**Viral Outbreak: The Science of Emerging Disease**
**Lecture 2 – The Virus Hunter's Toolkit**
**Joe DeRisi, Ph.D.**

**1. Begin of Lecture 2** (0:16)

**[ANNOUNCER:]** From the Howard Hughes Medical Institute. The 2010 Holiday Lectures on Science. This year's lectures," Viral Outbreak: The Science of Emerging Disease", will be given by Dr. Joseph DeRisi, Howard Hughes Medical Institute investigator at the University of California, San Francisco, and by Dr. Eva Harris, Professor of Infectious Diseases at the University of California, Berkeley. The second lecture is titled "The Virus Hunter's Toolkit". And now to introduce our program, the Vice President for Science Education of the Howard Hughes Medical Institute, Dr. Sean Carroll.

**2. Welcome by HHMI Vice President Dr. Sean Carroll** (1:08)

**[DR. CARROLL:]** Well, thanks for joining us for the 2010 Holiday Lectures on Science focusing on viruses, and emerging infectious diseases. I want to applaud the students on the great questions. Keep them coming. I think this audience and our home viewing audience might enjoy visiting a couple of HHMI websites. One is BioInteractive, which is loaded with animations, film clips, videos, virtual labs, and our Ask a Scientist website where you can ask a scientist a question and one of our volunteers will answer it. Our next speaker this morning is Joe DeRisi. Joe's a terrific example of a new breed of biologists, who has married a passion for computing and technology with a passion for biology. And Joe has led his research group to invent new technologies and to push existing technologies to their limits in the pursuit of a deeper understanding of biology. Joe's contribution to the invention of what he's dubbed the Virochip is one famous example that you'll hear about. In this lecture, Joe will emphasize that the basis for all of this sophisticated technology starts with understanding the fundamental chemical nature of DNA and building from there. So before he begins, a short video to introduce Joe.

**3. Profile of Dr. Joseph DeRisi** (2:38)

**[DR. DERISI:]** UCSF's a very special place. The focus is biomedical science. We're a little unusual in that we're the only UC campus that has no undergraduates. So, our entire focus is on graduate education, our post-docs with an explicit purpose of going after biomedical research problems that have an impact on human health in some way. I've always been interested in infectious disease. I went to high school in themid-'80s and this was the time of the HIV epidemic and was really a shocking realization to me as a student, as a high school student, that there were deadly viruses out there that we didn't know about. I thought Western medicine had characterized and understood all possible viruses that were in our environment. And the AIDS epidemic brought it to my mind that we weren't even close, that there were threats lurking out there that we had no concept of. So, I immediately set my sights on becoming initially a biochemist with an eye towards being a virologist. I actually went to graduate school to do virology, and I ended up doing more technology development in genomics than virology, but I was able to bring that back home when I became a faculty member here at the University of California, San Francisco, and team up with my collaborator Don Ganem, to create a real virology discovery-based program. I had seen Don Ganem give his Holiday Lecture on the DVD and I was amazed. That was one of the coolest things ever, and I'm very invested in education and in training students and in getting students to be excited about going into biotechnology, infectious disease research, and so on. And this is a medium through which you can directly talk to a large number of students and hopefully inject some enthusiasm and some excitement into this area because we need more people who are really smart and really good to get into infectious disease because they're not going away.

**4. Leading causes of death for children worldwide** (4:47)

All right, welcome back to the second part of today's lecture series. Eva Harris gave a great introduction to some of the reasons why we study infectious disease, and she's a tough act to follow. I'm not going to dance. I do have this cool virus tie, though. So, the reasons why we study infectious disease were hopefully made apparent in the previous lecture. My goal today is to tell you about some of the methodologies and techniques we use and our toolkit essentially to go , after the detection and discovery of viruses. And why is that important, you may ask? Well, let's start off with a little seriousness here. The most vulnerable segment of our population are children under five. If we consider what are the causes of death for children under five worldwide, do you know what they are? Let's go for the biggest cause. Anybody?

**[STUDENT:]** Diarrhea?

**[DR. DERISI:]** Diarrhea, that's a great guess and that's coming up. Any other guesses?

**[STUDENT:]** Malnutrition?

**[DR. DERISI:]** Malnutrition, an excellent suggestion. In fact, that is a major cause. I'll get right to the point. Neonatal causes, so neonates are children zero to 28 days and they die from preterm birth, they die from birth asphyxiation, they die from congenital malformations, and so on. They're incredibly vulnerable. They also die from infectious disease. And so, the largest cause of death in children under five is neonatal causes.

**5. Infectious disease is a major cause of child mortality** (6:23)

Next comes pneumonia. Now you might not have guessed that it would be pneumonia, but actually respiratory distress is a serious, serious life-threatening illness for very small children. It can be caused by bacteria, it can be caused by viruses, it can be caused by fungi, and the most common cause is *Streptococcus pneumonia*, actually a bacteria, but viruses like respiratory syncytial virus, parainfluenza, and others can also cause life threatening pneumonia. And what's really sad is that this is actually treatable. Immunization, nutrition --which was mentioned earlier -- breast feeding, all can reduce death due to pneumonia. Now one of the most striking facts about pneumonia which you may not have realized is that somewhere between 10 to 40% depending upon location and study are of unknown cause. We don't know what causes them, and that's a really striking fact because if we know what the cause is, maybe we can do something about it. So, the next cause, the next largest cause of death in children under five is in fact diarrhea, which was mentioned earlier. Again, it can be caused by bacteria, viruses, it can be caused by parasites. You may, if you've been watching the news lately, be aware of this horrible and growing epidemic of cholera in Haiti right now. That's caused by a bacteria, *Vibrio cholerae*. And again like pneumonia, it's largely treatable and preventable. In the case of diarrhea, it's just simple rehydration is the therapy, yet a significant fraction of severe diarrhea or gastroenteritisis of unknown cause. I have no clue what causes it. We test it for what we know of and it tests negative. And now if we take out the remaining causes of death for children under five, they include things like malaria -- which is a parasite and the other half of my lab actually works on malaria -- HIV, measles, and then finally you get to non-infectious causes like injuries and other random things that happen to kids unfortunately. I want to make two important points about this. First of all if you takeout neonatal causes, infectious disease is the largest reason why children under five die, and a disproportionate amount of childhood mortality due to infectious disease occurs in the developing world, not necessarily industrialized nations. So, let's consider Nicaragua.

**6. Video: Many Pathogens Affect Patients at a Nicaraguan Clinic** (8:40)

Let's just look at these images here of a clinic in Nicaragua. This is from Eva Harris' group. Kids come to these clinics and present with a wide spectrum of symptoms. They may be consistent with any number of infectious disease processes, and most of the time we don't know the cause. And why is it important to know? You might say it's not important to know. I mean, knowing the cause may not affect the course of therapy you give a patient. If they have diarrhea, you're still going to rehydrate them. It doesn't change it, but I would put to you that it is critically important. Let's take rotavirus, the most common cause of severe diarrhea in kids. It causes something around500, 000 deaths per year. So, two vaccines, live attenuated vaccines, are now internationally available and they are without a doubt going to save hundreds of thousands if not millions of lives. And I would put it to you that knowing your enemy is the first step to conquering it, and that's absolutely true with infectious disease.

**7. How do you determine the cause of an infectious disease?** (9:37)

Okay, so let's take a hypothetical case. A child presents in the clinic, it's a 12-year old boy -- totally hypothetical here-- shortness of breath, has a fever, maybe a long history of fever, a cough, night sweats, and a test in the clinic reveals low oxygen in the blood, really just meaning the kid's having a hard time breathing. Treatment with oral antibiotics -- the first thing that you would do -- did not improve this boy's situation. Now, that might suggest to you that it's not a bacteria. If antibiotics don't work, sometimes it's because it's a virus. It doesn't prove that, but it suggests that.

**8. Koch's postulates for identifying an infectious agent** (10:12)

And so how actually do you go about proving what the cause of an infectious disease is? Well, this story then turns to an older story. Over 100 years ago, Robert Koch published a set of postulates, and these postulates are the guiding principles by which most people establish whether a certain pathogen causes an infectious disease. So, let's just walkthrough these really quick. First of all, you have to find the microbe in every case of the disease. Second, you have to be able to propagate that microbe in culture. Third, if you take that purified microbe and you reintroduce it into a healthy recipient, they get the disease. And then four, you can re-isolate the microbe from that newly-infected individual. Now, this is very rigorous and I can also tell you that under most circumstances they're impossible to satisfy, many for very obvious reasons. I don't know who wants to be the healthy volunteer recipient. And so in the molecular area, I would put it to you that we can actually do things that get at these postulates from a different angle and help do it.

**9. Viral plaque assay** (11:21)

There's other problems with these postulates as well, and let's take our hypothetical case of this boy. Maybe you get a sample, a respiratory sample from the boy and you try to culture the virus. This is probably one of the gold standard ways of figuring out what causes something. You put it in a dish with a bunch of cells and see what happens, and if you're super lucky, that virus will grow. So, here's a dish in which a bunch of cells are growing on the bottom of the dish and they're stained purple, and where the virus has been active, you see cleared regions, a mottled whitish pattern. And this is where the virus has either been killing cells or modifying them and so on. In the business we say these viruses will have cytopathic effect. They mess up the cells, but you'd be very lucky because most of the time, maybe 90% or better of the time, you get a dish like this, in which the cells just grow just fine and the virus doesn't do a thing. The fact remains the vast majority of viruses that we know about that are out there, they can't be cultured. And that's not to say that they can never be cultured. It says we don't know enough how to culture them. We don't understand the conditions in which they grow. And this is not just for viruses, but many, many, many bacteria as well. And it can take years to figure a model system or some sort of way to grow a virus. Let's take leprosy for example. So, leprosy is caused by a microbacterium, and for the longest time no one could grow this bacteria. But in the '60s it was discovered that if you inject it into

the foot pads of mice, it'll grow a little bit and it will also grow in the nine-banded armadillo. So, if you happen to have a nine-banded armadillo around and you tried to inject it, you could get this thing to grow. Now, it turns out the nine-banded armadillo has a lower body temperature which makes it more conducive to growing this microbacterium. And so, this technically would not satisfy Koch's postulates actually, but no one doubts that this microbacterium is the causes of leprosy. Well, it turns out that there might be a second related microbacterium that can also causes a similar disease.

**10. Identifying viruses by their structure** (13:25)

All right, so let's get back to our hypothetical case. What if you can't grow it? How are we going to discover what the virus is in the case in which nothing grows in a dish? What features of viruses can be exploited? Well, viruses come in a wide variety of shapes and sizes and structures. They can look like complicated machines like this bacteriophage. It sort of looks like a lunar lander. They can be spherical objects like HIV, bullet shaped like rabies, they can be tiny,30 nanometers in the case of rhinovirus shown in the upper right, or absolutely massive, like mimivirus which is over 400 nanometers, the only virus I might add that you can see through an optical light microscope. Don't worry. It's not that scary. It only infects amoebas it turns out. They might have envelopes, meaning they contain lipid from the host, or they may be non-enveloped. So, let's go into a little bit more detail and I'll show you a couple of viruses.

**11. Animation: Viral Geometry and Structural Diversity** (14:20)

So, shown here are three of some of my favorite viruses. I picked them for no apparent reason other than that they look really cool. I've scaled them all to about the same size for easy viewing, and despite their apparent complexity, the structure of most viruses are actually formed by simple repeats of only a few proteins, sometimes in conjunction with lipids. The individual proteins here are shown in different colors. So, this first example is a non-enveloped virus. It means it lacks a lipid component. And the capsid or the shell is in the shape of an icosahedron. In geometrical terms, an icosahedron is formed by 20 identical triangular faces. Five faces meet at each of the 12 vertices forming a pentamer at each apex. And if you look closely, you can see it. All right, so take a look at your virus models. Who has this virus that I'm showing on the screen right now? Hold them up. Hold them up. If you're not holding up the red virus, put it down. So, the goal of this exercise is to correspond the images on the screen which are multicolored with the sort of the solid color here of the virus plastic model. So, I printed these on a 3-D printer from crystallographic data structures so that they're molecularly accurate, and if look very carefully you can spot the pentamers at each of the apexes. Now, your playing cards also have more details on each particular virus. This happens to be a nodavirus. The nodavirus is a small RNA genome virus.The RNA is in two parts. I think it's kind of cool because it infects both insects and fish, and under these certain precise conditions can infect mammals, which is kind of very unusual because there's not many insect viruses that can actually infect mammals. And so, my suspicion is there are a lot more nodaviruses out there than we realize, and what they're doing in the environment and what else do they infect is going to be an interesting mystery to solve soon. So let's put nodavirus back into the queue here and bring out our next contestant. It's a non-envelope virus just like nodavirus with an icosahedral capsid. The pentamers at this case in each vertex are fairly easy to see. The rhinovirus that causes the common cold is an RNA virus. It's a plus strand RNA virus. Once it gets in, it makes more of itself, and gets out fast. It's really fast and it does its thing with incredible efficiency. That was rhinovirus and I'm just going to move ahead to dengue virus.

**12. Animation: Structure of Dengue Virus** (17:01)

Dengue virus is more spherical, but it also has an icosahedral structure. In this case, the glycoproteins that are on the surface float on a lipid membrane that is host-derived. So, who here has dengue virus? I

better see the blue ones this time. Right, and so you can see it's very spherical, but if you rotate it very carefully you can see the pentamers in each of the vertices. All right, so now we're going to go into a little bit more detail on dengue virus since it's one of the subjects of our talk today. If you rotate it very carefully, I'm going to highlight for you the pentamer that is at each vertex -- so there it is there --and if you're very sharp, you can spot a threefold axis of symmetry. This corresponds to the center of each of the triangular faces. Now, this whole structure amazingly is formed by just dimers, pairs of glycoproteins, and they're highlighted right here. As I mentioned, this whole thing floats on a lipid bilayer. So if we strip away the glycoprotein, you can see the lipid underneath. And if we blow away the lipid bilayer, you can see a protein core, the nuclear capsid of the virus which encases the genome. So, now let's go right to the heart. Dengue is an RNA virus genome, and the RNA is packaged within this nucleocapsid. It looks like it's going to hit you right now.

**13. Antibodies can recognize specific viruses** (18:35)

So how do you detect viruses? Let's get back to our fundamental question. In the absence of culture, knowing that there are all of these shapes and sizes and different structures, what can you use to detect viruses? What tools do we have that can recognize one virus versus another? Any suggestions? I know some of you a recertified gene jockeys out here. There's one way in the back.

**[STUDENT:]** ...how they infect the cells and how they transfer their DNA into a cell?

**[DR. DERISI:]** The mechanism by which they replicate perhaps? Sure. That's one way we could distinguish. That would take some sophisticated tools. How about another suggestion?

**[STUDENT:]** Would it be antibodies?

**[DR. DERISI:]** I like the way you think. That's right. So, the structure and shape of these viruses will elicit different immune response, and that's absolutely correct. So when a person is infected with a virus, their immune system mounts a protective response, a defensive response to generate antibodies to these things. And antibodies as you probably know are very specific proteins that can find very specific structures on viruses. And so, you could use this as a tool. How would this work? Well, let's just say you could make a lot of that dengue virus or any kind of virus, and if you could put it on a solid substrate like a plate or something and then take serum from your patient and expose it to that virus that's stuck on the plate, if antibodies stuck to that virus which you put on the plate, you would say that they had mounted an immune response to that virus, therefore they must either have dengue virus or had dengue virus in the past. But how would you know that there's actually the dengue virus antibody really there? It's kind of tough to see the antibody. So, what we actually use are secondary antibodies, which are really just antibodies to an antibody. In fact, you could use it...these are human antibodies that stick to dengue. You could generate anti-human antibodies and label them with a colorimetric probe or fluorescent dye, and then you would get a result something like this, where the blue wells are the wells that had serum in them that had antibodies to dengue, and you're visualizing them by virtue of a secondary antibody. These sorts of immunoassays are incredibly popular. There's a wide diverse array of them. I don't have time to tell you about all of them today. And there are some features of them that are important. First of all, they're indirect. They don't say that the virus is actually in there right now. They say that somebody has had the virus at some point, and they tend to not be that sensitive or that specific because there might be cross-reactivity. Antibodies might not necessarily be able to tell one virus from another so well. So, how else can you detect viruses? Any suggestions?

**[STUDENT:]** Since specific viruses, you know, they infect different types of hosts, different types of cells, each virus infects a different part of the body, host range.

**[DR. DERISI:]** Right. So, tropism, that is the range of cells that a virus can infect. That is a great idea. In fact, that has been used to differentiate certain viruses from each other.

**14. Classifying viruses by their genome type** (21:39)

How about another suggestion? Or I'll throw one out there because it's my lecture. You could go for the genetic material itself. So instead of using the structure of the virus or the proteins, go for the material that constitutes its genetic material. And so, there are strict taxonomy rules for viruses, species, genus, and family, but there's also a separate classification, first proposed by David Baltimore for the classification of viral genomes based on how they replicate. And so, let me just take you through this because knowing the genetic systems of viruses will enable you to understand how some of the diagnostics work. So, first we have DNA viruses. There's double-stranded DNA viruses and they make more double-stranded DNA viruses. But there's also single-stranded DNA viruses that replicate into more single-stranded DNA viruses. But there are also purely RNA viruses like dengue that have no DNA intermediate, and in these there are both double-stranded RNA viruses as well as single-stranded RNA viruses, composing two different strands. Now finally, there's also the retro-transcribing viruses, and these are probably the most familiar to you because HIV is one of these. They start out as RNA, single-stranded RNA, and through an enzyme called reverse transcriptase, they reveres transcribe their genome into DNA. In HIV's case, it gets inserted into the genome of the host as double-stranded DNA. But there's also another class of virus that uses an RNA intermediate, like hepatitis B. It starts out as double-stranded DNA, it makes RNA, and then that's reverse transcribed back into DNA to create the genome once again to be transmitted. So, this is the general categorization of viruses. Now I want to throw out a special acknowledgment here, not to myself, but to Don Ganem. Don Ganem is one of my key collaborators and HHMI investigator, and together what we did is leverage the genetic components of viruses to create a massively parallel assay for simultaneously diagnosing and discovering viruses in literally any biological sample. And I'm going to tell you about that technology right after our question period here.

**15. Q&A: How can you stop viruses from spreading to other cells?** (23:47)

**[STUDENT:]** So, I wanted to know since like the virus, like it takes the shape of the real RNA or DNA, so how do you like stop it from spreading in a specific cell?

**[DR. DERISI:]** Your question is how can you stop a virus from spreading when it's already infected a set of cells, for example. Well... and I assume you mean and the question is how can we intervene in this process? Well, so therapeutics do exist for viruses, although they are not many. So for example, one can make a nucleotide analogue, a part of the DNA that's not really a legitimate part, and the virus might accidentally incorporate that into its genome and that could be a lethal event. It could stop it from replicating there. There could be protease inhibitors. A lot of viruses contain proteases. You could throw a drug in that might stop those. There are enzymes like integrase in retro-transcribing viruses like HIV, and you could have an inhibitor to that that would prevent it from getting in the genome in the first place. Many, many, many strategies, as well as drugs that could block entry of the virus into the cell in the very first place. So, there are avenues to block viral replication and propagation.

**16. Q&A: What kind of antigens do viruses have?** (25:03)

Another question? Question in the back.

**[STUDENT:]** I wanted to know why these viruses don't have antigens like any other normal virus?

**[DR. DERISI:]** Don't have antigens. Well, all viruses have antigens, actually. So, it turns out that the virus is an antigen that is a foreign protein component, and your body will make antibodies to these things. So, the one thing that might surprise you though is that just because your body makes an antibody to a virus doesn't mean that antibody is protective. It might be specific for the antigen in the virus, but not all antibodies do their job very well. In fact, a lot of antibodies don't do their job very well at all.

**17. Q&A: Where do secondary antibodies come from?** (25:46)

Question in the front here.

**[STUDENT:]** You said the way that they see the cell is through the antibodies, right? And then the antibodies to those antibodies or whatever, the secondary antibody, the ones that see. Why those are present in, like, humans, is the secondary antibody, you're saying. Why do we have those?

**[DR. DERISI:]** Sure. Let me... I must not have been a little too clear, and so I'll just make sure that everybody's on the same page there. So in these immunoassays, the primary antibody, the first antibody that comes in is from the serum of the affected patient and it reacts with the virus. The second antibody is not from the patient. The second antibody...good, I'm glad I clarified that. I didn't want you all to think you have anti-human antibodies in you or else you'd all have, you know, an autoimmunity, you know, problem. Those secondary antibodies that are anti-human, they usually are goat, rabbit, or something else, and they're generated specifically to be anti-human. Yeah, it's cool that you don't have anti-human antibodies in you.

**18. Q&A: Do lipid membranes come from other organisms?** (26:48)

Yeah, right here in the front.

**[STUDENT:]** Is the lipid envelope the viruses have, like can they hold lipids from other places, like from other organisms?

**[DR. DERISI:]** So, the question is where the origin of these lipids come from and can they get them from other organisms and so on. And that's a really interesting question actually that pertains...it's very closely linked-- your question -- to the actual viral life cycle in question. There are some viruses that are just human to human transmission, and for the viruses that are so-called enveloped that have a lipid component derived from the host, that lipid component will be in fact human all the time as it transmits from human to human, but that might not necessarily be so in the case of a virus that has an obligate stage in some other organisms like an insect. And maybe those lipid components could be exploited in some specific way if they differ. That's a very interesting question. I'd like to talk to you more about it afterward. Okay.

**19. Q&A: What are the advantages of different viral replication mechanisms?** (27:45)

A couple more questions. I've been too much in the front and then I'll get back.

**[STUDENT:]** What is the advantage for viruses that transcribe to RNA or DNA as opposed to being consistent?

**[DR. DERISI:]** The question is what is the advantage of having one replication strategy over another, that is double-stranded DNA viruses, single-stranded RNA viruses, and so on. Well, that's an interesting question. It may not be so much an advantage, that is they're different kinds of life cycles and it might be

comparing apples to oranges in some cases. For example, HIV, a retro-transcribing virus, inserts its genome into the host. It becomes a permanent part of the host, and in many cases it goes latent, that is the virus is quiet, it's not doing anything, it just inserts itself, and then infrequently might reactivate and re-emerge. That's an incredible successful strategy because it allows the virus to hide out for a while. Now, there are viruses like the rhinovirus that causes the common cold. That's the white virus that you guys - - all have here. The rhinovirus that causes the common cold is an RNA virus. It's a plus strand RNA virus. It has no DNA intermediate and it has no capability, at least that we know of, to become latent or hide out in the genome somewhere. It has to go. So once it gets in, it makes more of itself, and gets out fast. And so, it doesn't have this sort of strategy of hanging around. On the other hand, it's fast. It's really fast and it does its thing with incredible efficiency. And so, they've evolved along different fitness tracks for different kinds of lifestyles.

**20. Q&A: Are there innate defense mechanisms to detect viral DNA in the cell?** (29:19)

Question?

**[STUDENT:]** I know that some cells have protein receptors that recognize viral RNA as a signature for viruses, but then would viral DNA be more subtle or do they also have a signature that some cells can recognize?
**[DR. DERISI:]** These are great questions. You did a lot more reading than I suspected a normal high school kid would do. So, you're referring to innate defense systems, and basically there are numerous different innate defenses for viruses in host cells. One of those as you mentioned is detecting viral RNA, and a current area of intense study are in what are the DNA sensors because viruses, when they enter the cell, go through the cytoplasm, not necessarily to the nucleus first. And so, there are so-called cytoplasmic DNA sensors that then can react because there shouldn't be DNA there. And so, yes there are, and the actual mechanisms by which those sensors work and activate downstream, you know, defense against that virus are all being worked out right now, similarly to the way they were with RNA viruses.

**21. Concept of chip-based tool that can detect any virus** (30:22)

All right, so I foreshadowed this new technology that Don Ganem and I were going to build, and the idea was really simple. You'd like to have somebody just come up and sort of, you know, sneeze or cough up something onto a -- if they had a respiratory disease -- onto a chip, some piece of, you know, glass or something where you could do a simultaneous diagnostic for everything. Now, I'd like to tell you about how this chip works and what we did with it and so on, but I also want to make sure we're all on the same page with respect to how, you know, regular DNA diagnostics and genetic diagnostics work. Now, I know some of you probably work in labs or volunteer in labs and are familiar with all this molecular biology stuff, and I know some of you aren't. And so, let's first step back and discuss some basic principles of genetic detection and of DNA structure, and then we'll get into how we can exploit them and use them in new ways.

**22. Animation: The Chemical Structure of DNA** (31:22)

All right, so you're probably familiar with the general appearance of DNA as it's represented here in this computer model. I'll make it active. It's a double helix, two polymeric strands. Now, if we unwind those strands we can see the stacked nucleotides that form this linked structure. And what I'm going to do is take five of those nucleotides out of the DNA helix in diagrammatic form and bring them out here so that we can go through them in just a little bit more detail. So as I mentioned, DNA is a polymer of linked nucleotides, and these nucleotides are little repeating units that go down one strand and on the other strand they go the other way. So, these are anti-parallel strands. We'll get into why that is in a

minute. The major structural components of DNA is that they all share a deoxyribose sugar backbone and nitrogenous bases. The nitrogenous bases are the information carriers of the genetic material. The nucleotides themselves are linked together by phosphodiester bonds, and if we zoom in here you can see why DNA has a negative charge. Now, these DNA molecules have these nucleotides that are linked in a directional manner. So, if we zoom in on the ribose you can see that the linkage of the phosphodiester bonds occurs at the 5' carbon -- and you can see the numbering of the carbons around the ribose here -- as well as the 3' carbon. And so, DNA is inherently directional. We read and write DNA from the 5' direction to the 3' direction. That's not coincidentally the direction the enzymes which replicate DNA also extend a growing chain. All right, now back to the nitrogenous bases. There are two kinds. There are purines, adenine and guanine, and there are pyrimidines, thymine and cytosine. And there are also rules governing which can bind to which by hydrogen bonding, the so-called base pairing rules. And in the base pairing rules, cytosine can bind with guanine, and adenine can bind with thymine, and this forms the basis of the genetic code. Now, in RNA thymine is actually uracil, small point. Now, a beautiful little feature of knowing this hydrogen bonding and pairing rules is that if you know the sequence on one strand of DNA, you automatically know the sequence on the other strand of DNA. So, if this reads A-C-G-T-T across 5' to 3', on the other strand going the other direction it's going to read A-A-C-G-T. Now, put these back into the helix, I'll tuck them in. So, here are the nucleotides. We'll put them back in there and we twist it back up into this helical, right-handed helical spiral of anti-parallel strands where you can clearly see that the bases form sort of a semi-stacked ladder, like steps on a staircase.

**23. DNA hybridization an detect viral DNA in a complex mixture** (34:07)

All right, so by virtue of this structure, there's a really wonderful property, and that is if you heat up a double helix to a very high temperature -- let's say the boiling point of water -- eventually those helices will melt apart. The hydrogen bonds are non-covalent bonds, so if you put energy into it, they will fall apart. Now, that's not too surprising, but the wonderful thing is if you cool the temperature back down, these strands will come back together exactly the way they were. Those base pairs reform, the strands will reanneal or hybridize, and you'll get back what you started with. So, this is a pretty interesting property of DNA because it enables a lot of genetic diagnostics and enables this idea that we can pick out a sequence from a complex mixture. So, how might we do that? Any suggestions?

**[STUDENT:]** ...restriction enzymes.

**[DR. DERISI:]** Right, restriction enzymes, if you cut DNA at specific points, but there are diagnostics that are based on restriction enzymes. So, what I'm thinking of here is that one could actually make a DNA sequence that's complementary to the virus you're interested in and use it as bait, as a hook to pull out sequences that are in your sample. And if it pulls something out, then you know it was there. If it doesn't pull anything out, chances are it wasn't there, at least you hope. But there's a problem here. Let's take any ordinary infection. There might not be that much of the virus nucleic acid around. It might be one part in a million. It might be one part in 10 million. In fact, so little that you're not going to be able to detect it by the strategy of just yanking it out of the sample by hybridizing it to a complementary strand, but fortunately there are enzymes that can copy the DNA.

**24. Replicating viral DNA in vtiro** (35:45)

And the first step in replicating DNA with an enzyme and making more of it is just denaturing it like I said, and then we can actually put in a synthetic primer, a piece of synthetic DNA that's complementary to the sequence we're interested in, like a virus sequence. When we cool that stuff back down, the primers are going to sit down, hybridize, reanneal to that template strand, and form the basis of replication. There's a free 3' hydroxyl at the end of that primer, and if we add nucleotide triphosphates as

well as a polymerase like a bacterial polymerase or something and heat that guy up a little bit, the polymerase will dutifully copy the template strand. And if it does that, we've made another copy, which is great. So, now you've got...you started with one copy. Now you've got two copies. And if I repeated that process, I would have three copies, and then four copies, and five copies, and it would be a linear amplification process. Now, I said there might be one part in 10 million or something like this is, and to be able to do any reasonable diagnostic, I need like, you know, 10 million copies or a billion copies. So, this linear amplification process, while clever and tricky, isn't that efficient.

**25. Kary Mullis' contribution to polymerase chain reaction** (36:52)

So, who's this surfer guy and how did he solve this problem? Right. You guys are quick. So, this is Kary Mullis and he got the Nobel Prize for inventing the polymerase chain reaction in 1993. Here's the cover of his book. It's an interesting read. And what his brainchild really was was he said okay, you can put one primer in and you can make a copy. That's great, but if you put two primers in, one on the other side of the region you're trying to copy, going the other direction that is complementary to the other strand, then in the process of heating the sample up and copying it, you can achieve exponential amplification. So again, the situation is the same. You add nucleotides and polymerase and the polymerase will copy the strands, which is great, and now instead of having, you know, one copy, you have two copies. And the next cycle you will have four copies, and the cycle after that you'll have eight copies, and so on and so on and so on. So, if you did 30cycles of this, you would have a billion copies of that sequence that you're interested in. Now, one of the innovations of this whole process was to add a thermostable polymerase because it turns out that a lot of polymerases just fall apart at high temperature. But there are bacteria that live at phenomenally high temperatures. *Thermus aquaticus* naturally lives at 158 degrees Fahrenheit, and it turns out enzymes that are isolated from *Thermus aquaticus* are naturally thermostable, as you would guess. And so, if you added this to the mix, it creates a good reaction.

**26. Animation: The Polymerase Chain Reaction (PCR)** (38:30)

So, I'm going to show you a little animation here of how PCR works. So again, the cycle begins by heating your template strands up to a high temperature. You can see a little thermometer there and they bust apart. And then what happens next is primers are added that will then bind to the region you're interested in because you've designed them to do that. And then polymerases are added. You can't see the nucleotides here, but copies, it makes two copies. That's the end of the first cycle. After the next cycle, you heat it back up and repeat the process. Again, more primers come in, then polymerase will follow after that, strands are made, and we go on to the next cycle from there. And so, you can see in short order how this would make a lot of copies really, really quick, and for literally any target region that you're interested in, whether it be part of the human genome, a virus, a bacteria, or parasite.

**27. Various PCR machines** (39:25)

I wanted to show you what PCR machines look like. So, these are the typical PCR machines we use to amplify. The ones on the pictures on here are PCR machines from my lab. You know if I had remembered to turn this on, I would have said that in the time it took to give this lecture, this would have copied, you know, a billion copies of DNA. This is a rather old model. It doesn't have stickers like mine, so it's not as cool, but I like it. So, just because you amplified a sequence doesn't mean you know what the sequence was. One of the beauties of PCR is it's very specific, but another interesting feature of PCR is it also generates a lot of garbage. That's just the reality of doing it in lab. And so, the gold standard is reading the sequence. You never know really what you have until you read the sequence. So, you might do a PCR amplification, amplify part of a virus, think you have the virus, you sequence it, read the sequence, and it turns out not to be the virus or not to be what you thought it was.

**28. Sanger method of DNA sequencing** (40:21)

And so, I want to review Sanger sequencing. So, in the '70sFred Sanger came up with the predominant way DNA is sequenced today, and it's called Sanger sequencing named after him. It relies on a really interesting feature. Instead of just doing strict primer extension like we did before, where we just take an enzyme and extend from a primer, we'll also add a special nucleotide, a deoxy terminator. These are the same as regular nucleotides except the lack the 3' hydroxyl. So when they get incorporated, the polymerase can no longer extend beyond that point. Now, the next trick is we can also fluorescently label those terminators so when it stops, it can stop with a particular color. The next step is after amplifying or doing several cycles of this incorporation with these deoxy terminators, each one being incorporated in some random location because they're at low frequency in the total mixture, you end up with a population of molecules that are then terminated at essentially every base in your sequence for some number, some range, usually between 100 and 500 bases. You can then denature those templates and then separate them by size. So, we haven't covered gel electrophoresis, but basically if you put DNA in an electric field through some porous matrix, templates can be separated by their total size, and that's what's done here. And at the end of the separation matrix there's a little laser that illuminates the fragments coming down, and those will tell you which base is being read out at which position, and you end up with a chromatogram shown right here. That's the fluorescent signal at each one of these positions as a function of time or size of DNA. And then you can just look at those peaks and directly readout the sequence. So, that's the gold standard.

**29. PCR and sequencing for viral diagnostics** (42:09)

Now, how would this work in viral diagnostics? You would design primers for a specific virus, you would amplify it, use PCR to amplify from a biological sample like our hypothetical case, and then if you got a product, if you got something that amplified-- which you usually do-- you then would sequence that product to see if it's really real, there's something that's actually the virus there or not. And this can be a very tedious process because, well, there's literally thousands of viruses and if you want to do this on every single virus, you're going to need a lot of material, a lot of source. So in practice, people only do this kind of PCR amplifying sequence thing on a handful of viruses. You know if you do five viruses on a sample, you know you've gone pretty far, but we want to do more than that.

**30. Virochip designed to represent all known viruses** (42:51)

So, the technology that Don Ganem and I came up with was all right, let's make a DNA chip. Let's use that hybridization principle to our benefit. And so, the virus chip is essentially a solid substrate, a piece of glass in which we've printed synthetic pieces of DNA complementary to every virus that's ever been discovered. So, you can chemically synthesize DNA like a primer. Well, we made long ones, 70 nucleotides, and I built a robotic printer that could then deposit very small amounts, nanoliter-sized drops of different virus sequences on a piece of glass which we call a DNA chip. And then we can use that as the hybridization tool. What we do is we can then amplify the material from the virus and deposit it, let it rehybridize to the chip, and where it hybridizes, we can see what the signal is and what those spots correspond to. And so in this way, we can test for 22,000 viruses at the same time thereabouts.

**31. Evolutionarily conserved sequences can identify unknown viruses** (43:49)

Now, which sequences would we put down? Would we just putdown any old sequence? Not really because I want to build in a special feature to this chip. We wanted it to detect unknown virus as well. Can anybody guess how we managed to detect unknown, previously uncharacterized viruses for which we do not know the sequence?

**[STUDENT:]** I believe the way that you did it is you created a null set so if it didn't react to anything on the chip, you had something which would always react. So, it would react to, for example unknown virus X would react to the very last thing on the chip and that would be all it would react to. And every single other virus that we know reacts to let's say 52 and then your last one.

**[DR. DERISI:]** Yeah, so that's an interesting scheme, and that scheme relies on us knowing essentially... having a big knowledge of all of the sequences that are out there and all the possible sequence diversity. That's a clever idea. That's not how we did it, but I like it. I like it. I like the way you think. All right, any other ideas? Think evolution. Not evolution. No. Okay, so evolution is the key here. So, if you take a look at different viruses from a given family, let's say the common cold and polio which are in the same family, you realize very quickly that they share a lot of their genome in common. That's because they evolved from common ancestors. And so, many viruses, even ones that create vastly different diseases like polio and the common cold, actually share substantial segments of their genome, almost identical, that those sequences are so important that they haven't changed very much. So, this is a diagram of a picornavirus, a positive strand RNA virus here, and if we take that virus and compare it to another virus, virus one, we'll find that there are certain segments because they're in the same family that are highly related, almost identical. In this case it happens to be the 5' end of the virus and the downstream part of the virus where the polymerase is, and that's because these functions are so critical and some important and so well adapted that they don't change very much over the course of time. If we compare that to virus two, we'll see that there are segments that are a little bit different, but the general regions of conservation are essentially the same. And three and then four, and again we're nearing the region for which we can say it is evolutionarily conserved across a broad section of different viruses. And once we get to a certain point, we can say all right, I'm going to take those DNA sequences or those RNA sequences in this case that represent the most conserved sequences among these viral families, and I'm going to stick those on my chip because if there's a new virus in this family, chances are it's going to have that conserved sequence just like all the others. There's a limitation to this technology. The virus is from Mars, it doesn't belong to a family that we know about, or is really, really different like beyond about let's say 50 to 60% identical. The chip is going to miss it because hybridization relies on those base pairing rules and it can only tolerate so many errors before that process won't work.

**32. Animation: Running a Sample on a Virochip** (46:56)

Let me take you through a little video now of how the virus chip works. So again, it's a solid substrate of glass, and on that glass we've printed or deposited lots of DNA in little teeny spots. So, there's little spots obviously not to scale. Some of these spots might represent SARS for example, just highlighting them here. Some of those spots might represent sequences from dengue. There's no order to the spots on the array. And if we zoom in --obviously not to scale -- you can see little DNA molecules stuck to the glass, and these are deposited using robotic printers of our design in production in my lab. Now, a sample is applied to the chip. Come take a fantastic voyage with me now as we then drop onto this chip, sort of Road Runner style. And this sequence...yeah, you know what's going to happen. These sequences that are labeled with fluorescent dyes will then be allowed to hybridize to the sequences that are on there. So, here's this complex mixture of nucleic acid. A lot of it's not virus, a lot of it is host, only some of it is virus, and this is a virus spot and one sequence finds its match, complementary base pairs, hybridization occurs, and by virtue of hydrogen bonding and energy, that sequence stays there. Again, another sequence touches down and another and so on. Finally, the last step of this process is to just wash away all this material that doesn't belong, and we're left with the good stuff, the fluorescent labeled stuff. And then a giant laser beam-- it's really small actually but it looks giant here -- comes zooming by and illuminates the fluorescent dyes that are stuck to this nucleic acid, and that allows us to visualize which spots had hybridized material to them. So, we zoom out from the array. We should see some complex pattern of spots that are lighting up with fluorescent dyes to indicate what was in the sample. Now, you're going to have an opportunity for those of you that signed up in the bioinformatics exercise

to take data from real patient samples of actually-infected individuals and analyze the raw data from these chips, and your goal in that bioinformatics exercise is to tell me what virus was afflicting those individuals. I'm going to...it's going to be exactly what you do if you were a graduate student in my lab. I'm not going to pull any punches.

**33. Using the Virochip to identify viral pathogens** (49:11)

All right, so what do we use this chip for? We use this chip for a variety of explorations, some of which I'm going to tell you about tomorrow. We've investigated a wide range of respiratory diseases, gastroenteritis, diarrhea, we've discovered new viruses like human cardiovirus, we've used it in critical care settings like in emergency rooms settings, we've used it to investigate possible viral links to cancer. We might not know it, but about 15% of all cancers are due to infectious causes. We've used it on honey bees, snakes, birds, you name it. And importantly, we've also established good collaborations like with Eva Harris. We've leveraged existing cohorts to go in and look at other possible infectious causes of disease that weren't previously recognized to even be present.

**34. Q&A: How do you interpret the complicated pattern of Virochip output?** (49:53)

And so with that, I'm going to take questions. I'll start over here.

**[STUDENT:]** Okay, so you have the DNA on the substrate and you drop the virus onto it. Okay, and it lights up a specific pattern. So, you have these patterns memorized on like a computer, and that will tell you the virus, what virus it is?

**[DR. DERISI:]** Right. So, your question is, so you've got this complicated pattern. Are you just...is the graduate student just eyeballing the pattern and coming up with the answer? Hard, hard to do that. So, what we do is we actually wrote a lot of bioinformatics software to go after this, and the software works as follows. What we do is we take every virus that's ever been discovered in the database, and using an energy model we predict what the hybridization pattern would look like if we happened to have a sample of that virus. I'm never going to have Ebolavirus in my lab, I hope...it's not a good one to have in the lab, but I want to know what the hybridization pattern of that virus would look like on my chip if I did have it. So, we create a catalog, a matrix of all of these hybridization patterns, theoretical patterns, and then when we take an observation, a real patient sample, what we can do is systematically compare that pattern to all of our observed patterns and calculate some sort of similarity metric. You might use something like a Pearson correlation coefficient and ask the question how similar is my observed pattern to my theoretical pattern and which theoretical pattern would best explain the observables if any? This allows us to detect multiple infections in the same sample as well. And so, you are going to have to, if you signed up for the exercise, you're going to have a chance to run that algorithm called E-Predicton the samples that you process.

**35. Q&A: Is Virochip ready for mass production as a diagnostic tool?** (51:38)

**[STUDENT:]** Economically, how viable is a large-scale production of the Virochip?

**[DR. DERISI:]** So, the question is, is this economically feasible? I mean, how expensive is this stuff? So, you know, you do a PCR reaction and do a sequencing reaction. The real materials cost of that, not including the power for the building and, you know, paying the person to do it and all that, the real materials cost for that is probably somewhere around $5 to $10 realistically. On the Virochip, our realistic cost is probably around $30 to $50. So, it's a bit more expensive, but it's well within the realm of many diagnostics that are given today. If you send out, you want to test for SARS by PCR in a clinic and send it out to a diagnostic testing lab, maybe $300 to $1,000for a single virus test. So, the economic

feasibility of doing these massively parallel assays is quite real, and in fact at UCSF we setup a new viral diagnostic and discovery center that is pushing to try to roll out these large parallel viral diagnostic assays in a clinical setting. There are lots and lots of technical issues with rolling out a diagnostic. And so, that we set up this center which is headed up by Charles Chiu to try and dope out a lot of those issues.

**36. Q&A: How would you analyze a mixed sample of two viruses on a Virochip?** (52:56)

I'll go here and then I'll get to the back.

**[STUDENT:]** This is just back off of his question, what if two viruses share the same family though? Like what if I have two viruses from the same family infecting the person?

**[DR. DERISI:]** Yeah. No, that's great. So, let's say a person is co-infected with two, I don't know, rhinoviruses let's say and they're in the same family. How would you know you have two viruses there if they're pretty close? The fact of the matter is you might not, that the predict signature on that might come up as well, it's kind of close to this and close to that, but it's a mixed signature. And so, you would actually have to pull sequence out and sequence it to definitively say whether there's two there. That's a tricky situation and can occur. Now, it turns out for something like dengue with one, two, three, and four, we also put specific elements on the chip that are designed to tell apart close relatives. So, on our chip we can easily tell apart dengue one, two, three, and four because we know about that and we've designed it to do that.

**37. Q&A: Does Virochip analyze both RNA and DNA viruses?** (53:57)

Okay, I'll go in the back behind the cameras because you probably never get called on because you're like hidden behind the camera.

**[STUDENT:]** Thank you. I was wondering does the Virochip, is it able to deal with viruses that use RNA or reverse transcribing viruses instead of just DNA?

**[DR. DERISI:]** Absolutely. So, let's imagine that we don't know what the kind of virus it is. So, we take a sample from a patient. It might be a DNA virus. It might be an RNA virus. Who's to say? So, in practice if we don't know it and we suspect that something like that might be going on, we can isolate the nucleic acid, the total nucleic acid from the sample, the RNA plus the DNA and we can actually separate them. We can process the DNA separately and the RNA separately. The RNA fraction we would first reverse transcribe using reverse transcriptase, those enzymes derived from retroviruses themselves, to convert to cDNA, and then randomly amplify that material up. Likewise we could process the DNA with just regular DNA polymerases. And so, that's practically how we deal with that situation.

**38. Q&A: Can you use Virochip to find viruses that naturally attack cancer?** (54:57)

Way in the back on the right here.

**[STUDENT:]** I'm aware that there are some viruses that they're using to try to cure certain tumors, and I was wondering if this technology would be able to identify any new viruses that could potentially do this.

**[DR. DERISI:]** Well, that's a cool idea. So, the question is there have been biotechnology efforts to engineer viruses to fight cancer. There was an adenovirus developed by Onyx Pharmaceuticals and

others that targeted cancers that lacked a particular gene, p53, and there is ongoing work on those viruses right now. It hasn't been super successful yet, but there's definitely a lot of interest in the area. And your question is are there naturally occurring viruses perhaps that could infect cancers and have some preferential activity against those cancer cells and not the rest of the host. I don't know of a virus that has that specific activity, but it's plausible that one could exist by virtue of the fact that cancer cells are just so different and so messed up. It's an interesting concept to consider. Thanks. Okay, we've got to wrap right now, but I'll still be right here to take more questions if you want to come on down.

**39. Closing remarks by HHMI Vice President Dr. Sean Carroll** (56:16)

**[DR. CARROLL:]** Thanks, Joe, for a great talk. I'm starting to imagine all the dinner table conversation tonight. And those of you who have these cards and viruses, you know, don't slip them into your younger siblings bed or something like that tonight and, you know, tell them oh, that's a dengue virus and that's a rotavirus, et cetera. But seriously, thanks for showing us how the most sophisticated technologies really come from understanding fundamental principles, and we certainly look forward to hearing more tomorrow. So, please join us tomorrow morning for two more presentations from Eva Harris and Joe DeRisi. We'll hear more about dengue fever and some more successful virus hunting stories.