

Viral Outbreak: The Science of Emerging Disease
Lecture 4 – Solving SARS and other Viral Mysteries
Joe Derisi, Ph.D.

1. Begin of Lecture 4 (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 fourth lecture is titled Solving SARS and Other Viral Mysteries. And now to introduce our program, the President of the Howard Hughes Medical Institute, Dr. Robert Tjian

2. Welcome by HHMI President Dr. Robert Tjian (01:07)

[DR. TJIAN:] Welcome back to this final presentation of this year's Holiday Lectures on Science. It's a great pleasure once again to introduce Joe DeRisi to give our fourth and last lecture in the series. Previously, Joe told us about how using bioengineering, computers, and molecular biology, he has been able to combine these tools for a potent approach to hunt for new viruses. In this lecture, Joe is going to show you how he can use his Virochip in real-time and in real life situations to discover and quickly diagnosis new viral outbreaks. Joe will also, I think, give us a glimpse of what the future in biotechnology holds towards the end of his talk. And now a brief video about Joe.

3. Profile of Dr. Joseph DeRisi (02:07)

[DR. DERISI:] Science as we know it now is a highly interdisciplinary endeavor. There used to be this concept that what was required was multidisciplinary involvement. You have different people who are experts in different things. You had the electrician, the machinist, the biologist, and the clinical M.D., and the software programmer, and they spoke their own languages and somehow they were able to get together and work on projects in a way that produced good science. Now, what's quickly been realized is that we can make faster progress, better progress, more accurate progress by learning each other's language, becoming simultaneously conversant in the multiple languages of engineering, computer science, and biology, and that individuals can do all aspects of the individual projects and not simply rely on experts who speak a totally different language and come at the problem from a completely different framework.

NATHAN: So this is our, these are dengue-negative samples.

[DR. DERISI:] So there are advantages to being a specialist in one area or another, but there are also advantages to being a generalist, that is being able to involve yourself in multiple different fields and innovate in different fields to bring it together. And this is especially true in biotechnology and biochemistry and virology and other kinds of sciences that require new technology development. Of course, what you hear in the headlines is the biological achievements that are made, a virus is discovered and so on. But behind all that is all the methodologies and technologies that are required to do that and those don't necessarily gain the front line headlines. But without the right tools, without the right technologies, looking into the unknown is all but impossible.

4. SARS: A global epidemic (03:51)

And we're back. I always wanted to say that. It's not as good as Mike Myers. Yesterday I had the chance to tell you about different kinds of molecular biology techniques and tools that are used to develop

diagnostics and tell if there is a virus in a particular sample and so on. And so today what I'll be able to do is take you through a few examples of how we've used this in actual projects. And I hope to, by the end of the lecture, tell you about some exciting new technologies that are coming down the road that will make sort of the things that we've been doing here today go even faster. So our first example takes us back to 2003. That's probably like when you were in what, elementary school, or something like that? It's kind of scary, but let's consider this real-world example. In early 2003, a mystery illness appeared right out of nowhere seemingly. And no one knew what it was. It was a respiratory disease, but no one really had a clue what was causing it, and it seemed pretty severe. And the sort of anxiety and panic about this outbreak, while it was amplified further by the media without a doubt, and some of the headlines that were coming out of these areas were pretty scary, it's called SARS for severe acute respiratory syndrome. And by the time this reached its peak anxiety and panic period, you couldn't go anywhere without hearing about SARS. And sales of face masks went through the roof.

5. Origin of SARS epidemic (05:29)

So let's go to the heart of this epidemic, or this outbreak, it actually started in mainland China, but our story really starts in Hong Kong, a special administrative region of China. And if we go Google-style and zoom in on this outbreak, Hong Kong is here on the map, zoom in a little further and our story starts actually in one building, a particular hotel in Hong Kong. This is it right here, this is the Metropole Hotel. You can't get a reservation there anymore because it's not called that anymore. And you'll find out why in a minute. And so this story is centered on the ninth floor of the Metropole Hotel, and next I am going to show you a floor diagram of the ninth floor. So this is the ninth floor of the hotel. The index case, the person who brought the infection into this hotel, that's the first person recognizable with the infection in an epidemic, stayed ironically in room 911. You can't make this stuff up, this is real. And now I'm going to show you on the floor plan where the other guests staying on that floor, who got sick. We already know the index case was sick. It was a physician traveling from mainland China. He checked into the hotel for a single night to go to a wedding, and then left the next day. He was ill with this respiratory disease, but didn't really know the consequences of it. And people in these rooms and these rooms all got sick. And the ones in orange, in particular, went on to transmit that disease to others after they left the hotel. So can someone tell me, you know this pattern is not exactly random on the floor plan here, can someone tell me why is it that there is this non-random distribution of hotel rooms around 911 that contracted the disease? Do you have any ideas?

[STUDENT:] The environment that they were breathing in?

[DR. DERISI:] The environment they were breathing in? So ventilation. Ventilation; that's a great idea, that's one of the first things that you would look for in a hotel situation like this. It turns out that the rooms are generally under negative pressure such that the air is forced into the hallway and then into the rooms. So if the index case is in the room and breathing out a lot of, let's just say if this was transmitted by aerosol -- it really hasn't shown to be -- then those particles would stay in his room. They wouldn't go out into the hallway. So that was probably not the cause. Maybe they had a big party. Is that one of your ideas? Cleaning staff. So perhaps the cleaning staff transmitted it around. Interestingly enough, none of the cleaning staff got sick at this hotel. Any other ideas? One more, one more.

[STUDENT:] The common use of like the elevator and the stairs?

[DR. DERISI:] Common use of some sort of pathway through the floor, for example, like the elevator or the stairs. I like that answer a lot. So it turns out that when this whole thing was said and done, this is caused by a virus -- if it wasn't, I wouldn't be telling you about it -- they went in and looked for the virus in the hotel to see what happened with this case. It turns out there was a massive hot zone of the virus, tons of virus, in the carpet directly in front of the room. So two months after the outbreak, they could

still find virus in the carpet. And so can you guess what happened here? Well, no one really knows. This patient died. We'll never know. But what likely happened is the person was sick, he either coughed, or vomited and threw up in front of the room, on the carpet, embarrassed, he doesn't want to tell anybody, cleans it up himself and is back in the room. And then people going to their room walk through it and get it on their shoes or whatever and go back to their rooms. It doesn't explain all the cases in here. Maybe the guy in 938 got lost going down the wrong way, went back down the other hallway and that was a fatal mistake. So we'll never know, but this is the best that we can do in a retrospective study in a case like this is figuring out what happened.

6. Case definition of SARS (09:58)

So SARS was basically just a case definition. It was anybody with a history of a high fever after a certain date, November 1, 2002, because before that SARS was totally unknown. A dry cough, breathing difficulty, and importantly, close contact with somebody else who had SARS. That was the case definition. What was scary about this? No diagnostic; no one knew what it was. Every diagnostic that was put on this thing was negative. The transmission mode was unknown. Was it aerosol? Was it something on surfaces? And the incubation period was unknown. How long can people have it and spread the virus before they show symptoms? No idea. And the fatality rate at the time was actually pretty high; 14% is very high, actually. And that actually hides some details. It was age-stratified, so patients that were 65 and older had case fatality rates of over 50%, and so that's an extremely deadly virus.

7. SARS spreads worldwide (10:55)

So now let's zoom out from the Metropole Hotel, there it is. We're back out into space now, and where did the people in that hotel room then go after they checked out of the hotel? It went all over the world, so they walked through the hot zone, they checked out, got on planes, and went home. And this was kind of a big deal, so the folks that went to Canada, one individual from the Metropole Hotel who went back to Canada, then infected over 100 people at a hospital. Directly linked to that case at the Metropole Hotel. And, in fact, just to give you an idea of how efficient it was at spreading in the hospital in Canada, the attack rate, that is the frequency that people who are exposed to disease actually get the disease. The attack rate for medical school students who entered in the room, even to visit an adjacent bed of the index case, was 100%. If you went in the room, you got SARS, which is pretty outrageous, unheard of. And so that's why SARS was scary. Now this thing was just starting to go ballistic when one of the index cases who traveled to Vietnam checked into a hospital there, and the government in Vietnam noticed something funny about this case and contacted the WHO in late February of 2003. They said we think we have a problem, which was exactly the right thing to do. They sent an infectious disease specialist, Dr. Carlos Urbani. He came to Vietnam, arrived in a couple of days, checked it out and said, yes, this is something really different, and reported back to the WHO. So by March 8th or so, there was an emergency meeting at the WHO, they said we've got to do something about this, this is serious, quarantine the place. Dr. Carlos Urbani called the shots, quarantined the place, all the health workers kept themselves inside and were not allowed to go home. On their own choice they didn't go home, which is a very brave thing to do. And that probably stopped the wildfire of SARS from spreading in Vietnam, unlike some of the other places that there were outbreaks. But he paid the ultimate price; within a few weeks he was dead of SARS. He's a true hero of infectious disease and did the right thing. It's a tragic story, but it's one of ultimate self-sacrifice. What ensued after that global emergency call was an amazing collaboration. Over 13 laboratories from around the world, ten different countries at the minimum came together with different researchers, technologies and so on, to collaborate on solving SARS because it was a global emergency at this point.

8. Virochip identifies SARS virus (13:35)

And I'm going to tell you about our small part in this. We're not the only part in this, we are just one component, and I'll tell you what we did. So, some of that sample from Hong Kong was sent to the CDC here in the United States. The CDC then distributed samples to various different locations, different labs in the United States, and we were one of them, so one of the samples came to UCSF. These samples were actually supernatants of a cell culture that had been inoculated with a piece of lung from a person who had died. We put that supernatant onto our virus chip, so it told you about the virus chip yesterday. You guys who did the bioinformatics exercise get to analyze some data and get some results. And if you remember from the bioinformatics exercises, we used that algorithm that we developed called E-Predict, which basically compares observed signatures to known signatures to come up with the most likely result. And this is what Dave Wang, a post-doc in the lab at the time, saw when they did our little E-Predict algorithm. He saw that the signature was best explained by a coronavirus. A coronavirus from either birds, cows, people, and an astrovirus, a virus in a totally different family mind you, from turkeys, from cows, more birds, and some human. Okay, that's a weird signature, it didn't explain. If it was just a human coronavirus, it should have just come up human coronavirus, but this was mixed. Now a little bit of investigation reveals that certain astroviruses and a branch of the coronavirus family, the ones that are actually in birds, share a region that's almost identical, so astroviruses and coronaviruses have an evolutionary-conserved region, and that's what our array was picking up. So this was a novel coronavirus.

9. Sequencing confirms SARS is caused by a new coronavirus (15:18)

But, I'm not a believer in this kind of stuff unless I see sequence. Sequence is the gold standard, like I told you. And so how are we going to get sequence? Well, we could do the design primers, order them, they come in a few days, and then you do... this is an emergency, we want it now, so how are we going to get sequence fast? So a graduate student in my lab, Anatoly Urisman, had this crazy idea. He said, well, you know we know that SARS is on the array, right? So we hybridize this thing, it's sticking to the spot in the array, that's that signal you see when you analyze those arrays. Couldn't we just take like a sharp needle and pick the spot off the array and sequence that material right there that was stuck on the end of the needle. Basically use the array to purify SARS nucleic acid. That's nuts, but what the heck, let's try it. So we got a neuron dissection rig from a neighboring lab in the middle of the night. We returned it; they never knew anything. A tungsten needle used to dissect neurons was used then under a microscope, a confocal microscope, to actually scrape away one single spot that SARS had hybridized to. This is a rather insane little operation, but lo and behold, what we were able to do is just randomly amplify that material that's stuck to that spot, we don't need any specific kind of primers, and just sequence a bunch of clones that came out of that. And sure enough, we confirmed a novel coronavirus. It was 65% identical to any known coronavirus, which means it's really pretty different. It was definitely new. And so we reported those results to the CDC and to the Global Response Team, and around the same time other methods like electron microscopy, other degenerate PCR techniques and so on, all came together with the same conclusion, that there is a novel coronavirus here and it is the most likely candidate for an etiological agent of the outbreak. And so what was amazing here is this all happened within about one month, the entire thing, from the time WHO found out to the time that this thing was essentially revealed to be a novel coronavirus. From the time we received the package to the time we got the result I showed you, was 24 hours, so we were up all night as you can imagine. It was a really intense and exciting time.

10. SARS epidemic stopped in 6 months (17:32)

Now what happened with the outbreak? So here is a histogram of the outbreak, the number of cases as a unit of time, just like some of the histograms that Eva showed, and here is where the novel coronavirus

was discovered. So you can see from the beginning of the outbreak from the time it was discovered to the time that effective quarantine procedures and diagnostics were deployed and control measures were put in place, an epidemic was brought to heel and prevented from spreading further. So what could have been really horrible was actually put down pretty quick. Now you can see there is a little dip there. Does anybody have any idea what that dip is? No, it's not *Carnaval*. I don't know what that dip is.

11. SARS coronavirus originally came from horseshoe bat (18:15)

Okay, I have a question to ask you. This was not a human virus initially. Most viruses that burn this quickly in a population are not already there, we don't know about this. Does anybody know what the reservoir of the virus is? That is, its natural host, where it's not causing so much devastation. Any ideas?

[STUDENT:] Birds?

[DR. DERISI:] Birds, that's a great suggestion because there are a lot of viruses harbored in birds, like influenza. But in this case, no. Any other ideas?

[STUDENT:] A cow, bovine?

[DR. DERISI:] A cow. No, but that was a good guess. I'll give the answer. It was the horseshoe bat from central China. I know, that was a hard one, huh? It was initially thought that it might have been a civet cat, a kind of cat, but actually upon further investigation it turned out that the horseshoe bat actually had the coronavirus, which is extremely close to the SARS virus, and only a few changes were required to then transmit it to humans. And I want to make a point here, that many viruses that come freshly into the human population are so-called zoonotic transfers. They are the product of a transmission event between a non-human species to a human species through mutation or other changes. And so that is why it's also important to study viruses in different kinds of animals.

12. Proventricular dilatation disease (PDD): Mysterious parrot disease (19:44)

And so the next story I'm going to tell you is an animal story, non-human that's all about birds, and we heard -- there was a question earlier -- that the bird might be the reservoir. This is why we study some birds. So this story is about psittacines, though. Psittacines are birds like cockatoos and cockatiels and macaws and parrots and things like this. Many of these birds are critically endangered. Some of them have 100 animals left in the entire species before they go extinct. And some of them are also very popular pets, they live for a long time. There is a very big domestic trade in these and so on. And for over 20 years there was a disease spreading through psittacines, over 50 species affected. This disease was called proventricular dilatation disease, or macaw wasting disease. It produced results like the one you see on the right. I'll get back to that bird in a little bit. It was 100% fatal, there were no diagnostics available, and for literally two decades people beat their head on a wall trying to figure out what this thing is. There were several ideas, but none of them turned out to be true. And so how does this disease manifest itself? Well, here's a typical example. A breeder, an enthusiast hobbyist who breeds parrots and macaws in their backyard. Pretty common actually. And what the breeder did -- this is actually a true story -- what the breeder did is he bought a new bird to breed with his other birds. He bought an African Gray Parrot, brought it into the house, didn't know the bird had PDD, because a lot of times birds with PDD shed the virus before they're symptomatic. So that's why this one is colored red here. And he put it in with his other birds, and what he attempts to do was breed it with the other birds. So he put it in a cage with another bird. Well, the new bird, the African Gray, stopped eating which, by the way, is a sign and symptom of PDD. PDD is a gut motility disorder, which is why they waste away. And what was his response? His response was, well, maybe he's not eating because he doesn't like the other bird, some sort of sociological pressure. So what did he do? He moved that bird to a series of different cages in the

house, trying to find a better partner for that bird. Maybe he'll eat with this one, maybe he'll eat with that one. And you guess what's happening here. So he's progressively exposing other birds in the house to this one sick bird. So the bird really wasn't eating, so the bird died. And soon after that, all the other birds that that bird had contact with stopped eating, and you can guess where this is going to go. So pretty soon he lost a pretty significant part of his flock to this disease.

13. Virochip indicates bornavirus may cause PDD (22:24)

Now we got on this case from notifications from veterinarians around the country and other places, and we established two specific collaborations, one with a veterinarian in Florida, Susan Clubb, and another in Israel, Ady Gancz, and we exchanged samples independently from two different locations. That's also very helpful to do when you're investigating a mysterious disease to make sure it's the same thing. These are little biopsies of bird crop, which is part of their digestive tract. We put the samples on the virus chip, not unlike SARS, and when I say we, this is really a post-doc in my lab, Amy Kistler, who is actually here today, put it on the chip, let it hybridize, and she used E-Predict and other tools to look at it. And one sample in particular really stood out. It stood out with this signature, just like you did yesterday if you were in the bioinformatic exercise, Borna disease virus was the first hit. The second hit was a human virus, respiratory syncytial virus, which is a very distant relative of bornavirus. They're both negative strand, non-segmented, single strand, RNA viruses. And so there is something interesting going on here. Because...why is it interesting? bornavirus isn't a bird virus, it's a horse virus. In fact, all known cases of Borna disease, which is caused by bornavirus, were in horses in a very narrow region in Europe. Outside of that, never seen it. It causes this disease, sad horse disease, or Borna disease, and it makes, well, it makes the horses sad. They're sad because they have all kinds of neurological symptoms and they don't eat very well, and sometimes they're called pipe smokers because they put hay in their mouth, but they never swallow it. And so this was interesting to us because possibly this could be a different bornavirus specific to birds, which would be a whole new branch of the family. And so Amy, pictured here, actually did a lot of molecular biology techniques like the kind I described yesterday, make some primers, try and do some PCR, sequence some stuff; it all came back pretty much garbage with the exception of a couple of little fragments that told us we were on the right track. We want the whole genome, I don't want just a little bit of it. And so after months of trying, we could not pull this thing out because it was so different, which is probably why a lot of other people had trouble identifying it.

14. Ultra-deep sequencing reads hundreds of millions of bases at once (24:37)

So we tried a different technology, it's a new technology, and I'll tell you more about it. It's called ultra-deep sequencing. This is taking this whole sequencing thing to a different level. Ultra-deep sequencing is a way of doing a lot of sequence all at once. Instead of sequencing one template or two templates, you literally can sequence millions of templates. I'll tell you more about what this means for biology in a little bit. But let me walk you through it. What happens is you take your nucleic acid, your DNA, RNA, what have you. You chop it up into little pieces. They have to be little pieces because this deep sequencing technology is specific for sequencing short stretches. Then add primers, adapters, so these are little synthetic pieces of DNA that you ligate or stick on the ends of the DNA, and that allows them to stick down on a solid substrate, like a chip, like a DNA chip, except in this case you just flow them across this piece of glass, a flow cell, and they randomly stick down all over the place. And you end up with these little pieces of DNA stuck on the glass. Now remember Sanger sequencing? Well, this is a variation on Sanger sequencing. So now you have a million of these little molecules stuck all over the glass, now in one variation of the technique they can be amplified in place to make a little colony. But let's just represent them as single molecules now for clarity. Bring in a sequencing primer, a synthetic piece of DNA that's going to hybridize to the primer you stuck on the other DNA sequence. You don't need to know what the sequence is you're sequencing. You don't need to make a specific primer here.

And then you extend with a polymerase. But instead of using a bunch of nucleotides and some low frequency of a dideoxy terminator to terminate the strand, all the nucleotides are terminators. So it just goes one nucleotide and then stops. And those nucleotides, those terminators, are labeled with fluorescent dye so you can tell what they are. So then you take a picture of the flow cell and you see a little colored dot, and that tells you what base got extended. Here's the magic. The next cycle, you can then cut off the fluorescent dye and reverse the terminator, so now it can be extended again. So reversible terminators. You extend with the second base, image it, look at the colors, cleave it off and repeat. Third base, image it and so on. So it goes: extension, image, cleave, extension, image, cleave, over and over again, and you build up the sequence. You can do about 100 nucleotides this way. And so in cycle one, two, three, and four, if you look at the spot as it changes colors, you can read the sequence. Look at the one on the left. It's orange to begin with, then it's green, then it's blue; that's G, T, C, A, A, C, and so on. Now how many spots do you think you can do? What does the real data look like? Does it look like these nice little discrete spots? Not so much. I'll show you what raw data looks like. So this is actual raw data from the machine from my lab. Now it looks hard to analyze and that's why we have a lot of computers to do it, and there's a lot of informatics behind finding these spots. But the other point to make here is its dense. This is less than 1% of the whole flow cell, and so instead of sequencing just a few templates at a time, we literally sequence somewhere up to around 300 million sequences simultaneously. That's kind of a big difference. And so you might ask, well, if this is so great, why don't you just sequence everything? Forget the chips and PCR and all that stuff; just sequence everything all the time with this technology. Well, so it takes five or six days for this machine to run and it can take up to around three months of computer time to analyze the data, and that's not really fast. And I'm not even going to tell you what it costs right now.

15. Assembling sequences to get complete viral genome (28:17)

Okay, so after you have all this data, what are you going to do with it? You're going to assemble it, you've got to put it together. Just because you have 100 nucleotide reads doesn't -- and a read being one little piece of sequence that are read from one spot -- it doesn't mean you know what it is. You have to be able to put it together, so there are actually some sophisticated algorithms and bioinformatics that search for where the reads overlap, and then combine them into what's called a single contig, or contiguous sequence. And so one of the games here is being able to put all this stuff together, and I can tell you we're not really good at it. Most of the algorithms out there, they're kind of bad. They're getting better and there is a lot of really smart people working on it, so this is a tight marriage of bioinformatics, computer science, algorithm development, and technology.

16. Comparing sequences to known bornavirus genomes (29:02)

So what we did is we actually took all those reads from the bornavirus sample and tried to map them to the known map of bornavirus and guess where these things might map. We ended up with a histogram like this, so a histogram showing you how many reads we have per location in the genome. And what it shows is that we have sparse coverage. There are some areas where we have a lot of reads, then there are gaps and other things, so we can't really assemble the whole genome because we don't even know what's in there. And one of the main reasons is when we have a particular sequence, we don't have really good methods of assigning that to a virus. We may not know a sequence is from a virus because it's not long enough or has enough similarity to the known viruses, which has serious implications for using sequencing for discovery. Amy did, it just takes sequences right off the sequencer, then we designed those PCR primers, overlap them, and recover the whole genome. So what was taking her months took her a matter of days after the sequence was in.

17. Testing that the candidate bornavirus causes PDD (30:01)

Now we've got the bornavirus. Remember Koch's postulates? Does it really prove that the bornavirus caused the disease just because I found it there? No way. There are lots of endogenous viruses and viruses that we get asymptotically, so while it's a good candidate, we don't know. Again, the microbe has to be present in every case, that's what association studies are for. You've got to be able to propagate in culture. That's not necessarily true in the modern age, but you want to know that the microbe causes the disease you say it causes. And so I'll cut to the chase here. With our collaborators, we were able to do an experimental challenge. We picked three experimental birds and two control birds. And for the experimental birds, they were injected with virus that was isolated from an infected bird. And then at the end of the day, all three birds that were injected developed the disease. Now we were blind to which birds were infected and which were not when we were analyzing the data, which is also important. The two control birds remained disease free. Furthermore, from the infected birds, Amy was able to recover the bornavirus in the newly-infected birds, which is part of Koch's postulates. And for this reason, we believe that it really is an etiological agent, and this has been backed up by studies from other labs now. This is what the experimental bird looked like after it was infected. And what I'm showing you here on the right, you know, we can obviously see the bird was wasting away and starving, but on the right shows the guts of the bird. And that stuff in there is undigested seed. This virus prevents food from moving through the gut, so the bird eats, but can't move the food and that's why they waste away. It is a neurodegenerative disease; it basically attacks nerve cells that pattern gut motility. And the immune system responds inappropriately and destroys those nerves, the food stops moving. Interestingly enough, there is a disease correlate in humans called idiopathic achalasia. Whether there's a bornavirus involved in that or not, we don't know, but you can bet we're going to look. So the challenge study, have we proved that the cause of the things, the association between symptoms and virus? Yes. No virus in healthy birds? Yes. We can isolate it from the sick bird, pass it to a new bird. We can re-isolate it, we can culture it, and we have association by other groups as well, so that's where that story is at. Amy is actually now culturing the virus as well.

18. Q&A: What challenges did you face in developing the Virochip? (32:14)

So I'll take questions there.

[STUDENT:] Besides dealing with unknown viruses, during your development of the ViroChip, did you encounter any other challenges?

[DR. DERISI:] In science you always encounter challenges. That's a loaded question. One of the challenges that we faced initially was, how do we know this thing works because if you just take samples, you know a sick guy off the street, you don't know if that's a virus or a bacteria or something like that. Where are your positive controls? Getting the right positive controls for the virus chip was always a problem. So actually we were able to collaborate with a respiratory physician at UCSF, Dr. Homer Boushey, who actually had done a controlled infection experiment. He took healthy volunteers and infected them with rhinovirus as part of an asthma study under very controlled, safe conditions approved by a human subjects review board and so on. And so this was the perfect positive control because we knew exactly how many viruses were going in, we knew what the count was, the viral load was in the patients as a function of time and so on, so teaming up with other clinical researchers allowed us to overcome some of these technical challenges.

19. Q&A: Where did the index patient get SARS originally? (33:29)

[STUDENT:] Where do you think that the original, the person who was in 911, where would he have gotten that disease if no one else had ever had it?

[**DR. DERISI:**] So he was not the first. As I mentioned, the disease actually started somewhere in mainland China. He was a physician and had been attending to patients with SARS in China. What hadn't happened initially was that it wasn't recognized early in mainland China as a new respiratory disease and reported before this big, super-spreading event that occurred at the Metropole Hotel.

20. Q&A: Has treatment been developed for SARS? (34:09)

Question way in the back.

[**STUDENT:**] Has a treatment been developed for SARS?

[**DR. DERISI:**] A treatment? You know, in fact there has been a lot of people working on things like different kinds of inhibitors and so on as therapeutic agents for SARS. Some of that interest has died away because SARS itself has died away, effective control measures and so on have sort of made the job of making specific therapeutics and vaccines for SARS less important. But initially, during the peak of the outbreak, a lot of people got involved, and there were a lot of promising candidates. I think if it was still a problem, we'd see a robust drug or vaccine development effort, if it was still an issue.

21. Q&A: What makes SARS so much more transmittable? (34:48)

[**STUDENT:**] What makes SARS so much more transmittable than other viruses?

[**DR. DERISI:**] You know, this is an interesting question. What makes the virus more transmittable, what makes it worse, why is it so bad? Oftentimes these zoonotic transfer events, that is when a virus comes from some other non-human species into a human species, it's not well adapted. That is, a virus that goes too fast or kills too many of its hosts. That's actually not a very great virus, right? I mean, it's not great for us, especially. But a virus that persists in the human population that's allowed to replicate and spread through many, many individuals are those viruses that actually do not kill their hosts right away, or take long, long times, have long incubation times, or maybe even go latent, like a herpes virus, which are some of the most successful viruses on the face of the planet. So SARS actually was kind of in those -- in a fitness context -- kind of a lame virus, but scary nonetheless. And these are features of some of these viruses that come from non-human species. If they are given enough chances to infect lots of humans, they can adapt and become less virulent and perhaps persist longer in the population. Okay, the questions are fun.

22. Finding novel viruses in a Nicaraguan cohort (36:04)

All right, so my final story is one that's really a pleasure to tell you because it's a direct collaboration with Eva Harris. We're going to take you right back to Nicaragua. Still not going to dance, and the reason why we're interested in collaborating with Eva and others that have cohorts in different areas of the country is it's well and fine to analyze respiratory secretions of individuals walking down the street in San Francisco, but let's face it. In different environments, we're going to see different viruses. People are exposed to different things, they have different hygienes, there are different community water and sanitary structures and, of course, there is socioeconomics, demographics, all that stuff affects everything. And so we teamed up with Eva because, as you heard before in her lecture, she has well-established cohorts with kids that get yearly health visits, serum samples, and they are being seen for a variety of different things, not just dengue, like influenza and other respiratory disease. And so we were interested specifically in respiratory disease samples that were not influenza, we got that covered, we don't need to do that. And so what Nathan Yozwiak did, who is one of the TAs here for the bioinformatics exercise, he went down to Nicaragua, helped develop some of those RT-PCR assays and other things, and what they did is started screening samples, respiratory samples, that is, patients that

had a history of a cough or sore throat and had a fever in the last four days, he started screening those for different viruses. And out of at least 718 samples that they did from one season, you know there is a bunch that were flu, there were some that were respiratory syncytial virus, there was some parainfluenza one, two, and three, and so on. But a huge portion of those were unknown. Now some of this is going to be bacteria, some of it is not going to be a virus at all, so we know that. So what Nathan did was he actually took 200 of those samples and threw at it not only the virus chip, but a whole bunch of other diagnostic assays as well. And out of this he was able to assign a fraction of those to further viruses that we know about, picornavirus, some rhinoviruses, but there was also a signature. When I say signature, a sequence because remember I only believe in sequence at the end of the day. So I was able to sequence all of the stuff that was positive in the virus chip and PCR, and a small number of these had a new sequence for an enterovirus. Now enteroviruses are interesting and I'll tell you why in a minute. He also tried to get the sequence and, like Amy, had a tough time because it was very divergent. So he brought out the big guns, did some ultra-deep sequencing. And these were five samples. Now these five samples were from patients when we looked back at the clinical history, all of them had a history of fever and sore throat, that's how they got in the study, two of the patients were actually vomiting, and one of them had to be hospitalized, they were severely ill.

23. Bioinformatic filtering of deep-sequencing data to find viral DNA (38:48)

So let me just take you through a little of the informatics. I'm not going to go through the algorithmic details, but just sort of show you what kind of a data challenge it is. So only one-eighth of the output of the sequencer was used for this particular experiment, and that yielded 20 million sequences. So Nathan had 20 million sequences to read from, and this is from a respiratory secretion, a nasal swab from a patient. After filtering out stuff that he didn't believe was real sequence, simple repeats and other things that he couldn't possibly analyze, he was left with 14 million sequences. After taking out the human sequence, because most of that -- I mean, you swab somebody's nose, a bunch of it better be human, but I guarantee you it won't all be -- he was left with 6 million sequences, so 6 million of that was not human. And then he said, okay, we know a bunch of bacteria; let's take out all of the bacteria that we know. Now there is a lot of ongoing sequencing of bacteria. We were able to do the best we could and that was to only remove about 2 million sequences, so we had 4 million sequences we don't think are human, we don't think are bacteria, at least in the bacteria that we know of in the database, which there are certainly plenty more. And now he went straight together with Peter Skewes-Cox, another TA from the bioinformatic exercise, they analyzed that and asked the question, hey, which of these do you think are viruses out of that 4 million? And only 186 could we think were viruses. This is a needle in the haystack problem, and of those, only 119 did we think actually belonged to the picornavirus family, which is what enterovirus is, a picornavirus.

24. Discovery of EV109, a novel enterovirus (40:20)

Now what are those other viruses? It turns out that humans actually have a lot of endogenous retroviruses and things like that, and those are counted as virus hits, but they are part of your genome. That's another story for another day. So our perceived coverage of where we thought these reads mapped to the genome is shown here, this is a genome, sort of the typical genome map of a picornavirus. And so we have very sparse coverage. What Nathan did was he designed a bunch of PCR primers, a little more wild and crazy with the PCR design than what I showed you for Amy, but nevertheless was able to recover the whole genome of the picornavirus. And after it was said and done and sequenced by Sanger, closing all these gaps, we actually had many more reads, about twice as many reads, of this virus in this sequence than we knew about. But we couldn't tell that. There was no bioinformatic method that we had at that time, and we are trying to develop more sophisticated ones, that was able to ascribe those sequences to this new enterovirus. That's an interesting informatics challenge. Are you getting the theme here that you need informatics to do biology? Biology is becoming an information science. If you're

going to go into biology and you want to do grad school there, I encourage you to take a programming class, a statistics class, and some computer science. For real. Don't give me that look. All right, this is where enterovirus 109, as we named it, maps onto the phylogeny. This is the genetic relatedness of these viruses. There are a bunch of different families here. The enterovirus Cs are shown on the right, and just for reference, this is where polio is, so it's in the same subfamily as polio is. Rhinovirus is also over there, the common cold, and many of these viruses have very different phenotypes. Some can give you meningitis, some can give you polio and paralysis. Some give you the sniffles. It is an incredibly diverse family. What was interesting about that arm of the family that EV 109 is, no one has figured out how to grow them. None of those guys on that branch have been grown to date. We don't know; does it use a different receptor, what kind of cells does it infect? We can't figure it out quite yet and we're going to work on that.

25. Full-sequence analysis shows EV109 combines 3 viruses (42:20)

This is a plot of nucleotide identity of our virus versus three other viruses in that picornavirus family. And what it shows is that this isn't 100% identical. It's like 65% identical meaning it's very different, not unlike SARS. There is not a lot of information in this plot. If you look a little more closely, though and ask, all right, let's ignore, let's look at little sub-regions of the genome, and find something very interesting. So taking a small window of the genome, maybe 400 bases and moving it across the genome and asking what is this most similar to in the database, we find that actually the 5' end of the virus is more similar to a whole other branch of the family, and the rest of the virus is what we thought it was, an enterovirus C. That is the 5' end of the virus where all the regulatory stuff, the stuff that makes the virus go, translation, replication and so on, belongs to a different part of the family. And this is how viruses diversify, one of the ways, is that they can recombine. If you have two viruses that infect the cell simultaneously, they can exchange genetic information, polymerases can switch templates and actually then create a new virus that will be endowed with different properties, perhaps totally different disease phenotypes and so on, and so this is an exciting and interesting area to know what kind of virus recombination is really happening in the wild, and now we have an example of it from real individuals from real people. And the follow-up here, this is where the work just begins. I haven't proven to you that this causes the disease in those kids, we don't know that yet. But now we have a lead. We can look because Eva collected serum from yearly health samples, we could develop a serological assay and figure out who gets exposed to this and when in their lives. Is it a childhood disease or do you accumulate it as you grow older, who gets it and where? And what kind of disease phenotype is it linked to? Do they all have vomiting, or is it that healthy individuals have it, too? That has to be all worked out, and, of course, the molecular biology.

26. Comparison of advances in DNA sequencing to advances in telescopes (44:16)

All right, so now what does the future hold? I told you about ultra-deep sequencing and how this is changing things. You probably don't understand, though, or realize what kind of change this is. Let's discuss change in a context. For this example, I'm going to use astronomy, or the use of telescopes, as an example to show you how big this change is. So do people know what an arc second is? When you use telescopes and stuff. An arc second is the minimum angle that's required to tell two objects apart, so it's a measure of how good your telescope is. If you can tell the objects apart it's better than one in which you can't tell them apart. The human eye is about 60 arc seconds. You can tell things apart when they're 60 arc seconds apart. Now this all changed in the 1600's when Galileo made a telescope. His telescope reduced that to three arc seconds, a 20-fold improvement right away, and so this was kind of a big deal and that's why he is really famous. Now much, much later, nearly 400 years later, it only got three-fold better, so a big telescope on the ground, Mount Palomar, about one arc second. You can tell I'm stealing this example from NASA. This is great. What do you think the Hubble telescope can do, because I mean that's not on the ground, it's up in space, it's pretty cool. It can do 0.05 arc seconds. So this is pretty

amazing. So there are a couple of points here. Number one, the Hubble telescope is the single greatest increase in optical resolution since Galileo built his telescope. Nothing really happened for 400 years. That's a pretty interesting fact in and of itself. It turned out to be a 1,200-fold improvement over 400 years; the Hubble is really cool. This talk isn't about Hubble, it's a talk about sequencing. So here's the amount of sequence you can generate for a buck. Now you have to pay more than one buck, obviously, but if you reduce it to the cost of a dollar, this is how many bases you can sequence as a function of time. Do you notice how the curve is going exponential? In just the last ten years, there has been a 50,000-fold improvement in how much sequence you can do for a dollar, which is unbelievable because nothing in technology changes by 50,000-fold in five years, not semi-conductors, not computers, storage, data systems, none of that stuff, so we're really on a different trajectory here. What does that mean? It means we can view everything as sequence because now we have the possibility of literally sequencing everything. And that's coming soon. In fact, it's already here in many ways. Have any of you heard of personal genomics, or genetic testing?

27. Sequencing your personal genome (46:50)

Have any of you done it? No hands raised. I can understand that. But I did it. All right, and so I'll show you some personal genetic testing. Here's what happens. Right now, you can go on the web and you can get a bunch of different locations in your genome tested for various different disease associations. Now many of these disease associations are important and interesting, but a lot of them don't have a big-fold increase or a hazard ratio or sort of a risk ratio associated with them. I'll just show you a few. So it looks like it has highlighted red here on the screen that there's something going on with atrial fibrillation, not so much on the other things. The personal genetic testing also told me my eyes are blue; good to know that, and that I don't flush when I drink alcohol, like I didn't know that either. But, no, in all seriousness, this is sort of the results that are reported today with the kind of genetic testing that we have, and it's only going to get better. This right now is only reporting that I have a 1.2-fold risk over the general population for getting atrial fibrillation for European men of northern descent. Here's one for obesity. They say that I have an increased ratio -- it's only about 1.1-fold -- for being obese by the time I'm 59. So right now the point is that our genetic testing, while the technology is developing very quickly, the data isn't quite there yet. We sort of just have a rough idea of what's going on and it's only going to get better. We're in the infancy of personal genetic testing and it's going to happen; the cat is out of the bag.

28. Sequencing all the microbes associated with humans (48:34)

Now I also want to make a point that all this stuff is just testing the human genome, but we are more than just the human genome. There are all kinds of things around us, there are a huge number of bacterial cells in our body, there are more bacteria in you than there are human cells, and there are viruses. We are only now learning that there is a whole virome, an associated set of viruses that inhabit you, not just endogenous retroviruses, but many other viruses that don't cause disease, and we don't actually know what they do. And so we want to look at these as sequence as well. You can imagine the day when you get sequenced and every bacteria and every virus you have in you gets sequenced, perhaps even as a function of time, and that is also going to have impact. And so what we now do with genetic testing and test 30,000 genes for different associations may expand to testing the whole human microbiome, which includes bacteria and viruses, and that space may be much, much larger than your genome. And the context of that conversation between the bacteria and the viruses and our cells and our immune systems and our lifestyles and the foods we eat and so on, that's where there is going to be interesting new discoveries, new ideas, new things coming along. And I hope that you guys get interested in this and come along for the fun. All right.

29. Q&A: Are there other sequencing technologies? (50:03)

Questions? Come on, don't be shy.

[STUDENT:] Even with ultra-deep sequencing, there is still a limited number of base sequences you can actually put into the machine. Are there any possible new technologies coming out that are commercially viable that can sequence longer base pair strands?

[DR. DERISI:] You bet. The sequencing technology I show is just one of many deep sequencing technologies. This one is limited to about 100 base pairs. And what they do to increase sequence coverage is just sequence more 100 base pair sequences. So, for example, my machine can do about 300 million. I'm upgrading it in a couple of months -- thanks, Tj -- to something that's quadruple that and so you can imagine something that quadruples every year. It's going to get deeper and deeper. There are other technologies that allow the monitoring of single polymerases to extend a template, so I can actually watch with the microscope a single polymerase putting nucleotides into a template in real time, not with dye terminators. And that technology has the capability of reading much longer reads, maybe perhaps 1,000 base pairs, 2,000 base pairs, 4,000 base pairs. They probably don't even know the limit yet, and that's technology is actually coming on line right now. The disadvantage is that you sequence less templates. They actually seek longer ones, but fewer total templates. But like I said, all these things double or triple every year in this time of change right now.

30. Q&A: Are we born colonized by bacteria and viruses? (51:32)

There's one back over here, and then I'll get back to the center.

[STUDENT:] You were talking about all the bacteria and non-endogenous viruses that live in people. When in your lifetime did those come in? Are you born with them?

[DR. DERISI:] No, actually. So, for example, we just take the GI tract, and this can be at the top of an entire other lecture, Holiday Lecture. When you're born, you're born sterile. That GI tract has no bacteria in it. And I actually have a colleague at Stanford who has actually collected diapers from babies as they were growing...that's a pleasant experiment. I have kids, so I know. And then analyzed how the gut becomes colonized and with which bacteria, and more importantly how the population structure of those bacteria change when you take an antibiotic, and whether the population comes back or how does it change. It's now believed that this population structure that's in your gut, for example, has an impact on your disease state and on your life. It's been associated with things like obesity, for example, so that's really interesting and a lot of that is going to come out soon.

31. Q&A: Can DNA chips be used to identify diverse HIV types? (52:40)

Okay, I had one in the center, way in the back.

[STUDENT:] Do you use the microchip to sequence HIV and find out how it's occurring in people, and is there a way to stop it?

[DR. DERISI:] So the DNA chip technology that I described is really for detection. There are DNA chip technologies that can re-sequence the virus or detect whether there are changes in the virus, so-called quasi species analysis. There are chips that exist like that and, in fact, there are a lot of people sequencing using deep sequencers, mostly, to analyze how a virus is mutating, like HIV. And this is important because as we develop drugs and deploy them into the population, viruses adapt, mutate, and change often in response to these drugs, and we want to have sort of predictive values to know where the changes are going to occur and so on. So this is actually an intense area of research right now.

32. Q&A: How does zoonosis occur? (53:37)

[STUDENT:] I was wondering, do people know why, or have a reason why, animal viruses becomes zoonotic. Like what mutations there are and why do those mutations occur?

[DR. DERISI:] Well, you know in many animals the cells that they have, the cells that we have from mice to people and so on, they're all based on a common evolutionary lineage. They share many components. And so for many viruses, the mechanisms by which they use to enter cells and so on are very similar, just maybe a few amino acid changes away. And so it's not inconceivable that it just takes a little bit of mutation and a chance interaction with an animal -- whatever that may be -- to result in a transfer. And the more contact with animals in those kinds of contexts, like for example, when you're exposed to the blood of an animal or something like this, the more chance that's likely to happen.

[STUDENT:] ...mutations like, I don't know, UV rays or whatever?

[DR. DERISI:] Right, so let's talk about mutation in RNA genomes, for example. So RNA viruses are some of the most mutable genomes that we know, and they, on purpose, have polymerases that are sloppy, that insert errors on purpose. Their machinery is built to mutate themselves. Just to give you an idea of that scale, you know, rhinovirus or something like that, 8,000 nucleotides approximately, and during an infection of an adult, every single position of the entire genome is mutated in every infection. There is one to two mutations made in the genome per round of replication, and there are ten to the tenth or greater viruses there; that's way bigger than the genome. So the virus is actually making every mutation and every position every time, looking for new opportunity. And that's part of the evolutionary strategy of those viruses. It's actually a really exciting field. And it's a wrap.

33. Closing remarks by HHMI President Dr. Robert Tjian (55:43)

[DR. TJIAN:] Well, Joe, that was a fantastic set of lectures. What an inspiring and, I would say, optimistic, encouraging close to this series. We certainly have tools that will at least allow us to hunt down these new viruses. The stories that you and Eva have told us really help us to connect what goes on in the laboratory to what all of us care about in the real world, and I think that's been a special feature of these lectures this year. Certainly many of us remember the SARS epidemic and the hysteria that was sweeping the world. I remember being in the airport feeling very naked without my mask, but it's fantastic to hear how quickly the timing of finding out what the source of the infection was. So let me extend my final thanks to both Joe and Eva for this incredible, lucid, certainly high-energy presentation. And now I'd like to take a little time to give you a little sort of advertisement because, of course, for next year we're going to have another set of exciting series of lectures. And the title, it's a very interesting title, it's called "Bones, Stones, and Genes: The Origin of Modern Humans." And this is going to be such an amazing topic, we actually couldn't just find us two good speakers, we had to find three great speakers for this topic, and they're going to be John Shea of Stony Brook University; Sarah Tishkoff of the University of Pennsylvania; and Tim White of the University of California in Berkeley. So, from all of us at HHMI, I extend to you a healthy, virus-free, happy holiday season. Thank you.