Prof. Ron Shamir
School of Computer Science
Tel Aviv University
March 2006

Motif Finding References

References

[1] T. Bailey and C. Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology (ISMB 1994)*, pages 28–36, 1994.

[2] T.L. Bailey and C. Elkan. Unsupervised learning of multiple motifs in biopolymers using expectation maximization. *Machine Learning*, 21(1-2):51–80, 1995.

[3] J. A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimatio for Gaussian mixture and hidden Markov models. Technical report, International Computer Science Institute, Berkeley CA, 1998.

[4] M. Blanchette, B. Schwikowski, and M. Tompa. Algorithms for phylogenetic footprinting. *Journal of Computational Biology*, 9(2):211–223, 2002.

[5] J. Buhler and M. Tompa. Finding motifs using random projections. *Journal of Computational Biology*, 9(2):225–242, 2002.

[6] R. Elkon, C. Linhart, R. Sharan, R. Shamir, and Y. Shiloh. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Research*, 13(5):773–780, 2003.

[7] D. Gusfield. *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*. Cambridge University Press, 1997.

[8] G. Z. Hertz, G. W. Hartzell III, and Gary D. Stormo. Identification of consensus patterns in unaligned DNA sequences known to be functionally related. *Computer Applications in the Biosciences*, 6(2):81–92, 1990.

[9] U. Keich and P. Pevzner. Finding motifs in the twilight zone. *Bioinformatics*, 18(10):1374–1381, 2002.

[10] C.E. Lawrence, S.F. Altschul, M.S. Boguski, J.S. Liu, A.F. Neuwald, and J.C. Wootton. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*, 262:208–214, 1993.

[11] C. Linhart and R. Shamir. The degenerate primer design problem: Theory and applications. *Journal of Computational Biology*, 12(4):431–456, 2005.

[12] G. Pavesi, G. Mauri, and G. Pesole. An algorithm for finding signals of unknown length in DNA sequences. *Bioinformatics*, 17, Suppl 1:S207–S214, 2001.

[13] P.A. Pevzner and S. Sze. Combinatorial approaches to finding subtle signals in DNA sequences. In *Proc. 8th International Conference on Intelligent Systems for Molecular Biology (ISMB 2000)*, pages 269–278, 2000.

[14] A. Prakash, Blanchette M., S. Sinha, and M. Tompa. Motif discovery in heterogeneous sequence data. In *Proceedings of the 9th Pacific Symposium on Biocomputing (PSB 2004)*, pages 348–359, 2004.

[15] E. Rocke and M. Tompa. An algorithm for finding novel gapped motifs in DNA sequences. In *RECOMB*, pages 228–233, 1998.

[16] FP. Roth, JD. Hughes, PW. Estep, and GM. Church. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat Biotechnol*, 16:939–945, 1998.

[17] S. Sinha, M. Blanchette, and M. Tompa. PhyME: A probabilistic algorithm for finding motifs in sets of orthologous sequences. *BMC Bioinformatics*, 5:170, 2004.

[18] S. Sinha and M. Tompa. Discovery of novel transcription factor binding sites by statistical overrepresentation. *Nucleic Acids Research*, 30(24):5549–5560, 2002.

[19] G. Stormo. DNA binding sites: representation and discovery. *Bioinformatics*, 16:16–23, 2000.

[20] M. Tompa, N. Li, T. L. Bailey, G. M. Church, B. De Moor, E. Eskin, A. V. Favorov, M. C. Frith, Y. Fu, W. J. Kent, V. J. Makeev, A. A. Mironov, W. S. Noble, G. Pavesi, G. Pesole, M. Regnier, N. Simonis, S. Sinha, G. Thijs, J. van Helden, M. Vandenbogaert, Z. Weng, C. Workman, C. Ye, and Z. Zhu. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol*, 23(1):137–44, 2005.

[21] T. Vavouri and G. Elgar. Prediction of cis-regulatory elements using binding site matrices–the successes, the failures and the reasons for both. *Current Opinion in Genetics and Development*, 15:1–8, 2005.

[22] C. Workman and G. Stormo. ANN-Spec: a method for discovering transcription factor binding sites with improved specificity. In *Proceedings of the 5th Pacific Symposium on Biocomputing (PSB 2000)*, pages 467–478, 2000.