Advancing Cancer Genomics:
The Impact of Personalized Genome Sequencing
Webinar
25 February 2013

[0:00:10]
Slide 1
Sean Sanders:
Hello and welcome to this Science/AAAS audio webinar. I’m Sean Sanders editor for custom publishing at Science.

Recent advances in high throughput sequencing technologies has helped to clarify the genetic mechanisms certain of cancers, revealing variants associated with novel pathways in tumor formation and response to anti-tumor therapies. The clinical potential of sequencing technologies in the diagnosis and treatment of diseases has now been demonstrated, but certain challenges remain, especially in sample preparation and data interpretation.

During this webinar, our expert panel will describe recent discoveries in cancer genomics and inherited diseases, and how these advances are enabling personalized medicine and the individualization of treatment.

Slide 2

I have three esteemed speakers on the line with me today to talk about this topic. They are Dr. Elaine Mardis from Washington University in St. Louis Missouri and Dr. Michael Snyder from Stanford University in California, along with Dr. Philip Stephens from Foundation Medicine Cambridge, Massachusetts. A very warm welcome to all of you and thank you for being with us.

Before we get started, I have some information that our audience might find helpful. Note that you can resize or hide any of the windows in your viewing console. The widgets at the bottom of the console control what you see. Just click on these to see speaker bios or additional information about technologies related to today’s discussion or to download a PDF of the slides.
Each of our speakers will give a short presentation followed by a Q&A session during which our guests will address the questions submitted by our live online viewers. So if you're joining us live, start thinking about some questions now and submit them at any time by typing them into the box on the bottom left of your viewing console and clicking the submit button. If you can't see this box, just click the red Q&A widget at the bottom of the screen. Please remember to keep your questions short and concise, as this will give them the best chance of being put to our panel.

You can also log in to your Facebook, Twitter, or LinkedIn accounts during the webinar to post updates or send tweets about the event, just click the relevant widgets at the bottom of the screen. For tweets, you can add the hash tag sciencewebinar.

Finally, thank you to PerkinElmer for their sponsorship of today's webinar.

Slide 3

Now, I'd like to introduce our first speaker for today, Dr. Elaine Mardis. Dr. Mardis graduated with a Ph.D. in chemistry and biochemistry from the University of Oklahoma. She spent four years as a senior researcher at BioRad Laboratories in Hercules, California before joining The Genome Institute at Washington University School of Medicine. As director of Technology Development, she has helped to create methods and automation pipelines for sequencing the human genome. Dr. Mardis has research interests in the application of DNA sequencing to characterize cancer genomes and transcriptomes, and in using these data to support therapeutic decision-making. A warm welcome to you, Dr. Mardis.

Dr. Elaine Mardis: Thanks very much, Sean. It's great to asked to do the webinar and I'm excited to be included with the other two members of the panel, Mike Snyder and Phil Stephens. So I'm going to provide some introductory remarks and a little bit of an indication of what we've been doing along the lines of cancer sequencing in our work at Washington University trying to cover the areas of prognosis, diagnosis, and cancer care for patients.

Slide 4

To begin with, I had a couple of slides to just talk about the notions and logic and philosophy really around cancer genomics in particular focusing on the work that we've done with whole genome sequencing. This slide that's being shown currently is just a few
comments about why cancer whole genome analysis is easy. In particular, this has of course been greatly facilitated by the availability and scale of next-generation sequencing platforms that have been commercially available and scaling in output since about 2007.

In this work that I’ll describe today, what we really have is this innate ability to compare at a very high level the patient’s tumor genome to their normal genome. We’ve known well in advance of these types of instruments being available just by microscopy that cancer is a disease of the genome and in this very careful look and characterization with next-gen methods, we can very completely and comprehensively determine what’s new or novel about the cancer genome of an individual compared to their normal.

In addition, we have lots of supporting resources to help in this interpretation of the data in particular some good online information such as the Cosmic Database that tells us about frequently mutated genes and that database as well as others is becoming extraordinarily well populated as a result of large scale efforts. So they’re using next gen methods to catalogue mutated genes and other aspects of cancer genomics such changes in methylation, RNA sequencing and so on and a couple of the large scale projects. The Cancer Genome Atlas and the International Cancer Genome Consortium are mentioned here and they are really increasing the knowledgebase tremendously.

But as people will know just from this next slide there are a lot more reasons about why cancer genome analysis is challenging. I’ll just go through a few of them here. So most people will focus on solid tumors because that’s the largest portion of cancer cases worldwide, but what we know about solid tumors is that they’re rarely 100% tumor cells. Rather there are normal cells that are present to differing degrees and some tumor types are actually quite challenging in this regard because they have a very large amount of normal cells present. In some cases, to enhance the tumor content, you have to isolate tumor cells away by laser capture microscopy, which is what LCM stands for or by flow sorting.

Another aspect of cancer genome analysis that’s challenging and this will be mentioned by the other speakers as well is use of formalin fixation and paraffin embedding. This is historically something that’s been used by pathology to preserve the cellular structure but it
results in DNA and RNA degradation and that increases as time goes on so that older samples and FFPE are much more difficult to obtain high quality nucleic acids from.

There’s also two aspects that contribute to the coverage of the genomes that we’re sequencing namely chromosomal amplification and aneuploidy, determines how much sequencing we have to do to make sure that even the diploid regions of the genome are covered.

Cellular heterogeneity, this notion that not all cells contain all mutations and in particular the mutations that are druggable or actionable, which will be covered by most speakers today, might be present actually at low levels especially in primary tumors but you would want to be able to detect them if they’re there.

Then switching gears to liquid tumors or leukemias, myelomas, etc., when you take those types of tumors, they’re contained in the blood and so you need to compare it to normal, which is typically a skin biopsy. But this may be challenging because just like the first example for solid tumors, in this case the normal may actually contain a lot of the tumor itself if the high circulating white count for the patient is there. So in those cases, we often go to either cheek swab or a mouthwash sample, but these are challenging as well because there are bacterial genomes that come along for the ride in those particular cases.

**Slide 6**

So having sequenced now through over a thousand cancer tumor normal comparison by whole genome methods, this slide is just meant to illustrate that essentially for every patient we find different combinations of point mutations, structural rearrangements, and copy number alterations. These circos plots as they’re called are representative of just the myriad of different cancer genomes that are present in which we are left to interpret in terms of which mutations are so-called drivers versus passengers, a very challenging aspect of this work as well.

**Slide 7**

We started our examination of tumor normal genomes early on with sort of old school or old technology methods as exemplified by this paper that we published in PNAS 2004 with William Pao and Harold Varmus and others and their group at the time at Memorial Sloan-Kettering, really beginning to understand for the first time the
therapeutic response to cancer in terms of new small molecule inhibitors such as tyrosine kinase inhibitors.

In this study, we used old technology such as PCR and capillary sequencing and then newly minted human genome to design PCR primers to amplify out specific genes such as the epidermal growth factor receptor shown here. To illustrate that in particular about 80% of patients who responded to these drugs, these tyrosine kinase inhibitors of which Iressa listed here is one, actually had EGFR mutations in the tyrosine kinase domain. As most people on the phone understand, today this has now moved forward as a typical clinical workup for lung adenocarcinoma patients is to determine EGFR mutation status because of the directability to correlate that mutational status with response to tyrosine kinase inhibitors.

Slide 8

If you jumped forward then about six years in time, we’re really at a point now where we are approaching diagnostic cancer sequencing from the very comprehensive aspect that’s illustrated here. This is an illustration that came from a Nature Genetics review that I wrote last year. It really shows this notion that rather than solely looking single genes or, you know, in isolation, we get a much more comprehensive picture for each patient if we isolate DNA and RNA from the tumor, DNA from blood, this is a solid tumor example, construct libraries and do sample quality control for example using SNP arrays for the genomic DNA and produce an exome capture as well as a whole genome sequence as well as RNA sequencing from the tumor.

So you just see the coverage levels listed here that we typically are pursuing in an effort to as shown at the bottom of the slide really integrate the data across whole genome, exome, and RNA sequencing to give the fullest possible picture not only of genomic aberrations, but also of their manifestation in the RNA population that’s produced in the tumor cells.

Slide 9

So the next slide just gives a visual of kind of how this might look in an idealized way where you can see the alignment of normal and tumor for whole genome, for exome and then for whole transcriptome of the tumor only as you work your way from the top to the bottom of the slide. Across the top, we just have a pseudo-reference genome depiction here with a green gene and a brown gene where what you can see is that for the whole genome, you’re
able to identify for example specific point mutations as shown in red and blue, intronic or intergenic sequence change such as shown in green and then aberrant read per mapping such as shown in pink at the far left for the tumor, which is our fundamental approach to identifying these structural variants that often lead to deletions for example or translocations that produce fusion genes.

In the targeted sequencing or exome approach that’s shown in the middle, we do have the ability to pick some of these aspects such as the red variant that’s been identified in the tumor. But in case of the structural variant or of course anything that falls outside of the region that’s been not captured by the hybridization reagent, we’re completely unable to detect those types of events by just focusing in on specific portions of the genome.

Lastly, the whole transcriptome sequencing shown at the bottom gives a nice example of the green gene variant being well expressed and very evident from the read repopulation that’s mapped to that gene, but by contrast the brown gene really has no representation and indeed no expression whatsoever in that tumor. This is a discriminator that tells us which mutated genes to pay attention to as we proceed in our analysis.

**Slide 10**

So we then go forward to combine these data and annotate the somatic alterations that are identified. Several years ago, we proposed this tiering system, tiers 1 through 4 represented on this pie graph where the exome is about 1.3% of the genome in red, the conserved or regulatory region about 9%, and about 42% of the remainder is actually unique sequenced but not yet annotated. Then you can see that a large proportion of the genome as we all know is highly repetitive. This becomes important in a second approach to validate these variants that are identified in the tiers 1 through 3 whereas the ability to validate in tier 4 is really not possible.

But in addition to the point mutations and other alterations as shown in this figure on the right-hand part of the slide, this ability to identify structural alterations is quite important. In particular, this is a figure from a paper that we published in the Journal of the American Medical Association in 2011 identifying a patient from a research study who actually had a previously undetected TP53 deletion as shown in the upper panel for her tumor, for her normal genome rather. Then this went to homozygosity in her tumor as shown on this second panel down with the read depth illustrating
that by annotating the region, we were able to identify that exon 7, 8, 9 were completely missing in the tumor and that by definition she was a lead from any patient due to the situation in her normal DNA.

[0:15:09]

We take all of these information and combine in the circos plot, which I mentioned earlier where we can see here one up close identifying that the chromosomes are labeled from number 1, the large brown chromosome at the upper right of the circle all the way around the circle for the autosomes and the sex chromosomes that point mutations are listed on the outside of this circle. The gray area in the center just under the colored bar shows the various copy number aberrations and then any of the lines inside the circle illustrate deletions, translocations, and so on forth. As I mentioned earlier for every cancer patient, this is a completely different story.

Slide 11

So we’ve used this information in a variety of ways that are diagnostic and prognostic, I’ll just illustrate those with a few quick examples. Here’s another paper from that same issue of JAMA where we identified a patient who had from pathology examination what looked like acute promyelocytic leukemia typically caused by a translation between chromosomes 15 and 17, but in this patient’s case, that chromosomal translocation could not be identified. The reason why is we were able to ascertain by whole genome sequencing is illustrated in this figure where she actually had an insertional event that juxtaposed PML the first three exons with the remaining three prime portion of RAR alpha so that the net result this production of PML-RAR alpha fusion that occurs normally with a translocation in acute promyelocytic leukemia in fact was what was driving her disease.

As a result of this, she was actually treated differently by the conventional approach to acute promyelocytic leukemia and she remains disease free about three years after this work was done.

Slide 12

In the next example we’ve looked at the comparison between the primary presentation of AML, which you can see from this figure of our Nature 2012 paper can be quite heterogeneous. This patient has four different subclone populations that we’ve been able to detect. Only one of which survived though clinical remission where most leukemic clones are lost. But one survives, acquires additional mutations in the relapse presentation of the patient’s disease.
By looking here at the transition/transversion ratio and across seven other patients, we were able to illustrate that DNA damaging chemotherapy used in these patients to get them into remission actually leaves a signature of DNA damage on the resulting relapse population.

The other conclusion from this paper is cited at the bottom where we identified either multiclonal or monoclonal primary disease in these eight patients, but in every case relapse always involved new mutations. However, some of these we have still not been able to ascertain the relevance of the mutation.

Then moving forward now to just finish up, we’ve begun developing a clinical genome analysis and interpretation pipeline. This is shown here and really is the result of the hard work of Malachi and Obi Griffith assisted by Scott Smith in producing this very complex analysis pipeline. As you can see in blue on the left of this slide, we’re taking in all of this information that we are generating from the approaches that I’ve already talked about, putting it through some initial functional annotation as well as trying to identify the activating or driver mutations in the cancer. I think Phil will talk about this a bit more in his presentation just in terms of how difficult this truly is to do to identify the activating or driver mutations that is.

We then used this information in the context of not only the genes that are identified but also the pathways in which they participate, which is a critically important way to look at the data since gene products never work in isolation. The way that we examine these is illustrated by the orange panel on the right, which gives you the beginning indications of all of the different approaches, databases and do on that are being utilized in this very complex analysis of for example what can be identified to approach these pathways that are being upregulated in that individual’s cancer. Coming out of that then we identified the clinically actionable events by looking at website are databases rather such as drug bank or PHarmGKB or clinicaltrials.gov. We can come up with a report that can be passed to the oncologist to give them information about what drugs might be applicable to that patient’s treatment.

This next slide really just shows for illustration purposes the application of this pipeline as published in a paper in Cell last year.
looking at nonsmall lung cancers for smokers and never smokers. I won’t spend a lot of time on this slide. The paper is in publication. You can take a look at it. But in particular, we just point out to the audience that the green bars represent over expression of genes, which is obtained from the transcriptome data and this particular case shows a large number of gene targets being identified for never smokers. So transcriptome data in particular can become important especially in tumors such as without a known oncogenic agent such as cigarette smoke where the mutational load maybe quite low and there aren’t a lot of apparent candidates from DNA sequencing.

Slide 15

So just to finish up these thoughts, some of the challenges that exist are that targeted therapies often lead to therapy resistance. If we go back to the EGFR treatment that we first identified mutations for in the mid-2000s, what we now know a few years later is that most of these patients do respond well to the tyrosine kinase inhibitors and achieve remarkable relief from their tumor burden, but also progress to metastatic disease while on therapy due to acquired resistance. This brings up the specter of the need for repeated testing in cases where biopsies can be obtained and that’s actually medically sometimes quite challenging.

Slide 16

Another story that’s come out of our work and reflects yet another challenge is exemplified by the story of Lukas Wartman who is affected by B cell ALL, acute lymphoblastic leukemia. Our work examining his tumor genome and transcriptome by the methods I’ve already talked about indicated he would respond to Sutent due to FLT3 over-expression of the wild type gene. But Sutent is not an FDA approved drug for use in ALL treatment and even in spite of two compassionate use request to Pfizer, he was turned down for access to this very expensive drug. Fortunately, of course, we were able to put together our money and help him to buy the drug and he is alive today because of that.

Slide 17

So just one last thought and I will finish up and hand off to Mike Snyder and that is that I think there’s a very exciting new area developing around personalized immunotherapy. You can see the approach that’s shown here where we’re combining sequencing of the tumor by RNA capture through an exome reagent to identify genes that have missense mutations and are being expressed. We combined that with the class 1 peptide binding predictions for that
patient’s HLA class 1 and use a portal on the web called NetMHCpan to prioritize potential tumor unique antigens for that patient alone. We then validate those and go through it in vitro functional evaluation using leukapheresis specimens from that patient to identify whether they will be able to activate T cells around this particular set of stimulatory neo-antigens as we call them.

This is very early on right now but I think provides an exciting example that may be able to obtain durable responses for patients with very little therapeutic side effects. But we still have a lot of work to do to get this entire workflow sorted out. In the meantime, we remain very excited that this will be a tractable solution for patient and that they won’t encounter the therapeutic resistance that I talked about earlier.

Slide 18
Sean Sanders:
Fantastic. Thank you so much, Dr. Mardis, for a really great introduction and kicking us off in this webinar. We’re going to move right on to our second speaker as you mentioned Dr. Michael Snyder.

Slide 19
Dr. Michael Snyder:
Excuse me, thank you for having me here. So I thought what I would do is just start by reiterating some similarities to what Elaine just said. That is we’ve been doing quite a bit of cancer genome sequencing, a variety of different types.

[0:25:08]
Slide 20
To walk you through two examples of these, if we go to this slide here, these are two examples from metastatic colon cancer where we did whole genome sequencing from FFPE sections, which as Elaine pointed is a bit challenging, your coverage is not quite as
uniform. We did deep Illumina sequencing on these and then with copy number and other somatic variations.

Slide 21

These turned out to suggest the therapeutic avenues based on the information so as shown on this next slide here. Basically, this one individual it turns out from the deep sequencing had amplifications in EGF receptor gene, which is not very common, although the pathway is activated quite a bit in colorectal cancer. The gene itself is not known to be altered very often yet the sequencing revealed that in fact this region is amplified along with another region that is targetable CDK8. So based on this information, we have used drugs that are available to the EGF receptor pathway that are now being used to treat this individual.

I might also point out of the prognosis of this patient was informative from the genome sequence. He had a deletion in the short arm of the chromosome 8, which is associated with metastatic cancer and that was useful to see.

Slide 22

Here’s another example from another patient that had amplification of CDK8 gene, which again is an actionable result. There are now drugs available for CDK8 that can be used to treat this patient.

So those relatively simple cases that came out from genome sequencing. What happens when you get a complex one and so I thought I’ll walk you through one example that we have here.

Slide 23

This is esophageal cancer. This is actually a research patient that’s one that we did not use the information for treating the patient.

Slide 24

Basically, they had esophageal cancer and we did laser capture to pull out, excuse me, the cancer cells along with the adjacent normal tissue and then did in this case complete genomic sequencing along with as you’ll see in a minute RNA sequencing which turned out to be quite informative for what was going on here.

Slide 25

So basically, this individual as they say we did deep genome sequencing from and it turns out they had over 20,000 small nucleotide variants, single nucleotide variants and indel so small
mutations and over a thousand structural variants had mostly copy number changes along with some fusion gene changes and some inversions as well. So how do you go from this complex mixture to trying to find potential driver and activating mutations? Well RNA sequencing again was very useful in integrative analysis here.

Slide 26

So as shown on this next slide, just doing a very simple comparison of the amplified regions along with their expression. So the expression is shown on the Y axis and the amplified region is on the X axis. There is a general correlation between amplification and expression, although there are some interesting exceptions that are worth pursuing from a scientific standpoint.

Slide 27

But when you start zooming in on these, you do see that the upregulation of genes if you look at the genes whose expression is changed either in the amplified sample as we’ll get to later in the single nucleotide variants, you can get informative information. In this case, the upregulated genes were very consistent with cell cycle genes and cell division genes being upregulated as they’re responsible for driving proliferation of this tumor.

Slide 28

One of the interesting results is if we not only just look at upregulated genes but look at genes that are undetectable in the normal sample but then upregulated in the tumor, we actually find metastatic genes show up. So these are things involved in cell migration, extracellular signaling, extracellular space. So you can actually get some information from the expression of again the types of prognosis and things that are going on in this individual.

Slide 29

Likewise, if we look at the down regulated genes as shown on this next slide, you actually find that cell differentiation genes are down regulated in the tumor relative to the normal. So the expression information was very useful just even looking at the structural variant.

Slide 30

We can also apply this to the single nucleotide variants as well where we drill in and look at the changes, the gene whose mutations are affected by the single nucleotide variants. Again, here we zoom in on those 22 inner circle genes. These are things that are actually
upregulated in the tumor relative to the normal for these small nucleotide variants. When zooming in on those, we actually found that there are quite a few interesting potential driver mutations in this. There are five candidates in there, one of which is recurrent in esophageal and other cancers and had not been reported previously.

[0:30:22]
So we’re using expression information if you will to filter the list. It’s basically extension of what Elaine talked about to be able to get better information about the kinds of changes and what their effect might be. This turned out to be particularly valuable in this case because there were three fusion genes that you might have thought were helping drive this, now maybe they were at one point in the cancer but these fusion genes are not expressed in the cancer, it fell to the normal. So it’s not clear that they are in fact having relationships and the expression was very, very useful for sorting all that out.

Slide 31
So that’s my very brief story. In conclusion, you can find druggable targets by whole genome sequencing. Transcriptome we think is an effective filter both for upregulated and down regulated genes and I guess I’ll leave it at that. Given the recurrent rates of these, we think that doing combined approaches will be very, very useful by bringing all sorts of information Elaine talked about.

Slide 32
I just wanted to acknowledge people who did the work. Hanlee Ji and George Fisher were collaborative in this effort and Hassan Chaib in my lab was responsible for this for the colorectal cancer patients and the esophageal cancer work was done by Jason Reuter.

Sean Sanders: Fantastic. Thanks so much, Dr. Snyder. Very interesting stuff and we’ve got a number of questions in, which I’ll be posing to you later about your work.

Slide 33
We’re going to move on to our final speaker now and that’s, Dr. Philip Stephens.

Slide 34
Dr. Stephens studied at Oxford University in the United Kingdom, where he received his Ph.D. Now, as vice president of Cancer Genomics, Dr. Stephens leads research and development at Foundation Medicine. Since joining the company in early 2011, Dr.
Stephens has overseen the development of FoundationOne, a comprehensive next-generation sequencing diagnostic assay that accurately profiles the entire coding sequence of over 200 cancer-related genes in a CLIA setting. Prior to joining Foundation Medicine, he held various senior positions during his 11-year tenure with the Cancer Genome Project at the Wellcome Trust Sanger Institute in the United Kingdom. Thanks for joining us, Dr. Stephens.

Dr. Philip Stephens: Thanks, Sean, for the kind introduction and I would also like to echo my appreciation in being able to participate in this webinar today. My presentation will focus on meeting the challenges of translating NGS into the clinic.

Slide 35

It is becoming increasingly clear that therapies targeting the alteration that drive an individual patient’s disease can show remarkable clinical utility when compared to non-targeted therapy. But as the number of clinically relevant genomic alterations increases, matching the correct targeted therapy to the correct cancer patient presents some tremendous diagnostic challenges.

Slide 36

The main challenge as you’ve heard from the two previous speakers is that the genomes of solid cancers can become very complex in nature. Now whether this is due to mutagenic exposure, genetic instability, genomic instability, or perhaps a combination of all three it is now believed that the genomes of the common epithelial cancers such as breast cancer, lung cancer, and ovarian cancer may have as few as one to two clinically relevant cancer genomic alterations buried amongst a sea of tens of thousands of passenger mutations that have absolutely nothing to do with the development of that cancer – sorry, with the treatment of that patient.

Slide 37

Now while the number of clinically relevant cancer genes within in an individual patient maybe low, across an entire disease indication for example nonsmall cell lung cancer, they can be very high.

Slide 38

In addition, there are at least five clinically relevant types of genomic alterations including base substitutions, short insertions and deletions, focal gene amplifications, homozygous deletions and gene fusions and each of these pose unique and different diagnostic challenges.
As you’ve heard, you know, I’ve got low tumor purity that is present in many clinical specimens requires diagnostic tests with extremely high accuracy.

Again, you’ve heard that many clinical cancer specimens are formalin fixed and paraffin embedded which can damage DNA and also many include needle biopsies, which are very, very small.

Finally, to maximize clinical utility, we’ve got to take this complex information and translate it into a comprehensive but easily interpretable report to a busy community oncologist who may only have 30 seconds to review that report prior to seeing the patient and who may have trained in the pregenomic era.

So our solution is one comprehensive genomic profile to simultaneously detect all classes of clinically relevant genomic alterations within a single assay.

It is focused on the 236 known clinically and biologically relevant cancer genes, although some of the data that I will present later in this talk will be from a previous assay focused on 182 cancer genes.

It is validated for high accuracy from routine clinical specimens from as low as 20% tumor nuclei.

It requires very small amounts of tissue from routine FFPE samples including needle biopsies and we’ve recently been able to get this to perform robustly on fine needle aspirates.

Finally, we’ve developed some customized and highly sensitive computational biological algorithms, which again have been validated for high sensitivity and specificity in samples with low purity.
Slide 47

So this slide it’s a schematic overview of our workflow, which currently takes 14 to 21 days but we were hoping to drive down consistently to 14 days and below by Q3 this year. Essentially, we do a sample preparation of nucleic acid extraction. You make a library and hybrid capture the 3000 odd exons and introns that we want to sequence.

Again, this really doesn’t work straight out of the box. This has required extensive optimization to get the high but also the uniform coverage that’s required to maximize sensitivity. Our sequencing box is currently the Illumina HiSeq 2000. We have a highly customized analysis pipeline, which identifies the genomic alterations, which are then used to populate the clinical report that is sent back to the ordering physician.

Now the take-home point from this slide that translating, you know, and especially of research grade NGS into a clinical cancer diagnostic assay requires significant optimization and investment.

Slide 48

So prior to use in the clinic, any diagnostic assay should undergo a rigorous analytical validation, but this posed us with a problem. How does one go about validating an assay that interrogates over 1.2M nucleotides and can be somatically altered by multiple mechanisms across the entire span of that assay?

This slide here is an overview of our approach. Essentially, we took up the ten HapMap cell lines and then we poured them in such a way that the 2056 base substitutions present throughout our assay were there in a wide range of mutant allele frequencies and we biased this deliberately toward a low mutant allele frequency.

Slide 49

When we sequenced these pools in a blinded fashion multiple times, we were able to achieve an accuracy of greater than 99% for both sensitivity and specificity for a base substitution that’s there and an allele frequency between 5% and 100%.

Slide 50 to slide 51

We performed similar validation experiments for insertions and deletions, copy number alterations and again it says that sample preparation and algorithms optimized to detect these genomic alterations with high accuracy that is key to working with clinical
specimens with as few as 20% tumor nuclei. So if we look at our performance for short insertions and deletions in this wide range of 1 to 40 base pairs, it has an allele frequency of 10% to 100%. We have a sensitivity over 98% and approaching 100% specificity.

To put this result into context, when we took some of the best publicly available algorithms from the genome centers, not a single one of them was able to get either a sensitivity or specificity above 40%. Perhaps this should not surprise us because these algorithms in the sample population has been optimized for research grade specimens, which typically contains very high amounts of tumor nuclei often in excess of 80%.

For the final part of this presentation, I want to give an overview of the statistics of first 2200 cases that were profiled in our CLIA certified, CAP accredited lab. The first statistic is that we have a failure rate of only 4.9%. Now this is quite remarkable because if you compare to this to some of the single biomarker tests they have a failure rate of somewhere between 10% and 20%. The number of samples with at least one actionable alteration was around 76%. The mean number of alterations per sample was 3 and the mean number of actionable alterations per sample was around 1.5.

And we have a very conservative definition of actionable, either an FDA approved therapy in that tumor type or another tumor type or an open clinical trial of a therapy targeting that alteration.

If we look at the tumor types that comprise the first 2200 specimens, you know, they were from all stages of disease from early diagnosis to late recurrent metastasis.

We next went on to look at, you know, the clinical relevance of the alterations that we identified and we compared this to a hypothetical test combining AmpliSeq, SNaPshot, OncoCarta, OncoMap, HER2 FSH, and EML4-ALK and we assume 100% sensitivity. When we did this, our assay identified nearly four times the number of actionable alterations.
The reason for this could be seen on this slide. We call this the long tail effect. This slide shows only 62 of 155 most commonly altered genes and while the rate of alteration within a single patient or a disease indication maybe very low to some of the genes on the right, cumulatively they account for the vast majority of the actionability across all solid tumors.

Slide 57

In our third clinical patient, we identified a novel KIF5B-RET fusion. In collaboration with Pasi Jänne, we went on to show that this gene fusion was recurrent. It was oncogenic and KIF5B-RET transformed cells are sensitive to RET inhibitors.

In the first 2200 cases, we’ve identified RET fusions with multiple partners in around 2.2% of nonsmall cell lung cancer patients and we know of at least two patients with RET fusions that have responded to multi kinase inhibitors that inhibit RET.

Slide 58

So ERBB2 amplification and over expression is routinely tested for in breast and gastroesophageal cancers.

Slide 59

We were able to identify ERBB2 alterations in 102 specimens across 13 tumor types and importantly over 40% of these alterations were either mutation or fusion in non-ERBB2 amplified specimens, i.e., these would not have been picked up with currently used methodologies.

Slide 60

Elaine Mardis’ team has recently demonstrated ERBB2 mutations that cluster in the kinase domain in breast cancer.

Slide 61

We were able to identify these and a novel ERBB2-GRB7 fusion. When we looked at the clinical and the genomic data, we found around a 10-fold enrichment of ERBB2 mutation or fusion in the CDH1 mutated disease compared to the CDH1 wild type disease.

Slide 62

This slide here is showing you ERBB2 alterations for all tumors. You can see a cluster at serine 310. Again, we know of at least one invasive lobular breast cancer patient with an ERBB2 mutation that has responded to an ERBB2 TKI.
So my final data slide moving on to EGFR, alterations were identified in 151 cases across 13 different tumor types.

When we plot, we can plot the alterations on EGFR protein but there’s two points that I would like to make from this slide. One, that around 1 in 5 nonsmall cell lung cancers harbored an EGF alteration that was missed by other diagnostic assays and we know of a breast cancer patient with an EGFR sensitizing mutation of exon 21 that had had a good response to an EGFR TKI.

So to conclude, I think that translating a research grade NGS into a clinical cancer diagnostic assay requires extensive optimization and investment.

I believe that rigorous analytical validation to determine test accuracy and reproducibility is required prior to clinical use.

That the complex information must be conveyed into an easily interpretable comprehensive report to maximize utility for busy oncologists.

The initial results with our assay are encouraging, 76% of clinical cases harbored at least one actionable genomic alteration.

We believe that our assay can identify nearly four times the number of targeted treatment options compared to Hotspot assays.

Finally, the patients who have exhausted treatment options may show dramatic responses to targeted therapies identified as rational candidates in our clinical reports. And that’s the end of my presentation.

Fantastic. Thank you very much, Dr. Stephens, and we’re going to move right on to the question that we’ve received from our
audience. We got some fantastic questions so I hope everyone can stick with us to answer them.

Slide 72

Just I want to say thanks to our speakers for the very informative and engaging presentations. A quick reminder to those watching us live, if you still want to submit questions, you can just type them into the text box and click the submit button. If you don’t see the box on your screen, just click the red Q&A icon and it should appear.

So the first question that I’m going to put out to all of you and perhaps I’ll have Dr. Mardis answer this first. What is the current thinking around how tumor heterogeneity related to sequencing data output and how sequencing sample prep may affect this relationship?

Dr. Elaine Mardis: So heterogeneity sort of takes two forms in my thoughts about it in general. So one is the type of heterogeneity I mentioned during my presentation, which is cellular heterogeneity in that not all tumor cells contain the same genome if you will or the same genotype. The other type of heterogeneity is maybe what could be called regional heterogeneity namely lager tumor growth, mass, you get differentiation across the tumor mass itself where the areas of necrosis, advancing margins and so on and so forth.

So in terms of sample preparation, you know, the conventional thinking, although the number of examples is incredibly small, is that the larger the tumor mass, the more sites of the tumor you may want to sample to get the fullest picture of heterogeneity in a regional sense. In terms of the cellular heterogeneity if you think maybe about smaller tumors firstly in that regard with a single sample available, there the idea would be that you would want to sequence at higher depth to be able to pick out these very low lying mutations such as the ones that Phil was referring to where the lack of sensitivity for research grade software usually does tend to miss those as well as the fact that the overall depth on whole genome data is typically not entirely adequate for high sensitivity for those types of mutations.

Sean Sanders: Dr. Snyder?

Dr. Michael Snyder: Yeah, I agree completely with Elaine. I think deep sequencing is useful for capturing the heterogeneity within a sample and multiple samples is useful for capturing what’s going on especially for
metastatic cancer. I think in the future we may see single cells sequencing especially for liquid types of tumors and cancers. So I think that will be very, very valuable for understanding heterogeneity and also for ultimately adding back so called normal cells.

Sean Sanders: Dr. Stephens?

Dr. Philip Stephens: I agree with both Mike and Elaine, but I’d just like to point out that it’s not the high sequencing depth that’s required. It’s also a high but uniform depth so you can actually look for heterogeneity across the entire genome. There’s just one point that I just also like to make. I think from the work, a lot of it coming from Elaine’s lab others from TCGI and the big genome institutes, you know, while heterogeneity does exist across the majority of tumor types, a lot of that heterogeneity is within the passenger mutation. The actual driving mutation, the one, which is driving the oncogenic process within the tumor in that patient, and the ones, which are the most therapeutically targetable, there does appear to be a growing body of literature that there is less heterogeneity within these driver alterations. Indeed if you think of the renal cancer in the New England Journal of Medicine paper, they showed enormous heterogeneity but in that ambiguous driver alterations of renal cancer there was a very high concordance. Even when there was a discordance, there was convergence evolution. For example, there were multiple mutations in set B2 and KDMD5 across the tumor but pretty much every single site assayed had one.

[0:50:35]

Sean Sanders: Excellent. Dr. Stephens, I’m going to stay with you for a second because just to get a question out of the way that I’ve had from a number of viewers is are the clinical markers that you use available in the public domain? If so, where can they be accessed?

Dr. Philip Stephens: Okay. So when we publish, we publish exactly what we put on our assay. So our current assay we screen the entire coding sequence of 236 cancer related genes, which is around three and a half thousand exons and we also sequenced the introns of 50 genes, which are frequently rearranged in cancer. So we have those publicly available.

The validation exercise that I presented today that is currently under review at a very high profile journal. We deliberately used publicly available cell lines and tools so people could actually recapitulate our
results and use this tester as a benchmark to really determine how analytical validation should be performed for NGS sequencing.

Sean Sanders: Perfect. Dr. Snyder, I’m going to put this question to you and maybe Dr. Mardis would like to chime in as well. Is there any way to capture the mutational heterogeneity in the tumor without isolating individual cells?

Dr. Michael Snyder: Well technically, yes, you can. I guess what you can do is if you have linked technologies or so, there’s some experimental ways of doing this, although I guess they ultimately do isolate things, either molecules or cells in a subcellular environment. But certainly if you do deep sequencing you’ll get a pretty good feel for the heterogeneity that’s within a cancer just by the frequency with which those mutations are present because they’re at varying degrees, the more heterogeneous, the more varying degrees.

Sean Sanders: Dr. Mardis, anything to add?

Dr. Elaine Mardis: Mike’s exactly right. If people want to an example of how this could be done, they could look at the Ding et al reference from Nature 2012, that figure that I showed for the AML de novo relapse. That’s exactly the approach that we utilized for those eight patients is a deep sequencing approach across the tier 1 through 3 mutations that we identified in their tumors looking at the changes of fluctuations in prevalence of all of those mutations through the transition from de novo to remission to relapse.

Sean Sanders: Dr. Snyder, I’m going to come back to you again. You mentioned in your presentation this viewer asks that differentiation genes are down regulated I guess in some of the samples that you were looking at. Are you implying that the cell is more stem like in this case and if so, do you have any data to this effect?

Dr. Michael Snyder: Well I think the data mostly comes from the RNA expression results, which we think is probably one of the best indicators of that. The genes are down regulated. I do think it’s consistent with some of the histological scenes and things that have been put on. Not always the same samples but samples of a similar type. So I do believe they are stem like.

Sean Sanders: Great. A question for all of you but I think I’ll start with Dr. Stephens because this I guess really refers to what you’re trying to do with your assays is are there possibly cancer genomic hotspots and might
it only be necessary that somebody sequence these hotspots when looking at a variety of patients?

Dr. Philip Stephens: So that’s a great question and, you know, I think I advocate for all genomic testing in cancer patients. I think for the patients who don’t get access to a comprehensive assay such as ours or whole genome and whole exome sequencing, I think that the hotspot assays have tremendous clinical utility for patients. When you think about the patient with nonsmall cell lung cancer, you know, only 40 of them in the United States get any biomarker testing at all for EGFR, KRAD, and EML4-ALK. I just think that as I’ve hopefully demonstrated within my presentation that by doing a comprehensive genomic profiling of around 200 cancer genes, you can actually identify four times the number of therapeutic treatment options for patients with cancers across all the genotypes.

[0:55:13]

Sean Sanders: Dr. Snyder?

Dr. Michael Snyder: Yeah. I mean I think hotspot directed sequencing does have a lot of value but I have to admit I do believe going whole genome has value as well. For example you can get pharmacological information by analyzing the cytochrome P450 genes and things like that. I think the combinatorial effect, which Phil alluded to in his talk, is going to be very, very valuable. So I know a lot of it’s research right now but ultimately gaining all of the information about the exact mutations that are there as well as the combinatorial information is ultimately going to be extremely valuable and that’s why I’m a strong advocate for getting whole genome information not just hotspots.

Sean Sanders: And Dr. Mardis, do you have anything, any words on that?

Dr. Elaine Mardis: Yeah. Well Mike and I drinking the same Kool-Aid I think here. You know, the bottom line is the panel tests are great. They are easier to interpret, you know. But even from the numbers that Phil presented, right, 76% of patients are getting information back, it makes me wonder about the 24% that aren’t and sort of what we’re missing. I think what we’ve tried to illustrate most recently in a Nature paper this summer in breast cancer is that while it’s informative for many patients to look at the significantly mutated genes, you know, as I tried to point out in my presentation pathway hits don’t necessarily always fall in those so-called hotspots. So you can have a pathway upregulated by a number of hits that may not be statistically significant but to that patient’s tumor they’re incredibly significant.
So I think comprehensivity is best. It’s just more expensive so that’s a problem.

Dr. Philip Stephens: So I agree with you, Elaine, but I just would like to make one point of caution is, you know, that if you have all clinically relevant genomic alterations on an assay and you want to make a treatment – a doctor wants to make a treatment decision, when you then go off into the whole genome and whole exome sequencing space, yes you can find a lot more but how do you interpret that mutation perhaps in a lung cancer which may have tens of thousands of genomic alterations? How do you know which one of these genomic alterations are driving the disease?

I’ll answer that question myself. I think by integrating the RNA sequencing can give you guidelines. But for a doctor to make a treatment decision and then to apply these to tens and perhaps hundreds of thousands of cancer patients annually I think that’s where the real challenge is. I think you know, I agree with you in certain – you know, perhaps in a research setting that whole genome and whole exome sequencing maybe better but translating this to large number of patients I see tremendous challenges.

Dr. Elaine Mardis: Right. I wasn’t talking –

Dr. Michael Snyder: Yeah, well your point – oh, go ahead, Elaine.

Dr. Elaine Mardis: I wasn’t talking about unknown mutations. All I was talking about is that you can alter a gene in a number of ways firstly and secondly that pathways are more informative than single point mutations. I mean I firmly believe that and I think the RNA expression being coincident with the hypothesis of pathway upregulation is incredibly powerful. I’m not suggesting that gets applied across thousands of patients. I’m suggesting that this is the end of one diagnostic. So for that particular patient, it may be incredibly important.

Dr. Philip Stephens: Uh-hum.

Sean Sanders: Great. Dr. Snyder, anything that you want to jump in with in?

Dr. Michael Snyder: Well again, I guess to use Elaine’s phrase, we are drinking the same Kool-Aid. I do think – I mean Phil’s point is well taken, doctors want to standardize tests and that’s why I think his assay is very, very successful. I do think there is a lot of other information in a genome sequence that can be useful and I think if you’re that patient, you
want that information. So I do think a first pass you know, after the hotspots can be achieved very rapidly but I do think the actual information will be valuable. I think this is why RNA sequencing and these sorts of things will ultimately be part of standard of care in the future as well.

Sean Sanders: Great. Well we’re just at the top of the hour but I’m throwing one more question at you and that is so we’ve talked about whole genome sequencing and exome sequencing and looking at the transcriptome and RNA. But broadening that a little bit and, Dr. Snyder, I know this is possibly right in your valley work is what about the other types of data that we can possibly add to this such as epigenetic data, proteomics, metabolomics? Maybe each of you can speak a little bit about how you see that happening in the future? So maybe we’ll start with you, Dr. Snyder?

Dr. Michael Snyder: Yeah. Well obviously, this is our cup of tea. We think all these will be very, very valuable, metabolomics, proteomics and especially epigenomics, which is a little bit easier to interpret right now. But it is all research right now, I’ll have to say, that is we have to collect the information and see how well it correlates with prognosis and treatments and things like that. But I do think ultimately it’s going to be incredibly valuable because the combinatorial information again is much more valuable than any individual data type.

[1:00:30]
Sean Sanders: Great. Dr. Stephens?

Dr. Philip Stephens: I complete agree. You know, our test at the moment is DNA based. We’re going to be adding a transcriptome sequencing component to it for our hematologic malignancy tests. But I think the challenge we have thinking about epigenomics, metabolomics is how you’re going to get this reimbursed for patients with cancer. I think that is the central challenge. So I think yes in an ideal world, you would want to know absolutely everything about this patient with cancer but if the cost then becomes prohibitive to be able to apply that to tens of thousands of patients, I think that’s a challenge.

Sean Sanders: And Dr. Mardis, you get the last word for today.

Dr. Elaine Mardis: Thanks. I do agree that the proteomic extension from DNA, RNA is going to be important. What we’re working on presently, as I think Mike is, is making that extension. So it’s not straightforward, the mass spectrometry instruments do not have the same dynamic
range as sequencing based approaches do and so there’s a bit of a translation in bioinformatic computational space that needs to take place. But I think in particular the proteomic prediction and then validation of our predictions in the research space will be an important precursor to the diagnostic use and in particular looking at protein modifications such as phosphorylation, etc., which is now much more possible. These new instruments is going to really I think open that up dramatically in research and ultimately there will be a translation into the clinic as Phil pointed out very sagely as long as the cost model is there and it’s time and cost effective for patient’s care.

Sean Sanders: Fantastic. Unfortunately, we are out of time for this webinar, but I would very much like to thank our speakers for providing such fascinating talks and very lively and interesting discussion, Dr. Elaine Mardis from Washington University in St. Louis, Dr. Michael Snyder from Stanford University, and Dr. Philip Stephens from Foundation Medicine.

Many thanks to our online audience for the questions you submitted. I’m sorry that we didn’t have time to get to all of them.

Slide 73

Please go to the URL that is now up in your slide viewer to learn more about resources related to today’s discussion, and look out for more webinars from Science available at webinar.sciencemag.org. This particular webinar will be made available to view again on-demand within approximately 48 hours from now.

We would love to hear what you thought about the webinar. Feel free to send us an email at the address up in your slide viewer, webinar@aaas.org.

Again, thank you so much to our panel and to PerkinElmer for their kind sponsorship of today's educational seminar. Goodbye.