6. References

(1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium. Science *282*, 2012-2018.

(2001). Creating the gene ontology resource: design and implementation. Genome Res *11*, 1425-1433.

Afif, H., Allali, N., Couturier, M., and Van Melderen, L. (2001). The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. Mol Microbiol *41*, 73-82.

Aitman, T. J., Glazier, A. M., Wallace, C. A., Cooper, L. D., Norsworthy, P. J., Wahid, F. N., Al-Majali, K. M., Trembling, P. M., Mann, C. J., Shoulders, C. C.*, et al.* (1999). Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. Nat Genet *21*, 76-83.

Akiyama, T., Ohuchi, T., Sumida, S., Matsumoto, K., and Toyoshima, K. (1992). Phosphorylation of the retinoblastoma protein by cdk2. Proc Natl Acad Sci U S A *89*, 7900-7904.

Alaoui-Ismaili, M. H., Lomedico, P. T., and Jindal, S. (2002). Chemical genomics: discovery of disease genes and drugs. Drug Discov Today *7*, 292-294.

Albala, J. S., Franke, K., McConnell, I. R., Pak, K. L., Folta, P. A., Rubinfeld, B., Davies, A. H., Lennon, G. G., and Clark, R. (2000). From genes to proteins: high-throughput expression and purification of the human proteome. J Cell Biochem *80*, 187-191.

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X.*, et al.* (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503-511.

Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K., Lam, Y. W., Steen, H., Mann, M., and Lamond, A. I. (2002). Directed proteomic analysis of the human nucleolus. Curr Biol *12*, 1-11.

Anderson, N. L., and Anderson, N. G. (2002). The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics *1*, 845-867.

Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet *30*, 41-47.

Ashrafi, K., Chang, F. Y., Watts, J. L., Fraser, A. G., Kamath, R. S., Ahringer, J., and Ruvkun, G. (2003). Genome-wide RNAi analysis of Caenorhabditis elegans fat regulatory genes. Nature *421*, 268-272.

Bader, G. D., and Hogue, C. W. (2002). Analyzing yeast protein-protein interaction data obtained from different sources. Nat Biotechnol *20*, 991-997.

Banerjee, N., and Zhang, M. Q. (2002). Functional genomics as applied to mapping transcription regulatory networks. Curr Opin Microbiol *5*, 313-317.

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.

Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A*., et al.* (2000). Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature *406*, 536-540.

Blanco, M. A., Sanchez-Diaz, A., de Prada, J. M., and Moreno, S. (2000). APC(ste9/srw1) promotes degradation of mitotic cyclins in G(1) and is inhibited by cdc2 phosphorylation. Embo J *19*, 3945-3955.

Boulton, S. J., Gartner, A., Reboul, J., Vaglio, P., Dyson, N., Hill, D. E., and Vidal, M. (2002). Combined functional genomic maps of the C. elegans DNA damage response. Science *295*, 127-131.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., and LaBaer, J. (2002). Proteome-scale purification of human proteins from bacteria. Proc Natl Acad Sci U S A *99*, 2654-2659.

Brizuela, L., Braun, P., and LaBaer, J. (2001). FLEXGene repository: from sequenced genomes to gene repositories for high-throughput functional biology and proteomics. Mol Biochem Parasitol *118*, 155-165.

Brown, P. O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. Nat Genet *21*, 33-37.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J Mol Biol *268*, 78-94.

Burley, S. K. (2000). An overview of structural genomics. Nat Struct Biol *7 Suppl*, 932-934.

Bussemaker, H. J., Li, H., and Siggia, E. D. (2001). Regulatory element detection using correlation with expression. Nat Genet *27*, 167-171.

Cash, P. (2002). Proteomics: the protein revolution. Biologist (London) *49*, 58-62.

Chambers, S. P. (2002). High-throughput protein expression for the post-genomic era. Drug Discov Today *7*, 759-765.

Chance, M. R., Bresnick, A. R., Burley, S. K., Jiang, J. S., Lima, C. D., Sali, A., Almo, S. C., Bonanno, J. B., Buglino, J. A., Boulton, S.*, et al.* (2002). Structural genomics: a pipeline for providing structures for the biologist. Protein Sci *11*, 723-738.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J. R., Booth, V., Mackereth, C. D., Saridakis, V., Ekiel, I.*, et al.* (2000). Structural proteomics of an archaeon. Nat Struct Biol *7*, 903-909.

Clark, E. A., Golub, T. R., Lander, E. S., and Hynes, R. O. (2000). Genomic analysis of metastasis reveals an essential role for RhoC. Nature *406*, 532-535.

Claros, M. G., and von Heijne, G. (1994). TopPred II: an improved software for membrane protein structure predictions. Comput Appl Biosci *10*, 685-686.

Cook-Deegan, R. M. (1989). The Alta summit, December 1984. Genomics *5*, 661-663.

Cornelis, G. R., and Van Gijsegem, F. (2000). Assembly and function of type III secretory systems. Annu Rev Microbiol *54*, 735-774.

Costanzo, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Olsen, P., Robertson, L. S., Skrzypek, M. S., Braun, B. R., Hopkins, K. L., Kondu, P.*, et al.* (2001). YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. Nucleic Acids Res *29*, 75-79.

Costanzo, M. C., Hogan, J. D., Cusick, M. E., Davis, B. P., Fancher, A. M., Hodges, P. E., Kondu, P., Lengieza, C., Lew-Smith, J. E., Lingner, C.*, et al.* (2000). The yeast proteome database (YPD) and Caenorhabditis elegans proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. Nucleic Acids Res *28*, 73-76.

Cutler, D. J., Zwick, M. E., Carrasquillo, M. M., Yohn, C. T., Tobin, K. P., Kashuk, C., Mathews, D. J., Shah, N. A., Eichler, E. E., Warrington, J. A., and Chakravarti, A.

(2001). High-throughput variation detection and genotyping using microarrays. Genome Res *11*, 1913-1925.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet *14*, 457-460.

Doyle, S. A., Murphy, M. B., Massi, J. M., and Richardson, P. M. (2002). High-Throughput Proteomics: A Flexible and Efficient Pipeline for Protein Production. Journal for Proteome Research *1*, 531 -536.

Dreger, M. (2003). Emerging strategies in mass-spectrometry based proteomics. Eur J Biochem *270*, 569.

Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P., and Trent, J. M. (1999). Expression profiling using cDNA microarrays. Nat Genet *21*, 10-14.

Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Collins, J. E., Bruskiewich, R., Beare, D. M., Clamp, M., Smink, L. J*., et al.* (1999). The DNA sequence of human chromosome 22. Nature *402*, 489-495.

Eickhoff, H., Birch-Hirschfeld, E., Scheef, J., Hoyer, C., Drexhage, K. H., and Greulich, K. O. (1996). Sequence verification by hybridisation with fluorescent octanucleotides as a first step to a fluorescent sequencing by hybridisation protocol. J Biochem Biophys Methods *32*, 59-68.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A *95*, 14863-14868.

Engelman, D. M., Steitz, T. A., and Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Biophys Chem *15*, 321-353.

Fairhead, C., Thierry, A., Denis, F., Eck, M., and Dujon, B. (1998). 'Mass-murder' of ORFs from three regions of chromosome XI from Saccharomyces cerevisiae. Gene *223*, 33-46.

Fambrough, D., McClure, K., Kazlauskas, A., and Lander, E. S. (1999). Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. Cell *97*, 727-741.

Felleisen, R., Zimmermann, V., Gottstein, B., and Muller, N. (1996). Use of a 96-well format for the affinity purification of maltose- binding protein (MBP) fusion proteins. Biotechniques *20*, 616-620.

Fields, C., Adams, M. D., White, O., and Venter, J. C. (1994). How many genes in the human genome? Nat Genet *7*, 345-346.

Fields, S., Kohara, Y., and Lockhart, D. J. (1999). Functional genomics. Proc Natl Acad Sci U S A *96*, 8825-8826.

Fujibuchi, W., and Kanehisa, M. (1997). Prediction of gene expression specificity by promoter sequence patterns. DNA Res *4*, 81-90.

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M*., et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Ge, H., Liu, Z., Church, G. M., and Vidal, M. (2001). Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. Nat Genet *29*, 482-486.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M*., et al.* (1996). Life with 6000 genes. Science *274*, 546, 563-547.

Gopal, S., Schroeder, M., Pieper, U., Sczyrba, A., Aytekin-Kurban, G., Bekiranov, S., Fajardo, J. E., Eswar, N., Sanchez, R., Sali, A., and Gaasterland, T. (2001). Homology-based annotation yields 1,042 new candidate genes in the Drosophila melanogaster genome. Nat Genet *27*, 337-340.

Guignard, L., Ozawa, K., Pursglove, S. E., Otting, G., and Dixon, N. E. (2002). NMR analysis of in vitro-synthesized proteins without purification: a high-throughput approach. FEBS Lett *524*, 159-162.

Guigo, R., Knudsen, S., Drake, N., and Smith, T. (1992). Prediction of gene structure. J Mol Biol *226*, 141-157.

Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H., and Hard, T. (2002). Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Sci *11*, 313-321.

Hartley, J. L., Temple, G. F., and Brasch, M. A. (2000). DNA cloning using in vitro site-specific recombination. Genome Res *10*, 1788-1795.

He, M., and Taussig, M. J. (2001). Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). Nucleic Acids Res *29*, E73-73.

Hino, M., Shinohara, Y., Kajimoto, K., Terada, H., and Baba, Y. (2002). Requirement of continuous transcription for the synthesis of sufficient amounts of protein by a cell-free rapid translation system. Protein Expr Purif *24*, 255-259.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K*., et al.* (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180-183.

Holden, A. L. (2002). The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome. Biotechniques *Suppl*, 22-24, 26.

Holz, C., Hesse, O., Bolotina, N., Stahl, U., and Lang, C. (2002). A micro-scale process for high-throughput expression of cDNAs in the yeast Saccharomyces cerevisiae. Protein Expr Purif *25*, 372.

Huber, L. A. (2003). Is proteomics heading in the wrong direction? Nat Rev Mol Cell Biol *4*, 74-80.

Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science *292*, 929-934.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A *98*, 4569-4574.

Ivanov, I. G., Saraffova, A. A., and Abouhaidar, M. G. (1997). Unusual effect of clusters of rare arginine (AGG) codons on the expression of human interferon alpha 1 gene in Escherichia coli. Int J Biochem Cell Biol *29*, 659-666.

Jeong, H., Mason, S. P., Barabasi, A. L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. Nature *411*, 41-42.

Jermutus, L., Ryabova, L. A., and Pluckthun, A. (1998). Recent advances in producing and selecting functional proteins by using cell-free translation. Curr Opin Biotechnol *9*, 534-548.

Jung, E., Heller, M., Sanchez, J. C., and Hochstrasser, D. F. (2000). Proteomics meets cell biology: the establishment of subcellular proteomes. Electrophoresis *21*, 3369-3377.

Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Curr Opin Biotechnol *6*, 494-500.

Kapust, R. B., and Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci *8*, 1668-1674.

Khrapko, K. R., Lysov Yu, P., Khorlyn, A. A., Shick, V. V., Florentiev, V. L., and Mirzabekov, A. D. (1989). An oligonucleotide hybridization approach to DNA sequencing. FEBS Lett *256*, 118-122.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., and Yokoyama, S. (1999). Cell-free production and stable-isotope labeling of milligram quantities of proteins. FEBS Lett *442*, 15-19.

Kigawa, T., and Yokoyama, S. (2002). [High-throughput cell-free protein expression system for structural genomics and proteomics studies]. Tanpakushitsu Kakusan Koso *47*, 1014-1019.

Kim, D. M., and Swartz, J. R. (1999). Prolonging cell-free protein synthesis with a novel ATP regeneration system. Biotechnol Bioeng *66*, 180-188.

Kleffe, J., Hermann, K., Vahrson, W., Wittig, B., and Brendel, V. (1998). GeneGenerator--a flexible algorithm for gene prediction and its application to maize sequences. Bioinformatics *14*, 232-243.

Klein, J., and Dhurjati, P. (1995). Protein aggregation kinetics in an Escherichia coli strain overexpressing a Salmonella typhimurium CheY mutant gene. Appl Environ Microbiol *61*, 1220-1225.

Knudsen, E. S., and Wang, J. Y. (1996). Differential regulation of retinoblastoma protein function by specific Cdk phosphorylation sites. J Biol Chem *271*, 8313-8320.

Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L., Deciu, C., Haynes, P. A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J.*, et al.* (2002). Proteomic survey of metabolic pathways in rice. Proc Natl Acad Sci U S A *99*, 11969-11974.

Kumar, A., Agarwal, S., Heyman, J. A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y.*, et al.* (2002). Subcellular localization of the yeast proteome. Genes Dev *16*, 707-719.

LaBaer, J., Garrett, M. D., Stevenson, L. F., Slingerland, J. M., Sandhu, C., Chou, H. S., Fattaey, A., and Harlow, E. (1997). New functional activities for the p21 family of CDK inhibitors. Genes Dev *11*, 847-862.

Lamphere, L., Fiore, F., Xu, X., Brizuela, L., Keezer, S., Sardet, C., Draetta, G. F., and Gyuris, J. (1997). Interaction between Cdc37 and Cdk4 in human cells. Oncogene *14*, 1999-2004.

Lander, E. S. (1996). The new genomics: global views of biology. Science *274*, 536-539.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lesley, S. A., Graziano, J., Cho, C. Y., Knuth, M. W., and Klock, H. E. (2002a). Gene expression response to misfolded protein as a screen for soluble recombinant protein. Protein Eng *15*, 153-160.

Lesley, S. A., Kuhn, P., Godzik, A., Deacon, A. M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H. E., McMullan, D., Shin, T.*, et al.* (2002b). Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci U S A *99*, 11664-11669.

Letunic, I., Goodstadt, L., Dickens, N. J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R. R., Ponting, C. P., and Bork, P. (2002). Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Res *30*, 242-244.

Liao, H. H. (1991). Effect of temperature on the expression of wild-type and thermostable mutants of kanamycin nucleotidyltransferase in Escherichia coli. Protein Expr Purif *2*, 43-50.

Lilie, H., Schwarz, E., and Rudolph, R. (1998). Advances in refolding of proteins produced in E. coli. Curr Opin Biotechnol *9*, 497-501.

Lin, K., Kurland, I., Xu, L. Z., Lange, A. J., Pilkis, J., el-Maghrabi, M. R., and Pilkis, S. J. (1990). Expression of mammalian liver glycolytic/gluconeogenic enzymes in

Escherichia coli: recovery of active enzyme is strain and temperature dependent. Protein Expr Purif *1*, 169-176.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. Nat Genet *21*, 20-24.

Lo Conte, L., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2002). SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res *30*, 264-267.

MacBeath, G., and Schreiber, S. L. (2000). Printing proteins as microarrays for high-throughput function determination [see comments]. Science *289*, 1760-1763.

Madin, K., Sawasaki, T., Ogasawara, T., and Endo, Y. (2000). A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. Proc Natl Acad Sci U S A *97*, 559-564.

Maglott, D. R., Katz, K. S., Sicotte, H., and Pruitt, K. D. (2000). NCBI's LocusLink and RefSeq. Nucleic Acids Res *28*, 126-128.

Makalowska, I., Ryan, J. F., and Baxevanis, A. D. (2001). GeneMachine: gene prediction and sequence annotation. Bioinformatics *17*, 843-844.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science *298*, 1912-1934.

Martzen, M. R., McCraith, S. M., Spinelli, S. L., Torres, F. M., Fields, S., Grayhack, E. J., and Phizicky, E. M. (1999). A biochemical genomics approach for identifying genes by the activity of their products. Science *286*, 1153-1155.

Milanesi, L., D'Angelo, D., and Rogozin, I. B. (1999). GeneBuilder: interactive in silico prediction of gene structure. Bioinformatics *15*, 612-621.

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.

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H., Remington, K. A*., et al.* (2000). A whole-genome assembly of Drosophila. Science *287*, 2196-2204.

Parinov, S., Barsky, V., Yershov, G., Kirillov, E., Timofeev, E., Belgovskiy, A., and Mirzabekov, A. (1996). DNA sequencing by hybridization to microchip octa-and decanucleotides extended by stacked pentanucleotides. Nucleic Acids Res *24*, 2998-3004.

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A.*, et al.* (2000). Molecular portraits of human breast tumours. Nature *406*, 747-752.

Pilpel, Y., Sudarsanam, P., and Church, G. M. (2001). Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet *29*, 153-159.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y.*, et al.* (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet *20*, 207-211.

Possee, R. D., Thomas, C. J., and King, L. A. (1999). The use of baculovirus vectors for the production of membrane proteins in insect cells. Biochem Soc Trans *27*, 928-932.

Pradel, E., and Schnaitman, C. A. (1991). Effect of rfaH (sfrB) and temperature on expression of rfa genes of Escherichia coli K-12. J Bacteriol *173*, 6428-6431.

Pryor, K. D., and Leiting, B. (1997). High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system. Protein Expr Purif *10*, 309-319.

Rappsilber, J., and Mann, M. (2002). What does it mean to identify a protein in proteomics? Trends Biochem Sci *27*, 74-78.

Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T., Jackson, C., Shin-i, T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J.*, et al.* (2001). Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in C. elegans. Nat Genet *27*, 332-336.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E.*, et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., He, Y. D., Dai, H., Walker, W. L., Hughes, T. R.*, et al.* (2000). Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science *287*, 873-880.

Roberts, L. (2001). The human genome. Controversial from the start. Science *291*, 1182-1188.

Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M.*, et al.* (2000). Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet *24*, 227-235.

Rumlova, M., Benedikova, J., Cubinkova, R., Pichova, I., and Ruml, T. (2001). Comparison of classical and affinity purification techniques of Mason-Pfizer monkey virus capsid protein: the alteration of the product by an affinity tag. Protein Expr Purif *23*, 75-83.

Savchenko, A., Yee, A., Khachatryan, A., Skarina, T., Evdokimova, E., Pavlova, M., Semesi, A., Northey, J., Beasley, S., Lan, N.*, et al.* (2003). Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches. Proteins *50*, 392-399.

Sawasaki, T., Ogasawara, T., Morishita, R., and Endo, Y. (2002). A cell-free protein synthesis system for high-throughput proteomics. Proc Natl Acad Sci U S A *99*, 14652-14657.

Shen, X., Hu, P. P., Liberati, N. T., Datto, M. B., Frederick, J. P., and Wang, X. F. (1998). TGF-beta-induced phosphorylation of Smad3 regulates its interaction with coactivator p300/CREB-binding protein. Mol Biol Cell *9*, 3309-3319.

Shih, Y. P., Kung, W. M., Chen, J. C., Yeh, C. H., Wang, A. H., and Wang, T. F. (2002). High-throughput screening of soluble recombinant proteins. Protein Sci *11*, 1714-1719.

Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R., and Wiemann, S. (2000). Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. EMBO Rep *1*, 287-292.

Smith, D. B. (2000). Generating fusions to glutathione S-transferase for protein studies. Methods Enzymol *326*, 254-270.

Smith, D. B., and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene *67*, 31-40.

Solomon, M. J., and Kaldis, P. (1998). Regulation of CDKs by phosphorylation. Results Probl Cell Differ *22*, 79-109.

Sonnhammer, E. L., Eddy, S. R., and Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins *28*, 405-420.

Steczko, J., Donoho, G. A., Dixon, J. E., Sugimoto, T., and Axelrod, B. (1991). Effect of ethanol and low-temperature culture on expression of soybean lipoxygenase L-1 in Escherichia coli. Protein Expr Purif *2*, 221-227.

Stevenson, L. F., Kennedy, B. K., and Harlow, E. (2001). A large-scale overexpression screen in Saccharomyces cerevisiae identifies previously uncharacterized cell cycle genes. Proc Natl Acad Sci U S A *98*, 3946-3951.

Strausberg, R. L., Feingold, E. A., Klausner, R. D., and Collins, F. S. (1999). The mammalian gene collection. Science *286*, 455-457.

Strugnell, S. A., Wiefling, B. A., and Deluca, H. F. (1997). A modified pGEX vector with a C-terminal histidine tag: recombinant double-tagged protein obtained in greater yield and purity. Anal Biochem *254*, 147-149.

Tabuchi, M., Hino, M., Shinohara, Y., and Baba, Y. (2002). Cell-free protein synthesis on a microchip. Proteomics *2*, 430-435.

Tintignac, L. A., Leibovitch, M. P., Kitzmann, M., Fernandez, A., Ducommun, B., Meijer, L., and Leibovitch, S. A. (2000). Cyclin E-cdk2 phosphorylation promotes late G1-phase degradation of MyoD in muscle cells. Exp Cell Res *259*, 300-307.

Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H.*, et al.* (2001a). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364-2368.

Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H.*, et al.* (2001b). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364-2368.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P.*, et al.* (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623-627.

Vaillancourt, P., Zheng, C. F., Hoang, D. Q., and Breister, L. (2000). Affinity purification of recombinant proteins fused to calmodulin or to calmodulin-binding peptides. Methods Enzymol *326*, 340-362.

Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., and Hunkapiller, M. (1998). Shotgun sequencing of the human genome. Science *280*, 1540-1542.

Vitkup, D., Melamud, E., Moult, J., and Sander, C. (2001). Completeness in structural genomics. Nat Struct Biol *8*, 559-566.

Walhout, A. J., Sordella, R., Lu, X., Hartley, J. L., Temple, G. F., Brasch, M. A., Thierry-Mieg, N., and Vidal, M. (2000a). Protein interaction mapping in C. elegans using proteins involved in vulval development [see comments]. Science *287*, 116-122.

Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S., and Vidal, M. (2000b). GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Methods Enzymol *328*, 575-592.

Walhout, A. J., and Vidal, M. (2001). High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods *24*, 297-306.

Wall, J. G., and Pluckthun, A. (1995). Effects of overexpressing folding modulators on the in vivo folding of heterologous proteins in Escherichia coli. Curr Opin Biotechnol *6*, 507-516.

Wasserman, W. W., Palumbo, M., Thompson, W., Fickett, J. W., and Lawrence, C. E. (2000). Human-mouse genome comparisons to locate regulatory sites. Nat Genet *26*, 225-228.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P*., et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Weickert, M. J., Doherty, D. H., Best, E. A., and Olins, P. O. (1996). Optimization of heterologous protein production in Escherichia coli. Curr Opin Biotechnol *7*, 494-499.

Wigley, W. C., Stidham, R. D., Smith, N. M., Hunt, J. F., and Thomas, P. J. (2001). Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. Nat Biotechnol *19*, 131-136.

Wu, L. F., Hughes, T. R., Davierwala, A. P., Robinson, M. D., Stoughton, R., and Altschuler, S. J. (2002). Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters. Nat Genet *31*, 255-265.

Xu, Y., and Uberbacher, E. C. (1996). Gene prediction by pattern recognition and homology search. Proc Int Conf Intell Syst Mol Biol *4*, 241-251.

Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G. M., Bhattacharyya, S., Gutierrez, P.*, et al.* (2002). An NMR approach to structural proteomics. Proc Natl Acad Sci U S A *99*, 1825-1830.

Yokoyama, S., Matsuo, Y., Hirota, H., Kigawa, T., Shirouzu, M., Kuroda, Y., Kurumizaka, H., Kawaguchi, S., Ito, Y., Shibata, T.*, et al.* (2000). Structural genomics projects in Japan. Prog Biophys Mol Biol *73*, 363-376.

Zahn, K. (1996). Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. J Bacteriol *178*, 2926-2933.

Zhao, J., Dynlacht, B., Imai, T., Hori, T., and Harlow, E. (1998). Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S- phase entry. Genes Dev *12*, 456-461.

Zheng, X. F., and Chan, T. F. (2002). Chemical genomics in the global study of protein functions. Drug Discov Today *7*, 197-205.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T.*, et al.* (2001). Global analysis of protein activities using proteome chips. Science *293*, 2101-2105.