

Genes and Genomes: Impact on Medicine and Society

Genes, Genomes, and Evolution
October 16, 2003

Sydney Brenner, Ph.D., Salk Institute, La Jolla, CA
From Genes to Organisms

Introduction by Isidore S. Edelman

Isidore S. Edelman: Dr. Sydney Brenner will speak to us on a topic from genes to organisms. Dr. Brenner was born and achieved his baccalaureate in the Union of South Africa. He earned a doctor of philosophy at Oxford University in the UK, and soon thereafter joined the MRC Unit in Cambridge. His successes in the MRC are quite remarkable, spanning cell biology, developmental biology and molecular genetics. He introduced and exploited one of the most powerful model systems, the worm, with the marvelous name of *Caenorhabditis elegans*. Dr. Brenner served as director of MRC Laboratories for a decade and a half. In the 1990s, he translocated to California, where he is now a distinguished professor in the Salk Institute in La Jolla. He played a particularly important role in the International Genome Initiative. He introduced the pufferfish as a model system for studies on gene expression, owing to its paucity of introns. Some of the difficulty in understanding molecular genetics has to do with the fact that the active genes are interspersed between regions which are not involved directly in gene function. Two of the model systems which were chosen for total genome sequencing, *C. elegans* and the pufferfish, Fugu, were championed by Dr. Brenner.

Time does not permit a listing of all of Dr. Brenner's honors and awards, but I will read his 2002 Nobel Prize citation, which he shared with Dr. Robert Horvitz and John Sulston. "For their discoveries concerning genetic regulation of organ development and programmed cell death," he was awarded the Nobel Prize, and his cane is right here.

The Natural Complexity of Genes

Sydney Brenner: Well, I'm very pleased to be attending yet another celebration at Columbia. The last one I came to was the 200th anniversary of Columbia medical school, not of the university. That was in 1968 and we were recalling that it had been held in the Armory and during the course of the lectures people started to drive trucks around behind the screen. I think it was only Paul Marks who put on a uniform and ordered them out of the lecture hall. But of course there are other anniversaries that we are celebrating this year, and that is the fiftieth anniversary of the discovery of DNA. And I understand that next year we will be celebrating the fifty-first anniversary of the discovery of DNA.

So I've had quite a load of celebration, and it's given me a lot of time to think about what it is we actually did. And when I saw that I was asked to speak on a topic called "Genes Are History," I didn't know how to interpret that, whether it was going to be the history of the genes themselves or whether it was going to be a history of genetics, a talk about genes. And in fact then I decided I would give one of my standard titles, because I discovered you only need two titles to talk about the whole of genetics. The first lecture is called "From Organisms to Genes," that deals with the past, most of the present, and "From Genes to Organisms" deals with the future. So, if you like, this is a history of the future. I am still waiting for a misprint in my title where the letters NI are left out. You can work that out. The solution of that conundrum is left to the reader.

So I think we are at a very interesting junction in genetics, but first we should state what the project is all about, what is the project of genetics? And it's quite simple. Biological systems are unique in the world of natural complexity; they are the only naturally complex systems that contain an internal description. I once went to a meeting where we were addressed by a Buddhist—he was called an archbishop, but I'm sure they have other names—and he was asked, "What is the Buddhist definition of life?" And he said in true Buddhist flexibility, he said, "Well, some Buddhists think everything is alive; mountains are alive, rivers are alive." So I stopped him and said, "Mountains are not alive." So he asked a good question, he said, "How do you know?" I said, "You can't clone a mountain." A mountain contains no internal description, as all living systems do. And I think our task has been always, is, how do we map, if you like, genotype onto phenotype? Because if you think about this lineal sequence of bases within all of us, and if you think about the complexity of structure and function that we all have, that we human beings have a nervous system, a very complex thing, and what are, we might ask, the transformations of going from this linear sequence into the structure of the organism?

The Sequencing Revolution

Now, of course, there was a second revolution in biology, in molecular biology, which took place about 1975, and that was the development of techniques of

cloning DNA, and in particular development of techniques of sequencing DNA. Because up to that point we had only been able to study the DNA sequence through the rather opaque lens of genetics, that is, the only way you could tell changes in the sequence was by mutation. We could map mutation, with fine structured genetics with a kind of failed attempt to sequence DNA by its effects on the phenotype. But now for the first time we could get direct access to the genetic material, and we could read the base sequence there. This alters the problem into one which goes from an inverse problem to a forward problem, and science is always interested, as I shall point out, in solving forward problems.

And of course it raises the question of can we compute organisms? That's the forward problem, can we compute organisms from DNA sequences? If we solve that we solve all the questions in biology. And this is the issue that I want to take up here, because it is the issue of what is the theoretical basis that we will need to construct in the future in order to put that problem at our disposal and aim to work towards a solution of it.

Now, of course, just to reflect back again, the ability to sequence DNA, to get the direct access to the sequence, generated the Human Genome Sequencing Project in 1985 when, I should remind you, that the biggest genome sequence to date was 45,000–46,000 bases, the genome of lambda bacteriophage. And this rather heady group decided they would take a jump from 45,000 to 3 billion. Of course that will need technology, that will need money. And it was put together. I for one thought the technology was not yet able to deal with that size, but of course they hadn't realized the importance of the techniques of DNA sequencing which have come home. So I must point out to you that DNA sequencing is a totally unique technique, technology, nothing else is like it. You can take DNA from any source, plants, animals, bacteria, your mother-in-law, and you can put it through a machine which extracts the essential information, which is the lineal sequencing bases. That means you can essentially make a factory of this, and this means that if you tackle big projects you just have big—more machines. So it's something I called 3M science. It stands for money, machine, and management. And the concept of factory production, which I have to say was first put forward by Wally Gilbert actually, and he saw very clearly that effectively you could expand this technology and obtain the sequences of anything, simply because it's unique. Nothing else in biology is like that; proteins are not like that, cells are not like that. We cannot do everything by this method because the information, relevant information, that we have to extract is not a lineal sequence of bases but is at a different level of complexity. So I disbelieved strongly in all –*omic* science beyond genomics simply because of the uniqueness of this.

Now, make no mistake, I don't disbelieve in parallel observations, that's a different thing; it's a subject we used to call spectroscopy. So I think we should do spectroscopy which would make measurements as effective as possible, but I don't think, as you will see in a moment, that we can extract anything significant from it. Now that's, of course, a hard thing and probably a daring thing to say at

this stage, but I would like to develop the logic of this. There's now a great movement that we will pin everything we know, hang everything we know, onto the genome. It is in fact true that ultimately the fundamental description lies at the level of the sequence of bases. But it is not true that it is implemented in that way. So we have to distinguish between something that in fact Schrödinger has failed to distinguish between, that is, the nature of the information that exists in the DNA and it's the execution, or the implementation, of that information. In fact Schrödinger has stated in his book *What Is Life?* that the DNA contains, as he says, a program for development and the means of executing it. It doesn't—contains a program for development and it contains a description of the means of executing it, but not the means themselves. And that was, of course, earlier enunciated very beautifully in the '40s by John von Neumann, who showed what was essentially required in terms of logic for self-reproducing automata of which you could see biological systems as one example. In fact, I strongly believe that that is probably the first, most important theoretical concept to have been produced in which we can have a level of abstraction that could have actually led one to consider DNA simply from a logical basis, but never did, because von Neumann's work was not known to biologists.

Starting the Cell-Map Project

Very well. So what is the best way to look at this information in terms of how we are going to put it all together and how we're going to actually use it? Now I believe very strongly that the fundamental unit, the correct level of abstraction, is the cell and not the genome. In other words, I've been quoted as saying "forget the genome," you know, we don't want to forget it, we'd like to thank all those people for their sterling work and give them all a gold watch and send them home, or better still send them back to the factory to sequence more genomes. But what we've got to do now is to get away from that and look at how we're going to give the true biological picture of it.

And I want to say this because this is quite important, and a student many years ago came up to me after a lecture and said, "Dr. Brenner, what is going to be the breakthrough in the nervous system?" And I said, "You're about fifty years late, it's already happened. It's called the neuron hypothesis." And of course if you look at what people thought about brains before the neuron hypothesis, it was very clear they would absolutely nowhere. But it was the idea that there were these units called neurons, that they were connected in various ways that underpin the framework of all explanation in the future. And of course the neuron is just another cell. And the cell theory, which was put forward more than a century ago, 150 years or so ago, that is the one that just said everything's just made out of cells, you start off with one cell, have cell division, cells grow, produce a lot of different cells, and so that's clearly the way to get back to looking at organisms.

And I think if I can explain. So of course all projects should get a name, because then they get—I call this the Cell Map Project. It exists only as a plan, but I think it is one I hope that everybody will participate in the future, because that is the way, I think, the only way to organize our activity. And I want to just sketch out roughly how we—I would see this developing in the future.

The Cell as a City

Well, let me begin with an analogy, because I think that's very useful to think about. Let's look at a city, New York. Now, all of the people who live in New York know how New York works. All of the strangers who come here have to learn how it works; otherwise, they just don't understand it at all. And cities in general work at least to, let us say, a martian observer as follows. There are a whole lot of units, they're called houses. And every morning they dissociate subunits. The subunits then go in all different directions where they aggregate again into other complex assemblages, which have names on them like banks, hospitals, stores, etcetera, etcetera. And if you study the function of the city, you can see that the function must be defined in terms of these assemblages. So for example, you'll see that banks, money goes in, lots of it, very little comes out. Schools, all of these functions. Right. So I want you to envisage all of that complexity, all of those, if you like, all of those interactions. And they're not to be considered as a matrix of interactions of all the people in New York with each other, but simply that there are—that matrix indeed is quite sparse, but there are, if you like, strong interactions and weaker interactions. The strong interactions between all the employees of the bank and another set of strong interactions of all the employees in the hospital.

Now, I would like you to think of the genome sequence of the white pages of the telephone directory. Now, what we hope is that they've got it all accurate, there are no mistakes. When you call the dean, you won't find yourself in the House of Pleasure, for example. So we will have that. This is a list and of course we want that, and of course the one thing you can do with sequencing is obey what I call the CAP principle: complete, accurate, and permanent. You never have to do it again, it's a permanent resource in biology once we have done it.

What people are thinking about doing now is to compile the yellow pages; this is the annotated genome. Now I mean it's a great thing to know there are seven plumbers on one block, in fact, just the existence of plumbers might make you think there must be something for them to plumb, so maybe there are pipes hidden underneath those streets, which plumbers will get into. But, you see, you will still have a fragmentary description, just glimpses of pieces of it. So you must reflect the organization of it, and you must put that organization in right from the start.

Well, let me give you one quick example of what I propose, and that is the following. Everybody will hold their hands up in horror and say, "Look in a cell!"

Maybe we have 20,000 genes expressing. How are we going to cope with 20,000 different polypeptide chains?" The one lesson you should learn is that if you can't cope, neither can a cell. And cells have learnt to view the world as composed of something like income tax. In other words, it's criminal to evade, but there's a legal means of avoiding paying income tax. And I think what we have to say is when we find things that prima facie would just seem impossible, you can be sure that evolution has found the way of avoiding them. All right. And this is exactly what nature has done.

The Cell as a System of Gadgets

So the first thing to realize is a cell is not a bag of protein molecules all buzzing around and interacting with each other, it's not like that at all. When you go into it you find that no gene product hardly ever acts on itself; it usually acts within a molecular assemblage. And some of these assemblages are quite complex. The assemblage that actually splices out introns contains the products of 65 genes. And so when you go through this, you'll realize immediately that there is a strong reduction of the complexity, perhaps by an order of magnitude, so that if I call these things "devices" or "gadgets" rather than the term "molecular machine" because I think these have much more important functions to reveal than you might have expected from a mechanical device, if you call these "devices," we have condensed the problem immediately into 2,000 such gadgets that exist in the cell.

And the next thing you observe is that the cell is not homogeneous, it has a topography. What goes on in the nucleus is different from what goes on in the cytoplasm, and what goes on at a membrane is different from what goes on in the mitochondria. And just for argument's sake, let us just think of this as again ten topographical regions in the cell. So that the problem is reduced then by thinking about it through this kind of modularization. And now we're only thinking of two hundred gadgets or so per topographical region, and that means that we can start to look separately at how these gadgets communicate with each other, and the problem then becomes totally digestible, in my opinion. And if you look at the way it's done, it's done stepwise. And let me say that it's all written in the genome like that, because it's written in the genome of a protein that it should have a certain sequence of amino acids in a certain configuration so it can bond with another protein which has similar properties, and that these then can bond with yet others. So that is what I call strong interactions.

And so you come to a model of—the cell model being a theory that effectively there is a—we would look at it as a graph of gadgets connecting with signals passing between these, and you'll see into that logical structure we can easily assemble not only true signaling, but even metabolic pathways can be thought of as signals, chemical signals, passing between gadgets and connecting between them.

The Next Fifty Years: Networks of Cells

Now, of course one question we'd like to know, and this has been purely a sketch, is how many different kinds of cells are there in the body? Well, you can go to a textbook of histology—I used to teach histology—and maybe there are two hundred different cell types, smooth muscle, striated muscle, etcetera, etcetera. You can write a list. Now when we come to the nervous system we just don't know. It is quite feasible that in the retina that 26 kinds of amacrine cells have been characterized. So there may be, for all we know, in terms of what I call noncontingent states, that is, states that do not depend on environment; there may be as many as a thousand such states in the brain. That remains to be found out, and effectively I think that is the first task that we have to do, is to discover how many of the noncontingent states there are.

Now of course, there are contingent differences as well, differences due to stimulation, differences due to learning. But I don't consider a neuron that has learnt something as a different cell type from its naïve neighbor, it is just something—because the capacity to learn, the capacity to learn is in effect a contingency—a noncontingent state, which allows the thing then to, so to speak, fill in the form with what its connections with environment [are].

So if I were to return to Earth, as I'm threatening to do for one day on April 26, 2053, which is the hundredth anniversary of the publication of the Watson-Crick model, in order just to have a look around and see what's happened—promise to go back from wherever I came from—just want to have a look, but I hope that by 2053 we would not be groaning as we are today saying, "What are we going to do? What's all this information?" Because I think that if the last year, the last fifty years, has seen us elucidate the nature of biological information at the chemical level, if you like, it's just been the chemistry, which means also the physics of biological information—I think what we are going to do in the next fifty years is to elucidate the biochemistry, if I can call it that, of biological organization. And I think we will come to see that we will be preoccupied with networks, which are graphs, that we will see, I hope, organisms as a network of cells, with cells connecting with each other, just as we will see the cell as a network of molecules. And that once more, I hope, that we will use the structure into which we will feed the genome, and not try to hang everything on it.

Rethinking the Gene

So I just want to finish up with one last remark. When after Herculean efforts it had been discovered that we only have 30,000 genes, amazing to some, insulting to most, that they had eight times the number as *E. coli*, people were very puzzled as to why we have so few genes. When it was discovered effectively as could be well predicted that all vertebrates will have the same number of genes, because evolution can only take place at this level, at the level of reorganizing what you do with the same genes rather than having explicitly

human genes—in fact, the way we think about a gene is totally wrong today. We think about a gene for some function, and I like to tell the story—which I happen to have invented—about the difference between chimpanzees and humans. And everybody has felt that when they get the final sequences humans will have one extra gene, the one for language, what we call the Chomsky gene. But of course what you don't realize is that chimpanzees have learnt that talking gets you into trouble—just look what's happening to us—and so they evolved a language-suppressor gene. So they'll have the extra one, and it'll be called the Chimpsky gene, of course. But I give this as a lesson, you see, because now language must surely be a very complex thing; it involves vocalization, you can find a huge list of things, and you don't go from not speaking to speaking with one gene. I mean otherwise we would see the mice, the perfect model for dyslexia. Mice cannot read, and therefore we should study dyslexia with your ordinary mouse.

But one thing that's interesting is complex systems can be broken in many different ways and [in] the way of breaking them, each of these appears as a unique gene. So what we are really interested in are the genes for normality, if you like, and we're interested in finding out how those work. And only when we find out how those work will we understand then how you do this.

Solving the Forward Problem

Now, I said I would follow up on something which I think has only come home to me in the last few months. There are people who will tell you the following: you should study the system. In fact, I believe a whole subject has been created, called systems biology, which we used to call physiology, actually, but if they want a new name for it, fine. But systems biology has a program, or at least certain proponents of it have a program, which we should ask whether it's feasible at all to do it this way. Now the idea being that when you study a lot of things at the same time, there are some things called emergent phenomena, things will happen there. Because everybody will tell you that a system is greater than the sum of the parts, and when you put a lot of things together new things emerge.

Now that statement itself is nonsense; it's nonsense because they haven't quite the correct definition of a system. A system, the whole, is the sum of the parts and their interaction, okay? Because it is true that the whole is greater than the sum of the parts studied in isolation, that *studied in isolation* is what has been left out. The whole can not be more than the sum of the parts and their interactions, or else we're getting off into a nonscientific view of how things work. There can't be mysterious essences flowing around.

So, if you make a set of multiple observations, could you deduce what's going on in the system? Could you model the system as it's expounded and deduce what's going on? These are classic inverse problems—look at the result, try to work out the causes, look at the effects, try to work out the causes.

Now, such systems are generally also called ill-posed systems, because they don't give rise to unique solutions, they're ambiguous, they're not continuous; this whole thing was specified as to what a well-posed explanation will be. And what I want you to do is to think of the following: if someone plays a drum and you record the sound and you make an analysis of the sound, could you work out the structure of the drum, the physical structure of the drum? Now that is a classic inverse problem; it is also classically ill-posed because information has been lost due to interference and all sorts of other things. And so therefore only if you make certain assumptions, which is called regularization, can you effectively use the sound to tell you about the drum.

But there's another way, and that's the way I think science should go. Get hold of the drum and find out what it looks like; what's it built out of, how big is it, and then you can solve the forward problem which is easy, you can play the drum yourself. And that is what I think is going to be the essential methodological difference. So if we were to call systems biology of going from the end result, from the phenotype, to try to deduce what is inside the organism, I think that it is doomed to failure. On the other hand, if we find out what is in the organism and solve the forward problem, that is the only path to success. So if [we] want names, we'll call the latter part computational biology because we're going to compute what there is.

And as a last word I would say the following: although people have said "more is better," I would like to point out that the least is best. That is, find the least you have to find out and you can predict the rest. That's what science is all about, and I hope that we will encourage people very strongly not to become factory hands in some vast institute, but to really get out there, find those drums, you can play them yourself.

Thank you very much.

Isidore S. Edelman: We're now scheduled for what is generally called a coffee break, but you're not obligated to drink coffee. The—it is my intention to restart the symposium at 10:30 sharp. Thank you.