

So this lecture is called the future of biology. And I want to associate somewhat freely. I'm not going to write on the board. I am going to post this presentation so you can pull down stuff from it. And I think there are two basic threads where we're going in biology. One of them is a basic understanding, what don't we know about biology? And the second is how are we going to use that information to go somewhere profound in the future? So the first one is how life works. And the obvious place to start is the Human Genome Project. The Human Genome Project was a very bold initiative that was first spoken about in the late 1980s. And I remember being at some of the first conversations. I was at the time finishing off my graduate work. I remember being at some of the first conversations about the Human Genome Project where the goal was to sequence the human genome, identify all the genes and DNA, and then use them to do a bunch of things that I'll talk about in a moment. And at the time, I have to say, it seemed like a really stupid idea. It was very difficult to determine the sequence of DNA. It took hours and days and days to read even a thousand base pairs. And, as you know, we have more than ten to the ninth base pairs. So the notion of sequencing the entire human genome seemed incredibly expensive and incredibly stupid. But that was a reflection, I think, of my naivety. And, in fact, it's been a very useful exercise. Sequencing has gotten better, largely because it had to in order for this project to succeed. And I think there's a real lesson there. If something has to happen, if you have to get a project done, there are people, you guys, who can make techniques better and get things done. So DNA sequencing is much, much orders of magnitude faster than it was ten years ago. And, in fact, this project was initiated in 1990. The sequencing, per se, was completed a couple of years ago. But, still, there are many people who are looking at the sequence and trying to figure out what it means. Because, as you remember from everything we've talked about, DNA sequence is a code. And even if you get three times ten to the ninth base pairs of the human genome, all you've gotten is a code. And now you have to crack the code. And we know how to crack the code kind of, that's what we've been talking about over and over. What does a promoter look like? What does an RNA look like? What does a coding region look like? What are the signals in a coding region that tells a protein synthesis to begin and to end? But when you've just given the, when you're given this huge mass of information that contains maybe 5% genes and 95% other stuff, to actually find the genes in the human genome from all the sequence data is not trivial. And so there is still analysis going on trying to figure out the identity of a bunch of genes and indeed the gene number. And you may notice every now and then revised estimates for gene numbers in the human genome, and it hovers somewhere around 20,000 to 30,000. Before the sequence was obtained, it was thought that there were at least 100,000 distinct genes in the genome. The number came down and down and down. And now we think they're somewhere around 25,000 genes. That's the latest estimate. It doesn't really matter. Give or take a few thousand. But still the Human Genome Project is not complete. But it's complete enough that many people, including Professor Lander who is here at MIT who is over at the Broad Institute that's being built across Main Street. If you guys walk up towards the Stata Center and look across Main Street, there's a new building going up. That is the Broad Institute that is being organized by Professor Lander who was instrumental, one of the pivotal people in sequencing the human genome. And he and others are now doing the takeoff from the Human Genome Project, and it goes like this. Basically describe everything else about molecular biology. And it's a daunting list of what people want to do. Find all the RNAs, all the proteins in every cell type at every time during a cell's life. Figure out all the DNA-protein interactions, so all the transcription factors that bind to DNA. Figure out all the proteins that bind to RNA and might regulate their translation or might regulate their stability. Figure out all the protein-protein interactions, all those enzyme complexes, all those proteins that interact in all those signal transduction cascades you've been talking about. We have no idea what all the protein-protein interactions that go on in every cell at every point in a cell's life are. All the signal transduction events. All the regulatory circuits. I'll talk more about that in a moment. All gene function. All diseased genes. This is an enormous list. It's going to take decades of many people to get through this list and get all this information. And, of course, in the end this is just information. And you have to do something with it and put it together so that you do land up with understanding gene function and being able to build circuits of the kind that bioengineers like to do. One thing I want to point out for those of you who are interested in computer science. One of the things that has come out of the Human Genome Project is a lot of data, but it's actually not that much data. It's a few terabytes. OK? So 80,000 CDs will store all the information coming from the Human Genome Project. But that's just DNA sequence. OK? If you're starting to look at protein-protein interaction, all the RNAs, everything that I just went through on that list, we're talking about billions and billions and probably trillions of terabytes. Where is that information going to go? Is there a good way to store the information now? There probably ought to be some real reevaluation of data storage. And there is. There is some interesting work being done to try to figure out how to store and how to access the information that's going to come out of the follow-up of the Human Genome Project. How do you find proteins that are present in all cells at different times in a cell life? So here's a piece of real data. In the study of proteomics the notion is to look for proteins that are present in one cell type and not in another cell type. This is a technique that you know, gel electrophoresis. It's called 2-dimensional gel electrophoresis. In the first dimension you separate proteins by charge. And then you actually turn your gel around, rerun it and separate proteins according to their size. And what you get are a constellation of spots, each of which represents a protein. And you can look at the spectrum of proteins from one cell type and from another cell type and ask what's different and what's similar between the two cell types. So, for example, this arrow up here. Actually, let's look at this one. In this cell type one there's one, two, three spots that are in the circle and an arrow pointing to nothing. If you look at cell type two, here are one, two, three, the same spots. And here one, two, three. And here the arrow is pointing to another spot which is a protein that's present in cell type two and not cell type one. And this kind of method is the way that people are figuring out which proteins are present in which cell type. What you can do now is to actually cut this little spot out of the gel of cell type two, put it through the mass spec and figure out the identity of that protein. So you can do this stuff. It's just a lot of

work. And there are more sophisticated methods than this to go about finding all the proteins, but basically you have to look and you have to identify the protein. And then you have to store that data and use it somehow. Here's something else that's being done by Professor Young at MIT. Professor Young is trying to figure out all the regulatory networks between all the genes in yeast. So yeast is a small organism. It has just a few thousand genes. And it has actually just a few hundred transcription factors. And their names are arrayed around the outside of the circle. And what he's done, using various techniques, is to figure out which transcription factor activates the expression or changes the activity of which other transcription factor. And so every arrow indicates that there is some kind of interaction between these different transcription factors. And this gives a kind of regulatory network of the circuitry involved in controlling yeast transcription. Now, yeast is a single cell with very few genes. We are, as you know, multicellular organisms with many genes. And so the regulatory maps for humans are going to look many, many orders of magnitude more complex than this one. That's where we're going. And part of going there is using computational biology. One of the things that there is focus on in a number of departments at MIT, including Course 7. , is the question of computational biology or that include systems biology. And how can you use computational methods to work together with real data to predict these circuits, to describe these very, very complex circuits, to describe the circuit of life? And I can tell you something. It sounds as though it's a doable task, and in theory it is, but actually we are not able to describe the circuit of life for even the simple viruses. So there is a virus called phage Lambda that's been mentioned to you. It has been studied for many, many, many decades. And we know about its lifecycle in great detail. It doesn't have that many genes. We know which genes turn on and off. And yet we still don't have a completely reliable computer model of how this phage responds to various environmental inputs. We don't quite know when it's going to lyse the cell or when it's going to incorporate into the bacterial cell chromosome. We don't even have a complete computational description for a simple virus. So to get it for a cell is a daunting task. And this is where computational biology is going to have to work with the real data and where you guys come in to try to bring things together so we can actually get reasonable equations of life. Here's an equation that I took from one of my colleagues, Professor Eric Davidson who is at Caltech, who has been working with someone else to look at one of the regulatory circuits in drosophila. And again this is just the tip of the iceberg of gene interactions. You can look at this on the PowerPoint later. These are gene interactions, and this is just a little bit of the circuitry that sets up a little bit of the body plan in the fruit fly drosophila. Here's another frontier of biology. Imaging. Imaging in biology is fantastic right now. So we can do stuff like look at fish that has got its red blood cells fluorescently labeled. And we can actually see in real-time the movement of the red blood cells through the different parts of the body. We can put various drugs on the fish. We can use various fish mutants that are defective in components of the extracellular matrix, for example, or some other aspect of the animal that might control red blood cell movement or function. And we can look in real-time at what happens to the animal. This works great in fish, but it's only the beginning because there are things we still cannot see clearly enough, even in fish which are transparent. So these methods work well. The challenges become very great in mammals where development occurs inside the mother and where the animal is opaque. And how do you actually follow single cells through the animal as they're doing whatever they're doing? So, for example, if one wants to know what happens to a cancer cell when it's introduced into an animal, does it go directly to the place where it's going to make a tumor or does it wander around the body until it actually finds where it's going? You have to be able to image single cells in a very profound way. And this is one of the current frontiers of biology. But then you can go deeper down into the cell. You can expand that by a few orders of magnitude and say, well, it's not just looking at the outside of the cell. You really want to be looking inside the cell in real-time to see proteins interacting, to see transcription happening in real-time in the cell. It's not quite clear how to do that right now. It's a combination of physics. So if you're thinking of a Course 8.0 major, this is something you might think about. It's a combination of physics and biology. How do you get imaging on a resolution high enough that you can look in real-time at these events that are occurring in a cell? Fascinating problem. Neurobiology we touched on. Where is neurobiology going? Well, lots of places. How do you make the brain? We have no idea how you construct the 3-dimensional brain. And we don't understand the significance of the 3-dimensional structure of the brain. In neurobiology we talked about the circuitry in the brain and about the billions and billions of circuits that there are in the brain, probably ten to the fifteenth circuits within the brain itself. How on earth are we going to actually figure out that circuitry? I have no idea. I have no idea what we can do in the mammalian brain. We cannot even do it properly in something fairly simple like the fruit fly, so how are we going to do it in the human brain? This is a real frontier of biology that, again, brings together multiple disciplines; physics, biology, brain and cognitive science. What's the molecular basis for thought and how can we think about ourselves thinking about ourselves thinking about ourselves? What does that mean? OK? You've had one test on action potentials. You know about synapses. It's got something to do with channels and synapses, right? But what goes beyond there? How does it come back to something that allows us to think in such complex ways? Here's one. Why do we sleep? Simple question. We sleep. You have to sleep. If you don't sleep, this is something for you guys to bear in mind. If you don't sleep, after two weeks you will drop dead. That is true of rats. If you prevent a rat sleeping for two weeks it drops dead. There is something that happens during sleep, and it's not clear what, it's really not clear what. It's thought that it might be some kind of metabolic restoration of the brain that maybe you run out of some essential components that you need in order to get normal circuitry. But you literally have to sleep or you will die. But we don't know why. OK? So that's a frontier that is particularly interesting. OK. This is work I wanted to show you from my own laboratory. We're interested in the 3-dimensional structure of the brain and why you have a 3-dimensional structure and how you make the 3-dimensional structure. We're looking in the zebra fish. This is a normal zebra fish brain. And you can see it's got these three red regions, which are actually cavities in the brain. And then this is a whole series of mutant fish we've isolated that have got really messed up brains. They turn out to be really messed up animals. They've got abnormal behavior, their neurons grow in the wrong place. and there's something profoundly wrong with both the architecture of the brain and the function of the brain.

This is one kind of approach we can take but, in fact this is asking a rather simple question. It's not asking the question of how the zebra fish thinks about itself, if it does. OK. Oh, so what I'd like to do in the last lecture also is to point you in the direction of relevant movies, some of which you'll have heard of and some of which you won't. This is one you've probably heard of because it's a new one. So I particularly liked this movie with respect to neurobiology because, actually, I didn't like this movie. I thought it was a really depressing movie, but the part that I thought was really relevant to this class is that there is a company called Lacuna Incorporated that will go in and selectively erase memories from your memory banks. And they actually can plug in, you know, they put a thing on your head with electrodes coming out. And then they have a TV monitor. And they can actually see the circuits that correspond to a particular memory. And then they hit the erase button or the delete button and that memory goes. And this movie is about this guy, you look in so pain, Dr. Gardel. This movie is about this guy, Jim Carey, who is trying not to be erased. Anyway, it's very interesting because I thought, gee, will there ever come a time when we actually can have a TV screen and we actually can see the circuits that correspond to a particular memory? So see it if for no other reason. OK, here we go, basic understanding. Something we have not talked much about in this course but is really very important with regard to the future of biology is evolution. What does evolution mean, especially in molecular terms? And I actually wanted to throw this out at you because this is in the news presently, this term "intelligent design" and the contrast to evolution. I think that you guys, even now but certainly as you develop and go through MIT, really become spokespeople for science and become commentators on current issues in science. And I think you really should be aware of some current issues, so I'm throwing this out at you to increase your awareness. There is a term floating around called "intelligent design" which is sort of, I would say, an extension of creationism where the sense is that things are just so complex and so interesting and seem to be so well designed that how could this have happened by the process of evolution? And so if you look in the news, if you do a Google news or if you just look in the newspapers, you'll see there are raging controversies about the notion of intelligent design versus evolution around the country. And indeed evolution is complex, and we cannot explain how everything occurs. This is a picture of Darwin's finches, the thing that got him thinking about evolution. These finches that live in the Galapagos Islands and are believed to have arisen from a single pair of finches that the wind blew astray about 100,000 years ago. And that turned into a bunch of different species that can be picked out by their head shape and their beak size. And it's really not clear how you actually got this set of different beak shapes and head size. The sense of selection for particular beaks that allowed the birds to eat particular foods and so on is a very compelling one. And there certainly is no doubt in my mind, or I would say in most people's mind who work in biology, that natural selection and evolution is the way to go. But I want to raise with you an interesting question and then I want to tell you about a new paper that I read concerning natural selection. So natural selection leading to evolution is thought to act on three different kinds of changes in DNA. Single based mutations, you remember those, frame shifts, missense, nonsense mutations and so on. Cis-regulatory mutations, those refer to mutations in the promoter regions of genes. So those would change the transcription of a gene. A single base mutation would change whether a protein is made and what the actual sequence of the protein is and therefore its potential function. The cis-regulatory mutations would change how much of a message was made, how much of a protein was made. And here's one that you've touched on a bit, but I want to touch on a bit more, which is the repeat number of motifs within one coding sequence. So what am I talking about? Well, I'll tell you what I'm talking about. And I'll use it to describe the example of dog evolution. So dogs are really different from one another. They're extraordinarily different from one another. If you look at their size and their actual faces and the bones of their face, the shapes of the bones, the size of the snout and so on are really different. You know, not only are they cute, but they're really different from one another. OK. And if you actually look, over the past 150 years, there has been a huge increase in the number of breeds and a huge increase in the changes that you see in dog facial skeleton. Now, this bothers people who think about evolution. Because if you look at the number of single based mutations that are around in coding sequences, it would not seem to be enough to accomplish these rapid changes in dog morphology. And so a very interesting paper came out last year that will lead to a conclusion I'll tell you a moment. The conclusion has to do with variations in the number of motif repeats within a protein. So what am I talking about? So forget this. This is actually on your handout. So if you look at, I did give you a handout and I haven't been referring to it, but this in fact is number seven on your handout. So if you look at protein A and allele A of protein A or allele B of protein A, they may differ in the following way. In protein A there may be a small amino acid stretch that is repeated a couple of times. It could be directly contiguous or it could be a little far apart from each other. And then if you look at allele B of protein A, you might have five copies of that repeat sequence. And, in fact, that change in number of copies of a particular part of a protein can profoundly change the function of the protein. It can change confirmation. It can change enzymatic activity. It can change localization in the cell. It can change stability of the protein and so on. These repeats and variation in the number of repeats are actually very easy to get. They are about 100,000 times more frequent than point mutations if you look in genomes. And they occur during recombination where the DNA sequences might misalign with one another. And I'm not going to get into this now, but if you want to come ask me later I'll email you. We can go into this more. But they occur because the DNA sequences don't quite align properly during recombination, and you get the protein changing a bit with respect to these repeat sequences. And so Fondon and Garner looked at 92 breeds of dogs, and they looked in 17 genes that they thought might important in shaping the facial skeleton because we know what those genes are. And they found, very interestingly, that these 17 genes had 37 repeat regions amongst them which is actually much higher than you find in just your general spread of genes. And when they looked from breed to breed they found there was a huge variation in the numbers of repeats in different genes from one breed of dog to another breed of dog. Now, I don't know really what this means. It's very interesting potentially for looking at how breeds of dogs have evolved or how we have forced their evolution. Does this have something to do profoundly with evolution and changes in form in general? Don't know that. But it's something that you should be aware of as you go on because it's a slightly different way to

think about how rapid evolution can occur. OK. Clinical understanding. Disease taxonomy, you've talked a bit about this in cancer. Over the last few decades there has been an enormous increase in the number of genes that can be assigned to be associated with a particular disease. And my chart here only goes up to 2002. It would probably be somewhere out here on the roof for 2005. How do you do this? Well, this is where the Humane Genome Project comes in. One can look. And each of these squares represents a gene and its expression, and the level of expression is proportional to the color or is associated with the color. It doesn't matter which. But you can look in different cancers. And each of these lines is a cancer. And you can look at different genes. And you can see that different tumors have got different patterns of gene expression. And you can use those patterns of gene expression to classify the tumors. And this has been done by Professor Lander and Dr. Golub over at the Genome Center. So, for example, in acute lymphocytic leukemia, you can see one pattern of gene expression. Again, each of these squares represents a gene and the color represents the level of expression of the gene. And you can see in acute myelogenous leukemia there's a completely different pattern of expression. This is fantastic because it starts to allow you to classify a disease in molecular detail. The old way of pathologists looking at diseases, looking at cells and trying to classify both cancers and other disorders on the basis of morphology of cells and of staining of cells is actually not that precise. It's much better than nothing. But being able to do it at a molecular level really lets you know what disease you're dealing with and what spectrum of drugs might be appropriate to treat that disorder. And so that segues nicely into the future of prediction. What can we predict in biology? So here are a couple. Will your specific disorder respond to particular drugs? If you have acute Lymphocytic leukemia, will it respond to a particular spectrum of drugs? And if you have a particular variant of acute Lymphocytic leukemia, will it respond to particular variants of drugs? We're already on the cusp of classifying cancers in a way that you can give a particular spectrum of drugs for a particular kind of cancer. And this is really going to escalate to everything. There is almost no disorder that is treatable by medication that where different people are not sensitive at different levels to a particular medication. So some people might respond very well and some people might respond very poorly, not just in cancer but in almost all disorders. In the future, and I think it's going to be in the near future, really in the next few years, I think it's going to be possible to say what specific disorder do you have and should you be taking this particular combination of drugs? And here's another one. Are you genetically predetermined to get a specific disease? That's a really tough one. Maybe you want to know. Maybe you don't want to know. We'll come to that to in a moment. Here's a movie that I particularly liked that has to do with predicting who you're going to be and what you're going to get or not get. And it's called Gattaca. You guys may or may not have seen it, or used to see it long ago. Have you guys seen Gattaca? Yes. OK. Good. Fine. I'm not that far out. I was trying to gauge your level here. I liked that. I particularly liked A—there is no gene for the human spirit. So Gattaca falls into the prediction aegis quite well. In the next few years, and I would say within ten years easy, you are going to be about to get your personal DNA profile, including information about the approximately 1, 00 bad alleles of genes that we all carry. So that's good, I guess. And that's bad. Do you want this information? Do you want to know what you're going to get? Do you want to know that you're going to get a neurological disease as you get older? Do you want to know that you're likely to have a heart attack before you're 50? Do you want others to have this information? Would you like your perspective partner to know that you're likely to get some horrible neurological disease? Would you like your insurance company to know this? Would you like your children to know it? I don't know. I don't know what the answer is. Myself, I prefer actually not to know and to go through life day-to-day having as good a time as I can and letting whatever, God or chance take care of the rest. But there is certainly merit in trying to prevent some things. And this is really going to be a reality very soon. Already, actually, you can get for your dog a DNA profile that's actually fairly detailed. It is RFLP mapping that we talked about in class where you can make sure that your dog is who you think it is and who its parents are and who it thinks it is. So you can get your own in the future. I don't know what they're going to call the American Kennel Club equivalent for humans, but you'll be able to get your certificate of DNA analysis. OK. Design. We talked long ago about rational drug design. And Gleevec is really one of the shining examples of being able to look at the structure of a protein and saying, hey, this protein is a bad protein, it's an abnormal protein, and it's causative of leukemia. And wouldn't it be great to inhibit its function? And so let's design the screen molecule which looks as though it will inhibit ATP binding and prevent this kinase from acting. And, in fact, Gleevec works really well that way. And this is something that is being, this approach is being very, very actively pursued, and will only be more actively pursued. And if you're thinking of a Course 5. major or if you're thinking of many of the engineering majors, you may well get into rational drug design and figuring out how to get it to work. Here's another one, prediction for the future and Jurassic Park. Xenotransplantation, using pigs that have immune systems engineered to look like the human immune system for organ transplants. There are companies trying to do this. Bionics. Here is something for all of you interested in mechanical engineering and other, bionics. Robocop. Here's a movie to see if you haven't. Artificial hearts. Artificial hearts are a disaster presently. AbioCor has a heart now that is fully implantable and that will take over ventricular function, but it is a lousy heart. And if one of you could go and make a new heart that actually would work properly that would be a real service to humankind. Blood substitutes. We still do not have a blood substitute that really works. There is one patented. It's called Oxygent. It's a perfluorocarbon that will carry oxygen around the blood in the body for a while, but it's not very good. And finally aging. Here's a movie you probably haven't heard of, Zardoz. Zardoz, a great movie about a community where nobody aged and they were really, really unhappy. And then Sean Connery comes along and introduces this community that doesn't age and doesn't have sex and is just generally miserable to the joys of procreation. And they go for it. And then they age and they die happily ever after. So that's OK. And if you're looking for a summer book to read, Professor Guarente at MIT has written a great book about aging, which is what his research focuses on. And you might want to look at that. So the challenge I give to you is where are you going to come in? I wish you all the best of luck. And it's been a pleasure to teach you. [APPLAUSE]