SPEAKER: I'd like to welcome you to today's CTSA Grand Rounds presentation. It's my pleasure to introduce Dr. George Vasmatzis as our speaker today. Dr. Vasmatzis is assistant professor of Laboratory Medicine and Pathology and serves as director of the Biomarker Discovery Program in Development and Translation. Dr. Vasmatzis shares with me an Engineering degree from a Big 10 school, his from Purdue. Also had a degree from Texas Tech.

Went on to get his PhD in Biomedical Engineering at Boston University, had postdoctoral experiences at Boston University and at the National Institutes of Health before coming to join us here at Mayo. He has been a very active investigator. While he's been here, he has been the PI and coinvestigator of several grants from the NIH and other sources.

He's served on multiple study sections at the level of NIH and many different administrative responsibilities here with the institution and has over 50 publications. Very glad to have him today, and I think the fact that we have a very full house indicates the interest in this area. So we look forward to hearing from you, George.

GEORGE VASMATZIS: Well-- am I on? Good. Thank you for having me. It's a long journey from engineering to giving CTSA Grand Round, so I hope you'll be pleased with what I have to tell you.

A quick disclosure is that many of us that we are working in this area have interest, financial interest, in some of the biomarkers that are coming out. We're trying to protect IP. So, that's my disclosure.

I am here to give you an overview of the Biomarker Discovery Program under the Center of Individualized Medicine. And this is a program that I codirect with Dr. John Cheville, who's a pathologist.

The vision of our program is to become the world leader in biomarker discovery. We are improving patient care through seamless translation to the clinic. Our goal is to facilitate collaborative and team research and discover and validate biomarkers and develop clinical tests. The three major goals is to diagnose disease earlier and more accurately, to predict outcomes, and to tailor therapy using the patient's genetic profile.

We have organized an oversight committee, an executive committee that oversights our activities. The group is seen on the right. Our primary expectations from the institution is to deliver biomarker to DLMP and to deliver radiological markers to radiology. Secondary expectations is publications, secure external funding, and develop collaborations with other academic institutions and industry. The executive committee is supposed to keep us to our vision and our goals, to understand the projects, and make changes if it's needed.

So what is our value proposition? What do we have better than the broad institute and the foundation medicine who are actually leaders in this field, out of Boston? What can we do here to compete with them or even be better than them?

Well, we have an integrated research and practice model at Mayo that can be tapped in. It can be used to do a good biomarker discovery. We have a very large surgical practice. Tens of thousands of samples would come per year, from our 100-and-over surgical rooms.

We have the Frozen Section practice, which is pathologists looking at the samples right after surgery. And if they want to pick up the right sample for us to profile, they can do it right then and freeze it for proper profiling. We have many clinical expertise clinicians, oncologists, that see patients and take all the clinical information from those patients, put them in databases, and link them to the samples that we're using.

We have developed protocols here at Mayo and algorithms, both protocols to interrogate even small samples of cells and algorithms to look at the sequencing data. And, through a lot of help from the research community for the last couple of years, Dr. Riza, Dr. Prendergast, we have actually the technologists that we need do this type of interrogations. We have more than seven [INAUDIBLE] machines, and that makes us one of the world leaders in sequencing.

So what is our research model? We have a little bit different research model than the usual RO1 type of investigation. We call it a "product-driven," versus a "hypothesis-driven." We usually start with a clinical question.

So a clinician comes with a clinical dilemma where a biomarker could help them do a better decision. Then our team designs the right experiments. We go through a pipeline, a complex, perhaps expensive discovery pipeline, which has both experimental analytical work flows. And then we went-- go through different steps of verification and independent validation of the biomarkers.

That's the time we try to publish our findings and also try to protect IP for the clinic. And we send some of those biomarkers for test development.

Now, this model does not work in isolation from the typical model, which is hypothesis-driven. It starts with a biological question, and it goes through different design experimental steps to develop preliminary data and towards publication and securing external funding. As a matter of fact, we think that, with many of our findings from discovery, we can generate new hypotheses and we can help and guide the biological questions. And also, as some of the biologists will take some of those questions and work on them in different models, like cell lines and mouse models, they can give us knowledge that we can take back to our clinical design in consideration. So, the idea is the two models to work together and not in competition of one another.

Of course, the biomarker discovery is a very complex process, and I'm taking a very simplistic pipeline, as I showed you before, to a more complex one, here. And it many times starts with a market and financial analysis for the clinical objective and also reviewing patents and license rights. Then we go through a complex discovery evaluation, clinical validation and translation pipeline. And eventually we go to test development. And each one of those steps could lead to different publications.

Our business model is to develop teams of multiple expertise working together around patient needs, around a patient problem. And most of our teams have oncologists, molecular biologists, mathematicians, epidemiologists, pathologists, and surgeons working together in an equal setting. And our business model works through a design discovery, verification validation, and test development process. And it is organized as shown here.

So the three major groups working together are those three boxes. This is our main engine, where you have a group of pathologists extracting the right cells of interest from frozen or paraffin embedded tissues. We have a molecular biologist who will develop the protocols to sometimes take small amounts of DNA and amplify it and take it to a high-throughput technologist like next-generation sequencing in the micro record.

The millions of bits of information goes into a bioinformatics group where we have expert bioinformaticians taking the large amounts of data and funneling them through, filtering the noise out, and coming to a small number that can be more manageable. And then we have two other groups that will take those biomarkers, one group that will do more clinical development, test development, and work with DLMP to translate a test, and another group that will take biomarkers towards functional studies so that they can both help us publish much better papers as well as maybe go for more external funding using the RO1 model.

And this group works with all the core facilities around the clinic, the TACMA, the Biorepository, the BAP, as well as the bioinformatics corps. And it uses a lot of the databases that are readily available to us, like the CoPATH, for example, for pathology.

We have a number of projects going on through the program right now. There are some major projects. T-cell lymphoma, prostate cancer, lung cancer, pancreatic cancer are somewhat larger projects that we have funded this year. Endometrial cancer is an intermediate project. We also have two radiology projects. And we have an RFA process where we take proposals for pilot projects, give them about $50,000 to $60,000, and then have them get the preliminary data that they need to take it to a larger project and then convince all of your group to do that.

I'm going to go through each one of the projects real quick. And then I'm going to try to go into more highlights, the things that happened this year. So the T-cell lymphoma project, the team is shown above. Dr. Feldman is the leader, Dogan, myself, Ansell, Johnson, Slager, Caride.

As I'm going through the projects, I'm going to mention most of the names because there's more than 100 people working in the program right now. Somehow we managed to take little money and expand it to many products. So you will see the acknowledgements of the people who work on the projects, as I go through.

The clinical objective for the T-cell lymphoma project is to do better diagnosis and also better prognosis for certification of patients for more aggressive treatment, versus chemotherapy or other treatments. We have analyzed more than 100 samples using microarrays and next-generation sequencing. We have actually two publications in [INAUDIBLE].

We have many biomarkers. More than five biomarkers came out of the process, already. Two tests developed, and there are two active grants on the area already. So it is a pretty mature project and is doing very well.

Another major project is in prostate cancer. Again, the, team is listed above. The major political objectives is to predict risk for disease in the indolent, in the low-risk group, and also risk of recurrence in the high-risk group, as well as find field-effect biomarkers.

I'm going to go into a little bit more detail on the prostate-cancer project, so I'm just going to skim through it. And again, this is another project we've done hundreds of profiling with microarrays and next-generation sequencing. It cost, at the beginning $400,000 or $500,000 to do that, but you'll see all the knowledge that comes out of it later on.

Many publications already, biomarkers discovered, two tests developed. As a matter of fact, you can order them right now through the catalog. And there are two active grants.

Also, in lung cancer, here we have a pretty large team. And we are looking for a prognostic test to identify aggressive lung cancer, but we're also trying to figure out, when somebody has multiple nodules in their lung, if those are dependent or independent. And that is important for clinical decision. And if I have the time, I'll be able to tell you a little bit more about this project, as well.

Also four biomarkers have been discovered. A couple of them are in development. One of them can be ordered already. And tests have been developed.

Endometrial cancer. Here is the project. This is intermediate. There are only 30 samples on. And we are at the process right now, we're trying to figure out if we can find the biomarkers for the clinical objectives that they are shown here. As a matter of fact, the first one, "Identify risk factors for lymph-node metastasis," is really the objective that we are focusing. And we are trying to see, by analyzing the data, if we can find biomarkers for this objective.

In pancreatic cancer, we work with Fergus Couch, and we have a Minnesota Partnership Grant to look for early detection markers in cytology in [INAUDIBLE] specimens. We've done more than 50 next-generation sequencing experiments already. There is one publication almost accepted. And, as I said, that is one grant on it.

There are two major imaging projects. Dr. Ehman and Dr. McGee are working on fibrosis of imaging biomarkers. And Dr. Calson and Lowe are working on imaging biomarkers for pancreatic cancer. And we are trying to take this project and the other project with [INAUDIBLE] and somehow make them work together so that we don't duplicate the experiments and somehow use the funding more appropriately.

We do have many other, smaller projects going on. We work with Dr. Grebe in a couple of very interesting projects. We have taken the eight genes that they do sequencing on those, one at a time. And we have been able to put them in a panel, under a next-generation sequencing, multifaceted approach. And we can look at the mutations of all eight of those genes at the same time. And that is now-- will be becoming eventually a clinical test pretty soon.

Another interesting question that Dr. Grebe came about is that-- says we have no way right now to find, if we see two mutations in a gene, if they are in the same allele or if they're in different alleles. And I guess this is an extremely important question for when you have a double mutation for the patient, because this might be-- this patient will have still a good allele to have, in this case, but this patient might have two better alleles, and that would be much worse.

So we have found ways to solve the problem actually with some of our protocols. And here's just a quick example, showing a bunch of sequences, the mate-pair approach, where we can link now the mutations from distant parts of the gene. And something like that could be, from what I'm told, transferred to hundreds of tests. Because this is an important problem in many situations. And we've been able to solve this.

We also have a very elaborate project-management team and an administrative team in the Sim and biomarker discovery which tracks all the projects and also tracks the biomarker translation. And I'm showing you easily 20 different biomarkers, here, from the different diseases, at the point where they are at either with DLMP into a catalog or delivered to DLMP or in a codevelopment stage with DLMP. So we have project management to track the projects and keep us on track.

And right here is actually the MML code of the test that will develop. Every time you see a number here, you can actually order this test right now. So we have already shown success.

The way that we have decided to bring new projects in is through an RFA process. Now we are down to four pages. And there was just an announcement, today, I believe, for a new set of RFAs. And there is a template that I can provide, or maybe it's provided by the administration, to pull in new projects. So there will be funding for this project for next year. And we will take them through--

We have a review group, shown here, expert reviewers, both from DLMP the Biomarker Discovery, and from OIP, to make sure that this is something that can be taken to the clinic. So the reviewers will look at the proposals and make recommendations for funding, and we'll go from there.

For example, last year we received 17 applications. And we reviewed them for these different reasons. They are not exactly the NIH type of critiques, because they do look for clinical significance and especially the little extra property that you won't see at NIH, but this is what we think that we have to have for the Biomarker Discovery Program. We were able to fund five of those, and now we're in the process to seeing how they are doing and if they can provide biomarkers for next year.

So, this is-- what I try to do in the first part of the talk is to give you a quick overview of the whole program, so what it is. But what I'm going to try to do now, in the 30 minutes that I am left, is take you through some of the highlights, so that you can see some of the science which is going on. As you can see, there are so many projects, and it's impossible for me to go through all the projects and show you the science, so I'll just pick a couple of projects and put them here. And there is a chance I'm not going to be able to go through everything I have here, as well. So I might be going fast, so that I can leave five minutes of the end for questions.

And the subjects that I like to touch on is lineage relationships of concurrent tumors and mixed grades, the synchronous lung tumors-- are they independent, or are they related in a lineage fashion? And also I want to talk about context gene sets that they associate to clinical information.

This last one, there is a chance that I'm not going to get to it. But it is very interesting, because, with all these discoveries we are doing, we are coming up with sets of genes, or gene alterations, that are associated with clinical outcomes. So, if a new patient comes in this disease, we can then contrast it with those databases, their gene alterations, and figure out how to treat them. So it's very informative and very important data sets which we want to move on especially for next year.

So the first example that I'm going to go into is on prostate cancer. Out of the 200,000 surgical patients in the United States, about half of the people have what we call "low-risk disease," of which many of them are overtreated right now. They have positive biopsy, they have high PSA, positive biopsy, and the recommendation, not that long ago, was to go for surgery.

But then about half of them, after surgery they find a very small nodule, and they ask the question, should we even have taken that surgery? So it's a half-a-billion-dollar question, really, for the health system. If we can come up with a test to figure out who would progress versus who wouldn't progress, and can be just watchful waiting regimen, then we will save a lot of money. And we can make Mayo Clinic a destination center for this type of testing.

Also, on the other side, on the left upper corner, we have a group of patients that, even after surgery, they're not cured [INAUDIBLE]. And if you knew that at surgery time, that would be a good group to go and do more aggressive treatments right after surgery. These are the two places we are trying to focus on prostate cancer.

Now, prostate cancer has multiple different grades. And Gleason has given the names of those grades. And sometimes you have two different grades in the same patient.

Now I'm showing here a patient that has both Gleason pattern 3 and Gleason patter 4 in the same time. Those are the most interesting patients. We know Gleason pattern 6 is-- that means that they have only two patents of Gleason pattern 3-- are usually not that aggressive. The question is, what do we do with patients that they have mixed patterns?

And the other question, the more research question, is, do cells from this side and cells from this side have any relationship with each other? Did something from-- did this cell come from a cell from here? Did they have lineage relationship? And I'm going to try to get into that subject with this talk.

But the important clinical objective in here is actually the indolence test. Here's two different prostates. This is all cancer. And, in this case, there is very little cancer on that prostate. Yet both those two patients went through surgery. And at the biopsy level, they had exactly the same clinical characteristics. OK?

One core was positive. This patient had only this few glands. OK, so a very indolent cancer. And here is actually the needle, as it went into the prostate.

So you can see it picked up a few glands, and said, oh, you've got cancer. You've got to go to surgery. Where this patient also had a positive thing but had a big cancer, perhaps quite aggressive. And we know that big cancers or cancers that they have mixed grades have much worse outcome that the ones that they are small with no mixed grades.

So, can we make that decision at the biopsy level? Well, because at the biopsy level you don't have near the volume of the cancer nor you know all the different grades that the cancer has, you just have a few cells, the question is, can you look at the molecular profile of those cells and tell if they have alterations that they are aggressive or not?

But how do you do this, especially when you-- at the biopsy level, or even at the [INAUDIBLE] level-- you only have a few hundred cells to work with. You cannot just take bulk tissue, because that would be contaminated with a lot of normal cells. So we have developed, here at Mayo, protocols to be able to go and pick up a small number of cells with laser-capture microdissection. And this is shown here.

This is-- some of those slides, Steve Murphy has made. So I just stole them from him. But you can see here the cap falling on the tissue. And then there is a laser that hits this polymer, makes it adhesive. And when you lift the cap, you take the little piece of tissue with you.

And here is a real thing, when you have a piece that you want to interrogate, you can lift it to the cap. So, this way, you can isolate the cancer cells from the normal cells.

Now, we were able to get good-quality DNA from frozen sections. And Steve has developed an insight to whole-genome amplification technique. So he does amplification right on the cap, using all the cells, as opposed to taking this through DNA isolation, where you would lose a lot of DNA.

So we can take that DNA, and people like Bruce, who is sitting out here, has developed techniques to sequence this DNA. And myself and a few others in the group have developed algorithms to analyze the data and find alterations. As a matter of fact, we, here at Mayo, have our own algorithm based on binary indexing. That's coming from my engineering background. I thought that I can take nucleotides and change them to zeros and ones, and that conversion makes the analysis much, much faster, actually. But I'm not going to go into the algorithm, in this talk.

So we do have the protocols. We have the techniques. Now we have to design the experiment. And usually what we want to do is look at the whole spectrum of the disease.

So we take indolent cancers with a significant [INAUDIBLE] aggressive. And we also get normal cells, normal adjacent to the tumor. So we are able to compare normal and cancers of all these different individuals. We find changes between normal cancer, and then we take them through case controls, where everything else is matched-- other than the outcome, let's say, or the significance of the tumor-- and find, and then associate each one of the changes with badness or with indolence. That's the idea.

And here is a real experiment. This part of the tumor is Gleason pattern 3. This part of the tumor is Gleason pattern 4. We're able to isolate these two groups separately into two different caps. And now we can profile this, and we can profile that separately.

And what we did is we took that through different sequencing techniques. Now, if you did what we call "paired-end sequencing," you break the DNA to 200- to 500-long pieces-- so, pretty small pieces. And then what you want is to go and cover the DNA and see whether there are things like rearrangements.

Like, one of the interesting rearrangements that we found some time ago is the BCR-ABL rearrangement. It creates the BCR-ABL fusion gene. And this is the Philadelphia chromosome. If you wanted to find out with next-generation sequencing, you can, as long as you cover the whole genome and find pair ends that they hit before and after that break point. OK? So that's how the pair-end works.

The problem is that, because these are small pieces, you need to be covering the genome quite a lot to see them, which means that we might need a few lanes of the sequencing-- mixing, making this quite expensive to do for many patients. So what we've done here is we developed another protocol called "mate-pair sequencing"-- and Bruce and Steve have worked on that, and published on it, actually-- where we break DNA on 2Kb--to-5Kb pieces, now, and circularize it and then just phase out the ends, only, make them 500 long, and then sequence them like a pair end. So, instead of having small pieces trying to cover a junction, we now have these larger pieces where we find many more pieces covering it. In this way, we can make this--

So it's almost like tiling, let's say, the floor. Instead of using small tiles, you use larger tiles, so you need less tiles to cover it. Which means a less-expensive experiment.

Now, out of this next-generation sequencing [INAUDIBLE], you get hundreds of millions of sequences. So it is a big task to figure out how to take all those sequences and analyze them and map them to the genome and so forth. We have developed, here at Mayo, a set of algorithms that can do that very fast. And we are working closely with IT-- Travis is here, working on the algorithm-- through the Sim and the bioinformatics corps to make those readily available to everybody.

But the idea is this funneling approach, where you start with something, with sometimes a billion sequences, and you have to go down to a few that you can handle, a few biomarkers you can handle, and you can take them to DLMP. All right? So, that's the idea. How do you develop all these different steps to find, in this vast amount of data, the important alterations that matter for the disease?

So what comes out of this pipeline is things like that, where, for example, here is one case, we have rearrangements. These positions of chromosome-- in this case, it's chromosome 2, in both positions, showing perhaps a deletion or an inversion on that chromosome. And you have a slew of this type of breakpoints. Some of them are actually between different chromosomes and sometimes within the same chromosome. OK?

So this is the type of data we generate. And each one of those breakpoints could be an important one. For example, here you have the EML4-ALK translocation, which now we know it is in about 5% of lung-cancer patients. And it could be a possible target for lung cancer. And now we have a test for it, both a FISH test as well as an [INAUDIBLE] stain test for it, here.

So these type of experiments can provide these type of breakpoints. Now, we also can take the data and look at the structure of the genome. For example, if you look at-- this is chromosome 3-- you do these frequency plots across the genome, and you can see, you know, chromosome 3 is pretty OK, in most part, except right here-- looks like there is a dip, indicating a possible deletion at that point.

And, not only that, we have made pairs that cover that deletion. So then we have two independent ways of figuring out that there is something deleted in that region. It could be that, right here, you have two genes that they are fused, because of that deletion. And that might be an important either target or a biomarker. And we are looking for things like this.

For example-- and I'm showing now the same thing but in about 25 different patients-- if you look at chromosome 21 of these different patients, right here, there is a deletion right at this point. I think it's about 40, 45 million. This deletion brings two genes together, [INAUDIBLE], creates this fusion gene. And that's one of the most common deletions in prostate cancer. It was found about five years ago. And it is becoming a test now in many different ways.

But, as you can see, this line here separates the indolent cancers from the aggressive cancers. And it is also in the indolent cancers. So that wouldn't be a good marker for indolence, because it is equally in both groups.

Also, every you see a number here, that means that there is a translocation between chromosome 21 and, in this case, chromosome 9. But what is obvious now is that there is a lot of heterogeneity. Now, there is some commonality between the patients. Obviously this one. But there is a lot of heterogeneity.

Here is chromosome 10. This particular spot is around the gene p10, which is a tumor suppressor. And you can see that there is many different ways of hitting that area-- small deletions, large deletions, sometimes there are inversions. Sometimes a whole arm goes away. And all what it's trying to do is actually delete p10. OK? Many different ways.

Interestingly, with p10 you can only see it in the significant cancers and not in the indolent cancers-- maybe showing that this could be a biomarker for indolence. So, if, you have a positive biopsy, Gleason score 3, Gleason pattern 3, and you are positive for p10, you'd go for surgery. But if you're negative for p10, you might think you have an indolent cancer. OK?

Unfortunately, this is not 100%, because there are some significant ones that do not have deletions. Which means we have to find multiple genes together, that we'll be able to make that decision correctly. One biomarker won't do it, is the point.

Now, this is another way of visualizing the data. One could ask, OK, so if we take, we go through all the patients, what is the gene that is being hit most? OK? And other genes that are being hit multiple times.

For example, here, I'm showing 14 patients. And I have sorted the genes according to their recurrence or how recurrent and they are. And what that shows is that ERG, the chromosome 21 one, is the one that is the most recurrent in about eight of the patients. OK? And when there is a blue here, it means that it is both in the Gleason pattern 3 and the Gleason pattern 4 of the patient. OK? Which means that it is probably an early event and not a progressive event from 3 to 4.

On the other hand, when something is red that means it was only on the 4 and not on the 3. And that can give us indications of alterations that they are late, versus other ones, like the green ones, that they are only on the 3 and maybe not important. OK? So this is one way for us to look for driver mutations.

But what is clear here is that there is a lot of heterogeneity between patients. You take two patients, nothing is-- they're not identical. Which shows how important individualized medicine is. We need to figure out what is the molecular profile before we know how to treat them correctly, both for the prognosis as well as for the treatment.

So, what we are trying to do now, the more patients we get, what we are trying to do is find the recurrence of genes and find the most recurrent ones and see if they are diagnostic. So we develop this type of plots, where, for example, here, out of 27 prostate patients, the ERG is in 15 patients, p10 is in 5 patients, and there are some in 4 patients, 20 genes in 3 patients, and so forth. So we are looking for those ones that they are the most recurrent genes.

Now, what is obvious here is that there is no gene that is in all patients. That's one. And the other thing is that, if you are looking for diagnostic markers, especially the ones that they are in the lower recurrence, you actually have to do a lot of patients before you can associate the clinical outcome to that gene.

Which is why, for most of those projects, we need to go to more than 100 patients. The more patients we do, the more genes we find that we can associate to clinical outcome.

The other thing we can do is we can look at shared mutations between different patterns and different grades in the same disease. The bar plot on the top shows, for each patient, how many shared mutations, how many shared alterations, there are there. So black, here, is the shared, the light grey is the GP3, and the darker is the GP4. And shows that most of the patients have some shared mutations, proving that there is lineage relationships between these different patterns.

But sometimes, many times, you have additional mutations in both the patterns. So it looks like the tumors are boiling, and they not only have started together from somewhere but they are also gaining more mutations.

Now, if you used a phylogenetic tree analysis, using all those dimensions, you can see how different samples link together. When you have a big line here, that means you have a lot of shared mutations. And samples will come together if you have a lot of shared and will go away if they have nothing shared.

As you can see, most of the shared mutations, what they do is they bring samples together, like PR25 A and B here means the Gleason pattern 3 and 4 from the same patient. So, a GP3 is closer to its GP4, next to it, than it is to GP3s from another patient.

Using this type of data, we can come with different models of progression. So here's the different scenarios we see. There are situations where something started, let's say, from a normal-- sometimes we do see background mutations. You have GP3 and then, from then, spin off GP4.

But some other times what we see is that the GP3 and the GP4 are independent from each other. And sometimes there is just some dependence. With this type of data, we can plot these scenarios for different diseases.

Now, let me touch on another theme, here, on these gene sets. What we can also do is, by knowing when a gene hits, does it hit in a higher grade? Or if it hits in an indolent cancer, we can put a value to the gene.

So, for example in that value as shown here, something that has zero, it means that it has no significance with indolence or with earlier versus late. Where something which is on the left would be more like an indolent marker, something on the right would be more of an aggressive marker. Using the more data we have and the more samples we can get, we can give this value to the gene, which is actually extremely important for diagnostic purposes.

Up on top, we have the most recurrent genes. Again, ERG and [INAUDIBLE] are the most recurrent. And, as you can see, they kind of fall into the middle, which means that this particular marker will tell you little about indolence or aggressiveness. Where other things, like p10 or this PDE4D, are more on the right. And those would be the more aggressive markers. These are the ones we go after to develop a test.

So we see thousands of alterations. And one can say, bueno, you can find any biomarkers? I can tell you, I can find thousands of biomarkers. And many of them are going to be important. So which ones of those are you going to go after?

Well, we're looking for two things. One is biological significance. The other one is clinical significance. If we can use these type of scenarios to find the ones that they seem to be biologically important and clinically important, those are the ones we try to codevelop with DLMP as a test.

So, let me go to another project for a little bit. This is a lung-cancer project that has started with a large team from Pathology-Oncology-- Dr. Aubry, Jin Jen, Julian Molina, Charles Thomas, myself, Dennis Wigle, Ping Yang, and many other people actually are coming together to work on this project. A a mater of fact, one of the major accomplishments that we have as a problem is that we were able to pool people together. Just with a little bit of funding, it's amazing how many people actually find value on this type of data and they want to work together.

Now, lung cancer is a very heterogeneous disease. There are actually three major subtypes-- squamous, adenocarcinoma, and small-cell. And sometimes you see those same subtypes in the same patient. Like I'm showing here, this is actually a mixed one. This person had both adenocarcinoma and small-cell. And they were close to each other, too. And, as I was showing you from the prostate cancer, many of those grades that they were close to each other had lineage relationship.

But that doesn't mean that's always the case, because you could have a collision tumor, as well, what is called, a tumor grows and eventually they collide. So they might not be. Question is, are they related, or not? And can we use the technologies we have to solve these type of problems?

So, in this project, we looked for this type of analysis. When we have two tumors next to each other, adjacent to each other, are they related in a lineage formation? Or, if they are far away from each other, are they independent?

Now, this is an important biological question, because it tells you if something grows from something else. For example, if that is AIS or BAC, and this is an invasive adenocarcinoma, you want to know if adenocarcinoma came from the BAC. Now that was always assumed from pathologists, because they many times saw them together. But it was never proven in a molecular basis. But with the techniques we have right now, we can actually prove this. And I'll show you the data.

Also, in another clinical scenario, where you have two different nodules and they are far away from another, if they are metastasis from one another, that would be a high-stage disease and need to be treated differently than if they were independent from one another. Then it would mean two stage-1 disease and could be cut independently, for example, or be treated independently. All right?

So, can we use the technologies we have to resolve some of those questions? Here's a real case, where Dr. Aubry has given us this case, where this in situ component and the invasive component are next to each other. OK? There is some separation between them, but obviously there is-- well, you could say that, although they look very different with one another, they're so close, somebody might ask that may be related in a lineage formation. So that's the question.

And the clinical significance of that is that, if the different 100% IS versus 100% invasive have very different clinical outcomes. This is quite good, where this is quite bad. And the question is, what is happening on those intermediates?

So, again, we went through different ways of doing that. We have whole-genome sequencing, which is very expensive. But it can give you complete information. You have exome sequencing, where you can actually capture only the exome parts of the genome and look for somatic mutations between normal and cancer. And that is intermediate pricing. It's about $2,000, isn't it, to do a good exome capture, Bruce? What is it?

BRUCE: It's coming down.

GEORGE VASMATZIS: Coming down [INAUDIBLE] that? And then we have the rearrangement sequencing protocol that I showed you earlier for prostate cancer, which is more on the $1,000-per-patient pricing.

So we use the rearrangement sequencing. We thought that this would be a better way to look for this type of lineage relationship and, thus, independence. And it turns out that, when you look in the mutational space with exome capture, you see a lot of mutations. And sometimes you don't know if they are a common mutation-- common somatic mutation-- if it came from a background, or if it was a germline mutation. So it's not so easy to look at these lineage relationships in that space.

The rearrangement space, where you look for translocations or deletions, seems to be a more specific way to look for things. Like, for example, if, in the two different regions, you find an ELM4-ALK inversion, which makes this fusion gene, and it is exactly the same place on the DNA, it is very unlikely that that would be by chance. As a matter of fact, when we look for rearrangements across different genes, even if they hit the same gene it is never at the same spot on the genome. It's always off by some. Where a mutation can be exactly the same mutation between different patients, it never happens in the rearrangement space.

So the first part of the study was to look at lineage relationships between the AIS component and the adenocarcinoma component. We use 14 cases that they had both the components in. And, again, we did laser-capture microdissection, to make sure we don't have contamination between the two groups of cells, populations of cells. And we have no contamination from normal cells.

We did in situ whole-genome amplification. We did the mate-pair sequencing for rearrangements. And we did the next-generation sequencing. And, just to prove to you how clean the experiment is, here's the staining of the case before LCM. This part is the AIS component. This is the adenocarcinoma component. And it looks like there is actually a region here that separates the two quite well. OK?

So, this is the after-post-LCM picture, showing how this and this were collected. This is the cap from this site, and this is the cap for this site. Better than 95%, I would say, better than 95% purity, to do this. OK? When, if you didn't do that, and you used bulk type of experiments, you might have contamination between the populations. You wouldn't know what's going on.

Again, we go through the mate-pair protocol, next-generation sequencing. We've come up with slews of rearrangements, in tables like this, as I showed you before. And then, here, from the same case-- that's LU6-- what we are showing is that there is a group of mutations that they are almost identical between the BAC and the adenocarcinoma component.

There is a number of mutations that they are only in the adenocarcinoma component, perhaps showing progression markers or showing progression of the adenocarcinoma, but there are so many common mutations between the two, showing that there is-- into the molecular level, they are somehow related. So, again, for these 14 patients, we can see blue is when there is a certain mutation, green is if IS only, and red is in the invasive only.

And this particular case, for example, has many common mutations. And so does this. Almost proving that we don't know if the adenocarcinoma progressed from the AIS, but this is a pretty good proof to show that those two came from the same precursor cell, having so many common mutations.

And then, going and looking at the shared mutations for all the samples, you can see that some have many shared mutations, showing the lineage relationship. Right here we have three cases that are pretty interesting, in a sense that they have shared mutations and adenocarcinoma mutations, showing a real progression of the disease, that the AIS, the noninvasive component, has no more mutations than the adenocarcinoma component, where the AD has a lot more, showing real progression.

So, what we have done is, we see all this from bioinformatics and from these tables. But are they real? Well, we have used techniques to go and validate those, by PCR. So we go-- since we know the two ends, we can go and make primers and amplify the junction point. And then we can see that can be done with a different experimental technique. And if it is in the adjacent normal or genomic control. And we can take this band and go and sequence it and find the exact breakpoint.

And this happens, you know, I don't know, more than 90% of the cases, it's validating, Steve? Yeah. So, really, the whole experimental process works quite well.

Now, there are some times, though, that we do see situations where the adjacent normal has the same mutation. And sometimes the adjacent normal has a faint band. We started thinking, what is going on? I mean, does that mean that the background had that mutation, beforehand? ?

So we went in and analyzed that a lot more and found that the ones that seem to be more lineage relationship-- they looked like this-- were the ones that they seemed to have less relationship. They have a lot of background common mutations. OK? So imagine the following-- that, in some case, you have background mutation becomes AIS, and that becomes AD. Or sometimes you have background mutations, and two different cancers rise from it. OK?

Now, that's a pretty interesting observation. This background, if you look, to the pathologists that looks completely normal. It looks like a normal phenotype.

And, even more, we were-- going back to this faint band, told us that not all background is the same. So we did the following experiment. We actually cut different parts of the background of the adjacent normals from different parts of the block and then showed that only one of those was positive for the mutation, where all the other ones were negative, showing that the background problem is a zonal issue. So that there will be something happening at some part of the normal. Cancer comes from there, but not necessarily everywhere around.

So now we come with this hypothesis, that cancer arises in zones. And the big question here is that, if you were to take this piece, somebody's cancer, out, have you actually taken the background which was responsible for that cancer? Is that positive for abnormalities that you cannot see, by the pathology?

I'm getting close to the end, here, so I'm going to quickly skip through a few of those and, in a couple of minutes bring up another problem. Here is a person with two positive nodes in their lungs, and they're very far away from one another. The question is, can we use the same technologies to tell if those are metastases from one another or are they completely independent?

And we've done that for seven cases already. And, for this pilot study, we used things that they looked very, very similar to the pathologies. Like, for example, this is LU27a and LU27b. Those are two distant nodes. But you can see how close they look from each other.

So a pathologist could say, those are probably metastases, one from another, or they have a common precursor somewhere. And so did this. When there were other cases that they looked very different-- like, this is case 26, a and b-- those are very different from one another. So obviously they would be-- we would say that the molecular changes here would be different from those two.

Fair enough. We went through our whole mate-pair process, and we came with all the alterations. And we were able to split the cases that we have to ones that they are likely not related versus the ones that they are likely related, using the same approaches. Now, that's kind of an important clinical question, many times, where somebody needs to know if they should be treated with chemotherapy, if they have stage IV disease, versus go to surgery, if they have two stage-I tumors. OK? And now we can do this, with this type of test.

So I think I'm going to stop here. As I said, there is a large number of people that we work together to solve these type of problems. We have many biomarkers in translation. And we hope that whoever has any good ideas they can go through our RFA process and come and work with us to codevelop more biomarkers and do this type of research. Thank you.

[APPLAUSE]