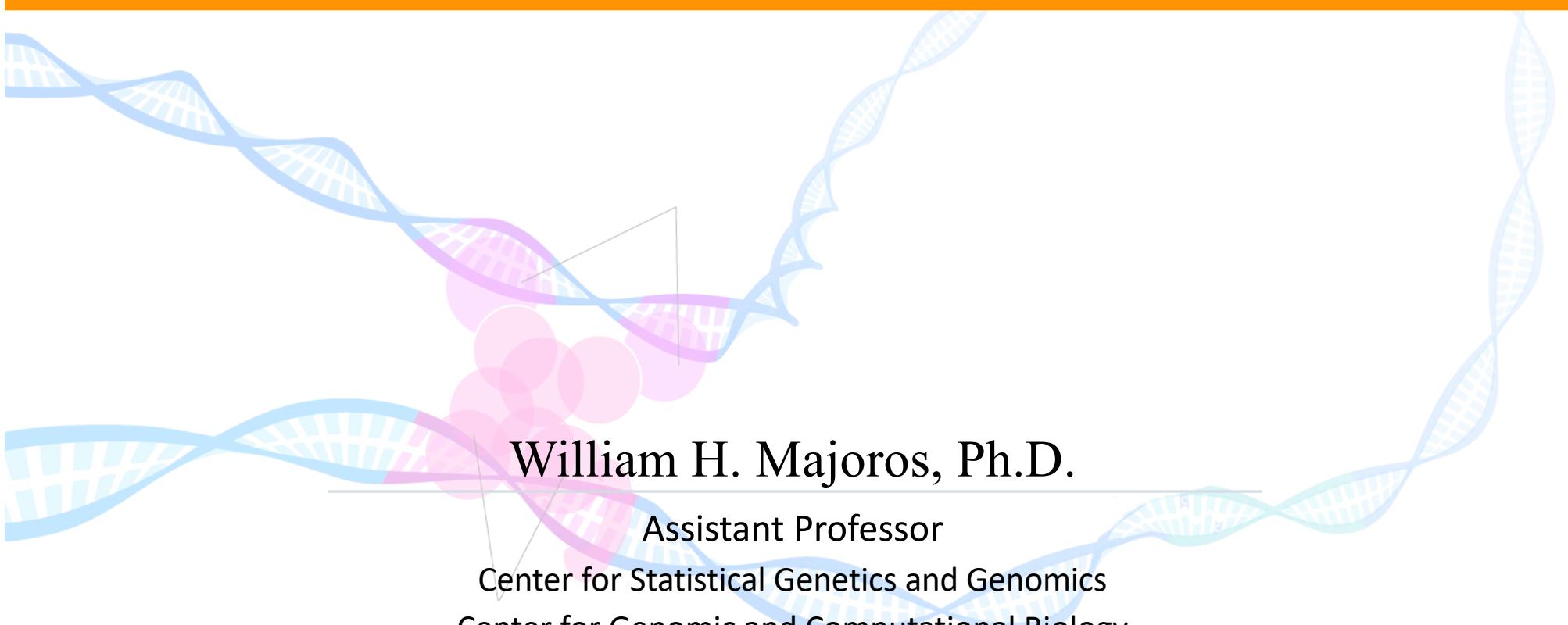


# Eukaryotic Gene Structure and Its Role in Genetic Disease

## Part 2: The Impact of Variants on Gene Structure



William H. Majoros, Ph.D.

Assistant Professor

Center for Statistical Genetics and Genomics

Center for Genomic and Computational Biology

Center for Advanced Genomic Technologies

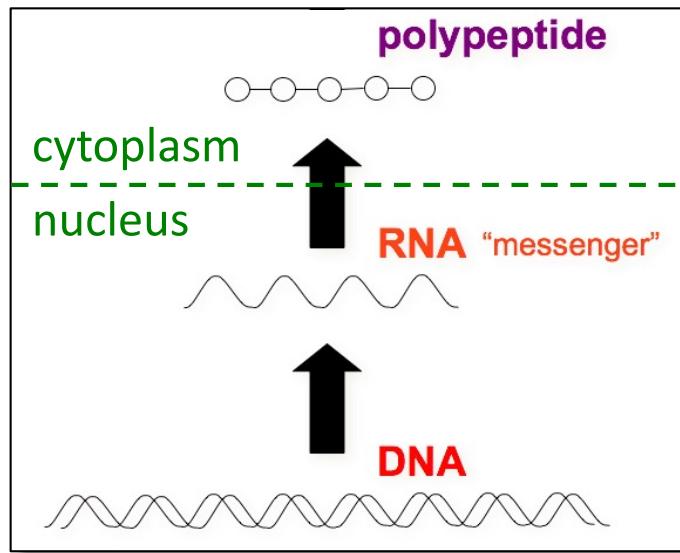
Duke University School of Medicine

[bmajoros@duke.edu](mailto:bmajoros@duke.edu)

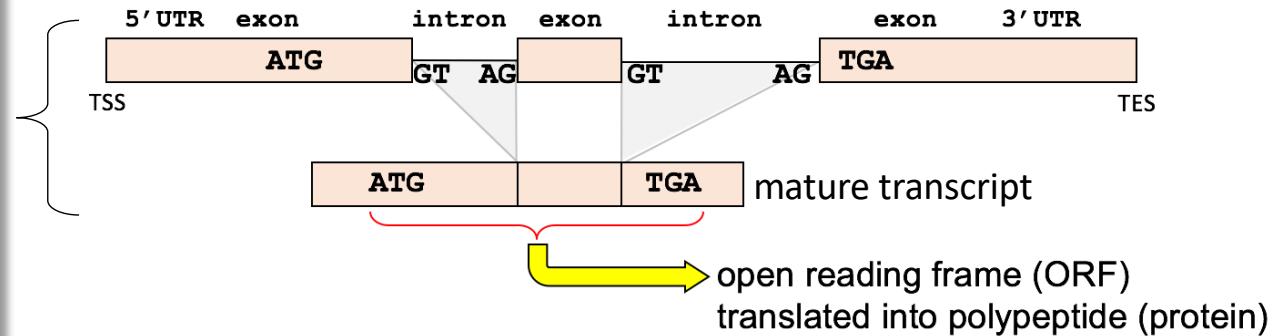
# Outline

- 1. Overview: Variants and Gene Structures**
2. Impact on Protein-coding Reading Frames
3. Impact on Splicing
4. Other Impacts on RNA

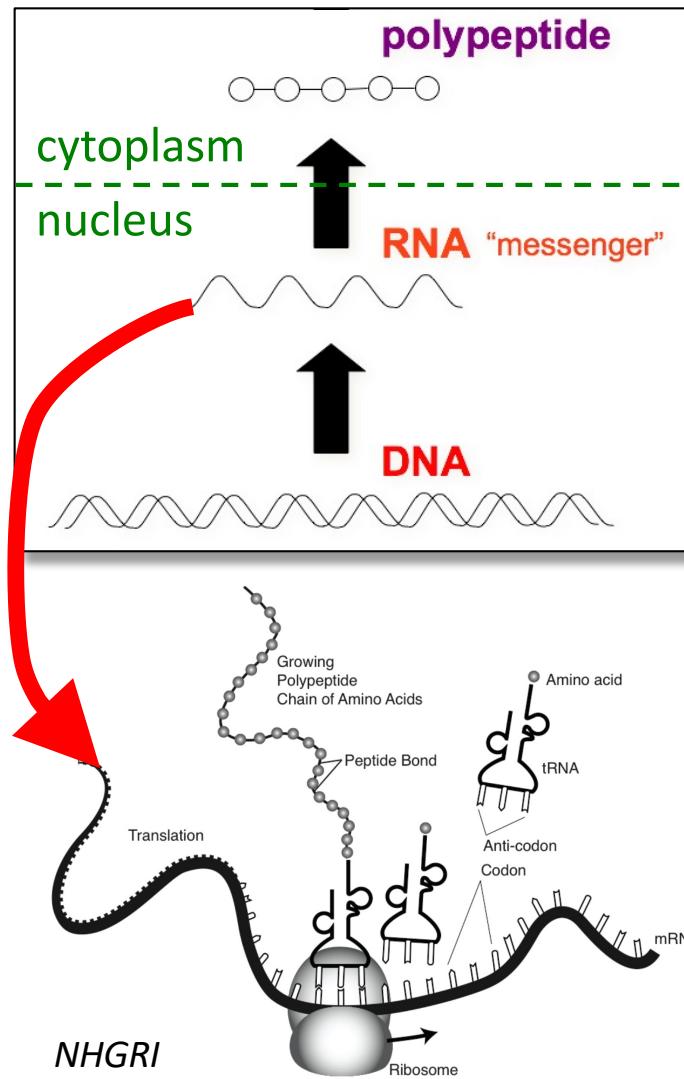
# Recall: Gene Structure



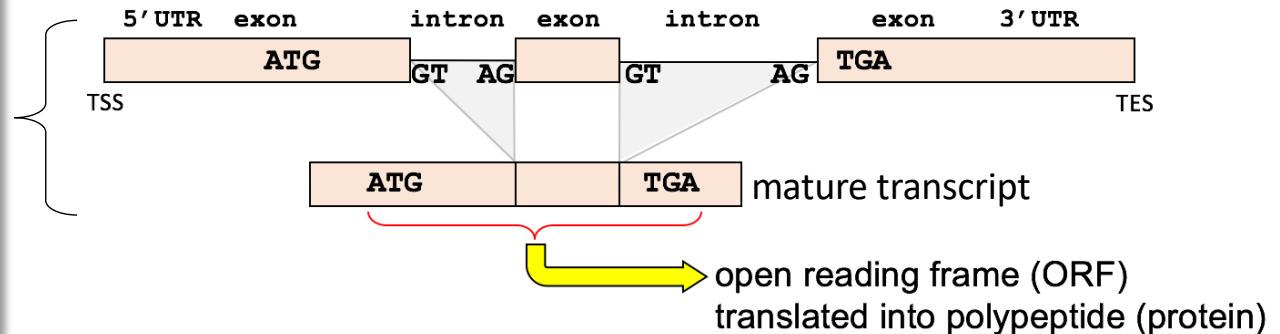
95% of human genes contain introns.



# Recall: Gene Structure



95% of human genes contain introns.



The way that a gene is spliced can impact how it is later translated.

Translation reading frames:

phase 0:

ATG GAC CAC CCA ATT GTG GTT GAG CAG CCA GAT GCC TGG ACA GAG GAC AAT GGC TTC **TGA**  
met asp his pro ile val val glu gln pro asp ala trp thr glu asp asn gly phe \*\*\*

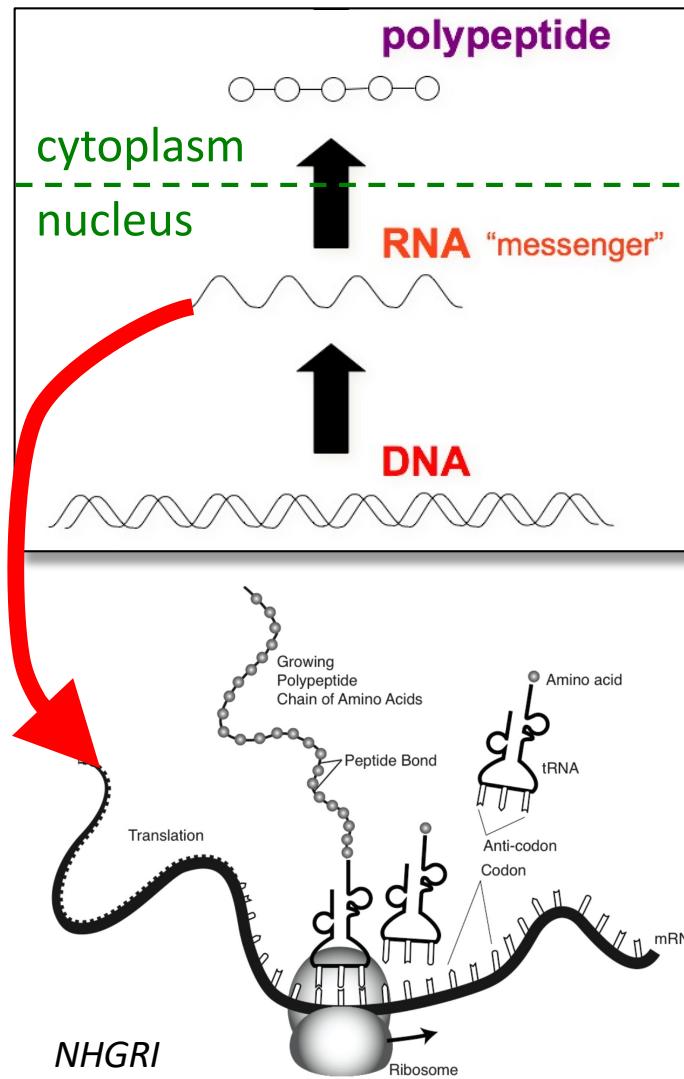
phase 2:

A TGG ACC ACC CAA TTG TGG TTG AGC AGC CAG ATG CCT GGA CAG AGG ACA ATG GCT TCT GA  
trp thr thr gln leu trp leu ser ser gln met pro gly gln arg thr met ala ser => no stop

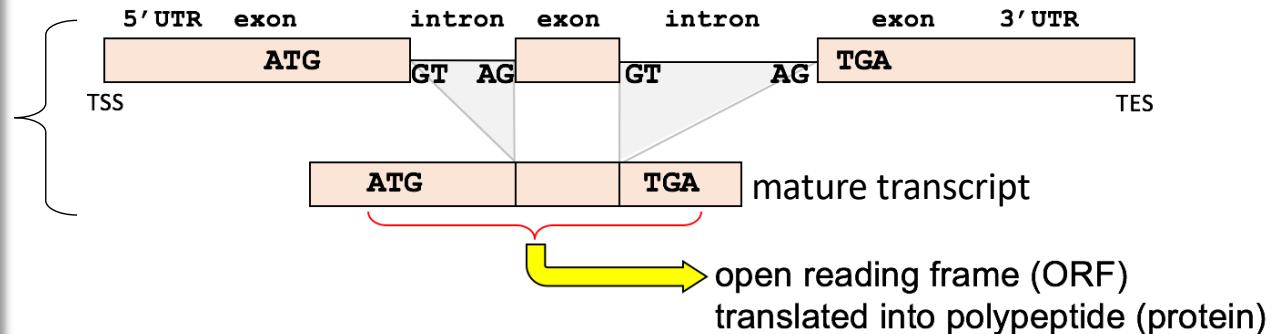
phase 1:

AT GGA CCA CCC AAT TGT GGT **TGA** GCA GCC AGA TGC CTG GAC AGA GGA CAA TGG CTT CCA TG  
gly pro pro asn cys gly \*\*\*

# Recall: Gene Structure



95% of human genes contain introns.



The way that a gene is spliced can impact how it is later translated.

Translation reading frames:

phase 0:

ATG GAC CAC CCA ATT GTG GTT GAG CAG CCA GAT GCC TGG ACA GAG GAC AAT GGC TTC **TGA**  
met asp his pro ile val val glu gln pro asp ala trp thr glu asp asn gly phe \*\*\*

phase 2:

A TGG ACC ACC CAA TTG TGG TTG AGC AGC CAG ATG CCT GGA CAG AGG ACA ATG GCT TCT GA  
trp thr thr gln leu trp leu ser ser gln met pro gly gln arg thr met ala ser => no stop

phase 1:

AT GGA CCA CCC AAT TGT GGT **TGA** GCA GCC AGA TGC CTG GAC AGA GGA CAA TGG CTT CCA TG  
gly pro pro asn cys gly \*\*\*

In order to know what protein is produced by a gene, we need to know the exact splicing pattern + reading frame = gene structure.

# Variants $\leftrightarrow$ Gene Structure

- Gene structure influences how we interpret the function of genetic variants
- Variants can alter gene structure

# Interpreting Variants in Genes

## **Gene structure influences variant interpretation:**

# **BRCA1**

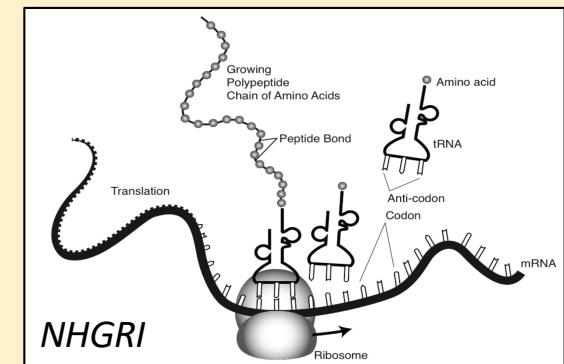
# Interpreting Variants in Genes

## Gene structure influences variant interpretation:

# BRCA1

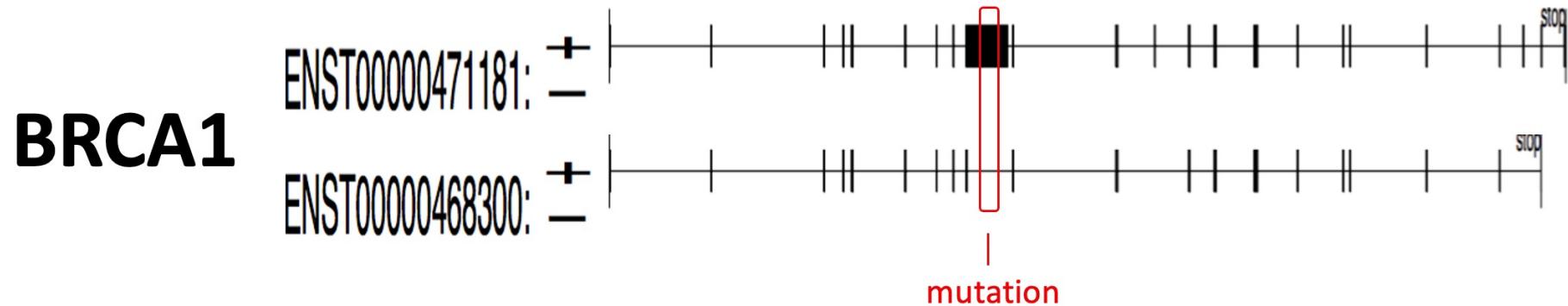
## Variants in exons can:

- Alter amino acids
  - Modify protein domains or signal peptides
  - Modify the reading frame
  - Alter splicing
  - etc.



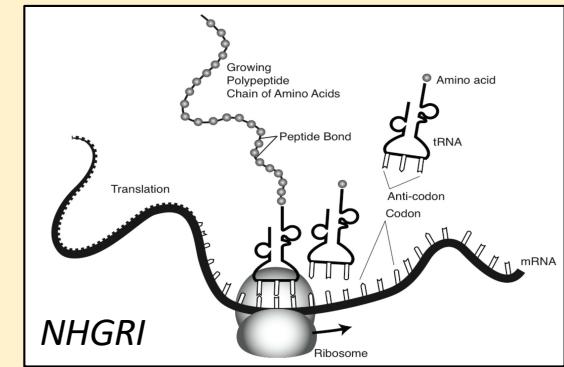
# Interpreting Variants in Genes

## Gene structure influences variant interpretation:



## Variants in exons can:

- Alter amino acids
  - Modify protein domains or signal peptides
  - Modify the reading frame
  - Alter splicing
  - etc.



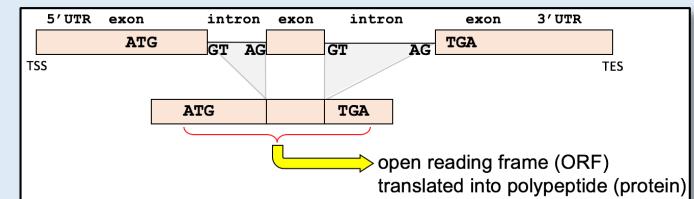
## **Variants in introns can:**

- Alter splicing via splicing regulation
  - Impact gene regulation (intronic enhancers)
  - Impact a miRNA gene within the intron

# Variants Can Alter Gene Structure

Variants can alter the protein-coding reading frame (“ORF”):

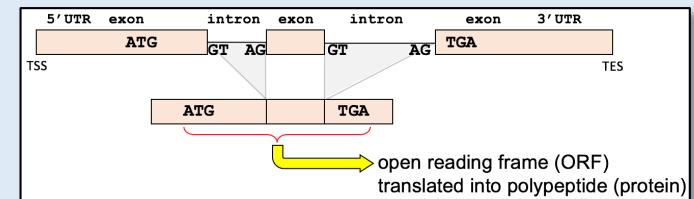
- Interrupt a start codon, or create a new start codon
- Interrupt a stop codon, or create a new stop codon
- Cause a frameshift (indels)



# Variants Can Alter Gene Structure

## Variants can alter the protein-coding reading frame (“ORF”):

- Interrupt a start codon, or create a new start codon
- Interrupt a stop codon, or create a new stop codon
- Cause a frameshift (indels)



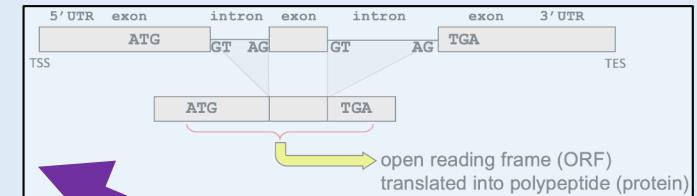
## Variants can alter splicing:

- Interrupt a splice site, or create a new splice site
- Alter splicing regulatory elements

# Variants Can Alter Gene Structure

## Variants can alter the protein-coding reading frame (“ORF”):

- Interrupt a start codon, or create a new start codon
- Interrupt a stop codon, or create a new stop codon
- Cause a frameshift (indels)



cascading  
effects

## Variants can alter splicing:

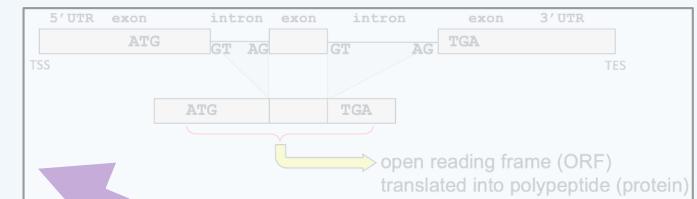
- Interrupt a splice site, or create a new splice site
- Alter splicing regulatory elements

Note that changes to splicing can change the reading frame!

# Variants Can Alter Gene Structure

## Variants can alter the protein-coding reading frame (“ORF”):

- Interrupt a start codon, or create a new start codon
- Interrupt a stop codon, or create a new stop codon
- Cause a frameshift (indels)



cascading  
effects

## Variants can alter splicing:

- Interrupt a splice site, or create a new splice site
- Alter splicing regulatory elements

Note that changes to splicing can change the reading frame!

## Variants can do other things:

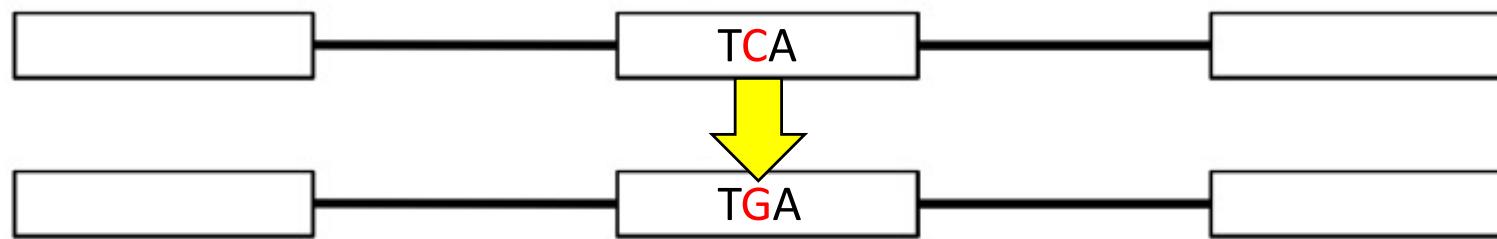
- Create secondary structures such as hairpin loops
- Create or interrupt upstream open reading frames (uORF)
- Create or interrupt internal ribosome entry sites (IRES)

# Outline

1. Overview: Variants and Gene Structures
- 2. Impact on Protein-coding Reading Frames**
3. Impact on Splicing
4. Other Impacts on RNA

# Variants Can Modify the Reading Frame

A common way to modify the reading frame is to introduce a pre-termination codon (PTC):

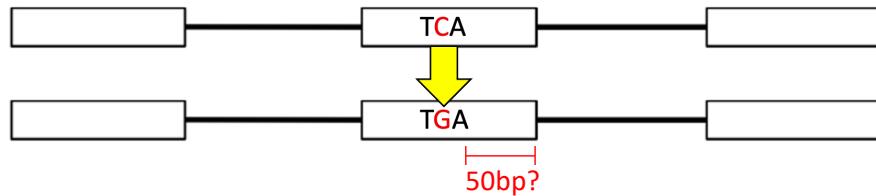


- A PTC is only functional if it occurs in the translation reading frame.
- If the PTC occurs late in the gene, it can result in protein truncation, which may or may not be deleterious—via loss of function or gain of function (e.g., *dominant-negative*).
- If it occurs early in the gene, it can result in nonsense-mediated decay (NMD) . . .

# Nonsense Mediated Decay (NMD)

The 50 bp “rule”:

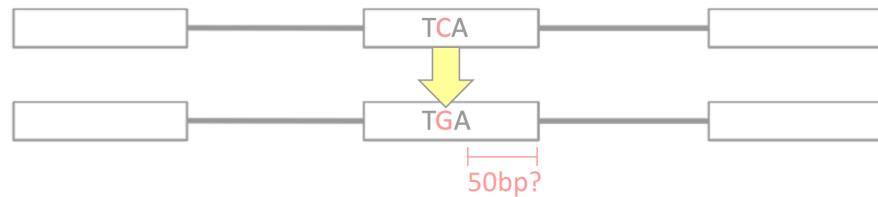
- PTC at least 50-55 bp upstream of last exon junction
- But PTCs far upstream can escape NMD



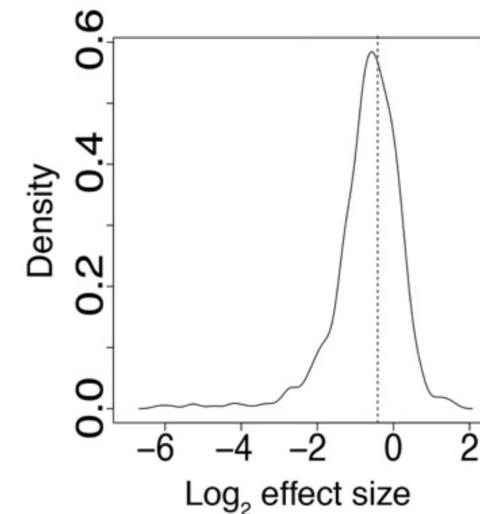
# Nonsense Mediated Decay (NMD)

## The 50 bp “rule”:

- PTC at least 50-55 bp upstream of last exon junction
- But PTCs far upstream can escape NMD



## NMD effect size varies:



(Majoros, 1017)  
(also: Rosenberg, 2015)

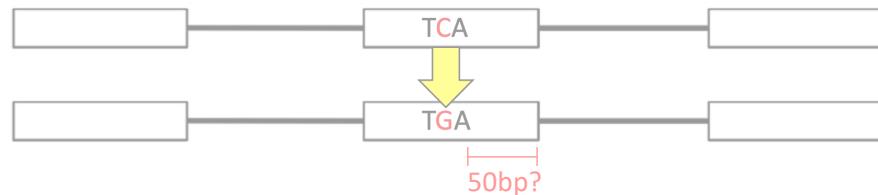
→ On average, NMD results in roughly a halving of expression of each affected copy

→ The remaining undegraded transcripts can have deleterious gain-of-function effects (e.g., dominant-negative)

# Nonsense Mediated Decay (NMD)

## The 50 bp “rule”:

- PTC at least 50-55 bp upstream of last exon junction
- But PTCs far upstream can escape NMD



## Some PTCs escape NMD:

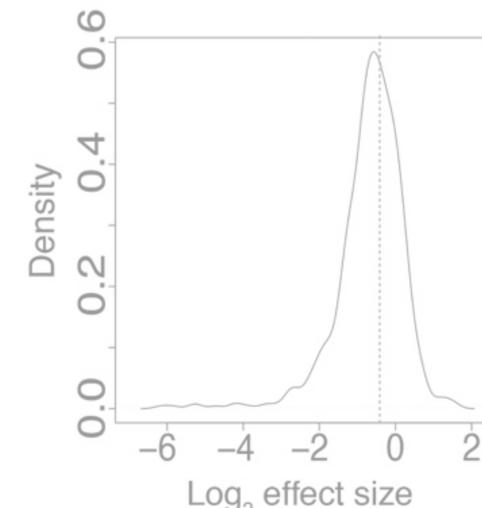
### A Novel *FLCN* c.1489\_1490delTG Mutation that Escapes the Nonsense-Mediated Decay System

Yong-Jin Park<sup>1</sup>, Seog-Ki Lee<sup>2</sup>, Seong-Ho Kang<sup>3</sup>, Sook-Jin Jang<sup>3</sup>, Dae-Soo Moon<sup>3</sup>, and Geon Park<sup>3</sup>

<sup>1</sup>Department of Emergency Medicine, Chosun University College of Medicine, Gwangju, <sup>2</sup>Department of Thoracic and Cardiovascular Surgery, Chosun University College of Medicine, Gwangju, and <sup>3</sup>Department of Laboratory Medicine, Chosun University College of Medicine, Gwangju, South Korea

**Abstract.** A novel *FLCN* c.1489\_1490delTG (p.Val497Glyfs\*22) mutation at the genomic DNA and mRNA levels was identified in a 43-year-old woman with complaining of recurrent primary spontaneous pneumothorax. The aberrant *FLCN* mRNA escaped the nonsense-mediated decay system (NMD) because of a premature termination code located in an NMD-incompetent region. To the best of our knowledge, this is the first case report of an *FLCN* mutation escaping the NMD.

## NMD effect size varies:



(Majoros, 2017)  
(also: Rosenberg, 2015)

→ On average, NMD results in roughly a halving of expression of each affected copy

→ The remaining undegraded transcripts can have deleterious gain-of-function effects (e.g., dominant-negative)

# Variants Can Cause Frameshifts

## Translation reading frames:

phase 0:

```
ATG GAC CAC CCA ATT GTG GTT GAG CAG CCA GAT GCC TGG ACA GAG GAC AAT GGC TTC TGA
met asp his pro ile val val glu gln pro asp ala trp thr glu asp asn gly phe ***
```

phase 2:

```
A TGG ACC ACC CAA TTG TGG TTG AGC AGC CAG ATG CCT GGA CAG AGG ACA ATG GCT TCT GA
trp thr thr gln leu trp leu ser ser gln met pro gly gln arg thr met ala ser => no stop
```

phase 1:

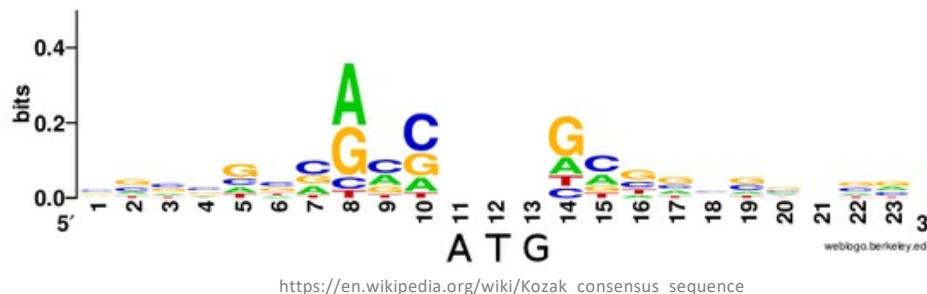
```
AT GGA CCA CCC AAT TGT GGT TGA GCA GCC AGA TGC CTG GAC AGA GGA CAA TGG CTT CCA TG
gly pro pro asn cys gly ***
```

- Indels (insertions/deletions) can cause a frameshift if the length is not divisible by 3
- Splicing changes can also cause frameshifts
- Changes to start codons can cause frameshifts

# Variants Can Change the Start Codon



**Ribosome scanning model:** the ribosome scans for the first start codon with a strong Kozak signal (“5' cap-dependent translation”)



Heterozygous *SSBP1* start loss mutation co-segregates with hearing loss and the m.1555A>G mtDNA variant in a large multigenerational family ♂

Peter J Kullar, Aurora Gomez-Duran, Payam A Gammage, Caterina Garone, Michal Minczuk, Zoe Golder, Janet Wilson, Julio Montoya, Sanna Häkli, Mikko Kärppä ... Show more

Brain, Volume 141, Issue 1, January 2018, Pages 55–62, <https://doi.org/10.1093/brain/awx295>

Published: 22 November 2017 Article history ▾

[PDF](#) [Split View](#) [Cite](#) [Permissions](#) [Share ▾](#)

## Abstract

The m.1555A>G mtDNA variant causes maternally inherited deafness, but the reasons for the highly variable clinical penetrance are not known. Exome sequencing identified a heterozygous start loss mutation in *SSBP1*, encoding the single stranded binding protein 1 (*SSBP1*), segregating with hearing loss in a multi-generational family transmitting m.1555A>G, associated with mtDNA depletion and multiple deletions in skeletal muscle. The *SSBP1* mutation reduced steady state *SSBP1* levels leading to a perturbation of mtDNA metabolism, likely compounding the intra-mitochondrial translation defect due to m.1555A>G in a tissue-specific manner. This family demonstrates the importance of rare *trans*-acting genetic nuclear modifiers in the clinical expression of mtDNA disease.

# Variants Can Disrupt the Stop Codon



No stop codon – ribosomes pile up at the poly-A tail, triggering “non-stop decay” (NSD).

## Non-stop decay—a new mRNA surveillance pathway

Shobha Vasudevan, Stuart W. Peltz, and Carol J. Wilusz\*

BioEssays 24:785–788, © 2002 Wiley Periodicals, Inc.

BioEssays 24.9 785

### Summary

Gene expression is an inherently complex process and errors often occur during the transcription and processing of mRNAs. Several surveillance mechanisms have evolved to check the fidelity at each step of mRNA manufacture. Two recent reports describe the identification of a novel pathway in eukaryotes that recognizes and degrades mRNAs that lack a stop codon.<sup>(1,2)</sup> The non-stop decay mechanism releases ribosomes stalled at the 3' end of a mRNA and stimulates the exosome to rapidly degrade the transcript. *BioEssays* 24:785–788, 2002.

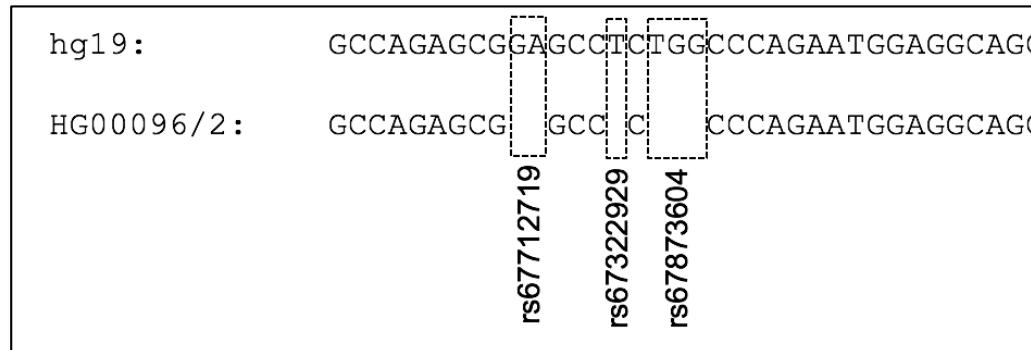
© 2002 Wiley Periodicals, Inc.

# Variants are Not Independent!

The first two deletions shown below would shift the reading frame, but together they affect only two amino acids:

hg19:	GCCAGAGCGGAGCCTCTGGCCCAGAATGGAGGCAGC
HG00096/2:	GCCAGAGCGGCC C CCCAGAATGGAGGCAGC

rs67712719      rs67322929      rs67873604



# Variants are Not Independent!

The first two deletions shown below would shift the reading frame, but together they affect only two amino acids:

hg19:	GCCAGAGCGGAGCCTCTGGCCCAGAATGGAGGCAGC
HG00096/2:	GCCAGAGCGGCC C CCCAGAATGGAGGCAGC

rs67712719      rs67322929      rs67873604

4869 / 5008 = 97% of  
Thousand Genomes  
haplotypes contain the  
first two variants

In Thousand Genomes, the first two variants overwhelmingly occur together. Because that haplotype is so common, it is unlikely to be deleterious.

# Variants are Not Independent!

The first two deletions shown below would shift the reading frame, but together they affect only two amino acids:



4869 / 5008 = 97% of  
Thousand Genomes  
haplotypes contain the  
first two variants

In Thousand Genomes, the first two variants overwhelmingly occur together. Because that haplotype is so common, it is unlikely to be deleterious.

**Ensembl's VEP tool predicts that the first two variants would be deleterious.**

**This is not anecdotal!** Every individual in the 1000 Genomes Project sample has one or more compensatory frameshifts (median = 7 per individual) affecting ≤30 amino acids.

# Example Continued

hg19:	GCCAGAGCGGAGCCTCTGGCCCAGAATGGAGGCAGC
HG00096/2:	GCCAGAGCGGCC CCCCAGAATGGAGGCAGC

rs67712719      rs67322929      rs67873604

reference: AGA GCG GAG CCT CTG GCC CAG AAT GGA GGC AGC ... (1662bp) ... TGA  
1st deletion: AGA GCG **GCC TCT GGC CCA GAA TGG AGG CAG CAG** ... (651bp) ... **TAG** => truncation

reference: AGA GCG GAG CCT CTG GCC CAG AAT GGA GGC AGC ... (1662bp) ... TGA  
2nd deletion: AGA GCG GAG **CCC TGG CCC AGA ATG GAG GCA GCA** ... (1032bp) ... **TGA** => truncation

All three variants:

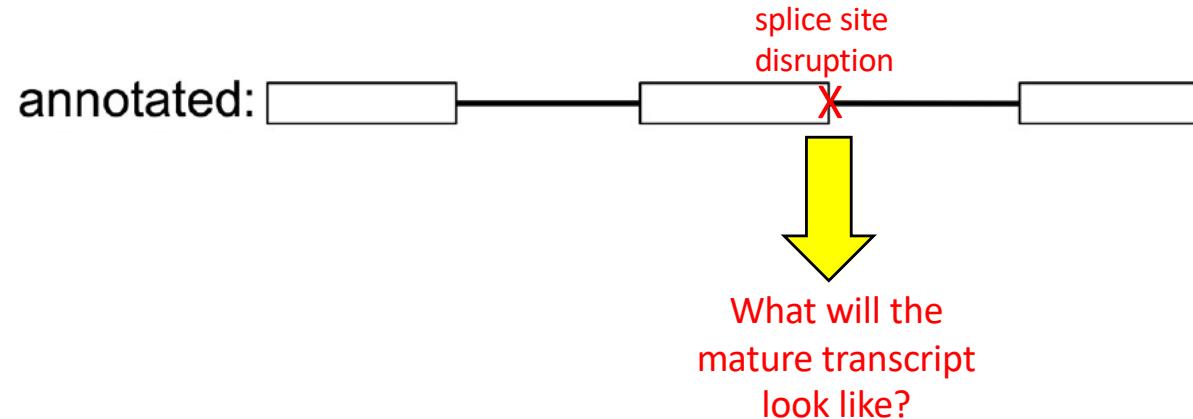
reference: AGA GCG GAG CCT CTG GCC CAG AAT GGA GGC AGC ... (1662bp) ... TGA  
HG00096/2: AGA GCG **GCC CCC CAG AAT GGA GGC AGC** ... (1662bp) ... TGA => local protein change (4aa)

# Outline

1. Overview: Variants and Gene Structures
2. Impact on Protein-coding Reading Frames
- 3. Impact on Splicing**
4. Other Impacts on RNA

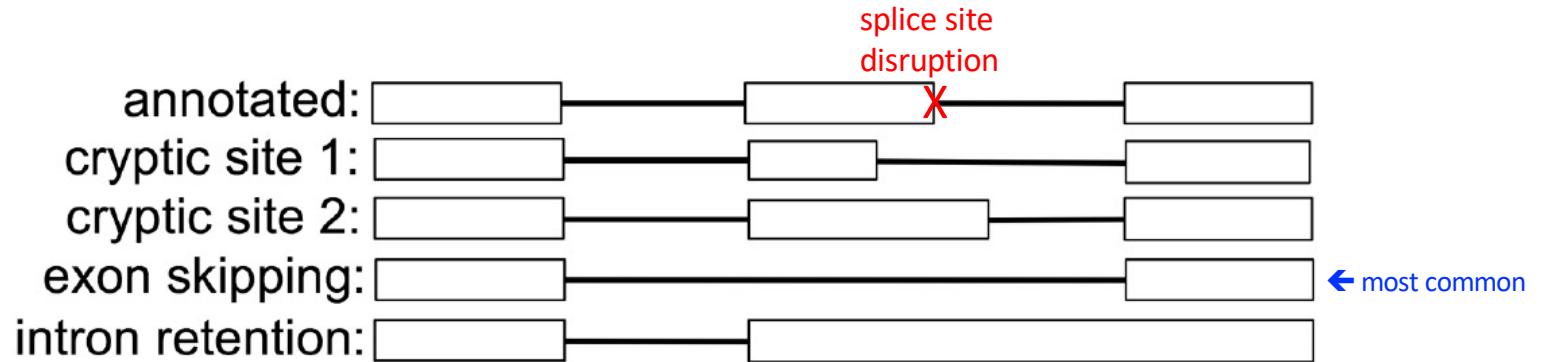
# Changes to Splicing

Variants that disrupt an existing splice site can have multiple possible outcomes:



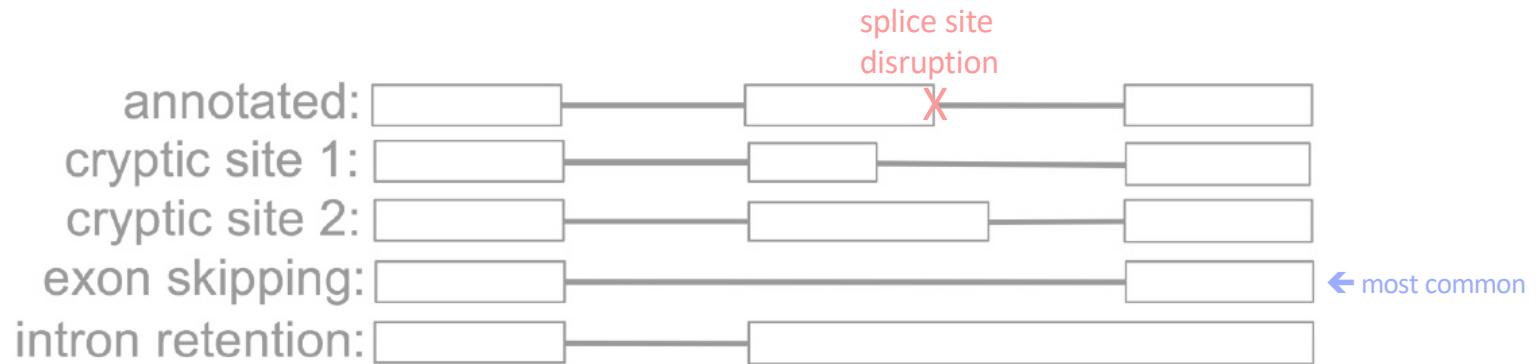
# Changes to Splicing

Variants that disrupt an existing splice site can have multiple possible outcomes:



# Changes to Splicing

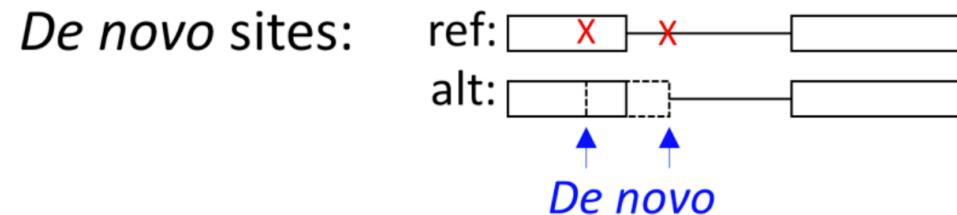
Variants that disrupt an existing splice site can have multiple possible outcomes:



**Splicing is stochastic!** One variant can result in a mixture of transcripts with different structures in the same cell.

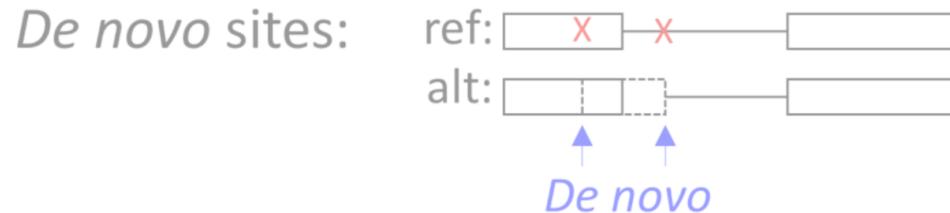
# De Novo Splice Sites

A variant can create an entirely new splice site – this is not as unlikely as you might think!



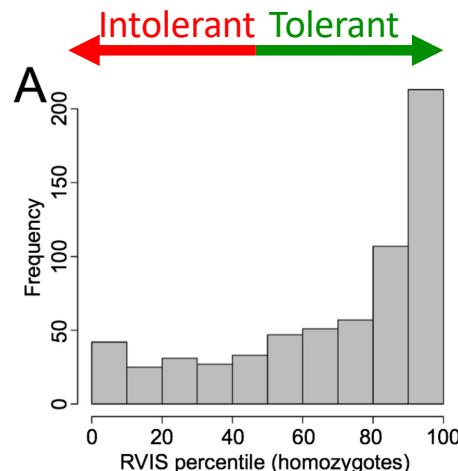
# De Novo Splice Sites

A variant can create an entirely new splice site – this is not as unlikely as you might think!



**In Geuvadis (Thousand Genomes), each individual has on average 127 *de novo* splice sites supported by RNA-seq data.**

There is a strong bias for those variants to be in genes tolerant to mutation:



# De Novo Splice Sites Can Cause Disease!

**DBASS**  
Data Base of  
Aberrant  
Splice  
Sites

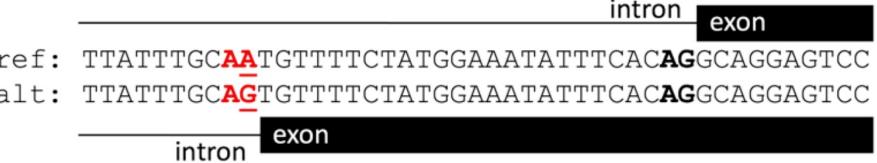
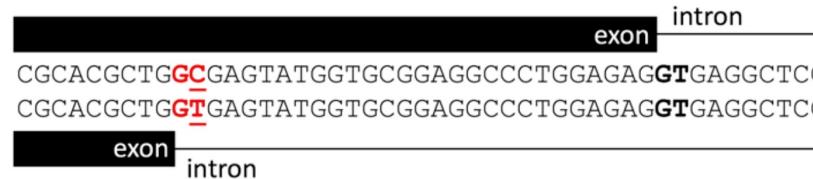
intron exon  
ref: TTATTTGCAATGTTTCTATGGAAATATTCACAGGCAGGAGTCC  
alt: TTATTTGCAGTGTTTCTATGGAAATATTCACAGGCAGGAGTCC

intron exon

**disease:** cystic fibrosis  
**gene:** CFTR  
**mutation:** IVS17a-26A>G  
**ACE+ probability:** 0.64  
**predicted fate:** NMD

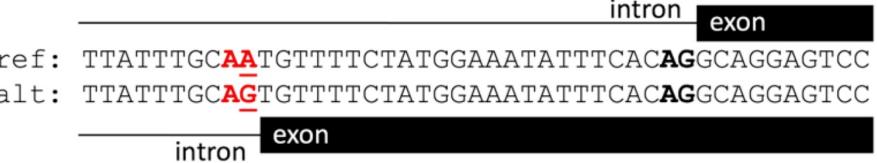
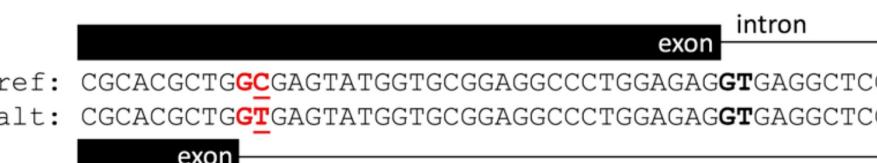
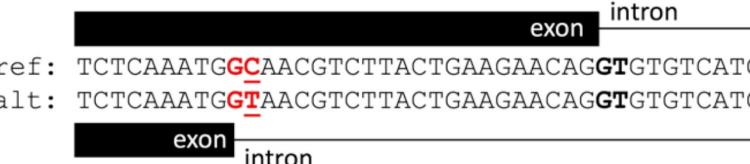
# De Novo Splice Sites Can Cause Disease!

**DBASS**  
Data Base of  
Aberrant  
Splice  
Sites

	<b>disease:</b> cystic fibrosis <b>gene:</b> CFTR <b>mutation:</b> IVS17a-26A>G <b>ACE+ probability:</b> 0.64 <b>predicted fate:</b> NMD
	<b>disease:</b> alpha-thalassemia <b>gene:</b> HBA2 <b>mutation:</b> E1+135C>T <b>ACE+ probability:</b> 0.62 <b>predicted fate:</b> NMD

# De Novo Splice Sites Can Cause Disease!

**DBASS**  
Data Base of  
Aberrant  
Splice  
Sites

	<b>disease:</b> cystic fibrosis <b>gene:</b> CFTR <b>mutation:</b> IVS17a-26A>G <b>ACE+ probability:</b> 0.64 <b>predicted fate:</b> NMD
	<b>disease:</b> alpha-thalassemia <b>gene:</b> HBA2 <b>mutation:</b> E1+135C>T <b>ACE+ probability:</b> 0.62 <b>predicted fate:</b> NMD
	<b>disease:</b> Duchenne musc. dystr. <b>gene:</b> DMD <b>mutation:</b> E14+82C>T <b>ACE+ probability:</b> 0.47 <b>predicted fate:</b> NMD

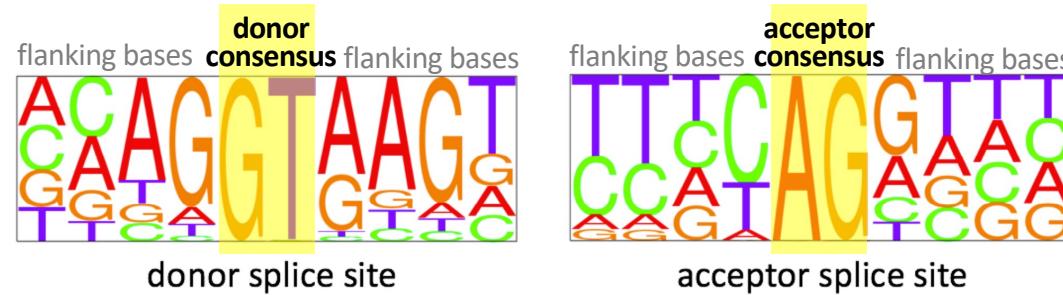
# De Novo Splice Sites Can Cause Disease!

**DBASS**  
Data Base of  
Aberrant  
Splice  
Sites

<p>ref: TTATTTGC<u>AA</u>TGTTTCTATGGAAATATTCAC<u>AGG</u>CAGGAGTCC alt: TTATTTGC<u>AG</u>TGTTTCTATGGAAATATTCAC<u>AGG</u>CAGGAGTCC</p>	<b>disease:</b> cystic fibrosis <b>gene:</b> CFTR <b>mutation:</b> IVS17a-26A>G <b>ACE+ probability:</b> 0.64 <b>predicted fate:</b> NMD
<p>ref: CGCACGCTG<u>GC</u>GAGTATGGTGC<u>GGAGG</u>CCCTGGAGAG<u>GT</u>GAGGCTCC alt: CGCACGCTG<u>GT</u>GAGTATGGTGC<u>GGAGG</u>CCCTGGAGAG<u>GT</u>GAGGCTCC</p>	<b>disease:</b> alpha-thalassemia <b>gene:</b> HBA2 <b>mutation:</b> E1+135C>T <b>ACE+ probability:</b> 0.62 <b>predicted fate:</b> NMD
<p>ref: TCTCAAAT<u>GC</u>AACGTCTTACTGAAGAACAG<u>GT</u>GTGT<u>CATG</u> alt: TCTCAAAT<u>GT</u>AACGTCTTACTGAAGAACAG<u>GT</u>GTGT<u>CATG</u></p>	<b>disease:</b> Duchenne musc. dystr. <b>gene:</b> DMD <b>mutation:</b> E14+82C>T <b>ACE+ probability:</b> 0.47 <b>predicted fate:</b> NMD
<p>ref: TTCTCAAAC<u>AA</u>TTTAATT<u>TC</u>AGGAGCCTACAA alt: TTCTCAAAC<u>AG</u>TTTAATT<u>TC</u>AGGAGCCTACAA</p>	<b>disease:</b> breast cancer <b>gene:</b> BRCA1 <b>mutation:</b> IVS5-12A>G <b>ACE+ probability:</b> 0.21 <b>predicted fate:</b> NMD

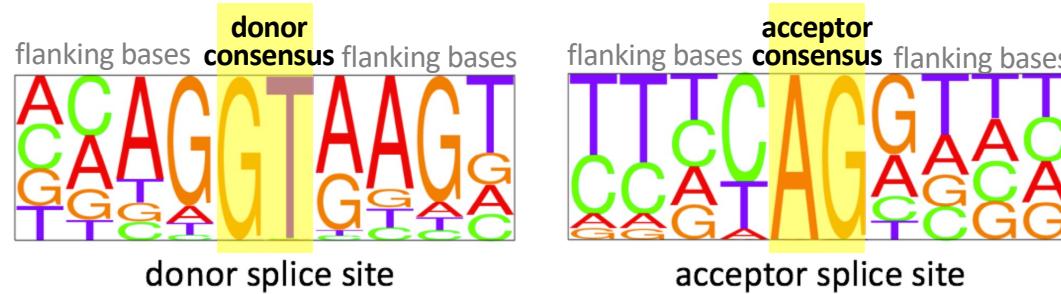
# Variants Can Alter Splice Site Strength

A mutation that does not change the donor (GT) or acceptor (AG) consensus, but changes one or more flanking bases can strengthen or weaken an existing splice site:

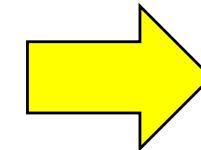


# Variants Can Alter Splice Site Strength

A mutation that does not change the donor (GT) or acceptor (AG) consensus, but changes one or more flanking bases can strengthen or weaken an existing splice site:



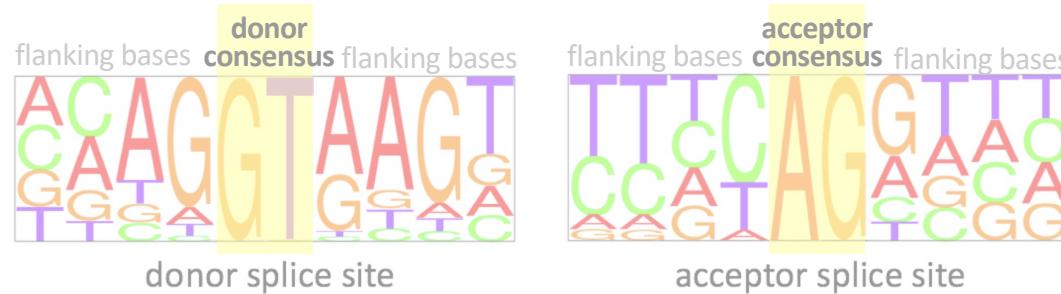
- Weaken an existing splice site
- Strengthen a cryptic splice site



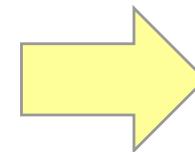
Change isoform ratios

# Variants Can Alter Splice Site Strength

A mutation that does not change the donor (GT) or acceptor (AG) consensus, but changes one or more flanking bases can strengthen or weaken an existing splice site:



- Weaken an existing splice site
- Strengthen a cryptic splice site



Change isoform ratios

› Nat Biotechnol. 2004 May;22(5):535-46. doi: 10.1038/nbt964.

## Alternative splicing in disease and therapy

Mariano A Garcia-Blanco <sup>1</sup>, Andrew P Baraniak, Erika L Lasda

Affiliations + expand

PMID: 15122293 DOI: [10.1038/nbt964](https://doi.org/10.1038/nbt964)

→ mutations that lead to even subtle changes in the ratio of MAPT isoforms 3R and 4R cause an inherited form of dementia

# Context Is Important!

The local context of a splice-altering variant (SAV) is important.

# Context Is Important!

The local context of a splice-altering variant (SAV) is important.

In the example below, an entire splice site is deleted, but there is a “cryptic splice” site nearby, and after the deletion the cryptic site is actually predicted to be stronger than the original site.

hg19:	TGTGTACAg <b>GT</b> GTGGGTGTGTGTGGG
HG00096/1/2:	TGTGTACA----- <b>GT</b> GTGTGTGGG
	<small>rs11278302</small>

A dashed box highlights the deleted sequence in hg19, and a red label "cryptic site" points to the new splice site in HG00096/1/2.

# Context Is Important!

The local context of a splice-altering variant (SAV) is important.

In the example below, an entire splice site is deleted, but there is a “cryptic splice” site nearby, and after the deletion the cryptic site is actually predicted to be stronger than the original site.

hg19: TGTGTACAgGTGTGGGTGTGTGTGGG  
HG00096/1/2: TGTGTACA-----GTGTGTGTGGG  
rs11278302



New splice site is stronger than the original

Allele frequency = 0.23

The result is a loss of two amino acids and no change to the reading frame.

Because the variant is common, it is unlikely to be deleterious.

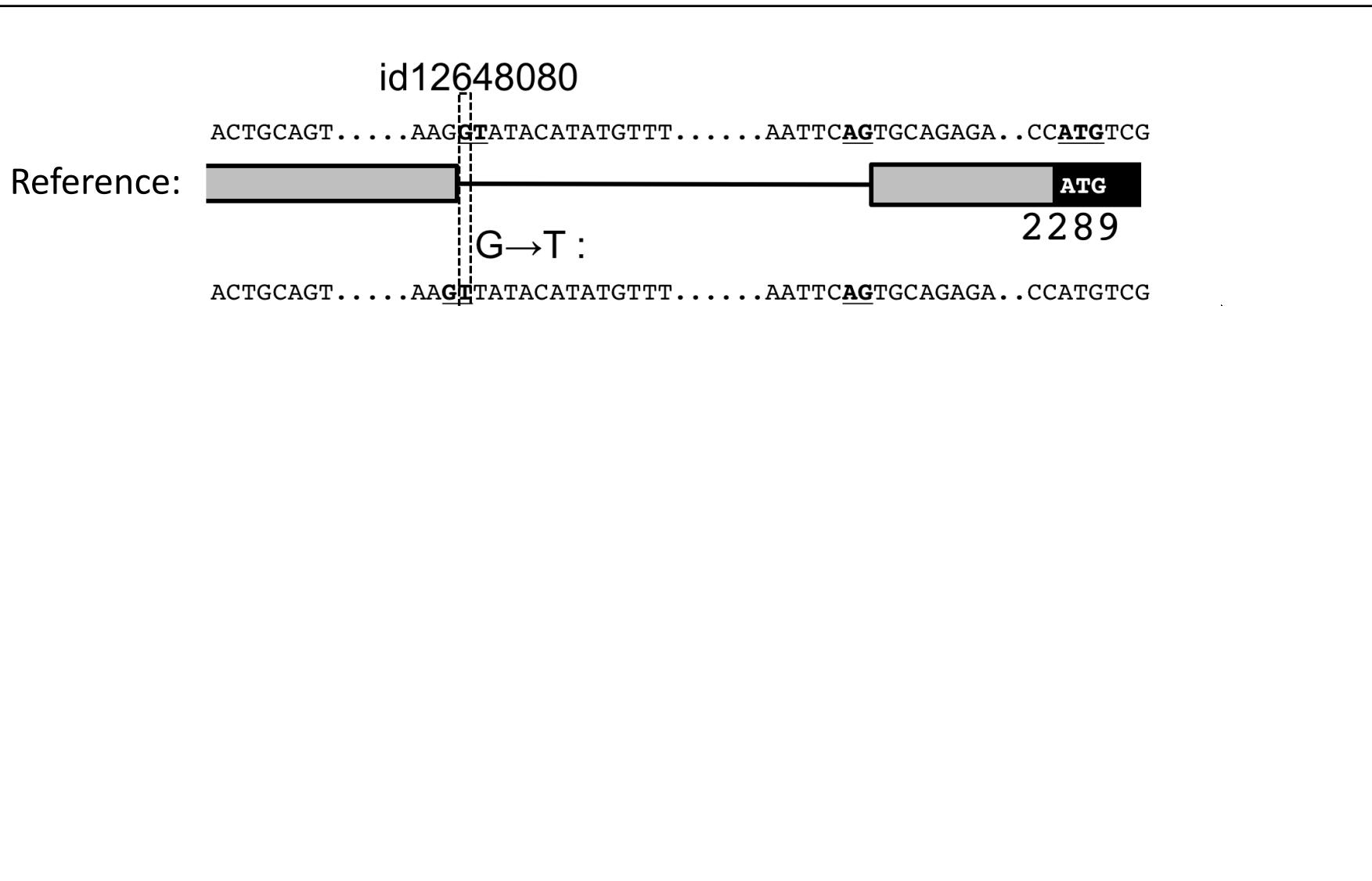
**Ensembl's VEP tool predicts that this variant would be deleterious, because it deletes an entire splice site.**

# Cascading Effects

**Changes to splicing often have cascading downstream effects:**

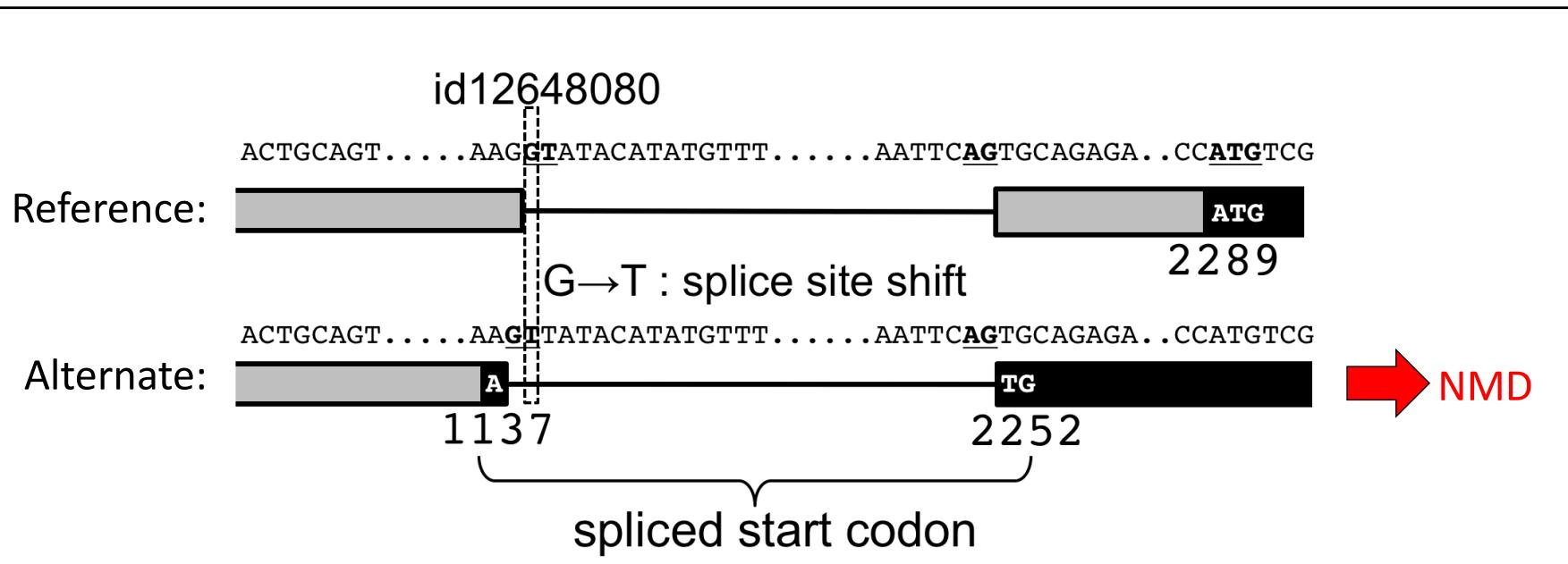
# Cascading Effects

Changes to splicing often have cascading downstream effects:



# Cascading Effects

Changes to splicing often have cascading downstream effects:



Example:

- 1) Changing G to T disrupts a splice site.
- 2) The same variant also creates a new splice site 1 bp upstream.
- 3) That shortens the exon by 1 bp, and after splicing a new start codon is created.
- 4) That new start codon establishes a different reading frame.
- 5) The new reading frame is shorter and triggers NMD.

# Variants Can Alter Splicing Regulatory Elements

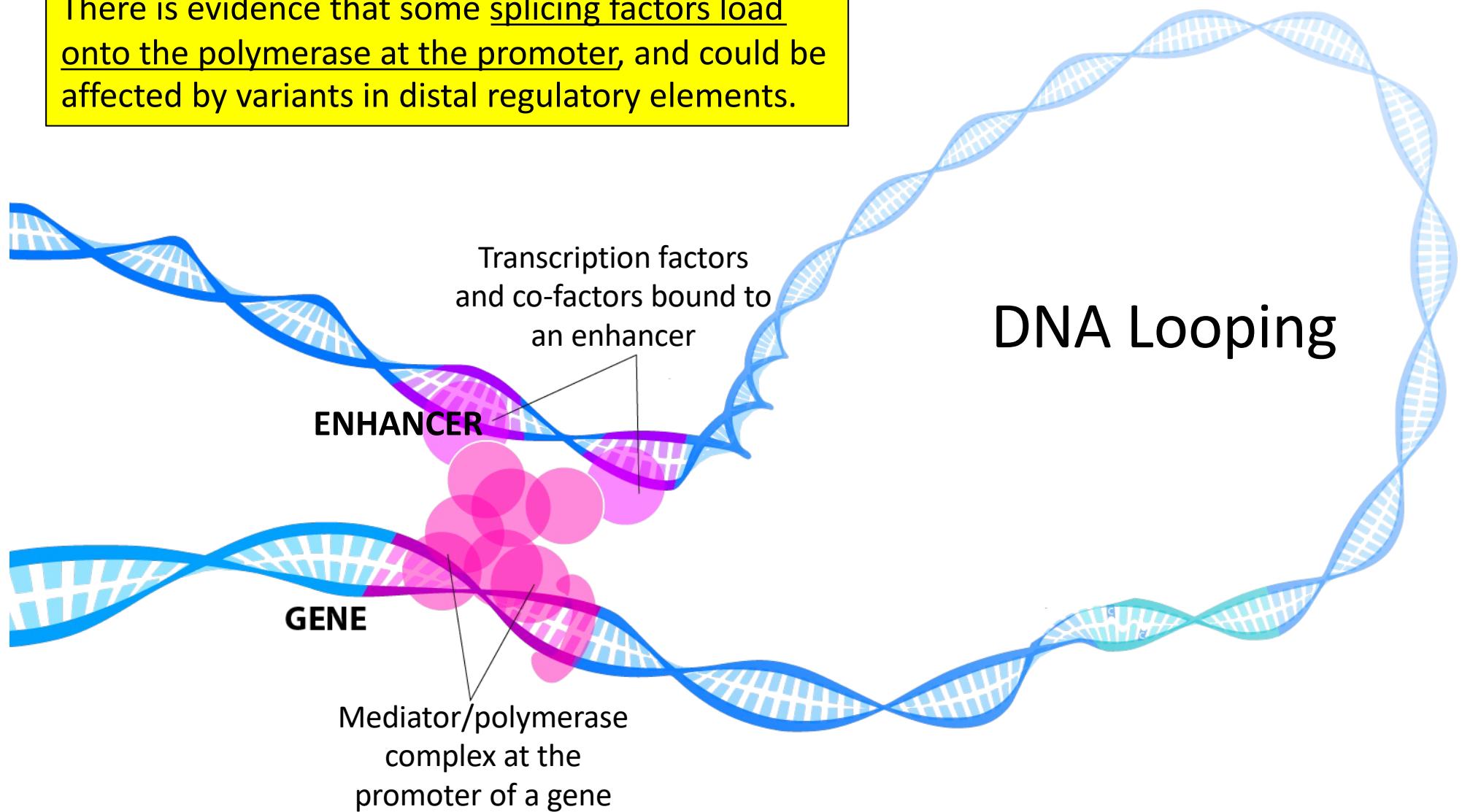
Variants that alter splicing enhancers/silencers can result in repression or activation of splice sites:

See: Kornblihtt *et al.*, 2013 for  
full description and figures

These are the most difficult cases to predict. Current methods are focusing on deep-learning neural networks, but their predictive accuracy isn't perfect, and they aren't always interpretable.

# Distal Variants Might Impact Gene Structure

There is evidence that some splicing factors load onto the polymerase at the promoter, and could be affected by variants in distal regulatory elements.



# Outline

1. Overview: Variants and Gene Structures
2. Impact on Protein-coding Reading Frames
- 3. Impact on Splicing**
4. Other Impacts on RNA

# Secondary Structures

- Single-stranded RNA can form secondary structures
- Variants can increase the probability of a secondary structure forming
- These can stall the ribosome and trigger a decay pathway

For full description and figures, see:  
April 2006 Nature 440(7083):425-6

# Changes to IRES

For full description and figures, see:  
Periodicum Biologorum 114(4):471-478

## A mutation in the *c-myc*-IRES leads to enhanced internal ribosome entry in multiple myeloma: A novel mechanism of oncogene de-regulation

Stephen A Chappell, John PC LeQuesne, Fiona EM Paulin, Matthew L deSchoolmeester, Mark Stoneley, Richard L Soutar, Stuart H Ralston, Miep H Helfrich & Anne E Willis✉

Oncogene 19, 4437–4440 (2000) | Cite this article

2247 Accesses | 113 Citations | 3 Altmetric | Metrics

### Abstract

The 5' untranslated region of the proto-oncogene *c-myc* contains an internal ribosome entry segment (IRES) (Nanbru et al., 1997; Stoneley et al., 1998) and thus *c-myc* protein synthesis can be initiated by a cap-independent as well as a cap-dependent mechanism (Stoneley et al., 2000). In cell lines derived from patients with multiple myeloma (MM) there is aberrant translational regulation of *c-myc* and this correlates with a C-T mutation in the *c-myc*-IRES (Paulin et al., 1996). RNA derived from the mutant IRES displays enhanced binding of protein factors (Paulin et al., 1998). Here we show that the same mutation is present in 42% of bone marrow samples obtained from patients with MM, but was not present in any of 21 controls demonstrating a strong correlation between this mutation and the disease. In a tissue culture based assay, the mutant version of the *c-myc*-IRES was more active in all cell types tested, but showed the greatest activity in a cell line derived from a patient with MM. Our data demonstrate that a single mutation in the *c-myc*-IRES is sufficient to cause enhanced initiation of translation via internal ribosome entry and represents a novel mechanism of oncogenesis.

# Changes to uORFs

## Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans

Sarah E. Calvo<sup>a,b,c,d,1</sup>, David J. Pagliarini<sup>a,b,c,1</sup>, and Vamsi K. Mootha<sup>a,b,c,2</sup>

PNAS

Upstream ORFs (uORFs) are mRNA elements defined by a start codon in the 5' UTR that is out-of-frame with the main coding sequence. Although uORFs are present in approximately half of human and mouse transcripts, no study has investigated their global impact on protein expression. Here, we report that uORFs correlate with significantly reduced protein expression of the downstream ORF, based on analysis of 11,649 matched mRNA and protein measurements from 4 published mammalian studies. Using reporter constructs to test 25 selected uORFs, we estimate that uORFs typically reduce protein expression by 30–80%, with a modest impact on mRNA levels. We additionally identify polymorphisms that alter uORF presence in 509 human genes. Finally, we report that 5 uORF-altering mutations, detected within genes previously linked to human diseases, dramatically silence expression of the downstream protein. Together, our results suggest that uORFs influence the protein expression of thousands of mammalian genes and that variation in these elements can influence human phenotype and disease.

# References

- Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol.* 2013 Mar;14(3):153-65. doi: 10.1038/nrm3525. Epub 2013 Feb 6. Erratum in: *Nat Rev Mol Cell Biol.* 2013 Mar;14(3). doi:10.1038/nrm3560. PMID: 23385723.
- Tollervey D. Molecular biology: RNA lost in translation. *Nature.* 2006 Mar 23;440(7083):425-6. doi: 10.1038/440425a. PMID: 16554791.
- Ozretić, Petar & Bisio, Alessandra & Inga, Alberto & Levanat, Sonja. (2012). The growing relevance of cap-independent translation initiation in cancer-related genes. *Periodicum Biologorum.* 114. 471-478.
- Majoros WH, Campbell MS, Holt C, DeNardo EK, Ware D, Allen AS, Yandell M, Reddy TE. High-throughput interpretation of gene structure changes in human and nonhuman resequencing data, using ACE. *Bioinformatics.* 2017 May 15;33(10):1437-1446. doi: 10.1093/bioinformatics/btw799. PMID: 28011790; PMCID: PMC5860548.
- Rosenberg AB, Patwardhan RP, Shendure J, Seelig G. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell.* 2015 Oct 22;163(3):698-711. doi: 10.1016/j.cell.2015.09.054. Epub 2015 Oct 22. PMID: 26496609.
- Buratti E, Chivers M, Hwang G, Vorechovsky I. DBASS3 and DBASS5: databases of aberrant 3'- and 5'-splice sites. *Nucleic Acids Res.* 2011 Jan;39(Database issue):D86-91. doi: 10.1093/nar/gkq887. Epub 2010 Oct 6. PMID: 20929868; PMCID: PMC3013770.

# Questions?