Readings

COURSE HOME
SYLLABUS
CALENDAR
THIS COURSE AT MIT
READINGS
LECTURE SLIDES
VIDEO LECTURES
RECITATIONS
ASSIGNMENTS

This textbook is recommended for the course:
Buy at Amazon Zvelebil, Marketa J., and Jeremy O. Baum. *Understanding Bioinformatics*. Garland Science, 2007. ISBN: 9780815340249. [Preview with Google books]

The instructors have also selected various texts as particularly useful in specific areas, if you are looking for more information. See the textbook section on the syllabus.

LEC #	TOPICS	READINGS
1	Course Introduction; Overview	No readings for this lecture.
2	Local Alignment; Statistics	National Center for Biotechnology Information. "The Statistics of Sequence Similarity Scores." BLAST Tutorial. Metzker, Michael L. "Sequencing Technologies—The Next Generation." *Nature Reviews Genetics* 11, no. 1 (2010): 31–46.
3	Global Alignment of Protein Statistics	No readings for this lecture.
		Sabeti, P. C., S. F. Schaffner, et al. "Positive Natural Selection in the Human Lineage." *Science* 312, no. 5780 (2006): 1614–20. Read the first three pages. Bejerano, Gill, Michael Pheasant, et al. "Ultraconserved Elements in the Human Genome." *Science* 304, no. 5675 (2004): 1321–5. Pennacchio, Len A., Nadav Ahituv, et al. "In Vivo Enhancer Analysis of Human

PROJECT

TOOLS

DOWNLOAD COURSE MATERIALS

4	Comparative Genomics	Conserved Non–coding Sequences." *Nature* 444, no. 7118 (2006): 499–502. Visel, Axel, Shyam Prabhakar, et al. "Ultraconservation Identifies a Small Subset of Extremely Constrained Developmental Enhancers." *Nature Genetics* 40, no. 2 (2008): 158–60. Bejerano, Gill, Craig B. Lowe, et al. "A Distal Enhancer and an Ultraconserved Exon are Derived from a Novel Retroposon." *Nature* 441, no. 7089 (2006): 87–90. Lareau, Liana F., Maki Inada, et al. "Unproductive Splicing of SR Genes Associated with Highly Conserved and Ultraconserved DNA Elements." *Nature* 446, no. 7138 (2007): 926–9. Lewis, Benjamin P., I–hung Shih, et al. "Prediction of Mammalian MicroRNA Targets." *Cell* 115, no. 7 (2003): 787–98. Lewis, Benjamin P., Christopher B. Burge, et al. "Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets." *Cell* 120, no. 1 (2005): 15–20. Kheradpour, Pouya, Alexander Stark, et al. "Reliable Prediction of Regulator Targets Using 12 Drosophila Genomes." *Genome Research* 17, no. 12 (2007): 1919–31. Friedman, Robin C., Kyle Kai–How Farh, et al. "Most Mammalian mRNAs are Conserved Targets of MicroRNAs." *Genome Research* 19, no. 1 (2009): 92–105. Graveley, Brenton R. "Mutually Exclusive Splicing of the Insect Dscam Pre–mRNA Directed by Competing Intronic RNA Secondary Structures." *Cell* 123, no. 1 (2005): 65–73. Jansen, Ruud, Jan Embden, et al. "Identification of Genes that are Associated with DNA Repeats in Prokaryotes." *Molecular Microbiology* 43, no. 6 (2002): 1565–75. Bolotin, Alexander, Benoit Quinquis, et al. "Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) have Spacers of Extrachromosomal Origin." *Microbiology* 151, no. 8 (2005): 2551–61.
5	Read Alignment	Langmead, Ben, Cole Trapnell, et al. "Ultrafast and Memory–efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10, no. 3 (2009): R25. Li, Heng, and Richard Durbin. "Fast and Accurate Short Read Alignment with Burrows–wheeler Transform." *Bioinformatics* 25, no. 14 (2009): 1754–60. Trapnell, Cole, and Steven L. Salzberg. "How to Map Billions of Short Reads onto Genomes." *Nature Biotechnology* 27, no. 5 (2009): 455. Burrows–Wheeler Aligner

		Bowtie: An ultrafast memory–efficient short read aligner
6	Genome Assembly	Simpson, Jared T., and Richard Durbin. "Efficient De Novo Assembly of Large Genomes Using Compressed Data Structures." *Genome Research* 22, no. 3 (2012): 549–56. Zerbino, Daniel R., and Ewan Birney. "Velvet: Algorithms for De Novo Short Read Assembly Using De Bruijn Graphs." *Genome Research* 18, no. 5 (2008): 821–9.
7	ChIP-seq / IDR	Guo, Yuchun, Georgios Papachristoudis, et al. "Discovering Homotypic Binding Events at High Spatial Resolution." *Bioinformatics* 26, no. 24 (2010): 3028–34. Li, Qunhua, James B. Brown, et al. "Measuring Reproducibility of High–throughput Experiments." *The Annals of Applied Statistics* 5, no. 3 (2011): 1752–79.
8	RNA–seq Analysis	Trapnell, Cole, Brian A. Williams, et al. "Transcript Assembly and Quantification by RNA–seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation." *Nature Biotechnology* 28, no. 5 (2010): 511–5. Anders, Simon, and Wolfgang Huber. "Differential Expression Analysis for Sequence Count Data." *Genome Biology* 11, no. 10 (2010): R106. Licensed under CC–BY. Wang, Zhong, Mark Gerstein, et al. "RNA–Seq: a Revolutionary Tool for Transcriptomics." *Nature Reviews Genetics* 10, no. 1 (2009): 57–63. Shalek, Alex K., Rahul Satija, et al. "Single–cell Transcriptomics Reveals Bimodality in Expression and Splicing in Immune Cells." *Nature* 498 (2013): 236–40. Smith, Lindsay I. "A Tutorial on Principal Components Analysis." (PDF) February 26, 2002.
9	Modeling and Discovery of Sequence Motifs (Gibbs Sampler, Alternatives)	D'haeseleer, Patrik. "What are DNA Sequence Motifs?" *Nature Biotechnology* 24, no. 4 (2006): 423–25. ———. "How does DNA Sequence Motif Discovery Work?" *Nature Biotechnology* 24, no. 8 (2006): 959–61. Eddy, Sean R. "What is Bayesian Statistics?" *Nature Biotechnology* 22, no. 9 (2004): 1177–8. Bailey, Timothy L., and Charles Elkan. "Unsupervised Learning of Multiple Motifs in

		Biopolymers Using Expectation Maximization." *Machine Learning* 21, no. 1–2 (1995): 51–80. Lawrence, Charles E., Stephen F. Altschul, et al. "Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment." *Science* 262, no. 5131 (1993): 208–14.
10	Markov and Hidden Markov Models	Eddy, Sean R. "What is a Hidden Markov Model?" *Nature Biotechnology* 22, no. 10 (2004): 1315–6. Rabiner, Lawrence. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE* 77, no. 2 (1989): 257–86.
11	RNA Secondary Structure Prediction	Eddy, Sean R. "How do RNA Folding Algorithms Work?" *Nature Biotechnology* 22, no. 11 (2004): 1457–8.
12	Introduction to Protein Structure	Buy at Amazon Scheeff, Eric D., and J. Lynn Fink. "Fundamentals of Protein Structure." In *Structural Bioinformatics*. Edited by Philip E. Bourne and Helge Weissig. Wiley–Liss, 2003, pp. 15–39. [Preview with Google Books]
13	Predicting Protein Structure	Moretti, Rocco, Sarel J. Fleishman, et al. "Community-wide Evaluation of Methods for Predicting the Effect of Mutations on Protein–protein Interactions." *Proteins: Structure, Function, and Bioinformatics* 81, no. 11 (2013): 1980–7.
14	Predicting Interactions	Tuncbag, Nurcan, Attila Gursoy, et al. "Predicting Protein–protein Interactions on a Proteome Scale by Matching Evolutionary and Structural Similarities at Interfaces Using PRISM." *Nature Protocols* 6, no. 9 (2011): 1341–54. Zhang, Qiangfeng Cliff, Donald Petrey, et al. "Structure–based Prediction of Protein–protein Interactions on a Genome–wide Scale." *Nature* 490, no. 7421 (2012): 556–60. Jansen, Ronald, Haiyuan Yu, et al. "A Bayesian Networks Approach for Predicting Protein–protein Interactions from Genomic Data." *Science* 302, no. 5644 (2003): 449–53.
15	Gene Regulatory Networks	Marbach, Daniel, James C. Costello, et al. "Wisdom of Crowds for Robust Gene Network Inference." *Nature Methods* 9, no. 8 (2012): 796–804.
16	Protein Interaction Networks	No readings for this lecture.
		Morris, Melody K., Julio Saez–Rodriguez, et al. "Logic–based Models for the Analysis of

17	Logic Modeling of Cell Signaling Networks. Guest Lecture: Doug Lauffenburger	Cell Signaling Networks." *Biochemistry* 49, no. 15 (2010): 3216–24. Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al . "Discrete Logic Modelling as a Means to Link Protein Signalling Networks with Functional Analysis of Mammalian Signal Transduction." *Molecular Systems Biology* 5, no. 1 (2009): 331.
18	Analysis of Chromatin Structure	Hoffman, Michael M., Orion J. Buske, et al. "Unsupervised Pattern Discovery in Human Chromatin Structure through Genomic Segmentation." *Nature Methods* 9, no. 5 (2012): 473–6. Zhou, Vicky W., Alon Goren, et al. "Charting Histone Modifications and the Functional Organization of Mammalian Genomes." *Nature Reviews Genetics* 12, no. 1 (2010): 7–18. Sherwood, Richard I., Tatsunori Hashimoto, et al. "Discovery of Directional and Nondirectional Pioneer Transcription Factors by Modeling DNase Profile Magnitude and Shape." *Nature Biotechnology* 32, no. 2 (2014): 171–8. Dostie, Josée, and Job Dekker. "Mapping Networks of Physical Interactions between Genomic Elements Using 5C Technology." *Nature Protocols* 2, no. 4 (2007): 988–1002.
19	Discovering Quantitative Trait Loci (QTLs)	Bloom, Joshua S., Ian M. Ehrenreich, et al. "Finding the Sources of Missing Heritability in a Yeast Cross." *Nature* 494, no. 7436 (2013): 234–7. Buy at Amazon Broman, Karl W., and Saunak Sen. "Single–QTL Analysis." Chapter 4 in *A Guide to QTL Mapping with R/qtl*. Springer, 2009. ISBN: 9780387921242. [Preview with Google Books]
20	Genome Wide Associate Studies	Li, Heng. "A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data." *Bioinformatics* 27, no. 21 (2011): 2987–93. Roberts, Nicholas J., Joshua T. Vogelstein, et al. "The Predictive Capacity of Personal Genome Sequencing." *Science Translational Medicine* 4, no. 133 (2012): 133ra58. 1000 Genomes. "Variant Call Format." Goldstein, David B., Andrew Allen, et al. "Sequencing Studies in Human Genetics: Design and Interpretation." *Nature Reviews Genetics* 14, no. 7 (2013): 460–70.
21	Synthetic Biology: From Parts to Modules to Therapeutic Systems.	No readings for this lecture.

	Guest Lecture: Ron Weiss	
22	Causality, Natural Computing, and Engineering Genomes. Guest Lecture: George Church	No readings for this lecture.

FIND COURSES

- » Find by Topic
- » Find by Course Number
- » Find by Department
- » Audio/Video Courses
- » Courses with Subtitles
- » Online Textbooks
- » New Courses
- » Most Visited Courses
- » OCW Scholar Courses
- » This Course at MIT
- » Supplemental Resources
- » Translated Courses
- » View All Courses

ABOUT

- » About OpenCourseWare
- » Site Stats
- » OCW Stories
- » Media Coverage
- » Newsletter
- » Press Releases

DONATE

- » Make a Donation
- » Why Donate?
- » Our Supporters
- » Other Ways to Contribute
- » Shop OCW
- » Become a Corporate Sponsor

FEATURED SITES

- » Highlights for High School
- » OCW Educator
- » MIT Crosslinks and OCW
- » MITx Courses on edX
- » Teaching Excellence at MIT
- » Open Education Consortium

TOOLS

- » Help & FAQs
- » Contact Us
- » Advanced Search
- » Site Map
- » Privacy & Terms of Use
- » RSS Feeds

ABOUT MIT OPENCOURSEWARE

MIT OpenCourseWare makes the materials used in the teaching of almost all of MIT's subjects available on the Web, free of charge. With more than 2,200 courses available, OCW is delivering on the promise of open sharing of knowledge. Learn more »

Don't show me this again

Need help getting started?