BCHB 533: EMERGING BIOTECHNOLOGIES
ULTRA HIGH THROUGHPUT SEQUENCING
SPRING 2008 2 credits

Professor for the course::
Larry Millstein, Ph.D., J.D.
Department of Biochemistry, Molecular & Cellular Biology
School of Medicine, Georgetown University
lmillstein@millsteintaylor.com
202-363-6740 Phone
202-363-6772 Fax

Summary

In general this course covers emerging technological advances with substantial and important implications for biotechnology. The particular advance(s) to be addressed will be decided upon each time the course is offered. For Spring, 2008 the course will cover ultra high throughput sequencing technologies ("UHTS"), revisiting the topic of Fall 2006. UHTS promise to increase the speed and decrease the cost of DNA sequencing by several orders of magnitude. The course will cover, primarily: the technologies themselves (and the underlying science); their state of development and potential performance; their economics and cost challenges; intellectual property issues; impacts on existing products and markets; new opportunities; ancillary technologies, such as advanced sensors and ultra high capacity informatics systems; and potential societal impacts.

Prerequisites

Advanced/upper level undergraduate or graduate student standing and an understanding of basic nucleic acid structure and chemistry

Credit and Hours

Two credits
Two class hours per week

Schedule and Location

Tuesdays, 6 PM – 8 PM
BSB 3rd Floor Library

Goals

- Understand the technologies, how they work, their strengths and weaknesses.
- Become familiar with the major contributors and their approaches.
- Analyze throughput and cost of present and future technologies.
- Understand reasons for developing the technologies.
- Understand the potential impacts of the technologies.
- Examine and characterize IP relating to the technologies.
- Identify and study efforts at commercializing the technologies.
- Develop a picture of the potential markets for the technologies.
- Identify and understand obstacles to fully exploiting the new technologies.
- Develop a skill set for understanding and assessing existing and emerging biotechnologies in general.

Subject Matter

This semester (Spring 2008) the course will examine emerging technologies for ultra high throughput DNA sequencing. A number of particular aspects of the technologies will be examined and discussed in depth. Some aspects of sequencing technologies that will examined this semester are listed below.

- The underlying technical basis of each technology and its implementation(s).
- The published literature, including published patent applications and issued patents relevant to each of the technologies.
- The history of RNA and DNA sequencing and a review of the historical development of current and emerging technology that will be addressed in the course.
- The current markets for sequencing-related devices and reagents, the main suppliers and the main consumers in the markets, segmentation of the markets, and examination of the size of the market and its segments.

- The performance parameters of currently available technologies and those expected for emerging technologies, including:
  - *Inherent capabilities and limitations (read length, accuracy)*
  - *Required equipment and reagents*
  - *Required personnel and expertise*
  - *Required facilities*
  - *Capacity factors (up time / down time)*
  - *Throughput – bases/run; bases/hr; bases/µg DNA, etc.*
  - *Cost – $ per base; bases per $*
  - *Capital costs, facilities requirements and maintenance expense*
  - *Performance in context*
    - *Cost/project or cost/test*
    - *Overall suitability to different applications*
- The likely short and long term development of these market(s) based on:
  - Evolution of current technologies, and
  - The advent of transformative technologies that improve on current technologies by at least two orders of magnitude in the short term, and more in the longer term.
- Literature on the potential societal impacts of the technologies.

Course Work

- Reading assigned literature, generally technical scientific publications and publications on business aspects of UHTS technologies and their commercialization.
- Attending lectures on the technologies, companies, and related subject matter.
- Participating in class discussions.
- Researching a specific technology and company, and preparing a report thereon.
- Presenting the technology and the company to the class.

Materials

Science publications, business publications, patent documents, company literature, lecture notes, power point presentations, and internet accessible information.

Assignments and Grading

(1) The Report – 60% of the Final Grade

Each student must prepare a report on a technology and a company that is commercializing that technology from a list prepared by Dr. Millstein. Company assignments will be made by Dr. Millstein taking into consideration any preference you may have, if possible. Preferences should be communicated by email to Dr. Millstein before the due date, which will be communicated in class well in advance. Reports are due when noted in the class schedule.

Each report is to describe all of the following:

(a) The technology, its advantages and projected capacity, its potential disadvantages, and its present state of commercial development.

(b) The company, its principal scientists, founders, and investors.

(c) The economics of the technology.

(d) The company's business model and targeted markets.

(e) The company's IP and IP strategy, and how the strategy serves the company's business model.

(f) The student's assessment of the company's competitive edge (or lack thereof), as to both technology and business model, and its likely fate (as to exit strategy or market place success or failure).

Report are to be approximately 30 pages in length (not less than 25 pages nor longer than 50 pages), on 8 1/2 x 11 paper, with 1 inch margins on all sides, using 12 point font and 1.5 line spacing. The report should include meaningful diagrams that depict the technology and tables that set forth the operative parameters of the technology along with benchmark values, including values for throughput and cost.

Grading will be based primary on content, particularly how well the report covers the above-listed subject matter. Grading also will be based on the quality of the student's research and analysis and on the quality of the student's writing.

(2) The Presentation - 25% of the Final Grade

Each student will make a presentation to the class on the same subject matter as their report. The presentation is to describe the key substantial points about the technology being covered – particularly as to its technical

capabilities, state of development, economics at present, and potential for improvement. And also is to cover key aspects of the company involved, including but not limited to key personnel, history, business model, financing sources and rounds, partners and deals,, and stage of commercialization. As the presentations will be done before the due date of the report, students should view the presentation as an opportunity to refine their major points and get feedback from the class and Dr. Millstein. Each presentation must be 25 – 35 minutes. Presentations will be followed by 5 minutes of class discussion, led by Dr. Millstein. The presentation schedule will be assigned by lottery. Students may trade assignments by mutual agreement only in the first week after the assignments are made, and only if all students party to a trade notify Dr. Millstein by the end of that first week.

(3) Class Preparation and Participation - 15% of the Final Grade

Showing up counts. Absence does not. Informed and intelligent participation counts. The rest does not.

Schedule – Spring 2008

All classes meet from 6 to approximately 8 PM on Tuesdays
This is a tentative schedule that likely will be modified to suit the on-going needs of the class.

(**1**)

INTRODUCTION

Introductory discussion of the purposes and topics to be covered by the class, the class schedule, the course report and presentation, and how the course grade will be determined. Materials for the class will be distributed. There will be a very general description of some sequencing techniques, some of the companies to be covered in later lectures, some current applications of these technologies and potential impacts they may have on biotechnology and other fields. There also will be a short discussion of some basic principles of patent and of searching the patent literature using the USPTO web site. This class session may end early.

Read the following to prepare for this and for subsequent classes.

(A) General Background Information

Review, using any good undergraduate biochemistry textbook, each of the following: (a) DNA structure, transcription, and replication; (b) electrophoresis and other commonly employed methods for separating polynucleotides; (c) methods for labeling and for detecting labeled DNAs and RNAs; and (d) cloning, amplifying, and sequencing DNAs. Stryer, *Biochemistry*, is one good text you might use. Its illustrations are especially effective. Read the chapters on DNA and RNA, the introductory chapter on genes, and any chapters or appendices devoted to recombinant DNA techniques, electrophoresis, and sequencing of polynucleotides. Weaver, *Molecular Biology*, is another text you might use, especially, in the 2nd Ed., Chapters 3, 4 and 5.

(B) Specific Reports and Reviews

(1) Glaser, "Pushing Toward a $1,000 Genome;" *Genetic Engineering News* 27 (19): pp 1 and 34-35 (Nov. 2007).

(2) Hall, "Advanced sequencing technologies and their wider impact in microbiology;" *J. Experimental Bio.* 209: 1518 – 1525 (2007).

(3) "Next-Generation Sequencing, Updates in Technology" Supplement to *Genetic Engineering News* 26(12) 1-34 (June2006)

(4) Shendure *et al.*, "Advanced Sequencing Technologies: Methods and Goals;" *Nature Reviews/Genetics* 5: 335-344 (May 2004).

(5) Marziali and Akeson: "New DNA Sequencing Methods," *Ann. Rev. Biomed. Eng.* 3: 195-123 (2001).

(**2**)

HISTORY OF NUCLEIC ACID SEQUENCING METHODS

Lecture on the development of sequencing methods. This lecture will cover the historical context in which the methods were developed, their principles of operation (the underlying science), how they worked (what was involved in using them), and what was achieved by using them. The following methods will be covered: wandering spot sequencing of RNA, Maxam Gilbert sequencing, Plus and Minus Sanger sequencing, one color primer/terminator Sanger sequencing, four color Sanger sequencing, alternative dyes and systems, automation, throughput, and cost analysis. This historical treatment will attempt to convey a sense of the richness in the variety of sequencing techniques that have been developed, and the depth and breadth of current knowledge about the

physics and chemistries of RNA and DNA. The impact of IP on the commercial development of sequencing technology will be discussed as well.

Reading and Preparation:

Copies of the presentation and notes will be provided.

(3)

THE STATE OF THE ART: THE NIH SEQUENCING CENTER

Lecture and discussion on the current standards for DNA sequencing, primarily automated Sanger/chain terminating sequencing by synthesis, especially as implemented on ABI sequencers at the NIH core sequencing facility. The current state of the art technologies utilized by the NIH Intramural Sequencing Center will be discussed. The discussion will cover sequencing methods, instrumentation, information capture and processing, and the economics of sequencing. The programs that utilize the sequencing center will be discussed. The accomplishments of such state of the art facilities will be summarized as well as their limitations. Applications that are beyond the practical limitations of the current state of the art; but, might be feasible with the next generation of sequencers (including those based on emerging technologies) will be sketched out.

Discussion of the current facilities will cover, in particular, a variety of topics relating to operations and costs, including but not limited to (a) the technologies involved in carrying out state of the art methods; (b) the inherent capabilities and limitations (read length, accuracy) of the methods as presently implemented; (c) the equipment and reagents required to carry out the center's sequencing operations; (d) the personnel and costs involved, and the facilities required; (e) capacity factors (bottlenecks, equipment time up/down); (f) throughput – bases/run; bases/hr; bases/μg DNA, *etc.*; (g) IT and other ancillary and downstream requirement; (h) cost, and (i) prospects for incremental improvements to the current technology in the coming years.

Reading and Preparation:

Copies of the presentation and a bibliography will be provided.

(4)

SURVEY OF OTHER CURRENT METHODS

Lectures and discussion on other sequencing methods and platforms, including but not limited to Sequencing by Hybridization (Hyseq); MS sequencing (Sequenom, others); array sequencing and variation mapping (Affymetrix), pyrosequencing, and others. The technologies, available information on their operating parameters, and IP covering them will be addressed.

Reading and Preparation:

Copies of the presentation and notes will be provided at the start of the class. Companies that may be covered include: Applied Biosystems, Li-Cor, Affymetrix, Hyseq, Lynx, and Roche (pyrosequencing). Students should look at the literature on these companies, their technologies, and the US patents that have issued to them, using the companies' web sites, and by going to the USPTO patent database advanced search page and searching for issued patents in which the companies are the named assignees.

(5)

SURVEY OF EMERGING TECHNOLOGIES AND COMPANIES - 1

Lecture and discussion on emerging sequencing technologies. Labs and companies engaged in the development and commercialization of the technologies will be investigated, analyzed, and discussed.

The technologies that may be discussed include those that fall into the following five categories: microelectrophoretic sequencing, sequencing by hybridization, cyclic array sequencing on amplified molecules, cyclic array sequencing on single molecules and real time methods. More particularly, the technologies that may be specifically discussed may include the following: sequencing by synthesis using polonies; sequencing by synthesis using picoliter reactors; sequencing by synthesis on beads; sequencing by ligation addition synthesis on beads; sequencing by subtraction; sequencing by transit through nanopores; sequencing by subtraction; and sequencing by direct visualization, such as sequencing by atomic force microscopy.

Specific labs and companies that *may* be discussed include: 454 & Roche (Jonathan Rothenberg, Marcel Margulies); Solexa; Agencourt (George Church, Gina L. Costa); Microchip Biotechnologies (Roger McIntosh, Stevan Jovanovich); Helicos BioSciences (Steven R. Quake, Stan Lapidus); Diversa (GigaMatrix).

Reading and Preparation:

References (1) through (5) from lecture 1 above.

Specific references on all of the technologies to be covered will be provided, as will a select list of patent references to be reviewed.

(6)

SURVEY OF EMERGING TECHNOLOGIES AND COMPANIES - 2
SURVEY OF MARKETS, COMPANIES, IP, AND OPPORTUNITIES

(a) Continuation of the topics partially covered in the previous class.

(b) Lecture and discussion of the major current applications of sequencing platforms, with an emphasis on scientific principles and on utilitarian purposes. Among the applications that *may* be covered are the following: De novo sequencing; Resequencing; Small genomes; Large genomes; SNP discovery; SNP testing; and Expression profiling.

(c) Lecture and discussion on the markets for sequencing-based applications and their requirements, the costs of R&D to develop commercially significant advantages over current sequencing methods, and the current state of issued and pending patent rights on technologies pertinent to these issues. Among specific topics that *may* be examined in the survey are the following: Performance in context; Cost/project or cost/test; Overall suitability to different applications; IP surveys (who owns what and how much does it cover).

(d) Lecture and discussion on the research and commercial opportunities that cannot be reached by current sequencing platforms.

(7)

TECHNOLOGIES AND COMPANIES - GROUP 1 (1 – 4)

Student presentations and discussion on leading UHTS technology companies. Each presentation will cover a single company. The principal managers, scientists, and investors will be presented and discussed. The core technology of each company will explained. Its technical merits and drawbacks will be outlined. Its economics will be analyzed. The company's IP and apparent IP strategy will be presented. Target markets (business model) will be explored. And a prediction for the company's exit strategy or market place failure or success will be presented and justified. Each company will be covered in a single presentation 20 to 30 minutes long, followed by a discussion period led by Dr. Millstein.

(8)

TECHNOLOGIES AND COMPANIES - GROUP 2 (5 – 8)

Student presentations and discussion on leading UHTS technology companies, as described for the immediately preceding session.

(9)

TECHNOLOGIES AND COMPANIES - GROUP 3 (9 – 12)

Student presentations and discussion on leading UHTS technology companies, as described for the immediately preceding session.

(10)

TECHNOLOGIES AND COMPANIES - GROUP 4 (13 – 16)

Student presentations and discussion on leading UHTS technology companies, as described for the immediately preceding sessions.

(11)

TECHNOLOGIES AND COMPANIES - GROUP 5 (17 – 20)

Student presentations and discussion on leading UHTS technology companies, as described for the immediately preceding sessions.

(12)

UHTS – Microbioversity studied by Roche/454 technology

A user of the Roche/454 Sequencing System will describe the system, its capabilities and limitations, its manner of use in actual practice, what it costs to procure and to use the system, the interaction between the

vendor, the buyer and the user (with a look at what it takes to make a sale), the use of the system to study microbial population in the human gut in health and in disease, and the prospects for using UHTS systems for clinical diagnostic applications.

(13)

BIOINFORMATICS – HANDLING DATA AND ESTABLISHING MEANINGS?

An expert experienced in handling sequencing data flows and analyzing sequencing information will discuss the data handling and analysis challenges the UHTS methods will present, and how they might be met.

(14)

UHTS – USING UHTS TECHNOLOGIES AT NIH

A speaker from NIH will discuss the UHTS platforms in use at the Intramural Sequencing Center, their use and the experience of NIH users of the devices. Impacts of the platforms on research at NIH may be discussed as well.

NOTE: Reports are due the beginning of this class.

(15)

UHTS – Microbioversity studied by Roche/454 technology

A user of the Roche/454 Sequencing System will describe the system, its capabilities and limitations, its manner of use in actual practice, what it costs to procure and to use the system, the interaction between the vendor, the buyer and the user (with a look at what it takes to make a sale), the use of the system to study microbial population in the human gut in health and in disease, and the prospects for using UHTS systems for clinical diagnostic applications.

(16)

WHITHER UHTS AND WITH WHAT EFFECT?

The potential effects of UHTS – both short term and long term – will be discussed, focusing on basic research in biology and biochemistry, pharmaceutical R&D, clinical diagnosis and medicine. There will be time for questions and class discussion.