7.345 Are There Inherent Limits to Our Understanding in Biology? A Challenge and Exploration Based on Diseases of the Nervous System

Instructor: Sepehr Ehsani (ehsani@csail.mit.edu; laboratory of Bonnie Berger)

Spring 2016. Wednesdays 3 pm – 5 pm. (Class day and time are flexible.) Room 68-150.

Molecular biology over the past two decades has experienced significant changes in both methods and understanding, with major technical innovations facilitating diverse breakthroughs. For example, high-throughput techniques and genome sequencing, introduced in the 1990s, have generated vast quantities of data and valuable insights concerning the workings of the cell under normal and disease conditions. The impact of these findings in the context of human disease has been greatest in the case of single-gene disorders (e.g., cystic fibrosis), which in general are relatively rare. However, most common human diseases, ranging from solid tumors (e.g., sarcomas and carcinomas) to cardiovascular, neurodegenerative and neuropsychiatric pathologies, have remained refractory to non-symptomatic therapeutic interventions, mostly because researchers have been unable to identify simple causative mechanisms. In other words, most common diseases have proved to be both heterogeneous in origin and mechanistically complex. Why is this the case, and what is preventing us from reaching an understanding of the pathologies of these disorders -- a scientific understanding that is not merely descriptive but rather founded on mechanism? This course aims to examine current challenges in the field of pathobiology (the study of the molecular and physiological mechanisms of disease). Students will discuss, through detailed analysis of the primary research literature, whether these challenges possess an underlying commonality. For example, have ultimate causes of many diseases remained elusive because of (i) limitations in experimental or computational methodology, (ii) limitations in our ability to interpret complex data, and/or (iii) some unknown facet of the diseases themselves? Can we identify a common thread in the answers to these questions for multiple diseases? In our efforts to answer such questions, might we discover some inherent limitation to human understanding -- a cognitive limitation similar to that which a rodent faces when fruitlessly attempting to learn to navigate a prime-number maze? If the answer is yes, can we do anything to overcome that limitation? If the answer is no, does that mean that there are no upper limits to what science can reveal and to what we can comprehend, e.g., concerning the etiology of a disease? We will focus on disorders of the nervous system, such as neurodegenerative diseases and cancers of the central nervous system. Our discussions will be framed by two general themes: (i) the quantification and meaning of uncertainty in experimental biology and (ii) a potential limit to scientific understanding. The primary goals of this course are for students to enhance their skills in critically evaluating the primary research literature and to think about the relationship between objective realities as typified by experimental data and human cognitive abilities and limits. The course will include a field trip to a computational/theoretical biology laboratory focused on the structures of proteins to observe how theoretical studies of protein structures can help reveal novel facets of pathological protein-protein interactions in neurodegenerative disorders.

COURSE OBJECTIVES

The primary objective of the course is:

- To enhance the study and assessment of recent primary research articles in the field of neuro-pathobiology (the study of the molecular and physiological mechanisms of neuronal diseases) through discussion-based analysis and critique

The secondary objectives are:

- To illustrate the use of pathobiology in understanding normal biology
- To discuss known and unknown complexities of a number of common human diseases, with an emphasis on neuropathology
- To discuss current examples of the need to determine inherent boundaries in our ability to understand and intervene in biological systems

Lastly, it should be noted that in addition to the stated objectives of the course and the material covered, the individual knowledge that students construct throughout the course will be of paramount importance and a main emphasis of the course assignments. In the allegorical words of the late MIT physics professor Victor Weisskopf, "it doesn't matter what we *cover*, it matters what you *discover*."

COURSE FORMAT AND REQUIREMENTS

- Grading: **P/F**; **6 credits**
- The recommended prerequisites for 7.345 are 7.03, 7.05, 7.06 or 7.28.
- This seminar course will consist of twelve 2-hour weekly sessions, and two related assignments: a written essay and an oral presentation. The topics of the assignments will be chosen by students based on their reflections on and interests in the weekly critical reading of the assigned literature (*please see last two bullet points for more details*).
- **Session 1** will consist of student introductions and a primer to the course format and expectations. In **Sessions 2 to 11** (with the exception of **Session 7** which will be a field trip), one student will begin introducing the first article of the session for approximately 10 minutes. Each member of the class is expected to have read the two articles of that session and **prepared two questions to be raised for discussion** after the introduction (questions should be e-mailed to the instructor at least three hours before class). The main focus of the interactive discussions will be to identify the key experiment and control of each study and to determine whether the results and methodology of the article support the conclusions of the paper. Moreover, the usage of style and argumentation in the article will be analyzed. This process will be repeated a second time for the second article of the day. **The two articles will randomly be assigned to class members to lead the discussion each session**. A short presentation will then be given by the instructor to introduce and set up the context of the following week's articles and discussion.
- During the **last 10 minutes of each class**, students will write **1-2 sentences** about what (if any) **new idea** they have discovered during that day's discussion. In addition, 1-2 sentences should be written about what aspect of the discussed articles of that day really stood out. If no aspect of the articles was surprising and/or interesting, that should also be noted. This material will be given to the instructor at the end of the session and returned the following session. The students will be encouraged to use these ideas in choosing the topics for and completing their assignments.
- **Assignment 1** will be a written essay to be handed in **one day before Session 8**. By the end of Session 5, each student should have identified one "uncertain" point in (i) an experiment, (ii) a methodology or (iii) a hypothesis in one of the papers discussed in the course thus far, which they suspect can be "resolved" (i.e., made "less uncertain" had they been the authors of the paper) by better reasoning, more meticulous experimental design, etc. The topics should be e-mailed to the instructor for approval and/or feedback. The essay for Assignment 1 should be a maximum of 4 double-spaced pages (Times New Roman 11 pt., 2-cm margins) and be in the format of a scientific critique of the chosen paper, i.e., it would contain an introduction, a contextual framing of the paper in question, statement of the problem at hand, the proposed solution and future directions (*to be discussed in more detail in class*).The emphasis of the essay should be on point(s) of uncertainty in one or more experiments and the relevant controls in the chosen paper followed by proposed solutions and/or alternatives.
- **Assignment 2** will be an oral presentation in **Session 12** based on the Assignment 1 essay, and should include revisions as per feedback from the instructor about Assignment 1 and refinements based on the literature discussed in Sessions 8 to 11. Each presentation should be 10-15 minutes, and can utilize PowerPoint or be in a chalk-talk format. There will be 5-10 minutes of questions and answers following the presentation. Similar to the essay, the presentation should have an emphasis on uncertainty in experimental design and controls.

SESSION ORGANIZATION

SESSION 1 Feb. 3	**(i) Introduction to course** **(ii) Why are the themes of "limits" and "uncertainty" important in biology?**	*Session 1 will begin with an introduction to the course theme and expectations, its format and the articles chosen for discussion.* *[Index cards handed out for student introductions: please include major, year, biology courses taken, lab experience (if any) and reason(s) for taking this course]* *A general introduction will also be given to (i) current trends in research publications, (ii) the use of statistics and the quantification of uncertainty, (iii) graphical representation of data in manuscripts and (iv) publication venues in biology.*
SESSION 2 Feb. 10	**(i) At present, what are some of the knowns and unknowns in our understanding of the central dogma of molecular biology?**	*Session 2's discussion will revolve around the central dogma of molecular biology, its utilization over the past several decades as a model to explain molecular phenomena within the cell, and recent expansions and modifications which are beginning to cover some of the missing links. The relevance of the new findings will be discussed in the context of disease biology in general.* ***ARTICLE 1:*** "**Genome-wide consequences of deleting any single gene**" [Teng X, Dayhoff-Brannigan M, Cheng WC, Gilbert CE, Sing CN, Diny NL, Wheelan SJ, Dunham MJ, Boeke JD, Pineda FJ, Hardwick JM. *Mol Cell* 2013; 52:485-94] Single gene deletions are usually thought to provide a suitable controlled framework for studying the functions of genes, one gene at a time. This is a critical method for linking genes to proteins and phenotypes. But every deletion appears to have widespread consequences. This article provides a systematic analysis of this issue. ***ARTICLE 2:*** "**Protein synthesis rate is the predominant regulator of protein expression during differentiation**" [Kristensen AR, Gsponer J, Foster LJ. *Mol Syst Biol* 2013; 9:1-12] Proteins perform the vast majority of "work" functions of the cell. However, we do not yet know exactly what factors determine protein synthesis rates. This paper provides a plausible model.

SESSION 3 Feb. 17	**(i) Neurodegeneration: the brain and the aging process**	*In the previous session we covered some aspects of the current state of the central dogma of molecular biology. We will now begin our discussion of the importance of pathobiology, specifically neurodegeneration, in understanding normal biology. The two articles are concerned with two key aspects of neurodegenerative diseaes: aging and unique genetic characteristics of neurons.* ***ARTICLE 1:*** "**Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging**" [Ben-Zvi A, Miller EA, Morimoto RI. *PNAS* 2009; 106:14914-9] Aging is the single most important risk factor for neurodegenerative diseases. However, the aging process is not well understood at the cellular level. This article focuses on protein homeostasis dysfunction in a model of aging. ***ARTICLE 2:*** "**Constitutional aneuploidy in the normal human brain**" [Rehen SK, Yung YC, McCreight MP, Kaushal D, Yang AH, Almeida BS, Kingsbury MA, Cabral KM, McConnell MJ, Anliker B, Fontanoz M, Chun J. *J Neurosci* 2005; 25:2176-80] This article raises a critical issue about gene dosage effects specific to the central nervous system, and therefore potentially of paramount importance to the initiation of neurodegenerative diseases.
SESSION 4 Feb. 24	**(i) Neurodegeneration: "uncertainties" in Alzheimer's disease research**	*We will continue our disease-specific discussion of neurodegeneration with articles on Alzheimer's disease, the most common neurodegenerative disease. What is the current state of knowledge, and what aspects of the disease are the most uncertain at present?* ***ARTICLE 1:*** "**Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study**" [Alzheimer Disease Genetics Consortium. *JAMA Neurol* 2014; 71:1394-404] One of the main challenges in Alzheimer's disease research, as with other common human diseases, is to identify new susceptibility genes and determine the function of known susceptibility genes (or loci). This genome-wide association study attempts to determine the connection between some of the known loci and the age at onset of the disease. ***ARTICLE 2: *long article; discussion to be led by two students**** "**Physical and functional interaction between the α- and γ-secretases: A new model of regulated intramembrane proteolysis**" [Chen AC, Kim S, Shepardson N, Patel S, Hong S, Selkoe DJ. *J*

		Cell Biol 2015; 211:1157-76] One of the main avenues of intervention for Alzheimer's disease has been the modulation of the γ-secretase complex. Clinical results over the past decade, however, have been perplexing. This new study sheds light on how the dominant pathobiological model should be modified.
SESSION 5 Mar. 2	**(i) Neurodegeneration: "uncertainties" in Parkinson's disease research**	*In this session our focus will shift to Parkinson's disease, with an emphasis on the role of alpha-synuclein and recent clinical trial results. In our discussion and critical reading of the articles, we will look for possible common themes with our focus on Alzheimer's disease last session.* ***ARTICLE 1: *long article; discussion to be led by two students**** "**Nonaggregated α-synuclein influences SNARE-dependent vesicle docking via membrane binding**" [Lai Y, Kim S, Varkey J, Lou X, Song JK, Diao J, Langen R, Shin YK. *Biochemistry* 2014; 53:3889-96] Alpha-synuclein is a small lipid-binding protein that has been implicated as a risk factor in familial cases of Parkinson's disease, and has become one of the most studied proteins in general. This is partly due to its enigmatic "disordered" structure. This paper sheds some light on this protein's mechanism of action in neurons. ***ARTICLE 2:*** "**A randomized clinical trial of high-dosage coenzyme Q10 in early Parkinson disease: No evidence of benefit**" [The Parkinson Study Group QE3 Investigators, *JAMA Neurol* 2014; 71:543-52] This is an important clinical trial investigating the potential benefit of an antioxidant in Parkinson's disease patients.
SESSION 6 Mar. 9	**(i) Neurodegeneration: "uncertainties" surrounding the mammalian prion protein (PrP)**	*Our last session on neurodegeneration will turn to prion diseases and the mammalian prion protein, one of the most-studied proteins in the human proteome (similar to alpha-synuclein), and the only known protein-only means of disease transmission in mammals.* ***ARTICLE 1:*** "**Identification of a gene regulatory network associated with prion replication**" [Marbiah MM, Harvey A, West BT, Louzolo A, Banerjee P, Alden J, Grigoriadis A, Hummerich H, Kan HM, Cai Y, Bloom GS, Jat P, Collinge J, Klöhn PC. *EMBO J* 2014, 33:1193-284] Although PrP can self-replicate, its ability to propagate in different cell lines and conditions depend on unknown factors. This paper sheds light on this aspect by identifying extracellular matrix proteins as important aides to PrP propagation.

		ARTICLE 2: "**Axonal prion protein is required for peripheral myelin maintenance**" [Bremer J, Baumann F, Tiberi C, Wessig C, Fischer H, Schwarz P, Steele AD, Toyka KV, Nave KA, Weis J, Aguzzi A. *Nat Neurosci* 2010; 13:310-8] One of the critical questions in prion biology is what the normal function of PrP is. This is an important paper from 2010 that links PrP to myelin maintenance.
SESSION 7 Mar. 16	**Lab visit**	*Session to be held at the computational/theoretical biology lab of Prof. Bonnie Berger at MIT CSAIL. The session is focused on the structures of proteins, with protein misfolding being a common hallmark of neurodegenerative diseases. Discussions with group members will include: What is theoretical biology? Can experimental biology provide ultimate verification of theories in all circumstances? How can theoretical studies of protein structures help reveal novel facets of pathological protein-protein interactions (in neurodegenerative and other diseases)? The session will end with a short presentation about current research on the disordered region of alpha-synuclein.* ***RELEVANT READING:*** "**Compressive genomics for protein databases**" [Daniels NM, Gallant A, Peng J, Cowen LJ, Baym M, Berger B. *Bioinformatics* 2013; 29:i283-90] This paper is an example of a computational biology study that focuses on a new area called 'compressive genomics.' This is a method of dealing with large sequence data that can be applied to genomic, proteomic and other datasets. **[Midterm course feedback; Essay due before Session 8]**
SESSION 8 Mar. 30	**(i) Cancer: why is there a need to add randomness (stochasticity) to the models of gene-environment interaction?**	*After our focus on neurodegeneration, we will shift to cancers of the central nervous system, the cancer genome, and problems of overdiagnosis and overtreatment. Are there common themes of uncertainty in these two research areas?* ***ARTICLE 1:*** "**Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells**" [Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, Lander ES. *Cell* 2011; 146:633-44] This paper focuses on the theme of single-cell variation and heterogeneity in cancer. Is a tumor mass composed of different cell "states"? If so, why does the tumor "look" and "behave" as though it were a single entity? ***ARTICLE 2:*** "**Twenty five year follow-up for breast cancer incidence and**

		mortality of the Canadian National Breast Screening Study: randomised screening trial" [Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. *Brit Med J* 2014; 348:g366] One of the biggest questions in the oncology field has been whether annual mammography is beneficial both at an individual and a population level in reducing breast cancer mortality. This clinical trial goes some way in answering this question.
SESSION 9 Apr. 6	**(i) Systematics 1: an attempt at gaining certitude?**	*For the next two sessions, we will turn to a new development that is gaining popularity in different areas of pathobiology, i.e., a move toward systematic and large-scale methods that aim to identify dysfunctional "circuits" from a vast collection of data.* ***ARTICLE 1:*** "**Entropy-scaling search of massive biological data**" [Yu YW, Daniels NM, Danko DC, Berger B. *Cell Syst* 2015; 1:130-140] The amount of different "omic" data is increasing so rapidly that different algorithms, each geared for a specific purpose, are needed to search these databases. This article, published in the first volume of the new journal *Cell Systems*, "raises important questions as to what the theoretically optimal bounds for querying a dataset are." ***ARTICLE 2:*** "**A mesoscale connectome of the mouse brain**" [Oh SW, Harris JA, Ng L, Winslow B, Cain N, Mihalas S, Wang Q, Lau C, Kuan L, Henry AM, Mortrud MT, Ouellette B, Nguyen TN, Sorensen SA, Slaughterbeck CR, Wakeman W, Li Y, Feng D, Ho A, Nicholas E, Hirokawa KE, Bohn P, Joines KM, Peng H, Hawrylycz MJ, Phillips JW, Hohmann JG, Wohnoutka P, Gerfen CR, Koch C, Bernard A, Dang C, Jones AR, Zeng H. *Nature* 2014; 508:207-14] A currently popular trend in neuroscience is to collect vast amounts of data to create circuit maps of the brain. This is one example from the mouse brain. The question here is what exactly can be gained from these vast maps? Do we know all that we can know from smaller scale maps, such as the map of neurons in *C. elegans* or certain regions of the visual system of fruitfly?
SESSION 10 Apr. 13	**(i) Systematics 2: an attempt at gaining certitude?**	***ARTICLE 1:*** "**The human disease network**" [Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. *PNAS* 2007; 104:8685-90] In this paper the authors present a holistic model of linking human diseases in a network based on genetic and other similarities.

		ARTICLE 2: "**Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins**" [Wilhelm BG, Mandad S, Truckenbrodt S, Kröhnert K, Schäfer C, Rammner B, Koo SJ, Claßen GA, Krauss M, Haucke V, Urlaub H, Rizzoli SO. *Science* 2014; 344:1023-8] This paper presents a relatively realistic model of the immensely crowded and complex protein environment at neuronal synapses by pooling vast and disparate data.
SESSION 11 Apr. 20	**(i) Wrap up of literature discussion with a meta-analysis on neurodegeneration**	*We will discuss two meta-analyses (statistical pooling and synthesis of many studies on a single clinical topic) this session, which are meant to encapsulate some of the main themes of the course on uncertainty, protein aggregation, and neurodegenerative diseases. We will designate some time at the end of the session to go over any questions about the presentations next week and the expectations on the final draft of the essay.* ***ARTICLE 1:*** "**Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis**" [Amyloid Biomarker Study Group. *JAMA* 2015; 313:1924-38] Meta-analysis from the perspective of Alzheimer's disease and the lack of manifestation of dementia ***ARTICLE 2:*** "**Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease**" [International Parkinson's Disease Genomics Consortium. *Nat Genet* 2014; 46:989-93] Meta-analysis of genome-wide association data on thousands of Parkinson's disease cases
SESSION 12 Apr. 27	**Oral presentations**	*Students will give their final oral presentations, with an emphasis on (i) critical experimental design, (ii) dealing with uncertainty in choosing the right controls, and (iii) noting how their opinions about their selected topics and the initial introduction to the themes of the course might have changed compared to the beginning of the class and the first draft of the essays.* **[Final course evaluations and feedback about course themes]**