft, D.H., Selengut, J.D., and White, O. (2003). The TIGRFAMs database of protein families. Nucleic Acids Res 31, 371-373.

Haines, N., and Irvine, K.D. (2003). Glycosylation regulates Notch signalling. Nat Rev Mol Cell Biol 4, 786-797.

Han, J.H., Batey, S., Nickson, A.A., Teichmann, S.A., and Clarke, J. (2007). The folding and evolution of multidomain proteins. Nat Rev Mol Cell Biol 8, 319-330.

Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2, e100.

He, X., and Zhang, J. (2005). Gene complexity and gene duplicability. Curr Biol 15, 1016-1021.

Heger, A., Wilton, C.A., Sivakumar, A., and Holm, L. (2005). ADDA: a domain database with global coverage of the protein universe. Nucleic Acids Res 33, D188-191.

Henikoff, J.G., Pietrokovski, S., McCallum, C.M., and Henikoff, S. (2000). Blocks-based methods for detecting protein homology. Electrophoresis 21, 1700-1706.

Hillman, R.T., Green, R.E., and Brenner, S.E. (2004). An unappreciated role for RNA surveillance. Genome Biol 5, R8.

Holm, L., and Sander, C. (1994). Parser for protein folding units. Proteins 19, 256-268.

Hughes, A.L. (1999). Adaptive evolution of genes and genome (New York, Oxford University Press).

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., and Sigrist, C.J. (2006). The PROSITE database. Nucleic Acids Res 34, D227-230.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., et al. (2009). InterPro: the integrative protein signature database. Nucleic Acids Res 37, D211-215.

Hurles, M. (2004). Gene duplication: the genomic trade in spare parts. PLoS Biol 2, E206.

Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z., and Dunker, A.K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323, 573-584.

Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z., and Dunker, A.K. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 32, 1037-1049.

Jacob, F. (1977). Evolution and tinkering. Science 196, 1161-1166.

Johnson, M.E., Viggiano, L., Bailey, J.A., Abdul-Rauf, M., Goodwin, G., Rocchi, M., and Eichler, E.E. (2001). Positive selection of a gene family during the emergence of humans and African apes. Nature 413, 514-519.

Kaessmann, H. (2009). Genetics. More than just a copy. Science 325, 958-959.

Kaessmann, H., Vinckenbosch, N., and Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet 10, 19-31.

Kall, L., Krogh, A., and Sonnhammer, E.L. (2004). A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338, 1027-1036.

Knowles, D.G., and McLysaght, A. (2009). Recent de novo origin of human protein-coding genes. Genome Res 19, 1752-1759.

Kondrashov, F.A., and Koonin, E.V. (2003). Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. Trends Genet 19, 115-119.

Koonin, E.V. (2009). Darwinian evolution in the light of genomics. Nucleic Acids Res 37, 1011-1034.

Landry, C.R., Levy, E.D., and Michnick, S.W. (2009). Weak functional constraints on phosphoproteomes. Trends Genet 25, 193-197.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446, 926-929.

Letunic, I., Copley, R.R., and Bork, P. (2002). Common exon duplication in animals and its role in alternative splicing. Hum Mol Genet 11, 1561-1567.

Letunic, I., Copley, R.R., Pils, B., Pinkert, S., Schultz, J., and Bork, P. (2006). SMART 5: domains in the context of genomes and networks. Nucleic Acids Res 34, D257-260.

Lev-Maor, G., Goren, A., Sela, N., Kim, E., Keren, H., Doron-Faigenboim, A., Leibman-Barak, S., Pupko, T., and Ast, G. (2007). The "alternative" choice of constitutive exons throughout evolution. PLoS Genet 3, e203.

Li, H., Coghlan, A., Ruan, J., Coin, L.J., Heriche, J.K., Osmotherly, L., Li, R., Liu, T., Zhang, Z., Bolund, L., et al. (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res 34, D572-580.

Li, W.H. (1997). Molecular Evolution (Sunderland Massachusetts, Sinauer Associates, Inc.).

Linding, R., Russell, R.B., Neduva, V., and Gibson, T.J. (2003). GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31, 3701-3708.

Long, M. (2001). Evolution of novel genes. Curr Opin Genet Dev 11, 673-680.

Ludin, B., Ashbridge, K., Funfschilling, U., and Matus, A. (1996). Functional analysis of the MAP2 repeat domain. J Cell Sci 109 (Pt 1), 91-99.

Lukong, K.E., Chang, K.W., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends Genet 24, 416-425.

Makino, T., and McLysaght, A. (2010). Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc Natl Acad Sci U S A 107, 9270-9274.

Mann, M., and Jensen, O.N. (2003). Proteomic analysis of post-translational modifications. Nat Biotechnol 21, 255-261.

Marsh, J.A., and Teichmann, S.A. (2010). How do proteins gain new domains? Genome Biol 11, 126.

Melamud, E., and Moult, J. (2009). Structural implication of splicing stochastics. Nucleic Acids Res 37, 4862-4872.

Mendes Soares, L.M., and Valcarcel, J. (2006). The expanding transcriptome: the genome as the 'Book of Sand'. EMBO J 25, 923-931.

Modrek, B., and Lee, C.J. (2003). Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet 34, 177-180.

Moore, A.D., Bjorklund, A.K., Ekman, D., Bornberg-Bauer, E., and Elofsson, A. (2008). Arrangements in the modular evolution of proteins. Trends Biochem Sci 33, 444-451.

Neduva, V., and Russell, R.B. (2005). Linear motifs: evolutionary interaction switches. FEBS Lett 579, 3342-3345.

Nei, M., Rogozin, I.B., and Piontkivska, H. (2000). Purifying selection and birth-and-death evolution in the ubiquitin gene family. Proc Natl Acad Sci U S A 97, 10866-10871.

Neurath, H., and Walsh, K.A. (1976). Role of proteolytic enzymes in biological regulation (a review). Proc Natl Acad Sci U S A 73, 3825-3832.

Ohno, S. (1970). Evolution by gene duplication (Berlin, Springer-Verlag).

Pal, C., Papp, B., and Lercher, M.J. (2006). An integrated view of protein evolution. Nat Rev Genet 7, 337-348.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40, 1413-1415.

Parmley, J.L., Urrutia, A.O., Potrzebowski, L., Kaessmann, H., and Hurst, L.D. (2007). Splicing and the evolution of proteins in mammals. PLoS Biol 5, e14.

Pasek, S., Risler, J.L., and Brezellec, P. (2006). Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. Bioinformatics 22, 1418-1423.

Patthy, L. (1996). Exon shuffling and other ways of module exchange. Matrix Biol 15, 301-310; discussion 311-302.

Pawson, T. (2003). Organization of cell-regulatory systems through modular-protein-interaction domains. Philos Transact A Math Phys Eng Sci 361, 1251-1262.

Pils, B., and Schultz, J. (2004). Inactive enzyme-homologues find new function in regulatory processes. J Mol Biol 340, 399-404.

Reeck, G.R., de Haen, C., Teller, D.C., Doolittle, R.F., Fitch, W.M., Dickerson, R.E., Chambon, P., McLachlan, A.D., Margoliash, E., Jukes, T.H., et al. (1987).

"Homology" in proteins and nucleic acids: a terminology muddle and a way out of it. Cell 50, 667.

Resh, M.D. (1999). Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. Biochim Biophys Acta 1451, 1-16.

Rosenberg, H.F. (1995). Recombinant human eosinophil cationic protein. Ribonuclease activity is not essential for cytotoxicity. J Biol Chem 270, 7876-7881.

Roth, C., Rastogi, S., Arvestad, L., Dittmar, K., Light, S., Ekman, D., and Liberles, D.A. (2007). Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. J Exp Zool B Mol Dev Evol 308, 58-73.

Rueter, S.M., Dawson, T.R., and Emeson, R.B. (1999). Regulation of alternative splicing by RNA editing. Nature 399, 75-80.

Schadt, E.E. (2009). Molecular networks as sensors and drivers of common human diseases. Nature 461, 218-223.

Shimizu, K., and Toh, H. (2009). Interaction between intrinsically disordered proteins frequently occurs in a human protein-protein interaction network. J Mol Biol 392, 1253-1265.

Siepel, A. (2009). Darwinian alchemy: Human genes from noncoding DNA. Genome Res 19, 1693-1695.

Sorek, R., Ast, G., and Graur, D. (2002). Alu-containing exons are alternatively spliced. Genome Res 12, 1060-1067.

Tocchini-Valentini, G.D., Fruscoloni, P., and Tocchini-Valentini, G.P. (2005). Structure, function, and evolution of the tRNA endonucleases of Archaea: an example of subfunctionalization. Proc Natl Acad Sci U S A 102, 8933-8938.

Tokuriki, N., and Tawfik, D.S. (2009). Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459, 668-673.

Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett 579, 3346-3354.

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J 18, 1169-1175.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res 19, 327-335.

Vogel, C., Bashton, M., Kerrison, N.D., Chothia, C., and Teichmann, S.A. (2004). Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol 14, 208-216.

Vogel, C., Teichmann, S.A., and Pereira-Leal, J. (2005). The relationship between domain duplication and recombination. J Mol Biol 346, 355-365.

Vucetic, S., Obradovic, Z., Vacic, V., Radivojac, P., Peng, K., Iakoucheva, L.M., Cortese, M.S., Lawson, J.D., Brown, C.J., Sikes, J.G., et al. (2005). DisProt: a database of protein disorder. Bioinformatics 21, 137-140.

Walsh, C.T. (2006). Posttranslational Modification of Proteins (Englewood, Colorado, Roberts and Company Publishers).

Wang, W., Yu, H., and Long, M. (2004). Duplication-degeneration as a mechanism of gene fission and the origin of new genes in Drosophila species. Nat Genet 36, 523-527.

Wang, W., Zheng, H., Yang, S., Yu, H., Li, J., Jiang, H., Su, J., Yang, L., Zhang, J., McDermott, J., et al. (2005). Origin and evolution of new exons in rodents. Genome Res 15, 1258-1264.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337, 635-645.

Weiner, J., 3rd, Beaussart, F., and Bornberg-Bauer, E. (2006). Domain deletions and substitutions in the modular protein evolution. FEBS J 273, 2037-2047.

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., and Gough, J. (2009). SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res 37, D380-386.

Wojtowicz, W.M., Flanagan, J.J., Millard, S.S., Zipursky, S.L., and Clemens, J.C. (2004). Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell 118, 619-633.

Wootton, J.C. (1994). Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput Chem 18, 269-285.

Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293, 321-331.

Yeats, C., Lees, J., Reid, A., Kellam, P., Martin, N., Liu, X., and Orengo, C. (2008). Gene3D: comprehensive structural and functional annotation of genomes. Nucleic Acids Res 36, D414-418.

Yokoyama, S., and Yokoyama, R. (1989). Molecular evolution of human visual pigment genes. Mol Biol Evol 6, 186-197.

Zhang, J. (2003). Evolution by gene duplication: an update. Trends in Ecology & Evolution 18, 292-298.

Zhang, J., Rosenberg, H.F., and Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci U S A 95, 3708-3713.

Zhang, L., Ma, B., Wang, L., and Xu, Y. (2003). Greedy method for inferring tandem duplication history. Bioinformatics 19, 1497-1504.

Zheng, D., Zhang, Z., Harrison, P.M., Karro, J., Carriero, N., and Gerstein, M. (2005). Integrated pseudogene annotation for human chromosome 22: evidence for transcription. J Mol Biol 349, 27-45.

2.5 Bibliography

Babushok, D.V., Ostertag, E.M., and Kazazian, H.H., Jr. (2007). Current topics in genome evolution: molecular mechanisms of new gene formation. Cell Mol Life Sci *64*, 542-554.

Basu, M.K., Carmel, L., Rogozin, I.B., and Koonin, E.V. (2008). Evolution of protein domain promiscuity in eukaryotes. Genome research *18*, 449-461.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002). The Pfam protein families database. Nucleic acids research *30*, 276-280.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. (2000). The Pfam protein families database. Nucleic acids research *28*, 263-266.

Beaussart, F., Weiner, J., 3rd, and Bornberg-Bauer, E. (2007). Automated Improvement of Domain ANnotations using context analysis of domain arrangements (AIDAN). Bioinformatics (Oxford, England) *23*, 1834-1836.

Bjorklund, A.K., Ekman, D., and Elofsson, A. (2006). Expansion of protein domain repeats. PLoS computational biology *2*, e114.

Bjorklund, A.K., Ekman, D., Light, S., Frey-Skott, J., and Elofsson, A. (2005). Domain rearrangements in protein evolution. Journal of molecular biology *353*, 911-923.

Bru, C., Courcelle, E., Carrere, S., Beausse, Y., Dalmar, S., and Kahn, D. (2005). The ProDom database of protein domain families: more emphasis on 3D. Nucleic acids research *33*, D212-215.

Chothia, C., Gough, J., Vogel, C., and Teichmann, S.A. (2003). Evolution of the protein repertoire. Science (New York, NY *300*, 1701-1703.

Coin, L., Bateman, A., and Durbin, R. (2003). Enhanced protein domain discovery by using language modeling techniques from speech recognition. Proceedings of the National Academy of Sciences of the United States of America *100*, 4516-4520.

Coin, L., Bateman, A., and Durbin, R. (2004). Enhanced protein domain discovery using taxonomy. BMC bioinformatics *5*, 56.

Das, S., and Smith, T.F. (2000). Identifying nature's protein Lego set. Advances in protein chemistry *54*, 159-183.

Ekman, D., Bjorklund, A.K., Frey-Skott, J., and Elofsson, A. (2005). Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. Journal of molecular biology *348*, 231-243.

Enright, A.J., Iliopoulos, I., Kyrpides, N.C., and Ouzounis, C.A. (1999). Protein interaction maps for complete genomes based on gene fusion events. Nature *402*, 86-90.

Fong, J.H., Geer, L.Y., Panchenko, A.R., and Bryant, S.H. (2007). Modeling the evolution of protein domain architectures using maximum parsimony. Journal of molecular biology *366*, 307-315.

Forslund, K., Henricson, A., Hollich, V., and Sonnhammer, E.L. (2008). Domain tree-based analysis of protein architecture evolution. Molecular biology and evolution *25*, 254-264.

Gough, J. (2005). Convergent evolution of domain architectures (is rare). Bioinformatics (Oxford, England) *21*, 1464-1471.

Holm, L., and Sander, C. (1994). Parser for protein folding units. Proteins *19*, 256-268.

Itoh, M., Nacher, J.C., Kuma, K., Goto, S., and Kanehisa, M. (2007). Evolutionary history and functional implications of protein domains and their combinations in eukaryotes. Genome biology *8*, R121.

Kummerfeld, S.K., and Teichmann, S.A. (2005). Relative rates of gene fusion and fission in multi-domain proteins. Trends Genet *21*, 25-30.

Li, H., Coghlan, A., Ruan, J., Coin, L.J., Heriche, J.K., Osmotherly, L., Li, R., Liu, T., Zhang, Z., Bolund, L.*, et al.* (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic acids research *34*, D572-580.

Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O., and Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. Science (New York, NY *285*, 751-753.

Moore, A.D., Bjorklund, A.K., Ekman, D., Bornberg-Bauer, E., and Elofsson, A. (2008). Arrangements in the modular evolution of proteins. Trends in biochemical sciences *33*, 444-451.

Oliver, P.L., Goodstadt, L., Bayes, J.J., Birtle, Z., Roach, K.C., Phadnis, N., Beatson, S.A., Lunter, G., Malik, H.S., and Ponting, C.P. (2009). Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. PLoS genetics *5*, e1000753.

Pal, C., Papp, B., and Lercher, M.J. (2006). An integrated view of protein evolution. Nature reviews *7*, 337-348.

Pasek, S., Risler, J.L., and Brezellec, P. (2006). Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. Bioinformatics (Oxford, England) *22*, 1418-1423.

Patthy, L. (1999). Genome evolution and the evolution of exon-shuffling--a review. Gene *238*, 103-114.

Schlicker, A., Domingues, F.S., Rahnenfuhrer, J., and Lengauer, T. (2006). A new measure for functional similarity of gene products based on Gene Ontology. BMC bioinformatics *7*, 302.

Vogel, C., and Chothia, C. (2006). Protein family expansions and biological complexity. PLoS computational biology *2*, e48.

Weiner, J., 3rd, Beaussart, F., and Bornberg-Bauer, E. (2006). Domain deletions and substitutions in the modular protein evolution. The FEBS journal *273*, 2037-2047.

Yanai, I., Derti, A., and DeLisi, C. (2001). Genes linked by fusion events are generally of the same functional category: a systematic analysis of 30 microbial genomes. Proceedings of the National Academy of Sciences of the United States of America *98*, 7940-7945.

Zhang, J. (2003). Evolution by gene duplication: an update. TRENDS in Ecology and Evolution *18*, 292-298.

3.5 Bibliography

Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A., and Sorek, R. (2006). Transcription-mediated gene fusion in the human genome. Genome research 16, 30-36.

Arguello, J.R., Fan, C., Wang, W., and Long, M. (2007). Origination of Chimeric Genes through DNA-Level Recombination. Genome Dyn 3, 131-146.

Babushok, D.V., Ohshima, K., Ostertag, E.M., Chen, X., Wang, Y., Mandal, P.K., Okada, N., Abrams, C.S., and Kazazian, H.H., Jr. (2007a). A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. Genome research 17, 1129-1138.

Babushok, D.V., Ostertag, E.M., and Kazazian, H.H., Jr. (2007b). Current topics in genome evolution: molecular mechanisms of new gene formation. Cell Mol Life Sci 64, 542-554.

Bailey, J.A., and Eichler, E.E. (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. Nature reviews 7, 552-564.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. (2002). Recent segmental duplications in the human genome. Science (New York, NY 297, 1003-1007.

Basu, M.K., Carmel, L., Rogozin, I.B., and Koonin, E.V. (2008). Evolution of protein domain promiscuity in eukaryotes. Genome research 18, 449-461.

Bjorklund, A.K., Ekman, D., Light, S., Frey-Skott, J., and Elofsson, A. (2005). Domain rearrangements in protein evolution. Journal of molecular biology 353, 911-923.

Blomme, T., Vandepoele, K., De Bodt, S., Simillion, C., Maere, S., and Van de Peer, Y. (2006). The gain and loss of genes during 600 million years of vertebrate evolution. Genome biology 7, R43.

Buljan, M., and Bateman, A. (2009). The evolution of protein domain families. Biochemical Society transactions 37, 751-755.

Chothia, C., Gough, J., Vogel, C., and Teichmann, S.A. (2003). Evolution of the protein repertoire. Science (New York, NY 300, 1701-1703.

Conrad, D.F., and Hurles, M.E. (2007). The population genetics of structural variation. Nature genetics 39, S30-36.

Cordaux, R., Udit, S., Batzer, M.A., and Feschotte, C. (2006). Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proceedings of the National Academy of Sciences of the United States of America 103, 8101-8106.

Dehal, P., and Boore, J.L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. PLoS biology 3, e314.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome biology 4, P3.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. Journal of molecular biology 347, 827-839.

Ekman, D., Bjorklund, A.K., and Elofsson, A. (2007). Quantification of the elevated rate of domain rearrangements in metazoa. Journal of molecular biology 372, 1337-1348.

Farris, J.S. (1977). Phylogenetic analysis under Dollo's Law. Systematic Zoology 26, 77-88.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L., et al. (2008). The Pfam protein families database. Nucleic Acids Res 36, D281-288.

Fong, J.H., Geer, L.Y., Panchenko, A.R., and Bryant, S.H. (2007). Modeling the evolution of protein domain architectures using maximum parsimony. Journal of molecular biology 366, 307-315.

Forslund, K., Henricson, A., Hollich, V., and Sonnhammer, E.L. (2008). Domain tree-based analysis of protein architecture evolution. Molecular biology and evolution 25, 254-264.

Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E., et al. (2006). Copy number variation: new insights in genome diversity. Genome research 16, 949-961.

Gilbert, W. (1978). Why genes in pieces? Nature 271, 501.

Gsponer, J., and Babu, M.M. (2009). The rules of disorder or why disorder rules. Progress in biophysics and molecular biology 99, 94-103.

Gsponer, J., Futschik, M.E., Teichmann, S.A., and Babu, M.M. (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science (New York, NY 322, 1365-1368.

Gu, W., Zhang, F., and Lupski, J.R. (2008). Mechanisms for human genomic rearrangements. Pathogenetics 1, 4.

Hahn, M.W., Demuth, J.P., and Han, S.G. (2007). Accelerated rate of gene gain and loss in primates. Genetics 177, 1941-1949.

Han, J.S., and Boeke, J.D. (2005). LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? Bioessays 27, 775-784.

Itoh, M., Nacher, J.C., Kuma, K., Goto, S., and Kanehisa, M. (2007). Evolutionary history and functional implications of protein domains and their combinations in eukaryotes. Genome biology 8, R121.

Kaessmann, H., Zollner, S., Nekrutenko, A., and Li, W.H. (2002). Signatures of domain shuffling in the human genome. Genome research 12, 1642-1650.

Kawashima, T., Kawashima, S., Tanaka, C., Murai, M., Yoneda, M., Putnam, N.H., Rokhsar, D.S., Kanehisa, M., Satoh, N., and Wada, H. (2009). Domain shuffling and the evolution of vertebrates. Genome research 19, 1393-1403.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome research 12, 996-1006.

Kim, P.M., Lam, H.Y., Urban, A.E., Korbel, J.O., Affourtit, J., Grubert, F., Chen, X., Weissman, S., Snyder, M., and Gerstein, M.B. (2008). Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. Genome research 18, 1865-1874.

Knowles, D.G., and McLysaght, A. (2009). Recent de novo origin of human protein-coding genes. Genome research 19, 1752-1759.

Kummerfeld, S.K., and Teichmann, S.A. (2005). Relative rates of gene fusion and fission in multi-domain proteins. Trends Genet 21, 25-30.

Liu, M., and Grigoriev, A. (2004). Protein domains correlate strongly with exons in multiple eukaryotic genomes--evidence of exon shuffling? Trends Genet 20, 399-403.

Liu, M., Walch, H., Wu, S., and Grigoriev, A. (2005). Significant expansion of exon-bordering protein domains during animal proteome evolution. Nucleic acids research 33, 95-105.

Long, M. (2001). Evolution of novel genes. Curr Opin Genet Dev 11, 673-680.

Long, M., Betran, E., Thornton, K., and Wang, W. (2003). The origin of new genes: glimpses from the young and old. Nature reviews 4, 865-875.

Long, M., Rosenberg, C., and Gilbert, W. (1995). Intron phase correlations and the evolution of the intron/exon structure of genes. Proceedings of the National Academy of Sciences of the United States of America 92, 12495-12499.

Lynch, M. (2002). Intron evolution as a population-genetic process. Proceedings of the National Academy of Sciences of the United States of America 99, 6118-6123.

Madera, M. (2008). Profile Comparer: a program for scoring and aligning profile hidden Markov models. Bioinformatics 24, 2630-2631.

Magrangeas, F., Pitiot, G., Dubois, S., Bragado-Nilsson, E., Cherel, M., Jobert, S., Lebeau, B., Boisteau, O., Lethe, B., Mallet, J., et al. (1998). Cotranscription and intergenic splicing of human galactose-1-phosphate uridylyltransferase and interleukin-11 receptor alpha-chain genes generate a fusion mRNA in normal cells. Implication for the production of multidomain proteins during evolution. The Journal of biological chemistry 273, 16005-16010.

Milinkovitch, M.C., Helaers, R., and Tzika, A.C. (2010). Historical constraints on vertebrate genome evolution. Genome Biol Evol 2010, 13-18.

Nurminsky, D.I., Nurminskaya, M.V., De Aguiar, D., and Hartl, D.L. (1998). Selective sweep of a newly evolved sperm-specific gene in Drosophila. Nature 396, 572-575.

Parra, G., Reymond, A., Dabbouseh, N., Dermitzakis, E.T., Castelo, R., Thomson, T.M., Antonarakis, S.E., and Guigo, R. (2006). Tandem chimerism as a means to increase protein complexity in the human genome. Genome research 16, 37-44.

Patthy, L. (1996). Exon shuffling and other ways of module exchange. Matrix Biol 15, 301-310; discussion 311-302.

Patthy, L. (1999). Genome evolution and the evolution of exon-shuffling--a review. Gene 238, 103-114.

Patthy, L. (2008). Exons and Protein Modules. In Encyclopedia of life sciences (John Wiley & Sons, Ltd.).

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science (New York, NY 300, 445-452.

Peisajovich, S.G., Garbarino, J.E., Wei, P., and Lim, W.A. (2010). Rapid diversification of cell signaling phenotypes by modular domain recombination. Science (New York, NY 328, 368-372.

Ponting, C.P. (2008). The functional repertoires of metazoan genomes. Nature reviews 9, 689-698.

Przytycka, T., Davis, G., Song, N., and Durand, D. (2006). Graph theoretical insights into evolution of multidomain proteins. J Comput Biol 13, 351-363.

Roth, D.B., Porter, T.N., and Wilson, J.H. (1985). Mechanisms of nonhomologous recombination in mammalian cells. Molecular and cellular biology 5, 2599-2607.

Ruan, J., Li, H., Chen, Z., Coghlan, A., Coin, L.J., Guo, Y., Heriche, J.K., Hu, Y., Kristiansen, K., Li, R., et al. (2008). TreeFam: 2008 Update. Nucleic Acids Res 36, D735-740.

Ryan, T.J., Emes, R.D., Grant, S.G., and Komiyama, N.H. (2008). Evolution of NMDA receptor cytoplasmic interaction domains: implications for organisation of synaptic signalling complexes. BMC neuroscience 9, 6.

Sankoff, D., Cedergren, R.J., and McKay, W. (1982). A strategy for sequence phylogeny research. Nucleic Acids Res 10, 421-431.

Stankiewicz, P., and Lupski, J.R. (2002). Genome architecture, rearrangements and genomic disorders. Trends Genet 18, 74-82.

Thomson, T.M., Lozano, J.J., Loukili, N., Carrio, R., Serras, F., Cormand, B., Valeri, M., Diaz, V.M., Abril, J., Burset, M., et al. (2000). Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene. Genome Res 10, 1743-1756.

Turner, D.J., Miretti, M., Rajan, D., Fiegler, H., Carter, N.P., Blayney, M.L., Beck, S., and Hurles, M.E. (2008). Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. Nat Genet 40, 90-95.

Tzika, A.C., Helaers, R., Van de Peer, Y., and Milinkovitch, M.C. (2008). MANTIS: a phylogenetic framework for multi-species genome comparisons. Bioinformatics (Oxford, England) 24, 151-157.

van Rijk, A., and Bloemendal, H. (2003). Molecular mechanisms of exon shuffling: illegitimate recombination. Genetica 118, 245-249.

Vogel, C., and Chothia, C. (2006). Protein family expansions and biological complexity. PLoS computational biology 2, e48.

Weiner, J., 3rd, Beaussart, F., and Bornberg-Bauer, E. (2006). Domain deletions and substitutions in the modular protein evolution. FEBS J 273, 2037-2047.

Wilming, L.G., Gilbert, J.G., Howe, K., Trevanion, S., Hubbard, T., and Harrow, J.L. (2008). The vertebrate genome annotation (Vega) database. Nucleic acids research 36, D753-760.

Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. Journal of molecular biology 293, 321-331.

Yang, S., Arguello, J.R., Li, X., Ding, Y., Zhou, Q., Chen, Y., Zhang, Y., Zhao, R., Brunet, F., Peng, L., et al. (2008). Repetitive element-mediated recombination as a mechanism for new gene origination in Drosophila. PLoS genetics 4, e3.

Zhang, F., Khajavi, M., Connolly, A.M., Towne, C.F., Batish, S.D., and Lupski, J.R. (2009). The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nature genetics 41, 849-853.

Zhang, X.H., and Chasin, L.A. (2006). Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. Proceedings of the National Academy of Sciences of the United States of America 103, 13427-13432.

Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and Wang, W. (2008). On the origin of new genes in Drosophila. Genome research 18, 1446-1455.

4.5 Bibliography

Bell, S., Klein, C., Muller, L., Hansen, S., and Buchner, J. (2002). p53 contains large unstructured regions in its native state. J Mol Biol *322*, 917-927.

Bochkareva, E., Kaustov, L., Ayed, A., Yi, G.S., Lu, Y., Pineda-Lucena, A., Liao, J.C., Okorokov, A.L., Milner, J., Arrowsmith, C.H.*, et al.* (2005). Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. Proc Natl Acad Sci U S A *102*, 15412-15417.

Bourdon, J.C. (2007). p53 and its isoforms in cancer. Br J Cancer *97*, 277-282.

Brown, C.J., Takayama, S., Campen, A.M., Vise, P., Marshall, T.W., Oldfield, C.J., Williams, C.J., and Dunker, A.K. (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol *55*, 104-110.

Bult, C.J., Eppig, J.T., Kadin, J.A., Richardson, J.E., and Blake, J.A. (2008). The Mouse Genome Database (MGD): mouse biology and model systems. Nucleic Acids Res *36*, D724-728.

Chen, J.W., Romero, P., Uversky, V.N., and Dunker, A.K. (2006). Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. J Proteome Res *5*, 888-898.

Chumakov, P.M. (2007). Versatile functions of p53 protein in multicellular organisms. Biochemistry (Mosc) *72*, 1399-1421.

Consortium, U. (2009). The Universal Protein Resource (UniProt) 2009. Nucleic Acids Res *37*, D169-174.

Dawson, R., Muller, L., Dehner, A., Klein, C., Kessler, H., and Buchner, J. (2003). The N-terminal domain of p53 is natively unfolded. J Mol Biol *332*, 1131-1141.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Di Lello, P., Jenkins, L.M., Jones, T.N., Nguyen, B.D., Hara, T., Yamaguchi, H., Dikeakos, J.D., Appella, E., Legault, P., and Omichinski, J.G. (2006). Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. Mol Cell *22*, 731-740.

Dosztanyi, Z., Chen, J., Dunker, A.K., Simon, I., and Tompa, P. (2006). Disorder and sequence repeats in hub proteins and their implications for network evolution. J Proteome Res *5*, 2985-2995.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433-3434.

Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M., and Uversky, V.N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J *272*, 5129-5148.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol *6*, 197-208.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L.*, et al.* (2008). The Pfam protein families database. Nucleic Acids Res *36*, D281-288.

Forbes, S.A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J.W., Futreal, P.A., and Stratton, M.R. (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet *Chapter 10*, Unit 10 11.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat Rev Cancer *4*, 177-183.

Fuxreiter, M., Tompa, P., and Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. Bioinformatics *23*, 950-956.

Garcia, J., Gerber, S.H., Sugita, S., Sudhof, T.C., and Rizo, J. (2004). A conformational switch in the Piccolo C2A domain regulated by alternative splicing. Nat Struct Mol Biol *11*, 45-53.

Gnad, F., Ren, S., Cox, J., Olsen, J.V., Macek, B., Oroshi, M., and Mann, M. (2007). PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. Genome Biol *8*, R250.

Gsponer, J., and Babu, M.M. (2009). The rules of disorder or why disorder rules. Prog Biophys Mol Biol *99*, 94-103.

Gsponer, J., Futschik, M.E., Teichmann, S.A., and Babu, M.M. (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science *322*, 1365-1368.

Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol *2*, e100.

Holt, L.J., Tuch, B.B., Villen, J., Johnson, A.D., Gygi, S.P., and Morgan, D.O. (2009). Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. Science *325*, 1682-1686.

Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C., and Lempicki, R.A. (2003). Identifying biological themes within lists of genes with EASE. Genome Biol *4*, R70.

Hubbard, T.J., Aken, B.L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L.*, et al.* (2009). Ensembl 2009. Nucleic Acids Res *37*, D690-697.

Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z., and Dunker, A.K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol *323*, 573-584.

Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z., and Dunker, A.K. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res *32*, 1037-1049.

Ishimoto, O., Kawahara, C., Enjo, K., Obinata, M., Nukiwa, T., and Ikawa, S. (2002). Possible oncogenic potential of DeltaNp73: a newly identified isoform of human p73. Cancer Res *62*, 636-641.

Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R., Au-Young, J., and Stuve, L.L. (2004). PCR isolation and cloning of novel splice variant mRNAs from known drug target genes. Genomics *83*, 566-571.

Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet *11*, 345-355.

Kim, E., and Sheng, M. (2004). PDZ domain proteins of synapses. Nat Rev Neurosci *5*, 771-781.

Kondrashov, F.A., and Koonin, E.V. (2003). Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. Trends Genet *19*, 115-119.

Koscielny, G., Le Texier, V., Gopalakrishnan, C., Kumanduri, V., Riethoven, J.J., Nardone, F., Stanley, E., Fallsehr, C., Hofmann, O., Kull, M.*, et al.* (2009).

ASTD: The Alternative Splicing and Transcript Diversity database. Genomics *93*, 213-220.

Kriventseva, E.V., Koch, I., Apweiler, R., Vingron, M., Bork, P., Gelfand, M.S., and Sunyaev, S. (2003). Increase of functional diversity by alternative splicing. Trends Genet *19*, 124-128.

Kussie, P.H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J., and Pavletich, N.P. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science *274*, 948-953.

Lareau, L.F., Green, R.E., Bhatnagar, R.S., and Brenner, S.E. (2004). The evolving roles of alternative splicing. Curr Opin Struct Biol *14*, 273-282.

Lev-Maor, G., Goren, A., Sela, N., Kim, E., Keren, H., Doron-Faigenboim, A., Leibman-Barak, S., Pupko, T., and Ast, G. (2007). The "alternative" choice of constitutive exons throughout evolution. PLoS Genet *3*, e203.

Lukong, K.E., Chang, K.W., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends Genet *24*, 416-425.

Magidovich, E., Orr, I., Fass, D., Abdu, U., and Yifrach, O. (2007). Intrinsic disorder in the C-terminal domain of the Shaker voltage-activated K+ channel modulates its interaction with scaffold proteins. Proc Natl Acad Sci U S A *104*, 13022-13027.

Meszaros, B., Simon, I., and Dosztanyi, Z. (2009). Prediction of protein binding regions in disordered proteins. PLoS Comput Biol *5*, e1000376.

Modrek, B., and Lee, C.J. (2003). Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet *34*, 177-180.

Nourry, C., Grant, S.G., and Borg, J.P. (2003). PDZ domain proteins: plug and play! Sci STKE *2003*, RE7.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell *127*, 635-648.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet *40*, 1413-1415.

Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., and Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics *7*, 208.

Ranganathan, R., and Ross, E.M. (1997). PDZ domain proteins: scaffolds for signaling complexes. Curr Biol *7*, R770-773.

Resch, A., Xing, Y., Modrek, B., Gorlick, M., Riley, R., and Lee, C. (2004). Assessing the impact of alternative splicing on domain interactions in the human proteome. J Proteome Res *3*, 76-83.

Romero, P.R., Zaidi, S., Fang, Y.Y., Uversky, V.N., Radivojac, P., Oldfield, C.J., Cortese, M.S., Sickmeier, M., LeGall, T., Obradovic, Z.*, et al.* (2006). Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc Natl Acad Sci U S A *103*, 8390-8395.

Shaul, Y.D., Gibor, G., Plotnikov, A., and Seger, R. (2009). Specific phosphorylation and activation of ERK1c by MEK1b: a unique route in the ERK cascade. Genes Dev *23*, 1779-1790.

Shaul, Y.D., and Seger, R. (2007). The MEK/ERK cascade: from signaling specificity to diverse functions. Biochim Biophys Acta *1773*, 1213-1226.

Shimizu, K., and Toh, H. (2009). Interaction between intrinsically disordered proteins frequently occurs in a human protein-protein interaction network. J Mol Biol *392*, 1253-1265.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A., and Soreq, H. (2005). Function of alternative splicing. Gene *344*, 1-20.

Taylor, J., Schenck, I., Blankenberg, D., and Nekrutenko, A. (2007). Using galaxy to perform large-scale interactive data analyses. Curr Protoc Bioinformatics *Chapter 10*, Unit 10 15.

Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett *579*, 3346-3354.

Tompa, P., Szasz, C., and Buday, L. (2005). Structural disorder throws new light on moonlighting. Trends Biochem Sci *30*, 484-489.

Tress, M.L., Bodenmiller, B., Aebersold, R., and Valencia, A. (2008). Proteomics studies confirm the presence of alternative protein isoforms on a large scale. Genome Biol *9*, R162.

Tress, M.L., Martelli, P.L., Frankish, A., Reeves, G.A., Wesselink, J.J., Yeats, C., Olason, P.I., Albrecht, M., Hegyi, H., Giorgetti, A.*, et al.* (2007). The

implications of alternative splicing in the ENCODE protein complement. Proc Natl Acad Sci U S A *104*, 5495-5500.

Venables, J.P. (2004). Aberrant and alternative splicing in cancer. Cancer Res *64*, 7647-7654.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.

Wang, K., Geren, L., Zhen, Y., Ma, L., Ferguson-Miller, S., Durham, B., and Millett, F. (2002). Mutants of the CuA site in cytochrome c oxidase of Rhodobacter sphaeroides: II. Rapid kinetic analysis of electron transfer. Biochemistry *41*, 2298-2304.

Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol *293*, 321-331.

Wu, G., Feng, X., and Stein, L. (2010). A human functional protein interaction network and its application to cancer data analysis. Genome Biol *11*, R53.

Xing, Y., and Lee, C.J. (2005). Protein modularity of alternatively spliced exons is associated with tissue-specific regulation of alternative splicing. PLoS Genet *1*, e34.

Yura, K., Shionyu, M., Hagino, K., Hijikata, A., Hirashima, Y., Nakahara, T., Eguchi, T., Shinoda, K., Yamaguchi, A., Takahashi, K.*, et al.* (2006). Alternative splicing in human transcriptome: functional and structural influence on proteins. Gene *380*, 63-71.