

# biosvm

October 12, 2009

## References

- [1] J. Aires-de Sousa and J. Gasteiger. Prediction of enantiomeric excess in a combinatorial library of catalytic enantioselective reactions. *J Comb Chem*, 7(2):298–301, 2005.
- [2] C.F. Aliferis, D.P. Hardin, and P. Massion. Machine Learning Models For Lung Cancer Classification Using Array Comparative Genomic Hybridization. In *Proceedings of the 2002 American Medical Informatics Association (AMIA) Annual Symposium*, pages 7–11, 2002.
- [3] C. Ambroise and G.J. McLachlan. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc. Natl. Acad. Sci. USA*, 99(10):6562–6566, 2002.
- [4] D.C. Anderson, W. Li, D.G. Payan, and W.S. Noble. A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. *J Proteome Res*, 2(2):137–146, 2003.
- [5] Yindalon Aphinyanaphongs, Ioannis Tsamardinos, Alexander Statnikov, Douglas Hardin, and Constantin F Aliferis. Text categorization models for high-quality article retrieval in internal medicine. *J. Am. Med. Inform. Assoc.*, 12(2):207–16, 2005.
- [6] Rieko Arimoto, Madhu-Ashni Prasad, and Eric M Gifford. Development of CYP3A4 inhibition models: comparisons of machine-learning techniques and molecular descriptors. *J Biomol Screen*, 10(3):197–205, Apr 2005.
- [7] Tomasz Arodz, Marcin Kurdziel, Erik O D Sevre, and David A Yuen. Pattern recognition techniques for automatic detection of suspicious-looking anomalies in mammograms. *Comput. Methods Programs Biomed.*, 79(2):135–49, Aug 2005.
- [8] V. Atalay and R. Cetin-Atalay. Implicit motif distribution based hybrid computational kernel for sequence classification. *Bioinformatics*, 21(8):1429–1436, Apr 2005.

- [9] L. Bao and Z. Sun. Identifying genes related to drug anticancer mechanisms using support vector machine. *FEBS Lett.*, 521:109–114, 2002.
- [10] Lei Bao. Identifying genes related to chemosensitivity using support vector machine. *Methods Mol Med*, 111:233–40, 2005.
- [11] Lei Bao and Yan Cui. Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. *Bioinformatics*, 21(10):2185–90, May 2005.
- [12] C. Baumgartner, C. Bohm, D. Baumgartner, G. Marini, K. Weinberger, B. Olgemoller, B. Liebl, and A. A. Roscher. Supervised machine learning techniques for the classification of metabolic disorders in newborns. *Bioinformatics*, 20(17):2985–2996, 2004.
- [13] A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, and D. Romani. An SVM classifier to separate false signals from microcalcifications in digital mammograms. *Phys Med Biol*, 46(6):1651–63, Jun 2001.
- [14] N. Beerenwinkel, T. Lengauer, M. Daumer, R. Kaiser, H. Walter, K. Korn, D. Hoffmann, and J. Selbig. Methods for optimizing antiviral combination therapies. *Bioinformatics*, 19(Suppl. 1):i16–i25, 2003.
- [15] N. Beerenwinkel, B. Schmidt, H. Walter, R. Kaiser, T. Lengauer, D. Hoffman, K. Korn, and J. Selbig. Geno2pheno: Interpreting Genotypic HIV Drug Resistance Tests. *IEEE Intelligent Systems*, 6(6):35–41, 2001.
- [16] A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer, and Z. Yakhini. Tissue classification with gene expression profiles. *J. Comput. Biol.*, 7(3-4):559–583, 2000.
- [17] A. Ben-Hur and D. Brutlag. Remote homology detection: a motif based approach. *Bioinformatics*, 19(Suppl. 1):i26–i33, 2003.
- [18] A. Ben-Hur and W. S. Noble. Kernel methods for predicting protein-protein interactions. *Bioinformatics*, 21(Suppl. 1):i38–i46, Jun 2005.
- [19] M. Bern, D. Goldberg, W. H. McDonald, and III Yates, J. R. Automatic Quality Assessment of Peptide Tandem Mass Spectra. *Bioinformatics*, 20(Suppl. 1):i49–i54, 2004.
- [20] M. Bhasin and G. P. S. Raghava. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci.*, 13(3):596–607, Mar 2004.
- [21] M. Bhasin and G. P. S. Raghava. Classification of Nuclear Receptors Based on Amino Acid Composition and Dipeptide Composition. *J. Biol. Chem.*, 279(22):23262–23266, 2004.

- [22] M. Bhasin and G. P. S. Raghava. ESLpred: SVM-based method for sub-cellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. *Nucl. Acids Res.*, 32(Suppl. 2):W414–419, 2004.
- [23] M. Bhasin and G. P. S. Raghava. GPCRpred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors. *Nucl. Acids Res.*, 32(Suppl.2):W383–389, 2004.
- [24] M. Bhasin and G. P. S. Raghava. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*, 22(23-24):3195–3204, 2004.
- [25] M. Bhasin and G. P. S. Raghava. SVM based method for predicting HLA-DRB1\*0401 binding peptides in an antigen sequence. *Bioinformatics*, 20(3):421–423, 2004.
- [26] M. Bhasin and G. P. S. Raghava. GPCRclass: a web tool for the classification of amine type of G-protein-coupled receptors. *Nucleic Acids Res.*, 33(Web Server issue):W143–7, Jul 2005.
- [27] M. Bhasin and G. P. S. Raghava. Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Res*, 33(Web Server issue):W202–7, Jul 2005.
- [28] J. R. Bock and D. A. Gough. Predicting protein-protein interactions from primary structure. *Bioinformatics*, 17(5):455–460, 2001.
- [29] J. R. Bock and D. A. Gough. A New Method to Estimate Ligand-Receptor Energetics. *Mol Cell Proteomics*, 1(11):904–910, 2002.
- [30] J. R. Bock and D. A. Gough. Whole-proteome interaction mining. *Bioinformatics*, 19(1):125–134, 2003.
- [31] Andrew J Bordner and Ruben Abagyan. Statistical analysis and prediction of protein-protein interfaces. *Proteins*, 60(3):353–66, Aug 2005.
- [32] K.M. Borgwardt, C.S. Ong, S. Schönauer, S.V.N. Vishwanathan, A.J. Smola, and H.-P. Kriegel. Protein function prediction via graph kernels. *Bioinformatics*, 21(Suppl. 1):i47–i56, Jun 2005.
- [33] James R Bradford and David R Westhead. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics*, 21(8):1487–94, Apr 2005.
- [34] Hans Briem and Judith Günther. Classifying ”kinase inhibitor-likeness” by using machine-learning methods. *Chembiochem*, 6(3):558–66, Mar 2005.

- [35] M. P. Brown, W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Ares, and D. Haussler. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA*, 97(1):262–7, Jan 2000.
- [36] R. Bunescu, R. Ge, R. J. Kate, E. M. Marcotte, R. J. Mooney, A. K. Ramani, and Y. W. Wong. Comparative experiments on learning information extractors for proteins and their interactions. *Artif. Intell. Med.*, 33(2):139–55, Feb 2005.
- [37] R. Burbidge, M. Trotter, B. Buxton, and S. Holden. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput. Chem.*, 26(1):4–15, December 2001.
- [38] T. Burckin, R. Nagel, Y. Mandel-Gutfreund, L. Shiue, T. A. Clark, J.-L. Chong, T.-H. Chang, S. Squazzo, G. Hartzog, and M. Ares. Exploring functional relationships between components of the gene expression machinery. *Nat. Struct. Mol. Biol.*, 12(2):175–82, Feb 2005.
- [39] S. Busuttill, J. Abela, and G. J. Pace. Support vector machines with profile-based kernels for remote protein homology detection. *Genome Inform Ser Workshop Genome Inform*, 15(2):191–200, 2004.
- [40] E. Byvatov, U. Fechner, J. Sadowski, and G. Schneider. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J Chem Inf Comput Sci*, 43(6):1882–9, 2003.
- [41] E. Byvatov and G. Schneider. Support vector machine applications in bioinformatics. *Appl Bioinformatics*, 2(2):67–77, 2003.
- [42] Evgeny Byvatov and Gisbert Schneider. SVM-based feature selection for characterization of focused compound collections. *J Chem Inf Comput Sci*, 44(3):993–9, 2004.
- [43] C. Z. Cai, L. Y. Han, Z. L. Ji, X. Chen, and Y. Z. Chen. SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res*, 31(13):3692–7, Jul 2003.
- [44] C.Z. Cai, L.Y. Han, Z.L. Ji, and Y.Z. Chen. Enzyme family classification by support vector machines. *Proteins*, 55(1):66–76, 2004.
- [45] C.Z. Cai, W.L. Wang, L.Z. Sun, and Y.Z. Chen. Protein function classification via support vector machine approach. *Math. Biosci.*, 185(2):111–122, 2003.
- [46] Y.-D. Cai, X.-J. Liu, X.-B. Xu, and K.-C. Chou. Support vector machines for prediction of protein subcellular location by incorporating quasi-sequence-order effect. *J. Cell. Biochem.*, 84(2):343–348, 2002.

- [47] Y.-D. Cai, X.-J. Liu, X.-B. Xu, and G.-P. Zhou. Support Vector Machines for predicting protein structural class. *BMC Bioinformatics*, 2(3):3, 2001.
- [48] Y.-D. Cai, X.-J. Liu, X.-B. Xu, and G.-P. Zhou. Prediction of protein structural classes by support vector machines. *Comput. Chem.*, 26(3):293–296, 2002.
- [49] Y.-D. Cai, G.-P. Zhou, and K.-C. Chou. Support Vector Machines for Predicting Membrane Protein Types by Using Functional Domain Composition. *Biophys. J.*, 84(5):3257–3263, 2003.
- [50] Y.D. Cai, K.Y. Feng, Y.X. Li, and K.C. Chou. Support vector machine for predicting alpha-turn types. *Peptides*, 24(4):629–630, 2003.
- [51] Y.D. Cai and S.L. Lin. Support vector machines for predicting rRNA-, RNA-, and DNA-binding proteins from amino acid sequence. *Biochim. Biophys. Acta*, 1648(1-2):127–133, 2003.
- [52] Y.D. Cai, S.L. Lin, and K.C. Chou. Support vector machines for prediction of protein signal sequences and their cleavage sites. *Peptides*, 24(1):159–161, 2003.
- [53] Y.D. Cai, X.J. Liu, Y.X. Li, X.B. Xu, and K.C. Chou. Prediction of beta-turns with learning machines. *Peptides*, 24(5):665–669, 2003.
- [54] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support vector machines for prediction of protein subcellular location. *Mol. Cell Biol. Res. Commun.*, 4(4):230–234, 2000.
- [55] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support Vector Machines for predicting HIV protease cleavage sites in protein. *J. Comput. Chem.*, 23(2):267–274, 2002.
- [56] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support vector machines for predicting the specificity of GalNAc-transferase. *Peptides*, 23:205–208, 2002.
- [57] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support vector machines for the classification and prediction of beta-turn types. *J. Pept. Sci.*, 8(7):297–301, 2002.
- [58] Y.D. Cai, X.J. Liu, X.B. Xu, and K.C. Chou. Support vector machines for prediction of protein domain structural class. *J. Theor. Biol.*, 221(1):115–120, 2003.
- [59] Y.D. Cai, P.W. Ricardo, C.H. Jen, and K.C. Chou. Application of SVM to predict membrane protein types. *J. Theor. Biol.*, 226(4):373–376, 2004.
- [60] Y.D. Cai, G.P. Zhou, C.H. Jen, S.L. Lin, and K.C. Chou. Identify catalytic triads of serine hydrolases by support vector machines. *J. Theor. Biol.*, 228(4):551–557, 2004.

- [61] Yu-Dong Cai and Andrew J Doig. Prediction of *Saccharomyces cerevisiae* protein functional class from functional domain composition. *Bioinformatics*, 20(8):1292–300, May 2004.
- [62] G. Camps-Valls, A.M. Chalk, A.J. Serrano-Lopez, J.D. Martin-Guerrero, and E.L. Sonnhammer. Profiled support vector machines for antisense oligonucleotide efficacy prediction. *BMC Bioinformatics*, 5(135):135, 2004.
- [63] E. Capriotti, P. Fariselli, and R. Casadio. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, 33(Web Server issue):W306–10, Jul 2005.
- [64] R. J. Carter, I. Dubchak, and S. R. Holbrook. A computational approach to identify genes for functional RNAs in genomic sequences. *Nucl. Acids Res.*, 29(19):3928–3938, 2001.
- [65] Y. Chen and D. Xu. Understanding protein dispensability through machine-learning analysis of high-throughput data. *Bioinformatics*, 21:575–581, Mar 2005.
- [66] Y.C. Chen, Y.S. Lin, C.J. Lin, and J.K. Hwang. Prediction of the bonding states of cysteines using the support vector machines based on multiple feature vectors and cysteine state sequences. *Proteins*, 55(4):1036–1042, 2004.
- [67] K.-C. Chou and Y.-D. Cai. Using Functional Domain Composition and Support Vector Machines for Prediction of Protein Subcellular Location. *J. Biol. Chem.*, 277(48):45765–45769, 2002.
- [68] M. L. Chow, E. J. Moler, and I. S. Mian. Identifying marker genes in transcription profiling data using a mixture of feature relevance experts. *Physiol. Genomics*, 5(2):99–111, Mar 2001.
- [69] Nigel Collier and Koichi Takeuchi. Comparison of character-level and part of speech features for name recognition in biomedical texts. *J Biomed Inform*, 37(6):423–35, Dec 2004.
- [70] Q. Cui, T. Jiang, B. Liu, and S. Ma. Esub8: A novel tool to predict protein subcellular localizations in eukaryotic organisms. *BMC Bioinformatics*, 5(66):66, 2004.
- [71] M. Cuturi and J.-P. Vert. A mutual information kernel for strings. In *Proceedings of IJCNN 2004*, pages 1904–1910, 2004.
- [72] M. Cuturi and J.-P. Vert. The context-tree kernel for strings. *Neural Network.*, 18(4):1111–1123, 2005.
- [73] S. Degroeve, B. De Baets, Y. Van de Peer, and P. Rouze. Feature subset selection for splice site prediction. *Bioinformatics*, 18(Suppl. 1):S75–S83, 2002.

- [74] S. Degroeve, Y. Saeys, B. De Baets, P. Rouze, and Y. Van de Peer. SpliceMachine: predicting splice sites from high-dimensional local context representations. *Bioinformatics*, 21:1332–1338, 2005.
- [75] M. Deshpande and G. Karypis. Evaluation of Techniques for Classifying Biological Sequences. In *PAKDD '02: Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, pages 417–431. Springer Verlag, 2002.
- [76] C.H.Q. Ding and I. Dubchak. Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics*, 17:349–358, 2001.
- [77] P.D. Dobson and A.J. Doig. Distinguishing enzyme structures from non-enzymes without alignments. *J. Mol. Biol.*, 330(4):771–783, 2003.
- [78] P.D. Dobson and A.J. Doig. Predicting enzyme class from protein structure without alignments. *J. Mol. Biol.*, 345(1):187–199, Jan 2005.
- [79] I. Donaldson, J. Martin, B. de Bruijn, C. Wolting, V. Lay, B. Tuekam, S. Zhang, B. Baskin, G.D. Bader, K. Michalickova, T. Pawson, and C.W.V. Hogue. PreBIND and Textomy - mining the biomedical literature for protein-protein interactions using a support vector machine. *BMC Bioinformatics*, 4(1):11, Mar 2003.
- [80] S. Doniger, T. Hofmann, and J. Yeh. Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms. *J. Comput. Biol.*, 9(6):849–864, 2002.
- [81] P. Dönnes and A. Elofsson. Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics*, 3(1):25, Sep 2002.
- [82] P. Dönnes and O. Kohlbacher. Integrated modeling of the major events in the MHC class I antigen processing pathway. *Protein Sci.*, 14:2132–2140, Jun 2005.
- [83] G. Dror, R. Sorek, and R. Shamir. Accurate identification of alternatively spliced exons using support vector machine. *Bioinformatics*, 21(7):897–901, Apr 2005.
- [84] Anshul Dubey, Matthew J Realff, Jay H Lee, and Andreas S Bommarius. Support vector machines for learning to identify the critical positions of a protein. *J Theor Biol*, 234(3):351–61, Jun 2005.
- [85] J. H. Fong, A. E. Keating, and M. Singh. Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol.*, 5(R11), 2004.
- [86] C. C. Friedel, K. H. V. Jahn, S. Sommer, S. Rudd, H. W. Mewes, and I. V. Tetko. Support vector machines for separation of mixed plant-pathogen EST collections based on codon usage. *Bioinformatics*, 21:1383–1388, 2005.

- [87] B. Fritz, F. Schubert, G. Wrobel, C. Schwaenen, S. Wessendorf, M. Nessling, C. Korz, R. J. Rieker, K. Montgomery, R. Kucherlapati, G. Mechttersheimer, R. Eils, S. Joos, and P. Lichter. Microarray-based Copy Number and Expression Profiling in Dedifferentiated and Pleomorphic Liposarcoma. *Cancer Res.*, 62(11):2993–2998, 2002.
- [88] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10):906–914, Oct 2000.
- [89] C. Furlanello, M. Serafini, S. Merler, and G. Jurman. Entropy-based gene ranking without selection bias for the predictive classification of microarray data. *BMC Bioinformatics*, 4(54), 2003.
- [90] Rajeev Gangal and Pankaj Sharma. Human pol II promoter prediction: time series descriptors and machine learning. *Nucleic Acids Res*, 33(4):1332–6, 2005.
- [91] J. L. Gardy, M. R. Laird, F. Chen, S. Rey, C. J. Walsh, M. Ester, and F. S. L. Brinkman. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, 21(5):617–623, Mar 2005.
- [92] A. Garg, M. Bhasin, and G.P. Raghava. SVM-based method for subcellular localization of human proteins using amino acid compositions, their order and similarity search. *J. Biol. Chem.*, 280(15):14427–32, Apr 2005.
- [93] T. Gärtner, J.W. Lloyd, and P.A. Flach. Kernels and distances for structured data. *Mach. Learn.*, 57(3):205–232, 2004.
- [94] S. Gaudan, H. Kirsch, and D. Rebholz-Schuhmann. Resolving abbreviations to their senses in Medline. *Bioinformatics*, Jul 2005.
- [95] S. M. Gomez, W. S. Noble, and A. Rzhetsky. Learning to predict protein-protein interactions from protein sequences. *Bioinformatics*, 19(15):1875–1881, 2003.
- [96] L. Gordon, A. Y. Chervonenkis, A. J. Gammerman, I. A. Shakhmuradov, and V. V. Solovyev. Sequence alignment kernel for recognition of promoter regions. *Bioinformatics*, 19(15):1964–1971, 2003.
- [97] Y. Guermeur. Combining Discriminant Models with New Multi-Class SVMs. *Pattern Anal. Appl.*, 5(2):168–179, 2002.
- [98] Y. Guermeur, A. Lifschitz, and R. Vert. A kernel for protein secondary structure prediction. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 193–206. MIT Press, 2004.



- [99] Y. Guermeur, G. Pollastri, A. Elisseeff, D. Zelus, H. Paugam-Moisy, and P. Baldi. Combining protein secondary structure prediction models with ensemble methods of optimal complexity. *Neurocomputing*, 56:305–327, 2004.
- [100] J. Guo, H. Chen, Z. Sun, and Y. Lin. A novel method for protein secondary structure prediction using dual-layer SVM and profiles. *Proteins*, 54(4):738–743, 2004.
- [101] Ting Guo, Yanxin Shi, and Zhirong Sun. A novel statistical ligand-binding site predictor: application to ATP-binding sites. *Protein Eng Des Sel*, 18(2):65–70, Feb 2005.
- [102] T. Gururaja, W. Li, W.S. Noble, D.G. Payan, and D.C. Anderson. Multiple functional categories of proteins identified in an in vitro cellular ubiquitin affinity extract using shotgun peptide sequencing. *J Proteome Res*, 2(394-404):394–404, 2003.
- [103] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Mach. Learn.*, 46(1/3):389–422, Jan 2002.
- [104] Torsten Haferlach, Alexander Kohlmann, Susanne Schnittger, Martin Dugas, Wolfgang Hiddemann, Wolfgang Kern, and Claudia Schoch. AML M3 and AML M3 variant each have a distinct gene expression signature but also share patterns different from other genetically defined AML subtypes. *Genes Chromosomes Cancer*, 43(2):113–27, Jun 2005.
- [105] Torsten Haferlach, Alexander Kohlmann, Susanne Schnittger, Martin Dugas, Wolfgang Hiddemann, Wolfgang Kern, and Claudia Schoch. A global approach to the diagnosis of leukemia using gene expression profiling. *Blood*, 106(4):1189–1198, Aug 2005.
- [106] J. Hakenberg, S. Schmeier, A. Kowald, E. Klipp, and U. Leser. Finding kinetic parameters using text mining. *OMICS*, 8(2):131–152, 2004.
- [107] Jrg Hakenberg, Steffen Bickel, Conrad Plake, Ulf Brefeld, Hagen Zahn, Lukas Faulstich, Ulf Leser, and Tobias Scheffer. Systematic feature evaluation for gene name recognition. *BMC Bioinformatics*, 6 Suppl 1:S9, 2005.
- [108] L.Y. Han, C.Z. Cai, Z.L. Ji, Z.W. Cao, J. Cui, and Y.Z. Chen. Predicting functional family of novel enzymes irrespective of sequence similarity: a statistical learning approach. *Nucl. Acids Res.*, 32(21):6437–6444, 2004.
- [109] L.Y. Han, C.Z. Cai, Z.L. Ji, and Y.Z. Chen. Prediction of functional class of novel viral proteins by a statistical learning method irrespective of sequence similarity. *Virology*, 331(1):136–143, 2005.

- [110] L.Y. Han, C.Z. Cai, S.L. Lo, M.C. Chung, and Y.Z. Chen. Prediction of RNA-binding proteins from primary sequence by a support vector machine approach. *RNA*, 10(3):355–368, 2004.
- [111] Sangjo Han, Byung-Chul Lee, Seung Taek Yu, Chan-Seok Jeong, Soy-oung Lee, and Dongsup Kim. Fold recognition by combining profile-profile alignment and support vector machine. *Bioinformatics*, 21(11):2667–73, Jun 2005.
- [112] D. Haussler. Convolution Kernels on Discrete Structures. Technical Report UCSC-CRL-99-10, UC Santa Cruz, 1999.
- [113] C. Helma, T. Cramer, S. Kramer, and L. De Raedt. Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. *J. Chem. Inf. Comput. Sci.*, 44(4):1402–11, 2004.
- [114] S. Hochreiter and K. Obermayer. Gene selection for microarray data. In B. Schlkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 319–355. MIT Press, 2004.
- [115] Oliver Hofmann and Dietmar Schomburg. Concept-based annotation of enzyme classes. *Bioinformatics*, 21(9):2059–66, May 2005.
- [116] Y. Hou, W. Hsu, M. L. Lee, and C. Bystroff. Efficient remote homology detection using local structure. *Bioinformatics*, 19(17):2294–2301, 2003.
- [117] Y. Hou, W. Hsu, M. L. Lee, and C. Bystroff. Remote homolog detection using local sequence-structure correlations. *Proteins*, 57(3):518–530, 2004.
- [118] C. Hu, X. Li, and J. Liang. Developing optimal non-linear scoring function for protein design. *Bioinformatics*, 20(17):3080–3098, 2004.
- [119] H.J. Hu, Y. Pan, R. Harrison, and P.C. Tai. Improved protein secondary structure prediction using support vector machine with a new encoding scheme and an advanced tertiary classifier. *IEEE Trans. Nanobioscience*, 3(4):265–271, 2004.
- [120] J. Hua, Z. Xiong, J. Lowey, E. Suh, and E. R. Dougherty. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics*, 21(8):1509–1515, Apr 2005. To appear.
- [121] S. Hua and Z. Sun. A Novel Method of Protein Secondary Structure Prediction with High Segment Overlap Measure: Support Vector Machine Approach. *J. Mol. Biol.*, 308(2):397–407, April 2001.
- [122] S. Hua and Z. Sun. Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, 17(8):721–728, 2001.

- [123] J. Huan, W. Wang, A. Washington, J. Prins, R. Shah, and A. Tropsha. Accurate classification of protein structural families using coherent sub-graph analysis. In *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 411–422, 2004.
- [124] Jing Huang and Feng Shi. Support vector machines for predicting apoptosis proteins types. *Acta Biotheor.*, 53(1):39–47, 2005.
- [125] N. Huang, H. Chen, and Z. Sun. CTKPred: an SVM-based method for the prediction and classification of the cytokine superfamily. *Protein Eng. Des. Sel.*, Jun 2005.
- [126] Shao-Wei Huang and Jenn-Kang Hwang. Computation of conformational entropy from protein sequences using the machine-learning method—application to the study of the relationship between structural conservation and local structural stability. *Proteins*, 59(4):802–9, Jun 2005.
- [127] T. M. Huang and V. Kecman. Gene extraction for cancer diagnosis by support vector machines—An improvement. *Artif. Intell. Med.*, Jul 2005.
- [128] B. Hutter, C. Schaab, S. Albrecht, M. Borgmann, N. A. Brunner, C. Freiberg, K. Ziegelbauer, C. O. Rock, I. Ivanov, and H. Loferer. Prediction of Mechanisms of Action of Antibacterial Compounds by Gene Expression Profiling. *Antimicrob. Agents Chemother.*, 48(8):2838–2844, Aug 2004.
- [129] T. Jaakkola, M. Diekhans, and D. Haussler. A Discriminative Framework for Detecting Remote Protein Homologies. *J. Comput. Biol.*, 7(1,2):95–114, 2000.
- [130] T. S. Jaakkola, M. Diekhans, and D. Haussler. Using the Fisher Kernel Method to Detect Remote Protein Homologies. In *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*, pages 149–158. AAAI Press, 1999.
- [131] T. S. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *Proc. of Tenth Conference on Advances in Neural Information Processing Systems*, 1999.
- [132] Barbara Jarzab, Malgorzata Wiench, Krzysztof Fajurewicz, Krzysztof Simek, Michal Jarzab, Malgorzata Oczko-Wojciechowska, Jan Wloch, Agnieszka Czarniecka, Ewa Chmielik, Dariusz Lange, Agnieszka Pawlaczek, Sylwia Szpak, Elzbieta Gubala, and Andrzej Swierniak. Gene expression profile of papillary thyroid cancer: sources of variability and diagnostic implications. *Cancer Res.*, 65(4):1587–97, Feb 2005.
- [133] S. Jiang-Ning, L. Wei-Jiang, and X. Wen-Bo. Cooperativity of the oxidation of cysteines in globular proteins. *J. Theor. Biol.*, 231(1):85–95, 2004.

- [134] R. N. Jorissen and M. K. Gilson. Virtual screening of molecular databases using a support vector machine. *J Chem Inf Model*, 45(3):549–61, 2005.
- [135] R. Karchin, K. Karplus, and D. Haussler. Classifying G-protein coupled receptors with support vector machines. *Bioinformatics*, 18:147–159, 2002.
- [136] R. Karchin, L. Kelly, and A. Sali. Improving functional annotation of non-synonymous SNPs with information theory. *Pac Symp Biocomput*, pages 397–408, 2005.
- [137] Y. Karklin, R. F. Meraz, and S.R. Holbrook. Classification of non-coding RNA using graph representations of secondary structure. *Pac. Symp. Biocomput.*, pages 4–15, 2005.
- [138] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized Kernels between Labeled Graphs. In T. Faucett and N. Mishra, editors, *Proceedings of the Twentieth International Conference on Machine Learning*, pages 321–328, New York, NY, USA, 2003. AAAI Press.
- [139] H. Kashima, K. Tsuda, and A. Inokuchi. Kernels for graphs. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 155–170. MIT Press, The MIT Press, Cambridge, Massachusetts, 2004.
- [140] H. Kim and H. Park. Protein secondary structure prediction based on an improved support vector machines approach. *Protein Eng.*, 16(8):553–560, Aug 2003.
- [141] H. Kim and H. Park. Prediction of protein relative solvent accessibility with support vector machines and long-range interaction 3D local descriptor. *Proteins*, 54(3):557–562, Feb 2004.
- [142] J. H. Kim, J. Lee, B. Oh, K. Kimm, and I. Koh. Prediction of phosphorylation sites using SVMs. *Bioinformatics*, 20(17):3179–3184, 2004.
- [143] T. Kin, K. Tsuda, and K. Asai. Marginalized kernels for RNA sequence data analysis. In R.H. Lathrop, K. Nakai, S. Miyano, T. Takagi, and M. Kanehisa, editors, *Genome Informatics 2002*, pages 112–122. Universal Academic Press, 2002.
- [144] A. Kohlmann, C. Schoch, S. Schnittger, M. Dugas, W. Hiddemann, W. Kern, and T. Haferlach. Pediatric acute lymphoblastic leukemia (ALL) gene expression signatures classify an independent cohort of adult ALL patients. *Leukemia*, 18(1):63–71, 2004.
- [145] A. Koike and T. Takagi. Prediction of protein-protein interaction sites using support vector machines. *Protein Eng. Des. Sel.*, 17(2):165–173, Feb 2004.

- [146] D. Komura, H. Nakamura, S. Tsutsumi, H. Aburatani, and S. Ihara. Multidimensional support vector machines for visualization of gene expression data. *Bioinformatics*, 21(4):439–444, Feb 2005.
- [147] R. Kondor and J.-P. Vert. Diffusion kernels. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 171–192. MIT Press, 2004.
- [148] R. I. Kondor and J. Lafferty. Diffusion Kernels on Graphs and Other Discrete Input. In *ICML 2002*, 2002.
- [149] Zsofia Kote-Jarai, Richard D Williams, Nicola Cattini, Maria Copeland, Ian Giddings, Richard Wooster, Robert H tePoele, Paul Workman, Barry Gusterson, John Peacock, Gerald Gui, Colin Campbell, and Ros Eeles. Gene expression profiling after radiation-induced DNA damage is strongly predictive of BRCA1 mutation carrier status. *Clin. Cancer Res.*, 10(3):958–63, Feb 2004.
- [150] S. Kramer, E. Frank, and C. Helma. Fragment generation and support vector machines for inducing SARs. *SAR QSAR Environ Res*, 13(5):509–23, Jul 2002.
- [151] V. G. Krishnan and D. R. Westhead. A comparative study of machine-learning methods to predict the effects of single nucleotide polymorphisms on protein function. *Bioinformatics*, 19(17):2199–2209, 2003.
- [152] B. Krishnapuram, L. Carin, and A. Hartemink. Gene expression analysis: joint feature selection and classifier design. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 299–317. MIT Press, 2004.
- [153] B. Krishnapuram, L. Carin, and A. Hartemink. Joint Classifier and Feature Optimization for Comprehensive Cancer Diagnosis Using Gene Expression Data. *J. Comput. Biol.*, 11(2-3):227–242, 2004.
- [154] Balaji Krishnapuram, Alexander J Hartemink, Lawrence Carin, and Mrio A T Figueiredo. A bayesian approach to joint feature selection and classifier design. *IEEE Trans Pattern Anal Mach Intell*, 26(9):1105–11, Sep 2004.
- [155] R. Kuang, E. Ie, K. Wang, K. Wang, M. Siddiqi, Y. Freund, and C. Leslie. Profile-based string kernels for remote homology detection and motif extraction. *Proc IEEE Comput Syst Bioinform Conf*, pages 152–160, 2004.
- [156] R. Kuang, E. Ie, K. Wang, K. Wang, M. Siddiqi, Y. Freund, and C. Leslie. Profile-based string kernels for remote homology detection and motif extraction. *J. Bioinform. Comput. Biol.*, 3(3):527–550, Jun 2005.

- [157] R. Kuang, C. S. Leslie, and A.-S. Yang. Protein backbone angle prediction with machine learning approaches. *Bioinformatics*, 20(10):1612–1621, 2004.
- [158] M. Kumar, M. Bhasin, N. K. Natt, and G. P. S. Raghava. BhairPred: prediction of beta-hairpins in a protein from multiple alignment information using ANN and SVM techniques. *Nucleic Acids Res*, 33(Web Server issue):W154–9, Jul 2005.
- [159] G. R. G. Lanckriet, T. De Bie, N. Cristianini, M. I. Jordan, and W. S. Noble. A statistical framework for genomic data fusion. *Bioinformatics*, 20(16):2626–2635, 2004.
- [160] G.R. Lanckriet, M. Deng, N. Cristianini, M.I. Jordan, and W.S. Noble. Kernel-based data fusion and its application to protein function prediction in yeast. In *Proceedings of the Pacific Symposium on Biocomputing*, pages 300–311, 2004.
- [161] G.R.G. Lanckriet, N. Cristianini, M.I. Jordan, and W.S. Noble. Kernel-based integration of genomic data using semidefinite programming. In B. Schlkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 231–259. MIT Press, 2004.
- [162] Dongkwon Lee, Sang Wook Choi, Myengsoo Kim, Jin Hyun Park, Moonkyu Kim, Jungchul Kim, and In-Beum Lee. Discovery of differentially expressed genes related to histological subtype of hepatocellular carcinoma. *Biotechnol Prog.*, 19(3):1011–5, 2003.
- [163] Y. Lee and C.-K. Lee. Classification of multiple cancer types by multicategory support vector machines using gene expression data. *Bioinformatics*, 19(9):1132–1139, 2003.
- [164] C. Leslie, E. Eskin, and W.S. Noble. The spectrum kernel: a string kernel for SVM protein classification. In Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Kevin Lauerdale, and Teri E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 564–575, Singapore, 2002. World Scientific.
- [165] C. Leslie, E. Eskin, J. Weston, and W.S. Noble. Mismatch String Kernels for SVM Protein Classification. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.
- [166] C. S. Leslie, E. Eskin, A. Cohen, J. Weston, and W. S. Noble. Mismatch string kernels for discriminative protein classification. *Bioinformatics*, 20(4):467–476, 2004.
- [167] D. Lett, M. Hsing, and F. Pio. Interaction profile-based protein classification of death domain. *BMC Bioinformatics*, 5(75), 2004.

- [168] H. Li, C. Ung, C. Yap, Y. Xue, Z. Li, Z. Cao, and Y. Chen. Prediction of genotoxicity of chemical compounds by statistical learning methods. *Chem. Res. Toxicol.*, 18(6):1071–1080, Jun 2005.
- [169] L. Li, W. Jiang, X. Li, K.L. Moser, Z. Guo, L. Du, Q. Wang, E.J. Topol, Q. Wang, and S. Rao. A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset. *Genomics*, 85(1):16–23, 2005.
- [170] L. Li, H. Tang, Z. Wu, J. Gong, M. Gruidl, J. Zou, M. Tockman, and R.A. Clark. Data mining techniques for cancer detection using serum proteomic profiling. *Artif. Intell. Med.*, 32(2):71–83, 2004.
- [171] T. Li, C. Zhang, and M. Ogihara. A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, 20(15):2429–2437, 2004.
- [172] L. Liao and W. S. Noble. Combining pairwise sequence similarity and support vector machines for remote protein homology detection. In *Proceedings of the Sixth International Conference on Computational Molecular Biology*, 2002.
- [173] L. Liao and W.S. Noble. Combining Pairwise Sequence Similarity and Support Vector Machines for Detecting Remote Protein Evolutionary and Structural Relationships. *J. Comput. Biol.*, 10(6):857–868, 2003.
- [174] K. Lin, Y. Kuang, J. S. Joseph, and P. R. Kolatkar. Conserved codon composition of ribosomal protein coding genes in Escherichia coli, Mycobacterium tuberculosis and Saccharomyces cerevisiae: lessons from supervised machine learning in functional genomics. *Nucl. Acids Res.*, 30(11):2599–2607, 2002.
- [175] P. Lind and T. Maltseva. Support vector machines for the estimation of aqueous solubility. *J Chem Inf Comput Sci*, 43(6):1855–9, 2003.
- [176] J. Listgarten, S. Damaraju, B. Poulin, L. Cook, J. Dufour, A. Driga, J. Mackey, D. Wishart, R. Greiner, and B. Zanke. Predictive Models for Breast Cancer Susceptibility from Multiple Single Nucleotide Polymorphisms. *Clin. Cancer Res.*, 10(8):2725–2737, 2004.
- [177] H. X. Liu, C. X. Xue, R. S. Zhang, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. Quantitative prediction of logk of peptides in high-performance liquid chromatography based on molecular descriptors by using the heuristic method and support vector machine. *J Chem Inf Comput Sci*, 44(6):1979–86, 2004.
- [178] H. X. Liu, R. S. Zhang, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. QSAR study of ethyl 2-[(3-methyl-2,5-dioxo(3-pyrrolinyl)amino]-4-(trifluoromethyl) pyrimidine-5-carboxylate: an inhibitor of AP-1 and

NF-kappa B mediated gene expression based on support vector machines. *J Chem Inf Comput Sci*, 43(4):1288–96, 2003.

- [179] H. X. Liu, R. S. Zhang, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. Prediction of the isoelectric point of an amino acid based on GA-PLS and SVMs. *J Chem Inf Comput Sci*, 44(1):161–7, 2004.
- [180] H. X. Liu, R. S. Zhang, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. QSAR and classification models of a novel series of COX-2 selective inhibitors: 1,5-diarylimidazoles based on support vector machines. *J Comput Aided Mol Des*, 18(6):389–99, Jun 2004.
- [181] Huiqing Liu, Hao Han, Jinyan Li, and Limsoon Wong. An in-silico method for prediction of polyadenylation signals in human sequences. *Genome Inform Ser Workshop Genome Inform*, 14:84–93, 2003.
- [182] Huiqing Liu, Hao Han, Jinyan Li, and Limsoon Wong. Using amino acid patterns to accurately predict translation initiation sites. *In Silico Biol.*, 4(3):255–69, 2004.
- [183] Huiqing Liu, Jinyan Li, and Limsoon Wong. Use of extreme patient samples for outcome prediction from gene expression data. *Bioinformatics*, Jun 2005.
- [184] Jane Jijun Liu, Gene Cutler, Wuxiong Li, Zheng Pan, Sihua Peng, Tim Hoey, Liangbiao Chen, and Xuefeng Bruce Ling. Multiclass cancer classification and biomarker discovery using GA-based algorithms. *Bioinformatics*, 21(11):2691–7, Jun 2005.
- [185] Y. Liu. Active learning with support vector machine applied to gene expression data for cancer classification. *J. Chem. Inf. Comput. Sci.*, 44(6):1936–1941, 2004.
- [186] Y. Liu. A comparative study on feature selection methods for drug discovery. *J Chem Inf Comput Sci*, 44(5):1823–8, 2004.
- [187] Zhenqiu Liu, Dechang Chen, and Halima Bensmail. Gene expression data classification with kernel principal component analysis. *J Biomed Biotechnol*, 2005(2):155–9, 2005.
- [188] Siaw Ling Lo, Cong Zhong Cai, Yu Zong Chen, and Maxey C M Chung. Effect of training datasets on support vector machine prediction of protein-protein interactions. *Proteomics*, 5(4):876–84, Mar 2005.
- [189] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. J. C. H. Watkins. Text classification using string kernels. *J. Mach. Learn. Res.*, 2:419–444, 2002.
- [190] H. Lodhi, J. Shawe-Taylor, N. Cristianini, and C. J. C. H. Watkins. Text Classification using String Kernels. In *Adv. Neural Inform. Process. Syst.*, pages 563–569, 2000.



- [191] B. Logan, P. Moreno, B. Suzek, Z. Weng, and S. Kasif. A Study of Remote Homology Detection. Technical Report CRL 2001/05, Compaq Cambridge Research laboratory, June 2001.
- [192] Y.J. Lu, D. Williamson, R. Wang, B. Summersgill, S. Rodriguez, S. Rogers, K. Pritchard-Jones, C. Campbell, and J. Shipley. Expression profiling targeting chromosomes for tumor classification and prediction of clinical behavior. *Genes Chromosomes Cancer*, 38(3):207–214, 2003.
- [193] Feng Luan, Ruisheng Zhang, Chunyan Zhao, Xiaojun Yao, Mancang Liu, Zhide Hu, and Botao Fan. Classification of the carcinogenicity of N-nitroso compounds based on support vector machines and linear discriminant analysis. *Chem Res Toxicol*, 18(2):198–203, Feb 2005.
- [194] P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret, and J.-P. Vert. Extensions of marginalized graph kernels. In R. Greiner and D. Schuurmans, editors, *Proceedings of the Twenty-First International Conference on Machine Learning (ICML 2004)*, pages 552–559. ACM Press, 2004.
- [195] P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret, and J.-P. Vert. Graph kernels for molecular structure-activity relationship analysis with support vector machines. *J. Chem. Inf. Model.*, 45(4):939–51, 2005.
- [196] M.Z. Man, G. Dyson, K. Johnson, and B. Liao. Evaluating methods for classifying expression data. *J. Biopharm. Stat.*, 14(4):1065–1084, 2004.
- [197] Yong Mao, Xiaobo Zhou, Daoying Pi, Youxian Sun, and Stephen T C Wong. Multiclass cancer classification by using fuzzy support vector machine and binary decision tree with gene selection. *J Biomed Biotechnol*, 2005(2):160–71, 2005.
- [198] F. Markowetz, L. Edler, and M. Vingron. Support Vector Machines for Protein Fold Class Prediction. *Biometrical Journal*, 45(3):377–389, 2003.
- [199] S. Martin, D. Roe, and J.-L. Faulon. Predicting protein-protein interactions using signature products. *Bioinformatics*, 21(2):218–226, Jan 2005.
- [200] A. Matsuda, J.-P. Vert, H. Saigo, N. Ueda, H. Toh, and T. Akutsu. A novel representation of protein sequences for prediction of subcellular location using support vector machines. *Protein Sci.*, 14(11):2804–2813, 2005.
- [201] T. Mattfeldt, H.W. Gottfried, H. Wolter, V. Schmidt, H.A. Kestler, and J. Mayer. Classification of prostatic carcinoma with artificial neural networks using comparative genomic hybridization and quantitative stereological data. *Pathol. Res. Pract.*, 199(12):773–784, 2003.
- [202] Larry McKnight and Padmini Srinivasan. Categorization of sentence types in medical abstracts. *AMIA Annu Symp Proc*, pages 440–4, 2003.

- [203] P. Meinicke, M. Tech, B. Morgenstern, and R. Merkl. Oligo kernels for datamining on biological sequences: a case study on prokaryotic translation initiation sites. *BMC Bioinformatics*, 5(169), 2004.
- [204] S.I. Meireles, A.F. Carvalho, R. Hirata, A.L. Montagnini, W.K. Martins, F.B. Runza, B.S. Stolf, L. Termini, C.E. Neto, R.L. Silva, F.A. Soares, E.J. Neves, and L.F. Reis. Differentially expressed genes in gastric tumors identified by cDNA array. *Cancer Lett.*, 190(2):199–211, Feb 2003.
- [205] M. Middendorf, E. Ziv, C. Adams, J. Hom, R. Koytcheff, C. Levovitz, G. Woods, L. Chen, and C. Wiggins. Discriminative topological features reveal biological network mechanisms. *BMC Bioinformatics*, 5(181), 2004.
- [206] Sven Mika and Burkhard Rost. NLProt: extracting protein names and sequences from papers. *Nucleic Acids Res*, 32(Web Server issue):W634–7, Jul 2004.
- [207] Sven Mika and Burkhard Rost. Protein names precisely peeled off free text. *Bioinformatics*, 20(Suppl. 1):i241–i247, 2004.
- [208] Tomohiro Mitsumori, Sevrani Fation, Masaki Murata, Kouichi Doi, and Hirohumi Doi. Gene/protein name recognition based on support vector machine using dictionary as features. *BMC Bioinformatics*, 6 Suppl 1:S8, 2005.
- [209] F. Model, P. Adorjan, A. Olek, and C. Piepenbrock. Feature selection for DNA methylation based cancer classification. *Bioinformatics*, 17(Suppl. 1):S157–S164, 2001.
- [210] E. J. Moler, M. L. Chow, and I. S. Mian. Analysis of molecular profile data using generative and discriminative methods. *Physiol. Genomics*, 4(2):109–126, Dec 2000.
- [211] S. Mukherjee, P. Tamayo, J. P. Mesirov, D. Slonim, A. Verri, and T. Poggio. Support vector machine classification of microarray data. Technical Report 182, C.B.L.C., 1998. A.I. Memo 1677.
- [212] K.-R. Müller, G. Rätsch, S. Sonnenburg, S. Mika, M. Grimm, and N. Heinrich. Classifying 'drug-likeness' with Kernel-based learning methods. *J Chem Inf Model*, 45(2):249–53, 2005.
- [213] E. Myasnikova, A. Samsonova, M. Samsonova, and J. Reinitz. Support vector regression applied to the determination of the developmental age of a *Drosophila* embryo from its segmentation gene expression patterns. *Bioinformatics*, 18(Suppl. 1):S87–S95, 2002.
- [214] Rajesh Nair and Burkhard Rost. Mimicking cellular sorting improves prediction of subcellular localization. *J Mol Biol*, 348(1):85–100, Apr 2005.

- [215] N.K. Natt, H. Kaur, and G.P. Raghava. Prediction of transmembrane regions of beta-barrel proteins using ANN- and SVM-based methods. *Proteins*, 56(1):11–18, 2004.
- [216] M. N. Nguyen and J. C. Rajapakse. Two-stage multi-class support vector machines to protein secondary structure prediction. *Pac Symp Biocomput*, pages 346–57, 2005.
- [217] Minh N Nguyen and Jagath C Rajapakse. Multi-class support vector machines for protein secondary structure prediction. *Genome Inform Ser Workshop Genome Inform*, 14:218–27, 2003.
- [218] Minh N Nguyen and Jagath C Rajapakse. Prediction of protein relative solvent accessibility with a two-stage SVM approach. *Proteins*, 59(1):30–7, Apr 2005.
- [219] W. S. Noble. Support vector machine applications in computational biology. In B. Scholkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 71–92. MIT Press, 2004.
- [220] Rebekah K O’Donnell, Michael Kupferman, S. Jack Wei, Sunil Singhal, Randal Weber, Bert O’Malley, Yi Cheng, Mary Putt, Michael Feldman, Barry Ziober, and Ruth J Muschel. Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene*, 24(7):1244–51, Feb 2005.
- [221] R. A. O’Flanagan, G. Paillard, R. Lavery, and A. M. Sengupta. Non-additivity in protein-DNA binding. *Bioinformatics*, 21(10):2254–63, May 2005.
- [222] Tapio Pahikkala, Filip Ginter, Jorma Boberg, Jouni Jarvinen, and Tapio Salakoski. Contextual weighting for Support Vector Machines in literature mining: an application to gene versus protein name disambiguation. *BMC Bioinformatics*, 6(1):157, Jun 2005.
- [223] Fei Pan, Baoying Wang, Xin Hu, and William Perrizo. Comprehensive vertical sample-based KNN/LSVM classification for gene expression analysis. *J Biomed Inform*, 37(4):240–8, Aug 2004.
- [224] K.-J. Park and M. Kanehisa. Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. *Bioinformatics*, 19(13):1656–1663, 2003.
- [225] A. Passerini and P. Frasconi. Learning to discriminate between ligand-bound and disulfide-bound cysteines. *Protein Eng. Des. Sel.*, 17(4):367–373, 2004.
- [226] D.J. Patterson, K. Yasuhara, and W.L. Ruzzo. Pre-mRNA secondary structure prediction aids splice site prediction. In Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Kevin Lauerdale, and Teri E. Klein,

editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 223–234. World Scientific, 2002.

- [227] S. Pavey, P. Johansson, L. Packer, J. Taylor, M. Stark, P.M. Pollock, G.J. Walker, G.M. Boyle, U. Harper, S.J. Cozzi, K. Hansen, L. Yudt, C. Schmidt, P. Hersey, K.A. Ellem, M.G. O’Rourke, P.G. Parsons, P. Meltzer, M. Ringner, and N.K. Hayward. Microarray expression profiling in melanoma reveals a BRAF mutation signature. *Oncogene*, 23(23):4060–4067, May 2004.
- [228] P. Pavlidis, T. S. Furey, M. Liberto, D. Haussler, and W. N. Grundy. Promoter Region-Based Classification of Genes. In *Pacific Symposium on Biocomputing*, pages 139–150, 2001.
- [229] P. Pavlidis, J. Weston, J. Cai, and W.N. Grundy. Gene functional classification from heterogeneous data. In *Proceedings of the Fifth Annual International Conference on Computational Biology*, pages 249–255, 2001.
- [230] P. Pavlidis, J. Weston, J. Cai, and W.S. Noble. Learning gene functional classifications from multiple data types. *J. Comput. Biol.*, 9(2):401–411, 2002.
- [231] S. Peng, Q. Xu, X.B. Ling, X. Peng, W. Du, and L. Chen. Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. *FEBS Lett.*, 555(2):358–362, 2003.
- [232] Tho Hoan Pham, Kenji Satou, and Tu Bao Ho. Prediction and analysis of beta-turns in proteins by support vector machine. *Genome Inform Ser Workshop Genome Inform*, 14:196–205, 2003.
- [233] Tho Hoan Pham, Kenji Satou, and Tu Bao Ho. Support vector machines for prediction and analysis of beta and gamma-turns in proteins. *J. Bioinform. Comput. Biol.*, 3(2):343–58, Apr 2005.
- [234] Dariusz Plewczynski, Adrian Tkacz, Adam Godzik, and Leszek Rychlewski. A support vector machine approach to the identification of phosphorylation sites. *Cell Mol Biol Lett*, 10(1):73–89, 2005.
- [235] Dariusz Plewczynski, Adrian Tkacz, Lucjan Stanislaw Wyrwicz, and Leszek Rychlewski. AutoMotif server: prediction of single residue post-translational modifications in proteins. *Bioinformatics*, 21(10):2525–7, May 2005.
- [236] N. Pochet, F. De Smet, J. A. K. Suykens, and B. L. R. De Moor. Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction. *Bioinformatics*, 20(17):3185–3195, Nov 2004.

- [237] J. Prados, A. Kalousis, J.C. Sanchez, L. Allard, O. Carrette, and M. Hilario. Mining mass spectra for diagnosis and biomarker discovery of cerebral accidents. *Proteomics*, 4(8):2320–2332, 2004.
- [238] J. Qian, J. Lin, N. M. Luscombe, H. Yu, and M. Gerstein. Prediction of regulatory networks: genome-wide identification of transcription factor targets from gene expression data. *Bioinformatics*, 19(15):1917–1926, 2003.
- [239] J. Qin, D. P. Lewis, and W. S. Noble. Kernel hierarchical gene clustering from microarray expression data. *Bioinformatics*, 19(16):2097–2104, 2003.
- [240] Gajendra P S Raghava and Joon H Han. Correlation and prediction of gene expression level from amino acid and dipeptide composition of its protein. *BMC Bioinformatics*, 6(1):59, Mar 2005.
- [241] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J.P. Mesirov, T. Poggio, W. Gerald, M. Loda, E.S. Lander, and T.R. Golub. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. USA*, 98(26):15149–15154, Dec 2001.
- [242] H. Rangwala and G. Karypis. Profile-based direct kernels for remote homology detection and fold recognition. *Bioinformatics*, 21(23):4239–4247, Dec 2005.
- [243] G. Ratsch and S. Sonnenburg. Accurate splice site detection for *Caenorhabditis elegans*. In B. Scholkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 277–298. MIT Press, 2004.
- [244] G. Ratsch, S. Sonnenburg, and B. Scholkopf. RASE: recognition of alternatively spliced exons in *C.elegans*. *Bioinformatics*, 21(Suppl. 1):i369–i377, Jun 2005.
- [245] Stefan A Rensing, Dana Fritzowsky, Daniel Lang, and Ralf Reski. Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens*. *BMC Genomics*, 6(1):43, Mar 2005.
- [246] I. Res, I. Mihalek, and O. Lichtarge. An evolution based classifier for prediction of protein interfaces without using protein structures. *Bioinformatics*, 21(10):2496–501, May 2005.
- [247] Simon B Rice, Goran Nenadic, and Benjamin J Stapley. Mining protein function from text using term-based support vector machines. *BMC Bioinformatics*, 6 Suppl 1:S22, 2005.

- [248] Henning Riedesel, Björn Kolbeck, Oliver Schmetzer, and Ernst-Walter Knapp. Peptide binding at class I major histocompatibility complex scored with linear functions and support vector machines. *Genome Inform Ser Workshop Genome Inform*, 15(1):198–212, 2004.
- [249] Thorsteinn Rögnvaldsson and Liwen You. Why neural networks should not be used for HIV-1 protease cleavage site prediction. *Bioinformatics*, 20(11):1702–9, Jul 2004.
- [250] J. R. Rose, Jr. Turkett, W. H., I. C. Oroian, W. W. Laegreid, and J. Keele. Correlation of amino acid preference and mammalian viral genome type. *Bioinformatics*, 2005.
- [251] S. Rudd and I. V. Tetko. Eclair—a web service for unravelling species origin of sequences sampled from mixed host interfaces. *Nucleic Acids Res*, 33(Web Server issue):W724–7, Jul 2005.
- [252] S. Ruepp, F. Boess, L. Suter, M. C. de Vera, G. Steiner, T. Steele, T. Weiser, and S. Albertini. Assessment of hepatotoxic liabilities by transcript profiling. *Toxicol Appl Pharmacol*, Jun 2005.
- [253] J. Saeh, P. Lyne, B. Takasaki, and D. Cosgrove. Lead hopping using SVM and 3D pharmacophore fingerprints. *J Chem Inf Model*, 45(4):1122–1133, Jul 2005.
- [254] P. Saetrom. Predicting the efficacy of short oligonucleotides in antisense and RNAi experiments with boosted genetic programming. *Bioinformatics*, 20(17):3055–3063, 2004.
- [255] Y. Saeyns, S. Degroeve, D. Aeyels, P. Rouz, and Y. Van de Peer. Feature selection for splice site prediction: A new method using EDA-based feature ranking. *BMC Bioinformatics*, 5(64), 2004.
- [256] Y. Saeyns, S. Degroeve, D. Aeyels, Y. Van de Peer, and P. Rouze. Fast feature selection using a simple estimation of distribution algorithm: a case study on splice site prediction. *Bioinformatics*, 19(Suppl. 1):ii179–ii188, 2003.
- [257] H. Saigo, J.-P. Vert, N. Ueda, and T. Akutsu. Protein homology detection using string alignment kernels. *Bioinformatics*, 20(11):1682–1689, 2004.
- [258] Marta Sanchez-Carbayo, Nicholas D Socci, Juan Jose Lozano, Wentian Li, Elizabeth Charytonowicz, Thomas J Belbin, Michael B Prystowsky, Angel R Ortiz, Geoffrey Childs, and Carlos Cordon-Cardo. Gene discovery in bladder cancer progression using cDNA microarrays. *Am. J. Pathol.*, 163(2):505–16, Aug 2003.
- [259] Deepak Sarda, Gek Huey Chua, Kuo-Bin Li, and Arun Krishnan. pSLIP: SVM based protein subcellular localization prediction using multiple physicochemical properties. *BMC Bioinformatics*, 6(1):152, Jun 2005.

- [260] B. Schölkopf, K. Tsuda, and J.-P. Vert. *Kernel Methods in Computational Biology*. MIT Press, The MIT Press, Cambridge, Massachusetts, 2004.
- [261] Holger Schwender, Manuela Zucknick, Katja Ickstadt, Hermann M Bolt, and G. E. N. I. C. A. network. A pilot study on the application of statistical classification procedures to molecular epidemiological data. *Toxicol Lett*, 151(1):291–9, Jun 2004.
- [262] M. Seeger. Covariance Kernels from Bayesian Generative Models. In *Adv. Neural Inform. Process. Syst.*, volume 14, pages 905–912, 2002.
- [263] M. R. Segal, K. D. Dahlquist, and B. R. Conklin. Regression approaches for microarray data analysis. *J. Comput. Biol.*, 10(6):961–980, 2003.
- [264] N. H. Segal, P. Pavlidis, C. R. Antonescu, R. G. Maki, W. S. Noble, D. DeSantis, J. M. Woodruff, J. J. Lewis, M. F. Brennan, A. N. Houghton, and C. Cordon-Cardo. Classification and Subtype Prediction of Adult Soft Tissue Sarcoma by Functional Genomics. *Am. J. Pathol.*, 163(2):691–700, Aug 2003.
- [265] N. H. Segal, P. Pavlidis, W. S. Noble, C. R. Antonescu, A. Viale, U. V. Wesley, K. Busam, H. Gallardo, D. DeSantis, M. F. Brennan, C. Cordon-Cardo, J. D. Wolchok, and A. N. Houghton. Classification of Clear-Cell Sarcoma as a Subtype of Melanoma by Genomic Profiling. *J. Clin. Oncol.*, 21(9):1775–1781, May 2003.
- [266] M. Seike, T. Kondo, K. Fujii, T. Okano, T. Yamada, Y. Matsuno, A. Gemma, S. Kudoh, and S. Hirohashi. Proteomic signatures for histological types of lung cancer. *Proteomics*, Jul 2005.
- [267] T.Z. Sen, A. Kloczkowski, R.L. Jernigan, C. Yan, V. Honavar, K.M. Ho, C.Z. Wang, Y. Ihm, H. Cao, X. Gu, and D. Dobbs. Predicting binding sites of hydrolase-inhibitor complexes by combining several methods. *BMC Bioinformatics*, 5(205), 2004.
- [268] Pasak Senawongse, Andrew R Dalby, and Zheng Rong Yang. Predicting the phosphorylation sites using hidden markov models and machine learning methods. *J Chem Inf Model*, 45(4):1147–52, 2005.
- [269] J.R. Serra, E.D. Thompson, and P.C. Jurs. Development of binary classification of structural chromosome aberrations for a diverse set of organic compounds from molecular structure. *Chem. Res. Toxicol.*, 16(2):153–163, 2003.
- [270] Roded Sharan and Eugene W Myers. A motif-based framework for recognizing sequence families. *Bioinformatics*, 21 Suppl 1:i387–i393, Jun 2005.
- [271] R. She, F. Chen, K. Wang, M. Ester, J.L. Gardy, and F.S.L. Brinkman. Frequent-subsequence-based prediction of outer membrane proteins. In

*KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 436–445. ACM Press, 2003.

- [272] Lei Shi and Fabien Campagne. Building a protein name dictionary from full text: a machine learning term extraction approach. *BMC Bioinformatics*, 6(1):88, Apr 2005.
- [273] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. T. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G. A. Pinkus, T. S. Ray, M. A. Koval, K. W. Last, A. Norton, T. A. Lister, J. Mesirov, D. S. Neuberg, E. S. Lander, J. C. Aster, and T. R. Golub. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.*, 8(1):68–74, 2002.
- [274] J. A. Siepen, S. E. Radford, and D. R. Westhead. Beta Edge strands in protein structure prediction and aggregation. *Protein Sci.*, 12(10):2348–2359, 2003.
- [275] S. Sonnenburg, G. Rätsch, A. Jagota, and K.-R. Müller. New methods for splice-site recognition. In JR. Dorronsoro, editor, *Proc. International conference on artificial Neural Networks ? ICANN?02*, number 2415 in LNCS, pages 329–336. Springer Berlin, 2002.
- [276] M. J. Sorich, J. O. Miners, R. A. McKinnon, D. A. Winkler, F. R. Burden, and P. A. Smith. Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms. *J Chem Inf Comput Sci*, 43(6):2019–24, 2003.
- [277] Michael J Sorich, Ross A McKinnon, John O Miners, David A Winkler, and Paul A Smith. Rapid prediction of chemical metabolism by human UDP-glucuronosyltransferase isoforms using quantum chemical descriptors derived with the electronegativity equalization method. *J Med Chem*, 47(21):5311–7, Oct 2004.
- [278] B.J. Stapley, L.A. Kelley, and M.J. Sternberg. Predicting the sub-cellular location of proteins from text using support vector machines. In Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Kevin Lauerdale, and Teri E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 374–385. World Scientific, 2002.
- [279] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics*, 2005. To appear.
- [280] Alexander Statnikov, Constantin F Aliferis, and Ioannis Tsamardinos. Methods for multi-category cancer diagnosis from gene expression data: a



comprehensive evaluation to inform decision support system development. *Medinfo*, 11(Pt 2):813–7, 2004.

- [281] Guido Steiner, Laura Suter, Franziska Boess, Rodolfo Gasser, Maria Cristina de Vera, Silvio Albertini, and Stefan Ruepp. Discriminating different classes of toxicants by transcript profiling. *Environ. Health Perspect.*, 112(12):1236–48, Aug 2004.
- [282] A. I. Su, J. B. Welsh, L. M. Sapinoso, S. G. Kern, P. Dimitrov, H. Lapp, P. G. Schultz, S. M. Powell, C. A. Moskaluk, H. F. Jr. Frierson, and G. M. Hampton. Molecular Classification of Human Carcinomas by Use of Gene Expression Signatures. *Cancer Res.*, 61(20):7388–7393, 2001.
- [283] Yang Su, T.M. Murali, Vladimir Pavlovic, Michael Schaffer, and Simon Kasif. RankGene: identification of diagnostic genes based on expression data. *Bioinformatics*, 19(12):1578–1579, 2003.
- [284] Y.F. Sun, X.D. Fan, and Y.D. Li. Identifying splicing sites in eukaryotic RNA: support vector machine approach. *Comput. Biol. Med.*, 33(1):17–29, 2003.
- [285] S. J. Swamidass, J. Chen, J. Bruand, P. Phung, L. Ralaivola, and P. Baldi. Kernels for small molecules and the prediction of mutagenicity, toxicity and anti-cancer activity. *Bioinformatics*, 21(Suppl. 1):i359–i368, Jun 2005.
- [286] Y. Takaoka, Y. Endo, S. Yamanobe, H. Kakinuma, T. Okubo, Y. Shimazaki, T. Ota, S. Sumiya, and K. Yoshikawa. Development of a method for evaluating drug-likeness and ease of synthesis using a data set in which compounds are assigned scores based on chemists’ intuition. *J Chem Inf Comput Sci*, 43(4):1269–75, 2003.
- [287] Koichi Takeuchi and Nigel Collier. Bio-medical entity extraction using support vector machines. *Artif. Intell. Med.*, 33(2):125–37, Feb 2005.
- [288] Thomas Tang, Jinbo Xu, and Ming Li. Discovering sequence-structure motifs from protein segments and two applications. *Pac Symp Biocomput*, pages 370–81, 2005.
- [289] Reiji Teramoto, Mikio Aoki, Toru Kimura, and Masaharu Kanaoka. Prediction of siRNA functionality using generalized string kernel and support vector machine. *FEBS Lett.*, 579(13):2878–82, May 2005.
- [290] M. Tobita, T. Nishikawa, and R. Nagashima. A discriminant model constructed by the support vector machine method for HERG potassium channel inhibitors. *Bioorg. Med. Chem. Lett.*, 15(11):2886–90, Jun 2005.
- [291] Richard W Tothill, Adam Kowalczyk, Danny Rischin, Alex Bousioutas, Izhak Haviv, Ryan K van Laar, Paul M Waring, John Zalcborg, Robyn Ward, Andrew V Biankin, Robert L Sutherland, Susan M Henshall, Kwun Fong, Jonathan R Pollack, David D L Bowtell, and Andrew J Holloway.

An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. *Cancer Res.*, 65(10):4031–40, May 2005.

- [292] C.A. Tsai, C.H. Chen, T.C. Lee, I.C. Ho, U.C. Yang, and J.J. Chen. Gene selection for sample classifications in microarray experiments. *DNA Cell Biol.*, 23(10):607–614, 2004.
- [293] A. Tsirigos and I. Rigoutsos. A sensitive, support-vector-machine method for the detection of horizontal gene transfers in viral, archaeal and bacterial genomes. *Nucleic Acids Res.*, 33(12):3699–707, 2005.
- [294] K. Tsuda, S. Akaho, and K. Asai. The em Algorithm for Kernel Matrix Completion with Auxiliary Data. *J. Mach. Learn. Res.*, 4:67–81, 2003.
- [295] K. Tsuda, M. Kawanabe, G. Rätsch, S. Sonnenburg, and K.-R. Müller. A new discriminative kernel from probabilistic models. *Neural Computation*, 14(10):2397–2414, 2002.
- [296] K. Tsuda, T. Kin, and K. Asai. Marginalized Kernels for Biological Sequences. *Bioinformatics*, 18:S268–S275, 2002.
- [297] K. Tsuda and W.S. Noble. Learning kernels from biological networks by maximizing entropy. *Bioinformatics*, 20:i326–i333, 2004.
- [298] G. Valentini. Gene expression data analysis of human lymphoma using support vector machines and output coding ensembles. *Artif. Intell. Med.*, 26(3):281–304, Nov 2002.
- [299] J.-P. Vert. Support vector machine prediction of signal peptide cleavage site using a new class of kernels for strings. In R. B. Altman, A. K. Dunker, L. Hunter, K. Lauerdale, and T. E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*, pages 649–660. World Scientific, 2002.
- [300] J.-P. Vert. A tree kernel to analyze phylogenetic profiles. *Bioinformatics*, 18:S276–S284, 2002.
- [301] J.-P. Vert. Kernel methods in computational biology. Technical Report ccscd-00012124, CNRS-HAL, Oct 2005.
- [302] J.-P. Vert and M. Kanehisa. Graph-driven features extraction from microarray data. Technical Report 0206055, Arxiv physics, 2002.
- [303] J.-P. Vert and M. Kanehisa. Extracting active pathways from gene expression data. *Bioinformatics*, 19:238ii–234ii, 2003.
- [304] J.-P. Vert and M. Kanehisa. Graph-driven features extraction from microarray data using diffusion kernels and kernel CCA. In S. Becker, S. Thrun, and K. Obermayer, editors, *Adv. Neural Inform. Process. Syst.*, pages 1449–1456. MIT Press, 2003.

- [305] J.-P. Vert, H. Saigo, and T. Akutsu. Local alignment kernels for biological sequences. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 131–154. MIT Press, The MIT Press, Cambridge, Massachusetts, 2004.
- [306] J.-P. Vert, R. Thurman, and W. S. Noble. Kernels for gene regulatory regions. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Adv. Neural Inform. Process Syst.*, volume 18, pages 1401–1408, Cambridge, MA, 2006. MIT Press.
- [307] J.-P. Vert, K. Tsuda, and B. Schölkopf. A primer on kernel methods. In B. Schölkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 35–70. MIT Press, 2004.
- [308] J.-P. Vert and Y. Yamanishi. Supervised graph inference. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Adv. Neural Inform. Process. Syst.*, volume 17, pages 1433–1440. MIT Press, Cambridge, MA, 2005.
- [309] A. Vinayagam, R. Knig, J. Moormann, F. Schubert, R. Eils, K.-H. Glattig, and S. Suhai. Applying Support Vector Machines for Gene Ontology based gene function prediction. *BMC Bioinformatics*, 5(1):116, Aug 2004.
- [310] Kristian Vlahovicek, Lszl Kajn, Vilmos Agoston, and Sndor Pongor. The SBASE domain sequence resource, release 12: prediction of protein domain-architecture using support vector machines. *Nucleic Acids Res*, 33(Database issue):D223–5, Jan 2005.
- [311] M. Wagner, D. Naik, and A. Pothen. Protocols for disease classification from mass spectrometry data. *Proteomics*, 3(9):1692–1698, 2003.
- [312] M. Wagner, D.N. Naik, A. Pothen, S. Kasukurti, R.R. Devineni, B.L. Adam, O.J. Semmes, and G.L. Wright Jr. Computational protein biomarker prediction: a case study for prostate cancer. *BMC Bioinformatics*, 5(26), 2004.
- [313] J. Wang, W.-K. Sung, A. Krishnan, and K.-B. Li. Protein subcellular localization prediction for Gram-negative bacteria using amino acid sub-alphabets and a combination of multiple support vector machines. *BMC Bioinformatics*, 6(1):174, Jul 2005.
- [314] Long-Hui Wang, Juan Liu, Yan-Fu Li, and Huai-Bei Zhou. Predicting protein secondary structure by a support vector machine based on a new coding scheme. *Genome Inform Ser Workshop Genome Inform*, 15(2):181–90, 2004.
- [315] M. Wang, J. Yang, and K-C. Chou. Using string kernel to predict signal peptide cleavage site based on subsite coupling model. *Amino Acids*, 28(4):395–402, Jun 2005.

- [316] M. Wang, J. Yang, G.-P. Liu, Z.-J. Xu, and K.-C. Chou. Weighted-support vector machines for predicting membrane protein types based on pseudo-amino acid composition. *Protein Eng. Des. Sel.*, 17(6):509–516, 2004.
- [317] M-L. Wang, W-J. Li, M-L. Wang, and W-B. Xu. Support vector machines for prediction of peptidyl prolyl cis/trans isomerization. *J Pept Res*, 63(1):23–8, Jan 2004.
- [318] Ming-Lei Wang, Hui Yao, and Wen-Bo Xu. Prediction by support vector machines and analysis by Z-score of poly-L-proline type II conformation based on local sequence. *Comput. Biol. Chem.*, 29(2):95–100, Apr 2005.
- [319] Yu Wang, Igor V Tetko, Mark A Hall, Eibe Frank, Axel Facius, Klaus F X Mayer, and Hans W Mewes. Gene selection from microarray data for cancer classification—a machine learning approach. *Comput. Biol. Chem.*, 29(1):37–46, Feb 2005.
- [320] J. J. Ward, L. J. McGuffin, B. F. Buxton, and D. T. Jones. Secondary structure prediction with support vector machines. *Bioinformatics*, 19(13):1650–1655, 2003.
- [321] Jeffrey F Waring, Roger G Ulrich, Nick Flint, David Morfitt, Arno Kalkuhl, Frank Staedtler, Michael Lawton, Johanna M Beekman, and Laura Suter. Interlaboratory evaluation of rat hepatic gene expression changes induced by methapyrilene. *Environ Health Perspect*, 112(4):439–48, Mar 2004.
- [322] M. K. Warmuth, J. Liao, G. Rätsch, M. Mathieson, S. Putta, and C. Lemmen. Active learning with support vector machines in the drug discovery process. *J Chem Inf Comput Sci*, 43(2):667–673, 2003.
- [323] M. K. Warmuth, G. Rätsch, M. Mathieson, L. Liao, and C. Lemmen. Active learning in the drug discovery process. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Adv. Neural Inform. Process. Syst.*, volume 14, pages 1449–1456. MIT Press, 2002.
- [324] C. Watkins. Dynamic alignment kernels. In A.J. Smola, P.L. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 39–50. MIT Press, Cambridge, MA, 2000.
- [325] E. A. Weathers, M. E. Paulaitis, T. B. Woolf, and J. H. Hoh. Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. *FEBS Lett.*, 576(3):348–352, 2004.
- [326] J. Weston, F. Pérez-Cruz, O. Bousquet, O. Chapelle, A. Elisseeff, and B. Schölkopf. Feature selection and transduction for prediction of molecular bioactivity for drug design. *Bioinformatics*, 19(6):764–771, 2003.

- [327] R.D. Williams, S.N. Hing, B.T. Greer, C.C. Whiteford, J.S. Wei, R. Natrajan, A. Kelsey, S. Rogers, C. Campbell, K. Pritchard-Jones, and J. Khan. Prognostic classification of relapsing favorable histology Wilms tumor using cDNA microarray expression profiling and support vector machines. *Genes Chromosomes Cancer*, 41(1):65–79, Sep 2004.
- [328] D. Wilton, P. Willett, K. Lawson, and G. Mullier. Comparison of ranking methods for virtual screening in lead-discovery programs. *J Chem Inf Comput Sci*, 43(2):469–74, 2003.
- [329] S. Winters-Hilt, W. Vercoutere, V.S. DeGuzman, D. Deamer, M. Akeson, and D. Haussler. Highly accurate classification of Watson-Crick basepairs on termini of single DNA molecules. *Biophys. J.*, 84(2):967–976, 2003.
- [330] B. Wu, T. Abbott, D. Fishman, W. McMurray, G. Mor, K. Stone, D. Ward, K. Williams, and H. Zhao. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics*, 19(13):1636–1643, 2003.
- [331] Dan Xie, Ao Li, Minghui Wang, Zhewen Fan, and Huanqing Feng. LOCSVMPSI: a web server for subcellular localization of eukaryotic proteins using SVM and profile of PSI-BLAST. *Nucleic Acids Res.*, 33(Web Server issue):W105–10, Jul 2005.
- [332] M. Xiong, X. Fang, and J. Zhao. Biomarker Identification by Feature Wrappers. *Genome Res.*, 11(11):1878–1887, 2001.
- [333] Xiu-Qin Xu, Chon K Leow, Xin Lu, Xuegong Zhang, Jun S Liu, Wing-Hung Wong, Arndt Asperger, Sren Deininger, and Hon-Chiu Eastwood Leung. Molecular classification of liver cirrhosis in a rat model by proteomics and bioinformatics. *Proteomics*, 4(10):3235–45, Oct 2004.
- [334] C. X. Xue, R. S. Zhang, H. X. Liu, M. C. Liu, Z. D. Hu, and B. T. Fan. Support vector machines-based quantitative structure-property relationship for the prediction of heat capacity. *J Chem Inf Comput Sci*, 44(4):1267–74, 2004.
- [335] C. X. Xue, R. S. Zhang, H. X. Liu, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. An accurate QSPR study of O-H bond dissociation energy in substituted phenols based on support vector machines. *J Chem Inf Comput Sci*, 44(2):669–77, 2004.
- [336] C. X. Xue, R. S. Zhang, H. X. Liu, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan. QSAR models for the prediction of binding affinities to human serum albumin using the heuristic method and a support vector machine. *J Chem Inf Comput Sci*, 44(5):1693–700, 2004.
- [337] C. X. Xue, R. S. Zhang, M. C. Liu, Z. D. Hu, and B. T. Fan. Study of the quantitative structure-mobility relationship of carboxylic acids in

capillary electrophoresis based on support vector machines. *J Chem Inf Comput Sci*, 44(3):950–7, 2004.

- [338] Y. Xue, Z. R. Li, C. W. Yap, L. Z. Sun, X. Chen, and Y. Z. Chen. Effect of molecular descriptor feature selection in support vector machine classification of pharmacokinetic and toxicological properties of chemical agents. *J Chem Inf Comput Sci*, 44(5):1630–8, 2004.
- [339] Y. Xue, C. W. Yap, L. Z. Sun, Z. W. Cao, J. F. Wang, and Y. Z. Chen. Prediction of P-glycoprotein substrates by a support vector machine approach. *J Chem Inf Comput Sci*, 44(4):1497–505, 2004.
- [340] Y. Yabuki, T. Muramatsu, T. Hirokawa, H. Mukai, and M. Suwa. GRIF-FIN: a system for predicting GPCR-G-protein coupling selectivity using a support vector machine and a hidden Markov model. *Nucleic Acids Res.*, 33(Web Server issue):W148–53, Jul 2005.
- [341] Y. Yamanishi, J.-P. Vert, and M. Kanehisa. Heterogeneous data comparison and gene selection with kernel canonical correlation analysis. In B. Scholkopf, K. Tsuda, and J.P. Vert, editors, *Kernel Methods in Computational Biology*, pages 209–230. MIT Press, 2004.
- [342] Y. Yamanishi, J.-P. Vert, and M. Kanehisa. Protein network inference from multiple genomic data: a supervised approach. *Bioinformatics*, 20:i363–i370, 2004.
- [343] Y. Yamanishi, J.-P. Vert, and M. Kanehisa. Supervised enzyme network inference from the integration of genomic data and chemical information. *Bioinformatics*, 21:i468–i477, 2005.
- [344] Y. Yamanishi, J.-P. Vert, A. Nakaya, and M. Kanehisa. Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis. *Bioinformatics*, 19(Suppl. 1):i323–i330, 2003.
- [345] C. Yan, D. Dobbs, and V. Honavar. A two-stage classifier for identification of protein-protein interface residues. *Bioinformatics*, 20(Suppl. 1):i371–i378, 2004.
- [346] C. Yan, V. Honavar, and D. Dobbs. Identification of interface residues in protease-inhibitor and antigen-antibody complexes: a support vector machine. *Neural Comput. & Applic.*, 13:123–129, 2004.
- [347] Z. R. Yang and K.-C. Chou. Bio-support vector machines for computational proteomics. *Bioinformatics*, 20(5):735–741, 2004.
- [348] X. J. Yao, A. Panaye, J. P. Doucet, R. S. Zhang, H. F. Chen, M. C. Liu, Z. D. Hu, and B. T. Fan. Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression. *J Chem Inf Comput Sci*, 44(4):1257–66, 2004.

- [349] C. W. Yap, C. Z. Cai, Y. Xue, and Y. Z. Chen. Prediction of torsade-causing potential of drugs by support vector machine approach. *Toxicol Sci*, 79(1):170–7, May 2004.
- [350] C. W. Yap and Y. Z. Chen. Prediction of Cytochrome P450 3A4, 2D6, and 2C9 Inhibitors and Substrates by Using Support Vector Machines. *J Chem Inf Model*, 45(4):982–92, 2005.
- [351] C.H. Yeang, S. Ramaswamy, P. Tamayo, S. Mukherjee, R.M. Rifkin, M. Angelo, M. Reich, E. Lander, J. Mesirov, and T. Golub. Molecular classification of multiple tumor types. *Bioinformatics*, 17(Suppl. 1):S316–S322, 2001.
- [352] S. M. Yiu, Prudence W. H. Wong, T.W. Lam, Y.C. Mui, H. F. Kung, Marie Lin, and Y. T. Cheung. Filtering of Ineffective siRNAs and Improved siRNA Design Tool. *Bioinformatics*, 21(2):144–151, Jan 2005. To appear.
- [353] Y. Yoon, J. Song, S.H. Hong, and J.Q. Kim. Analysis of multiple single nucleotide polymorphisms of candidate genes related to coronary heart disease susceptibility by using support vector machines. *Clin. Chem. Lab. Med.*, 41(4):529–534, 2003.
- [354] C. Yu, N. Zavaljevski, F. J. Stevens, K. Yackovich, and J. Reifman. Classifying noisy protein sequence data: a case study of immunoglobulin light chains. *Bioinformatics*, 21(Suppl 1):i495–i501, Jun 2005.
- [355] C.-S. Yu, C.-J. Lin, and J.-K. Hwang. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.*, 13(5):1402–1406, 2004.
- [356] C.S. Yu, J.Y. Wang, J.M. Yang, P.C. Lyu, C.J. Lin, and J.K. Hwang. Fine-grained protein fold assignment by support vector machines using generalized npeptide coding schemes and jury voting from multiple-parameter sets. *Proteins*, 50(4):531, 6 2003.
- [357] J.-k. Yu, S. Zheng, Y. Tang, and L. Li. An integrated approach utilizing proteomics and bioinformatics to detect ovarian cancer. *J Zhejiang Univ Sci B*, 6(4):227–31, Apr 2005.
- [358] J. S. Yu, S. Ongarello, R. Fiedler, X. W. Chen, G. Toffolo, C. Cobelli, and Z. Trajanoski. Ovarian cancer identification based on dimensionality reduction for high-throughput mass spectrometry data. *Bioinformatics*, 21(10):2200–9, May 2005.
- [359] J.K. Yu, Y.D. Chen, and S. Zheng. An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics. *World J. Gastroenterol.*, 10(21):3127–3131, 2004.

- [360] Z. Yuan, K. Burrage, and J.S. Mattick. Prediction of protein solvent accessibility using support vector machines. *Proteins*, 48(3):566–570, 2002.
- [361] Z. Yuan, J.S. Mattick, and R.D. Teasdale. SVMtm: support vector machines to predict transmembrane segments. *J. Comput. Chem.*, 25(5):632, 6 2004.
- [362] N. M. Zaki, S. Deris, and R. Illias. Application of string kernels in protein sequence classification. *Appl. Bioinformatics*, 4(1):45–52, 2005.
- [363] N. Zavaljevski, F.J. Stevens, and J. Reifman. Support vector machines with selective kernel scaling for protein classification and identification of key amino acid positions. *Bioinformatics*, 18(5):689–696, 2002.
- [364] V. V. Zernov, K. V. Balakin, A. A. Ivaschenko, N. P. Savchuk, and I. V. Pletnev. Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions. *J Chem Inf Comput Sci*, 43(6):2048–56, 2003.
- [365] Qidong Zhang, Sukjoon Yoon, and William J Welsh. Improved method for predicting beta-turn using support vector machine. *Bioinformatics*, 21(10):2370–4, May 2005.
- [366] S.-W. Zhang, Q. Pan, H.-C. Zhang, Y.-L. Zhang, and H.-Y. Wang. Classification of protein quaternary structure with support vector machine. *Bioinformatics*, 19(18):2390–2396, 2003.
- [367] X. H-F. Zhang, K. A. Heller, I. Hefter, C. S. Leslie, and L. A. Chasin. Sequence Information for the Splicing of Human Pre-mRNA Identified by Support Vector Machine Classification. *Genome Res.*, 13(12):2637–2650, 2003.
- [368] Z. Zhang, S. Kochhar, and M. G. Grigorov. Descriptor-based protein remote homology identification. *Protein Sci.*, 42(2):431–444, 2005.
- [369] Y. Zhao, C. Pinilla, D. Valmori, R. Martin, and R. Simon. Application of support vector machines for T-cell epitopes prediction. *Bioinformatics*, 19(15):1978–1984, 2003.
- [370] GuoDong Zhou, Dan Shen, Jie Zhang, Jian Su, and SoonHeng Tan. Recognition of protein/gene names from text using an ensemble of classifiers. *BMC Bioinformatics*, 6 Suppl 1:S7, 2005.
- [371] Xin Zhou and K. Z. Mao. LS Bound based gene selection for DNA microarray data. *Bioinformatics*, 21(8):1559–64, Apr 2005.
- [372] A. Zien, G. Rätsch, S. Mika, B. Schölkopf, T. Lengauer, and K.-R. Müller. Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics*, 16(9):799–807, 2000.