

Host:

This is the podcast from *Clinical Chemistry*. I am Bob Barrett. Rapid advances in molecular diagnostics not only enable basic research, they also result in practical diagnostic tests. This central position of molecular diagnostics in translational research is the topic of the April issue of the journal *Clinical Chemistry*, devoted to molecular diagnostics at the cutting edge of translational research.

DNA sequencing is at the heart of molecular diagnostics and the April issue includes a review by Dr. Karl Voelkerding on next-generation sequencing. Dr. Voelkerding is an Assistant Professor of Pathology at the University of Utah and serves as Medical Director for Advanced Technology at ARUP Laboratories and is our guest.

Dr. Voelkerding, what is the history of sequencing, in comparison, how is it done today for clinical diagnostics?

Dr. Karl Voelkerding:

Well, the history of sequencing really begins in the mid-1970s, and in 1977, there were two very landmark publications. One introduced the sequencing method that became known as Maxam and Gilbert Sequencing. In the same year, there was a second publication that introduced the sequencing method that was referred to as DNA sequencing with chain-terminating inhibitors, and this landmark publication came from the laboratory of Frederick Sanger and it quickly became known as Sanger Sequencing.

In the years that followed, the Sanger Sequencing technology underwent a series of modifications and improvements that took it from its original format that used radioactive dideoxy terminator molecules for chain termination to fluorescently labeled dideoxynucleotides, and it moved from using slab-gel electrophoresis to capillary electrophoresis, and out of the sequential progression and improvement of the Sanger methodology came what is being used worldwide, both by research scientists and in the clinical diagnostic setting.

The Sanger Sequencing technology was essentially through the use of automated instrumentation developed by Applied Biosystems, has become widely disseminated. So it's that combination of Sanger Sequencing and automation and chemistry improvements that are used today for clinical diagnostics.

Host: What kinds of information do clinicians obtain from sequencing? Why do they use that for medical diagnostics?

Dr. Karl Voelkerding: The Sanger Sequencing technology provides us the actual sequence of a nucleic acid molecule, so the actual ATCG sequence, and it's used for a variety of medical diagnostic purposes.

For example, sequencing of the human immunodeficiency virus, the HIV virus, is done on a clinical basis to determine the sequence of the virus with respect to susceptibility or resistance to HIV antiviral medications.

Another area that is being used is to sequence a variety of different genes in the human genome that have been implicated in disease pathogenesis. An example would be sequencing of the cystic fibrosis gene in the setting of an individual with suspected cystic fibrosis.

One can begin to list a large number of areas where sequencing is actually employed, but genetic diagnostics, viral diagnostics have really led the way. Another area that many people may be familiar with is sequencing of the BRCA1 and BRCA2 genes, which are predisposition genes for breast cancer.

Host: What about next-generation sequencing, how does that differ compared to sequencing today?

Dr. Karl Voelkerding: Well, the paradigm of Sanger Sequencing is a single sample that has undergone a Sanger reaction of sequencing by chain-terminating dideoxynucleotides, which are then resolved, the differential length fragments are resolved in a capillary electrophoresis system.

(00:04:50)

The advent of next-generation sequencing has shifted that paradigm in which we now take DNA and create basically a library of those DNA molecules, which are then sequenced in a massively parallel fashion on a flow cell as opposed to a capillary electrophoretic platform.

This massive parallel high throughput sequencing approach, that's been referred to as "next-generation sequence," has with it the features of being very high throughput relative to the instrumentation platform

for Sanger Sequencing, and allows you to derive both qualitative sequence information and quantitative information regarding the population of DNA in the sample that you are sequencing.

Host: How has next-generation sequencing impacted biomedical research?

Dr. Karl Voelkerding: The interesting thing to think about in this regard is at the first peer-reviewed publication, using the first of several commercially available next-generation sequencing platforms, was published in 2005, and since 2005, there have been slightly over 600 publications that have come out of the research community using the next-generation sequencing technologies. These publications have spanned the spectrum of applications, including complete or full genome sequencing: everything from plants to microbes to human genomes.

It has been used to sequence the entire RNA complement of cells or tissues, referred to as whole transcriptome sequencing. It's been used to map binding of proteins to genomic regions and deriving the sequences that the proteins are binding to.

It's also been used to study populations of microbial organisms in everything as diverse as a sample of seawater to samples from the human gut.

So the high throughput nature of the technology and the ability to derive both quantitative and qualitative information has allowed investigators to pursue experiments that were either technically not possible to do or so impractical to do by traditional Sanger Sequencing that they were never attempted. So in that regard, it has definitely been a technology that has opened many, many investigative avenues as evidenced by the rapid number of publications that have come forth from investigators.

Host: How might next-generation sequencing be used for clinical diagnostic applications?

Dr. Karl Voelkerding: Clinical diagnostic applications for next-generation sequencing in many ways will follow the types of applications that have been developed for basic research studies.

For example, one can envision sequencing multiple genes in the human genome, all of which when they have mutations in them confer an overlapping

phenotype. This has traditionally been very challenging to do by Sanger Sequencing.

One can use this technology to examine the population of viruses that are circulating in an individual, for example, infected by the HIV virus or the hepatitis C virus, the technology can be used to enumerate and identify subspecies of the virus that may be resident in the individual that would confer drug resistance.

The technology has the potential, both at the research level, but going forward to sequence the genomes of individual tumors in patients that could be used to direct both diagnostic considerations, prognostic, and potentially therapeutic considerations.

So although we are in the very early phase of translating this technology from the basic research laboratory into the clinical laboratory, we are already seeing examples that have great future potential for clinical diagnostics.

Another area on this topic is the use of the technology to sequence cell-free DNA that's circulating in the blood circulation that can be isolated and sequenced, and this has been shown in two separate publications to be a mechanism to identify chromosomal abnormalities such as Trisomy 21 or Down Syndrome in pregnant mothers in a noninvasive fashion.

(00:10:00)

Host:

Are there more new sequencing technologies on the horizon?

Dr. Karl Voelkerding:

Yes, there are two things ongoing. One is that, of the commercial platforms that are currently available, each of these platforms are undergoing refinement, leading to improved chemistries and sequencing accuracies. They are also demonstrating higher throughput capabilities for each of the existing platforms through engineering modifications.

Furthermore, there are several additional technologies on the horizon that are being developed and will be commercially available within the next two to five years, and these technologies, in particular, may again take us to another level of

throughput and also promise potentially to streamline the overall technical processes.

Host:

Well, that's being said, in the not-too-distant future we'll be able to sequence a person's entire genome for less than \$5,000. When is that going to be a reality and how will the information be used medically?

Dr. Karl Voelkerding:

Yes, I think there certainly has been a lot of interest in this area. The technologies to allow us to sequence a person's entire genome for less than \$5,000 are still in their early stages.

Interestingly, there are several companies that anticipate that they will be able to do this within either this next year period or within the next two- to three-, to most five-year period. These technologies are either utilizing very, very high throughput approaches, with engineering modifications that are advances beyond the commercially available instruments, or they are going to be with technologies and instruments that are not yet commercially available.

The latter would include methodologies that actually will fall under the category of real-time DNA sequencing, so that either a preliminaries molecule will be actually synthesizing a DNA and it will be essentially monitored real-time using optics and fluorescent nucleotides that are being incorporated, or single molecules of DNA will be digested successively with essential exonucleolytic enzymes, and the individual basis will be channeled through a nano-pore structure and registered as a variation incurring.

So if we anticipate that this will actually occur, the most important questions will be, if we can do this, what will be the accuracy of the sequence that is derived, and if we assume that the accuracy is sufficient for diagnostic purposes, then how will we use that information?

I can think of several caveats to this. One is that currently we only know in detail about a very small fraction of the human genome. So the technology may come first before we have a full understanding of the genome. I think that will be how it will come to pass.

But that kind of information will allow us then to have a much better understanding of the genetic makeup

of the individual. We will then need to begin to correlate that with known clinical phenotypes, and probably one way to think about this is that an individual may have their genome sequenced in the year 2015, and we will have a certain amount of information that we can make sense of from that genome.

But as the individual proceeds through life and as the basic biomedical research community develops more information and understanding of various disease processes and the genetics underlying them, one can then revisit that individual's genome and essentially do a reinterpretation of the individual's genome based on the new medical knowledge. So I think it will be an iterative process.

We already have several genomes that have been sequenced, that are available in public databases, and you can draw a certain amount of conclusion from that information for that individual genome. But if we anticipate 10, 15, 25, to 50 years out, we will be on a constant reinterpretation of the genetic information.

Host:

Dr. Karl Voelkerding is an Associate Professor of Pathology at the University of Utah and has been our guest in this podcast from *Clinical Chemistry*. I am Bob Barrett. Thank you for listening.

Total Duration: 15 Minutes