TITLE: Sanger Sequencing

PRESENTER: Jason Y. Park, MD, PhD

Slide 1:
Hello, my name is Jason Park. I am an Assistant Professor at UT Southwestern Medical Center and Children's Medical Center Dallas. Welcome to this Pearl of Laboratory Medicine on "Sanger Sequencing."

Slide 2: Outline
This teaching session will describe the basics of DNA sequencing by the Sanger Method. I will describe the historical context of the original radioactive method and then I will compare the fluorescence-based method that is currently used in clinical laboratories.

Slide 3: Historical Overview
Sanger Sequencing is a method to determine the order of DNA nucleotides. Each subunit of the DNA double-helix has complementary basepairs. In DNA testing, the fundamental questions are: What are the bases present? And what order are they in?

The term 'Sanger' refers to Frederick Sanger, a British scientist who invented the method in the 1970s. Dr. Sanger was a well-established scientist who had already won a Nobel Prize in 1958 for protein structures. His DNA sequencing method resulted in his 2nd award of a Nobel Prize in 1980.

Slide 4: Key Components of Sanger Sequencing
There are six key components of Sanger Sequencing:
1. The DNA target of interest. For example, an exon or specific pathogenic variant of a gene;
2. Polymerase Chain Reaction (PCR) amplification of the target of interest;
3. A synthetic oligonucleotide known as a primer. The primer is necessary to target a specific region of interest;
4. DNA polymerase: this enzyme joins the nucleotides into a chain;
5. Unlabeled deoxynucleotides: these are the subunits that will be incorporated as individual bases; and finally,
6. Labeled dideoxynucleotides: these are the critical molecules that define sequencing by the Sanger method.

Slide 5: DNA Chain

Before we go further into the details of Sanger Sequencing, let's review some basics of DNA structure. The cartoon on the right depicts a chain of two deoxyribonucleic acid subunits. The term 'Base' describes each of these subunits. Each subunit is comprised of a nitrogenous base, a deoxyribose sugar, and a phosphate group. The DNA chain is assembled from the 5' carbon to the 3' carbon of each sugar ring. Each phosphate group, shown with a red 'P', is linked to the 5' carbon of the sugar ring. The Base location is on the opposite side of the ring; this is where an adenine, thymine, guanine, or cytosine is attached. Each subunit is linked by a phosphodiester bond which forms between the 3' carbon of the sugar ring and the phosphate group of the next subunit. At the very bottom of the chain, you'll see a hydroxyl group at the 3' position highlighted in yellow. This hydroxyl group is important for the formation of the phosphodiester bond and the addition of another nucleotide triphosphate to the chain.

Slide 6: Dideoxynucleotides: Chain terminators

As I mentioned before, the key component of Sanger Sequencing is the dideoxynucleotide. These are synthetic molecules that have the 3' hydroxyl group replaced with hydrogen. In the figures shown, the left figure is the usual deoxynucleotide triphosphate subunit used by DNA polymerase to create DNA chains. On the right is a dideoxynucleotide where the hydroxyl has been replaced by hydrogen. When a dideoxynucleotide is incorporated into a DNA chain, no further bases can be added, and the chain is terminated.

Slide 7: Schematic of Sanger Sequencing

This figure summarizes what occurs in a Sanger Sequencing reaction. First, the pink block letters on the top arrow represent the bases of the DNA template of interest. We are going to determine the order of bases from this template. Second, the white block letters on the bottom arrow represent the synthetic oligonucleotide known as a primer. I've also added a 5' and 3' to the arrow to emphasize that the primer is the starting point of a chain that will grow in a 5' to 3' direction. It is important to recall that DNA is double-stranded, with each strand having bases complementary to the other. Thus, Adenine and Thymine will always be paired together and Guanine and Cytosine will always be paired together. This complementary relationship between bases is used in the design of the primer and the interpretation of the final sequencing result.

A third component, in orange, represents the DNA polymerase enzyme. This enzyme will incorporate either deoxynucleotides or dideoxynucleotides into the chain initiated by the primer. When a deoxynucleotide is incorporated, the DNA polymerase will continue to add additional subunits. When a dideoxynucleotide is incorporated, the DNA polymerase will stop forming the chain. For the next couple of slides, deoxynucleotides are colored white and dideoxynucleotides are colored yellow.

Slide 8: Original Radiolabeled Sanger

Sanger sequencing using radioactive tracer molecules is no longer used. However, understanding this original version of Sanger sequencing is helpful in understanding the modern version. For radiolabeled Sanger, the sequence of nucleotides from a single template of DNA needs to be examined by four separate reactions. Each reaction includes the same DNA template and a mixture of all four deoxynucleotides: dATP, dCTP, dTTP, and dGTP. The difference between each of the reactions is the addition of a specific dideoxynucleotide. One reaction will have ddATP, and second will have ddCTP, a

third will have ddTTP, and a fourth will have ddGTP. All of the dideoxynucleotides are labeled with the same radiotracer.

In the figure, the chain lengthening and termination for each dideoxynucleotides reaction is shown. All four reactions with dideoxynucleotides have the exact same template in pink block letters. Let's examine each chain formation from left to right. We will examine the chains growing from the bottom to the top in a 5' to 3' direction.

In the ddATP reaction, the first three bases in white moving from bottom to top are T, T, and C; these are the bases from the original primer. The subsequent yellow A is a dideoxynucleotide and would terminate the DNA chain. Next to this short chain, a different strand is shown where the T, T, C primer is followed by A, C, and T in white followed by a yellow A. In this example, the first A position had dATP added which permitted further lengthening of the chain until the final dideoxy-ATP was incorporated. In the ddCTP reaction, the first three bases of the primer are again T, T, and C, which are followed by A and then a yellow C. The chain is terminated with the addition of dideoxy CTP. The next chain has the primer T, T, C followed by a deoxy-CTP, then by T, A and then a yellow C – dideoxy CTP. In the ddTTP reaction, the first three bases of the primer, T, T and C, are followed by A, C, and then a yellow – dideoxy TTP. The next chain has the primer T, T and C, followed by A, C, T, A, C, and then a yellow – dideoxy TTP. In the final ddGTP reaction, the first three bases of the primer, T, T and C, are followed by A, C, T, A, C, T, and then a yellow – dideoxy GTP.

Each of these reactions generates fragments of a specific size that corresponds to the location of the dideoxynucleotides that are incorporated.

Slide 9: Fluorescently Labeled Nucleotides
A major innovation in Sanger sequencing occurred in the 1980s when Dr. Leroy Hood invented automated instrumentation and used fluorescent dyes instead of radioactivity to detect the DNA fragments. With the radioactive method, all of the ddNTPs had the same radioactive label and needed to be used in separate reactions. The chief advantage of using fluorescence is that each dideoxynucleotide fragment has a different label and can be combined into a single reaction. The figure shows that each nucleotide position where a dideoxynucleotide is incorporated has a fluorophore color that specifies the nucleotide-type. In this picture, ddATP is shown in green, ddCTP is shown in blue, ddTTP is shown in red, and ddGTP is shown in yellow.

Slide 10: Automated Sanger with Fluorescence
The fragments of DNA labeled with fluorescently labeled nucleotides were initially run on acrylamide slab gels similar to what had been done previously for traditional radioactive Sanger. However, another innovation was transitioning to capillary electrophoresis for fragment separation. The capillary electrophoresis used in today's automated Sanger sequencing instrumentation comes in reusable bundles of capillaries filled with acrylamide. In the figure shown is a cartoon of a single capillary for a single sample. Each nucleotide position corresponds to a DNA chain that ends with a labeled nucleotide. The longest chain moves the slowest and is closest to the starting point of where the sample is applied; in this case, the longest chain ends with a G, highlighted in yellow. The shortest chain moves the fastest and is furthest from the starting point of where the sample is applied; in this case, the shortest chain ends with a A, highlighted in green.

Slide 11: Data Output from Fluorescent Sanger
The left figure shows the fluorescent output an automated capillary electrophoresis sequencing run. The automated capillary electrophoresis image is an assembly of the fluorescent signals from multiple individual capillaries. Each vertical column is data from a single capillary column; this picture is the computer-generated image of data from five capillary columns. The figure on the right is the chromatogram trace file. Each colored peak has a height that is proportional to the intensity of the signal.

Slide 12: Two Signals at the Same Position
When sequencing human DNA for clinical or research purposes, a common finding is base substitution. If we recall from genetics, each person has two copies of a gene and those two copies may have variation in the bases. As each fluorescently labeled fragment progresses through the capillary, it is examined for the identity of the nucleotide: A, T, G, or C. When the patient sample has base differences between their two copies of the gene – two signals will be detected at the same position.

In the right figure is represented the dye signal at each position within the capillary. At the fastest moving position, at the bottom, only the green signal corresponding to Adenine is detected. However, at the next position, the signal for both Cytosine – blue and Adenine – green are detected together. In the third position, there is only a red – Thymine signal. At the fourth position, there is only a green- Adenine signal. At the fifth and slowest moving position, there are two signals: Thymine –red and Cytosine – blue.

Slide 13: Example of Heterozygous C to T
This is an example of the fluorescent signals that are seen when there is a heterozygous change. The chromatogram of the sample being tested is on the top and the reference sample is on the bottom. As we look across the chromatograms, the peaks vary in height depending on the signal intensity. The reference sequence and the sample tested differ at the sixth nucleotide position. The reference sequence has a C here and the test sample has both a C and a T (depicted by overlapping blue and red peaks). The letter shown over the double peak is Y, which is an abbreviation meaning either C or T. At the very bottom of this slide, I've put a reference key for many of the common abbreviations that are used in base calling.

Slide 14: Limitations of Sanger Sequencing
The limitations of Sanger sequencing can be considered in terms of analytical-issues or laboratory management issues. The key analytical limitations are difficulties in designing and optimizing primers. Some target sequences have base compositions (e.g., GC-rich content) that create difficulties in designing robust clinical tests. In addition, target regions of interest may be surrounded by polymorphic sequence; polymorphic sequence also adds to assay design challenges. An important limitation of Sanger sequencing is that it will not detect large deletions or duplications of sequences.

In terms of laboratory management, the cost of Sanger sequencing must be carefully considered. Many of the steps of Sanger sequencing can be automated but there are multiple steps that still require manual processing. In addition, in a low volume laboratory, the cost of Sanger sequencing is still significant. Indeed, for low volume laboratories, the cost of Sanger sequencing two or three genes

sometimes exceeds the cost of performing sequencing of panels of hundreds of genes by next-generation sequencing methods.

Slide 15: Summary
DNA sequencing by the Sanger method has existed for almost thirty-years. Initially, the method was dependent on radioactively labeled nucleotides. A key innovation in the method was the switch to fluorescently labeled nucleotides. Modern instrumentation is automated and relies on capillary electrophoresis with fluorescent detection.

As discussed in other teaching modules, next-generation sequencing has begun to replace Sanger sequencing in research and clinical applications. However, there is still a significant analytical role for Sanger sequencing. At the very least, understanding the mechanism of Sanger sequencing is helpful for understanding more complex DNA sequencing technologies.

Slide 16: References

Slide 17: Disclosures

Slide 18: Thank You from www.TraineeCouncil.org
Thank you for joining me on this Pearl of Laboratory Medicine on "Sanger Sequencing." My name is Jason Park.