From DNA to TAZ variants

... is this a mutation or not?

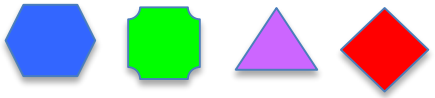
Iris L. Gonzalez

2016 Conference

DNA basics

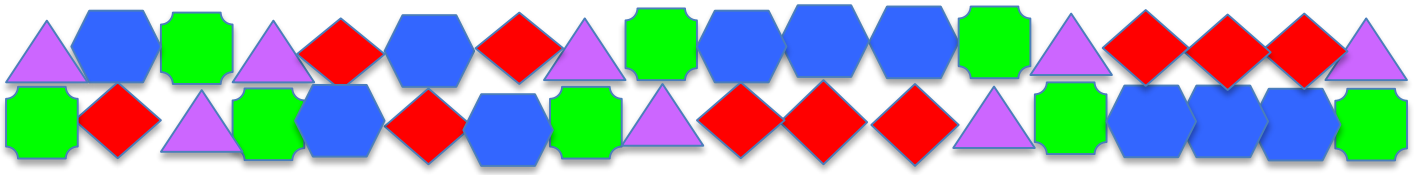
- DNA (deoxyribonucleic acid) is a huge molecule made up of units called "nucleotides" strung together into a chain.

- There are 4 different nucleotides with different shapes; we will call them by their abbreviated names: A, G, C, T

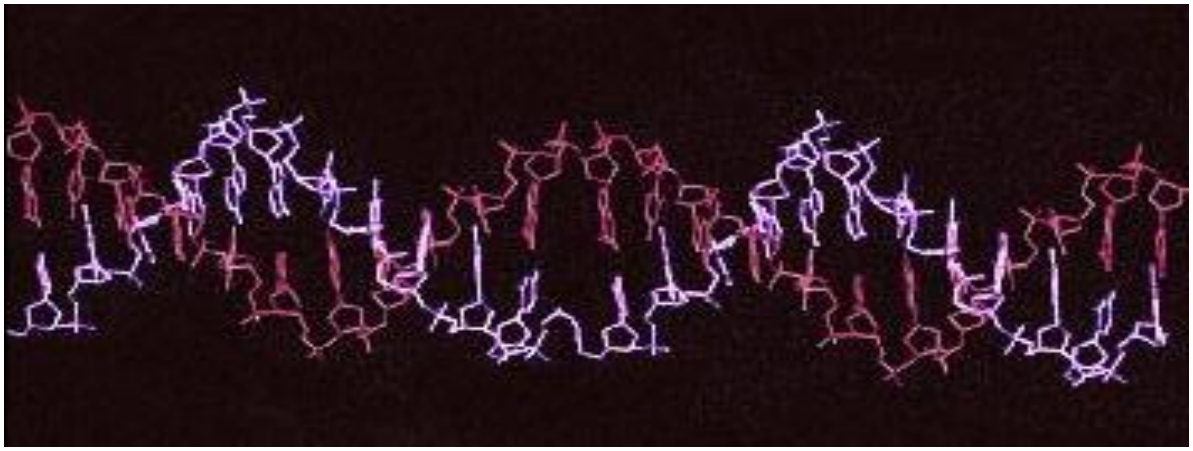


- In our cells, DNA consists of 2 side by side chains such that A on chain 1 pairs with T on the other chain, and G pairs with C

A, G, C, T



The 2 chains wind around each other, to form the "DNA double helix".



This diagram shows the shapes of the pairing nucleotide bases.

The bases are the parts that protrude inward like the rungs of a ladder

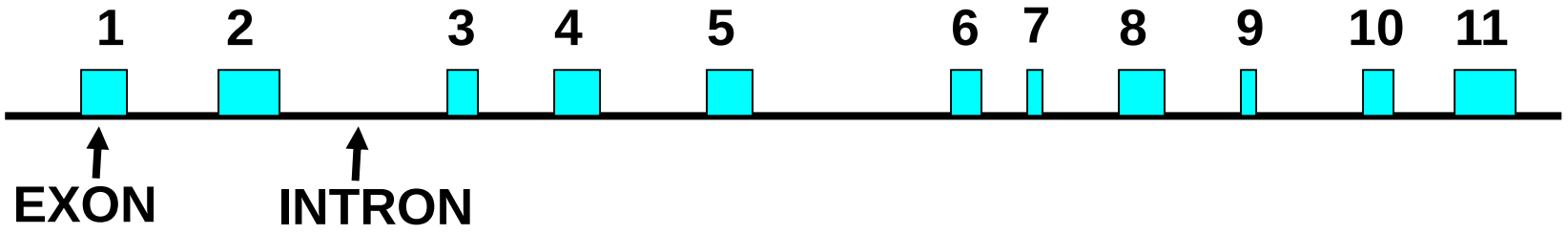
DNA resides in the nucleus of the cell and contains the instructions (recipes) for making the thousands of products that cells need.

Each recipe is a GENE.

One of these genes is TAZ, which codes for a protein called taffazin.

Barth syndrome is caused by changes / variations in the TAZ gene that lead to a taffazin protein with altered function.

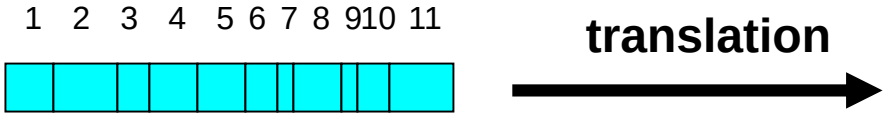
TAZ gene basic structure



Only the exons contain the "coding" for taffazin

In order to make the taffazin protein, TAZ DNA is first copied ("transcribed") into mRNA that looks like the DNA. The mRNA is then "spliced" to remove the introns so that the "coding text" is continuous and can be "translated" into protein

mRNA after intron removal



Tafazzin protein
292 amino acids

The TAZ gene DNA is a very specific recipe: the gene is made up of specific nucleotides in a specific order.

Groups of 3 nucleotides form a word (codon) in DNA and in the mRNA copy, specifying an amino acid of the TAZ protein.

Codons are just like words in the ingredient list of a cooking recipe

Genetic code

Note that the "T" of DNA becomes "U" in mRNA

First Position (5´end)	Second Position: U	Second Position: C	Second Position: A	Second Position: G	Third Position (3´ end)
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA *Stop*	UGA *Stop*	A
	UUG Leu	UCG Ser	UAG *Stop*	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

The beginning of the TAZ gene (DNA) and its coded amino acids:

ATG CCT CTG CAC GTG AAG TGG CCG TTC CCC ...
met pro leu his val lys trp pro phe pro ...

The beginning of the TAZ mRNA:

AUG CCU CUG CAC GUG AAG UGG CCG UUC CCC ...
met pro leu his val lys trp pro phe pro ...

There may be errors in a cooking recipe, like substituting garlic powder instead of baking powder – can you imagine the result ?

Or an important ingredient like yeast is skipped

Or an ingredient that does not belong is added, like oregano to a banana bread recipe

Similar errors can occur in DNA

1. Base substitutions are 1-letter typos –
their effect depends on the location of the change:

Change within a codon can lead to

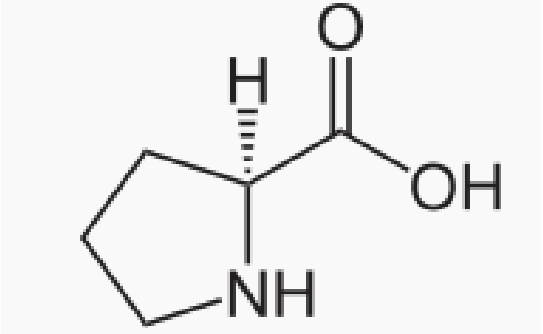
- **no amino acid change (synonymous), or**
- **amino acid change in the protein, or**
- **generation of a termination signal**

Change at or near the end of an exon can lead to

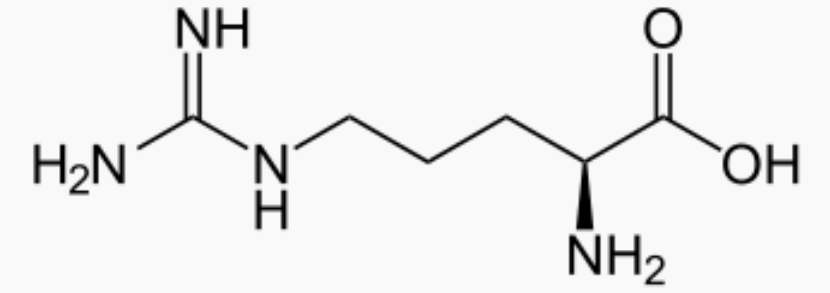
- **incorrect mRNA splicing such that an exon is skipped or an intron is retained**

1a. Single base change within a codon leading to amino acid substitution:

CCG (coding for proline)



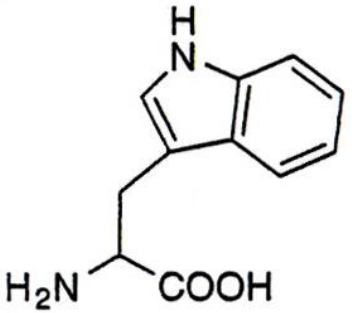
is changed to GCG (arginine)



Notice the difference in size and shape of these amino acids: this may or may not affect the structure and function of the tafazzin protein

1b. Single base change within a codon leading to termination (STOP) signal:

TGG codes for Tryptophan



Tryptophan (Trp, W)
MW: 186.21

G changed to A = TGA, a termination signal.
Protein product will be truncated after this point

2. Deletions and insertions:

a) One or more bases may be deleted from DNA.
b) One or more bases may be inserted into DNA.
Consequence: reading frame is shifted

2a. Normal beginning of TAZ:

ATG CCT CTG CAC GTG AAG TGG CCG TTC CCC ...
met pro leu his val lys trp pro phe pro ..

After deletion of 1 base, reading frame is shifted and termination generated:

ATG CTC TGC ACG TGA AGT GGC CGT TCC CC ...
met leu cys thr STOP

2b. Insertion of one base:

Normal beginning of TAZ:

ATG CCT CTG CAC GTG AAG TGG CCG TTC CC ...

met pro leu his val lys trp pro phe pro ..

After inserting the letter in red: coding is changed:

ATG CCT CTG CAC TGT GAA GTG GCC GTT CCC C

met pro leu his cys glu val ala val pro

2c) Large fragments of TAZ or even the entire gene may be deleted

1c single base change at/near splice signal:

Special splice signals (green) at edges of exons.

Example normal:

GCTTCTGGACCAgtgagtgggccca

EXON intron

Example change in DNA:

g to a change abolishes signal

GCTTCTGGACCAatgagtgggccca

EXON intron

Consequence: exon skipped or intron retained

DNA changes used to be called "mutations" and are now called "variants"

Both mutation and variant mean "a change", but "mutation" has a negative connotation in common parlance.

Recent sequencing of the DNA of many thousands of people has shown that our DNA is loaded with variants, yet we are all walking around and are mostly healthy

The term variant is used now because we have learned that not all changes are deleterious.

A deleterious change is a pathogenic variant.

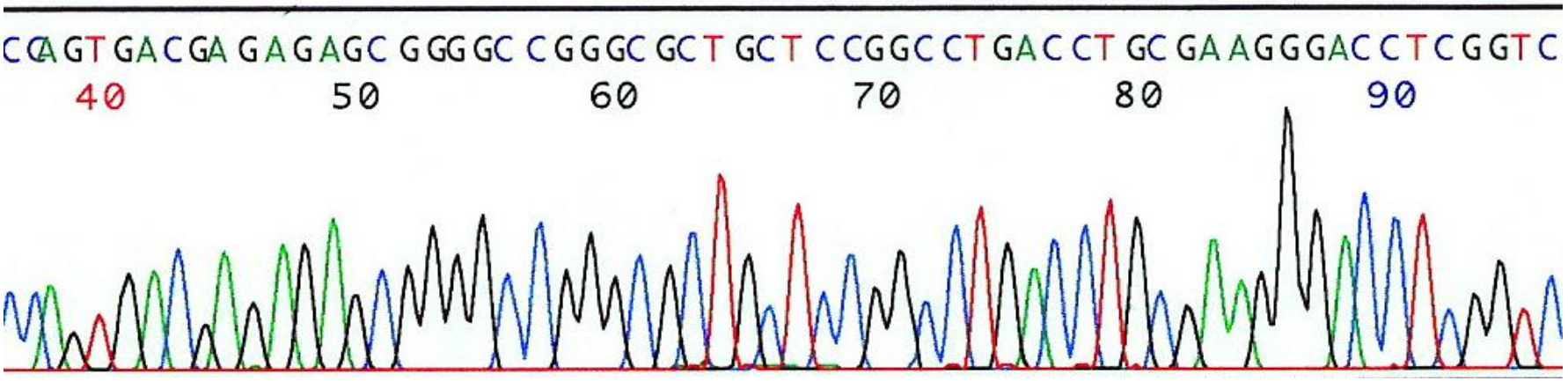
A benign variant is one that allows function.

And then there are "variants of unknown significance" (VUS), that are found in a gene and you don't know if they affect function.

How is it decided if a variant is pathogenic or benign?

First, how are variants found?

The diagnostic begins with "sequencing" the gene. This is a raw result (chromatogram) that comes from the "sequencer"



The computer also generates a text version with just the sequence that you see above the color peaks

The patient's DNA sequence is compared to the normal reference sequence using a computer program.

If there is a difference, determine location in exon or intron.

Does the variant change the coding?

Or may it affect the splicing?

... now comes VARIANT ASSESSMENT -- often not so easy!

VARIANT ASSESSMENT

There are national guidelines to help with assessment (on next slide).

A variety of analysis algorithms are used, some of which classify the variant as benign, likely benign, likely pathogenic, pathogenic, or unknown significance.
Depending on the variant, they may not all agree with each other.

A computer-generated "opinion" can only "suggest" -- the classification must be supported by specific patient information.

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very Strong**
Population Data	MAF frequency is too high for disorder *BS1* OR observation in controls inconsistent with disease penetrance *BS2*			Absent in 1000G and ESP *PM2*	Prevalence in affecteds statistically increased over controls *PS4*	
Computational And Predictive Data		Multiple lines of computational evidence suggest no impact on gene /gene product *BP4* Missense in gene where only truncating cause disease *BP1*	Multiple lines of computational evidence support a deleterious effect on the gene /gene product *PP3*	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before *PM5* In-frame indels in a non-repeat region or stop-loss variants *PM4*	Same amino acid change as an established pathogenic variant *PS1*	Truncating variant in a gene where LOF is a known mechanism of disease *PVS1*
Functional Data	Well-established functional studies show no deleterious effect *BS3*	In-frame indels in a repetitive region without a known function *BP3*	Missense in gene with low rate of benign missense variants and path. missenses common *PP2*	Located in a mutational hot spot and/or known functional domain *PM1*	Well-established functional studies show a deleterious effect *PS3*	
Segregation Data	Non-segregation with disease *BS4*		Co-segregation with disease in multiple affected family members *PP1*	Increased segregation data →		
De novo Data				*De novo* (without paternity & maternity confirmed) *PM6*	*De novo* (paternity & maternity confirmed) *PS2*	
Allelic Data		Observed in *trans* with a dominant variant *BP2* Observed in *cis* with a pathogenic variant *BP2*		For recessive disorders, detected in *trans* with a pathogenic variant *PM3*		
Other Database		Reputable database = benign *BP6*	Reputable database = pathogenic *PP5*			
Other Data		Found in case with an alternate cause *BP5*	Patient's phenotype or FH highly specific for gene *PP4*			

Pathogenic

Likely pathogenic

Uncertain significance

Likely benign

Benign

So, lab personnel will review the results of the analysis and ask questions:

Is there information about population frequency of the variant?

If population frequency is high, more than 5%, the variant is considered "likely benign" or "benign" -- that is a fast and easy one, and we all carry many such variants.

The clinical and family history information about the patient are VERY important:

- **does the presentation support the clinical diagnosis of Barth syndrome?**
- **Is there a suggestive family history of male infant deaths?**
- **A pedigree showing inheritance thru females and showing multiple affected males?**

This information is most often missing or scant because little or no information is given to the lab on the submission form, even when it is requested.

Has a similar variant been published in the literature?

Or are similar variants already in genomic databases? Note that genomic databases usually include frequency of the variant

This information is included in the patient's report along with the published classification

Have “functional studies” been described in the literature?

A study of the gene product activity can tell us either that the variant does not reduce activity, therefore benign,

or that it abolishes activity and is pathogenic

Every bit of information is assigned weight in order to include a classification in the report.

Sometimes there is insufficient support for either benign or pathogenic and this leads to the dreaded classification as Variant of Unknown Significance (VUS)

(*the techs in the photo were dealing with a VUS*)

In such cases, the lab may suggest a family study to follow "segregation" of the variant and of the condition thru generations:

If the variant is found in both affected and unaffected, it would be classified as "likely benign".

If found only in affected, it would be classified as "likely pathogenic".

Since the advent of diagnostic panels that test dozens of genes related to DCM at once, one reason for a VUS classification can be the finding of an additional variant in another gene that causes a syndrome with overlapping presentation.

In such a case, which variant "wins"?

Are both genes causing symptoms?

Does the patient have a "hybrid phenotype"?

?

?

?

?

?

?

?

The future will bring us answers to these and other questions

THE END