

## **Panel V: Drug Discovery**

**Moderator: Liu Keliang, Beijing Institute of Pharmaceutical & Toxicology**

Please come in and sit down. We have to move to the next session, Panel V: Drug Discovery.

First of all, I'm going to introduce myself. I'm Keliang Liu from Beijing Institute of Pharmaceutical & Toxicology. Basically my background is pharmaceutical . . . . In this panel, we will see the new ideas and the new techniques in the drug discovery for infectious diseases. In this panel, we have four speakers. Our first speaker is Dr. Bruce Taillon from 454 Life Sciences.

**Bruce Taillon, 454 Life Sciences**

***454 Life Sciences Next Generation Sequencing Technology: Applications to Infectious Disease Research***

Thank you very much and, like the other speakers, I'd like to thank the organizers for putting this conference together. It has been very exciting. I've learned a lot and thanks to the two groups from National Academy and MOST for organizing it and getting us here and making it a very comfortable place to be to learn so much about global approaches to infectious disease research.

I also have the distinct pleasure, it would appear, to represent the entire United States biotechnology field. So, I will try to do my best to do that, just as Eric represents the entire U.S. not-for-profit research organizations. I will try to dutifully answer why we do things for profit in a scientific environment where knowledge should be the end-all. I am a scientist. I just happen to play a businessman on Monday to Friday.

Let me tell you a little bit about 454 Life Science, first of all, to head off any questions about what the name means. We don't know. We were founded by an entrepreneur and he is not going to tell us what the name means until we are successfully a public company. So, hopefully we will know that in the not too distant future. So, we will put that question aside.

What we are dedicated to doing – our vision as a corporation – is we want to make routine human sequencing a reality. That is a very grand vision. To get there, we have to take incremental steps and I'll be telling you about the incremental steps we've made to get toward our vision. The other part of our corporation resides around our mission statement. Our mission statement is to democratize sequencing. So, we want to make it so that one doesn't have to, if you don't want to, go to an organization like TIGR or JGI or Brode Institute to get your sequencing done. We want every laboratory or every university to have the capacity to do high throughput sequencing. Again, I'm going to give you some background as to how that's possible.

So, 454 is not as it was a few years ago. We were founded in 2000 – we're only six years old. A few years ago we were a pipe dream, to use an American term. But, now we are actually proven and scaleable technology. So, this is the inch men that underlies the technology that we bring forth. This is referred to as the genome sequencer 20 or the GS20 – that is how I'll abbreviate it.

Just to set the baseline as to what this technology can do, in a single 4-5 hour run, it will generate 20 million bases of data. These are hundred-based pair reads and at least 200,000 of those reads. So, basically we are looking at a system that has 10-fold the output of conventional sequencers.

So, we've developed the system and it is commercially available to go from library preparation through de novo assembly – so both the reagent stream, the instrumentation, as well as the software, all of which is available to the customer.

In 2005, we placed 20 instruments in the field. We didn't give them away – we sold them. So, people have purchased the instruments. We have repeat purchases. Most of them are in the genome centers. However, we have also penetrated the academic markets.

We have many peer review publications, some from our own work, but more excitingly, we have peer review publications from those folks who own an instrument and are using it day-to-day.

In order to take the technology to the next level commercially, we got into a distribution agreement with Roche Applied Sciences so we could penetrate the worldwide market quickly. They've done a great job in making the instrument available worldwide.

Finally under the category of this is a real product and generates real data, the NCBI (the National Center for Biotechnology Information) is now accepting the sequence file format that the 454 instrument, the GS20, generates. So, that means that any data that is generated by those who own the instrument can be uploaded into NCBI and be available for manipulation just as those runs that come off of the conventional sequencers.

So, just to underscore the point that we are a proven technology – this is not something that is “fly-by-night”. We published our seminal application in nature last year. This is a publication from our group where we demonstrate the utility of the product to sequence bacterial genomes. Last fall, a group from Penn State University published this first large amount of sequence from the wholly mammoth. Not only did they generate sequence from the wholly mammoth, but it also turned out to be a metagenomic application because this ancient DNA contained bacteria and viral as well as plant DNA. So, they were able to get this publication in science and essentially it was a single run off of the GS20. They generated over three billion bases of DNA which was the largest contribution of ancient DNA ever.

A group from Johnson & Johnson did a collaboration with 454 and in January of 2005, they published the use of the technology to do drug development. This is some of the material that I'll present to stay on the drug development path.

Then most recently – this came out now – I actually put this bit in the slide on Saturday – where the 454 system again, in collaboration with the sequencing center at 454, was used to do sequencing of deep mine microbial ecology. So, the technology has applications across many different areas. It can be drug development all the way to ecology, which spans the life science arena, if you were.

We also have applications in place to not only look at drug development from an infectious disease standpoint, but other applications such as oncology because fundamentally the technology will allow for sequence analysis no matter what the source of the DNA is. Once one can get it to DNA, we can sequence it.

I'm going to give a few slides that talk about the technology so that we are all on the same framework before we go into data. So, the core of the technology is this – what we call a peakatiter plate. A peakatiter plate is a fiber optic bundle so there are a million or so fiber optics bundled together. What we are seeing here is individual fibers which have been chemically etched to form a well which is approximately 40-50 microns in width and about the same in depth. So, we selectively remove that core material so what we are left with is very small reaction vessels – mini test tubes if you like. We do all of our sequencing chemistry within those wells. The power of having it be a fiber optic bundle is that then the sequence which is generated transmits through that fiber optic bundle and we place this fiber optic plate directly against the CCD camera which is an astronomy grade CCD camera, and I'll describe why that is important.

So, just to give you the very background – I'll be glad to go into detail, but time limits me to tell you too much detail how the chemistry works. We begin with a DNA library of any source. We've done everything from mammalian genomes down to micro RNA samples. The idea is to take a DNA library and shear it down into a size that is acceptable for the chemistry. We like somewhere between 200-500 base pairs, except for micro RNA – we can do them directly. What we will do is basic molecular biology chemistry. We have tried to keep this technology as simple as possible so that any group can use it. So, we do a basic linker ligation. Again, I would be glad to go into details about this if you like. But, the end product is a series of single-stranded

DNA molecules in which we have adapted on their primers sequences that we know in A sequence on one end and a B sequence on the other end. So, now we have single-stranded DNA molecules where the ends are defined and specific.

What we then do is we have invented and improved a process for clonal amplification of the single-stranded DNA molecules within an emulsion. So, the output then are going to be many beads in which all the DNA that decorates those beads arise from a single DNA strand. I am going to have one slide where I can describe this in a little more detail.

We then take this bead and a collection of millions of these beads, and load them onto the peakatiter plate, as well as loading in beads that contain reactions. Now, the sequencing signal is generated through a couple biochemical reactions. Fundamental molecular biology says that if you are able to extend a sequence, if . . . is able to add a sequence or nucleotide to a growing strand of a sequence, one of the reaction products of that is pyrophosphate. So, pyrophosphate is a very nice reagent for the enzyme sulferialase in the presence of APS. The reaction there is a generation of ATP. Well, ATP is a brilliant substrate for the enzyme luciferase in the presence of luciferon. The action of luciferase on ATP and luciferon is the generation of light. It generates light in a linear fashion depending on the presence of ATP. So, we have a couple biochemical reaction here, which is linear to the amount of nucleotides that are added.

So, let's talk about a key and fundamental part of this technology is the fact that we are able to clonally separate the complex genome and get them associated with a bead so that we can get that product into one of these micro. . . . A primer and a B primer. So, we know the ends of these.

We bring in beads in which there is the complement of one of these primers. So, for the sake of argument, let's say it is the complement of A. We mix the DNA in limited dilution with the beads so that there is many more beads than there is DNA. We add to that a proprietary mix of emulsion oils and all of the reagents one needs to do PCR, and we mix these up in a very special way – we shake them. What we end up generating is a water and oil emulsion. In that emulsion we have two different types of mycells. We have large mycells which form around the beads,

and then we have very small mycells which essentially act as a buffer between the large mycells. And, because we placed the DNA fragment in limited dilution, some of these micro reactors that contain beads also contain a single strand of DNA. Based on Poisson distribution, most of these reactions contain no DNA. Those that do contain DNA, the vast majority of them contain a single molecule and then some will contain 2-3-4, etc. in a distribution going to the asymptote.

Then what we do is we simply perform PCR – PCR as done anywhere else. We put it into a PCR machine in a tube or micro-titer plate and what happens is the single strand becomes many strands and they begin to anneal to the bead so that now you have the DNA molecule associated with the bead. We then break this emulsion and isolate the DNA containing beads.

How do we isolate those beads that contain DNA away from all of those empty beads? Well, that is what the other primer is for. That other primer, when we do the PCR, we used biotin in that process. And, because all of these ends of these DNAs are decorated with biotin, it is simple molecular biology now to purify using streptavidin magnetic beads – we can purify those beads that contain DNA away from all of those other DNAs. So, now we have beads which are decorated with DNA – tens of millions of copies of DNA that all derive from a single single-stranded DNA molecule. So, we have clonally amplified a single strand of DNA and placed it on a bead, and now we can drop that bead into one of those micro chambers and sequence it.

So, what we have done is what the entire human genome project did by cutting DNA, cloning it, and preserving those clones in bacteria. We have completely eliminated that whole expense and labor of doing that by generating clonal libraries in a single reaction. So, any genome of any size we can make one library and then sequence it as much as we want to. That is a key fundamental part of the technology along with the other key fundamental parts that I've described – we only make one library. We don't make many libraries. We only make one library to sequence a genome. Then we just continue as much depth as we want to go into that genome and we will continue to sequence it.

So, then it is simply depositing this bead mix onto the plate and this is what it looks like before deposition, and this is what it looks like afterwards. These small beads actually not only contain the enzymes for the chemistry, but also serve to kind of lock the whole process into these wells.

Then, as I described, we basically anneal a sequencing primer onto the B strand of the DNA and sequence in this direction, as nucleotides are added, we generate light. Here is a snapshot of the light beam emitted. So, this plate is right up against an astronomy grade CCD camera which is designed to grab low levels of light emission and so we're not even looking at low levels of light emission. We are looking at fairly substantial levels of light emission. So, we are well within the dynamic range of the camera to do very sensitive analyses. So, now all it is is a matter of software to interpret the reaction.

So, we of course have to have a name for everything. So, when we describe our sequencing reactions, we call them flow grams. So, this is an example of a flow gram here. It is important to realize that we are looking at a digital output here. So, we are either getting zeroes or ones across the sequence – 0's, 1's, 2's, 3's 4', 8, 12 – depending on the DNA. So, basically our base calling counts photons being emitted from the individual wells and then we use software to look at those signals and adjust the threshold and make a base call.

So, what we have done is designed in these known sequences that I talked about, we also put in a key sequence – sequence of T, C, A, and G. A single nucleotide – when those are sequenced they give us a way to measure one quality of the sequence, two to set the threshold for what a single nucleotide extension looks like, and also to make sure that the nucleotides are flowing in a proper order. All of these things are important.

The way we carry out the sequencing is we flow a single nucleotide at a time. So, first we will flow a T, and then we will flow an A, then the C, then the G. So, when we are looking at the key sequence, we get signal for the first flow, no signal for the second, we get signal for the third, not for the fourth or fifth, for the sixth, not for the seventh, but for the eighth. Now we know we have had the proper flow, we have had the proper sequence, and we know how to adjust the signal.

We are not doing this for one of the hundreds of thousands of sequences we are generating. We do this in every single well because this is part and parcel to the DNA that is being sequenced. So, every single well gets quality controlled as opposed to doing the quality control on one well of a plate or for one round of sequencing or one day of sequencing. We are quality controlling every single sequencing well. So, our typical standard operating procedure is to cycle 42 cycles and we end up typically getting 100 bases of data off of that.

Let's talk about the utility of this. We have sequenced many different bacteria and we get overall accuracy in 99.99x% accuracy. Typically it only takes anywhere from .5 to three or four runs to sequence a bacteria. The only areas where we have problems with . . . taking an entire sequence out is where there are repetitive regions, the bane of all sequencing, and so somebody like strep-pneumo which has a lot of repetitive sequence, can be problematic.

So, we have done a variety of projects looking at microorganisms. We have taken nine million base pair evolved genome and sequenced it 20x coverage, and found 15 mutations within that evolved genome. They were all confirmed by Sanger sequencing – the technology works. And, we didn't miss any of the sequences.

So, let's talk about it in context of drug development. Johnson & Johnson had a novel anti-micro bacterial product that they were wanting to push forward. One of the problems was they didn't understand the mechanism of action. They knew it did not inhibit DNA gyrase, which is a typical target, so they set out and generated resistance mutants in not only M-tuberculosis but also M-smagmotis, and they sent us these strains – one resistant tuberculosis and two resistant smagmotis and a parent smagmotis and we sequence them to near completion. So, this is what the flow chart looked like. They did the biology. They confirmed our results by genetic complementation and we basically cranked out the sequence. This is a project that we can get done in a couple of weeks in our sequencing center, easily, because we do 20 million base pairs per run, and these genomes are only 4-6 million base pairs. So, we get 3-4 full coverage in a single five-hour run.

So, here is a take-home. This is work that was published in science and we found one mutation in the tuberculosis strain and in the smagmotis strain, and all of these resistance mutations fell within the membrane spanning region of the HPE gene and none of them were found in the DNA gyrase gene. So, these were the only mutations that were consistent across all of the strain sequences.

So, with this information we sequenced the genome. We identified resistance-confirming point mutations which pointed to the mechanism of action in the target for this novel drug. That turns out to be the ATPE gene. So, the findings then indicate that this novel anti-micro bacterial drug inhibits the proton pump of the ATPE synthase and J&J is now pushing that information forward to develop this compound. So, they will be going through their pre-clinical and clinical trials on it.

All of these findings were then confirmed by classic genetic complementation tests.

Just to hit on one more approach that we used in the drug development space is what we call ultra-deep sequencing of PCR amplicons. One of the powers of generating 20 million base pairs of sequences is that you get a lot of sequence. You can either consider a lot of sequence a bane or a bonus. So, we developed a PCR amplicon sequencing protocol which we think gains a huge amount of benefit. It is a relatively straightforward procedure. We do PCR amplification and put our A and B primers on the ends of the primers used for the PCR. Straightforward PCR amplification. Then we just put it into a standard 454 process. There are a variety of applications we have done to identify genetic mutations in tumor biopsies. I'll talk about some viral variance we identified. You can do population studies and bacterial strain identifications. All of these applications have been done by clients of 454.

So, we worked with Dr. Gazel at Yale to do an analysis of a HIV genome. We looked at these eight amplicons. Multiplexing is only limited by your imagination. Really, it is limited by the accuracy of quantification if you are trying to do any statistical power and the desired depth of coverage, depending on how deep you want to go.

So, what we found in one clinical sample was two mutations – one at 18% and one at 35%. This is the corresponding Sanger trace from the collaborator. You can see one of these was barely detectable, really not detectable, and the other did creep above where you might say there was a mutation there.

Now, one of the important things is that the digital mutation detection – this is a software aggregation of all the sequence we generated. This is actually the signal that was generated from this experiment. We can go back and look at the individual sequences that contributed to this aggregate analysis here.

To give you an example of that, here is a slightly more complex sample of the protease gene. You can see that it is unresolvable in the Sanger sequence trace, and it looks unresolvable in the 454 amplicon analysis software. However, it really isn't because when we dig a little deeper we see that these two mutations are actually from a single haplotype in 30% or 34% of the strains. This particular mutation is from this sequence in which you have a deletion and an insertion, or an AG to GC transition. The rest is from this event here, in which you have an insertion or deletion or some other sequence event going on.

When we look at it from the flow gram space here, as we like to say, they actually – I'm sorry – I thought I had gotten rid of all of these. So, this is what the individual peaks from the data look like. I think you could have seen that the sequence in the flow space, because it is a completely different look to the chiral flow grams that are generated by the 454 sequencing instrument, when you have any kind of change in the sequence, whether it is a . . . substitution, deletion or insertion, the actual sequencing spacer flow, as we call it, looks completely different.

So, what you can see here is, for example, this is wild type sequence here. This peak is 3-A's, and in this particular variance, it is just a single A. So, you can see the flow, the original data looks very different which allows us to fine-tune our software analysis of the data. There is a certain amount of power here that is not available with other . . . . .

So, I'll sum up. This technology allows us to do projects that had heretofore not been available because in a single four-hour run, we can do projects that we could not do in the past. Then we take this to the clinical level. Here is a sample from a patient who was beginning to see an increase in viral load after being on the anti-retroviral treatment for awhile. When we sequenced the HIV genome from this patient, we see the presence of this mutation at about 3.45%. Being good molecular biologists, of course we want to see it in both forward and reverse reads, and you see the mutation is very consistent.

So, we are looking at a technology that allows for generation of a lot of data – a huge amount of data – that that really wasn't possible before for a single individual. So, one person can generate 20 million base pairs of data in a five hour run. In our labs, we have basically four individuals who are running five instruments and they are doing two runs a day. So, we generate on average 200 million bases of sequence information in a day and we are not even trying hard yet. That is a minimum. Frequently we get 30-40 million bases in a given run.

What we have seen is the accuracy of the data is extremely high when one looks at consensus accuracy. When we're looking at getting 15-fold over sampling, we have very high consistent accuracy.

We basically took a flyer – we had one person sequence the . . . . . They were able to do one person, one machine, one week, and they were able to generate that whole genome over again. That information was presented at the Marco Island meetings a couple of months ago.

It is a proven technology so Roche has signed on to sell it. They are really happy. A variety of folks have purchased the instrument and continue to buy reagents. That is our razor and razorblade model. And, we have also won some pretty good awards – Popular Science said that we are the best of what is new in personal health in 2005. Possibly more relevant is The Wall Street Journal gave us a gold medal award in their technology innovations for 2005. Not only did they give us a gold medal for new technologies, but overall global gold medal award for new innovations.

So, in terms of drug development and other applications in infectious disease, the sequencing of bacteria or viruses fits the sweet spot of this technology very nicely. We can multiplex. We can divide the peakatiter plate up into different formats in order to generate multiple samples of sequence. So, the technology is now ready for helping in the global crisis of infectious disease. We hope that 454 and our partner, Roche, can play a role in that through both our scientific expertise and our instrumentation.

So, I'll stop there.

Moderator – Thank you, Dr. Taillon, for your wonderful talk. The next speaker will be Junzhi Wang, professor and deputy director of the National Institute for the Control of Pharmaceutical and Bio Products.

**Wang Junzhi, National Institute for the Control of Pharmaceutical and Bio Products**

*Current Situation of Medicine used for Infectious Disease in China*

Hello everybody. I would like to thank the organizer and the chairman. Today I have the opportunity to give you a presentation on the current situation of medicine used for infectious disease in China. I'm from the National Institute for the Control of Pharmaceutical and Biological Products, simply called NICPB. Our Institute . . . national laboratory responsible for the quality control of our drugs, including . . . . medical devices . . . .

**(not transcribed)**

**(Tape 11)**

Moderator – I have to remind the last two speakers to keep your presentation time within 20 minutes because we are already late. The next speaker is Professor Paul Ahlquist from the University of Wisconsin-Madison.

**Paul Ahlquist, University of Wisconsin-Madison**

***Functional Genomics of Virus-Host Interactions***

Thank you very much. I'd like to talk to you today about some functional genomics studies of virus-host interactions with a view to using those as potential targets for interfering with virus replication. I hope to show you some examples of the diversity of interactions with the host that viruses depend on, both in terms of their own replication and also in causing disease.

I'd like to tell you three little stories in the course of this. First off, by way of background, I'd like to tell you about some of our studies on the nature, the structure and assembly and function of positive strand RNA replication complexes and new parallels that we found between those and retroviruses and double stranded RNA viruses. I'll follow that up with some discussion using unique yeast virus systems to do genome-wide cataloguing of hosting that are involved in the replication of such positive strand viruses.

Finally, at the end, I will talk a little bit about Epstein-Barr virus and nasopharyngeal carcinoma.

Deleted: nasal fringal

Viruses are exceedingly diverse, but as you all know, they can be usefully divided into six major classes based on the kind of nucleic acid in the infectious virion particle. So, there are then three classes of viruses that replicate their genomes through RNA intermediates, two classes of DNA viruses, and then the reverse transcribing viruses that bounce their genomes back and forth between the two of these.

Deleted: So, v

Deleted: virian

Deleted: de

Deleted: our

Of these first, I'd like to talk a little bit about the positive strand RNA viruses. This is the largest of the six classes of viruses and encompasses over a third of the known virus genera and many important pathogens including the SARS virus, hepatitis C, West Nile, paralytic polio virus, most of the common cold viruses, and important veterinary viruses as well.

As one of the major targets of our studies of these viruses, we are working on understanding the molecular mechanisms of genome replication. That is chosen, in part, because this is a process that encompasses for all of these viruses more than half of the coding capacity of the virus. It is also the process that has the most highly conserved genes among these positive strand RNA viruses, so it has the best hope of being a target for a potentially broader spectrum antiviral action.

Deleted: coating

One of the difficulties in studying RNA replication in these systems is that unlike the other RNA viruses and the retroviruses, these positive strand viruses don't package their RNA polymerase and the other replication machinery in their infectious virion, so you don't get it in an easily purified form. Instead, these viruses just form strictly intra-cellular RNA replication complexes which turn out, in all cases studied so far, to be membrane associated – a point I'll return to several times.

Deleted: in

Deleted: virian

Deleted: out of it. So, i

Now, in order to understand the replication of these viruses and the broad principles therein, we've turned to several advanced model systems. One of those is the alpha virus-like agent, the Type Bromo-Virus (BMV). Some years ago we showed that BMV encodes two proteins involved in RNA replication – 1A and 2A. 2A is the viral polymerase; 1A has an NTPase/helicase domain that interacts with the polymerase and an N-proximal domain that has functions involved in capping the viral RNA's. All three of these domains are conserved throughout the large alpha virus super family of positive strand viruses.

Deleted: and s

Deleted: NTPA's

Deleted: end

Now, on the RNA side, we've mapped a number of cis-acting signals that are important to replication. These are two. Interestingly, these have a number of tRNA-like properties that I'd be happy to discuss subsequently.

Deleted: sys

This template recruitment element here initiates the replication process in the recognition of the RNA. It is acted on by the 1A protein to stop translation of the viral RNA and to physically recruit the RNA into the replication complex. Once the RNA is so recruited, then the polymerase can then act on the origin of replication at the three prime and to begin synthesizing minus strand which is, in turn, used as the template to produce large amounts of progeny plus strand.

Deleted: ...

Deleted: ...

So, we have the beginnings here of an understanding of the replication complex, but I have to stress that the actual operation and stoichiometry of the replication complex is much different than the simple diagram here. One of the important first steps in understanding that is to see the context of replication which turns out to be in virus-induced membrane invaginations approximately 60 nanometers in diameter. We see such structures formed specifically on par-nuclear ER membranes by BMV and another model that we have worked with – the best studied noda-virus FHV, does a very similar process on outer-mitochondrial membranes, making these invaginations of the outer membrane. Our immuno gold electron microscopy show that the interior of these invaginations are the sites of the RNA replication factors and of virus-specific RNA synthesis which we can label with bromoUTP.

Deleted: In both cases, o

Deleted: amino

Deleted: bromo

We are studying how these complexes assemble and our understanding is that most advanced at present for BMV. What we find is this helicase-like 1A protein that I mentioned before has not only all of the enzymatic properties that I mentioned, but also has a structural role. It localizes to the endoplasmic verticulum membrane, self-interacts, and hundreds of these 1A proteins cooperate together to induce the formation of these membrane compartments. 1A is the only protein you need in cells to form those structures.

Deleted: , and w

Deleted: helicais

The template recruitment element that I mentioned before on the viral RNA turns out to act essentially as a packing signal to direct the RNA into the interior of those compartments. It turns out the polymerase is not necessary for that process, but if you express the polymerase, then it interacts with the 1A protein and a few copies localized to these vesicle structures where they then initiate minus strand synthesis and then positive strand RNA synthesis. The progeny plus

strands are exported to the cytoplasm for translation and encapsidation. All of this is supported by publications that we've previously presented.

Deleted: site of ...

So, the virus generates, as it were, a mini-organelle for RNA replication that concentrates and sequesters the replication factors and templates, and organizes the replications processes and protects the potentially double-stranded intermediates of replication from host defenses.

This is a very useful understanding for us to get a better sense of what is going on in RNA replication. It is also quite interesting to us because upon seeing this, one immediately sees parallels with other systems, and many of you have probably seen the parallels with retroviruses already. What 1A is doing is essentially what the major capsid protein of retroviruses gag does – gag collects on at target membrane (usually the plasma membrane) and induces the formation of a viral capsid and its membrane envelopment. Similarly, like the template recruitment element, a packaging signal on retroviral RNA leads to a gag-dependent packaging of that RNA. Like the 2A polymerase, the retrovirus reverse transcriptase is not required for that packaging, but a few copies get directed into these particles to later direct CDNA synthesis which, for the foamy retroviruses, happens before the particle buds.

Deleted: , s

Deleted: is

Deleted: that

Deleted: d

Deleted: ... virus, this

There are also similarities to the double stranded RNA viruses which also make a multi-subunit core in which they capture a messenger RNA replication intermediate to synthesis progeny RNAs and export to the site of plasma.

So, the implications are that three of these distinct six classes of viruses then share substantial functional links in their genome replication processes. The similarities are extensive enough that we think one has to at least consider a possible common evolutionary origin. The major differences in these systems are that these structures then exit the cell in the virion for the double stranded RNA viruses and retroviruses whereas for the positive strand viruses, these stay in the cell as a little intra-cellular RNA factory that just keeps spewing out RNA into the cytoplasm.

Now, I'd like to turn on to the issue of host genes in the replication of the same viruses. As you will appreciate viruses are exceedingly gene poor and in consequence of that, most if not all

steps in virus replication depend on the interaction of virus factors with host factors which, to a large extent, are still undetermined. These host factors are then crucial determinants of host range in pathology, but for this session, they are also important avenues for virus control.

Now, as one approach to understand and to define such interactions, we deliberately set out some time ago to transfer the replication of some useful virus models into yeast. We have done that for the two model viruses that I mentioned, BMV and FHV. These are systems that been very useful and presently are nearly unique in terms of their ability to duplicate the natural replication pathways and assembly pathways of these viruses in attractable genetic model.

For genetic purposes, we can link useful colony level phenotypes to the RNA replication of these viruses and we typically do that by replacing viral genes, often the capsid gene, with suitable markers. When we do that, the expression of these marker genes becomes dependent on the assembly and the function of these replication complexes that I've mentioned previously.

We have used those tools quite successfully with classical yeast genetics, but also with the increasingly powerful functional genomics tools for working with yeast. One of these is the ordered yeast deletion strain array, which is a collection of 4,800 viable deletion strains, each of which has a pre-selected open reading frame replaced by a kanamycin resistance cassette. This allows you to test essentially 80% of all yeast genes. So, we put our model viruses into these yeast strains and then can basically just walk down the genome one gene at a time and see what happens with the loss of each gene.

Deleted: ...

In the case of BMV, for example, even under very highly permissive conditions, I would say really excessively permissive conditions for replication, one finds that 58 deletions inhibited virus replication to or above an arbitrary three-fold threshold. We also found about 39 deletions that actually stimulated virus replication. So, these 58 genes are genes that facilitate virus replication. The 39 genes are genes in the wild type cell have an antiviral effect which, of course, is also interesting.

Deleted: a tou

We have also collected quite a few genes, including a significant number of essential genes, that are important for virus replication through classical yeast genetics and consequently with those together with this collection of strains, we have identified well over 100 genes that are important in virus replication.

As a single laboratory, we have not been able to study each of those genes in complete detail, but nevertheless, we have been able to localize a significant number of them to different steps in the virus replication process that I've outlined so far, beginning from the translation of the incoming viral RNA through the assembly and operation of the replication complex that I've mentioned.

So, for example, we find a number of host genes that are required specifically to support the translation of the non-polydenylated viral RNA, and to also appropriate regulate the differential expression of viral genes. There are other host genes that appear to be involved in targeting and regulated stability. Some of these products, other cellular genes, are required for recruiting the viral template RNA out of translation and into the replication complex. Interestingly, some of those overlap with the translation pool. Other host genes turn out to be very important in providing a membrane environment that is suitable for supporting the formation and function of these replication complexes, and other genes are involved in the chaperone-mediated activation of the replication complex after it is formed, and still others in the survival in the fate of the progeny RNA's.

Now, each of the genes that is listed here we could discuss at some length and its involvement in replication. But, in the interests of time, I will just focus on this class of genes involved in membrane lipid composition, and the specific example of gene OLE1. OLE1 turns out to be delta 9 fatty acid desaturase. It is the key regulator both in yeast and in animal cells for controlling the ratio of unsaturated to saturated fatty acids. If you mutate OLE1, it turns out that you severely block replication just at this point. So, you assemble replication complexes, but they fail to carry out minus strand synthesis for reasons that we are still trying to understand in molecular detail, but for which there are a number of tantalizing possibilities.

Deleted: OLI1

Deleted: OLI1

Deleted: OLI1

Now, further work on this system shows that the virus has no interest in the OLE1 protein whatsoever. It only cares about the lipid composition of the target membrane. This and other studies show that the virus replication is exquisitely sensitive to the membrane lipid composition, much more sensitive than is cell growth providing a significant potential window for therapeutic intervention. So, we think this is a good example of how all of these virus dependencies on host functions are potential targets for antiviral action.

Deleted: OLI1

In the case of this situation, as we have discussed, we see the positive strand RNA viruses, retroviruses, and other viruses all depend on membranes on ways that we now find seem to be similar, and in fact, inhibitors of OLE1 or SCD1 in mammals are in pharmaceutical development now for use in heart disease and diabetes, and such.

Deleted: OLI1

So, to summarize this portion of the talk, systematically testing 80% of yeast genes has identified more than 100 genes that support or inhibit RNA replication. This is certainly not saturating analysis. The analysis so far shows that diverse functions and pathways are involved in expressing and activating the viral proteins, the assembly and function of the RNA replication complexes and again the survival and fate of the progeny RNA. All of these are potentially targets for virus control.

In the last minutes, I'd like to tell you a few introductory words about some of our work on other viruses, on DNA tumor viruses, and discuss Epstein-Barr virus and its association with nasopharyngeal carcinoma. Nasopharyngeal carcinoma is, of course, cancer of the epithelia of the nasopharynx. It is not a major cancer, but still causes about 70,000 deaths per year. NPC is interesting from a mechanistic standpoint because the predominant subtype of NPC is essentially always positive for Epstein-Barr virus which, as I mentioned yesterday, is a very widespread herpes virus that causes lymphomas and carcinomas.

Deleted: So, i

Deleted: little bit

Deleted: contributions to

Deleted: nasalfringal

Deleted: Nasalfringal

Deleted: nasal

Deleted: .

Deleted: . . .

Nasopharyngeal carcinoma is also interesting in that its incidence around the world is very non-random. The highest incidences are in Southern China, where the incidence is 25 to 50-fold higher than the world average. In addition to EBV, additional risk factors associated with NPC

Deleted: Nasalfringal

include an HLA subtype that is prevalent among Chinese but rare in Caucasians, and there is also a link to dietary nitrosamines which are found in the South Chinese diet.

Deleted: ...

In order to better understand the molecular nature of this cancer and EBV's contributions to it, we obtained access through Allan Hildesheim at NCI, to a fairly large collection of NPC tumor samples that had been collected in Taiwan. As an application of functional genomics to this problem, we are characterizing the expression profile of cell and EBV genes in these tumors, and expect in the near the future to be able to report more fully on the nature of genetic changes associated with this virus-associated cancer.

Deleted: Ellen

Deleted: r

So, to summarize the entire talk, unexpected parallels have now linked the genome replication processes of positive strand RNA viruses, retroviruses, and double-stranded RNA viruses. Functional genomics approaches suggest that the dependence of viruses on host genes for their replication is substantial and can provide many targets for intervention, and EBV interacts with additional risk factors to promote NPC tumors, and functional genomics approaches are being used to characterize the changes in these tumors and the contributions of the virus to oncogenesis.

Deleted: We processed a number of these and selected 31 tumor samples and 10 normal nasofringial samples, including some matched normals with some of the tumor samples for further analysis. All of these were then subjected to gene expression profiling using affamatrix chips to profile essentially all the human genes, and using real-time PCR, to profile the expression of all of the EBV latent genes and selected lytic genes as well.¶

¶ The basic result that I want to mention here is that the results imply that EBV regulates the expression of some specific genes in MPC as an important part of its contribution.¶

¶ So, we find under stringent statistical filtering, 831 genes that are consistently differentially expressed across this panel of tumors, and for some of those genes, both up and down, the extent of change in expression from normal as fairly strongly correlated with inter-tumor variations in expression of EBV genes. For example, expression levels of EBV ebda 1 correlate quite strongly with inhibition of multiple cell genes involved in class 1 antigen presentation, implying that EBV contributes to immune evasion by the tumors, and at the same time, EBV LMP1 expression levels correlate strongly with the activation of a number of genes that are implicated in metastasis and suppression of apoptosis and so forth.¶

For all of these projects, I'd like to acknowledge the principal workers and particularly Michael Schwartz, Chen Jianbo, Wang Xiaofeng . . ., Dave Kushner, Amine Noueir, Lee Waiming, Srikumar Sengupta, Bill Sugden, Allan Hildesheim, and Chen Chienjen.

Deleted: modulates a number of cell genes in MPC

Deleted: the proliferation and survival of MPC

Deleted: Chin

Deleted: . . ., Wong

Deleted: Kushner

Deleted: Mi Nuri

Deleted: Li Y Ming

Deleted: . . .

Deleted: Sugdin

Deleted: Allen

Deleted: J

Thank you very much.

Moderator – The last speaker is Professor Jiang, Deputy Director of the Shanghai Institute of Materia Medica. The title of his presentation is discovering anti-virus compounds from the existing drugs.

**Jiang Hualiang, Shanghai Institute of Materia Medica, CAS**

### *Discovering anti-SARS compounds from the existing drugs*

Thank you. Ladies and gentlemen, it is my great honor to have this opportunity to talk about our work. We are facing more and more . . . new virus and new bacterias. So, . . . method to deal with this new virus or new bacterias. Today, I would like to introduce strategy and related technologies to quickly find . . virus compound or antibacterial compounds from the existing drugs.

So, drug discovery and development is hard work and these slides show the process we have to screen tens of thousands of compounds and find a lead, and then optimize that lead and find a candidate and move to the pre-clinical tests and then clinical tests of Phase 1, Phase 2, Phase 3. It will take more than ten years and it costs a lot of money.

So, quickly discovering drugs for combating the emerging viruses and bacterias is a challenge. So, screening the non-drugs might be appreciated short-cut to discovery drugs . . . because all these drugs would have very good absorption distribution, . . . and even toxicity profiles. Once we find those active compounds, we can move to the clinical trial immediately. So, this is our idea – it is very simple, but now we have more than twenty-thousand drugs. For example, for HIV, one virus can turn a lot of your . . . If we collect all these examples and spring by using the experimental . . . or these compounds through all these targets is very hard work. It is almost impractical.

So, our strategy is . . . . first deputy . . . . virtual screening. The target is a specific protein structures, maybe this is a good strategy because you can narrow down to . . . potential candidates before expert . . . testing.

This slide shows the strategy for virtual screening. We collect target structures – most of them are proteins. Also, we collect . . . and using computation programs, . . . constructed into the

binding pocket according to the . . . . binding affinity, and . . . . molecules are possible, active compounds.

Our . . . . one million compounds per day. So, because of time and limitation, I have to finish my talk quickly and here I gave you a typical example produced from our laboratories. Used in this strategy, we find a . . . . drug code . . . . It is an inhibitor of SARS, . . . . protenase. . . . reduced virus application . . . in vitro.

So, when we started this project in 2003, they increased the structure of SARS . . . virus protenase has not been determined. After about three months, Professor . . . . in . . . University harvested a crystal structure. Before that, we construct a three-D structure model for the SARS . . . virus protenase. This is based on the variable structure . . . TGED . . . protenase structure.

Here comparing the binding site, the SARS . . . protenase structure is very similar to the TGVE protenase structure. So, we screen several databases, but then we were interested in mostly this database, CMC, because if this database . . . about 10,000 variable . . . structure information.

This left 10 candidates, but unfortunately, we cannot . . . chemical . . . . for all of these chemical structures. But, finally we find just one compound . . . . DDDC 001. But, it is actually . . . . It is an O-drug.

. . . indicate that this structure can fit in the binding site of protenase very well. This structure is very simple, but its . . . serotonin antagonists. This has been used in the clinical trial in the 1970's for this . . . . .

Based on the computational result, we do the experimental result and firstly, at the time of the binding affinity between these drug and protenase, the binding affinity is okay and also in agreement with our British prediction, and then we determined the enzymetrical inhibitor reactive data. We both . . . inhibits . . . enzyme very well.

**(Remainder not transcribed – too disjointed)**

Moderator – Next . . . . Any questions and comments?

### **Q&A Panel Discussion**

Question – I have question for Dr. . . .? You show the compounds have the effect on the . . . in the animal test . . .

Answer - . . . animal testing because the SARS and . . . No stable SARS model.

Question – I have a question for Dr. Taillon from 454 Life Sciences. In my opinion, one of the major advantages of your sequencing technique is that it can produce 20,000 read . . . very short in the Sanger sequencing. So, my concern is that if you want to finish a genome, in the finishing phase, as I know a . . . disadvantage. So, do we think it is necessary to combine your technique with the Sanger sequencing technique for . . . genome . . .

Taillon – Absolutely. Just to correct one thing – it is 200,000 reads per run – not 20,000. But, you're absolutely right. Recently there was a genome web article from JGI which has one of our instruments – Joint Genome Institute, which is a Department of Energy funded sequencing center which does a lot of bacterial sequencing now. What they have done is adopted the GS20 as part of their approach. So, what they are doing is generating 5 to 10x coverage of the bacterial genomes with our instrument and then doing the finishing with Sanger sequencing. We've seen that as a great approach and we certainly support it. You need all of these technologies to answer scientific questions. We bring something to the table that can get you there faster. So, what they purport, and these are their slides, that basically in a month they can accomplish using the combination approach what would have taken many months in the past to accomplish. Not only is it good to combine the two technologies for finishing, but we also saw in collaboration with the Brode Institute when we sequenced . . . gonorrhea, we saw a larger . . . organism, a larger not being mammalian – is that the overlap of the sequence between the two technologies

was such that 454 sequence was covering areas that had be unrepresented in their approach to do Sanger sequencing. I assume if we had started with GS20 sequencing first, then Sanger sequencing would have covered sequences that we missed. So, definitely the two technologies and any technologies that come behind us (hopefully not too soon) come behind us will also complement the approaches.

Question – I have a question for Dr. Ahlquist. . . . beautiful work by using the yeast system to do the analysis of the least minimum host components or host centers for the viral RN replication. So, in your systems – so far, how many viruses or family or groups of viruses you can use this yeast system to do this analysis?

Ahlquist – Unfortunately, the number is very small. So, I didn't mean to suggest that was a panacea with working with all viruses. We have the two that I mentioned. One other success has been obtained with viruses that are related to these noda viruses that I mentioned in passing. Beyond that, there is almost nothing. So, there are a number of other viruses that have been tried, but without success.

As you may know, there have been successes recently with several viruses in C. elegans, including the noda virus that I mentioned – FHV. But, in all of those cases in C. elegans so far, it is clear that one has to use fairly large amounts of virus to initiate an infection. We know from manipulations in our system and it is known in retrovirus interactions with host factors, that as you increase the inoculum load that you're placing on the system, you can alter what is happening. So, there are challenges in all such systems.

- Deleted: seaelegance
- Deleted: although one of them is
- Deleted: seaelegance
- Deleted: very
- Deleted: well-
- Deleted: radically

Question – Paul, could I get you to follow up – are there other simple eucariotic systems that might capture the other classes of virus? I'm thinking of seaelegance or maybe some other fungi. Do you know if that space has been explored?

Ahlquist – As I mentioned, there have been successes recently with C. elegans, with several viruses. So, C. elegans would be a next step up. So, one of the difficulties is that in many of these simple eukaryotes, there are not characterized viruses that would be good models for things

- Deleted: seaelegance
- Deleted: seaelegance
- Deleted: the
- Deleted: eucariots

that we're interested in. There are a few, for example, . . . virus-like agent, and . . . would be an example where there is something useful.

Question – I have two questions to Professor Paul. I came from the Institute of Microbiology. My name is . . . . I do appreciate your excellent work involved in host . . . identification. I know your system or BMV and . . . is one of two well-done systems now in the world so far. Since you have identified some host factors, now I want to know, how many host factors in your system have been identified? The second – I want to know . . . seems to be one nice host factor. . . for a potential . . . . So, I want to know whether you have found some polypeptide or some molecular inhibitors against these good targets?

Ahlquist – I discussed one examples under the constraints of the time, but we could have a similar discussion for many of the factors that have been characterized so far. One simple example would be that we have published that there is chaperone-mediated activation of the replication complex after its assembly and in a number of cases, there are chaperone inhibitors that might be tried. For example, geldanimycin is in trials in cancer and might be pursued in such regard. There would be other examples as well.

We have not done extensive work in terms of pharmaceutical modification. We have primarily pursued these things by genetic manipulation, but I think the implication that one could manipulate pharmaceutically is obvious.

In other work, we are also looking at host factors in the replication of some retroviruses and it happens there that we have picked up some things for which it was accessible to predict some chemical inhibition. In those cases, we have been successful in getting that. So, I think in general a genetic effect should be a good indication of a potential use as a pharmaceutical target.

Question – I want to ask you another question. Because I know in the . . . . has been identified and they have all served as host factors to support BMV replication in . . . . In . . . paper, we

**Deleted:** Question – I have a question for Dr. Paul too. You compared MPC tissues and the normal tissues. I would like to know how did you compare them? Did you pool the MPC tissues together or compare the MPC tissues with normal tissues . . . pull them together to compare them?¶

¶ Ahlquist – No, all of the tumor samples were assuredly analyzed separately.¶

¶ Question – How much . . . . is contained the tissue . . . .¶

¶ Ahlquist – That is a very significant issue and all the samples that we used with rare exceptions, we did laser capture micro-dissection of all of the samples to take only tumor tissue and remove surrounding normal tissue. This is a significant issue in MPC where there is typically tremendous amount of lymphocyte infiltration, so typically a tumor field has vast amounts of lymphocytes in it, in addition to potentially other cells.¶

¶ Question – May I ask you a favor? In your presentation you mentioned you thought about . . . . Could you – may I have this gene's name? I can give you an e-mail address.¶

¶ Ahlquist – Hopefully, much of this data will be published soon. So, we would hope at that point to provide you with such a list. But, we can also discuss . . . .¶

**Deleted:**

have found some homologue to . . . . So, I want to know, are they all homologues . . . in the other . . . .

Ahlquist – Yes. For example, in the case of OLE1, it is a well-known gene in mammals. The name most often used in mammals is SCD-1 for stearoyl co-A desaturase. There are knock-out mice and their phenotypes have been studied. So, yes.

Deleted: Oli-1

Deleted: . . . . disaturates

Question – OLI1 seem to be very popular in the . . . .?

Ahlquist – Yes. It is the first enzyme in making unsaturated fatty acids. So, yes.

Question – If so, it seems to be very universal . . . . But, I want to know, if we knocked out this gene, the mutant . . . . if in our humans we knock out this gene, I want to know whether we can acquire this disease . . . for HIV and HPV and others?

Ahlquist – It happens that nearly all organisms operate with a substantial excess of unsaturated fatty acids above their minimum needs, typically, in five-fold or greater excess. One of the reasons for that appears to be so that you can cope with drops in temperature. If you don't have excess unsaturated fatty acids and the temperature drops, then everything locks up in the membrane. So, that is a considerable range over which one can modulate the level of unsaturated fatty acid in a membrane. What we found is in the case of these viruses, that we can drop by mutation or by feeding experiments in a knock-out cell. We can easily drop the unsaturated fatty acid levels to a range in which the virus is essentially dead and the cell is quite happy.

Moderator – We have time for a last question. Any questions for Professor Wang? If no more questions, I want to say thanks to all the speakers again and we will have lunch?

**Panel VI: Case Studies – Prevention and Treatment**

**Moderator: Huntington F. Willard, U.S. GDEST Program Committee**

. . Institute for Genome Sciences and Policy at Duke University in the United States, and had the pleasure of being on the program committee for this conference and will chair this afternoon's session.

Our first speaker is Professor Wei Lai, from the Hepatology Institute at Peking University who will be speaking to us on hepatitis B.

**Wei Lai, Health Science Center, Peking University**

***HbeAg Negative Hepatitis B, Mutation Genotype, Genotyping Relationship, and Outcome***

Chairman and ladies and gentlemen, good afternoon. In China, the hepatitis B and all hepatitis such as hepatitis A, hepatitis C and . . . belongs to the infectious disease. If they have a different transmission route . . . hepatitis B and C, most patients . . . transfusion. Sometimes it was an unsafe injection.

Here I want to talk about a new concept, HbeAg negative hepatitis. Most of the patients fall under hepatitis B, or the positive HbeAg. However, . . . patients to have the higher . . and higher DNA, . . but they have the elective HbeAg. So, what is the relationship for the HbeAg negative to the genotyping and . . . . That is what I want to talk to you about this afternoon.

First, when you talk about genotyping, I would like to introduce the geographic distribution of hepatitis B virus genotyping in China. China is a huge country and the different areas have the different prevalence of hepatitis B. Also, for the difference to hepatitis B genotyping . . .

In this slide, we also concluded some of the large cities in China . . . . In some areas, most patients . . .

**(not transcribed)**

Moderator – Thank you, Professor Wei. All of the speakers will take questions at the end as we have done in the other sessions. Our next speaker is Dr. Steve O’Brien from the National Cancer Institute who will speak to us on genetic architecture of complex infectious diseases – lessons from AIDS.

**Steve O’Brien, National Cancer Institute**

***Genetic Architecture of Complex Infectious Diseases: Lessons from AIDS***

Thanks Hunt. It is great to be here and while I’m hooking up, I’d like to thank all the organizers listed behind me, but if I did, it would take up all my time because there are a lot of organizers, for allowing me to come here.

Halfway across the world and maybe a generation ago, the species conservation community learned an important lesson about the role of host genetics on the outcome of infectious diseases from an unlikely endangered species – the African cheetah. The species as, at the time, unusual because it had been shown to have survived a near extinction event some 10,000 years ago, which had homogenized its overall genetic variation intensely. But, notably, the genetic

variation that is involved in immune defenses, the different armaments which equip species to defend themselves against infectious agents they have seen in the past as well as new ones which are invented during their lifetime. An outbreak took place in the most successful breeding facility of cheetahs in North America, a small drive-through park called Wildlife Safari which had close to 100 cheetahs breeding successfully. A couple of animals were brought in and became ill while in quarantine, and died of a disease caused by a corona virus, a relative of the SARS virus called feline infectious peritonitis. The virus spread through the park within six months and infected nearly every animal in the park, leading to the increase in antibodies against the virus, as is shown in this family. Before the arrival of the first two cheetahs that were infected, all the animals were negative. But, within a year, everybody had become infected. There was not a normal stool in the park. All the animals had diarrhea. All the animals had nervous twitches. All the animals had jaundice like hepatitis. Within six months, 50% of the animals were dead and 70% of the animals were dead within three years. It was the worst outbreak of feline infectious peritonitis ever seen in any species.

Twenty years later, in the Gwong Dong province of China, a flu-like syndrome developed into the SARS epidemic late in the year 2002. It spread to 29 countries, and inflicted enormous emotional, cultural, and economic pressure upon many countries, particularly China and other Asian countries. When the epidemic burned out at the end of July, 2003, there had been a total of 8,000 people infected, and 700 deaths, a mortality rate of about 10%.

The SARS virus was a member of a corona virus family which we knew about because corona virus had been described in domestic animals, and had also been responsible for about one-third of human common colds. When the sequence was analyzed in a phylogenetic method, the SARS virus however did not fall in any of the traditional clusters of avian corona virus, or human corona virus related to bovine and mouse hepatitis virus, but rather came as an off-shoot which was distinct because it was evolving probably in a different host or reservoir species recently identified as several species of bats.

The cheetah virus, when sequenced retrospectively, turned out not to be a relative of the SARS virus, but to be almost indistinguishable from the domestic cat virus. The difference between the

mortality and the morbidity in cheetahs and humans was probably something to do with the genetic uniformity of the cheetah species because the mortality in humans, as bad as it was, was only about 5-10% and in outbred domestic cats, the same virus has less than 2% mortality. But, in cheetahs, this high uniformity is probably a consequence of a virus that overcame the defenses of a first individual, looked around and found the rest were immunological clones of each other, leading to a rather uniform homogeneous response. The genetics of the host can make a difference in how well a species does when it incurs an infectious agent that is deadly.

Like many historic epidemics, HIV/AIDS is very deadly. It does not respect geographic, ethnic, cultural, or social strata, but even so, displays some considerable heterogeneity of outcomes among different people. Today, it is my pleasure to present a brief overview of the genetic variance, AIDS restriction genes, or ARGs, that regulate an individual's response to HIV and infection. I will attempt to highlight a bit of the gene discovery process, the mechanisms by which the genes act, the implication of host genotypes on the progression of the epidemic, how the genes may help inspire new therapies, and the process for hopeful but not quite yet ready for prime-time future discoveries.

To begin with, I shall look back a decade, to 1996, the year that two similar discoveries gave us some needed hope for the horrors that we had known as AIDS. AIDS, of course, was first discovered in the early 1980's, about the time of the cheetah outbreaks that I described a moment ago, and led to ten years of a steady increase in the number of mortalities, of an incurable disease which still today we have no vaccine and no effective treatment.

The discoveries that came in 1996 were two. The first was the application of the powerful, evolution-inspired strategy of triple drug anti-retroviral therapies. HAART – highly active anti-retroviral therapy which for the first time actually caused a decrease in AIDS mortality, at least in the west, and the mortality had been steadily rising up until that time, since it was first recognized 15 years earlier.

The second discovery was a little bit more arcane. It was the contemporaneous announcements within a week of each other in July 1996, in back-to-back articles in *Science*, *Nature* and *Self*,

written by five outstanding research groups which identified CCR5 and CXCR4, heretofore innocuous . . . receptor molecules, as the required portals or doorways by which HIV enters macrophages, monocytes, and T-lymphocytes. The ligands which normally bind these chemokine receptors, RANTES or SDF1 physically interfere with the binding of HIV and HIV entry. That was how the receptors were identified.

By this time, our group had assembled a consortium of exquisitely curated AIDS cohort populations, extracted their DNA, and were screening human variance that might explain why not all exposed have become infected and why not all infected had developed AIDS. These included patients in epidemiological cohorts in the risk group of men who have sex with men, homosexual men, hemophiliacs, who receive contaminated clotting factor before the implementation of the HIV blood test in 1984, and IV drug users who shared needles in city slums where the HIV incidence in that risk group population approached 80%.

Well, we, and others, rapidly re-sequenced the coding genes for the chemokine receptors and the ligands to discover CCR5 delta 32, a natural genetic knock-out deletion that was carried in one in five European Americans, but was never seen in native Asians or in native-born Africans. People who carried two copies of CCR5 delta 32, between 1% and 2% of all European Americans, were near completely missing from HIV-infected people, suggesting that the CCR5 molecule on the surface of T-lymphoid cells, was a necessary requirement for HIV infection.

As illustrated in this slide, the frequency of homo . . . for delta 32 varies between 1% and 2% among European Americans, but is nearly completely absent in HIV-positive people because people without CCR5, people with two copies of delta 32, one from each parent, simply do not have an entry portal by which HIV to enter cells, and are completely or near completely resistant to HIV infection no matter how many times they become exposed.

People who carry one normal copy of CCR5 and one deletion copy do become infected. However, they progress to AIDS-defining conditions slightly more slowly, as is illustrated by the survival curve where sero-converted patients are plotted at the rate at which they go on to

develop AIDS, and individuals that are heterozygous actually delay the onset of AIDS by two to four years, a modest and very slight difference, but it is statistically significant.

In addition, when we inspect the type of AIDS-defining conditions that people develop when they carry CCR5 delta 32, what we discover is that HIV-infected people show a difference in how frequently they develop non-Hodgkin's lymphoma. A B-cell lymphoma where individuals who carry delta 32 are twice as frequent among those who avoid lymphoma as those who are not. So, those three effects: a recessive effect on infection, a dominant effect on rate of progression, and a dominant effect on AIDS-defining conditions were the hallmark of AIDS restriction genes, as illustrated by CCR5.

Now CCR5 was interesting because, as I mentioned, it probably was born after the divergence of ancestor of modern Europeans from the individuals who migrated out of Africa 150,000 years ago, and then split to the Asian and Caucasian derivatives. In addition, the frequency of delta 32 shows a gene frequency gradient across Europe where it is highest up north in Scandinavia and in Scotland, 10% in France and Germany, 5% in Greece, Europe and Turkey, and 0% in the Middle East and Sub-Saharan Africa. This gene frequency gradient is suggestive of some sort of favoring or strong selective pressure in the north. In addition, the rise of a mutation along this lineage from a number of a very high number of people to up a frequency of one in five Europeans carry this as heterozygotes, was rather unusual because 99% of all mutations go extinct. So, several pieces of evidence led me to speculate, and my colleagues, several years ago that CCR5 delta 32 had . . .

. . . to her offspring to a frequency of one in five carriers in Europe. The evidences have been summarized adequately in the literature, but they include the fact that CCR5 is a knock-out mutation which eliminates a perfectly good immune function, but yet instead of being eliminated, it rose up.

In addition to that, a founder effect, which would explain a rise quickly, probably was not occurring in the timeframe we're talking about, but rather than European populations have seldom been less than 20,000 over the last 50,000 years. There is no evidence for a population

bottleneck like we saw in the African cheetahs. Delta 32 frequency, defined north/south gradient, a signature of selection, and finally, if you look at the pattern of variation in other sites in the CCR5 molecule, most of the mutations are amino acid altering. A signature of a favoring of molecules, as has been seen in the pattern of variation in HLA, the human major . . . compatibility complex, and not seen in other genes that are not selected by outbreaks of infectious diseases in the past. So, CCR5 seems to have been jacked up probably by infectious disease, but what was it?

The next key to the puzzle takes into account tricks that the molecular revolution have taught us, the ability to recognize that there is a timekeeper, a chronometer if you will, in the genes of all human beings, that can allow us to estimate how long it has been since a particular mutation took place, or more precisely, how far back it has been since a particular mutation that was selected was selected. How far back it was. The measure is the randomization of the single nucleotide polymorphisms that are adjacent to the mutation that occurred. The phenomenon of non-random association of adjacent variances, called linkage of this equilibrium. It has been measured around CCR5 several times. Originally we used micro-satellites back in 1999, as have other workers in most recently single nucleotide variance or SNV-based analysis have indicated that the size of the haplotype or the non-random structure around CCR5 is about a centi-morgan long. That centi-morgan long haplotype translates using a simple equation to a time or estimate of the date at around 700 years ago. This is the period that has elapsed since the last delta 32 or that chunk would have been selected.

Of course, this immediately led us to wonder whether in Europe, 700 years ago, that the black death may have been responsible for causing the favoring and the transmission of the delta 32 molecule. It was an interesting hypothesis. It was indirect, and it was guesswork. But, it was the best candidate we had based upon the timing of the mutation and the fact that everybody seemed to think the delta 32 had been selected.

The black death was a massive epidemic which took about one-third of Europeans in a four-year period between 1348 and 1352 in the 14<sup>th</sup> century, that started in Mesina, Sicily and it migrated throughout Europe in a bit of a circle, a little bit like the SARS epidemic except them it quite

abruptly, just like SARS. During that period, it took many cities up to 50% mortality, 60% in other cities. It basically then stopped. Several generations after the black death, there have been other waves of Bubonic plague caused by . . . . that have passed through.

So, in addition to that, there have been waves of plague from several hundred years before, back to the Justinian period of the Roman Empire, which could also have contributed to selective advantages of carriers. If delta 32 actually had an effect on the cause of black death, which was . . . bacterium.

Having said that, there is not a lot of agreement that *Yersinia pestis* was, in fact, the cause of delta 32 rising up, even though we guessed that it might be. There have been other theories about what might have caused it, such as small pox, anthrax, cholera, Ebola, typhoid, even the Vikings having a strong influence. But recently, an interesting connection which connects *Yersinia pestis* directly to CCR5 was conducted by Stephen Albin and his colleagues and published in Nature last year. What they simply did was they took mice that were knock-outs for CCR5 and they isolated macrophages, and then they measured uptake of *Yersinia pestis* from cells that had a normal CCR5 and cells that didn't. What they saw was a 60-fold difference and a requirement for CCR5 for *Yersinia pestis* to be taken up, connecting directly *Yersinia pestis* biology to CCR5. I think this basically puts *Yersinia pestis* in the running in addition to some of these other candidates.

I'd like to change gears now and talk about some of the other AIDS restriction genes that we have been able to discover. But, before I do that, let me say that one of the reasons we were interested in identifying AIDS restriction genes was to point the new way to natural genetic resistance that might be useful for developing new kinds of therapy against this deadly disease.

Well, delta 32 discovery, which demonstrated that HIV's entry is necessary and can be inhibited by a natural mutation, actually has led to the development of a whole new generation of drugs which are different from the 16 heart drugs that are used right now. They are called entry inhibitors and this slide here is a summary from a recent discussion. There is a dozen of these made by different companies, all of which block HIV from interacting with CCR5 or CD4, or

facilitating the fusion process. Many of them have entered clinical trials. One of them, fusion or . . . , was actually approved by the FDA – the first non-viral drug ever approved for HIV. It is pretty exciting and I think the development of these things was precipitated and facilitated by the knowledge that delta 32 homozygotes don't become infected.

There are other kinds of human AIDS restriction genes that have also been discovered using the same approach since delta 32. We have discovered AIDS restriction genes that involve HIV entry that involve acquired immunity, innate immunity, and also HIV regulation. I'm going to show you three slides now which simply summarize these, and then talk a little bit about what they mean.

This is six genes that modify entry. I've talked about CCR5. There are others such as CCR2 and ligands, SDF, rantes, MCP, all of which have an effect on AIDS progression and have been published in pretty good journals and replicated in independent laboratories. These are genes that effect acquired immunity – that is, the ability to mount a humoral or cell mediated response. They are HLA homozygosity, the interaction of the natural killer cells in HLA, and then certain alleles of HLA that specifically regulate antigen presentation and epitope recognition that have been demonstrated directly to influence these cohorts and how well you do when you're infected with HIV. In addition to that, there are genes such as interferon gamma or cytokines like IL10 which influence AIDS progression, apobec, which is a molecule that is involved in interaction with HIV, the cure HLA, a cancer, a regulatory transcription factor – all of these have SNIP or mutation variations which up-regulate or down-regulate expression of these genes, which have like CCR5 delta 32, been linked to regulating quantitatively how rapidly a population progresses to AIDS or progresses to AIDS-defining conditions.

In addition to that, there is now a big interest in discovering the genes that influence why some people do better than others when they are treated with highly-active anti-retroviral therapy. The HAART therapy actually includes the idea of treating with three drugs so that mutational reversion is very rare, and they usually include a couple of nucleotide reverse transcriptase inhibitors two plus one not nucleotide reverse transcriptase inhibitors or other combinations.

But, unfortunately, it doesn't work for everybody. In fact, in the initial 18 months of a study population that we have looked at, the max, about 40% have succeeded and 60% fail. Within five years after changing the combinations and seeing what best, about 25% still fail. Some of that is due to individual compliance. They just don't take the drugs. But, much of it may not be. It may have a genetic explanation.

So, we guessed that there might be some genetic influences on three steps in HAART therapy. The first is failure to suppress the virus itself. The second is how slowly or rapidly a person goes on to get worse to develop AIDS. Third, the adverse events. We seem to be trading one disease, AIDS, in for hypertension and cardiac abnormalities, and lipodystrophy and drug toxicity. The kinetics of this has a genetic component that influences it.

This is a slide that simply lists the names of some of the genes I've just shown you a few minutes ago. They are genes that other laboratories have implicated in influencing what happens with HAART. We investigated these, plus the other genes that we have already discovered, the 20 or so AIDS restriction genes, and tried to find out whether or not we detect any influence of these AIDS restriction genes on survival. The answer, in short, is we do, and I'm going to show you three quick slides that are survival slides that show you kind of what I mean.

This, of example, is the discrimination between people that basically have different genotypes for delta 32 after they have had HAART therapy and how well they do. It is good to be heterozygous – it slows down the rapid progression of AIDS in the presence of HAART therapy. This is another mutation in rantees – one of the ligands for CCR5 that basically shows a rapid progression in the rare allyl. This is a slide which illustrates a different way of measuring survival, measuring the concentration of virus in hundreds of patients over the course of the time since they sero-converted, but before HAART therapy. This is starting at treatment and after HAART therapy. The delta 32 mutation shows a significant difference, although it is not a big deal.

Another way of looking at it is looking at the CD4 counts, which is the hallmark of AIDS development. In the absence of HAART, there is a difference in delta 32, and in the presence of HAART, there doesn't seem to be a very big difference, as indicated here.

For some of the other genes like rantees, the viral load seems to show a massive difference. These are just different ways at looking at how well a patient will do under these therapies.

In sum, these are the genes that we saw that had an effect in both HAART and in natural history before HAART. This is natural history before HAART. The red means highly significant and bad for you. You don't want to have this gene if you're on HAART. Green means it is good for you – it slows things down.

Viral suppression, progression to AIDS after HAART, CD4 slope – these are the summation of the results which show replication of some of these genes of being bad after HAART or good after HAART, as they were in the natural history studies and so basically the conclusion of this slide is not too surprising. But, one important point is that the replication of influences after HAART therapy means that ARGs continue to influence AIDS post-HAART – that means the virus is still replicating even in patients with it completely suppressed.

Two final points. The AIDS restriction genes – how important are they? This is a list of 20. They have different names. Some are dominant; some are recessive; they have different effects on HIV progression or infection. They have all been published, and many of them have been replicated in other laboratories.

Well, when epidemics vary in many ways, what are the guesses as to why there is a variation? Well, genetics of the virus. The genetics of the host. The patient's history. The non-genetic environmental influences, the poisons and the chemicals in nutrition, AIDS circumcision makes a difference, stochastic stuff – just bad luck for some people – or all of the above. Is it possible to quantify them? Is it possible to develop a yard stick so we know where we are in measuring genetic versus environmental influences?

Well, epidemiologists have a couple of things that they are used to talking about. One is relative risk, which is a measure of how strong the effect is. The second is attributable risk, which is when you go to the hospital and you get all the sick people, you saw what fraction of them are in there because they can carry a particular genotype. But, these numbers can be misleading for reasons that I don't have time to go into. A third, and I think more useful value, is something called explained fraction, which is to take all the epidemiological variance in AIDS survival – how much can we explain by the genotype of an individual, looking at all the AIDS restriction genes? In this we developed a method by mutual information theory using a small 2 x 2 contingency table. It is pretty simple. You can do it on a hand-held calculator.

What am I talking about here? These are the relative risks, the attributable risks, and the explained fraction for a number of AIDS restriction genes not the rate of progression. These are a bunch of genes that are good for you – that you'd like to have if you got infected with HIV, CCR5 and so forth. These are a bunch that are bad for you. What it does is it estimates the frequency of the genotype in the population, the relative risks, how much better off you are worse off. The attributable risks – what fraction of the patients in the hospital that are sick – you can explain by these. The explained fraction – which is how much of the overall variance can you explain.

Well, the totals are the bottom line, and the combined protective fraction is pretty high for the attributable risk – about 22% of the sick people we can explain by their composite genotype of one or more of these genes. But, the total variance – how much variability in the epidemic itself, taking everybody out three, is small – only about 10% of the variance we can explain. That means we have 90% more to explain. So, if we have a yardstick from 0 to 100 yards, we are only on the 10 yard line with the genes we know about. So, then 90% of the variation is not explained. So, how does this inference in form, the ability of our genotypes to inform clinical trials? Can we hand the genotypes of AIDS restriction genes to people doing clinical trials for vaccines and drugs and say, let's factor this into your equation so you can just subtract out the genetic noise? And, how well does the genotype imputed propensity index – this is the composite genotypic score, predict the time to develop AIDS?

Well, here is the answer. It is not very good. If we actually take the genotype of a bunch of patients and we predict how long it is going to take them to get AIDS, and then we actually measure how long it is going to take, this is the predicted. This is the observed for the same people. You get what looks like a circle. There is a slight regression, but it is tiny. That is what you'd expect since 90% of the variance we haven't explained this. We are only explaining 10%. So, this isn't ready for applications yet. So, comparing the predicted time to AIDS based upon the composite genotype, the actual time, we actually wind up with a disappointingly poor predictive value prognostic.

So, the short is only about 9% of the epidemiological variance is explained by the 17 genes. This is not enough for clinical trials.

What is the good news? Well, about 10% of . . . is actually quite a bit. If you do the same thing for smoking and lung cancer, we did that. Actually, smoking only explains about 10% of the variance too. So, these 20 genes that we have in the AIDS epidemic explains as much of the variance as smoking does in lung cancer, which we all know is a big deal.

In addition, genes with relatively modest influences can readily be discovered with these large cohorts. So, we think there may be some few more genes to discover and certainly other epidemiological things.

Now, my last five minutes will be developing into the future. We would like to take the haplotype map kind of idea and – let's go back. This is a summary of some of the milestones of the human genome project which epidemiologists want to use to discover new genes. One was in October of 2004. We had a human genome project which annotated the 20,000 odd genes in the human genome. Then, over the last several years, almost ten million single nucleotide variants have been annotated and a few months ago, the first one million snip level hap map was released, annotated in haplotype blocks. The average block size was about 14 kb in Caucasians. Beyond that, what is going on now is many species of mammals are being sequenced to identify regulatory reasons that are conserved through evolutionary inference.

Now, we wanted to see whether or not the haplotype-based analysis could be used and whether we could develop new tools, the subject of this conference, to help handle large amounts of genotyping data using the AIDS cohorts. So, what we did was we took 2,600 patients that were selected for highly informed clinical data and then we basically took eight known age-restriction genes and then we spaced snips across each of the genes going up 400 base pairs in one direction, and 400 in another. We built the haplotype structure around each gene, then we ran snip association tests for each snip using a real frequency, genotypes, dominant recessive, a number of different tests for AIDS outcomes. Total – 236 tests for each snip. They were basically broken down into infection effects, progression effects, age-defining conditions, and HAART therapy. What we wanted to do was evaluate the power of the causal or operative snip, and then what we call a proxy snips. These are the ones that are next to the operative snip, but are in linkages of this equilibrium, but track them in association – and then the haplotypes themselves.

These are the eight restriction genes we looked at – IL10, CCR2 and CCR5, chromosome . . . , rantees, SDF, and then two control regions: one region chromosome on #7 which had the CFTR regions which we didn't expect to effect AIDS, and then another region of chromosome #16 which was derived from one of these signatures of selection discovery panels.

Then we looked at each region for a number of things. This is a region of IL10 on chromosome #1. Basically, it is about 800 base pairs across, this is the linkage to the equilibrium pattern where the red indicates non-random linkage to equilibrium. These are the genes that are inside – IL10 is this one here. These are the snips that we interrogated. These are the haplotype blocks that are built by algorithms that allow us to know what the structure is. This is chromosome 3 – the same kind of idea – an LD scan here, the snips, the genes, chromosome CCR5 is here, CCR2 is here. These are the snips that were interrogated. Again, we are talking about hundreds of snips across all the genes, and these are the blocks with the number of haplotypes.

Now, a couple of new tools. First of all, . . . when you're talking about this number of snips and this number of tests and this number of genotypes, you're talking about 700,000 genotypes, you're talking about 200,000 combinations of tests. We need a better way to look at this other

than the way we did when we discovered CCR5. So, the first one that we developed was something that was based on a concept you've heard a lot about at this meeting – the affimetrix expression array. It is basically a heat plot where on one axis we list all the tests – the infection tests, the progression tests, the sequely or HAART test, and then on the other side we list the snips in the order at which they occur on the map. So, in all, the concept is based on the affimetrix express array chips.

So, what I'm going to show you is an expression array where the colors are indicating the P-value – how significant it is. The stronger the P-value, the better the color. The first one I'm going to show you is going to be the control region CFTR. Now, these are the snips that we assayed across the FTR and these are the 136 tests – chromosome 7. The orange means it is between .05 and .01. There are plenty of those by chance alone, and then there is a few .01. This is the background. This is what you expect when nothing is happening.

Now, let's go to chromosome #1 – IL10. In here, what you see is this operative snip, which is the red one, is this guy here and the blue ones are the proxy snips next to it. You can see the infection. You do see them infect, and if you blow it up, you can see which of the tests are actually firing and which of the data itself. Then we also look at progression and you see a bunch of highly significant tests here. Again, validating what we had hoped to see by simply viewing a whole bunch of tests and a whole bunch of snips at the same time.

Chromosome 3 was beautiful. This is where CCR5 is, and CCR5 promoter, and CCR2 – they are all linked closely together. As you can see, we see a very strong signal for infection here, a rather strong signal for a progression where you see lots of colors which means combinations of proxy snips, and different tests that are non-overlapping, non-independent. In addition, what is going on up here? This is not CCR5 or CCR2. It is a new region. They are showing us signals that we didn't know about before. That is basically what we expect in a genome scan.

If you compare CFTR, the chromosome 7 control region, with the CCR5 two regions, you can see this is what we want to see, define new genes, this is the background. That is really what the tool does – it gives us the chance to inspect it through a visual by eye.

We have another test which takes into account not only the P-value, but also the relative risk or the odds ratio – how much it is. What this does is it simply takes one snip and it looks at the 256 tests and it says, what is the lowest P-value we see for that test. Then, we take that P-value and we compare it to the next snip and the next snip. Then, we rank all 256 lowest snips and then we rank them.

Then, we do the same thing for the odds ratio divided by P. So, you take the highest odds ratio and divide it by the lowest P-value, and we rank it within one snip and then we rank all the snips compared to each other in the region and through all the other regions as well. So, these are five different kinds of ranks. One is snip the lowest P-value. The second is the number of tests where P-values are less than one. The third and fourth are odds ratio divided by P's, but at different significance levels. The fifth is just simply the average of the previous four.

What I'm going to show you is a plot which looks like the gene itself with all the snips along the baseline, but ranked for these five ranks, starting with the control region which is CFTR (cystic fibrosis). What you see here is the first, second, third, fourth, fifth rank. This is the gene itself, the gene region, 800 base pairs, about 50 snips ranked across. As you can see in this particular test which happens to be the fourth ranking, there is no dipping below 250. There is a couple here and there, but there is no consistency. This is a different analysis. This is for infection. This is for AIDS progression. You see a couple of things going on. This is a negative control. This is what you want to see in most regions.

Now, let's go to IL10 where there was a signal. Here what you can see is that there is a significant dip of the operative snip plus the proxy snips next to it for all five rankings. Bingo – that's a signal. That is what we want to see. When we look at the progression, it is also dramatic there. So, what you want to use is replication in the different kinds of rankings across that we can indicate. That is what we're looking for in this kind of test.

When we look at the chromosome 3 region, the one that was so bright on the . . . array, we also see a lot more action than we saw on the control region. There is a bunch of things around the

operative snip. They are showing it, including proxy snips, and there is one way out here that is in linkage to this equilibrium. For progression, this whole region is showing a big large signal.

So, the pilot conclusions, and I won't show you all the haplotypes, is the deoperative snips are detected pretty well. The proxy snips work about 80% of the time. And, the operative snip haplotypes actually work even if you removed the operative snip, they work 80% of the time. We do need better methods to quench the false-positive signal, but the prospects are pretty good for a full genome scan.

Final point – the signatures of recent selection in the human genome – there have been a bunch of papers out there looking for selection like delta 32 based upon different methods. Heterozygosity is something you expect to go down when a gene is selected because the genes will go to homozygosity and the markers around it will become homozygous.

FST – if you actually have two ethnic groups, say Asians versus Europeans, you would expect large differences if the gene were selected in Asia but not in Europe, and that would be illustrated by a big difference in the same snip in these two populations.

Linkages to equilibrium – take a look at the snip with respect to its adjacent markers. How long is the chunk? Remember CCR5 was about a centi-morgan long.

Then DSDN – remember the ratio of synonymous versus non-synonymous amino acid substitutions among polymorphic markers within coding regions. These are normal signals that people look at.

Up until this year there had been about a half-dozen really good studies that had found a number of regions across the human genome based on genathion, micro-satellites or snips or pearlagen. These are different big data sets, the hap map data set, that had indicated this many regions – the green – that were actually signatures of selection. Of those, the answer was when you ask how many of the Hutley ones were represented by Carlson or Nielson, the answer was not very many. There wasn't a lot of agreement.

. . . lexic and our lab basically did it with a multi-point analysis, a moving window where he looked at five to 70 snips in a window and ratcheted up each chromosome for multiple things, and was able to increase replication of all these studies so that about 10-15% were replicated. I'm telling you this because one of the regions that we selected for a control for the . . . and ARG rank was one of these regions on chromosome 16 – a big region with a lot of LD with one big gene, a glutamate receptor gene called GRIN 2A.

Why am I telling you this? When we put it down on ARG array, we discovered not a background signal, but a real strong positive signal. Here is progression and down here in a region and here in infection. So, basically this was a negative control region originally selected for a signature selection and it was showing an effect as if it might have been selected by an ancient disease waiting for HIV to come along, just like delta 32. When you compare with chromosome 16 array to the chromosome 7 control region, you can see that this one looks colorful, this one does not, this is what we're looking for.

When we look at ARG rank, the same five rankings for chromosome 16 region, we find again affirmation with multiple snips that are associated in this region together, there also in progression, and there also in progression – two regions of the region.

So, what do we know about this gene? Not a lot. It is an NMDA receptor, a class of glutamate receptors. It is thought to underlie certain kinds of memory and it has an interaction with HIV tat and this is basically the final point. I'm almost done. Once the HIV genome enters the cytoplasm, it traverses micro-tubulus to get to the cellular nucleus.

So, can we scale up a snip hap . . . whole genome? I think we can. If we take this kind of an array, it actually looks like this. It is about 14 inches long. Here is another one that is 14 inches long. That is about 800 kb. If we scale up 2,000 kb, it is about a meter. So, a chromosome of the same thing would be 75 meters long, which is up and down this room three or four times. A whole genome would be 1,400 meters or about three-quarters of a mile. So, a genomics ARG array video highway, we could drive down it for three quarters of a mile and see all these signals.

Conclusions – argus can translate to patients. Entry and integration inhibitors are under development which have been stimulated by these discoveries. Explain fraction for the epidemic is small, but growing. Whole genome scans with available cohorts are feasible . . .

**(Tape 13)**

. . . these are folks from my laboratory, principal investigators. These are the cohort directors, the Chinese collaborators -- . . . . . three of them are here. This is the folks in the laboratory – the principal investigators who have driven it, and this is the rest of the laboratory who does most of the work at our annual retreat.

The genome scan is a collaboration between LGD and the Brode Institute and is listed right here. We are looking forward to basically revealing all the age restriction genes we can find pretty soon.

Let me end with a quotation by Patty Stone. . . and Richard Luger who said in the 1960's – we launched the Apollo program to put a man on the moon; in the 1990's, we came together to map the human genome. In the decades ahead, why shouldn't we demand a similarly urgent effort, this time an international one to stop this deadly scourge. I kind of agree.

Thank you for your attention.

Moderator – Thank you very much, Steve. Our next speaker will be Li Taisheng. Professor Li is at the Peking Union Hospital and will speak about clinical outcomes in advanced Chinese AIDS patients.

**Li Taisheng, Peking Union Hospital**

***Clinical outcomes and Immune reconstitution in advanced Chinese AIDS patients undergoing  
12 months of highly active antiretroviral therapy***

Thank you Chairmen. Good afternoon. Thank for the organizers to give me this opportunity to give my presentation. My talk is clinical outcome and immune reconstitution . . . Chinese AIDS patients after one year after HAART.

**(not transcribed)**

Moderator – Thank you, Dr. Li. Our final speaker in this session is Lance Gable. Mr. Gable is a Senior Fellow at the Center for Law and Public's Health at Georgetown University speaking on avian influenza and the risk of a pandemic.

**Lance Gable, Center for Law and the Public's Health,  
Georgetown University**

***Avian Influenza: Preparing for and Responding to a Potential Human Pandemic***

Thank you. I'm happy to see so many people have stuck around to the end of the day. It has been a great honor to attend this meeting. I've learned so much from all of you. I want to thank the meeting hosts and sponsors very much for inviting me to attend. I want to thank everyone who has been a part of this because it really demonstrates some amazingly innovative work that is being done in genomics.

I am not a scientist, so it sets me apart a little bit from the other speakers. My presentation will have no graphs. It will have almost no numbers. It will be mostly words and a few pictures. But, what I want to address is, using the example of avian flu and in talking about preparedness for and responding to a potential human pandemic, whether it comes from H5N1 or some other strain of the virus, I want to talk about how law, and to some extent, policy can inform the decisions that are made to respond and prepare for a pandemic. I want to also talk about how some of the scientific discoveries that we've been talking about for the past two days and how it will affect how those discoveries are used in a practical response effort.

So, I want to start off just talking a little bit about what I mean by a public health law. It is a fairly broad definition which I'll explain in a second. Then I'm going to talk about a number of different approaches that have been proposed to intervene in a pandemic, and talk about how some of these approaches are either effected by or can be facilitated by well-developed laws.

So, first, what is public health law? I'm going to define public health law as a two-part definition. The first is the legal powers and duties of government used primarily to assure the conditions for people to be healthy. So, for example, any type of law that assists the government or the people in the private sector to identify, prevent, and ameliorate risks of health in human populations.

There is a second component to public health law which is that it is not only laws that enable public health to occur, but this concept can include laws that limit the ability to conduct public health-related activities. So, there can be structural components that are put into place by law such as governmental structures. A good example of this in the United States, some public health powers are at the national government level and some are at the state government level, and I'll talk a little bit more about this in a few minutes.

Then there are also some laws that affect what individuals can and cannot do, and in some cases, might have privacy protections or something like that which would affect how public health laws can be implemented.

So, this two-part definition is actually very broad. It encompasses a whole range of different types of potential issues. So, what I'm talking about here is not only laws and policies that directly authorize the government, for example, to go and conduct disease surveillance or to take information and share it with other nations with regard to a pandemic outbreak. Or, for example, to impose a quarantine on people. It is also about other related types of provisions that might also impact the implementation of those types of powers.

I also want to point out that this idea of law spans many levels and obviously each country has their own set of laws that affect how they might be able to respond in a pandemic-type situation. There is also international law which applies across multiple countries, although international law is much less enforceable and a little harder to get a grasp on. Laws also exist at multiple levels within countries, in many cases.

I'm not going to actually spend too much time talking about pandemic influenza from a scientific perspective because that has been very well done already by some of the previous presenters. Dr. Monto and others have already given a great description of some of the risks and some of the potential problems that can be posed if an influenza pandemic, whether it is H5N1 or some other variant, gets into the human population and becomes transmissible human-to-human.

From a policy perspective, and from a perspective of how laws might affect this situation, there are multiple public health challenges that can arise during a pandemic. Some of them will be scientific, as we have been talking about for the past two days, but there are legal challenges. Are there appropriate powers, for example, to conduct all the types of activities, all of the interventions that you might want to during a pandemic. There are ethical considerations. There are political considerations, of course. And, there are financial considerations and many others.

So, in all of these types of interventions which I'm going to talk about in a minute, laws are an important tool, as are policy decisions and how the law is implemented. There may be some ethical considerations as well.

This is a map from the WHO which just shows that as of about a week ago, where the human

cases of H5N1 had been since 2003. You see most of them are concentrated in southeast Asia, but there are increasingly human cases spreading to other parts of the world. We know the infection in bird populations has continued to spread around the world as well due to some of these bird flyways that overlap across the globe.

So, with that as the background, I want to turn now and identify a couple of issues that I think are worth talking about and to talk about how medical countermeasures and public health interventions that can be used during a pandemic or to prepare for a pandemic can be thought about in the context of how law might affect these types of interventions.

I believe Dr. Monto referred to these similar types of interventions as therapeutic and non-therapeutic. I'm using the terminology of medical versus public health. But, it's the same essential thing. The medical countermeasures are things like vaccines, antiviral medications that can be used to treat patients or potentially as prophylaxis to prevent infections.

The public health interventions, or non-therapeutic interventions, are other techniques and strategies to try to track the disease, to try to restrict the spread between people without actually using any kind of medication.

Of these different interventions, each of these potentially has scientific, political and legal components. I'm going to focus on the legal components. So, first, with either type of medical countermeasure, whether we are talking about vaccines or antivirals, there are several challenges that need to be overcome to make sure that even if you pass the first test which is to have the good science in place, there is going to be time lag in developing a new vaccine for a novel strain. But, even once you've succeeded in developing your new countermeasures that you're going to try to distribute to the population, there are three other challenges that present themselves. You need to have an adequate supply, first of all, to treat the population or to circulate to the population. You need to be able to distribute it to people all across the world potentially. And, there are also issues related to ethically appropriate and allocation of that medication to make sure that the decisions being made are appropriate and who gets the

medication. Undoubtedly, there will not be enough for everyone. Even the most optimistic scenarios, we don't have the production capacity if it truly is a worldwide pandemic.

I want to talk for a second about funding for these types of medical countermeasures. At least in the United States, there was a proposal that President Bush put forward that calls for \$7.0 billion in funding for pandemic influenza preparedness. Of this money, about six billion of that is for medical countermeasures. That is divided \$4.7 billion for cell-based vaccine technology, and stockpiling experimental vaccines, \$1.4 billion for antiviral medication. So far, Congress has allocated \$3.8 billion – about \$3.4 billion of that is towards medical countermeasures.

Now, one of the things that strikes me as interesting about this decision-making and this planning is that the vast majority of the money in this proposal goes towards medical countermeasures. Very little of it goes towards any of the other types of interventions – such as strengthening surveillance, for example, trying to detect the disease at an early stage. It is an interesting decision. It is very reliant on the medical components and not really as much money is going towards the public health infrastructure.

Outside the United States, there is a great deal of variability in terms of what countries are doing to prepare. The funding obviously varies quite a lot, depending on the resources of the country. There have been some international efforts, actually, to try to raise money that the World Health Organization and other international bodies can use to try to have a global approach to preparedness for a pandemic. Just a few months ago, in January 2006, here in Beijing a whole group of countries came up with the Beijing Declaration which calls for not only a great deal of cooperation between countries in preparing for a pandemic, but also receive commitments from the attending countries of \$1.9 billion to towards these efforts.

Another important point about the Beijing Declaration is that it explicitly notes that the World Health Organization and the International Health Regulations, which are an international set of regulations that require countries to report certain information to the WHO and to work with the WHO for certain types of diseases, and the Beijing Declaration actually explicitly says the countries should be using this mechanism to report.

One of the main issues in terms of vaccine supply, and this is where some of the legal concerns can come into play. Vaccine supply has been inconsistent for seasonal flu and since there isn't capacity for seasonal flu, there certainly isn't capacity for a potential pandemic. In the United States, in 1967, there were 26 vaccine manufacturers that were licensed in the United States. This year, there were only four. That is not the vaccine manufacturers who make flu vaccines. That is all vaccines – only four in the United States. Worldwide, there are many less than there used to be as well.

So, you might ask the question well why are companies not making vaccines? There is a couple of theories about why this is the case. They are listed on the slide. One is the market forces are a disincentive to creating vaccines. Companies can't money on vaccines, so they are making other things instead – pharmaceutical companies are trying to go towards other products that may be more financially lucrative because vaccine purchases are inconsistent.

The other three suggestions on this slide here are actually directly related to the legal structure. There is the idea that regulatory compliance with making vaccines is a very high cost on the manufacturers. So, they are less inclined to make them. Every country has a regulatory system that regulates the quality of medications like vaccines. In the United States, it's the food and drug administration and the regulatory process can take a long time. It can actually end up deterring some companies. So, there have been some proposals to change the regulatory laws either to make them less stringent, especially in an emergency situation like a pandemic, or there have also been proposals to try to streamline some of the regulations across different countries. If you're going to be distributing this vaccine around the world, you have to meet the regulatory standards in all the countries.

Of course the regulatory standards are there for a reason – to make sure the vaccines are safe and they are effective. So, there obviously can't be too much of a dilution of these requirements. But, there have been a number of legal proposals out there that have suggested doing just that.

Another two legal issues, which are the bottom two on that previous slide but are talked about in more detail on this slide, are the ideas of liability protection and patent protection. So, liability protection – the concept of liability is a legal concept that basically says that if you are injured by someone, you have some kind of right to go after them to file a lawsuit if you are injured. This happens a lot in the healthcare system in many countries around the world and with respect to vaccines, there is a fear by a lot of vaccine manufacturers that especially if they are making a novel vaccine that hasn't been tested very much.

. . . would be sued by the injured parties. So, there have been some legal proposals, both in the United States and elsewhere, to prevent lawsuits against vaccine manufacturers who are manufacturing vaccines for pandemic flu.

From an ethical perspective and from the perspective of fairness, there is also the other side of the argument which is that people who are injured by these vaccines, should they have some ability to receive compensation for their injuries. In the United States there is a system for certain vaccines that allows just that. You aren't permitted to sue the vaccine manufacturer, but you are permitted to go to what is called the vaccine injury compensation program and make a claim that you were injured as a direct result of exposure to the vaccine and receive some kind of compensation.

So far, an avian influenza, vaccine is not on that list of vaccines covered under that program, but it is an interesting idea and it is a way that the legal system can be used to not only encourage companies to manufacture needed medications, but also to make sure that people who are injured might actually have some way to be compensated for their injuries.

The second issue here, patent protections – this is a very complex issue that I'm not going to get into too much detail today, but every country, or most countries, have patent laws that give exclusive rights to companies who create a certain product to market that product exclusively. It is meant to encourage innovation. Most vaccines are under some sort of patent and the worry is that vaccines, or other interventions, like flu, for example, the patent is held by Roche. Roche has been very reluctant to allow anyone else to use its patent to increase capacity. Roche itself

does not have the capacity to make enough tama flu for the entire world. Actually, in the China Daily yesterday, there was an article about the fact that Roche did license tama flu to another firm in China. I think that brings internationally three other firms outside of Roche themselves who are permitted to make tama flu. But, under international trade law, there actually is a way that if a country decides it needs a medication urgently that is under a patent held by a private country, the country can do what is called compulsory licensing. They can say we are going to make this medication because it is an emergency for our population. And, it is legitimate to break that patent in that context. The limitation on that is that countries that do not have the capacity to make their own medications cannot buy it from third party countries. So, if the scientific and the technological capacity isn't there to make the vaccines, that doesn't make a difference.

I want to talk just for a second about allocation and I'm not going to dwell on this at all. But, there are a lot of considerations about how – and this can be addressed by law or just by policy. But, questions of how you decide who gets vaccines or tama flu. If you have a limited amount of medication and you have a population that greatly exceeds this amount, how do you decide who in the population that gets the supplies. There are many approaches that you could use. You could, for example, use a public health approach which would say that you're just going to go after areas that are already affected and try and do something like ring vaccination or a targeted approach to areas that are directly affected.

There could be an approach where you would favor people who are in the scientific or healthcare communities or perhaps in the government and other critical infrastructure. Then there is also other considerations as well.

I know my time is short and I'm not going to talk about this now, but I would be happy to talk about this in the question and answer session if anyone is interested.

I want to highlight two more issues that relate to some of the non-therapeutic interventions. The first is the issue of surveillance and how countries can conduct the type of disease detection to

try to identify whether there has been human-to-human transmission of influenza that could cause a pandemic.

Surveillance is obviously one of the key tools to make sure that there is the ability to quickly respond and given the nature of how flu spreads, it is vital to have this early detection in order to have any chance at stopping a large outbreak.

So, two legal issues that arise with regard to surveillance is first, who has the authority to conduct surveillance. In most countries, there may be laws that allow the national government to collect data such as this. In the United States, it is actually a little bit different. The CDC has some authority to collect data, but usually the U.S. CDC gets its information from state health departments that collect the data. Also, universities and private companies are sometimes involved through contracts in collecting some of the surveillance data and there are many reporting laws as well which require certain diseases to be reportable to the government.

Currently in the United States, H5N1 influenza is not a reportable disease. So, even if – that is not to say it wouldn't be reported, but there is no law requiring it to be reported to the government.

Now, once this information is in the hands of the government, who has access to it? That is another important question that is usually addressed under the law. The law, in some cases, will put limits on who can actually access this information. There has recently been some debate about the influenza database that is being put together by the World Health Organization in that there is limited access and there are others who wanted to gain access to this data. This wasn't based on any kind of law. It was just the policy of the WHO. But, in the United States and in the EEU, there are data privacy laws that limit, to some extent, what personal information about health can be transmitted. Of course, both of these data protection statutes include exceptions for disease surveillance and for public health practice.

We are going to skip a couple of these. But, I want to spend the remainder of my time on issues of isolation and quarantine because these are the issues that are probably the most contentious

legal issues. The vast majority of countries have laws on their books that authorize the government to isolate or quarantine individuals who become sick in certain circumstances. Sometimes these powers to isolate and quarantine are very broad. Sometimes they are not as broad, but I guess the first point I want to make about this is that these two terms are often used interchangeably and they shouldn't be. Isolation is when a person who already has indications of an illness is separated from the general population so that they don't spread that on this further. Quarantine are people who have been exposed, but have not yet demonstrated any kind of detectable symptom and the idea behind quarantine is to keep that person separate from everyone until you can determine whether or not they are going to become sick.

There is also a third concept which is a group quarantine idea. It is sometimes referred to as . . . . . and it is the idea that you would block off a whole section of a neighborhood or something like that to prevent people in that area from moving out and transmitting the disease to others.

Now, from a scientific perspective, there is a great deal of skepticism that quarantine will be effective at all in a flu pandemic, based upon the way that influenza spreads. So, the chances that these types of public health interventions will be effective has been questioned by many who know the ediology of influenza. However, it is very likely that many countries will try to use these powers, once there is an outbreak. The question from a legal perspective is who has these powers, when can they be exercised, are there any criteria that talk about how they could be exercised. So, it is just completely at the discretion of the President or the Prime Minister, or is it completely at the discretion of the leading health officer of the country. Or, are there some criteria that need to be met before people can be put under quarantine.

This is an issue at least in the United States, where the federal versus state government comes into play. Most quarantine power in the United States is at the state level. State health officers or the governors of the states actually have the power to authorize quarantine. It is sometimes delegated to officials in local health departments. But, the federal government only has quarantine power in very specific situations. People coming into the country at the borders, they can be quarantined by the federal government. People who are crossing state-to-state can be quarantined potentially by the federal government. But, people who are just in one place within

a state, the federal government does not have any kind of direct quarantine power. That has to be done at the state level. This is important because a lot of times there are plans proposed in the United States and elsewhere where political leaders will say, this is how we are going to conduct a quarantine. But, the reality is they might not have the legal power to do that, not to mention the fact that with any kind of restrictive measures like a quarantine, you have to take into consideration other factors about how can you do this successfully and how can you actually get people to comply. There are issues of safety and hygiene. There are issues of making sure the people who are put under quarantine have adequate medical care and necessities for life. If there is no food or water, people are not going to stay in quarantine. They are going to leave their houses or go to their jobs if they feel like they need the money to survive. That is an issue that most laws do not address and most policies do not address, which is how practically speaking can you implement this type of restrictive public health measure in a way that can allow it to be effective.

The final issue is just an issue about enforcement. Who enforces the quarantine. The people are not saying – if people are quarantined in their homes and they are staying in their homes, who actually is authorized to do that. Is it the military? Is it a police force? Is it some other entity? These are issues that all should be clarified ahead of time so there is not confusion if these powers do need to be used.

One final issue that I'm going to mention – there are other non-therapeutic interventions that could be used that may be more successful in a quarantine, but it is the flip side of the quarantine, which is to reduce opportunities for people to interact with each other. So, closing public places, canceling events, canceling schools for a time period potentially to delay the onset of the disease in whole sections of the population because they won't be interacting as much.

But, again, there are many legal questions that can be involved here. Who has the authority to close the schools? Who has the authority to make these decisions? In different countries, it is going to be different. Many times it will be the central government, but in other cases, it won't be. In the United States, for example, the schools are all done at the state level or sometimes even at the local level. So, just because one local school decides to close doesn't mean that the

others in that same state will necessarily close. So, there are a lot of things that are up in the air because of the variability of law.

I guess one of the things that I always try to tell people when I'm talking about these issues is that knowing the law ahead of time is important, and especially if the law is not set up in a way that allows for an effective response, thinking about maybe changing the law ahead of time is better. Laws that are changed in the midst of a crisis are usually not good laws.

There are also many issues that are related to international travel and border control. I know my time is running short, so I think I'll reserve this conversation for the question session if anyone wants to talk about some of the issues related to what countries can do to prevent people from entering into the country who may have the disease.

So, just to sum up, a well-developed legal structure can help with responding to a pandemic and these issues are important to think about ahead of time rather than after the fact.

Many of the scientific and logistical challenges that we have been talking about for the past few days, not only with avian influenza, but with any of the other diseases that we're talking about trying to work towards eradicating or reducing in the population, many of these also have legal components and it is good to keep that in mind.

The laws should be flexible enough to adapt to new threats. They shouldn't be rigid. Many of the old quarantine laws, for example, are directed at particular diseases. So, it will say you can quarantine for small pox, and that is not going to do anyone any good if there is any number of other diseases.

Finally, the laws should be as clear as possible to clearly authorize who has the power to engage in the different activities and what limits there are on those powers.

With that, I just want to thank you all again. If you have any questions, please send me an e-mail. This is my e-mail address and the website for the Law and the Public's Health. I also want

to thank two of my colleagues, Lauren . . . and Benjamin Berkman who helped with some of these ideas in this presentation. Thank you very much.

Moderator – Thank you very much for that. I think we have plenty of time for questions from the audience for all four of our speakers who I believe are here.

Question – This is a question for Dr. Li. Can you give us insight on what proportion of the HIV-infected patients are being treated with HAART therapy currently in China, and what are the restrictions that exist for broader use of HAART therapy?

Taisheng – The question is to the restrictions for the HIV in China? Please repeat your question.

Question – For patients that are infected with HIV, how commonly is HAART therapy being used in China and what are the restrictions to broad use of HAART therapy?

Taisheng – (can't understand)

Question – Can you give us the percentage – the number is just a . . . total number. What is the percentage of HIV positive persons?

Taisheng – About 20% of patients treated by HAART. . . . .

Question – I have one for Dr. O'Brien. As you expand your mapping to other regions of the genome, do you have plans to concentrate on certain regions like perhaps the “pharmaceutically attractable genome” to aid the pharmaceutical companies to get to the product faster?

O'Brien – Our major goal is not to get the pharmaceutical companies to the product faster. I guess our major goal is to discover as many verifiable gene regions that have translations of various sorts. We certainly are looking at the collection of pharmaceutically relevant genes involved in metabolism in our HAART population, and we are recruiting additional cohorts that

we're excited about, like the Vancouver cohort that has been sitting there for many years and has a thousand patients on treatment now.

So, the short-term goal is to really mount a full genome scan across the genome to get at the genes that we cannot guess at. Those are candidate genes – the ones that you described and we have put those down and they are in the queue to go on right now. But, I'd like to actually take advantage of the bioinformatics tools that have been developed now to scroll through the entire genome. So, that is what our major goal is, and hopefully it will include this group of genes as looked at from various ways.

Question – Steve, that raises an interesting point. The studies you and others have done are retrospective studies, and even under the best scenario where you identify another 20 ARGs that can account for 30% of the fraction of cases, how would you design a prospective study which is ultimately what you'd be interested in if you wanted to use these snips for any diagnostic purposes or to convert this to a prospective study would take an enormous period of time – much more time than any of us would want to devote to it.

O'Brien – I'm not sure I understand the question.

Question – In a retrospective study, you're identifying a series of snips that can explain why certain groups of people do better or worse or what have you. Before one would actually use that for any medical purpose, you'd want to do a prospective study.

O'Brien – If you're interested in diagnostics, yes, or in prognosis. The retrospective study can identify operative variance that quantitatively effects things in ways that we can then say that is useful for understanding the progression of the disease and gives us another avenue to develop therapies against, not by just simply using the gene, but by basically intervention – an intervention of some physiological phenomenon. That is one of the goals.

With respect to using the genotypes as useful for clinical trials, like for a new drug or for a prospective prediction of a patient at risk for something like that, it requires a pretty large

fraction of the explanation of the outcome to be a genetic. For example, 100% of cystic fibrosis is explained by mutations in the cystic fibrosis gene. So, finding that out is a very good prospective thing for prediction and counseling and things like that. But, if we're only going to explain 30% of the variance of AIDS with all the genes that are in the genome, then we are never going to be able to predict what is going to happen by just the genetics. We're going to have to know the other things that are non-genetic as well. So, I'm not so concerned about hitting that 100% target because I actually don't think we'll ever get there. As it is with any complex multi-factorial disease, where there is a component that is genetic and a component that is not genetic, just categorizing them and understanding them – that information can then be translated to therapy development, perhaps to understanding the heterogeneity indexing or drug trials, but not always. In each case, they would be different. So, I'm not optimistic that we will get to the prospective stage at all in this particular disease because I don't think it is all genetic.

Question – I have a follow-up question to that. Given that these ARGs represent interactions with both the virus and drugs, have you tried building any of those issues into your model, your explained fraction model, maybe looking at the genotype of the virus at the same time that you're looking at the ARG snips?

O'Brien – We thought about that a lot, but we haven't actually figured out how to do it yet. The reason is that most of the virus genetic associations that we see with disease are almost anecdotal. That is, occasionally there will be a mutation of the neph gene or something that will cause it. But, even with the major clades, the argument that one is more or less virulent than the other is controversial at best. So, because we're talking about a virus here that has an immense opportunity for variance in the sense that it produces something on the order of 10 billion copies a day after a few weeks, and every second virus has a new single mutation and a cross of 9,000 base pairs or so. So it is a swarm, if you will, of billions of different variants that are produced in every patient every day. It is a recipe for resistance of just about anything that we want to get at.

So, trying to capture that as part of the variation, as a lot of people have tried it, and it has not been very helpful because of the fact that there is so much going on that we don't know how to interpret it very well – too much.

Question – I have a similar question. I want to ask, based on the other results you got, did you see any difference compared with the normal HIV patient with HIV carrier but long-lasting and no symptom patient? The reason I'm asking about this is because I think if we can better understand this issue, then maybe we can better to develop a drug and vaccine.

O'Brien – That is a very good question. We have paid very close attention to the long-term, non-progressors – those lucky patients who comprise about 1% of the infected population that avoid CD4 depletion for 20 years or longer. These patients have an elevated frequency of many of the protective AIDS restriction genes such as delta 32 and some of the HLA, B27, and B57 and some of the other AIDS restriction genes. But, as I mentioned, although the attributable risk of this group is pretty high, on the order of 20%, we can explain by their genotype, unfortunately the total explanation for why some people get long-term survival and other people progress rapidly like within six months. We have two patients, for example, who are homozygous for HLA B35 which is a bad thing to have, and both of them progressed to AIDS in six months. The reason that they do that is the B35 produces an epitope that monopolizes cytotoxic t-cell activity to the extent that the immune system is almost circled by a decoy, if you will. These patients progress to AIDS, as I said, in six months or less when they have that particular genotype. So, the answer is yes – the long-term progressors and the rapid progressors are very important parts of this continuum of survival that we see, and they are part of the analysis that allowed us to detect and quantify each one of these genes.

Do we know how much of the long term survivors we can explain by their genotype of the genes we know about? We're estimating it is less than 10%.

Question – This is question for Dr. O'Brien. There are three ways for HIV infection – blood, sexual and mother to children. If the people carried homozygous CTR5 delta 32, . . . prevention of HIV infection by three ways.

O'Brien – I did not understand the question – please say it again.

Question – My question is, if the people carried homozygous CTR5 delta 32, these can prevent . . . HIV infection in three ways – blood transmission, sexual transmission, and from the mother to kid transmission. Just . . . one way . . .

O'Brien – People that are delta 32 homozygous don't become infected regardless of how they are exposed. Does that answer your question? It doesn't matter whether it is a blood transmission or whether it is a sexual transmission or a contaminated blood factor or how it is. The only way that a homozygous delta 32 people become infected, and we have about five . . .

**(Tape 14)**

. . . original type specimen for HIV. They had psoriasis and sores on their hands. They concentrated the virus. They became infected with HTLD 3B which is HIV X4 tropic virus became infected. Their CD4 cells were depleted and they developed AIDS. This actually was probably the best example for why HIV fulfills cokes postulates because it showed that it transmitted and caused the disease. That is important because there is a conspiracy that HIV doesn't cause AIDS – that it continues to surface every year . . . .

Question – I have a question for Lance. Could you evaluate, on a scale from maybe 1 to 10, how prepared you think the U.S. is in the event of a human-human transmission outbreak of H5N1, highlighting specifically the points you mentioned related to quarantine, and general overall response. As you mentioned, the overwhelming amount of funding is going to medical preparedness and how you think that is going to influence the effectiveness of any response.

Gable – I think it is a good question. It is really hard to estimate. It is one of those things that think if I had to say on a 1 to 10, I would say it is certainly right now probably anywhere from three to five. It seems to be getting better because there has been a lot of attention paid to this potential threat in the past few months. But, one thing that I always – at least for the past few years that I have been looking at some of the different legal approaches and policy approaches to preparing for different potential disease threats, I think that when so much attention is paid to a particularly disease, and in this case we're talking about the potential for a flu pandemic, and

also when so much attention and so much of the financing goes to specific medical interventions, or some kind of specific approach to target just one disease, I think that is not as effective as when you put the money toward more broad-based, structural preparedness types of initiatives. So, for example, building a really robust public health infrastructure that will be able to respond both to a flu pandemic and all of these other many types of disease threats that are out there, I think is probably the way that we will become the most prepared.

I think that at least with the focus on the medical interventions, they can be very effective and very important. The problem is, at least with one component of that, the vaccines, there is going to be a considerable lag time. If we can manage to make the breakthroughs necessary to do a cell-based vaccine, maybe that will solve some of that problem. But, I think right now, having the vast majority of the resources going to that is not putting the United States or others that are adopting that approach, in a great position because there is going to be far too many people if there truly is a large-scale outbreak for those medications to actually reach.

Moderator – We have time for one more question.

Question – I have a question for Dr. Li. My question is, I take care of patients who are primarily Caucasian in San Francisco. My question is, have you noticed any difference in the response to HAART in Chinese patients versus say a comparable cohort of Caucasian studies. There have been many studies on individual developed therapy. The second question I have is have you noticed side-effects from HAART therapy and is there any difference, again, between the side effects and potential side effects seen in the Caucasian population?

Taisheng – **(cannot understand)**

Moderator – Please join me in thanking all the speakers one more time. I believe we now have tea, followed by posters.