

Hello, everybody. Can we get started?

So my name is Andrew Chess. And I'm lecturing today replacing Eric Lander for the day. He had to be out of town, something that he could not reschedule. He really tries to arrange his very busy schedule so that he is here, but this was something that could not be rearranged. Anyway.

So I am a professor at Harvard Medical School, but I have a long history here at MIT, including being on the faculty here for a number of years. I used to teach undergraduates.

And even going back further than that, I actually took this class.

It was called 7.01 without any extra number at that time.

Now it is what, 7.012 or 7.13? 0-1-2. Anyway. So it was called 7.01. And it was an extremely interesting introduction to biology back then. It was some time over 20 years ago.

I'm not sure exactly how many years. I kind of stopped counting at 20.

So when Eric called me, when Professor Lander called me up to ask me to give this lecture, I talked to him for a while and I agreed to do it. And then I was thinking about what to present. And I went over the material that he has presented to earlier this week. And I discussed with him some of the things that he likes to do for the third lecture on neurobiology.

And I also thought about some of the things that I would like to do.

And so one of the things that occurred to me, so it's probably occurred to you from hearing Eric talk about neurobiology earlier this week that he is very enthusiastic about the subject. He always talks about it as being one of the driving forces that led him to enter biology from the realm of mathematics where he started his academic career. Anyway. So he's always talked to me, in the years I've known him, about how he loves neurobiology.

And I was thinking about it. It is, in some ways, kind of ironic that Eric loves neurobiology so much because in some ways he's been a big trouble-maker for neurobiology.

Let me explain. So Eric, Professor Lander is, of course, as you all know, was an instrumental driving force behind the sequencing of the human genome. Before all those efforts, over the last decade, people used to go around, biologists used to go around to each other and they would talk.

And they would say there are around 100,000 genes in the human genome, or 100,000 genes in a mouse genome. Mammalian genomes have around 100,000 genes. Sometimes some people would say 90,000 genes.

Sometimes they would say 110,000 genes.

But generally it was around 100,000 genes. And everybody was comfortable with that. In fact, right around the turn of the century when the stock market was going way up, the Internet bubble and biotech bubble and everything, estimates of the number of genes in a human genome actually went higher also. They went up as high as 120,000, 150,000. And this, I think, was because various companies were competing to have the most genes on their kind of micro array or whatever they were selling. They wanted to say we have the most, and so they kept saying more. The academic scientist usually still stayed around 100,000 in terms of their thinking.

OK. The other thing that was of a lot of use to neurobiologists in terms of them feeling that their problem of trying to figure out how the brain is set up as a tractable problem was that they thought there were going to be lots of genes available.

So there were 100,000 genes in the genome, and around half of them, people would say, are probably brain-specific.

So there were a variety of pieces of evidence that people would think that a lot of the genes in the genome would be brain-specific.

And that still remains the case. But, as you know from Eric Landers and other people's work sequencing the genome and mouse genome and other genomes now, it looks like mammals now have only around 30,000 to 40,000 genes. Now, if Eric told you a different number listen to his number. What did he say? [20,000 to 25,000?].

OK. Anyway. So many fewer than 100,000. OK?

A small number. A fly is thought to have only around 15,000 genes.

So Lander is now saying that humans don't have that many more genes than flies. So this presented a problem for neurobiologists, because even if you have half of them brain-specific, you still don't have nearly as many genes to play with to make the complex structure of the brain as you did when there were 100,000 genes in the genome. So the brain, of course, is an extremely complicated structure. There are thought to be somewhere between 100 billion and a trillion different neurons in the brain, and they also fall into many, many different neural types.

And so the developmental process, developmental biology is something that some of you will study in future courses and you'll have had some introduction to here, that's how you get from a single fertilized egg, a single cell to the complex organism. In the brain it's a particularly difficult problem because you have so many different types of cells and so many cells.

And each cell then makes all of these complex connections.

A given neuron might connect to 1,000 or a few thousand different other cells as a normal process. So forming all these different kinds of neurons and wiring up is a very daunting problem.

So what I thought I would do today would be to focus on two examples where in each case, starting with either one gene or a small number of genes you get a lot of complexity.

So this would then allow the smaller number of total gene number in the human genome to allow a lot of different kinds of proteins and maybe provide some explanations for certain parts of the complexity of the brain. Now, by no means am I going to attempt to explain all of how the brain develops.

That would take, well, at least one course, more likely a few different courses to actually get a good appreciation of that, but I'm going to go through a couple of very intriguing examples that are approachable at the level of this class.

OK. So the standard way that we think about how genetic information gets made into proteins is that there is DNA and then RNA and then proteins. I hope at this point in 7.

1 this is all familiar to you. Good. OK. So the DNA sequence gets transcribed into an RNA, and then there's a splicing event which takes bits and pieces of the RNA, puts them together, and then there's an area of the RNA that tells the ribosome to start making protein. And so you get the protein synthesis. Everything is following from the blueprint that was in the DNA. So what I'm going to talk about today are two different examples. One of them involves alternative splicing.

And that will be causes where instead of there being a static always reproduced way of going from DNA to RNA to a protein, that there are different alternative splicing events that can occur.

And this can allow one gene to make multiple different protein products.

And then I'm going to go over another way that you can violate this central dogma, this DNA, RNA to protein, which is something called RNA editing. So, as you might imagine from the common usage of the word editing, by editing what we mean is that the RNA sequence itself is actually changed so that it no longer reflects the exact nucleotide sequence of the DNA.

This can also add diversity to the number of potential encoded proteins.

I want to make sure that I, I'm going to talk about neural-specific examples, but I want to mention that these processes, alternative splicing and RNA editing are used also by other parts of the developing animal, and also in

other plants and other organisms, but not just by the brain to generate diversity.

So these are mechanisms that are widely used. But some of the most striking examples, as you'll see from my lecture and in further reading that you might do in the future, some of the most striking examples come from the nervous system.

And that's not surprising given the complexity of the nervous system and the fact that there are so many genes out there ready to help with this complexity. So the first I'll turn to alternative splicing.

So first, before getting into the extremely complex case that I'm going to focus on, I'm going to just briefly, by way of introduction, go over a standard alternative splicing scenario.

OK. So in a gene for which there is no alternative splicing, if I draw the exons as boxes and the introns as lines, what winds up happening is you have the first exon spliced to the second one, second to the third, third to the fourth. And so what you wind up with is a messenger RNA.

So here is the messenger RNA which has been spliced from the primary transcript. The primary transcript, of course, reflects the actual structure of the DNA in terms of sequence also, because in the genomic DNA you'd also have areas that are going to be exon, intron, exon intron exactly like this. So this is just a general example of alternative splicing, I'm sorry, of regular splicing. So then alternative splicing would involve something like this. You'd have 1, 2, 3A, 3B, 4. So then what happens is you have normal splicing from 1 to 2.

And then 2 could either go to 3A, which will then go to 4 leaving 3B out. Or the alternative is that 2 can skip 3A, go to 3B, which will then splice to 4. So this allows then two different messengers to be formed.

So this 3A and 3B might encode a slightly different sequence and might then allow two distinct proteins with different functions to be forwarded from one message. So this is an example, a simple example, a general example of alternative splicing.

So from one gene you have two proteins.

The example that I'm going to focus on today, instead of going from one gene to two proteins, allows you to go from one gene to 38,000 different possibilities.

It's actually 38,016 to be exact, and I'll explain to you why, but 38,000 will have occurred to you that this is larger than the number of genes that Eric Lander says are in the human genome.

It's certainly larger than the number of genes in the fly genome.

And this example I'm giving you is from a single gene in the fruit fly.

This one gene can come in 38,000 different forms.

The gene is called drosophila DSCAM.

It's named for a human gene which was cloned first and characterized first which was called just plain DSCAM.

What DSCAM stands for is Down syndrome cell adhesion molecule.

And let me just explain briefly why it has this name. I don't think that this name is actually relevant so much to the biology, and it certainly not relevant to the alternative splicing because the human gene and the mouse gene, neither of them have a lot of alternative splicing.

This is something that is particular to the fly.

That's something I will return to later in lecture.

This name Down syndrome cell adhesion molecule came about because this was cloned first from human and it's located on human chromosome 21.

Chromosome 21 is normally present in two copies in every individual.

In individuals who wind up with three copies of chromosome 21, something called trisomy 21, trisomy 21 causes Down syndrome, which is a syndrome that has some brain manifestations like mental retardation, and also has a number of other problems associated with it. When the people who found this gene found it they named it. So they gave the first part of its name, the DS comes from Down syndrome. Cell adhesion molecule comes from the fact that this gene is similar in structure to a lot of known cell adhesion molecules that are encoded by many different loci in the genome. And they initially thought that perhaps, and this gene is expressed in the brain. And they initially thought that perhaps having an extra copy, a third copy of this gene might be what's causing a lot of the brain phenotypes.

Subsequent work has not provided further evidence for that.

So at this point what I would say is that the name of the gene is Down syndrome cell adhesion molecule, DSCAM. It's on human chromosome 21.

It may play a role in Down syndrome but there isn't -- The name is really the best evidence that it plays a role, just the fact that they named it that. OK. But in the fly this is an extremely interesting molecule because, as I mentioned, it can come in 38,000 different forms. OK. So as for why you would have a cell adhesion molecule in

the brain, I just want to mention briefly. So Professor Lander went over with you the structure of a neuron. That neurons have cell bodies and axons and growth cones which allow them to get to wherever they're supposed to connect. One of the types of molecules that allows an axon as it's growing, the growth cone to lead an axon along a complex path is to interact with various structures that it encounters. And so cell adhesion molecule is one of the kinds of molecules that can allow the growth cone and then the rest of the axon to interact with various other cells or other extracellular substrates, proteins that have been deposited by other kinds of cells as they make their way and make the appropriate connections in the brain.

So cell adhesion molecules are one of the mechanisms.

There are also mechanisms that allow cells to respond to gradients of chemical signaling messengers. Question? Yes. Could you explain what a growth cone is? Oh, I'm sorry. That was not covered? OK. So the neuron has a cell body with a nucleus and all the other stuff that's in regular cells that you learned about. It then has an axon which then allows it to connect. And this connection could be very far away. In the case of, for example, a motor neuron in the spinal cord that's intermingling a muscle in the foot, that single cell would have its cell body in the spinal cord and its axon go all the way out to the foot. OK? So that's just an example of one very long neuron. Some of them are very long.

Some of them are shorter. So this is the axon. At the tip of the axon is this thing which looks sort of like my son's mitten when it's cold outside. But this is the growth cone. And basically what's going on is as the axon is growing out in this direction it's feeling its way.

And there might be cell adhesion molecules on these different protrusions that if they attach really well -- Like let's say that this area over here is stickier for this particular growth cone than this area over here. Then this growth axon is more likely to grow in that direction. OK. So cell adhesion molecules are well known to play important roles in axon guidance.

It's how axons grow in different directions.

So I will tell you know about the DSCAM gene, and that will give some insight into how the 38, 00 different forms might be used because they're going to provide different kinds of stickiness.

Let me explain.

So this is a drawing of the genomic organization of DSCAM that allows the extensive alternative splicing.

So this diagram is similar to the diagram that I drew by hand for the more simple cases. But basically in DSCAM what you start out with is exon 1 gets spliced to exon 2 gets spliced to exon 3, and then when you reach exon 4

there are 12 distinct possibilities. And only one of the 12 is chosen.

In this case the diagram shows it choosing, I don't know, the ninth one perhaps. Then exon 5 is regular so that always gets included. And exon 6 there are 48 distinct choices.

And again only one is chosen. Here, in this example, this one has been chosen at the expense of all of these other ones.

Exon 7 and 8 are normal. And exon 9 there are 33 choices.

Exon 17 there are two choices. And if you multiple $2 \times 33 \times 48 \times 12$ you wind up with 38, 16. There is evidence from a number of different types of studies, including coning and sequencing, lots of different messenger RNAs that are already spliced that basically almost all of these forms can be made. So what this structure allows is for there to be diversity generated in important areas of the cell adhesion part of this molecule.

The DSCAM molecule starts out with a number of domains which are called immunoglobulin-like domains.

Immunoglobulin domains are named for immunoglobulins which is another name for antibodies.

Antibodies help you fight infection. I don't know if that's been covered.

Has it been covered? Yes. And the particular fold that they form allows recognition of foreign antigens and it also allows stickiness of molecules in general. So cell adhesion molecules often have these immunoglobulin domains. The DSCAM starts out with, from the N-terminus towards the C-terminus it starts out, the first nine domains are these Ig type domains. That's then followed by another kind of domain called a fiberonectin type domain, which that's not important. All of the diversity is in this nine immunoglobulin domain. The exon 4 diversity allows diversity of the second of the nine. The exon 6 alternative splicing affects the third out of the nine immunoglobulin folds.

And the exon 9 diversity affects the seventh. So of these nine domains of immunoglobulin folds that allow for different kinds of stickiness, a lot of them are the same.

One is the same. 4, 5 and 6 are the same.

And 8 and 9 are the same. But 2, 3 and 7 have these differences which are encoded by this striking kind of genomic structure and alternative splicing.

So how is this diversity used?

So the early models for how DSCAM would be used stipulated that individual different kinds of neurons might express vastly reduced subsets out of the 38, 00. So let's say that one particular neuron type, of which there might

be many different neurons, might express maybe ten out of the 38,000 or even one out of 38,000. These were the different kinds of models that were tossed around by people who were thinking about this problem. But then people started to study it. And what turned out to be the case is that it looks like every kind of neuron population at first approximately expresses almost all of the different forms.

OK? So these are wrong, these models. And each different neuron type expresses a slightly different repertoire.

But at first approximation over 10, 00 or 20,000 forms are possible for each different neuron type. So that then caused people to scratch their heads and wonder, well, how is this used then? How is this used to make different kinds of neurons different from one another or anything in the function of them?

So the answer to this question has emerged in part from analyses of individual single cells. So it turns out that an individual cell, and I'm using the word cell and neuron interchangeably because neurons are cells. And not all cells are neurons but all neurons are cells. So for one cell or one neuron it makes somewhere in the range of 10 to 50 forms.

These are randomly chosen, apparently from the data that's available, from the tens of thousands of forms that are possible.

So you can imagine that two neighboring cells that are otherwise identical, that each are picking, let's say ten just to make it easy, ten different forms of DSCAM, are going to wind up with very different repertoires of DSCAM than an adjacent cell.

So what this allows is each individual cell to have a unique identity. The whole idea that individual neurons might need to have a unique identity actually is a new concept that's really been enlightened by this molecule. Because the way that people used to think of neurons is that they would wind up with unique identities based on the connections or experience, what they were exposed to in terms of different stimuli. But what this indicates is that from the splicing of an individual gene and the fact that each time this gene gets spliced you can wind up with a different form that at any given time each cell will have a unique set of messenger RNAs, and therefore proteins encoding this DSCAM gene.

OK. So I mentioned earlier that the human DSCAM does not have alternative splicing. We all like to think of ourselves, humans and other mammals as having brains that are on the level of complexity, at least on par with the fly. And so it's odd to think of all this complexity that's there for flies and other insects but why is it not there for humans? Well, it turns out that there are other kinds of genes that do have extensive alternative splicing in mammals. So one of them is called neurexins. These are genes that are involved in synapse, how the different kinds of cells communicate with each other at the interface. There are genes called protocadherins. And there are

also other kinds of genes that all have extensive alternative splicing in mammals.

Interestingly, these genes tend not to have extensive alternative splicing in flies. It's as if in each lineage certain genes have been chosen to get a lot of diversity by this mechanism of alternative splicing and other genes are left with their just standard single function where it's one gene, one RNA, one protein. OK. So I'm going to switch now to the second example which is RNA editing.

So, as I mentioned earlier, RNA editing involves an actual change in the RNA sequence so that it no longer reflects the exact DNA sequence. Now, this is different than splicing. Splicing takes different pieces of RNA and splices them together leaving out intervening sequences or introns. But in RNA editing you actually change the nucleotide sequence so that it no longer is identical to the DNA.

This is used in a number of parts of the brain. Most of the examples are brain-specific. There are some non-brain specific parts. Most of the time the editing event changes in adenosine, an A in the ACGT nomenclature, into an inosine.

This is read differently by the ribosome than the adenosine. So this leads, for example, in an important kind of channel called a glutamate receptor.

And the specific subtype that I'm talking about is something called an AMPA glutamate receptor. That's for a chemical ligand that activates this particular kind of glutamate receptor.

This leads to an important change of a glutamine to an arginine in the protein. So let me draw a quick diagram of what the protein looks like so we can see what the importance of this glutamine to arginine switch in this glutamate receptor is.

So in the absence of detailed structural information about different kinds of neurotransmitter receptors people often draw a diagram like this where this is the outside of the cell, this is the inside of the cell, this represents the cell membrane. And here's the amino terminus.

And then they draw a transmembrane portion. And this is what's called a reentrant loop. It doesn't quite pass through the membrane, but then it passes the membrane again. And here's the carboxy terminus. The glutamine to arginine change is here. It's in this area which is involved in making the pore.

So the pore of the channel has this change.

Glutamine to arginine change.

This vastly changes the properties of the channel, a channel that doesn't undergo this editing event.

So let me just state that for the GluR2 AMPA receptor, which is one of four different genes, it is 99% edited in adults.

So where over 99% of the time this adenosine is made into an inosine which leads to a glutamine becoming an arginine in the protein.

What this does to channel is it changes its permeability.

So this is a kind of channel that is mostly designed to let sodium in, but if it doesn't get edited, if the glutamine is there it also lets calcium in. So whether or not calcium gets into the cell is very important because they're both, both sodium and calcium are cat ions and can lead to membrane, potential disturbances like you learned about earlier this week, leading to an action potential.

But calcium also has other effects. It can lead ultimately to the turn on and off of genes and phosphorylation of various proteins to other kinds of effects in the neuron. So it has to be regulated very tightly. So these channels are designed to just let sodium through, to be involved in the transmission of an action potential from one neuron to the next.

So if you had a perturbation in this process you would then also let calcium in because the glutamine containing channel lets calcium in.

In fact, early in development it's probably true that you don't edit 100% and you let a little bit of calcium in.

And there are also other glutamate receptors that are not part of the AMPA family but they are related. And they're encoded by genes called, so this GluR1 through 4 encodes AMPA receptors, a gene called GluR5 and 6, they have editing which is more regulated. And by regulated I mean that the editing is sometimes present and sometimes not.

So even within a given neuron you might have some channels that have the glutamine and some channels that have the arginine.

There are also other sites.

I've talked about the main site. This is the one that has the most profound impact on the function of the protein. There are also other sites in the molecule that are edited. And then they also have important but slightly less prominent roles in the regulation of these channels. So what leads this gene to become edited? Why do most genes not become edited and this gene becomes edited? Well, people are starting to pursue that kind of mechanisms. And one of the mechanisms that's become very clear is that, for example, in the Q to R change in

the pore, or the adenosine to inosine change in the messenger RNA that leads to the Q to R change in the pore, if you look at the exon where that -- I'll draw it as an A.

Where that A is present. There's an area, and then here's the intron, of the intronic sequence which actually loops back and then allows base pairing to form between the messenger RNA and the intron.

And then the enzymes that are involved in mediating this adenosine to inosine change recognize the base pairing of this short area and some sequence specificity to the RNA sequence that's around here.

It's not just any base pairing, but the base pairing is critical.

And then the enzymes, which are called adenosine deaminase, can mediate this change of an adenosine to an inosine.

So this gene has been selected to have parts of its intron, in addition to allowing splicing to occur, to actually be able to base pair and allow this editing function to happen. And so this allows an individual gene then to make more than one form of protein.

OK. One other example that I'll tell you about of RNA editing involves the serotonin system. It involves a serotonin receptor which has RNA editing. And this is a serotonin receptor whose name is serotonin receptor 2C, so it's often written as 5-HT or 5 hydroxytryptamine 2C serotonin receptor. This receptor is a member of the G protein-coupled receptor super family.

G protein-coupled receptors are 7-transmembrane domain receptors.

And the editing of the serotonin receptor occurs in the second intracellular loop and affects the coupling to G protein.

So the way that these G protein-coupled receptors -- Did they do this already? G protein-coupled receptors transduce the signal through a G protein which often binds to the intracellular loops, particularly the second loop.

And the editing event which changes adenosines to inosines, there are actually a few of them in this region here, a few different sites which can get changed. That affects the efficiency of the transduction, when serotonin is present the transduction of a signal inside the cell.

There are other serotonin receptors that are ligand gated channels but they don't appear to have the same kind of editing as this serotonin receptor that is of the G protein-coupled variety has.

So I'll end with just one final note about this serotonin receptor which is that a drug called fluoxetine or Prozac, did

Eric go over that this year? Prozac?

No? He mentioned it. OK, he mentioned it. So it's mostly known as something which blocks reuptake of serotonin.

So one cell releases serotonin, which is a neurotransmitter.

Another cell responds to it. If you block the reuptake the neurotransmitter is present in the synapse for a longer amount of time leading to increased signaling. Prozac is widely thought to have its main effect, and that probably is its main effect to just block the reuptake leading to increased serotonin signaling.

Someone has studied the serotonin receptor in individuals who are taking Prozac versus individuals who are not taking Prozac.

And what they found was that there were differences in the amount of editing of various sites in this key area in individuals taking Prozac versus individuals not taking Prozac.

And the direction of the difference, whether it's switching from edited to unedited or back, and it varies for the different sites, the direction was the opposite of what was seen in a comparison of brains of victims of suicide versus brains of other accident victims. So it looks like Prozac is having an effect which is in the opposite effect to the skewing of editing that one sees in certain cases of depression.

So I bring that example up because it's important to know that, you know, if you can impact on some of these subtle differences between different kinds of messages, like whether it's edited or not or whether the splicing is more towards one kind of alternative splicing or another kind of alternative splicing, these might provide very interesting pharmacologic targets for therapies that might impact on a variety of different human diseases. So what I've hoped to do is give you a sense of a couple of different examples where you can take a single gene and make more than one protein, and this can lead to increases in the diversity of neurons, and therefore increases in the complexity of the brain. Thank you. [APPLAUSE] Are there questions?