Children's Mercy Research Institute
KANSAS CITY

CHILDREN'S MERCY RESEARCH INSTITUTE

DNA Decoding Challenge

Presented by Illumina Corporate Foundation

Children's Mercy has been providing care for children in our community since the hospital was founded by Drs. Katharine Berry Richardson and Alice Berry Graham in 1897. These two sisters believed that every child deserves access to quality health care and opportunities to reach their full potential, regardless of their station in life. They dreamed of one day building a research center focused solely on better understanding and treating the medical needs of children.

Today, that dream is a reality. At the Children's Mercy Research Institute (CMRI), scientists are embarking on a revolutionary journey to unlock discoveries that will allow speedy diagnoses, more precise treatments, cures, and prevention strategies for childhood illnesses. Using genetic information, the work happening inside the CMRI will elevate the level of pediatric care available to local families and influence the delivery of lifesaving care around the world.

To celebrate the opening of the CMRI, we have created an exciting opportunity for schools and learning centers to get involved. This real-world learning activity, designed for high school students, will challenge them to decode the DNA sequences displayed on the outside windows of the CMRI.

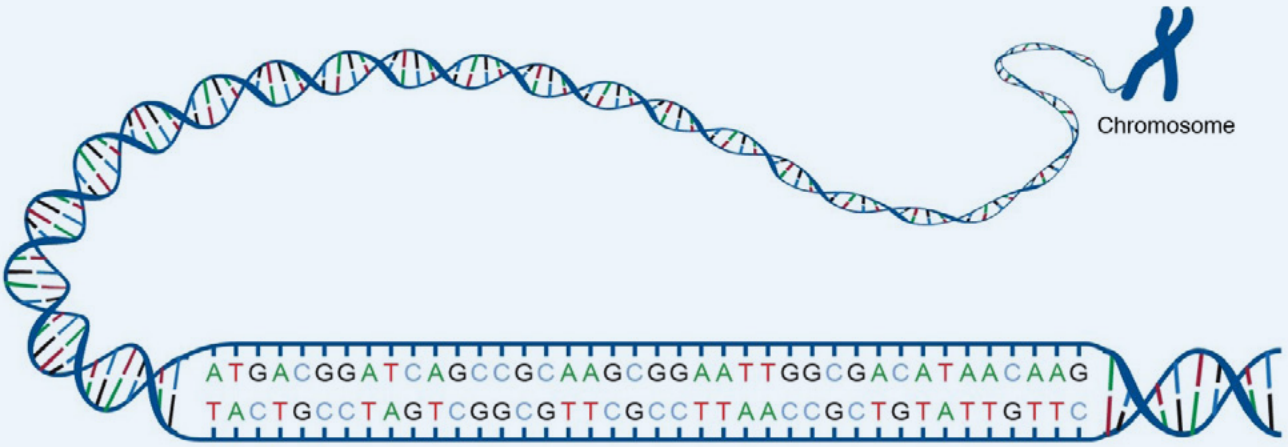
Students are invited to decipher the DNA sequences to determine what gene variants are represented. All these sequences and their respective variants were discovered in patients at Children's Mercy. The students can even take the investigation a step further by identifying the clinical implications of those variants.

Help us solve these genetic cases!

The lights of the Children's Mercy Research Institute aren't just another beautiful addition to the skyline of Kansas City. The lights represent the hidden genetic codes that the scientists inside had to crack. Often the best solutions to our most difficult problems come from many brains thinking in different ways, so we need your help! In these four cases, you will learn more about how understanding these genetic variations can help kids and their families live better lives. But first, let's review the basics of DNA and DNA Sequencing:

DNA Basics

- DNA, or deoxyribonucleic acid, is the hereditary material inside the cells of humans and almost all other organisms.
- DNA contains the instructions needed for an organism to develop, survive and reproduce. Genes are the part of our DNA that encode proteins, and proteins carry out these different functions in the body.
- DNA is made of chemical building blocks called nucleotides. These building blocks are made of three parts: a phosphate group, a sugar group, and one of four types of nitrogenous bases:
  - » Adenine (A), Guanine (G), Cytosine (C) Thymine (T)



DNA Sequencing:

- DNA sequencing determines the order of these bases, A G C T, similar to the way letters of the alphabet appear in a certain order to form words and sentences.
- The DNA sequence tells scientists the kind of genetic information that is carried in a particular DNA segment. Knowing the sequence can be helpful in a variety of ways, including:
  - » Determining which parts of DNA contain genes and which parts of DNA carry regulatory instructions.
  - » Highlighting changes in a gene that may cause a genetic disorder. A genetic disorder or genetic condition is a disruption in normal bodily functions caused in whole or in part by a change in the DNA sequence.
- You might hear other people refer to genetic changes as genetic mutations, but really, they are variations. Just because a change causes a difference, it doesn't mean that the outcome is necessarily harmful. Every human being has variants! Think about how brown and blue eyes are just variations in the genes that affect eye color.

How can DNA sequencing help determine if someone has a genetic disorder?
When scientists know the order of nucleotides, they can identify which change(s) contribute to or cause a condition.

- » A reference or reference sequence is the order of bases that scientists have generally agreed upon. A variant is a change that differs from this agreed upon sequence.
- » On the CMRI building, the **blue lights** represent bases that correspond with the reference sequence and the **red lights** represent bases that deviate from the reference (i.e. variants).

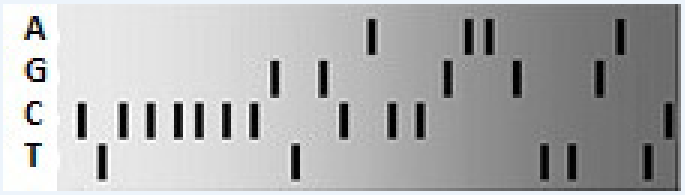
Here is a simplified example: Take a look at the example below. Notice that the variant DNA has a base change at the 2nd nucleotide. Scientists can now look at this variant to see if it could be the possible cause of a genetic disorder.

Reference DNA sequence	A-T-G-C-C-T-A-A-C
Variant DNA sequence	A-C-G-C-C-T-A-A-C

Sanger Sequencing:

- Sanger sequencing is one of the many methods scientists use to determine the nucleotide sequence of DNA.
- In short, four reactions are created, one tube for each nucleotide (A, G, C and T). The reaction ingredients combine inside the tube to form the bands we see on the gel, which represent the different positions of a nucleotide in a given sequence.
- The bands have different sizes depending on what nucleotide is present and how far into the sequence it is.
- However, all these different-sized bands are jumbled up in one reaction tube. They can be separated by adding the tube contents to a gel. When an electrical current is applied, the gel material acts like a mesh, separating out the different sized bands.
- Because each reaction tells us where that nucleotide is in a sequence, we can look at all the reactions together to form the full sequence.

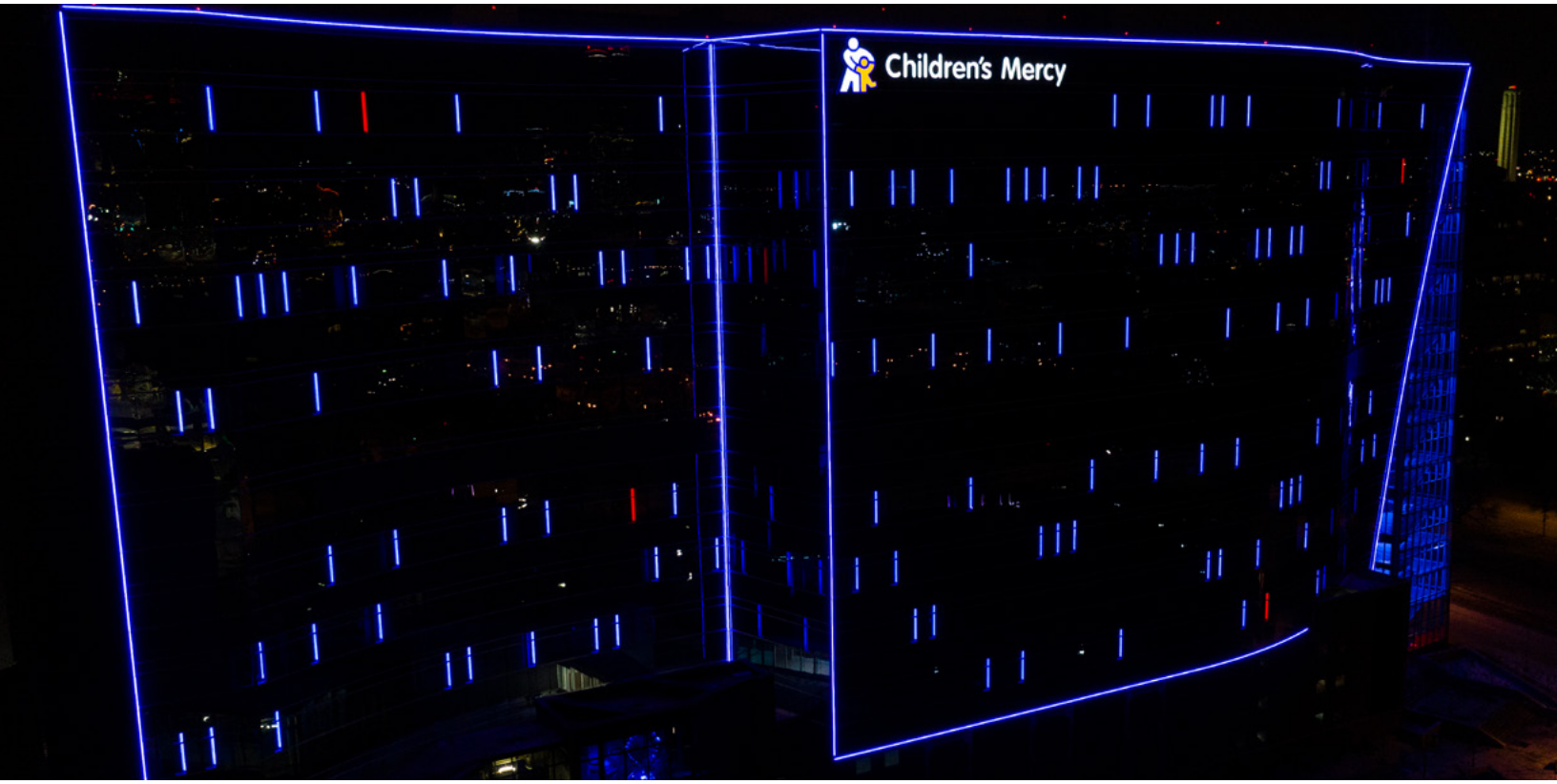
Below is an example of a Sanger sequencing gel. The sequence of the DNA can be identified by reading the bases left to right in order of which the bands appear. The sequence below would be read as: CTCCCCCCGTGCACCGAAGTTGATC.



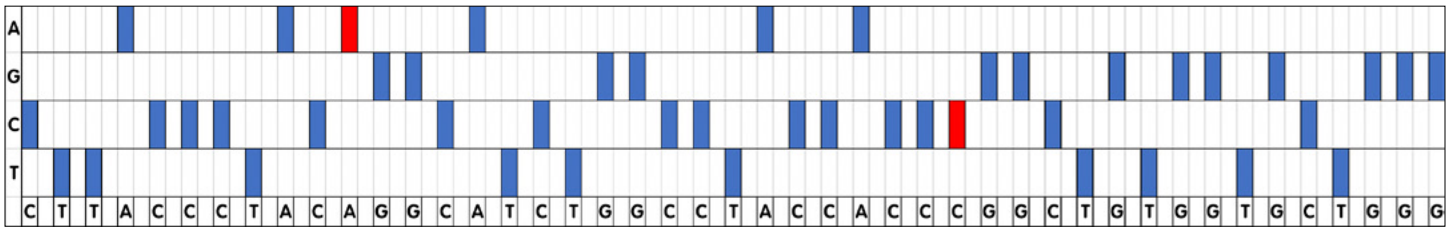
Want to learn more? Check out our Helpful Resources page at the end of this document!

Case #1:

Shameika started having problems with her hearing and speech since when she was 12 years old, and these symptoms have progressively gotten worse over time. Then shortly after she turned 17, she noticed that it was becoming more and more difficult for her to walk her favorite trails in the park. She also started having difficulty drinking water and swallowing food. Concerned for her health, Shameika's parents brought her to Children's Mercy to find out how they could help her get better. A team of scientists was assembled. They decided that looking at her DNA was the next step.



Below is a representation of what Shameika's sequence on the building (top left) would look like if it was run on a Sanger sequencing gel. Blue bands indicate bases that correspond with the reference sequence, and red bands correspond with variant bases (on a real gel, all the bands would be one color like what we saw in the example above). There is one additional "variant" in the sequence that is not in red. This first case has been decoded for you as an example.



Case #1 Questions

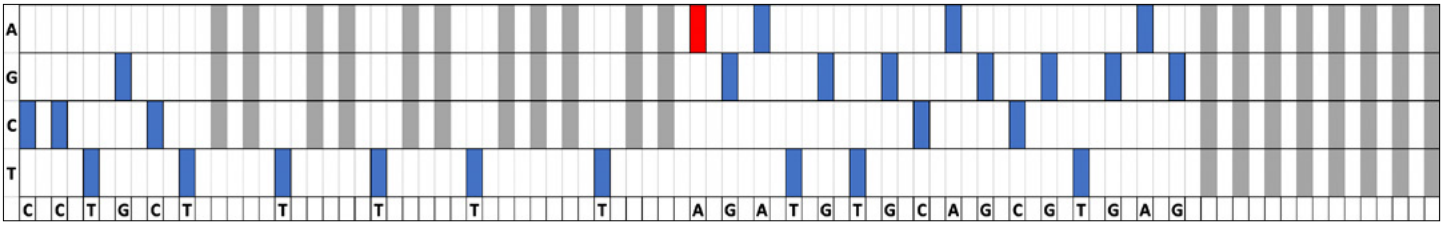
1. What sequence did the scientists see, which is represented on the building?
2. What gene are the scientists looking at?
3. If genes are like the blueprints for cells to make proteins, what does the protein for this gene do?
4. What genetic condition is associated with this protein when it isn't functioning properly?

- *Optional Challenge Question 1:* How do the specific variants (red bands) in the gene sequence change the function of the protein?
- *Optional Challenge Question 2:* What sort of treatments could the doctor prescribe to help Shameika?

Case #2:

Ben, a high school student with Attention Deficit Hyperactivity Disorder (ADHD), was prescribed Strattera® (also known as atomoxetine) to help him focus in school. A few weeks later he noticed that his heart was racing. At first, he thought it was anxiety for his upcoming finals, or maybe even something wrong with his heart? After enrolling in a research study at Children's Mercy, the scientists found that he had rare variants in a gene that break down certain medications, including atomoxetine. The result was that the amount of that medication in his body was too high, so the doctors reduced his atomoxetine dose. His heart rate returned to normal, and he ended up doing great on his finals. One of the rare variants is shown on the building (bottom left).

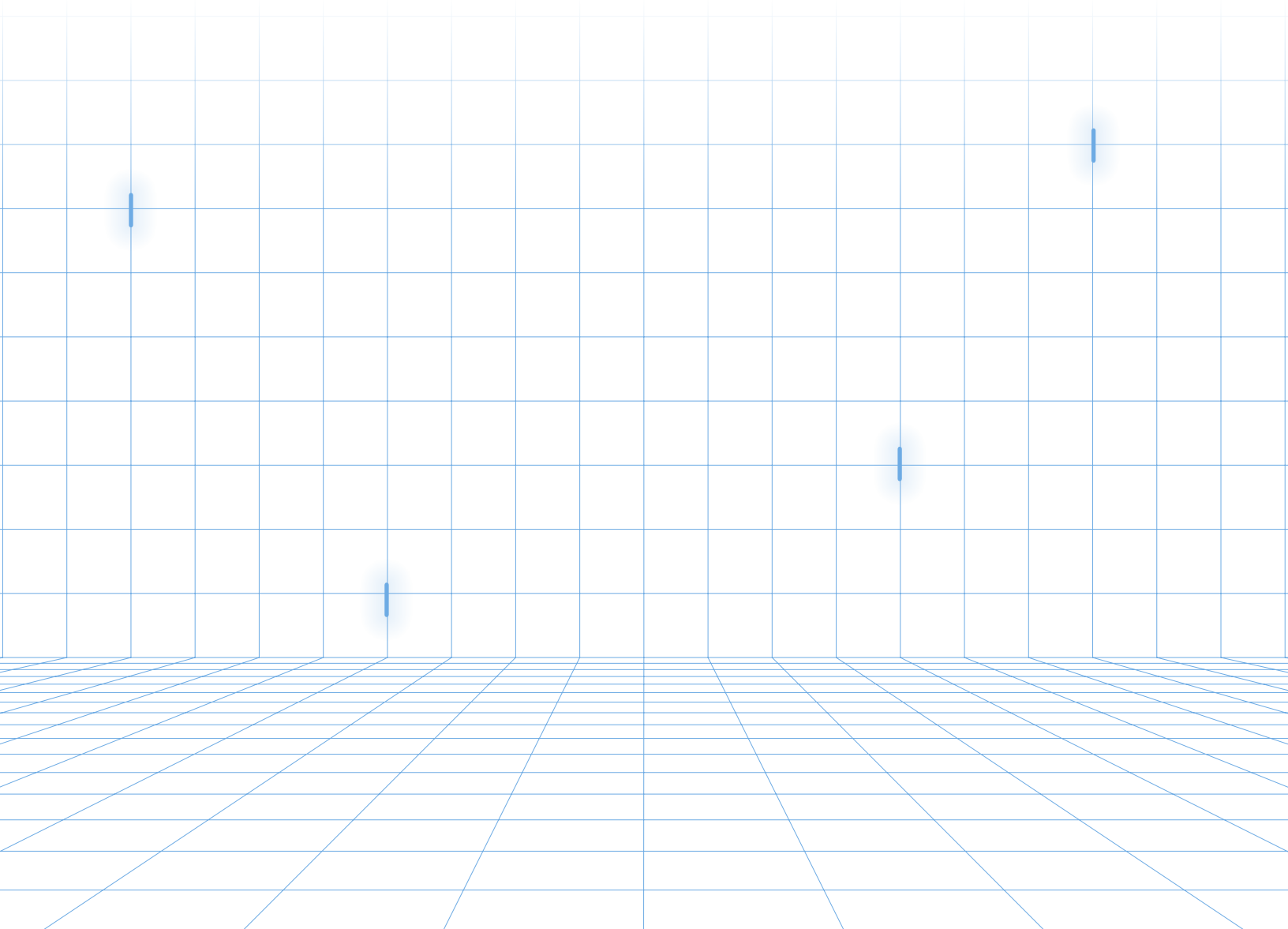
Below is the representation of the Sanger sequencing gel. Use the picture of the CMRI to fill in the missing bases. Some bases have already been decoded for you. There is also one additional variant included in Ben's sequence that is not shown in red, similar to the previous case. The grey areas below indicate what bases are possible in that position.



Case #2 Questions

1. What is the sequence the scientists looked at?
2. What is the gene that matches the sequence?

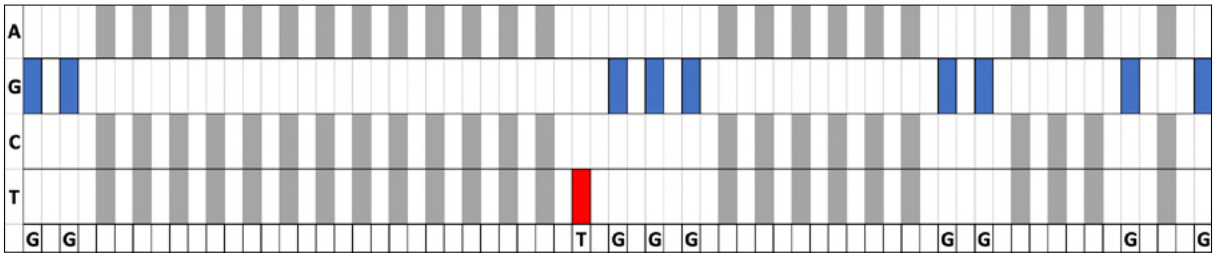
- *Optional Challenge Question 1:* What is Ben's specific rare variant?
- *Optional Challenge Question 2:* How does Ben's specific variant affect the way he would metabolize atomoxetine?
- *Optional Challenge Question 3:* Based on the effect of this variant, how might Ben be dosed compared to someone with another variant that does not cause a change in the function of the protein? (Assume the other copy of his gene produces a non-functional protein.)



Case #3:

Baby Dorian was admitted to the pediatric intensive care unit (PICU) at only 1 month old for hypoglycaemia, critically low blood sugar levels. Doctors tried to see if giving additional glucose or medication (diazoxide) would help, but these treatments did not seem to work. Hypoglycemia is harmful for babies and toddlers as it can cause permanent brain damage and affect development. After more laboratory testing, the scientists discovered that Dorian had excess levels insulin (hyperinsulinaemia), a hormone produced in the pancreas that regulates blood sugar levels. Genetic analysis was done to discover the underlying cause of these symptoms.

Below is the representation of the Sanger sequencing gel. Use the picture of the CMRI (bottom right) to fill in the missing bases. The variant (red) and the guanine nucleotides (G) have already been decoded for you. There is one additional variant included in the sequence that is not shown in red on the building. The grey areas below indicate what bases are possible in that position.



Case #3 Questions

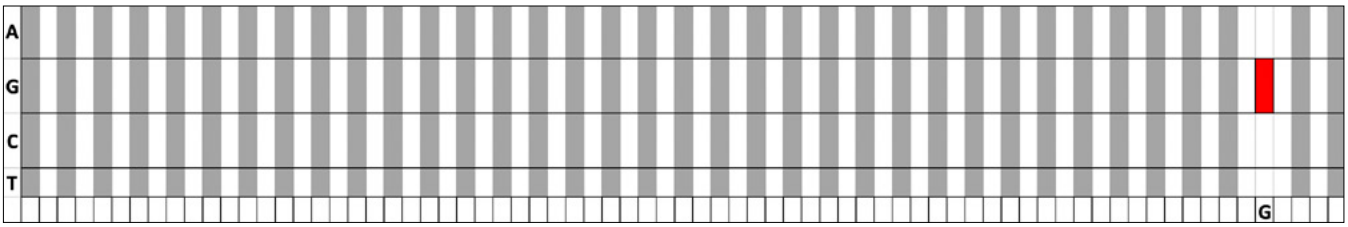
1. A rare variant was found in Dorian's DNA sequence, which is displayed on the CMRI windows. What is the decoded sequence?
2. What is the gene that corresponds to the sequence?

- *Optional Challenge Question 1:* What is the rare variant?
- *Optional Challenge Question 2:* Knowing the gene and the variant helped surgeons target the specific part of the pancreas that was causing the problem and remove it. After the surgery, Dorian's blood sugar level then stabilized, and the baby made a full recovery. Based on symptoms, laboratory and genetics results, what is the diagnosis of Dorian's condition?
- *Optional Challenge Question 3:* If Dorian's condition is considered autosomal dominant in this case, why is it that that both biological parents are healthy and have never shown any similar pattern of symptoms?

Case #4:

Tourette syndrome (TS) is a condition that affects a person's nervous system. It causes people to have involuntary "tics" that can be verbal like making clicking noises or movement-based like jerking their neck. Sometimes tics can even be as complex as yelling out phrases or repeatedly moving in a certain pattern without being able to stop. TS is more likely to develop in males and can be associated with other conditions like ADHD and obsessive-compulsive disorder (OCD). The exact cause of TS is unknown and is most likely the result of many factors, but part of the answer might be in a person's genetics. One study of TS showed that multiple affected family members had the same novel variant in their DNA sequence.

Below is a representation of the Sanger sequencing gel. Use the pictures of the CMRI (top right) to fill in the missing bases. Only the variant has already been decoded for you here. There is one additional variant included in the sequence that is not shown in red on the building. The grey areas indicate what bases are possible in a given position. This one is difficult! Good luck!



Case #4 Questions

1. What is the decoded DNA sequence from the family members?
2. What gene does this sequence belong to?
3. How might knowing this variant and its effects help others with Tourette syndrome? (Maximum 200 words)

Helpful Resources

BACKGROUND INFORMATION

- Genetics
  - » DNA -> RNA ->Protein
    - › Khan Academy: https://www.khanacademy.org/science/high-school-biology/hs-molecular-genetics
  - » YouTube – CrashCourse
    - › DNA structure and replication: https://www.youtube.com/watch?v=8kK2zwjRV0M
    - › DNA transcription and translation: https://www.youtube.com/watch?v=8kK2zwjRV0M
    - › Cold Spring Harbor Laboratory, DNA Learning Center: https://dnalc.cshl.edu/resources/animations/
- Sanger Sequencing
  - » Khan Academy: https://www.khanacademy.org/science/high-school-biology/hs-molecular-genetics/hs-biotechnology/a/dna-sequencing
  - » YouTube – Frank Lectures: https://www.youtube.com/watch?v=-QIMkQ4E_wE
  - » YouTube – My2Sense: https://www.youtube.com/watch?v=FvHRio1yyhQ

Decoding the Sequence

- Hint: From top to bottom, the bases are in order of adenine (A), guanine (G), cytosine (C), and then thymine (T).
- Hint: Each case has 1 additional random nucleotide change that is not the variant shown in red. These variants do not cause a functional change to the resulting protein.

Find the Gene

- NCBI BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi
  - » Hint: BLAST with the human genome.
  - » Hint: The more sequence that is decoded, the better the results will be.
  - » Hint: Try different optimization options under "Program Selection".

Helpful Resources

Find the Variant

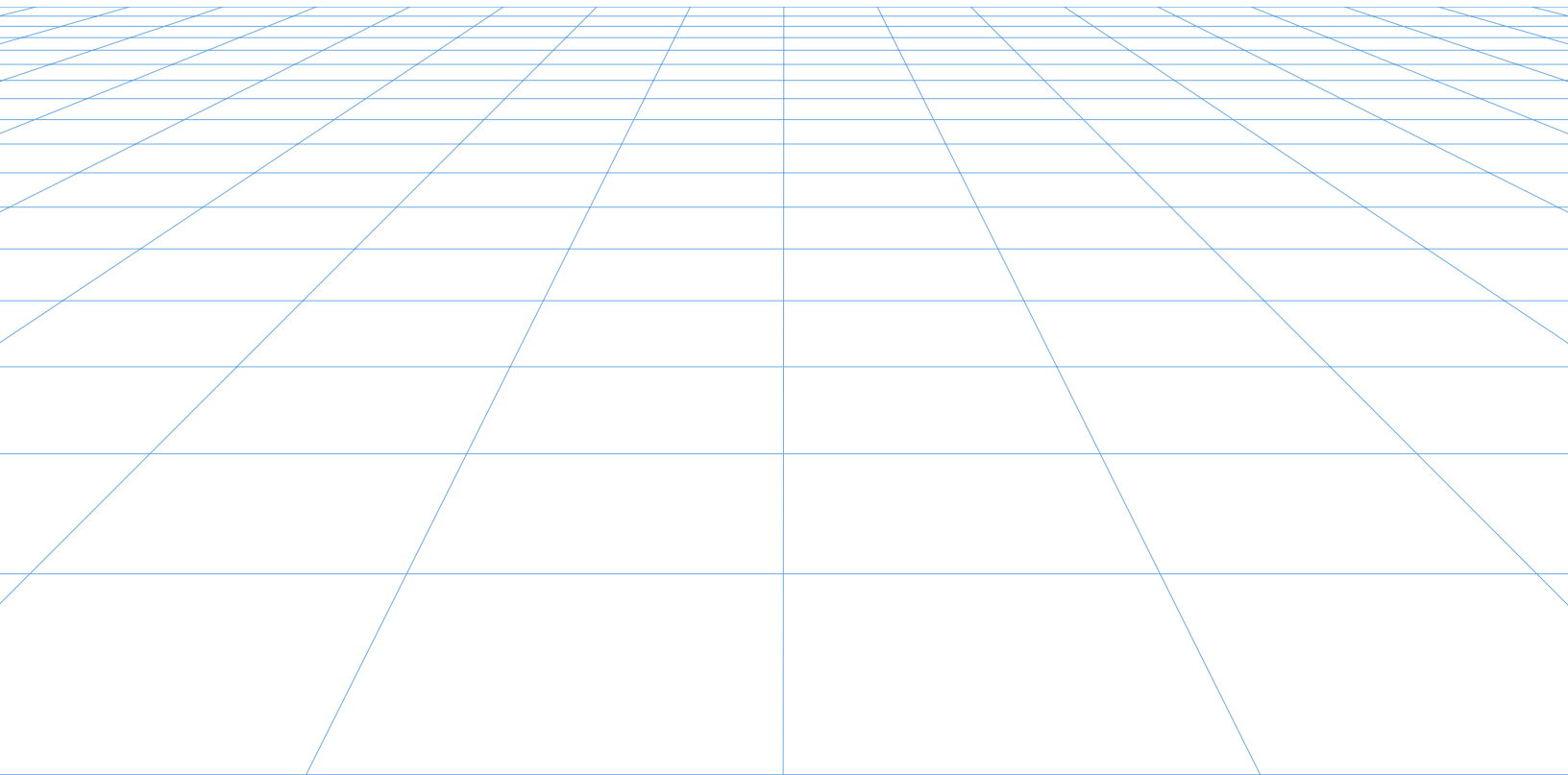
- NCBI dbSNP: https://www.ncbi.nlm.nih.gov/snp/
  - » Hint: Start by typing the gene name into the search bar.
  - » Hint: When in a specific reference SNP report, the sequence can be put into the "find" box on the interactive sequence viewer display.
  - » Hint: "rs#" or reference SNP (single nucleotide polymorphism a.k.a nucleotide variant) number. It is an identification number that is unique to every published variant.

Find the Function of the Variant

- NCBI ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/
  - » Hint: rs#'s can be cross-referenced to ClinVar
- Pharmacogene Variation Consortium: https://www.pharmvar.org/
  - » Hint: Useful for examining genes involved in drug metabolism

Gene Function/ Associated conditions

- NCBI Gene: https://www.ncbi.nlm.nih.gov/gene/
- NCBI OMIM: https://www.ncbi.nlm.nih.gov/omim/



Children's Mercy Research Institute
KANSAS CITY

Contact us

DecodeCMRI@cmh.edu

childrensmercy.org/research