Children's Mercy Research Institute
KANSAS CITY

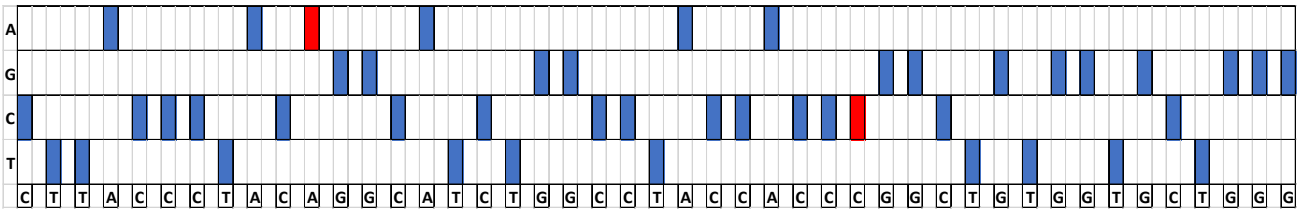
CHILDREN'S MERCY RESEARCH INSTITUTE

DNA Decoding Challenge Answer Key

Presented by Illumina Corporate Foundation

Case #1 - Shameika

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.



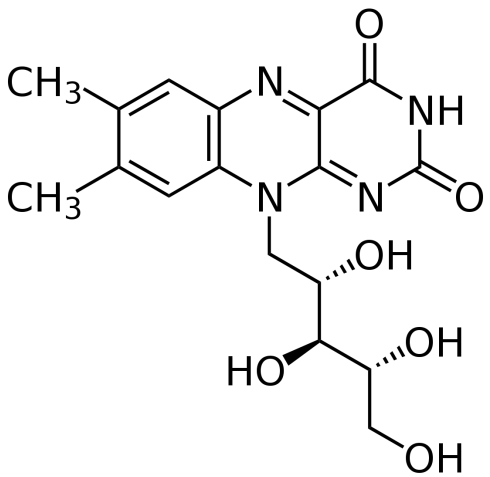
1. What sequence did the scientists see, which is represented on the building?
   CTTACCCTAC**A**GGC**A**TCTGGCCTACCACC**C**GGCTGTGGTGCTGGG

In the answer sequence above and in the rest of the answer key, the variant nucleotide change is represented in **red** text and the additional nucleotide change is represented in **blue** text.

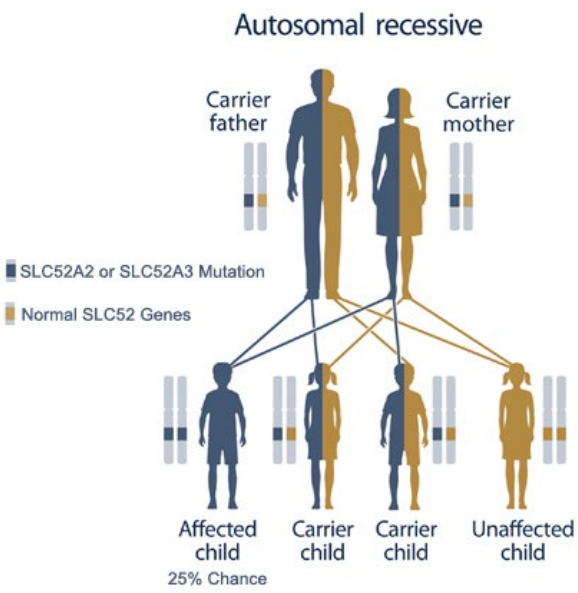
2. What gene are the scientists looking at?
   SLC52A2 (Solute Carrier Family 52 Member 2)

3. If genes are like the blueprints for cells to make proteins, what does the protein for this gene do?
   Mediates cellular uptake of riboflavin (Vitamin B2)

Why is Riboflavin important?
Riboflavin is an essential vitamin necessary for several processes like metabolizing carbohydrates, lipids, and amino acids. These are all necessary for life. Because our bodies do not make it internally, we must get it from food. *SLC52A2* encodes a protein called RFVT2 that helps us absorb riboflavin through our intestines. The loss of function of the riboflavin transporter results in riboflavin deficiency in different tissues.



4. What genetic condition is associated with this protein when it isn't functioning properly?
Riboflavin transporter deficiency (RTD) a.k.a Brown-Vialetto-Van Laere syndrome 2 (BVVLS).



Clinical Features of
Riboflavin Transporter Deficient (RTD):
The most frequent presenting symptoms are cranial neuropathy, sensory ataxia, muscle weakness and respiratory insufficiency due to diaphragmatic paralysis.

- The cranial neuropathy (which means the nerves are not functioning as they should, in this case in the brain) leads to sensorineural hearing loss, loss of vision due to optic atrophy, dysarthria, dysphagia, feeding problems, facial weakness and eye movement impairments.
- Peripheral neuropathy causes muscle weakness with atrophy and (sensory) ataxia.
- Diaphragmatic paralysis in combination with general muscle weakness may rapidly lead to respiratory insufficiency, especially in infants and young children. Often, feeding through a nasogastric tube or gastrostomy device and artificial ventilation and tracheotomy are required.

Optional Challenge Question 1: How do the specific variants (red bands) in the gene sequence change the function of the protein?
There are two amino acid changes: p.Gly306Arg (rs398124641) and Leu312Pro (rs754320812)

Our Expert: Dr. Ana Cohen, PhD
Assistant Director of the Clinical Molecular Genetics Laboratory
Department of Pathology and Laboratory Medicine

Expert Analysis:

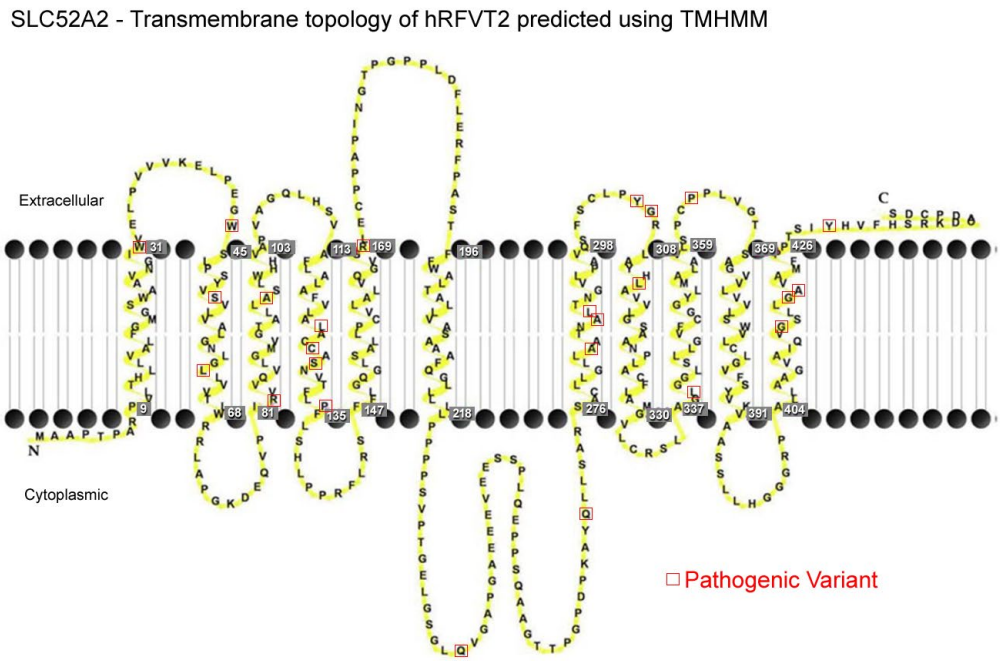
p.Gly306Arg:
The glycine amino acid is highly conserved and there is a moderate physicochemical difference between glycine and arginine. This variant has been reported in individuals with the clinical features of BVVLS and was found in >10 affected individuals from several families. Experimental studies have shown that this missense change decreased riboflavin transporter protein expression.

p.Leu312Pro:
The leucine amino acid is highly conserved and there is a moderate physicochemical difference between leucine and proline. This variant has been observed in the presence of a second *SLC25A2* variant in several individuals affected with BVVLS. This variant has been reported to affect SLC52A2 protein function.

Optional Challenge Question 2: What sort of treatments could the doctor prescribe to help Shameika?

Riboflavin supplementation, physical and speech therapy, and cochlear implants for hearing loss.

- Riboflavin supplementation is lifesaving. Daily administration of a high dose oral riboflavin should be started immediately upon suspicion of RTD; treatment is lifelong. Most patients show improvement over time and early treatment may even result in full recovery and normal development.
- In cases with early onset, assisted ventilation, tracheostomy, and maintenance of nutrition via gastrostomy may be needed.
- Regular clinical evaluation is required and includes neurological examination as well as evaluation of hearing and vision, respiratory function and skeletal deformities like scoliosis.
- Depending on the symptoms, physical and speech therapy should be initiated. Cochlear implants have proven highly successful for hearing loss in RTD.



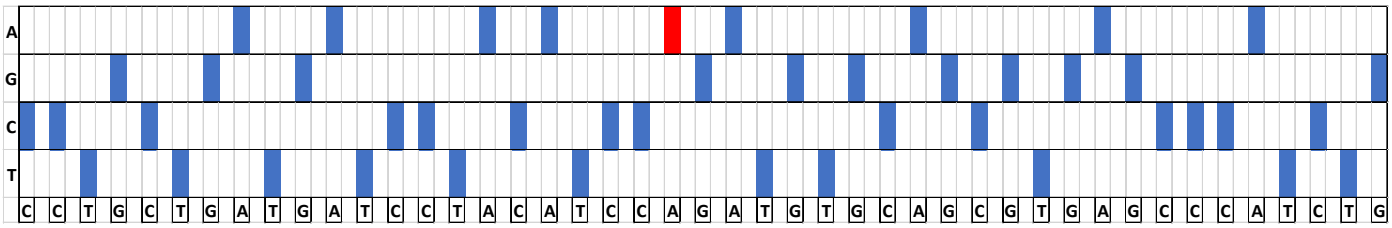
Hey, have you noticed that gene names are italicized?

Gene symbols (short name for a gene) are formatted in *italics* and proteins are not. Both are in upper-case for human genes and the rules are different for other organisms. These naming standards allow scientists to communicate easily in written materials. It is especially helpful because typically a gene and its associated protein(s) have the same name.

If you want to learn more check out the "Formatting Gene Names" summary in the Sources and Resources section at the end of this document – note that there are a lot more rules that can get updated over time

Case #2 - Ben

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.

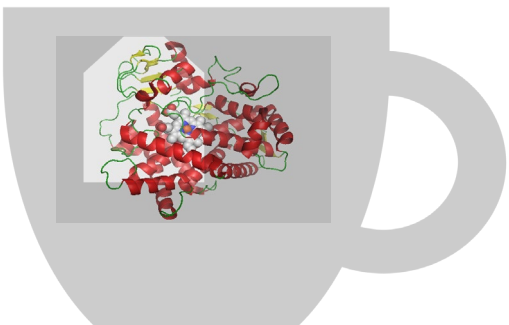


1. What sequences did the scientists see, which is represented on the building?
   CCTGCT**G**ATGATCCTACATCC**A**GATGTGCAGCGTGAGCCCATCTG

2. What is the gene that matches the sequence?
   CYP2D6 (Cytochrome P450 2D6)

Sip on some CYP2D6 Facts:

- CYP2D6 is an enzyme that metabolizes or interacts with almost a quarter of all pharmaceuticals that are prescribed including medication used to treat mental health, pain, cancer, and heart conditions.
- *CYP2D6* is also known to be highly polymorphic, meaning having lots of variants, which can differ substantially between individuals and people of different ethnicities.
- Knowing what variants a person has can help providers determine how much or what kind of drug will be most beneficial for a person.



Our World-Renowned Expert, Queen of *CYP2D6:*
Dr. Andrea Gaedigk, PhD
Director of the Pharmacogenetics Core Laboratory

Optional Challenge Question 1: What is Ben's specific rare variant?

rs79292917; this variant does not change an amino acid (Pro325=) but causes a splicing defect.

Sequence name	Change
chr 22 novel patch HSCHR22_5_CTG1	NW_009646208.1:g.13418C>T
chr 22 novel patch HSCHR22_7_CTG1	NW_014040931.1:g.21441C>T
chr 22 novel patch HSCHR22_8_CTG1	NW_015148968.1:g.5593C>T
CYP2D6 RefSeqGene	NG_008376.3:g.7140G>A
CYP2D6 RefSeqGene (LRG_303)	NG_008376.4:g.7959G>A
GRCh37.p13 chr 22	NC_000022.10:g.42523854C>T
GRCh38.p12 chr 22	NC_000022.11:g.42127852C>T
GRCh38.p12 chr 22 alt locus HSCHR22_2_CTG1	NW_004504305.1:g.50179C>T
GRCh38.p12 chr 22 alt locus HSCHR22_3_CTG1	NT_187682.1:g.50193C>T

Screenshot from dbSNP: https://www.ncbi.nlm.nih.gov/snp/rs79292917

dbSNP catalogues the variant or SNP (single nucleotide polymorphism) in different reference sequences. Let's look at the highlighted sequence for rs79292917 as an example.

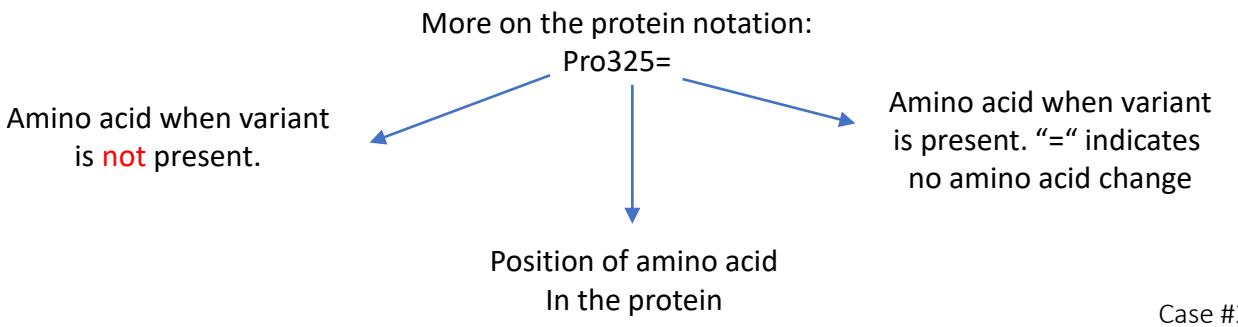
- CYP2D6 RefSeqGene (LRG_303) is the name of the reference sequence
- "NG_008376.4" is the genomic reference sequence ID; "NM" and "NP" prefixes indicate that the sequence corresponds to the gene's transcript and protein, respectively.
- "g." denotes that it is a genomic sequence
- 7959G>A indicates the position on that reference sequence and the nucleotide change

The reference sequence that is used can depend on: the field, a specific tool (bioinformatics) used, or what is most current.

Molecule type	Change	Amino acid[Codon]	SO Term
CYP2D6 transcript variant 1	NM_000106.6:c.975G>A	P [CCG] > P [CCA]	Coding Sequence Variant
CYP2D6 transcript variant 2	NM_001025161.3:c.822G>A	P [CCG] > P [CCA]	Coding Sequence Variant
cytochrome P450 2D6 isoform 1	NP_000097.3:p.Pro325=	P (Pro) > P (Pro)	Synonymous Variant
cytochrome P450 2D6 isoform 2	NP_001020332.2:p.Pro274=	P (Pro) > P (Pro)	Synonymous Variant

Screenshot from dbSNP: https://www.ncbi.nlm.nih.gov/snp/rs79292917

Similarly, this chart tells you the position of the SNP in reference to coding (mature RNA positions) and protein sequences . Under the change column, "c." indicates coding and "p." indicates protein.



Optional Challenge Question 2: How does Ben's specific variant affect the way he would metabolize atomoxetine?

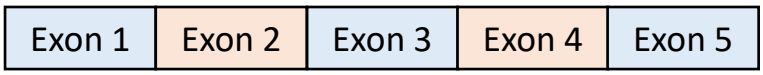
Because the variant (rs792929170) causes a splicing defect, there is a lower amount of correctly spliced messenger RNA (what is transcripted from the gene) and thus less of the CYP2D6 protein. As a result, the person's overall enzyme activity is decreased, which leads to higher amounts of the drug in Ben's body over time. This increase amount of active medication is what caused the increased heart rate.

What is a splicing defect?

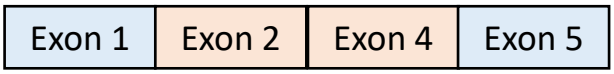
An mRNA (messenger RNA) splicing defect is defined as an error in how a mature RNA (coding exons only) is assembled from the pre mRNA (intron and exons).

Where introns are spliced out are in part due to specific sites in the DNA sequence. Thus, a change in the sequence can change where pre-RNA is spliced. In many cases, a splicing error produces nonfunction proteins, which is typically fast-tracked by the cell for degradation.

mRNA without splice variant defect



mRNA with splice variant defect



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.)

Lowering the dose of Ben's medication will help to bring the drug level within the normal range. Taking a lower dose of the medication means there will be less "excess" of it in his body to cause an unfavorable reaction.

PharmVar
Pharmacogene Variation Consortium

The PharmVar database keeps track of the many variants in important pharmacogenes.

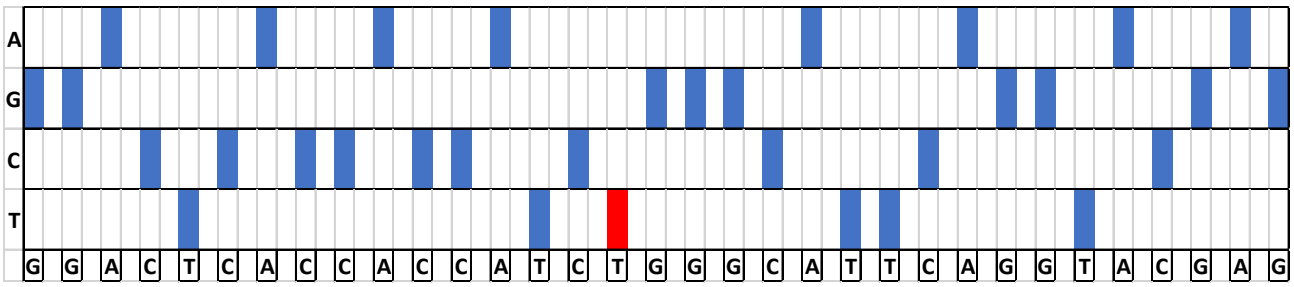
Ben's specific *CYP2D6* genotype is ***CYP2D6*4/*59*** indicating that he has variants on the gene copy inherited from both parents. The **4* allele is nonfunctional and the **59* allele has decreased function which predicts Ben to have diminished activity.

CPIC®
Clinical Pharmacogenetics Implementation Consortium

CPIC provides guidelines that help clinicians optimize drug treatment based on a pharmacogenetic test result. Ben's *CYP2D6* variants were determined to **decrease function** and recommends **a lower dose than normal metabolizers**.

Case #3 - Dorian

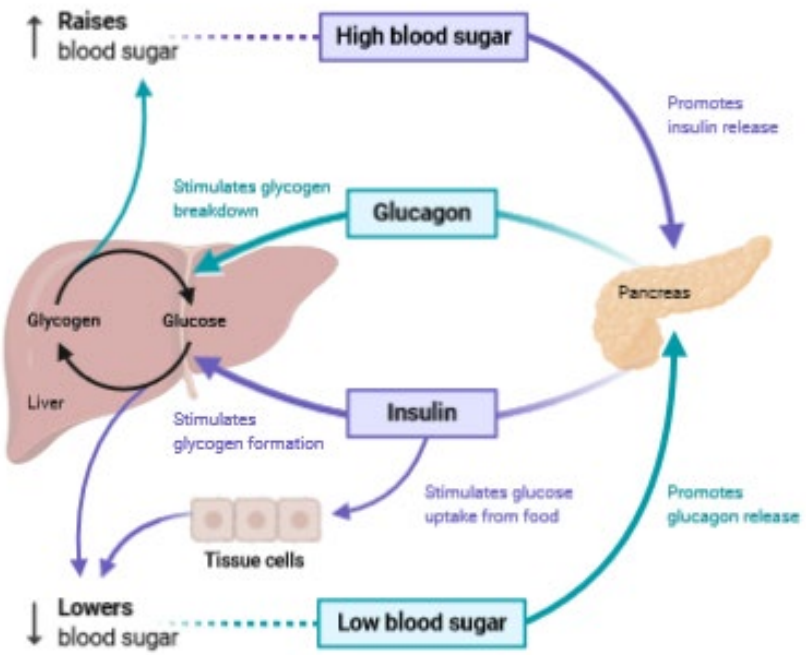
Baby Dorian was admitted to the pediatric intensive care unit (PICU) at only 1 month old for hypoglycaemia, critically low blood sugar levels. Doctors tried to see if giving additional glucose or medication (diazoxide) would help, but these treatments did not seem to work. Hypoglycemia is harmful for 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.



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

2. What is the gene that corresponds to the sequence?
   ABCC8 (ATP Binding Cassette Subfamily C Member 8)

Why is ABCC8 important?
ATP-binding cassette (ABC) transporters, in general, are a family of proteins that use ATP (energy) to transport something into or out of a cell. *ABCC8* codes for a protein called **SUR1**, which is a part of a larger protein complex responsible for regulating the release of insulin from the beta cells of the pancreas.

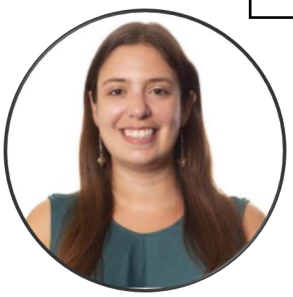


Optional Challenge Question 1: What is the rare variant?

rs139964066 which causes an amino acid change, p.Arg1214Trp

Expert Analysis:

- This sequence change replaces arginine with tryptophan at codon 1214 of the ABCC8 protein (p.Arg1214Trp).
- The arginine residue is highly conserved and there is a moderate physicochemical difference between arginine and tryptophan.
- This variant has been observed in several individuals affected with congenital hyperinsulinemia. Algorithms developed to predict the effect of missense changes on protein structure and function agree on the potential impact change
    - SIFT: "Deleterious"
    - PolyPhen-2: "Probably Damaging"
    - Align-GVGD: "Class C0"
- A different missense substitution at this codon (p.Arg1214Gln) has been determined to be pathogenic; this suggests that the arginine residue at the 1214 position is critical for *ABCC8* protein function and that other missense substitutions at this position may also be pathogenic.



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 fully recovery. Based on symptoms, laboratory and genetics results, what is the diagnosis of Dorian's condition?

Focal form familial hyperinsulinism type 1. This is also known as congenital hyperinsulinism.

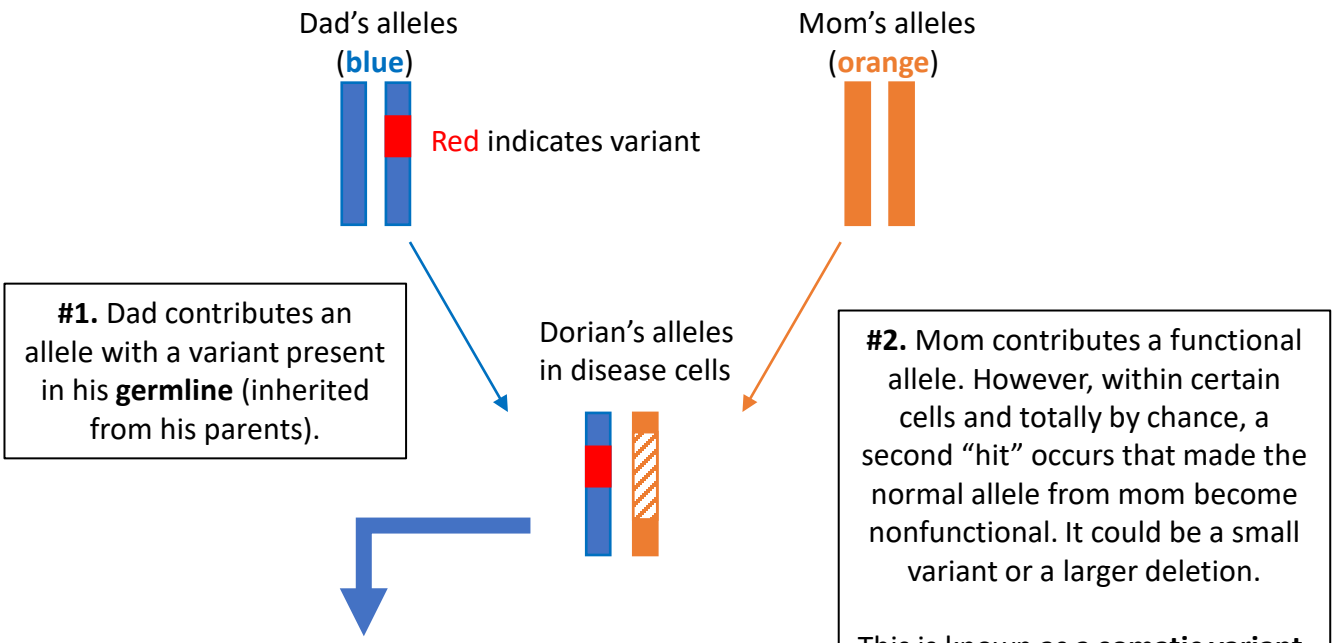
Definitions in context:	
Focal	Defined, discrete portions of the pancreas are affected (as opposed to "diffuse" meaning the entire pancreas is involved)
Familial	The increased insulin is hereditary.
Congenital	The condition is present from birth.
"Familial" and "congenital" are sometimes used interchangeably, but not all hereditary conditions appear at birth and not all conditions at birth have a hereditary component.	

Optional Challenge Question 3: If Dorian's condition is considered autosomal dominant in this case, why is it that both biological parents are healthy and have never shown any similar pattern of symptoms?

"Focal familial hyperinsulinism is inherited autosomal dominantly, but only manifests when the mutation is on the paternally derived allele and there is somatic loss of the maternal allele in a β-cell precursor." (Willig et al., 2015)

So, it's complicated. Let's break it down.

Focal familial hyperinsulinism only presents when **both** genetic events occur.



#1. Dad contributes an allele with a variant present in his **germline** (inherited from his parents).

#2. Mom contributes a functional allele. However, within certain cells and totally by chance, a second "hit" occurs that made the normal allele from mom become nonfunctional. It could be a small variant or a larger deletion.

This is known as a **somatic variant**. Somatic is different from germline because it means that the variant occurred only in the patient and was not passed down from their parents. Somatic variants are often tissue-specific as well.

Because Mom's allele is ultimately nonfunctional. The β-cells (cells in the pancreas that produce the insulin) only express dad's variant protein. This is what causes the disease state.

But wait! There's more!

It gets more complicated…

There are other features at play here like *epigenetic imprinting*. These concepts are entire fields of study, so we won't go too in-depth about them on the next page. If you're curious about specifics, check out the resources at the end of this document.

In short, just know that if Dad had the somatic variant and Mom carried the germline variant, we would not have the same disease presentation.

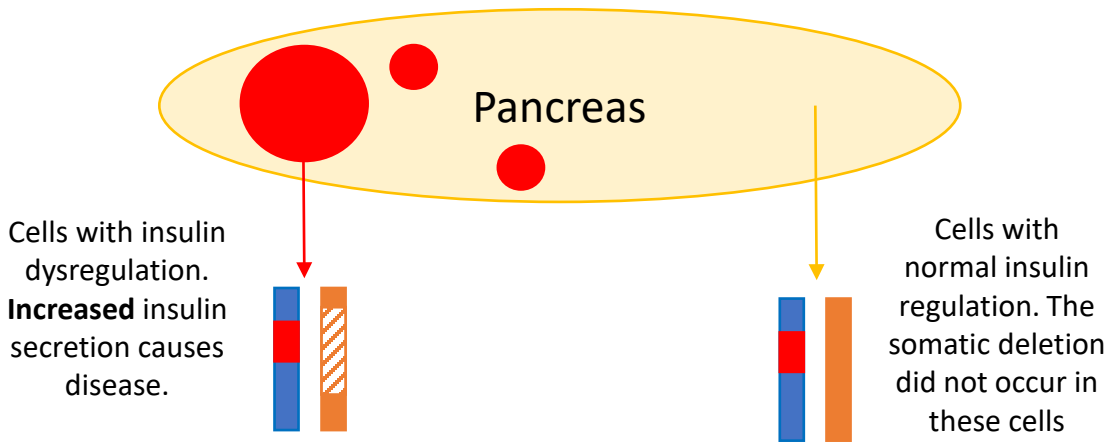
Epigenetics and Imprinting: A quick and simple explanation that absolutely does not give an entire scientific field any justice.

Epigenetics: If the DNA sequence is like the blueprint for cells to make proteins, epigenetics is like the architect's sticky notes on the blueprint. Epigenetic features change the way the genetic information is expressed, but do not actually change the DNA sequence itself.

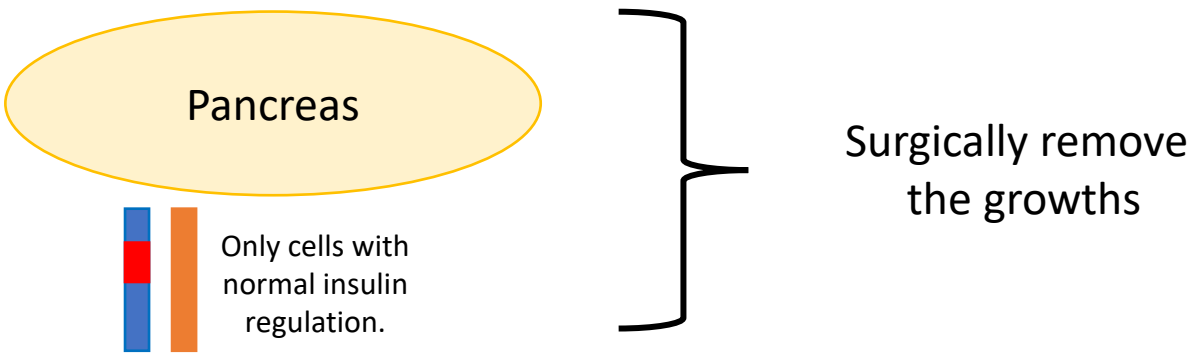
Imprinting: An epigenetic feature where it matters from which parent an allele is inherited, like in this case with Dorian. Prader-Willi Syndrome and Angelman Syndrome are other examples of genetic conditions associated with imprinting.

All together now!

- Dad's allele with a germline variant, which causes the production of a dysfunctional protein.
- Mom's somatic variant (in this specific case, it was a deletion of the *ABCC8 gene*)
    - No *ABCC8* gene; no SUR1 protein from mom's allele

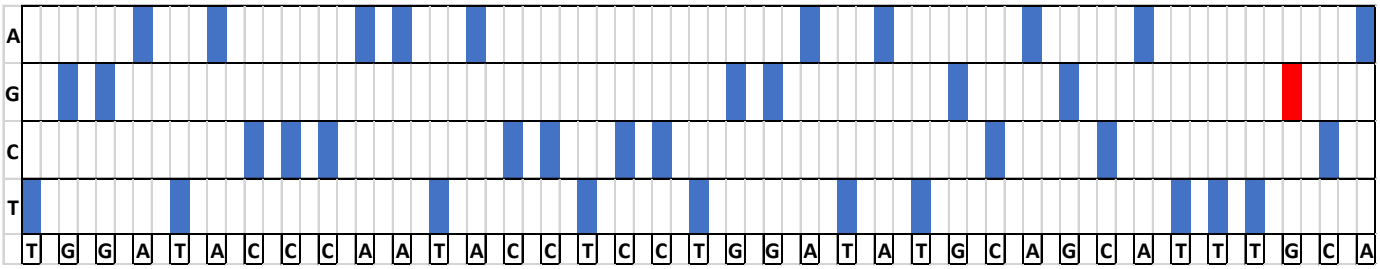


A simple cure for a complex condition:



Case #4 – Tourette Syndrome

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.



1. What is the decoded DNA sequence from the family members?
   TGGATACCCAATACCTCCTGGATATGCAGC**A**TTT**G**CA

2. What gene does this sequence belong to?
   ROR1 (Receptor Tyrosine Kinase Like Orphan Receptor 1)

In case you were curious…

This is the amino acid change: p.Pro832Ala

There is no rs# because it is very rare and/or it has not been reported before.

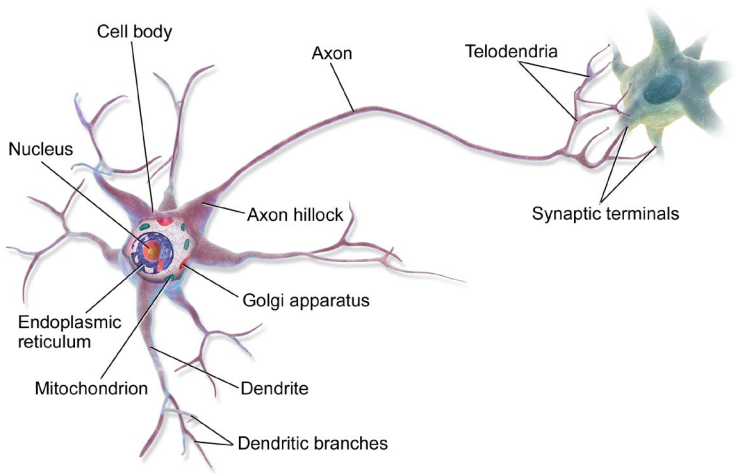
This is a *de novo* variant (Latin: "of new"), meaning that there was a spontaneous change in the DNA sequence that was not inherited from the previous generation.

This change is in the germline and will be present in future generations, which makes it different from the somatic changes like what we discussed in the previous case study.

Additional Information about *ROR1:*

Codes for a receptor protein known to regulate the growth of neuronal projections (neurite) in developing embryos.

Variants of *ROR1* are currently being studied in relation to cancer biology.



3. How might knowing this variant and its effects help with Tourette syndrome? (Maximum 200 words)

We intended for this question to be open-ended and wanted to hear what you all came up with! Below are just a few of the excellently thought-out responses that were submitted during the Decoding CMRI Challenge.

"In general, identifying the variants and understanding their functionality and effects allows the patient of that variant to receive specific and customizable care depending on the traits of the variants. In the case of Tourette Syndrome, there is no identifiable gene or cause because it is caused by many underlying factors. However, if you are able to identify a specific variant gene and its effect present in patients with Tourette Syndrome, there would be a correlation that can help scientists find a solution to the Tourette's using that specific gene variant."

Submitted by AC, SD, and EC

"Knowing this variant can assist genetic counsellors by allowing them to see whether expecting parents to see if there's a possibility to pass a genetic disorder to their children. Knowing this variant also will shed light on how to treat or cure Tourette syndrome with genetic treatments, making life easier for individuals with this disease. Identifying this variant may also shed light on other conditions with unknown causes, leading to an innovation of genetic engineering/treatments affecting not just those with Tourette, but those with other conditions as well."

Submitted by SR

"Knowing more about specific variants gives more information about what can cause Tourette syndrome, as there are most likely many contributing factors. Also knowing the specific variant could help with diagnosis."

Submitted by EC

"… Doctors can do prenatal tests looking for possible expressions of [Tourette Syndrome]. By knowing the effects of each mutation, doctors could potentially accurately predict the child's motor and speech function, in hopes of finding a specific treatment to help with their specific side effects. Tourette shows a wide variety of signs in every case but follows a similar pattern of repeated words or actions. Each person living with this condition has a different treatment plan that must be created after the child starts expressing different actions. By knowing how the child will be affected before developing fully, doctors can create a unique treatment plan to help with the upcoming effects before they present themselves in various forms. In summary knowing this condition can be passed down with similar mutations occurring, doctors can create a plan of action to benefit the child before onset symptoms occur."

Submitted by MR and KF

"By knowing the … variant and its effects, researchers can better understand and find potential treatments for Tourette syndrome. The original ROR1 gene controls neurite growth in the central nervous system, therefore identifying and understanding the variant in this gene can allow medical professionals to target the specific variant and work to modify a certain gene to aid in the prevention of Tourette syndrome. Knowing the variant can also allow medical professionals to look at other factors that may contribute to an individual being affected by Tourette syndrome."

Submitted by AP, JN, AR-O, and SA

Sources and Resources (By Case)

Case #1 – Shameika

Images

- **Riboflavin Chemical Structure:** (Wikipedia, April 2021) https://en.wikipedia.org/wiki/Riboflavin#/media/File:Riboflavin.svg
- **Autosomal Recessive Diagram:** (Cure RTD Foundation, April 2021) http://curertd.org/wp-content/uploads/2017/08/RFVT2-with-Varients-1200x802.jpg
- **SLC52A2 – Transmembrane Topology:** (Cure RTD Foundation, April 2021) http://curertd.org/wp-content/uploads/2017/08/rtdauto.jpg

Resources

- **Cure Riboflavin Transporter Deficiency Foundation:** http://curertd.org/
- **Formatting Gene Names:** https://www.biosciencewriters.com/Guidelines-for-Formatting-Gene-and-Protein-Names.aspx#:~:text=Protein%20symbols%20are%20identical%20to,case%20(e.g.%2C%20GFAP

Case #2 – Ben

Images

- **CYP2D6 Protein Structure:** (Wikipedia, April 2021) https://en.wikipedia.org/wiki/CYP2D6#/media/File:CYP2D6_structure.png

Resources

- **dbSNP:** https://www.ncbi.nlm.nih.gov/snp/?cmd=search
- **PharmVar (*CYP2D6*):** https://www.pharmvar.org/gene/CYP2D6
- **CPIC - Atomoxetine Guideline:** https://cpicpgx.org/content/guideline/publication/atomoxetine/2019/30801677.pdf

Case #3 – Dorian

Images

- **Glucose Regulation** (BioRender, April 2021) https://app.biorender.com/biorender-templates/figures/5c65c7b6bce1963300935374/t-5ed65555e0cb4c00aa170201-regulation-of-blood-glucose

Sources and Resources
(By Case)

Case #3 – Dorian

Resources

- **"Dorian's" Case Study:**

  Willig LK, Petrikin JE, Smith LD, Saunders CJ, Thiffault I, Miller NA, Soden SE, Cakici JA, Herd SM, Twist G, Noll A, Creed M, Alba PM, Carpenter SL, Clements MA, Fischer RT, Hays JA, Kilbride H, McDonough RJ, Rosterman JL, Tsai SL, Zellmer L, Farrow EG, Kingsmore SF. Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. Lancet Respir Med. 2015 May;3(5):377-87. doi: 10.1016/S2213-2600(15)00139-3. Epub 2015 Apr 27. PMID: 25937001; PMCID: PMC4479194.

- **Publications about the Inheritance Pattern of Congenital Hyperinsulinism:**

  Fournet, J. C., Mayaud, C., de Lonlay, P., Gross-Morand, M. S., Verkarre, V., Castanet, M., Devillers, M., Rahier, J., Brunelle, F., Robert, J. J., Nihoul-Fékété, C., Saudubray, J. M., & Junien, C. (2001). Unbalanced expression of 11p15 imprinted genes in focal forms of congenital hyperinsulinism: association with a reduction to homozygosity of a mutation in ABCC8 or KCNJ11. The American journal of pathology, 158(6), 2177–2184. https://doi.org/10.1016/S0002-9440(10)64689-5

  Verkarre, V., Fournet, J. C., de Lonlay, P., Gross-Morand, M. S., Devillers, M., Rahier, J., Brunelle, F., Robert, J. J., Nihoul-Fékété, C., Saudubray, J. M., & Junien, C. (1998). Paternal mutation of the sulfonylurea receptor (SUR1) gene and maternal loss of 11p15 imprinted genes lead to persistent hyperinsulinism in focal adenomatous hyperplasia. The Journal of clinical investigation, 102(7), 1286–1291. https://doi.org/10.1172/JCI4495

- **MedLine Plus: Summary of Congenital hyperinsulinism Inheritance Pattern:** https://medlineplus.gov/genetics/condition/congenital-hyperinsulinism/#inheritance

Case #4 – Tourette Syndrome

Image

- **Neuron Diagram:** (Wikipedia, April 2021) https://upload.wikimedia.org/wikipedia/commons/thumb/1/10/Blausen_0657_MultipolarNeuron.png/1920px-Blausen_0657_MultipolarNeuron.png

Resources

- **CDC – Tourette Syndrome Facts:** https://www.cdc.gov/ncbddd/tourette/facts.html

Children's Mercy Research Institute
KANSAS CITY

Contact us

DecodeCMRI@cmh.edu

childrensmercy.org/research