JIM INGLE: Welcome to CTSA Grand Rounds. I'm Jim Ingle from oncology, and it's my distinct pleasure to introduce the speaker today. Dr. Couch is a professor and consultant in the Division of Experimental Pathology in the Department of DLMP. He received his undergraduate degree and PhD at the University of Cork, in Ireland, and then came to the US and did a postdoc at the University of Michigan with Francis Collins. And that's important, because as you know, two of the most important breast cancer predisposition genes are BRCA1 and 2 and Dr. Couch was there this early work was being done. And as you know, Dr. Collins is now the Director of the NIH. So he selected the right things to study.

Dr. Couch then did another postdoc and was on the faculty at the University of Pennsylvania, and Mayo was fortunate enough to recruit him, I believe, in 1997. Now, I think the essence of Dr. Couch is that he's the quintessential translational researcher. There's no doubt that his laboratory research is very high quality, very good. But the thing that has stood out for me, as the director of our Mayo Clinic breast cancer SPORE, is that Dr. Couch can bring people together and understand the basic science, but then take it in to translational research. Because at the end of the day, the purpose of all this is to do something for the patients we serve every day.

On a global basis, I've learned more acronyms from Dr. Couch than I would want to say-- SIMBA, BCAC, COMPLEXO was the one that I really like, ENIGMA. These are all international collaborations to work on the genetics of breast cancer, looking at risk, and as you'll see, the complexities of this are really formidable, And it takes a worldwide effort, and he's on the steering committee of these major organizations. He has many publications, over 250 publications, most of them in very high impact journals. He's very well funded. He's a project leader in our breast cancer SPORE, which I'm very pleased about, has multiple RO1's. And I think that there's no doubt that he brings real credit to those of us involved with SPORE, involved with the cancer center, and he's very well qualified to give his talk on breast cancer susceptibility, Fergus.

FERGUS COUCH: Well thanks, Jim, for that wonderful introduction. We should tape record that one and keep it. That's good. And thanks to the organizers for inviting me to speak here, and thank you for everybody for turning up. I think we have a great crowd here, and so hopefully we'll have some little bit of discussion at the end of all of this.

So I'm going to talk today about breast cancer susceptibility, and I'm going to move through the various phases of what we know about breast cancer susceptibility. Giving some historical perspectives, and then also showing up to data that these various consortia that I'm involved in have produced. So just as a quick introduction heritability of breast cancer, the circle represents all of breast cancer and so these little quadrants here represent-- these are familial clusters, and these are the ones that we think are at really high risk.

So you see mutations tracking through families. These are sort of the BRCA1 and 2s. But you put those together, you can very clearly see that a fairly high proportion of all breast cancer is highly hereditary. It really does track through families, and there's very clearly dominant genes that are driving this in the population.

These are the type of pedigrees that are seen in high risk breast cancer clinics all the time. The color represents breast cancer, or at least cancer, because we've got ovarian cancer over here in the prostate. But you see a family like this, where you've got multiple breast cancers at young age of onset. You've got ovarian. You've got prostate. This looks very much like BRCA1 or a BRCA2 family, but you can see how devastating this is to a family. There's so many people with cancer, just in this small little family, and this is quite common. We see hundreds of pedigrees like this, hundreds of families like this, all the time.

So here is really an update on where we are with breast cancer genetics, and so just to fill you in, if you look on the y-axis here, you're talking about the level of risk. So these are very high risk genes up here. These are quite low risk genes down here, and then across the x-axis we're looking at the frequency in the population. So essentially, what we've got is very rare, high risk genes, very common, law risk genes, and then this other bunch in the middle. So there's really three stages, and I'll talk about each of these three stages as we go along.

So the first genes, as Dr. Ingle referred to, the first genes we really found out about in breast cancer were the BRCA1 and BRCA2, or BRCA1 and BRCA2 genes, cloned back in 1994 and 1995. And I was involved in the positional cloning strategies to try and identify those genes. And they have become really very important in the community at this point, because there's a lot of women who have a family history of breast cancer, get tested for the mutations in these genes. And when you find a truncating mutation, which is a stop codon or a frame shift, we know very well that inactivates these proteins. And then that's associated with very high risk, and if you look here, it gives the idea that breast cancer.

So the lifetime risk of breast cancer in a carrier of a BRCA1 mutation is, on average, 65% but very, very high. You think of population risk at about 10%. So 60%, 65% is huge. Second primaries are really elevated there, and ovarian cancer as well, 20% to 40%, in the carriers of these mutations. And then there are a number of other cancers that come up, certainly men, and men get breast cancer. It's rare, but it does happen, and they have a five to six-fold elevated risk of male breast cancer, when you carry a BRCA1 mutation. And then there's five-fold elevated risk for prostate and for pancreatic cancer and a few others. And that's BRCA2.

So again, a mutation in BRCA2, again very high risks. We've really thrown the same numbers in here, but the risk of breast cancer is a little bit lower. It's maybe 55% in this setting. And again, ovarian is slightly less as well, but you get all these other cancers coming through in the background at low frequency as well. So clearly mutations in these two genes are really important.

It's important to know that you have the mutation. So maybe you can do prescreening before you ever develop the cancer. That's the key to all of this. It's risk assessment and then prevention, if at all possible. And many women who have-- and men-- who are found to have mutations in these genes undergo prophylactic mastectomy, and many women undergo prophylactic oophorectomy. Because by taking out the ovaries, you reduce your risk of ovarian cancer, obviously, but it also reduces your risk of breast cancer, because of the estrogen stimulation that's involved there.

So surgical intervention is really the key in this community, at this point, in the BRCA1 and 2 community. And of course, we all wish that we could do something other than surgery, and so there's always the interest in chemo prevention. Coming up with agents that can deal with this and not having to do these types of surgeries.

Now, my lab for a number of years has been interested in a particular type of mutation that pops up in BRCA1 and BRCA2, and these are called Variants of Uncertain Significance or VUS. And essentially, what they represent is mutations that do not truncate the protein. They don't cause stop codons or frame shifts. So mainly missense mutations, single amino acid changes in these genes. So this is actually very common, and the numbers are shown here.

But basically, in the United States alone, we know there are 1,000 different mutations. Now, not different people, but different mutations within the BRCA1 gene, and 1,000 within the BRCA2 gene, that we don't know what they do. If you have a missense mutation in BRCA1, and it's one of these VUS, you have a genetic event, obviously, in your germline that you've inherited from one of your parents. But you don't know if it predisposes to cancer, or if it's just a random genetic neutral change in the background, and so this is the problem for a lot of these people who get these diagnosis. You pay your $3,300 for your clinical test at Myriad Genetics, and you get back one of these results, and you don't know what to do, and your clinician doesn't know what to do. So it's a bit of a problem, more than a bit of a problem, a very serious problem in the clinic.

Now, Myriad Genetics, as I mentioned, is the company in Utah that does a lot of the testing, and they have figured out some of these, what they actually mean. So they're down to about 3% of all the variance they find are unexplained. But for some reason in the African-American population there seems to be a higher frequency of these VUS tracking through, and nobody really knows what that means. Is that just that there's a lot more of these neutral changes in the African-Americans, or is it that they really have a higher level of predisposition than we really understand at this time? And then interestingly, in other countries, where they don't have the Myriad equivalents-- so they don't have as much knowledge about all these missense mutations, these VUS-- it's 20% of all variants that are found are unexplained.

And as you might imagine, any clinical test that you do, if 20% of the results were unexplained, I'm not sure we'd be too happy with that. So this needs improvement. And so the advantage of knowing these, if you can classify, as we call, it if you can classify a VUS as causing disease, as pathogenic. Then again, we're back to taking advantage of the risk assessment and the prophylactic surgery and the chemo prevention in the future. So that's the goal.

So what I'm going to describe now is how we actually study some of these things, how we've tried to classify some of these VUS. So one example is we focus on sequence conservation. So the idea here is we're looking at a piece of the BRCA1 gene here, at the amino acid level, and we've lined up all these various species going back to evolution, and in fact, we go down as far as puffer fish. So we go puffer fish all the way up to human, 2 billion years of evolution, and we actually find amino acids that are perfectly conserved, over here, perfectly conserved across all of evolution. There's only 165 of these in the entire BRCA1 protein. So that means that they're probably important functionally.

Then, when we see a mutation in a patient in that site, such as this F here, this phenylalanine, we then can say, well, this is likely probably fairly important. And it might be disease causing, because we think that amino acid was very important functionally. And so we can put a prediction on that. We can put an odds of a likelihood on that. We can make quantitative estimates.

We also do things, such as we track the mutation through the families. So we test as many people in the family, and if you look at the family on the bottom here, the plus sign means everybody who has the mutation. There's a particular mutation tracking through this. So I think there's lots, you can see there's lots of breast cancer in this family, and essentially, they all have the mutation. So by looking at that segregation, again we can come up with an odds, or a likelihood, that that mutation is causing the disease in that family, so again, a quantitative estimate.

But we do that to a number of other ways too. So the segregation I mentioned here, we also look at the family history, so breast, ovarian cancer. If you see a loaded family, with all these cancers, it's likely that the mutation is causing disease. And we put all of this information together to come up with a probability. So it's quite mathematical and statistical, but we come up with a probability of pathogenicity.

And then, we've also used that sequence conservation that I mentioned, and that gives us another probability, and we can combine these. And the final output is the posterior probability of pathogenicity, and that tells us is the mutation likely to be disease causing. Or is it likely to be neutral, or does it remain in this middle zone, where we don't know. It remains what we call unclassified.

So we formed the ENIGMA consortium, and this was one of the consortia that Dr. Ingle mentioned. This is a group of about 60 research labs from around the world, all over the world, that have come together, and they're combining all their information about pedigrees and families in these VUS mutations. And by combining all these together, we've managed to identify 24 of these BRCA1 VUS as clearly disease causing, and 15 BRCA2s that are clearly disease causing. And then actually, quite large numbers that are not disease causing, that are clearly just neutrals. And that negative information, the neutrals, are quite important clinically as well, because we can dismiss the mutation is not really meaning anything.

So that's really a tremendous effort to really to collect all this information, but it's proving useful. And that allows us to come up with this five level, 1, 2, 3, 4, 5 class of mutation, with the 5's being definitely disease causing. The ones being definitely not disease causing being neutrals, and then there's the other groups in between, and you can see these probability estimates that are on this slide here. But it's very clear to us, from those studies, that there's not enough family information to classify all the mutations that we want to study. You could get every family in the world, and you still wouldn't have enough information to classify a lot of those VUS.

So we've gone to functional studies. Is there any way that we could develop an assay in the lab that would allow us to figure this out? And we focused on homologous recombination repair. This is a DNA repair assay, because we know that both BRCA1 and BRCA2 are classical DNA repair proteins. They're right there, involved in the process of repairing double-stranded DNA breaks.

So Maria Jasin in New York developed this assay about 15 years ago, and we've adapted it for our purposes. Essentially, there's two genes here in this sort of gray to blue color. They're called GFP genes, Green Fluorescent Protein. If that protein is produced, the cell that you're looking at is green in color. So we can actually very easy to quantify how many green cells you have in a cluster of cells. So that's really the basic idea.

But in this case, this one has mutated, so it doesn't work. And this one has a unique restriction site in it, and when we cleave that, we have a double strand break. So now we have two inactive green fluorescent proteins. So there's no green color being produced. But if BRCA1 or BRCA2 are active, they can induce homologous recombination, where we borrow a bit of this gene, copy it, bring it over, and fix this break. And now we've got a fully constituted GFP gene, the cell turns green, and we can count the number of green cells.

So it's really a fairly simple assay, and it's done inside a cell, and we can then count the number of cells. And so we decided to apply that to one region of BRCA2, the DNA binding domain region, way down at the C terminus. We picked that because it's got many, many mutations, many hundreds of mutations, that are actually in amino acids, that are highly conserved across evolution. Remember, we talked about that a few minutes ago. So this was sort of a hot spot region that we wanted to go after.

So this just shows that we were able to make artificial genes, artificial constructs of all these mutations, put them in cells. They all seemed to work, and then we did our analysis with our homologous recombination assay. And what we very quickly learned is that this assay works really well in this setting. These guys down here have low activity, and as it turns out, they're all known neutral mutations. Sorry, they're all known pathogenic mutations, wrong way around. These guys up here all retain activity at some level, and they're all known neutrals from the genetic studies.

So you can see that there's this very nice separation in activity between known deleterious and known neutrals. And so that allows us to really put this forward as an assay with high sensitivity and specificity that could actually be used clinically. It's a complicated assay, and I don't think it'll ever be set up in a clinical lab, but we can certainly do it in our research labs. And so we've been trolling through many, many different mutations now, and we've done over 140 different mutations in the lab, from just this region of BRCA2.

And again, this is a graph just showing you the neutral guys, again up in the corner, to deleterious or pathogenic ones down in this corner, and then these are all the guys that we've been testing. And I think you can that there's another big batch here that looks like they've lost activity, and they're likely deleterious. And so we feel that we can give this data directly back to a clinician, and to a patient, because of the sensitivity and specificity of the assay. We can take this back and say, use this information to inform your decision making and your clinical management of this patient, who is potentially at very high risk of disease. So you can see the clinical application right away.

So now, moving on from that, I'm going to jump on and start talking about other high risk genes, especially these guys in the middle, these moderate risk type jeans and a few of the high risk ones. So all we know about BRCA1 and 2. We've talked about that, but is there anything else out there. Well, there are, and you can see in this diagram that there are a few other genes that have been identified. P53, many of you all have heard about that, PTEN, the ATM gene, which is another DNA repair gene. CHEK2 another DNA repair gene, and all these things.

So here's a list of the ones that we think really do predispose. So these are in your blood. If you have a mutation, one of these genes, you're at high risk of breast cancer, and we know that they are useful in that regard. So the question is can they be used for predictive type testing? Unfortunately all together, they still only account for a small amount of the remaining risk. So there's clearly many other genes still out there that we haven't identified yet, but I'm just going to focus on this batch that we know are predisposing at this point and tell you a little bit about what we know.

So one of the earliest studies of all these genes put together was done by Mary-Claire King, at the University of Washington. And this is a study where essentially she took about 340, I think, ovarian cancer families, ovarian fallopian tube and peritoneal cancers and screened the blood from those patients-- so germline genetics, not tumor genetics-- screened the blood for the presence of mutations. And the various genes are shown here in schematic form, and you should focus on the yellows and the reds. These are the ones that completely inactivate the protein, for sure inactivate the protein.

Now, I think you'll see the majority of the mutations are in BRCA1 and 2, but there are some sprinkled around these other genes as well. So it's very clear that there are individuals, and there are families, out there with mutations in some of these other genes that we could benefit from knowing about. And so that's indeed what Ambry Genetics has done. Ambry Genetics is a company in California. They have set up a special screening panel for all these genes.

So essentially, what happens is you go for your BRCA1 and 2 test with Myriad. If that's negative, if you don't find a mutation and you're still worried, you go off to Ambry, and you get all the other genes screened. And so to date what we've heard from them is that 10% of the samples they've been screening have again truncating mutations in these other genes, suggesting elevated risk, but 40% of the samples are yielding VUS. So it's the same problem as BRCA1 and 2 again, where we have all these VUS that we don't know what they do.

Now, we've got 40% of all samples being screened by Ambry, producing things that we don't understand. So this is a huge clinical problem, because again these results come back to the clinician, and again they do not know what to tell the patient. So we've actually set up a collaboration with Ambry, where we're going to use all the principles that we learned from BRCA1, all those methods that I talked about and the likelihood ratios and the probabilities. We're going to apply all that to all these other genes, with the help of Ambry. They will provide information on the pedigrees, information on the mutations they find, and we'll work together to do some estimation of risk and try to figure out which of the VUS in these genes actually cause disease.

And indeed, we're going to extend our functional studies as well. We're going to work with David Largaespada in the University of Minnesota, and we're also going to apply this homologous recombination assay to a lot of these VUS to see what they do. One thing I should point out is that nearly every gene on this list turns out to be a DNA repair gene. So the central theme to predisposition to breast cancer is a DNA repair defect, and it makes it a lot easier for us to do these types of studies.

So here's an example of a gene that we know a little bit about. So this is CHEK2. It's a cell cycle control, DNA repair type gene. It's mutated fairly infrequently, but if you live in Finland, it's in 1.7% of the general population. If you live in the United States, it's probably only about 0.2% of the general population. It's quite rare, but it is out there.

And essentially, what was done in this study is that a very large number of breast cancer cases, and a very large number of unaffected controls, were screened for the presence of mutations in the gene. And then essentially, it's an association study. If you have an excess of mutations in the cases over the controls, that suggests that the mutations are driving breast cancer at some level. And so that's indeed what they figured out, is that there could be as much as an observational six-fold elevated risk of breast cancer, if you have one of these mutations. And again, they were able to then break it down by the types of mutations, and they found that there was different levels, depending on how evolutionarily conserved that mutation was. So we know a little bit about one of the genes, and we can potentially tell patients something about those results.

Here's another one, the ATM gene, a very well known DNA repair gene, causes ataxia-telangiectasia. But in this case, it also can predisposed to breast cancer, and here we find that the truncating mutations in this gene are associated with about a three-fold elevated risk of breast cancer. Now, that's not a BRCA1 mutation, at 10 to 20-fold, but it still is an increased risk, a substantial increased risk of breast cancer.

So we've also been looking at these genes, here at the Mayo Clinic, these 12 genes, and one thing we decided to do was to set up our own screening panel, just like Ambry Genetics is doing, but this time at a research level. And we focused in on what's called triple negative breast cancer, and this is a very aggressive form of breast cancer. It accounts for about 12% to 15% of all breast cancers. It's estrogen receptor, progesterone receptor, and HER2 negative. These are three of a classical histological markers for breast cancer.

So it's an aggressive form. It doesn't respond well to therapy, and we wanted to know more about it. And one of the early studies suggested that there maybe were quite a few BRCA1 mutations in it. In fact, as many as 11% of triple negatives might have BRCA1 mutations. That led us to ask whether some of these other genes that we've been talking about might also be mutated at high frequency in that population.

So we set up what's called a Triple Negative Breast Cancer Consortium. It's a group of about 30 labs from around the world who study triple negative breast cancer. They've sent us a lot of their samples, and we actually have 5,000 samples collected from triple negatives at this point. In this study, we've been screening the 12 genes for 2,000 of those samples, plus 1,000 match controls. And so the whole idea is to figure out what is the frequency of mutations in these genes, in triple negatives.

Can we estimate risk? Can we estimate the penetrance? So what's the risk within the family? Do they get ovarian cancer as well as breast cancer? All these types of questions, and this work is ongoing right now, and Steve Hart is in the audience. He's helping us with a lot of these analyzes, and so in the next six months or so, we should see this project come to fruition.

It's a large undertaking, as you can imagine, collecting all the samples from these groups, but I think it will be really quite productive. Here's an example of just setting up that panel, and in fact, we got support from CIM, the Center for Individualized Medicine, helped us with this, and also with DLMP. So we've had some support internally to set up the panel and then to test it. And really what I'm showing here is just a variety of known mutations in different genes that we kind of spiked into our sample set as a test run, and in fact, we found every one of them.

So this technique works. It works really well, and we look forward to doing the study and then to potentially setting an assay like this up in the clinical labs in DLMP, so that people could actually purchase the test directly. We don't have to go through Ambry Genetics. There's no patent holding or anything like that. We could just as easily do it at Mayo Clinic. So those are the known additional predisposition genes.

Now, what about other predisposition genes? Well, this is the big gray area out here. Do we know any other genes, or can we even find any other genes? We estimate actually that there's a lot out there. There's maybe 50 to 100 additional predisposition genes that we haven't found yet, and can we go try and find them?

And I think the reason for looking at them is the same reason is BRCA1 and BRCA1. Can we do prevention? Can we do a risk assessment? Can we know what other cancers might be expected in the families? We have to really understand this. So will there be a BRCA3 or 4?

Well, there's work ongoing worldwide, and certainly we've been doing some work here at the Mayo Clinic. We've been sequencing DNA, doing exome sequencing. It's a big buzzword around these days, but we've been doing that on about 70 to 80 families at this point, and then we're collaborating with people in New York and Philadelphia to do this. And this has led to the COMPLEXO Consortium that Dr. Ingle mentioned, again, about 30 groups from around the world, slightly different to the ones I've mentioned before. All doing these exome sequences on these high risk families that don't have mutations in these other genes, throwing all the data together, because it's a numbers game.

If you look one family, you can never be certain that the mutations, the hundreds of mutations that you find, which one is the driver mutation. You can't figure it out. So you need hundreds, if not thousands, of families, combine them all together to get a better feel for what might be the driver genes. So that's ongoing.

There have been a few small success stories to date. It's early days yet, but there have been a few. Here's an example of how we found the XRCC2 gene, which turns out to be again a DNA repair gene. So in this case, here's a family, again the black represents cancer. So you can see there's several cancers in the family.

These two individuals were sequenced. And the idea is that, if you sequence relatives, you can compare them, and you might get a better clue as to what the driver gene in that family might be. So we sequenced the two with the blue circles, but it turns out that one of them didn't have the mutation and one of them did. The mutation is shown here. And then, when we tested this other individual, she had it. So in some ways, we picked the wrong person to sequence, but this is what happens in these families in complex diseases. You're never really sure who you should be testing. So it just shows the complexity of the issue. It's just in some ways you have to be lucky to pull these things out.

Here's another large family with several cancers in it that we found a mutation in XRCC2. Again, this person turned out to have the mutation, but this person did not. So again, we picked the wrong person to study. Whereas, there are other people with breast cancer in the family who had the mutation. So it's a very complex study. Again, the need for this combination of data, it's absolutely critical to combine the data sets.

And then, just to verify that this was the gene, again we did these large case control type studies and came up with a good P value, a significant P value here, suggesting that this thing really is associated. Mutations in this gene are associated with predisposition to breast cancer. But if you thought that was difficult, it's even worse, because here's an example of a pedigree where you have four young onset breast cancers. So you might say, well, you should easily be able to find the gene in that family. It's so striking with four breast cancers.

But in fact, we've exome sequenced all of these people. We didn't just pick two, we did them all, and then we combined them all, and so what we actually find is missense mutations. So again, we're not sure if they do anything, and it's three different genes, all with different missense mutations. And if you note, if you can see it, the mom up here has all three of the mutations, but each of the daughters have different constellations. This one over here does have all three, but the others have one but not the other two, the second one, but not the other two.

And we think where we're heading right now is lots and lots of moderate level of predisposition genes, and each family might have three or four or five of these things segregating. And we have to find them all, and then put them together back into complex risk models to say, well, you have two, so your risk is such and such. You have three, your risk is such and such, and I think this is where we're going to end up. Five years from now, I'll be back here talking about this, and we'll have proved this hypothesis. So this is what's ahead of us.

Now, I'm going to talk about the bottom corner here. See if I'm on time. I am. The bottom corner, we're going after now the common ones, the common ones in the population. So everybody in this room has common polymorphisms. We have 10 million common polymorphisms each. They don't do anything, for the most part, but there are some of them that do seem to influence risk of breast cancer, and so I'm going to talk a little bit about some of the studies we've been doing in that.

I mentioned the high risk genes and how much they account for, but clearly 65% to 70% of the familial risk of breast cancer is still unexplained. So could it be lots and lots and lots of these common polymorphisms that are driving some of that? Perhaps. So we went about this as a GWAS. Now a GWAS is Genome-Wide Association Study. It really means a huge case control study.

So again, you take many, many thousands of people with breast cancer, many, many thousands of people who don't have breast cancer. You combine them together, and you can compare the frequency of the SNP in the two populations. If it's more in the cases, it's probably involved in breast cancer at some level. And the more you do, and the better your P value or your level of significance, the more you're convinced that it's meaningful. And so we have this nice threshold, this unique threshold of 5 by 10 to the minus 8 P value, and it's just widely used in the community.

You have to get under that for your level of significance to convince someone that this SNP, this polymorphism, is really a driver of breast cancer. It's just many years of experience from various centers and various studies have told us that this is a good cutoff. So that's really where we're at.

So this is what we've done in breast cancer, and we got involved in what's called the COGS Initiative. This was really heavily funded by a huge grant, a 17 million euro grant, in Europe over the last four years, and we were involved in it through our collaboration. And it basically involves taking 211,000 different polymorphisms, from all over the genome, and putting them on this little silicon wafers, as you see here, and then genotyping 230,000 people. So somebody do the multiplication. It's massive.

You can see the amount of data here, and we brought on samples from BCAC, which is a breast cancer consortium. OCAC is ovarian. PRACTICAL is prostate. SIMBA is BRCA1 and 2. Endometrial and a few others. So we try to make as much use of this chip as possible.

We stuffed all the samples onto it to try and learn something. And if we look at just overall breast cancer, we actually came up with 41 novel locii, just through that one huge study. So again, about, 40,000 breast cancer cases 40,000 controls, massive numbers, but it was productive. Here's 41 polymorphisms, and that seemed to be associated with risk, and what you're looking at here is the vertical bar means no association with risk. If it's to the left, it means it's reducing the risk of breast cancer. If it's to the right, it's increasing the risk of breast cancer, and I just show two columns of those, trying to get all 41 onto the slide. But you can see there's some fairly striking effects here.

But what is the risk? This is a 1.2. That means-- So we were talking about 6, an odds ratio of 6 for the CHEK2. Now, we're down to 1.2. So you can see the way the risk is diminished now. We're way down in really tiny little effects, and that's why we think we need hundreds of these things, all combined together at 1.2 each. That will add up to the risk of breast cancer in the population.

There were a lot of interesting genes biologically came out of this. These are well-known tumor driving oncogenes, and now we know that they're important in the germline as well. The TET gene, which is involved in telomere biology, so aging of chromosomes, is involved here. BRCA2 and CHEK2 actually came up as having these common polymorphisms, as well as those rare high risk mutations. So that's hitting two different ways in the population, so lots of interesting stuff here.

One other thing, the FTO gene, so this has been heavily implicated in obesity and diabetes and a number of other diseases. So now, you're faced with a hypothesis that is there a link between obesity, diabetes, and breast cancer, because the same gene has been hit in both diseases? So that's sort of an interesting thing that needs to be followed up.

We can then split breast cancer by estrogen receptor positive and estrogen receptor negative. These are two major types of breast cancer at the tumor level. And what we find is that a lot of-- this was a study from 2011, with the first set of 27 or so risk factors that we knew, and you can see here that they behave quite differently in the ER positives than they do in the ER negatives. So what we're finding is that there are some that drive ER negative only, some that drive ER positive only, and some that do both, and this is what's shown in the boxes here. Here's this ER negative only, a really small little group down here, but clearly it does tell us that there's different etiology. There's different ways these tumors derive from the very, very beginning and how they grow out. They really are different things in the end.

And then, we wanted to extend that a little bit. So I talked about the Triple Negative Breast Cancer Consortium. As well as doing those rare mutations that I talked about, we've also been doing these common SNPs, these common polymorphism studies, in the Triple Negative Breast Cancer Consortium. And again, these are a lot of our collaborators I won't focus on that. So we've done all these studies and found about 12 of these polymorphisms that are relevant to Triple Negative Breast Cancer.

We have now a total in breast cancer of about 77 that we understand are risk factors. 12 of those seemed to be relevant to Triple Negative Breast Cancer. And interestingly, when you line triple negative up with ER negative, they look very similar for the most part, and that makes sense because of course triple negative is ER negative as well. It's a component. It's a part of the overall ER negative group. So it made sense that they were similar, for the most part, but there are a few that have jumped out, that are triple negative specific.

So we can go overall breast cancer, and we can break that down to ER negative specific, and we can break that down to triple negative specific, and not relevant to the rest of the ER negatives. So you can see that we can compartmentalize breast cancer down into smaller and smaller groups. And so to date, we've actually identified three risk factors that seem to be driving triple negative and triple negative only. They have no effect on any other form of breast cancer. So that's quite interesting.

Now, we're going ahead and we're actually setting up now a new very large study. That was the COG Study. Later this year, we'll be setting up a study called the OncoChip. This will have 400,000 polymorphisms on it, and we'll be running 400,000 DNA samples on it in an effort to learn even more. Because we think that there's a lot more out there, and it's a numbers game. The more you do, the better the chances you can actually prove, to yourself and to others, that a particular polymorphism is meaningful. And we think all of that put together will add another 14% to the risk.

So if you look at all the rare mutations and the rare genes I talked about, and then all these common polymorphisms that I talked about, were up to about 50%. There's still 50% out there that we have no clue, and these are genetic events. We know they're out there as genetics. So what is it in the genetic background that we're not finding? So this is still obviously an area of investigation for us.

With all that nice genetics and discovery, the question has to be how is this useful at the clinical level? And so that's what I'm going to talk about here. It's a Polygenic Risk Score, and it's being developed for all these various SNPs. So this is taking the 77 SNPs for breast cancer overall and trying to figure out how do they influence risk, and the graph on the left here is really the key graph. The average lifetime risk of breast cancer in the population is about 10%.

So this would be the middle quantile here of this group of patients. But using the SNPs, we can identify women that are almost at double that risk across their lifetime, and we can also identify women that are at half that risk across their lifetime. So you can see that we can take the population and break it up and start to put it into high risk compartments and low risk compartments. And that can have major implications for how these people are followed, how often they get their mammograms, et cetera, et cetera. And all that work still needs to be done to figure out what the recommendations should be, in combination with the SNPs, but you can see that it's coming.

This is the future, that we'll be able to risk stratify different people and put them in little compartments. It turns out that the SNPs are not particularly informative for the estrogen receptor negatives. You can see that the risk variation is not great. That's because we don't have very many SNPs that are specific to the ER negatives just yet, but that should improve.

Now, to give you a feel for how that relates to other risk factors for breast cancer, let's focus in on things like high estrogen levels, two-fold elevated risk in the population-- early menarche, late menopause, lack of parity, so not having kids. Again, 1.2 to 1.5, really fairly small effects on the overall breast cancer in the population, but still well known risk factors for breast cancer. But now, if we come up and we look at some of these genes we talked about, the CHEK2 and PALB2 maybe two to three-fold risk. So around the same level as these environmental risk factors, if you will. Mammographic breast density, so having high mammographic breast density upon mammographic screening. Odds ratio 4, this is the strongest risk factor for breast cancer that we know at this time.

Now, let's take all 77 SNPs, put them all together into a risk model, what do we get? Well, the top 20% in risk are at 3.4. So they're actually very similar. Their SNPs alone are very similar to the mammographic breast density, which makes this the second most powerful risk factor for breast cancer. So I think you can see that it has clinical implications and can be embedded within risk models.

The top 1%, now, it seems like a small percentage, but then if you think, I'm looking at 10,000 women, 1% is a lot of women. They're at 11-fold increased risk, just based on these SNPs, and that's as strong as having a BRCA1 mutation. So the impact is huge clinically, because these women really have to be carefully followed, because they're very, very elevated risk of breast cancer.

This translates into we do risk prediction curves, ROC, Receiver Operator Curves, to show some of this. Currently we're at, with the Gail model, which is a model in breast cancer, we're at about 0.65 0.67. All these factors together, including the SNPs, are going to move us up to 0.79. That is a huge deflection of a receiver operator curve. So I think in the next couple of years, we're going to see really tremendous advances in the modeling and the risk assessment that we can do in the clinic.

How does this have implications for Mayo Clinic studies and Mayo Clinic patients? Well, we now have plans to take these 77 known risk factors, and we're going to genotype them through the DLMP. We're working with them to try and set up an assay, and we'll incorporate various family history and stuff, and we'll try to build a Mayo-specific risk model for our patients. We'll also look at cohorts to validate. You've got to validate your findings, or replicate your findings. So we'll be doing that and cohorts that were collected at Mayo Clinic.

The Benign Breast Disease Cohort from Lynn Hartmann, the MMHS mammographic density cohort from Celine Vachon, these are really important research cohorts that we can apply this model to. And then, in the breast clinic population, so the breast clinic on Gonda 2, a lot of women who have lumps and various other complications that turn out not necessarily to be cancer are going there, or they're having their biopsies that are being checked out. And again, 90% of them turn out not to have breast cancer, but these women are obviously at elevated risk, and then they're followed over time.

We're wondering whether this risk model, with these 77 SNPs, can be used to stratify those. So if I take those 90%, and I say, well, you've got all these SNPs. You are at elevated risk. We should follow you more carefully or more frequently. Over here, you are at really low risk. Maybe it's OK if you only come in every couple of years.

So again, we're trying to develop. We'll be aiming to develop these sort of clinical strategies over the next few years. And then, just jumping back to the BRCA1 carriers, so this is the SIMBA Consortium. We took every BRCA1 and BRCA2 mutation carrier that we could find around the world and put them all together into this consortium. And it turns out, that we have about 20,000 BRCA1 and about 10,000 BRCA2s, which is really a huge effort on everybody, from everybody, about 60 groups.

The idea here is that, when I find an individual with a BRCA1 mutation, her average lifetime risk is 65%, but that means that 35% of those women never get cancer. So why do some get it, and some never get it? Some get it at 30 years of age. Some don't get it until they're 70. Why is that? It's not all environmental. There's some genetics driving that.

So what we're going after are the modifiers. We think these are genetic modifiers that influence the age of onset of this disease, or whether you even get the breast cancer or not. So that's what this was focused on. Again, it's a GWAS. It's these common polymorphisms in the genome.

What can we figure out? Well, we went through these two stage studies, and out of it we've actually found a series of SNPs that are associated with breast cancer, BRCA1 mutation carriers. I'm just showing you the names of them here. You can see the significant levels, and then we actually have this 1q32, which turns out to have the MDM4 oncogene in it. We looked in BRCA1, and you can see it's highly significant. We looked in BRCA2, it's only marginally significant.

So you can see that BRCA1s and 2s are different in the context of this SNP, and then we also looked at breast cancer overall, and it turns out it's highly significant in triple negatives only. So from this SNP, it sort of validates what we've already known for a while. That BRCA1 mutation carriers look very like triple negative breast cancers, and they don't look like any of the other forms of breast cancer. So we're finding more and more of these interesting observations. This is the sort of readout you get from a GWAS.

Every spot here is a different polymorphism, and all the red ones are the ones that are highly significant. So somewhere in there is one or two or three polymorphisms that are driving this influence on risk. We don't know who they are, and so there's a lot of laboratory previous research, looking at all these little red spots, trying to figure out which one is influencing disease. Many years of work still to be done to understand the basic etiology of some of these things.

Interestingly, in looking at the BRCA1s, we also discovered that there are SNPs that influence ovarian cancer risk in the BRCA1s, not just breast cancer. And in fact, it turns out that there's no overlap. The ones that influence ovarian have no role in breast. The ones that influence breast have no role in ovarian. So these diseases are clearly very different, and they're influenced by different underlying modifiers.

So here's a summary of what we have so far for breast. You can see the list that we have at this point, and then for ovarian, the list that we have at this point. And there's between 6 and 10 risk factors now, genetic risk factors that we have in that population. And then in the BRCA2s alone, we've actually found unique hits in the BRCA2s that are not in BRCA1s, are not in sporadic breast cancer. So BRCA2 is its own little thing as well, and so you can see the complexity that's coming up here. But it does turn out that there are about 20 of the known SNPs that are relevant to BRCA2s, and you can see them shown here, some with fairly reasonable effects on risk.

But where is this all going? Well, it's all going to risk modeling. Again, can we relate this back to the patient? Can we help the patient with this information? There's really not much use in understanding it if we can't help the patient in some way.

These are cumulative risk curves-- age across the bottom, increasing age across the bottom, up to a lifetime of 80, and then risk up the y-axis here. And so the average lifetime risk of a BRCA1 mutation carrier here is the blue line. So by age 80, we're at 65% lifetime risk. With the SNPs that we've identified, just those 10 SNPs that we've identified in BRCA1 when carriers, we can put women into these other curves. And so those at highest risk go from about 85% up to 100% risk lifetime of getting breast cancer, if you have BRCA1 mutation and the right combination of these SNPs. So you can see the value there. They're really at hugely elevated risk. And then down at the bottom end, we have women going from about 45% down to about 20% lifetime.

Now, you might argue that that's interesting to know, but in fact, what's the point in knowing what's going to happen to me by age 80? I really need to know it at a younger age. What happens at age 40? What happens at age 50? And so that's what these two lines represent.

At age 50, we're seeing this tremendous variation in risk. So basically it's about 30% for the average BRCA1 carrier. It's 45% for the elevated risk people, the highest risk people, and it's as low as 20% for the lowest risk people. So you can see with just 10 SNPs, we're pulling 30 down to 20 and up to 40. You can see this variation, the utility of knowing that information, and again, we feel that if we can find more and more SNPs, we can further pull that apart.

So we may actually be able to identify the few BRCA1 carriers that won't develop breast cancer, or the ones that are really at hugely elevated risk, like 90% risk, even at young age, even at the age of 50. And so those would be the people that we wanted to do some intervention on. BRCA2 is very similar. This is the curve for BRCA2. It's a lifetime risk of about 55%.

So that's just showing the lifetime, but here's the age 50, and it's a little bit less informative, but the idea is that the average here is at running at about 20%. Here, we can get it down to about 10% and up to about 30%. So just like in BRCA1, we can deflect by about 10% in each direction. And I think you can see that, in this setting, if a women is at age 50, and she thinks she's got a 20% risk at that age, she might undergo this prophylactic oophorectomy and mastectomy. But if I can tell her, basically on the level of the genetics, that she actually is only at 10%, she might decide to wait a little while.

This is particularly important for women of childbearing age. Let's back it up to 40. If we had an effect like that at 40, instead of undergoing the oophorectomy the woman might wait a few years, if she still wants to have kids or have more kids in that later age area. So you can see it has implications for the clinic, and so that's really, overall, what we're coming down to. These slides are repeating, unfortunately.

OK, so let me stop there and just summarize. So there's a lot going on in breast cancer genetics. We have the rare BRCA1s and 2s that we're learning a lot about, and we're looking at these funny VUS mutations. We've got 12 other genes that we know predispose at moderate to high risk of breast cancer, and we're trying to understand what those exact risks are, and what the VUS in those genes might be doing. So we can help to predict who's really at increased risk.

And then, we have all these common polymorphisms that I think I've shown you have, or can be, effective in a clinic, risks as high as 11-fold in the general population for breast cancer. And then, in the BRCA1 and 2 carriers, identifying the women at particularly high risk and particularly low risk, so that maybe we can change clinical management. And so the slide here shows just some of the consortium I'm involved in. I probably collaborate with about 500 groups around the world with all these consortia, but it's been very productive. It's been gratifying to work with these people.

They're great, great groups of people, and again, I also want to thank Dr. Ingle for support through the SPORE and then team that are here, Steve Hart and various other people from my group that are here. Everybody's been working really hard, and I think we're making tremendous advances, and stay tuned for more. It's changing every day. I'll stop there.