

BIBLIOGRAPHY

- Adema, G.J., Bovenberg, R.A., Jansz, H.S. and Baas, P.D. (1988) Unusual branch point selection involved in splicing of the alternatively processed Calcitonin/CGRP-I pre-mRNA. *Nucleic Acids Research*, **16**, 9513-9526.
- Agarwal, P. and Bafna, V. (1998) The ribosome scanning model for translation initiation: implications for gene prediction and full-length cDNA detection. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, **6**, 2-7.
- 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., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.
- Apic, G., Gough, J. and Teichmann, S.A. (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol*, **310**, 311-325.
- Aranda, A. and Proudfoot, N. (2001) Transcriptional termination factors for RNA polymerase II in yeast. *Molecular Cell*, **7**, 1003-1011.
- Aranda, A. and Proudfoot, N.J. (1999) Definition of transcriptional pause elements in fission yeast. *Molecular and Cellular Biology*, **19**, 1251-1261.
- Ashfield, R., Enriquez-Harris, P. and Proudfoot, N.J. (1991) Transcriptional termination between the closely linked human complement genes C2 and factor B: common termination factor for C2 and c-myc? *EMBO Journal*, **10**, 4197-4207.
- Ashurst, J. (2002) <http://vega.sanger.ac.uk>
- Ashurst, J.L. and Collins, J.E. (2003) Gene Annotation: Prediction and Testing. *Annual Review of Genomics and Human Genetics*, **4**, 69-88.

- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C. and Eddy, S.R. (2004) The Pfam protein families database. *Nucleic Acids Research*, **32**, D138-141.
- Beaumont, C., Porcher, C., Picat, C., Nordmann, Y. and Grandchamp, B. (1989) The mouse porphobilinogen deaminase gene. Structural organization, sequence, and transcriptional analysis. *Journal of Biological Chemistry*, **264**, 14829-14834.
- Bengal, E., Flores, O., Krauskopf, A., Reinberg, D. and Aloni, Y. (1991) Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. **11**, 1195-1206.
- Bentley, D.L. and Groudine, M. (1986) A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. *Nature*, **321**, 702-706.
- Berget, S.M. (1995) Exon recognition in vertebrate splicing. *Journal of Biological Chemistry*, **270**, 2411-2414.
- Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D. and Zardecki, C. (2002) The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr*, **58**, 899-907.
- Betton, J.M., Jacob, J.P., Hofnung, M. and Broome-Smith, J.K. (1997) Creating a bifunctional protein by insertion of beta-lactamase into the maltodextrin-binding protein. *Nat Biotechnol*, **15**, 1276-1279.
- Birney, E., Andrews, D., Bevan, P., Caccamo, M., Cameron, G., Chen, Y., Clarke, L., Coates, G., Cox, T., Cuff, J., Curwen, V., Cutts, T., Down, T., Durbin, R., Eyraes, E., Fernandez-Suarez, X.M., Gane, P., Gibbins, B., Gilbert, J., Hammond, M., *et al.* (2004) Ensembl 2004. *Nucleic Acids Research*, **32**, D468-470.

- Birney, E. and Durbin, R. (1997) Dynamite: a flexible code generating language for dynamic programming methods used in sequence comparison. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, **5**, 56-64.
- Birse, C.E., Lee, B.A., Hansen, K. and Proudfoot, N.J. (1997) Transcriptional termination signals for RNA polymerase II in fission yeast. *EMBO Journal*, **16**, 3633-3643.
- Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W. and Proudfoot, N.J. (1998) Coupling termination of transcription to messenger RNA maturation in yeast. *Science*, **280**, 298-301.
- Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, **72**, 291-336.
- Bogenhagen, D.F. and Brown, D.D. (1981) Nucleotide sequences in *Xenopus* 5S DNA required for transcription termination. *Cell*, **24**, 261-270.
- Bork, P., Downing, A.K., Kieffer, B. and Campbell, I.D. (1996) Structure and distribution of modules in extracellular proteins. *Q Rev Biophys*, **29**, 119-167.
- Bossemeyer, D. (1994) The glycine-rich sequence of protein kinases: a multifunctional element. *Trends Biochem Sci*, **19**, 201-205.
- Braddock, M., Muckenthaler, M., White, M.R., Thorburn, A.M., Sommerville, J., Kingsman, A.J. and Kingsman, S.M. (1994) Intron-less RNA injected into the nucleus of *Xenopus* oocytes accesses a regulated translation control pathway. *Nucleic Acids Research*, **22**, 5255-5264.
- Brendel, V. and Kleffe, J. (1998) Prediction of locally optimal splice sites in plant pre-mRNA with applications to gene identification in *Arabidopsis thaliana* genomic DNA. *Nucleic Acids Research*, **26**, 4748-4757.

- Brenner, S.E., Chothia, C. and Hubbard, T.J. (1998) Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships. *Proc Natl Acad Sci U S A*, **95**, 6073-6078.
- Briggs, D., Jackson, D., Whitelaw, E. and Proudfoot, N.J. (1989) Direct demonstration of termination signals for RNA polymerase II from the sea urchin H2A histone gene. *Nucleic Acids Research*, **17**, 8061-8071.
- Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E. and Palmiter, R.D. (1988) Introns increase transcriptional efficiency in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 836-840.
- Burge, C. (1998) *Modeling dependencies in pre-mRNA splicing signals*. Elsevier, Amsterdam, Netherlands.
- Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, **268**, 78-94.
- Burset, M., Seledtsov, I.A. and Solovyev, V.V. (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Research*, **28**, 4364-4375.
- Burset, M., Seledtsov, I.A. and Solovyev, V.V. (2001) SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Research*, **29**, 255-259.
- Calvo, O. and Manley, J.L. (2001) Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Molecular Cell*, **7**, 1013-1023.
- Calvo, O. and Manley, J.L. (2003) Strange bedfellows: polyadenylation factors at the promoter. *Genes and Development*, **17**, 1321-1327.
- Cavener, D.R. and Ray, S.C. (1991) Eukaryotic start and stop translation sites. *Nucleic Acids Research*, **19**, 3185-3192.

- Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K. and Gingeras, T.R. (2004) Unbiased Mapping of Transcription Factor Binding Sites along Human Chromosomes 21 and 22 Points to Widespread Regulation of Noncoding RNAs. *Cell*, **116**, 499-509.
- Celniker, S.E., Wheeler, D.A., Kronmiller, B., Carlson, J.W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S.P., Frise, E., Hodgson, A., George, R.A., Hoskins, R.A., Lavery, T., Muzny, D.M., Nelson, C.R., Pacleb, J.M., Park, S., Pfeiffer, B.D., Richards, S., *et al.* (2002) Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol*, **3**, RESEARCH0079.
- Chafin, D.R., Claussen, T.J. and Price, D.H. (1991) Identification and purification of a yeast protein that affects elongation by RNA polymerase II. **266**, 9256-9262.
- Chandonia, J.M., Walker, N.S., Lo Conte, L., Koehl, P., Levitt, M. and Brenner, S.E. (2002) ASTRAL compendium enhancements. *Nucleic Acids Res*, **30**, 260-263.
- Chinsky, J.M., Maa, M.C., Ramamurthy, V. and Kellems, R.E. (1989) Adenosine deaminase gene expression. Tissue-dependent regulation of transcriptional elongation. *Journal of Biological Chemistry*, **264**, 14561-14565.
- Chothia, C. (1992) Proteins. One thousand families for the molecular biologist. *Nature*, **357**, 543-544.
- Chothia, C. and Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *Embo J*, **5**, 823-826.
- Collins, J.E., Goward, M.E., Cole, C.G., Smink, L.J., Huckle, E.J., Knowles, S., Bye, J.M., Beare, D.M. and Dunham, I. (2003) Reevaluating human gene annotation: a second-generation analysis of chromosome 22. *Genome Research*, **13**, 27-36.

- Connelly, S. and Manley, J.L. (1989a) A CCAAT box sequence in the adenovirus major late promoter functions as part of an RNA polymerase II termination signal. *Cell*, **57**, 561-571.
- Connelly, S. and Manley, J.L. (1989b) RNA polymerase II transcription termination is mediated specifically by protein binding to a CCAAT box sequence. *Molecular and Cellular Biology*, **9**, 5254-5259.
- Cook, P.R. (1999) The organization of replication and transcription. *Science*, **284**, 1790-1795.
- Cooke, C., Hans, H. and Alwine, J.C. (1999) Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Molecular and Cellular Biology*, **19**, 4971-4979.
- Cosson, B., Couturier, A., Chabelskaya, S., Kiktev, D., Inge-Vechtomov, S., Philippe, M. and Zhouravleva, G. (2002) Poly(A)-Binding Protein Acts in Translation Termination via Eukaryotic Release Factor 3 Interaction and Does Not Influence [PSI⁺] Propagation. *Molecular and Cellular Biology*, **22**, 3301-3315.
- Cozzarelli, N.R., Gerrard, S.P., Schlissel, M., Brown, D.D. and Bogenhagen, D.F. (1983) Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. *Cell*, **34**, 829-835.
- Cramer, P., Caceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E. and Kornblihtt, A.R. (1999) Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Molecular Cell*, **4**, 251-258.
- Cramer, P., Pesce, C.G., Baralle, F.E. and Kornblihtt, A.R. (1997) Functional association between promoter structure and transcript alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 11456-11460.

- Cramer, P., Srebrow, A., Kadener, S., Werbajh, S., de la Mata, M., Melen, G., Nogues, G. and Kornblihtt, A.R. (2001) Coordination between transcription and pre-mRNA processing. *FEBS Letters*, **498**, 179-182.
- Crispino, J.D., Mermoud, J.E., Lamond, A.I. and Sharp, P.A. (1996) Cis-acting elements distinct from the 5' splice site promote U1-independent pre-mRNA splicing. *RNA*, **2**, 664-673.
- Cristianini, N. and Shawe-Taylor, J. (2000) *An introduction to support vector machines*. Cambridge University Press, Cambridge.
- CSC. (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium. *Science*, **282**, 2012-2018.
- Cuello, P., Boyd, D.C., Dye, M.J., Proudfoot, N.J. and Murphy, S. (1999) Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *EMBO Journal*, **18**, 2867-2877.
- Dantoni, J.C., Murthy, K.G., Manley, J.L. and Tora, L. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature*, **389**, 399-402.
- Das, A. (1993) Control of transcription termination by RNA-binding proteins. *Annual Review of Biochemistry*, **62**, 893-930.
- di Bernardo, D., Down, T. and Hubbard, T. (2003) ddbRNA: detection of conserved secondary structures in multiple alignments. *Bioinformatics*, **19**, 1606-1611.
- Dichtl, B., Blank, D., Ohnacker, M., Friedlein, A., Roeder, D., Langen, H. and Keller, W. (2002a) A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Molecular Cell*, **10**, 1139-1150.
- Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S. and Keller, W. (2002b) Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO Journal*, **21**, 4125-4135.

- Ditzel, L., Lowe, J., Stock, D., Stetter, K.O., Huber, H., Huber, R. and Steinbacher, S. (1998) Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell*, **93**, 125-138.
- Doi, N., Itaya, M., Yomo, T., Tokura, S. and Yanagawa, H. (1997) Insertion of foreign random sequences of 120 amino acid residues into an active enzyme. *FEBS Lett*, **402**, 177-180.
- Dong, S. and Searls, D.B. (1994) Gene structure prediction by linguistic methods. *Genomics*, **23**, 540-551.
- Dowell, R.D., Jokerst, R.M., Day, A., Eddy, S.R. and Stein, L. (2001) The Distributed Annotation System. *BMC Bioinformatics*, **2**, 7.
- Down, T. (2003) Computational localization of promoters and transcription start sites in mammalian genomes. *The Wellcome Trust Sanger Institute*. University of Cambridge, Cambridge, p. 149.
- Down, T. and Hubbard, T. (2004) Relevance Vector Machines for classifying points and regions in biological sequences.
- Down, T.A. and Hubbard, T.J. (2002) Computational detection and location of transcription start sites in mammalian genomic DNA. *Genome Research*, **12**, 458-461.
- Dunham, I., Shimizu, N., Roe, B.A., Chissole, S., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J., Ainscough, R., Almeida, J.P., Babbage, A., Bagguley, C., Bailey, J., Barlow, K., Bates, K.N., Beasley, O., Bird, C.P., Blakey, S., *et al.* (1999) The DNA sequence of human chromosome 22. *Nature*, **402**, 489-495.
- Duplay, P., Szmecman, S., Bedouelle, H. and Hofnung, M. (1987) Silent and functional changes in the periplasmic maltose-binding protein of *Escherichia coli* K12. I. Transport of maltose. *Journal of Molecular Biology*, **194**, 663-673.

- Durbin, R., Eddy, S.R., Krogh, A. and Mitchison, G. (1998) *Biological Sequence Analysis*. Cambridge University Press.
- Dye, M.J. and Proudfoot, N.J. (2001) Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell*, **105**, 669-681.
- Eddy, S.R. (1996) Hidden Markov models. *Curr Opin Struct Biol*, **6**, 361-365.
- Edwards-Gilbert, G., Prescott, J. and Falck-Pedersen, E. (1993) 3' RNA processing efficiency plays a primary role in generating termination-competent RNA polymerase II elongation complexes. *Molecular and Cellular Biology*, **13**, 3472-3480.
- Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*, **30**, 1575-1584.
- Enriquez-Harris, P., Levitt, N., Briggs, D. and Proudfoot, N.J. (1991) A pause site for RNA polymerase II is associated with termination of transcription. *EMBO Journal*, **10**, 1833-1842.
- Evers, R., Smid, A., Rudloff, U., Lottspeich, F. and Grummt, I. (1995) Different domains of the murine RNA polymerase I-specific termination factor mTTF-I serve distinct functions in transcription termination. *EMBO Journal*, **14**, 1248-1256.
- Ewing, B. and Green, P. (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nature Genetics*, **25**, 232-234.
- Friedman, D.I. and Court, D.L. (1995) Transcription antitermination: the lambda paradigm updated. *Molecular Microbiology*, **18**, 191-200.
- Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C. and Faye, G. (2003) Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *EMBO Journal*, **22**, 1588-1598.

- Gautheret, D. and Lambert, A. (2001) Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *Journal of Molecular Biology*, **313**, 1003-1011.
- Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141-147.
- Ge, H. and Roeder, R.G. (1994) Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell*, **78**, 513-523.
- Gene Ontology Consortium. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, **32**, D258-261.
- Gerber, J.K., Gogel, E., Berger, C., Wallisch, M., Muller, F., Grummt, I. and Grummt, F. (1997) Termination of mammalian rDNA replication: polar arrest of replication fork movement by transcription termination factor TTF-I. *Cell*, **90**, 559-567.
- Gerstein, M. (1997) A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure. *Journal of Molecular Biology*, **274**, 562-576.
- Gibson, W.T. and Dormor, D.J. (2003) Searching for the 'natural': the case for the gene 'for' homosexuality. *Journal of Human Reproduction and Genetic Ethics*, **9**, 30-35.
- Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6**, 1197-1211.
- Gray, N.K. and Wickens, M. (1998) Control of translation initiation in animals. *Annual Review of Cell and Developmental Biology*, **14**, 399-458.

- Greenblatt, J., Nodwell, J.R. and Mason, S.W. (1993) Transcriptional antitermination. *Nature*, **364**, 401-406.
- Greger, I.H., Demarchi, F., Giacca, M. and Proudfoot, N.J. (1998) Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. *Nucleic Acids Research*, **26**, 1294-1301.
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. and Eddy, S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Research*, **31**, 439-441.
- Grundy, W.N., Bailey, T.L. and Elkan, C.P. (1997) Meta-MEME: Motif-based Hidden Markov Models of Protein Families. *Computer Applications in the Biosciences*, **13**, 397-406.
- Guigo, R. (1997) Computational gene identification. *Journal of Molecular Medicine*, **75**, 389-393.
- Guigo, R., Agarwal, P., Abril, J.F., Burset, M. and Fickett, J.W. (2000) An assessment of gene prediction accuracy in large DNA sequences. *Genome Research*, **10**, 1631-1642.
- Guigo, R., Knudsen, S., Drake, N. and Smith, T. (1992) Prediction of gene structure. *J Mol Biol*, **226**, 141-157.
- Hager, C., Chen, G., Farmer, A., Huang, W., Inman, J., Kiphart, D., Schilkey, F., Skupski, M.P. and Weller, J. (2000) The Genome Sequence DataBase. *Nucleic Acids Research*, **28**, 31-32.
- Hartmuth, K. and Barta, A. (1988) Unusual branch point selection in processing of human growth hormone pre-mRNA. *Molecular and Cellular Biology*, **8**, 2011-2020.
- Hastings, M.L. and Krainer, A.R. (2001) Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *RNA*, **7**, 471-482.

- Hatzigeorgiou, A.G. (2002) Translation initiation start prediction in human cDNAs with high accuracy. *Bioinformatics*, **18**, 343-350.
- He, X., Khan, A.U., Cheng, H., Pappas, D.L., Jr., Hampsey, M. and Moore, C.L. (2003) Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes and Development*, **17**, 1030-1042.
- Hegyí, H. and Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J Mol Biol*, **288**, 147-164.
- Heinemann, M. and Wagner, R. (1997) Guanosine 3',5'-bis(diphosphate) (ppGpp)-dependent inhibition of transcription from stringently controlled *Escherichia coli* promoters can be explained by an altered initiation pathway that traps RNA polymerase. *European Journal of Biochemistry*, **247**, 990-999.
- Henkin, T.M. (1996) Control of transcription termination in prokaryotes. *Annual Review of Genetics*, **30**, 35-57.
- Hentze, M.W. (2001) Protein synthesis. Believe it or not-translation in the nucleus. *Science*, **293**, 1058-1059.
- Hernandez, N. (1992) *Transcription of vertebrate snRNA genes and related genes*. Cold Spring Harbor Laboratory, New York.
- Hillman, R.T., Green, R. and Brenner, S. (2004) An unappreciated role for RNA surveillance. *Genome Biology*, **5**, R8.
- Hobohm, U. and Sander, C. (1994) Enlarged representative set of protein structures. *Protein Sci*, **3**, 522-524.
- Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol*, **233**, 123-138.

- Holm, L. and Sander, C. (1996) Mapping the protein universe. *Science*, **273**, 595-603.
- Houdebine, L.M. and Attal, J. (1999) Internal ribosome entry sites (IRESs): reality and use. *Transgenic Research*, **8**, 157-177.
- Howe, K.L. (2003) Gene Prediction using a configurable system for the integration of data by dynamic programming. *The Wellcome Trust Sanger Institute*. University of Cambridge, Cambridge, p. 209.
- Howe, K.L., Chothia, T. and Durbin, R. (2002) GAZE: a generic framework for the integration of gene-prediction data by dynamic programming. *Genome Research*, **12**, 1418-1427.
- Hua, S., Guo, T., Gough, J. and Sun, Z. (2002) Proteins with class alpha/beta fold have high-level participation in fusion events. *J Mol Biol*, **320**, 713-719.
- Hubbard, T. (1994) Measuring distance between structural aligned residues. Cambridge. *Personal Communication*.
- Iborra, F.J., Escargueil, A.E., Kwek, K.Y., Akoulitchiev, A. and Cook, P.R. (2004) Molecular cross-talk between the transcription, translation, and nonsense-mediated decay machineries. *Journal of Cell Science*, **117**, 899-906.
- Iborra, F.J., Jackson, D.A. and Cook, P.R. (2001) Coupled transcription and translation within nuclei of mammalian cells. *Science*, **293**, 1139-1142.
- Ichiiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K. and Morikawa, K. (2000) Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. *J Mol Biol*, **300**, 889-901.
- Iida, Y. and Kanagu, D. (2000) Quantification analysis of translation initiation signal in vertebrate mRNAs: effect of nucleotides at positions +4(-)+6 upon efficiency of translation initiation. *Nucleic Acids Symposium Series*, 77-78.

- Johnson, M.R., Norman, C., Reeve, M.A., Scully, J. and Proudfoot, N.J. (1986) Tripartite sequences within and 3' to the sea urchin H2A histone gene display properties associated with a transcriptional termination process. *Molecular and Cellular Biology*, **6**, 4008-4018.
- Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Molecular Cell*, **12**, 5-14.
- Jurica, M.S. and Stoddard, B.L. (1999) Homing endonucleases: structure, function and evolution. *Cell Mol Life Sci*, **55**, 1304-1326.
- Kisselev, L.L. and Frolova, L. (1995) Termination of translation in eukaryotes. *Biochemistry and Cell Biology*, **73**, 1079-1086.
- Kiyosawa, H., Yamanaka, I., Osato, N., Kondo, S. and Hayashizaki, Y. (2003) Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome Research*, **13**, 1324-1334.
- Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, **15**, 8125-8132.
- Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene*, **234**, 187-208.
- Kozak, M. (2001) New ways of initiating translation in eukaryotes? *Molecular and Cellular Biology*, **21**, 1899-1907.
- Kozak, M. (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, **299**, 1-34.
- Krogh, A., Brown, M., Mian, I.S., Sjolander, K. and Haussler, D. (1994) Hidden Markov models in computational biology. Applications to protein modeling. *J Mol Biol*, **235**, 1501-1531.

- Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N.J. and Akoulitchiev, A. (2002) U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nature Structural Biology*, **9**, 800-805.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.
- Lang, W.H., Morrow, B.E., Ju, Q., Warner, J.R. and Reeder, R.H. (1994) A model for transcription termination by RNA polymerase I. *Cell*, **79**, 527-534.
- Lang, W.H. and Reeder, R.H. (1993) The REB1 site is an essential component of a terminator for RNA polymerase I in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, **13**, 649-658.
- Lanzotti, D.J., Kaygun, H., Yang, X., Duronio, R.J. and Marzluff, W.F. (2002) Developmental control of histone mRNA and dSLBP synthesis during *Drosophila* embryogenesis and the role of dSLBP in histone mRNA 3' end processing in vivo. *Molecular and Cellular Biology*, **22**, 2267-2282.
- Le Hir, H., Nott, A. and Moore, M.J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends in Biochemical Sciences*, **28**, 215-220.
- Legendre, M. and Gautheret, D. (2003) Sequence determinants in human polyadenylation site selection. *BMC Genomics*, **4**, 7.
- Levine, A. (2001a) Bioinformatics Approaches to RNA splicing. *The Wellcome Trust Sanger Institute*. University of Cambridge, Cambridge, p. 74.
- Levine, A. (2001b) <http://www.sanger.ac.uk/Software/analysis/stratasplice/>

- Lewis, B.P., Green, R.E. and Brenner, S.E. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 189-192.
- Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S.L. and Quackenbush, J. (2000) Gene index analysis of the human genome estimates approximately 120,000 genes. *Nature Genetics*, **25**, 239-240.
- Lindahl, E. and Elofsson, A. (2000) Identification of related proteins on family, superfamily and fold level. *J Mol Biol*, **295**, 613-625.
- Lin-Marq, N. and Clarkson, S.G. (1998) Efficient synthesis, termination and release of RNA polymerase III transcripts in *Xenopus* extracts depleted of La protein. *EMBO Journal*, **17**, 2033-2041.
- Liu, H., Hao, H., Li, J. and Wong, L. (2003) A New Method to Predict Translation Initiation sites. *Proceedings - European Conference on Computational Biology*.
- Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr. and Shenk, T. (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 8306-8310.
- Lois, R., Freeman, L., Villeponteau, B. and Martinson, H.G. (1990) Active beta-globin gene transcription occurs in methylated, DNase I-resistant chromatin of nonerythroid chicken cells. *Molecular and Cellular Biology*, **10**, 16-27.
- Lund, M., Tange, T.O., Dyhr-Mikkelsen, H., Hansen, J. and Kjems, J. (2000) Characterization of human RNA splice signals by iterative functional selection of splice sites. *RNA*, **6**, 528-544.
- MacDonald, C.C. and Redondo, J.L. (2002) Reexamining the polyadenylation signal: were we wrong about AAUAAA? *Molecular and Cellular Endocrinology*, **190**, 1-8.

- Mackay, D.J.C. (2003) *Information Theory, Inference and Learning algorithms*.
Cambridge University Press, Cambridge.
- Mangalam, H. (2002) The Bio* toolkits--a brief overview. *Brief Bioinform*, **3**, 296-302.
- Manley, J.L. (2002) Nuclear coupling: RNA processing reaches back to transcription.
Nature Structural Biology, **9**, 790-791.
- Maraia, R.J., Kenan, D.J. and Keene, J.D. (1994) Eukaryotic transcription termination factor
La mediates transcript release and facilitates reinitiation by RNA polymerase III.
Molecular and Cellular Biology, **14**, 2147-2158.
- Marshall, N.F. and Price, D.H. (1992) Control of formation of two distinct classes of RNA
polymerase II elongation complexes. *Molecular and Cellular Biology*, **12**, 2078-
2090.
- Martens, J.A. (2003) Expression of an intergenic RNA represses transcription of the
adjacent gene. New York. *Personal Communication*.
- Mathe, C., Sagot, M.F., Schiex, T. and Rouze, P. (2002) Current methods of gene
prediction, their strengths and weaknesses. *Nucleic Acids Research*, **30**, 4103-4117.
- Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence
dependence of thermodynamic parameters improves prediction of RNA secondary
structure1, *1. *Journal of Molecular Biology*, **288**, 911-940.
- Matsumoto, K., Wassarman, K.M. and Wolffe, A.P. (1998) Nuclear history of a pre-mRNA
determines the translational activity of cytoplasmic mRNA. *EMBO Journal*, **17**,
2107-2121.
- Mattick, J.S. (1994) Introns: evolution and function. *Current Opinion in Genetics and
Development*, **4**, 823-831.

- Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Reports*, **2**, 986-991.
- McCullagh, P. and Nelder, J.A. (1983) *Generalized Linear Models*. Chapman and Hall, London.
- Mechti, N., Piechaczyk, M., Blanchard, J.M., Jeanteur, P. and Lebleu, B. (1991) Sequence requirements for premature transcription arrest within the first intron of the mouse c-fos gene. *Molecular and Cellular Biology*, **11**, 2832-2841.
- Meinhart, A., Silberzahn, T. and Cramer, P. (2003) The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. *Journal of Biological Chemistry*, **278**, 15917-15921.
- Micklem, G. <http://www.rfcgr.mrc.ac.uk/Software/EMBOSS/Apps/cpgreport.html>
- Morea, V. (2001) Structural analysis of P-loops proteins. Cambridge. *Personal Communication*.
- Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L. and Proudfoot, N.J. (1998) The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes and Development*, **12**, 2522-2534.
- Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Research*, **10**, 459-472.
- Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*, **247**, 536-540.
- Namy, O., Hatin, I. and Rousset, J.P. (2001) Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Reports*, **2**, 787-793.

NCBI. <http://www.ncbi.nlm.nih.gov>

Neugebauer, K.M. (2002) On the importance of being co-transcriptional. *Journal of Cell Science*, **115**, 3865-3871.

Nussinov, R. (1978) Algorithms for loop matching. *Journal of Applied Mathematics*, **35**, 68-92.

Nussinov, R. (1986a) Sequence signals which may be required for efficient formation of mRNA 3' termini. *Nucleic Acids Research*, **14**, 3557-3571.

Nussinov, R. (1986b) TGTG, G clustering and other signals near non-mammalian vertebrate mRNA 3' termini: some implications. *Journal of Biomolecular Structure and Dynamics*, **3**, 1145-1153.

Nussinov, R. (1987) Asymmetry in the distributions of the four nucleotides at mRNA initiation and 3' termini sites: some geometrical implications. *Biochimica et Biophysica Acta*, **908**, 143-149.

Nussinov, R. (1990) Sequence signals in eukaryotic upstream regions. *Critical Reviews in Biochemistry and Molecular Biology*, **25**, 185-224.

Ohrt, M. (2004) <http://www.phpinsider.com/php/code/pmatch/>

Orengo, C.A., Bray, J.E., Buchan, D.W., Harrison, A., Lee, D., Pearl, F.M., Sillitoe, I., Todd, A.E. and Thornton, J.M. (2002) The CATH protein family database: a resource for structural and functional annotation of genomes. *Proteomics*, **2**, 11-21.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. and Thornton, J.M. (1997) CATH--a hierarchic classification of protein domain structures. *Structure*, **5**, 1093-1108.

Orozco, I.J., Kim, S.J. and Martinson, H.G. (2002) The Poly(A) Signal, without the Assistance of Any Downstream Element, Directs RNA Polymerase II to Pause in

- Vivo and Then to Release Stochastically from the Template. *Journal of Biological Chemistry*, **277**, 42899-42911.
- Orphanides, G. and Reinberg, D. (2002) A unified theory of gene expression. *Cell*, **108**, 439-451.
- Osheim, Y.N., Proudfoot, N.J. and Beyer, A.L. (1999) EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Molecular Cell*, **3**, 379-387.
- Osheim, Y.N., Sikes, M.L. and Beyer, A.L. (2002) EM visualization of Pol II genes in *Drosophila*: most genes terminate without prior 3' end cleavage of nascent transcripts. *Chromosoma*, **111**, 1-12.
- Ozawa, Y., Hanaoka, S., Saito, R., Washio, T., Nakano, S., Shinagawa, A., Itoh, M., Shibata, K., Carninci, P. and Konno, H. (2002) Comprehensive sequence analysis of translation termination sites in various eukaryotes. *Gene*, **300**, 79-87.
- Pain, V.M. (1996) Initiation of protein synthesis in eukaryotic cells. *European Journal of Biochemistry*, **236**, 747-771.
- Palm, G.J., Billy, E., Filipowicz, W. and Wlodawer, A. (2000) Crystal structure of RNA 3'-terminal phosphate cyclase, a ubiquitous enzyme with unusual topology. *Structure Fold Des*, **8**, 13-23.
- Pan, Q. and Simpson, R.U. (1999) c-myc intron element-binding proteins are required for 1, 25-dihydroxyvitamin D3 regulation of c-myc during HL-60 cell differentiation and the involvement of HOXB4. *Journal of Biological Chemistry*, **274**, 8437-8444.
- Park, J., Karplus, K., Barrett, C., Hughey, R., Haussler, D., Hubbard, T. and Chothia, C. (1998) Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *J Mol Biol*, **284**, 1201-1210.

- Park, J., Teichmann, S.A., Hubbard, T. and Chothia, C. (1997) Intermediate sequences increase the detection of homology between sequences. *J Mol Biol*, **273**, 349-354.
- Parra, G., Agarwal, P., Abril, J.F., Wiehe, T., Fickett, J.W. and Guigo, R. (2003) Comparative gene prediction in human and mouse. *Genome Research*, **13**, 108-117.
- Pearson, W.R. (1991) Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics*, **11**, 635-650.
- Pearson, W.R. (1995) Comparison of methods for searching protein sequence databases. *Protein Sci*, **4**, 1145-1160.
- Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A*, **85**, 2444-2448.
- Pedersen, A.G. and Nielsen, H. (1997a) <http://www.cbs.dtu.dk/services/NetStart>
- Pedersen, A.G. and Nielsen, H. (1997b) Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, **5**, 226-233.
- Pertea, M. (2001) <http://www.tigr.org/tdb/GeneSplicer/index.shtml>
- Pertea, M., Lin, X. and Salzberg, S.L. (2001) GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Research*, **29**, 1185-1190.
- Pesole, G., Grillo, G. and Liuni, S. (1996) Databases of mRNA untranslated regions for metazoa. *Computers and Chemistry*, **20**, 141-144.
- Pocock, M.R. (2003) Computational Analysis of Genomes. *The Wellcome Trust Sanger Institute*. University of Cambridge, Cambridge, p. 170.

- Pocock, M.R., Down, T. and Hubbard, T. (2000) BioJava: open source components for bioinformatics. *ACM SIGBIO Newsletter*, **20**, 10-12.
- Poland, B.W., Xu, M.Q. and Quioco, F.A. (2000) Structural insights into the protein splicing mechanism of PI-SceI. *Journal of Biological Chemistry*, **275**, 16408-16413.
- Proudfoot, N.J. (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends in Biochemical Sciences*, **14**, 105-110.
- Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501-512.
- Pruitt, K.D. and Maglott, D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Research*, **29**, 137-140.
- Ptitsyn, O.B. (1998) Protein folding: nucleation and compact intermediates. *Biochemistry (Mosc)*, **63**, 367-373.
- Rappsilber, J., Ajuh, P., Lamond, A.I. and Mann, M. (2001) SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome. *Journal of Biological Chemistry*, **276**, 31142-31150.
- Reed, R. (2000) Mechanisms of fidelity in pre-mRNA splicing. *Current Opinion in Cell Biology*, **12**, 340-345.
- Reeder, R.H. and Lang, W.H. (1997) Terminating transcription in eukaryotes: lessons learned from RNA polymerase I. *Trends in Biochemical Sciences*, **22**, 473-477.
- Reese, M.G., Kulp, D., Tammana, H. and Haussler, D. (2000) Genie--gene finding in *Drosophila melanogaster*. *Genome Research*, **10**, 529-538.

- Reines, D. (1994) Transcription: Mechanisms and Regulation. In Conaway, J.W. and Conaway, R.C. (eds.), *Transcription: Mechanisms and Regulation*. Raven Press, New York, pp. 263-278.
- Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H. and Price, D.H. (2001) A highly purified RNA polymerase II elongation control system. *Journal of Biological Chemistry*, **276**, 42601-42609.
- Rivas, E. and Eddy, S.R. (2001) Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*, **2**, 8.
- Rivas, E., Klein, R.J., Jones, T.A. and Eddy, S.R. (2001) Computational identification of noncoding RNAs in E. coli by comparative genomics. *Current Biology*, **11**, 1369-1373.
- Roberts, G.C., Gooding, C., Mak, H.Y., Proudfoot, N.J. and Smith, C.W. (1998) Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Research*, **26**, 5568-5572.
- Roberts, S. and Bentley, D.L. (1992) Distinct modes of transcription read through or terminate at the c-myc attenuator. *EMBO Journal*, **11**, 1085-1093.
- ROC-Curve. <http://gim.unmc.edu/dxtests/ROC1.htm>
- Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L. and Buratowski, S. (2000) Kin28, the TFIIF-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Molecular and Cellular Biology*, **20**, 104-112.
- Rogic, S., Mackworth, A.K. and Ouellette, F.B. (2001) Evaluation of gene-finding programs on mammalian sequences. *Genome Research*, **11**, 817-832.

- Rogozin, I.B., Kochetov, A.V., Kondrashov, F.A., Koonin, E.V. and Milanesi, L. (2001) Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics*, **17**, 890-900.
- Russell, R.B. (1994) Domain insertion. *Protein Eng*, **7**, 1407-1410.
- Salamov, A.A., Nishikawa, T. and Swindells, M.B. (1998a) Assessing protein coding region integrity in cDNA sequencing projects. *Bioinformatics*, **14**, 384-390.
- Salamov, A.A., Nishikawa, T. and Swindells, M.B. (1998b) <http://www.hri.co.jp/atgpr/>
- Salamov, A.A. and Solovyev, V.V. (2000) Ab initio gene finding in Drosophila genomic DNA. *Genome Research*, **10**, 516-522.
- Salamov, A.A., Suwa, M., Orengo, C.A. and Swindells, M.B. (1999) Combining sensitive database searches with multiple intermediates to detect distant homologues. *Protein Eng*, **12**, 95-100.
- Salzberg, S., Delcher, A.L., Fasman, K.H. and Henderson, J. (1998) A decision tree system for finding genes in DNA. *Journal of Computational Biology*, **5**, 667-680.
- Salzberg, S.L. (1997) A method for identifying splice sites and translational start sites in eukaryotic mRNA. *Computer Applications in the Biosciences*, **13**, 365-376.
- Sauder, J.M., Arthur, J.W. and Dunbrack, R.L., Jr. (2000) Large-scale comparison of protein sequence alignment algorithms with structure alignments. *Proteins*, **40**, 6-22.
- Schneider, C., Will, C.L., Makarova, O.V., Makarov, E.M. and Luhrmann, R. (2002) Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Molecular and Cellular Biology*, **22**, 3219-3229.
- Scholkopf, C., Burges, C. and Smola, A.J. (1999) *Advances in kernel methods - Support Vector Learning*. MIT Press, Cambridge.

- Schwer, B. (2001) A new twist on RNA helicases: DExH/D box proteins as RNAPases. *Nature Structural Biology*, **8**, 113-116.
- Sheets, M.D., Ogg, S.C. and Wickens, M.P. (1990) Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Research*, **18**, 5799-5805.
- Shindyalov, I.N. and Bourne, P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng*, **11**, 739-747.
- Slater, G. <http://www.ebi.ac.uk/~guy/exonerate/exonerate.man.1.html>
- Smit, A.F.A. and Green, P. (1996) <http://www.repeatmasker.org>
- Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. *J Mol Biol*, **147**, 195-197.
- Snyder, E.E. and Stormo, G.D. (1995) Identification of protein coding regions in genomic DNA. *Journal of Molecular Biology*, **248**, 1-18.
- Solovyev, V. and Salamov, A. (1997) The Gene-Finder computer tools for analysis of human and model organisms genome sequences. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, **5**, 294-302.
- Solovyev, V.V., Salamov, A.A. and Lawrence, C.B. (1995) Identification of human gene structure using linear discriminant functions and dynamic programming. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, **3**, 367-375.
- Sonenberg, N. and Dever, T.E. (2003) Eukaryotic translation initiation factors and regulators. *Current Opinion in Structural Biology*, **13**, 56-63.
- Spencer, C.A. and Groudine, M. (1990) Transcription elongation and eukaryotic gene regulation. *Oncogene*, **5**, 777-785.

- Staden, R. (1984) Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Research*, **12**, 505-519.
- Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D., *et al.* (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Research*, **12**, 1611-1618.
- Steinmetz, E.J., Conrad, N.K., Brow, D.A. and Corden, J.L. (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature*, **413**, 327-331.
- Sternberg. (2001) Clustering can be used to assign function to unknown protein. Cambridge. *Personal Communication*.
- Sun, Z.W. and Hampsey, M. (1996) Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. *Molecular and Cellular Biology*, **16**, 1557-1566.
- Tabaska, J.E. and Zhang, M.Q. (1999) Detection of polyadenylation signals in human DNA sequences. *Gene*, **231**, 77-86.
- Takagaki, Y. and Manley, J.L. (1998) Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Molecular Cell*, **2**, 761-771.
- Takagaki, Y., Seipelt, R.L., Peterson, M.L. and Manley, J.L. (1996) The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell*, **87**, 941-952.
- Talerico, M. and Berget, S.M. (1994) Intron definition in splicing of small Drosophila introns. *Molecular and Cellular Biology*, **14**, 3434-3445.

- Tantravahi, J., Alvira, M. and Falck-Pedersen, E. (1993) Characterization of the mouse beta maj globin transcription termination region: a spacing sequence is required between the poly(A) signal sequence and multiple downstream termination elements. *Molecular and Cellular Biology*, **13**, 578-587.
- Tarn, W.Y. and Steitz, J.A. (1996) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell*, **84**, 801-811.
- Tarn, W.Y. and Steitz, J.A. (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends in Biochemical Sciences*, **22**, 132-137.
- Teichmann, S.A., Park, J. and Chothia, C. (1998) Structural assignments to the Mycoplasma genitalium proteins show extensive gene duplications and domain rearrangements. *Proc Natl Acad Sci U S A*, **95**, 14658-14663.
- Thanaraj, T.A. (2000) Positional characterisation of false positives from computational prediction of human splice sites. *Nucleic Acids Research*, **28**, 744-754.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.
- Tipping, M.E. (2001a) The Relevance Vector Machine. *Advances in Neural Information Processing Systems*, **12**, 652-658.
- Tipping, M.E. (2001b) Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, **1**, 211-244.
- Uberbacher, E.C., Xu, Y. and Mural, R.J. (1996) Discovering and understanding genes in human DNA sequence using GRAIL. *Methods in Enzymology*, **266**, 259-281.

- Vagner, S., Vagner, C. and Mattaj, I.W. (2000) The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes and Development*, **14**, 403-413.
- Van Agtmael, T., Forrest, S.M., Del-Favero, J., Van Broeckhoven, C. and Williamson, R. (2003) Parametric and nonparametric genome scan analyses for human handedness. *European Journal of Human Genetics*, **11**, 779-783.
- Vapnik, V.N. (1995) *The nature of Statistical Learning Theory*. Springer-Verlag, New York.
- Wagner, R. (2000) *Transcription Regulation in Prokaryotes*. Oxford University Press, Oxford.
- Watson, R.J. (1988) A transcriptional arrest mechanism involved in controlling constitutive levels of mouse c-myc mRNA. *Oncogene*, **2**, 267-272.
- Westhead, D.R., Slidel, T.W., Flores, T.P. and Thornton, J.M. (1999) Protein structural topology: Automated analysis and diagrammatic representation. *Protein Sci*, **8**, 897-904.
- Wetlaufer, D.B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A*, **70**, 697-701.
- Wheeler, D.L., Church, D.M., Edgar, R., Federhen, S., Helmberg, W., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Suzek, T.O., Tatusova, T.A. and Wagner, L. (2004) Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Research*, **32**, D35-40.
- Wiest, D.K., Wang, D. and Hawley, D.K. (1992) Mechanistic studies of transcription arrest at the adenovirus major late attenuation site. Comparison of purified RNA polymerase II and washed elongation complexes. *Journal of Biological Chemistry*, **267**, 7733-7744.

- Will, C.L., Schneider, C., MacMillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Luhrmann, R. and Query, C.C. (2001) A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO Journal*, **20**, 4536-4546.
- Will, C.L., Schneider, C., Reed, R. and Luhrmann, R. (1999) Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science*, **284**, 2003-2005.
- Wilson, C., Hilyer, L. and Green, P. (1990) <http://cgap.nci.nih.gov/Genes/GeneFinder>
- Woychik, N.A. and Hampsey, M. (2002) The RNA polymerase II machinery: structure illuminates function. *Cell*, **108**, 453-463.
- WTSI. <http://www.sanger.ac.uk/Software/GFF>
- Yang, A.S. and Honig, B. (2000) An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. *J Mol Biol*, **301**, 665-678.
- Yeung, G., Choi, L.M., Chao, L.C., Park, N.J., Liu, D., Jamil, A. and Martinson, H.G. (1998) Poly(A)-driven and poly(A)-assisted termination: two different modes of poly(A)-dependent transcription termination. *Molecular and Cellular Biology*, **18**, 276-289.
- Yonaha, M. and Proudfoot, N.J. (1999) Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Molecular Cell*, **3**, 593-600.
- Yonaha, M. and Proudfoot, N.J. (2000) Transcriptional termination and coupled polyadenylation in vitro. *EMBO Journal*, **19**, 3770-3777.
- Yoo, C.J. and Wolin, S.L. (1997) The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. *Cell*, **89**, 393-402.

- Zanier, K., Luyten, I., Crombie, C., Muller, B., Schumperli, D., Linge, J.P., Nilges, M. and Sattler, M. (2002) Structure of the histone mRNA hairpin required for cell cycle regulation of histone gene expression. *RNA*, **8**, 29-46.
- Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L. and Hovorun, D.M. (2003) Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Research*, **31**, 1375-1386.
- Zeng, F., Yap, R.H. and Wong, L. (2002) Using feature generation and feature selection for accurate prediction of translation initiation sites. *Genome Inform Ser Workshop Genome Inform*, **13**, 192-200.
- Zhang, M.Q. (1997) Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 565-568.
- Zhang, M.Q. (2002) Computational prediction of eukaryotic protein-coding genes. *Nature Review Genetics*, **3**, 698-709.
- Zhang, M.Q. and Marr, T.G. (1993) A weight array method for splicing signal analysis. *Computer Applications in the Biosciences*, **9**, 499-509.
- Zhang, R., Evans, G., Rotella, F.J., Westbrook, E.M., Beno, D., Huberman, E., Joachimiak, A. and Collart, F.R. (1999) Characteristics and crystal structure of bacterial inosine-5'-monophosphate dehydrogenase. *Biochemistry*, **38**, 4691-4700.
- Zhang, X., Dong, G. and Wong, L. (2000) Using CAEP to Predict Translation Initiation Sites from Genomic DNA Sequences. *Manuscript*.
- Zhao, J., Hyman, L. and Moore, C. (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews*, **63**, 405-445.

- Zien, A., Ratsch, G., Mika, S., Scholkopf, B., Lengauer, T. and Muller, K.R. (2000) Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics*, **16**, 799-807.
- Zuker, M. (1994) Prediction of RNA secondary structure by energy minimization. *Methods in Molecular Biology*, **25**, 267-294.
- Zuker, M. (2000) Calculating nucleic acid secondary structure. *Current Opinion in Structural Biology*, **10**, 303-310.
- Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, **31**, 3406-3415.
- Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, **9**, 133-148.