

SYLLABUS
STAT 843: Next-Generation Sequencing and Systems Biology
October 23, 2015

Undergraduates are encouraged to take this course with permission!

Instructor: Istvan (Steve) Ladunga, Ph.D.

Professor of Computational Biology

Department of Statistics, also affiliated with the Department of Biochemistry and the School of Biological Sciences

346C Hardin Hall NORTH WING, Phone: (402) 472-6074, sladunga@unl.edu

Credit hours: 3.

Term Offered: Every Spring Semester.

Room: Tuesdays 12:30 - 1:45 PM, Room N172, Beadle Center, City Campus

**Thursdays 2:30 - 4:00 PM, Room 142 HARDIN HALL, East Campus Please note the time
and rom change!**

Course Prerequisites:

ANY introductory course in biology or statistics:

- STAT/BIOC 442/842 Computational Biology OR
- STAT 218 Introduction to Statistics OR
- BIOS 101 General Biology OR
- BIOS 103 Organismic Biology OR
- BIOS 206 General Genetics OR
- BIOC 321 Elements of Biochemistry OR
- BIOS 427 Practical Bioinformatics Laboratory OR
- BIOC 432/832 Gene Expression and Replication OR
- BIOS 477/877 Bioinformatics and Molecular Evolution OR

The PowerPoint presentations from 2015 will be made available from BlackBoard->Course Contents->Presentations. These presentations will be updated as we go.

The scope of the course. RNA sequencing, the primary focus point of the Course, has revolutionized the ways we can study biological and pathological processes in HIGH-THROUGHPUT technologies. RNA-Seq is much more accurate than microarrays, can work on any organism, and you can measure even the activity of genes which are expressed at the lowest levels. By using next-generation sequencing (NGS), new genomes have been sequenced, polymorphisms have been revealed in the human genome with potential in precision medicine. Unknown transcriptomes, complex gene regulatory networks and chromatin structures have been discovered. However, inappropriately designed and misinterpreted experiments may lead to wasting massive funds and efforts.

The major breakthrough of NGS is revealing systems properties such as complex biological

networks. We will learn principles and use practical tools of to understand complex biological systems.

Learning objectives that are measurable during tests and presentations:

- Understand, design and analyze large-scale next-generation sequencing experiments.
- Perform a COMPLETE analysis of an RNA-sequencing study.
- Assemble a high-quality *Bifidobacterium* genome from Illumina short sequencing reads and longer Pacific Biosciences reads.
- Learn to use the LINUX operating system at the beginner's level but be able to submit big data analysis tasks
- Understand and quantify false positives and False Discovery Rate.
- Understand the new paradigms of biological networks including metabolic and signaling pathways, transcriptional regulation.
- Display biological networks using the Cytoscape application.
- Communicate with experimental and computational biologists, computer scientists, statisticians, mathematicians, and engineers.

Who would benefit from taking this course?

This course is designed first of all for biology, agronomy, statistics, mathematics and computer science students. Students in physics, electrical engineering, and chemistry may also find it beneficial. This course is designed to benefit computational and experimental biologists as well as biostatisticians to understand the principles of analyzing biological data, building models and testing hypotheses using computer science paradigms. **This Course does NOT depend on any graduate course.**

Necessary background

- Basic knowledge of molecular biology is necessary but statisticians, computer scientists and mathematicians are exempted from that.
- Students should be able to navigate the Internet, use Microsoft Office including Powerpoint, Word, and Excel.
- *No programming or database skills are required.*

Lecture Topics/Course Outline and Assessment Plan:

I believe that one of the most critical but somewhat overlooked skill is reading, understanding and presenting scientific publications that are reasonably challenging and matching to your background (*e.g.*, biology, statistics, or computer science). Each student will be assigned a book chapter and/or a journal article to present during the computer labs using PowerPoint.

Your final grade will be based on the following scale:

Minimum Percent Score	Grade
97	A+
93	A
90	A-
87	B+
83	B
80	B-
77	C+
73	C
70	C-
<70	F

- 25 percent based on the publication presentations
- 30 percent based on the two test assignments
- 35 percent based on the final, written examination
- 10 percent based on class participation

Methods: a 75-minute lecture and another 2 hours (incl. break) of computer lab per week. High-performance computing will be performed on the LINUX servers and compute nodes of the Holland Computer Center.

OPTIONAL References/Textbooks

Next-generation sequencing and systems biology are among the most rapidly progressing fields of science. By the time a textbook comes out of press, it is already outdated. Do you really want to learn from outdated and frequently plainly useless textbooks?

The PowerPoint presentations, the exercises, and yes, attending to classes and actively participating in them, should be good enough even to receive an A+!

Next-generation sequencing and systems biology are among the fastest developing areas of life sciences. By the time a book leaves print it is already out of date, and this is particularly true for next-generation sequencing. Since we do not want to be dinosaurs, the Course will be taught on the basis of recent scientific research papers and reviews as well as online tutorials. We will NOT be using any chapters from the following books, it is NOT necessary to purchase any of them.

- **Stuart M. Brown (Editor)(2013) Next-Generation Sequencing and bioinformatics. Cold Spring Harbor Laboratory Press. Hardcover: \$49.18, Kindle: \$29.99.**
- **Hailing Jin and Walter Gassmann (May 10, 2012) RNA Abundance Analysis: Methods and Protocols (Methods in Molecular Biology, Vol. 883). \$86.39.**
- Ali Masoudi-Nejad, Zahra Narimani , Nazanin Hosseinkhan (2013) Next Generation Sequencing and Sequence Assembly: Methodologies and Algorithms (Springer Briefs in Systems Biology) [Paperback] ISBN-10: 1461477255.
- Eberhard O. Voit (2012) A First Course in Systems Biology. Garland Science. ISBN 978-0-8153-4467-4. It is also a very good introduction to biology for computer scientists,

statisticians, mathematicians and engineers.

- Björn H. Junker: *Analysis of Biological Networks* (Wiley Series in Bioinformatics).
- Uri Alon (2007) *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, ISBN 1-58488-642-0 and
- Istvan Ladunga (2010) *Computational Biology of Transcription Factor Binding*. Humana Press/Springer.

We will be using recent review and research papers and some protocols. They will be listed.

Tentative Schedule of lectures and computer labs

1. Jan. 12. How can next-generation sequencing reveal biological systems/networks? Databases for gene expression, next-generation sequencing and gene regulation.
2. Jan. 14. Technology that allows next-generation sequencing to change the life sciences.
3. Jan. 19 Next-generation sequencing principles and experimental design: how to avoid producing massive amounts of useless data and wasting your time and funds? Part I.
4. Jan. 21. Experimental design, Part II.
5. Jan. 26. Introduction to the LINUX operation system.
6. Jan 28. LINUX practice.
7. Feb. 2. Mapping reads to known genomes, theory.
8. Feb. 4. Mapping reads to known genomes under the LINUX operation system, hands-on.
9. Feb. 9. Mapping reads to known genomes, DNA polymorphisms, theory, Part II.
10. Feb. 11, Test No. 1, student presentations
11. Feb. 16-18, Genome Assembly, theory and hands-on
12. Feb 23. RNA sequencing: quantifying transcript levels and changes, theory. The curse (!) of alternative splicing. Part I.
13. Feb. 25. How to avoid wasting months and tens of thousands of dollars? Quality control and using statistical models implemented in DESeq2 and other Bioconductor packages for background correction, normalization, and analyses of differential expression, Part I, hands-on.
14. Mar. 1. Sequencing noncoding RNA. Lessons learned from the Encyclopedia of DNA Elements Project.
15. Mar. 3. Using statistical models implemented in DESeq2 and other Bioconductor packages for background correction, normalization, and analyses of differential expression, Part II, hands-on.
16. Mar. 8. Interpreting RNA sequencing in simpler network context: pathways, Gene Ontology and Gene Set Enrichment Analysis.
17. Mar. 10. RNA-seq, hands-on, Part III. Genome browsers.
18. Mar. 15. General features of biological networks and the regulation of transcription.
19. Mar. 17. Test 2, student presentations.
20. Mar. 20-27. Spring Vacation.

21. Mar. 29. Precision genome editing – CRISPR/CAS9.
22. March 31. Student presentations.
23. Apr. 5 & 7. Introduction to systems (network) biology.
24. Apr. 12. Protein-protein interaction networks.
25. Apr. 14. Analysis and visualization of networks using Cytoscape. Hands-on, Part I.
26. Apr. 19. Analysis and visualization of networks using Cytoscape. Hands-on, Part II.
27. Apr. 21. Regulatory networks. Systems biology: what can we do? Lessons learned from the Encyclopedia of DNA Elements Project.
28. Apr. 26-28. Reviewing what we have learnt, preparation for the final exam.
- 29. Final exam: TBD**

Students with disabilities are encouraged to contact the instructor for a confidential discussion of their individual needs for academic accommodation. It is the policy of the University of Nebraska--Lincoln to provide flexible and individualized accommodation to students with documented disabilities that may affect their ability to fully participate in course activities or to meet course requirements. To receive accommodation services, students must be registered with the Services for Students with Disabilities (SSD) office, 132 Canfield Administration, 472-3787 voice or TTY.