CS 260-1: “Pattern Discovery in Biosequences”
Reading list

Stefano Lonardi (stelo@cs.ucr.edu)
January 16, 2003

General References

Books

• Developing Bioinformatics Computer Skills by Cynthia Gibas, Per Jambeck; O’Reilly, 2001.
• Jason Wang, Bruce A. Shapiro, and Dennis Shasha, Pattern Discovery in Biomolecular Data Tools, Techniques, and Applications, Oxford University Press, 1999.

Papers

• Brona Brejova, Chryssanne DiMarco, Tomas Vinar, Sandra Romero Hidalgo, Gina Holguin, Cheryl Patten. ”Finding Patterns in Biological Sequences”. Unpublished TR. University of Waterloo, 2000 (available on the class website)
1 Motifs discovery

- Emotif http://motif.stanford.edu/emotif/
- Verbumculus http://www.cs.ucr.edu/~stelo/Verbumculus/
- Pratt http://www.ii.uib.no/~inge/Pratt.html
- Chapter 4 of Jason Wang, Bruce A. Shapiro, and Dennis Shasha, *Pattern Discovery in Biomolecular Data Tools, Techniques, and Applications*, Oxford University Press, 1999.
1.1 Motif discovery by Gibbs sampling

- Gibbs sampler http://bayesweb.wadsworth.org/gibbs/gibbs.html

1.2 Motif discovery by expectation maximization

- MEME http://meme.sdsc.edu/meme/website/
- Meta MEME http://metameme.sdsc.edu/
- Chapter 3 of Jason Wang, Bruce A. Shapiro, and Dennis Shasha, Pattern Discovery in Biomolecular Data Tools, Techniques, and Applications, Oxford University Press, 1999.

2 DNA segmentation


3 Finding genes

• Glimmer (TIGR) http://www.tigr.org/softlab/glimmer/glimmer.html
• GenScan http://genes.mit.edu/GENSCAN.html
• HMMGene http://www.cbs.dtu.dk/services/HMMgene/
• Grail http://compbio.ornl.gov/Grail-1.3/
• GenMark.hmm http://dixie.biology.gatech.edu/GeneMark/gmm2_prok.cgi
• VEIL http://www.tigr.org/~salzberg/veil.html
• John S. Chuang and Dan Roth, “Gene recognition based on DAG shortest paths”, Bioinformatics 2001 17: S56-S64

4 Finding regulatory elements

• PromoterInspector http://genomatix.gsf.de/cgi-bin/promoterinspector/promoterinspector.pl
• Promoter http://www.cbs.dtu.dk/services/promoter/
• Uwe Ohler, “Promoter prediction on a genomic scale - the Adh experience”, Genome Research, 10(4):539-542, 2000.
• WW Wasserman, M Palumbo, W Thompson, JW Fickett, CE Lawrence, ”Human-mouse genome comparisons to locate regulatory sites”, Nature Genetics, 26:225-227, 2000.

5 Finding splicing sites

• Splice Predication using ANN http://www.fruitfly.org/seq_tools/splice.html
• ISIS: http://isis.bit.uq.edu.au/a_splicers.html
6 Statistical analysis of gene expression data

- http://linkage.rockefeller.edu/wli/microarray/ is an excellent starting point for resources on micro array literature, groups, companies, etc.
- Jeffrey G Thomas, James M Olson, Stephen J Tapscott, Lue Ping Zhao (2001), "An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles", Genome Research, 11:1227-1236.

6.1 Clustering

• AV Lukashin, R Fuchs (2001), ”Analysis of temporal gene expression profiles: clustering by simulated annealing and determining the optimal number of clusters”, Bioinformatics, 17:405-414.

7 Protein classification

• SCOP http://scop.mrc-lmb.cam.ac.uk/scop/
• CATH http://www.biochem.ucl.ac.uk/bsm/cath_new/
• Blocks http://blocks.fhcrc.org/

8 Finding MARs

• MansFinder http://www.futuresoft.org/MAR-Wiz/

9 Compression & sequence analysis

• Chapter 1 of Jason Wang, Bruce A. Shapiro, and Dennis Shasha, Pattern Discovery in Biomolecular Data Tools, Techniques, and Applications, Oxford University Press, 1999.

10 Mining PubMed