<u>May 2016_Nussbaum_Genetics Web</u>

Mr. Nussbaum:

Thank you. Good afternoon everybody. So Sue Bellinson [spelled phonetically] asked me to talk about the whole exome, the whole genome testing. My disclosures: I work for a company that offers panel testing. I've consulted in the recent past for companies that offer whole exome and whole genome sequencing, and my disclaimer is I am going to talk about technology and products provided by various commercial entities. I am going to avoid any commercial bias. I'm not mentioning any brand names. And also the other disclaimer is this field is moving so rapidly that some data and conclusions have a short half-life.

So a few definitions [inaudible] the panel [unintelligible] is the sequencing and copy number analysis of a few to a few hundred genes, WES or whole exome sequencing. The sequencing of the two percent of the genome that codes for the approximately 20,000 proteins in the human genome, whereas whole genome sequencing is WGS. That's the sequencing of a [inaudible] DNA and is contained in a huge, you know [unintelligible].

Now just to remind you that the way genes are organized in the human genome is a little bit like calling this the Gettysburg Address. It's maybe not obvious that it's the Gettysburg Address until you suppress all the nonsense verbiage in between the key words here, and all of a sudden the "four score and seven years ago" pops out as being the obvious beginning of the Gettysburg Address. Well, this is very similar to what goes on in our genomes, in that the parts of the gene that do not code for any protein are like the spaces between the words, and what you have to do is remove those spacers and then splice the words together into a sentence for your "four score and seven years ago."

So here's an anatomy of a typical gene with the yellow gray rectangles are the exons and they actually contain the genetic code that is reflected in the sequence of a protein. Between those exons are introns which do not code for proteins. There are regulatory regions in the DNA between genes. About two percent of the human genome is contained within exons; 98 percent is not, and about 20 percent of the human genome is made up of exons; the introns between them and the regulatory regions. So 80 percent is in the DNA between genes.

When a gene is actually going to be read, in other words expressed, an enzyme RNA polymerase runs through the DNA sequence and makes an exact copy in RNA, and then all the individual exons need to be removed and spliced together so that you have one contiguous run of bases that actually code for a protein, and the introns in between have been removed.

The other acronym or definition I want to go into before I get into the meat of the talk had to do with what next-generation sequencing is or NGS. Another term for it is massively parallel sequencing and, essentially, what next-generation sequencing does is it attempts to sequence many, many millions of spatially separated DNA fragments, spatially separated in some sort of a chip or in bubbles or some other way of physically or spatially separating the fragments, and to sequence them simultaneously within some kind of a flow cell or other sort of reaction chamber for which the reagents can be delivered and removed through different cycles. So the idea is you can sequence many different fragments of DNA all at the same time, and that's why next-

generation sequencing allows you to generate such huge amounts of sequence, as compared to the previous methods of sequencing, which were essentially one molecule at a time.

There are different methods with different read lengths, accuracy, and throughput.   I don't want to get into any of those technical details, but suffice it to say this is the fundamental basis for next-generation sequencing; massively parallel.

Technologies are changing.   Panels are in transmission -- are in transition -- I'm sorry, from multiplex PCR where you amplify individual fragments and then Sanger sequence each of them. That is now transitioning in panels over to the greater use of next-generation sequencing, whereas for whole exome and whole genome sequence they all use next-generation sequencing at this point.   You really can't tackle a whole exome or whole genome without next-generation sequencing.

Panels by next-generation sequence rely on what we call hybrid capture.   And hybridization is when two single strands of DNA, base pair, A's with T's and G's with C's, can form a match to each other.   So panels rely on hybrid capture of the particular exons of the particular genes in the panel that you're trying to test, whereas whole exome sequencing also relies on hybrid capture, but there the capture is of as many exons as possible.   You try to pull down every exon in the -- in all the genes in the human DNA.

Whole genome sequencing does not rely on hybrid capture at all. It's just -- you just sequence all the DNA. Now it's possible to generate a whole exome sequence to extract the information bioinformatically or by software without doing a hybrid capture, which is essentially a biochemical or chemical step in the process.   But most whole exome sequencing done today is done by having a biochemical step that pulls down the two percent of DNA that's contained within exons and then subjecting that to next-generation sequencing.

Now let's talk about whole genome sequencing.   The "Holy Grail" of whole genome sequencing would be what we call De Novo assembly, which means the following: you start -- let's say you have a gene with four exons and there's all the DNA, and you cut the DNA up into fragments and shown here graphically are some fragments in the orange which are derived from the exons of this gene, and they're mixed in with millions and millions and millions of other fragments. You then, essentially, apply next-generation sequencing to generate the sequence of these fragments, and you can generate usually somewhere up to about 300 base pairs. So you try to create fragments that are about 300 base pairs in size, give or take a few, so that when you sequence in from each end you meet in the middle, and you end up with a representation of each fragment.   You then -- what you would ideally like to do is to be able to take all those fragments, run them through a computer, have the computer align them by having them match up at the base level, and create a long, contiguous run of sequence made up of all these individual fragments that have been aligned because they match.   They match their sequence.

The problem is that, although this is ideally what you'd like to be able to do and, therefore, be able to identify an exon, let's say; exon three shown here after you've assembled all of these little fragments into a single, contiguous sequence.   The problem is that there are types of DNA sequence which make this very, very difficult to do.

First of all, they are short read length. They're only 300 base pairs. So if you have, let's say, a deletion on one chromosome that's not present on the other chromosome, this will confuse your ability to align these things up because you'll be missing some of the sequence you need to make a contiguous sequence.

The other problem is repetitive DNA. There are millions of copies of types of repetitive DNA sequence which are the same sequence distributed throughout the genome. So when you do a 300 base pair sequence, and some of that at one end, let's say, or most of it is from a repeat DNA sequence. You have no idea which repeat of the many millions in the genome that actual fragment came from. So what that basically does is it blocks your ability to form a contiguous sequence because you can align it, but it will align to millions of different sequences. So you can't get a single, unambiguous alignment.

So instead what we actually do is called re -- it's basically resequencing. So, for example, suppose you have two genes; here gene one, gene two, each with four exons; extract the DNA and then here are these fragments of which some are fragments representing sequence of exon one through four of gene one, and other fragments are from gene two. Now beautiful sequence but instead of trying to align them to themselves, what you do is you align them to what's called a reference. And the reference sequence is something that was generated. It's kind of a consensus sequence. It's been generated -- it was the first "human genome sequence" that was generated 13 years ago and, of course, it's been worked on. It's been improved, but it still represents a consensus or a majority sequence. It doesn't actually reflect any one individual's sequence. And I call it a reference. It's not normal in the sense that there's normal sequence and non-normal. Everyone's DNA sequence is different on the planet, except for identical twins. So there are actually 7 billion reference sequences in the genome. But what we try to do is form a consensus, and use that then to align each of these little 300 base pairs against the reference; not against each other but we map them to the reference, and if a fragment doesn't map because it contains a repeat or for some other reason, you simply discard that fragment. And in that way you get around the problem of deletions and duplications. You get around the problem of repetitive DNA by aligning everything to a reference, and then you can see what the sequence from this DNA is that the aligns to exon three of gene two, and you can compare the sequence and see; is it the same or does it differ at any place along here from the reference? So that is in essence what we're doing with whole genome sequence; is we're aligning it to a reference and then asking does it differ anywhere from the reference?

Whole exome sequencing is the -- includes not just the exon, the coding region, but it also includes approximately 20 base pairs to either side of the exon so that you can see the parts of the intron that are intimately involved in splicing the exons. There are specific sequences that the cell uses to know that these are the ends of an intron that need to be spliced out and if those are mutated or altered, it will alter splicing and, therefore, change the messenger RNA and, therefore, change the protein made from that gene. So when you do whole exome sequencing you want not just the exons, but also the flanking intron to either side.

So here's an example of a whole exome sequence done by hybrid capture. Once again, we've got these two genes, four exons each, extract the DNA. Here are the fragments with the DNA

sequence but now what you do is you generate a series of short stretches of DNA representing all the exons in the human genome. And we call them oligonucleotides; it just me short pieces of DNA that are 30, 40 base pairs in length. They're complementary to the sequences to be captured, which are called the base and you have them on magnetic beads. You then mix these short oligos that are complementary to all the exons and have their magnetic beads attached to them. You mix them with all of these fragments, and these short oligonucleotides will find their perfect matches, their base matches, with the exons. You then use a magnet to pull them away from the rest of all the DNA that is not represented in your oligonucleotides. You then extract those off the magnetic beads, and you subject that to next-generation sequencing and, once again, map it against the reference so that you can find what these sequences from your DNA interest; whether they differ or not. So, for example, here's one where there's a key in the DNA from this individual, but the reference has a different letter there; a different base. You can do the same for the other exons. So this is an example of how whole exome sequencing works.

Now I want to talk about analytic validity issues. I know that you are very much involved with thinking about how to test the test. How do you know a test is worth doing, and the three main elements of any test are analytic validity, clinical validity, and clinical utility. So let's talk about those three issues in the context of whole exome sequence and whole genome sequencing and, by the way, by comparison, to what we call panel testing. So analytic validity issues: can you see what's there? Can you find what's there, and also, of course, do you not find what's not there? In other words, eliminate false positives.

Let me just say at the outset: the false positives are much less of a problem with next-generation sequencing than are false negatives. There are situations where there can be what appears to be a change seen by whole genome sequence or whole exome sequence that turns out not to be correct; that it's an artifact. But that's fairly rare but it's one of the reasons why, when doing genetic testing by whole exome or whole genome; when you find the variant that you think is of clinical significance people confirm it. They confirm it by going back and using the old method of polymerase chain reaction across that area where the mutation's been found and then doing Sanger sequencing to see whether, indeed, that sequence has been seen -- that was seen is actually present in the DNA.

Now I should mention that this has been going on now for years, and next-generation sequencing has been in place, and it is extremely rare to really have a false positives. This confirmation is kind of -- it's the way people do things in the laboratory. It's fairly standard. However, many laboratories have recognized that it's probably overkill. It's probably unnecessary but, at this point, confirmation is still being done and it still seems to be the accepted approach to dealing with false positives. False negatives, of course, are another issue.

So there are three sources of false negatives; in other words, of you not being able to see something by next-generation sequencing that's there in the DNA. The first are random effects. As I showed you, whole genome sequencing or whole exome sequencing; you're carrying a bunch of fragments, and then you are sequencing them at random, and the problem is that some of these fragments may be -- you may sequence more than others. And, in fact, it covers a fairly obvious normal distribution, a bell shaped curve, and for whole genome sequencing it's fairly typical to sequence enough of the fragments of DNA so that on average every base is actually

sequenced 60 times.

Well, you can see that the number of times that that same base is sequenced; the median or the average is 60 times but there are some parts of the genome where you're sequencing far less than 60 times, and there comes a point where you don't actually have enough sequence from the whole genome to be confident that you're reading the sequence correctly. So this is a source of missing sequence. You just don't have enough reads simply by random effects to be able to see it. If you use whole exome what you can do now is say, "Look, I'm only sequencing the two percent of the genome so I'm going to sequence more of that two percent. I'm going to sequence so that we have 350 fold or 400 fold average sequence." That, of course, means that out here they're still up and around 50X or 60X. You're still getting plenty of coverage but way down in the tail you still have a situation where there may be parts of an exome that you just do not have enough reads from to be able to have a confident call that you know what the sequence is. So there are random effects that affect your ability to read every base that you want to read.

There are also non-random effects, and this is -- remember that when you sequence you sequence fragments and then you align it to a reference. Well, there are some parts of the reference that are very hard to align to. If there are insertions or deletions of various sizes the software that's doing the aligning may have difficulty actually looking at the sequence you generated from your fragment, and being able to place it against the reference because the sequence that you generated is from an individual who has an insertion or a deletion of some sort that's not represented in the reference.

There are ways of getting around that. The tools are getting better and better for dealing with it, but it's still an issue; rinucleotide repeats as seen in myotonic dystrophy or in fragile X or in fredrei setaxia [spelled phonetically]. They are notoriously difficult because the next-generation sequencing has trouble getting through these long repeats. Remember each fragment is only 300 base pairs, and how do you align those to know whether they're actually 400, 800, or 2,000 copies of these repeats there? You just can't do it. And so trinucleotide repeats are well known to be not approachable at this point, completely approachable for next-generation sequencing.

And then finally repetitive sequences or closely similar pseudo-genes. There are parts of the human DNA where a gene -- perhaps it's a vestigial copy. It's one that, through evolution, has suffered a few mutations, and it's sitting there but it's actually not a functioning gene anymore. It has no functional activity. Yet when you sequence it the software may, by mistake, think that it's a sequence from the real gene and try to align it and will find differences with the reference, and will start to call them as mutations in the real gene when, in fact, what has happened is it has sequenced the pseudogene. So these are all things that are non-random effects that make hard -- what I call the hard to see regions of the genome by next-generation sequencing.

There are also non-random effects that come from hybrid capture, and this is specific to the biochemical approach to whole exome sequence where you generate these oligonucleotides, pull down the exomes and sequence them. Exons vary a lot in their base content. They may have lots of cytosines and guanines. They may have few. They may have a lot of adenines and thymines, et cetera.

Different sequences are more or less difficult to biochemically identify with the oligonucleotides and pull them down. So here's a picture of a bunch of -- of a gene and each point here represents an exon, and in between are -- would be the introns. This is the coverage or the capture success in blue for each of these different exons, and you can see it's very high; not as high. It's very low here. It's poor. It gets better, but it tapers off along that exon. It gets better. Here it starts very bad at the beginning of that exon, but improves as you go along. So, as you can see in blue, the capturing efficiency of different exons, and even different portions of different exons can be very different. What's shown here is that you can you can attempt to manipulate this; improve it by putting more oligonucleotides on; things that will do a better job of identifying various parts of the exon, and you can fill in so that you can improve the capture. But even for some places, even if you improve it, you still don't get it up to the level of the other exons. So this is a variable capture. This is also a non-random effect that has to do with sequence specificity.

How do we improve random coverage? Well, improving random coverage costs money. Whole exome sequencing sacrifices completeness of the genome for the sake of greater coverage of the most clinically valid portions of the genome at a given cost. So you might say, "Well, look. If it's a random problem just sequence more." The more you sequence, the more it costs, and so that you -- overcoming the random decrease in coverage for certain areas of the genome simply by increasing the amount of DNA you sequence will definitely work, but it raises the cost significantly each time you run a new flow cell or you do a new sequencing run.

Panels are a way of getting around this. It's even more cost conscious because you generate even higher coverage because you're looking at fewer genes. You're not looking at all the exons of 20,000 genes. You may be looking at all the exons of only 150 genes or 60 genes or 12 genes or four genes, depending on the size of the panel. So you can generate extremely high coverage to get around the random effects if you restrict what you're doing to a panel of a dozen to a few hundred.

Improving non-random coverage also costs money. First of all we have to improve the reference. We have to make sure the reference is accurate. When the National Center for Biotechnology, NCBI, changed -- it's what they call the genome build; in other words, what they changed what they called the reference. All of a sudden, things that were thought to be genes were demoted. Things that were thought to be exons turned out not to be. Other parts that were thought not to be exons became exons, and so these regions of the genome that we're interested in looking at may differ. The reference when the reference changes, and the reference is still continuing to change because of the problem that some people have deletions and duplications in the genome that other people don't have.

Seeing hard to do variants requires gene-specific custom solutions; getting around the pseudo-gene problem; getting around the trinucleotide repeats can be a real challenge and may require you to go outside of next-generation sequencing and actually go back to the original PCR and sequencing, and that is an expensive proposition, and that adds to the cost of doing genetic testing.

Now clinical validity: understanding what specific variants mean for health. This is definitely a work in progress. There are 20,000 genes in the human genome approximately. About 4,500

have been implicated in various diseases. There are millions of different -- there are millions of
different variants.   Some are implicated in disease. Others are not. Most are unknown, and of
the variants that we find -- remember a variant by definition simply means something different
from the reference. It doesn't mean it's disease causing. It doesn't mean anything. It just means
it's different from the reference.

Most of these variants, 85 percent of them, are rare. It may be specific to an ethnic group, a
family, or even an individual, and this allows me now to remind myself to make sure you know
that the reference genome is heavily weighted toward the sequences that have been done in the
population which, in turn, is heavily weighted toward Caucasians of northern European ancestry.
So that the reference genome under represents Africans. It under represents South Asians. It's a
little better for East Asians, but it's mostly European Caucasians, and so it is not unusual at all to
do a sequence for someone who comes from sub-Saharan Africa and find a lot of variants, as
compared to the reference; most of which is because the reference does not reflect the variation
that occurs in that ethnic group.

For whole genome sequence clinical validity is even a bigger problem because understanding
what specific variants in the non-coding portions of the genome is not in its infancy.   It's
embryonic.   We know very little about what the 98 percent of the human genome is non-coding;
what those sequences do. Not that we know nothing.   We've got some beautiful examples of
where we understand what a non-coding region is doing, in terms of human health.   But still, the
vast majority we do not know and, remember, 98 percent of the sequence obtained by genome
sequencing is in this category. So 98 percent of the sequence from a genome sequence -- we
simply don't know what it means, and there are many, many millions of different variants in a
whole genome sequence. Some are implicated in disease. Others are not. Most are unknown.

Now let's turn to clinical utility.   What test is most useful for making a diagnosis or changing
management?   If you do a panel there are certain advantages.   Panels target specific indicated
genes. These are genes that have already been implicated in or been involved in causing a
particular disorder. When you do a panel, in general, it costs you to sequence to great depth and
because panels, essentially, require the laboratory to sweat the details. If there's a hard to do
region in that, if there's a pseudogene of some kind, the laboratory has to figure out a way of
making that work. They can't just finesse it and say, "Oh, well, this is a hard to do what we can't
do next-generation sequencing."

That is not to say that there aren't some laboratories who simply state up front: "Exons 12
through 15 of this gene; we are simply not going to report on in this panel because we can't.
There's a pseudogene or there's some other complication that just makes it too difficult for us to
do it." But they have to say it up front. They have to state: "We don't guarantee certain exons or
certain regions of the genes on our panel." So, in other words, unless otherwise specified, the
sequence of the exons are all delivered.

Deletions and duplications are also delivered unless otherwise specified.   So for a panel a
negative result approaches being a true negative, but only for the genes on the panel. A true
negative means that you can sort of take to the bank that there's nothing there that you need to
worry about. But beware. There are false negatives even in panels because panels don't

sequence the entire gene. They only sequence the exons in the boundaries of the genes of interest, and there can be mutations deep inside an intron which affect the expression of that gene, and that mutation will be missed. So panels are not 100 percent sensitive, but they are more sensitive than exomes.

So why use a panel? Panels are most useful when a clinical evaluation suggests a particular diagnosis for which a set of genes has been identified as being implicated in causing that disease. You would not order a panel when a diagnosis is completely unclear or uncertain, and you're essentially shooting in the dark. A panel would not be appropriate in that situation.

Now how about for whole exome or whole genome sequencing? One of the advantages there: the advantage is that it allows you to discover new genes involved in known disease phenotypes. So maybe a disease is known, but we don't know all the genes that are responsible for it. For example, in thoracic aortic aneurysm we know the genes responsible for maybe 25 to 30, maybe 35 percent of all familial cases, but we don't know the other genes. So this is a way of being able to find them and the opportunity to find new disease phenotypes or solve diagnostic dilemmas is another potential advantage of whole genome or whole exome sequence, which panels cannot do because panels are only interrogating the genes you know about; not the genes you don't know about.

So the application of whole exome and whole genome sequencing; probably its greatest potential application, and I'm not stating this as being a clear indication, but it's something that is being used currently, and seems to have -- seems to show promise. So here is one example. This is a paper from 2015, Genetic Medicine, where they looked at 500 patients, and they found a positive or likely positive result in a characterized gene in 30 percent of the patients. And these were patients who had gone through a diagnostic odyssey; people that attempted to make a diagnosis were not able to do it for a variety of reasons, but whole exome or whole genome, in this case, was mostly all whole exome, was able to find the cause in 30 percent of patients.

A novel gene was found in another 7.5 percent. This was a gene that wasn't suspected before. The highest diagnostic rates were observed among patients with ataxia, hereditary imbalances, congenital anomalies, and epilepsy. Twenty-three percent of the positive findings were within genes characterized within the past two years. This is a striking piece of information. What this tells you is the rate of the accumulation of knowledge of which genes are responsible for particular diseases is continuing to grow so rapidly. This puts panels at a disadvantage because panels cannot keep up in real time with the new discoveries of new genes. You have to identify which genes you're going to study, and then you have to put a lot of time and effort into making that panel robust and consistent, and making sure that you can see everything that's there to be seen, and by that time additional genes responsible for that phenotype may have been identified.

So you're chasing your tail with panels, which is not -- which you're not doing with whole exome or whole genome, and the other interesting thing was -- these are done by doing a trio; doing the patient and both parents, as opposed to just the patient; him or herself, and the reason for that is if you find the change in a gene that's not suspected of being responsible for the phenotype in that patient, how do you implicate that gene? What is the data? What is the evidence? If that gene is not known before, as being responsible for a phenotype, how do you

implicate it? The main way is genetically; by finding a child, let's say, who has severe, congenital anomalies, that neither parent has, and you find a new change in that gene, not inherited from either parent. That's a new mutation and that is responsible for a substantial fraction; at least one fifth to one third of all of the successful diagnoses by whole exome and whole genome is because you've done a trio, and you found a new mutation in the proband [spelled phonetically] in the affected individual, not present in either parent.

So just to summarize differences between panels and exomes versus genomes. Panels are complete on a per gene basis. They essentially promise you'll get all the exons and you'll get all the coding and all the intron boundaries but, remember, not necessarily the entire gene, whereas exomes and genomes are comprehensive on a genome basis. They will pull in lots and lots of genes, but they don't guarantee, because of the random and non-random effects: they do not guarantee they are going to see every exon of every gene.

Panels: it's difficult to stay current, given the pace of new gene discovery, whereas exomes and genomes: it's difficult to guarantee each gene of interest will be covered every time because of these random and non-random effects. Panels are best when the clinical diagnosis is clear or the differential diagnosis is limited, and the genes involved are known or mostly known, whereas exomes and genomes are best when clinical diagnosis is obscure and the genes involved may be largely unknown, such as in undiagnosed diseases.

Panels: rare and rarer disorders add more and more content to the panels, but it's technically and economically demanding to add more and more genes for rare disorders when the testing will be available for fewer and fewer patients because these are rarer and rarer diseases. In other words, it almost becomes an issue of economic -- for the laboratory economic diminishing returns. How much more investment are you going to put in to add the next 1,000 genes responsible for a bunch of very rare disorders for which you may get few, if any, orders over the next few years. So you've put all this investment into making that panel, whereas whole exome or whole genome; you can see most of the genes and most of the parts of the genes without having to put a lot of extra effort into adding these genes for rare disorders. So, from a laboratory economic point of view, there comes a point where rare disorders become economically more and more difficult to justify.

The other issue I want to bring up in the last few minutes has to do with incidental and secondary findings. You know the different incidental finding is something that you find. It's not the indication for the test. They're discovered in the course of a diagnostic test, whereas a secondary finding is a result that is not related to an indication but, nonetheless, should be deliberately sought for, regardless of the indication from that test. And this really came to the fore in the genome and exome arena when the American College of Medical Genetics working group in 2013 reported -- made a recommendation that there are 56 genes and conditions that a clinical laboratory should seek out in the course of whole exome or whole genome, regardless of the indication, because these are conditions that are not diagnosable otherwise, are actionable, that it's important for patients to know about it because they can do something about it, et cetera.

It became quite controversial. It's controversial for a variety of reasons that we know enough about the natural history of these 56 genes, because most of what we know about them we know

from patients who have the disease; not from people who don't have that disease. Penetrants are not always known. Do we really know that we can intervene successfully? There are ethical issues because the recommendation was that these should go back to the clinician to report to the patient whether -- without giving the patient the ability to opt out and that we're actually foisting information on to these patients that maybe they don't necessarily want. So this was quite controversial.

So the ACMG revised these findings that patients be given the choice to opt out before testing takes place. So results they would not wish to receive are not even generated, and the issue just never comes up. And this is now the current standard. It's not observed. It was only recommendations; not a requirement, but most labs do now allow patients to opt in to receive a limited set of secondary findings that are incidental to the indication for the test, but fall into this category of things that you can do something about if only you knew about them.

There were a lot of studies done on how receptive are people to learning the secondary findings. One study in 2014 said that almost 93.5 percent of individuals undergoing a diagnostic, whole exome sequence said they did want one or more categories of incidental findings. So, most of the patients actually would opt in.

For children, however, parents were more ambivalent. If it's something that might predispose to a disorder that's treatable or preventable and it's preventable in childhood, they were much more likely to opt in. But if it's a predisposition of an untreatable or an adult onset condition or a carrier status for recessive conditions that actually doesn't put the child at risk but the child might be at risk once of having an affected child if they marry a spouse who is also a carrier, et cetera, parents were much less certain that they were interested in opting in for secondary findings under those conditions. In panels there are no incidental findings. You're only interrogating the genes that are relevant to the indication for the test, so this sort of eliminates this whole incidental finding or secondary finding quandary.

Consent for panels: they're basic genetics. You have to explain what basic genetics for patients if you're consenting them to have a panel test. You have to explain the difference between penetrants and expressivity that can carry a particular gene mutation without showing the disease, or two people in the family may carry the same mutation but, yet, not have exactly the same disorder; variable expressivity; what the different types of DNA variants one can find; ones that are disease causing; ones that we know are not disease causing, and ones we don't know anything about them; whether they cause disease or not. They have to know that there are false negatives available through panels, and they also need to be told about GINA, the Genetic Information Nondiscrimination Act.

Whereas, for whole genome or whole exome, you need to tell them all of this but we also have to explain to them that we may find pathogenic or damaging changes in genes that are of uncertain significance; genes we don't know whether they have anything to do with their disorder or not. There's the possibility of misattributed paternity that may come up when you do a trio, but also we need to tell them that there may be scientific discoveries that may result from test results. So the consent for whole genome or whole exome is more complex than for panels.

So the point I want to make here is that in human genetics we have always operated, and I've been in the field for four decades, and I'd say from the beginning of the field there has been a line between the clinical care of patients with hereditary disease, and research into the discovery of genes and the mutations in those genes that cause disease but that, as I'll show you -- that boundary is starting to be breached by whole genome or whole exome sequencing. I hope you understand from my explanation as to why that is.

So if we go over the history of genetic testing, and I'm just about done: we start with single genes and single variants, such as sickle cell disease. One gene, one mutation, or Huntington's Disease; one gene, one mutation causes that disease. So the number of genes involved is one, and the number of variants that can cause that disease is very limited. Just one, and you have a very high probability of a specific clinical diagnosis by doing the appropriate laboratory tests, or by doing the appropriate clinical evaluation and family history. Combine all those things. You have a pretty good idea, in advance, you've got a single gene, a single variant.

However, we now have a situation where you may have a disease where you don't actually know how many genes are involved. You can have different -- you can have beta thalassemia due to beta or the deletion of delta beta or a deletion of delta beta gamma. So there are multiple genes. There are also multiple mutations that can cause this disorder. It's not just a single one like sickle cell.

Then we get to things like gene panels like hereditary breast and ovarian cancer, where there are multiple genes that can predispose, or Lynch Syndrome; multiple genes that can predispose to that syndrome, and lots of different mutations that can be responsible and, when faced with a patient with an ovarian cancer or breast cancer or a colon cancer, the prior probability that they have this is not 100 percent. You can say, oh, yeah, given the family history and given the pathology of the polyps, or given the number of polyps, you can have an enriched idea that this is likely to be or it has a good chance of being hereditary breast, ovarian, or Lynch, but you don't know for sure. So you've moved up the axis here to the prior probability of this specific clinical diagnosis starting to drop, and then finally you get to whole genome or whole exome where the prior probability of a specific clinical diagnosis is very low. The number of genes involved may be so large or even unknown, and the application is primarily in the situation of undiagnosed disease, and you start to blur the line between clinical care and gene discovery in the undiagnosed diseases area, and I know this is something that you in the controversial care community are very involved with and very concerned with and, with that, we're going to stop.

[end of transcript]