Potential Consequences on the RNA Level and using prediction tools

Variant Effect Prediction Training Course

THE HUMAN VARIOME PROJECT
an NGO official partner of UNESCO
sharing data · reducing disease

Variant Effect Prediction Training Course

29 - 31 May 2019
Moscow, Russian Fed.

Andreas Laner

MGZ München

laner@mgz-muenchen.de

Potential Consequences on the RNA Level and using prediction tools

A. Variants altering the structure/ integrity: pre-mRNA splicing

B. Variants altering the stability/ turnover: mRNA (UTRs, 3D, miRNA binding)

C. Variants altering the translation dynamics: mRNA (codon usage, +/- ribosomal PS)

D. Prediction Tools

E. Functional RNA studies

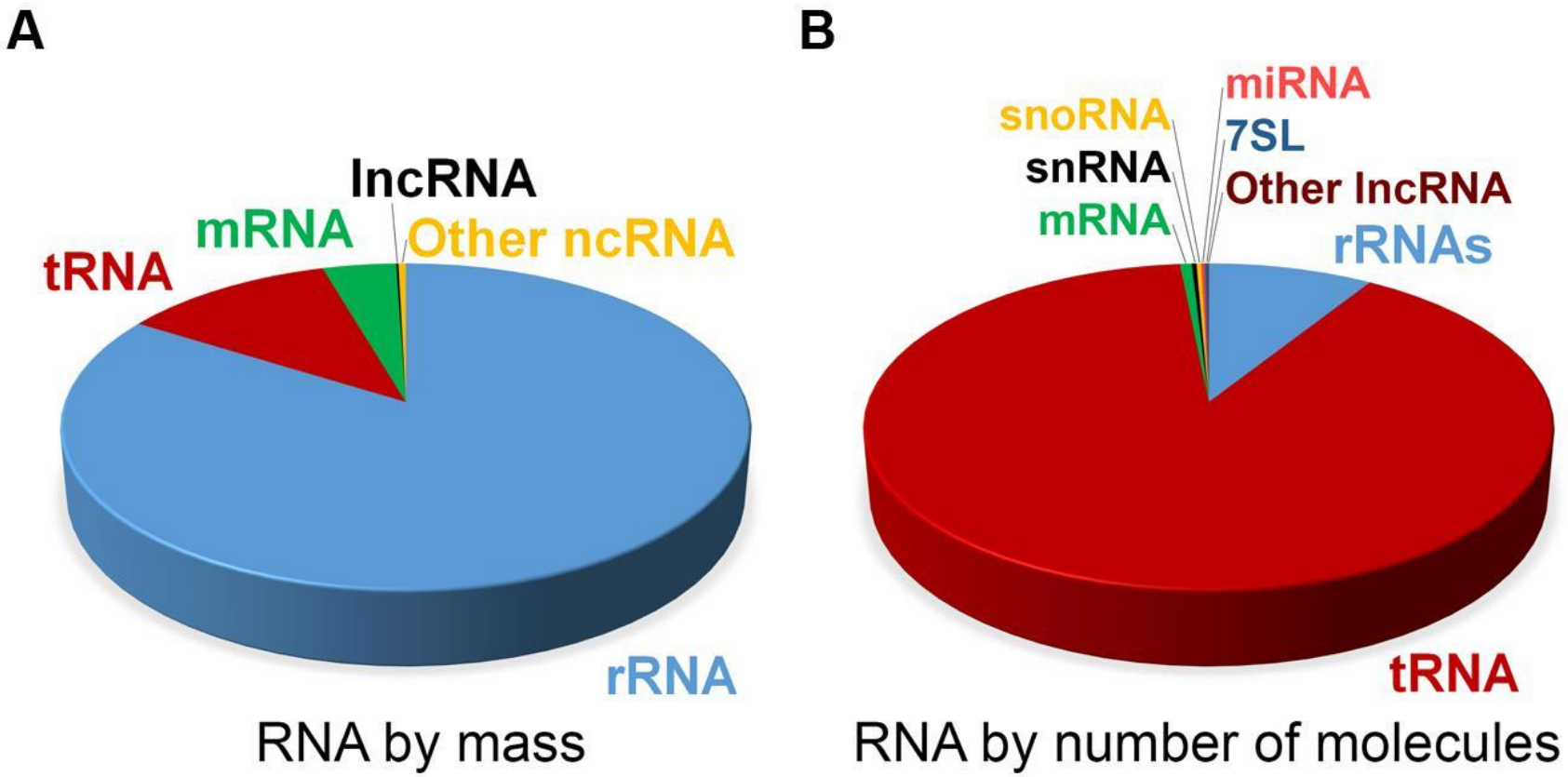
Speaking about RNA

ENCODE project (Nature 489, 57-74. 2012):

- Most of the genome (>85%) is transcribed

- 60.000 „genes“:
  - ~ 20.000 protein coding genes (>90% multiple isoforms)
  - ~ 16.000 long non-coding (lnc) RNAs
  - ~ 10.000 small non-coding (snc) RNAs
  - ~ 14.000 pseudogenes

Speaking about RNA



Palazzo et al.; Front. Genet., 26 January 2015

ncRNA, tRNA, rRNA, snRNA, siRNA, hnRNA, scRNA, RNA editing, Ribosome, RNP, mRNA surveillance/ decay

Potential Consequences on the RNA Level and using prediction tools

A. Variants altering the structure/ integrity: pre-mRNA splicing

B. Variants altering the stability/ turnover: mRNA (UTRs, 3D, miRNA binding)

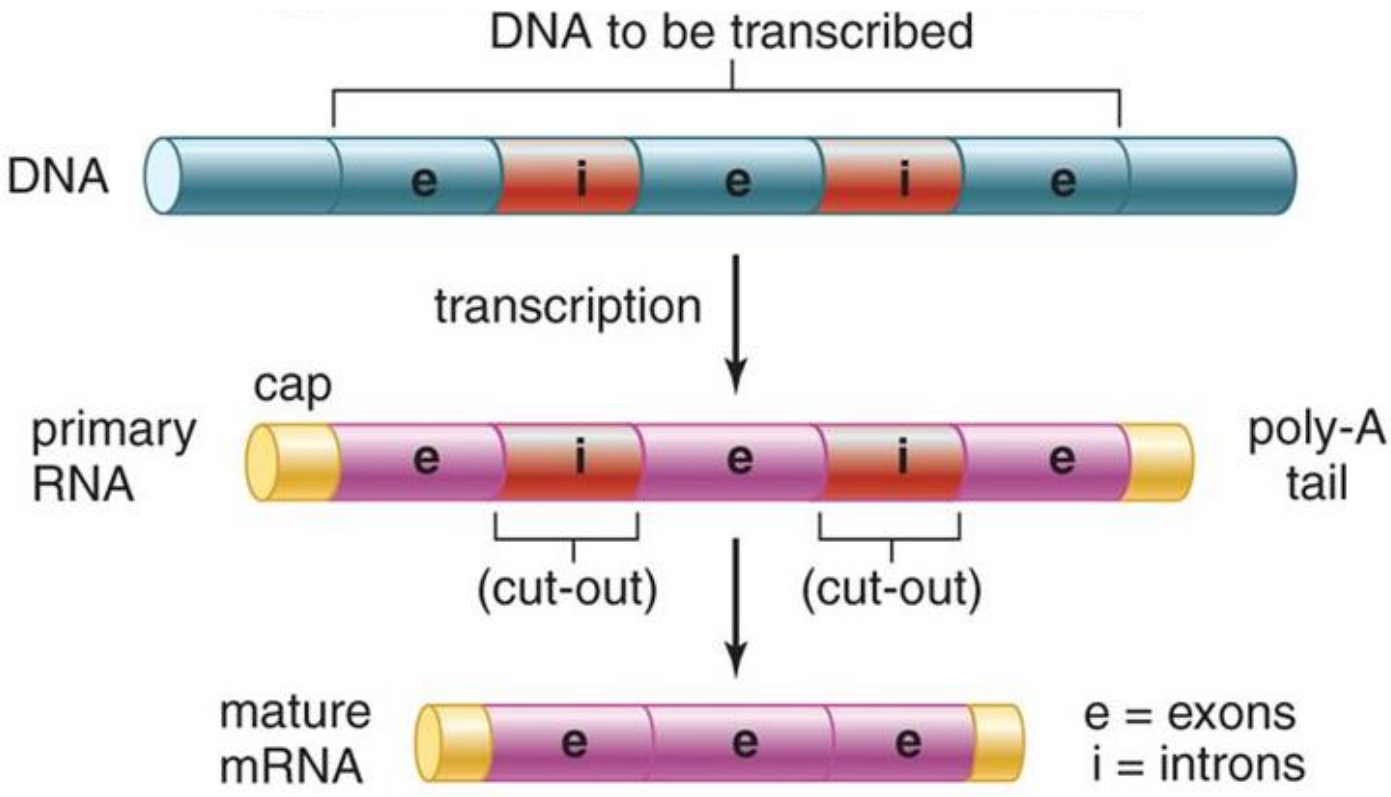
C. Variants altering the translation dynamics: mRNA (codon usage, +/- ribosomal PS)

D. Prediction Tools

E. Functional RNA studies

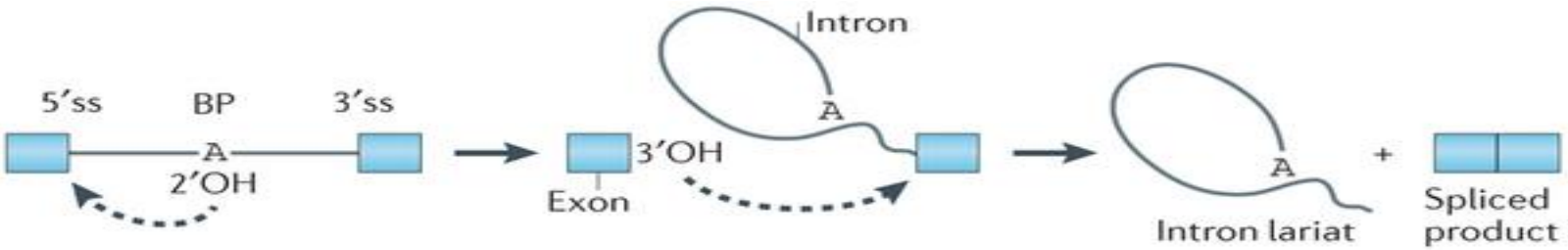
A. Variants altering the mRNA structure / integrity

pre-mRNA splicing



A. Variants altering the mRNA structure / integrity

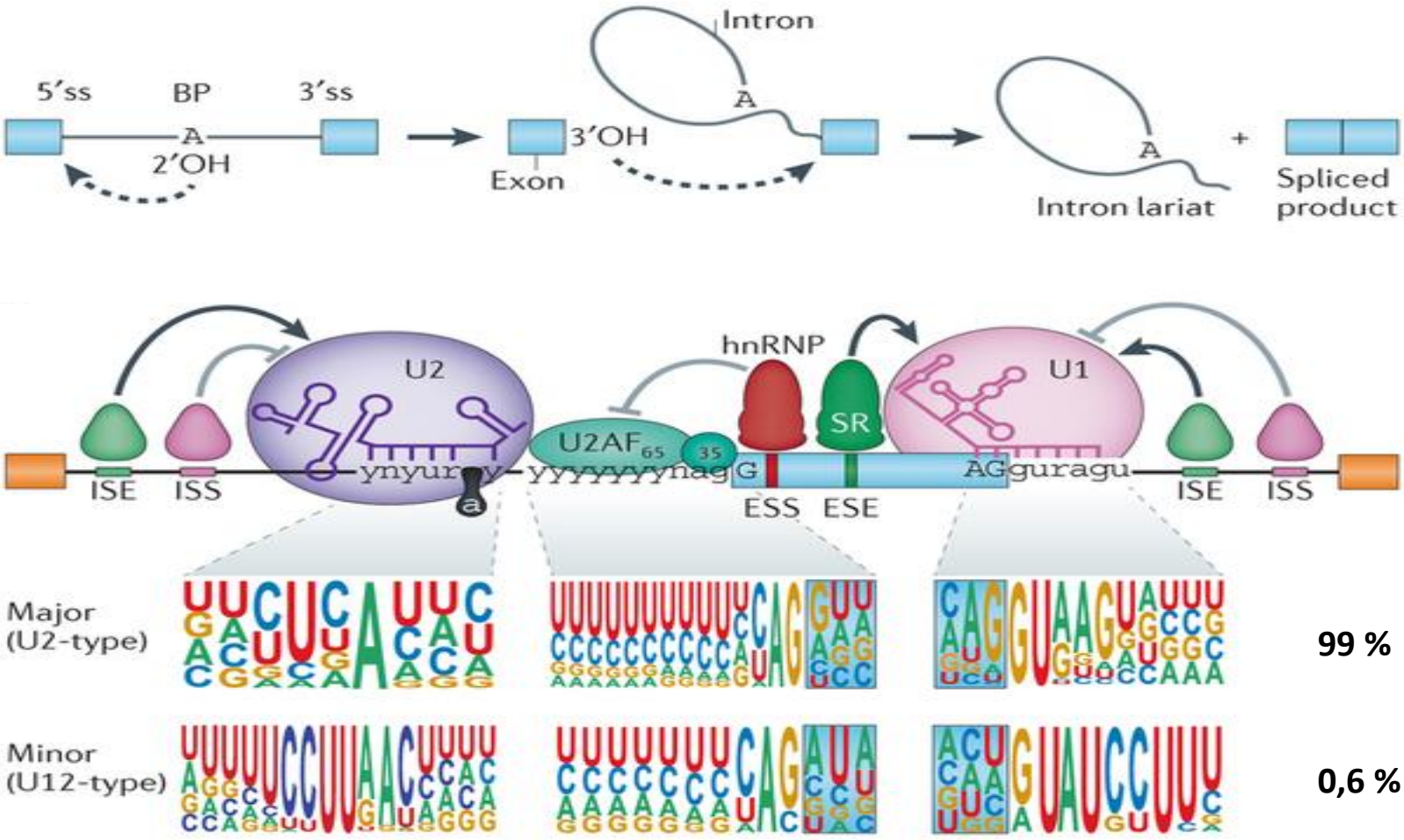
pre-mRNA splicing



Scotti & Swanson. Nature Reviews Genetics 17, 19–32 (2016))

A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

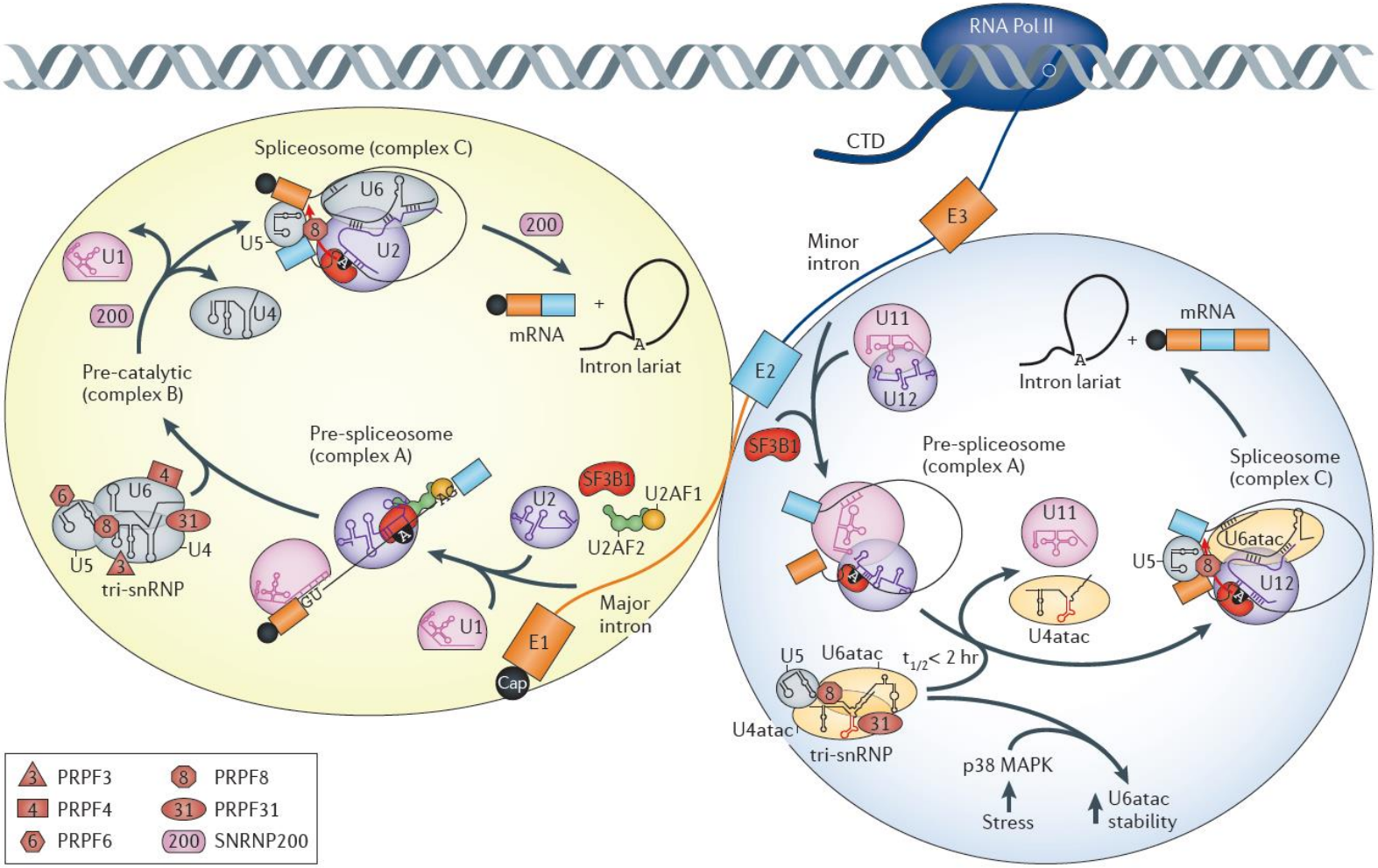


Nature Reviews | Genetics

Scotti & Swanson. Nature Reviews Genetics 17, 19–32 (2016))

A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

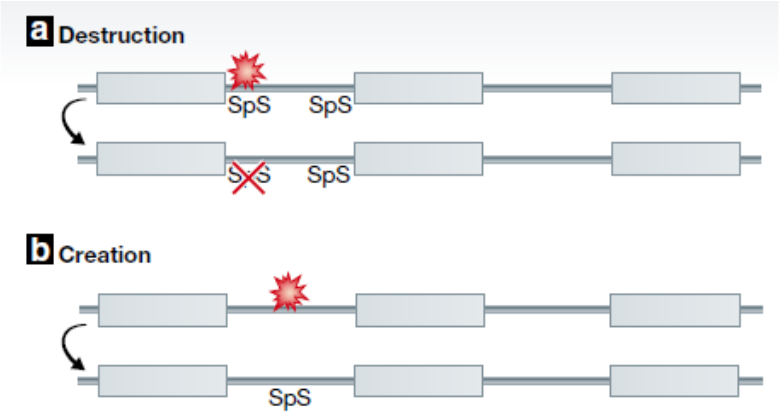


Scotti & Swanson. Nature Reviews Genetics 17, 19–32 (2016))

A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

Variants affecting splicing

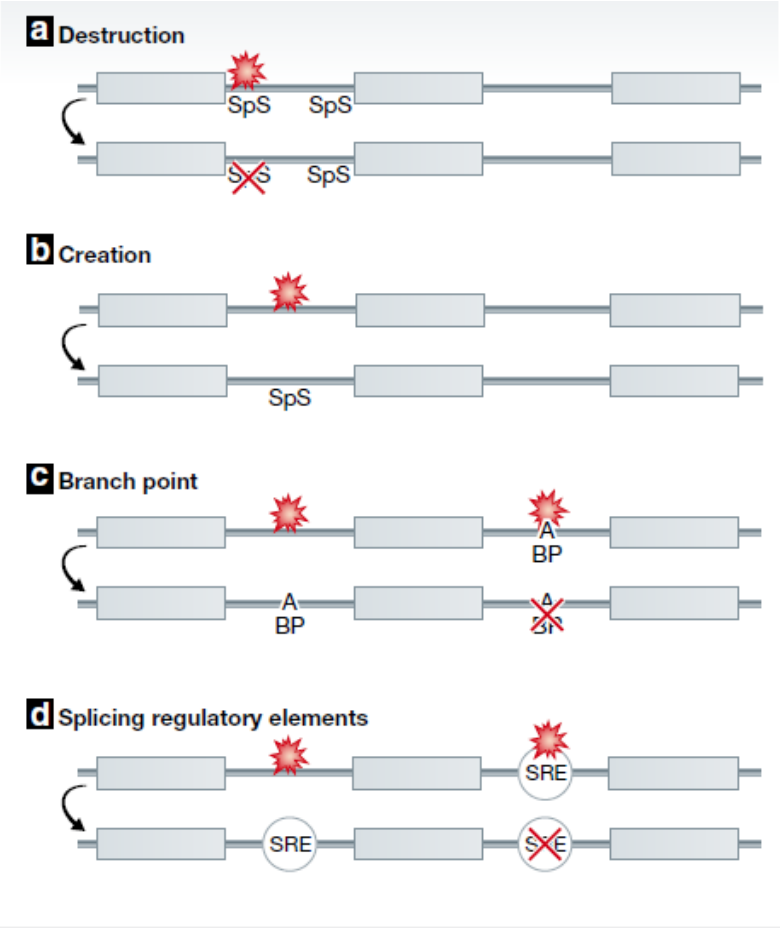


Diederichs et al.; 2016. EMBO Molecular Medicine 8(5):442-457

A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

Variants affecting splicing

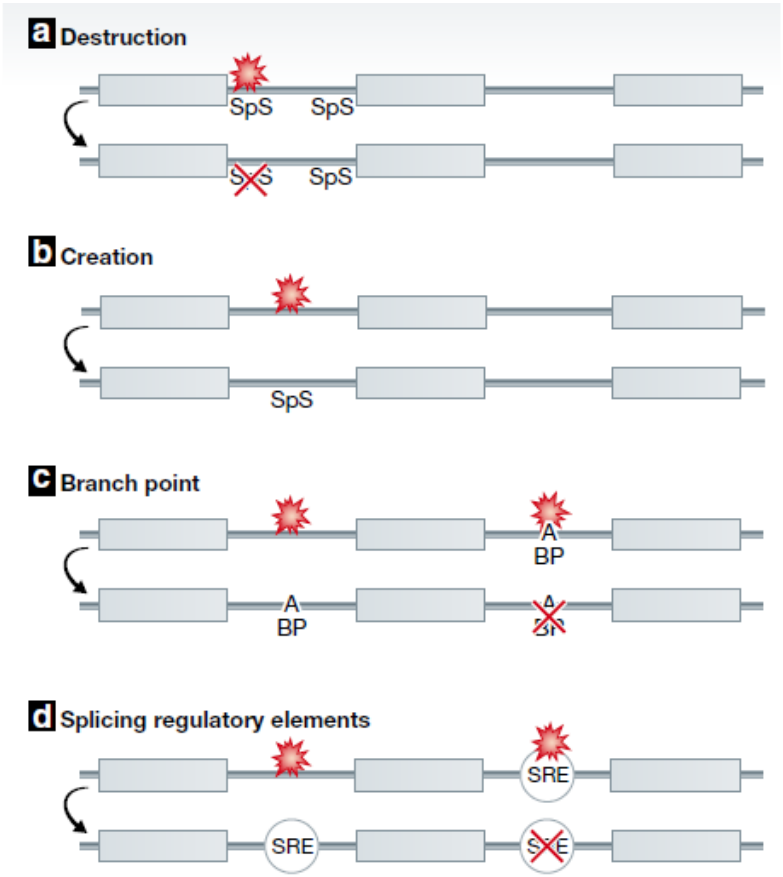


Diederichs et al.; 2016. EMBO Molecular Medicine 8(5):442-457

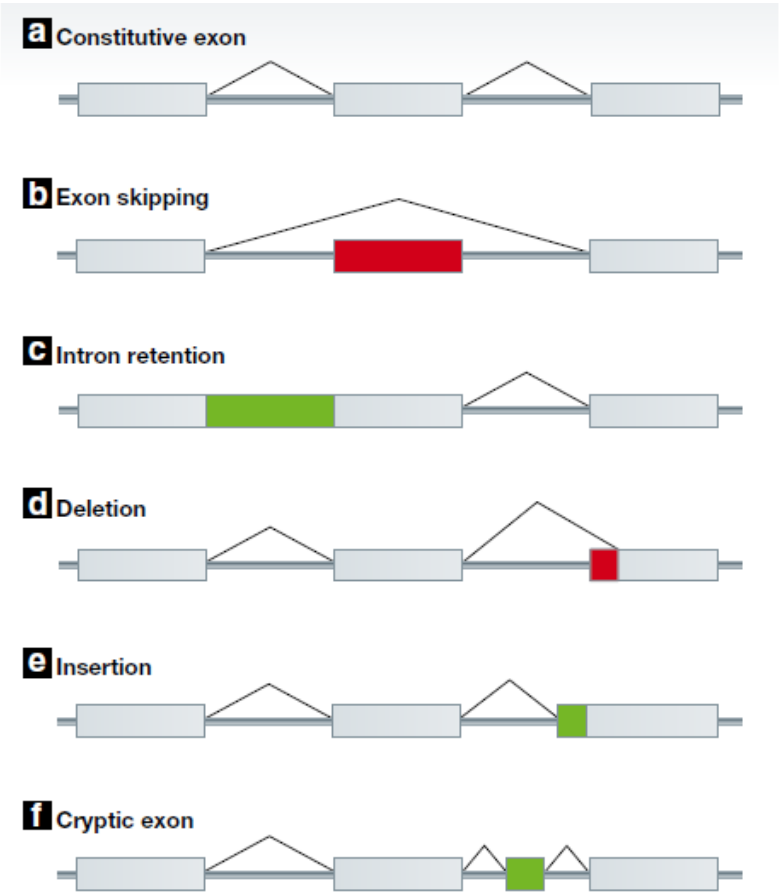
A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

Variants affecting splicing



Effect on RNA

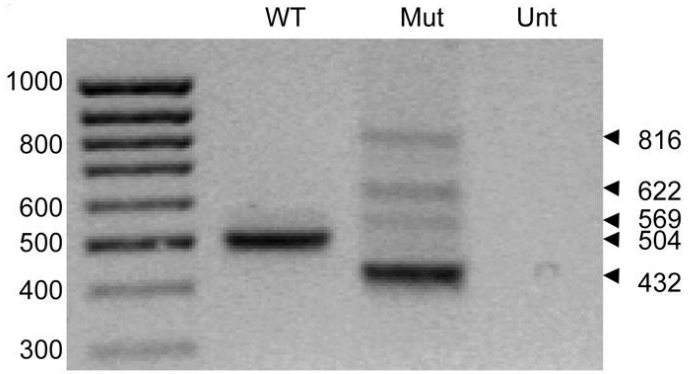


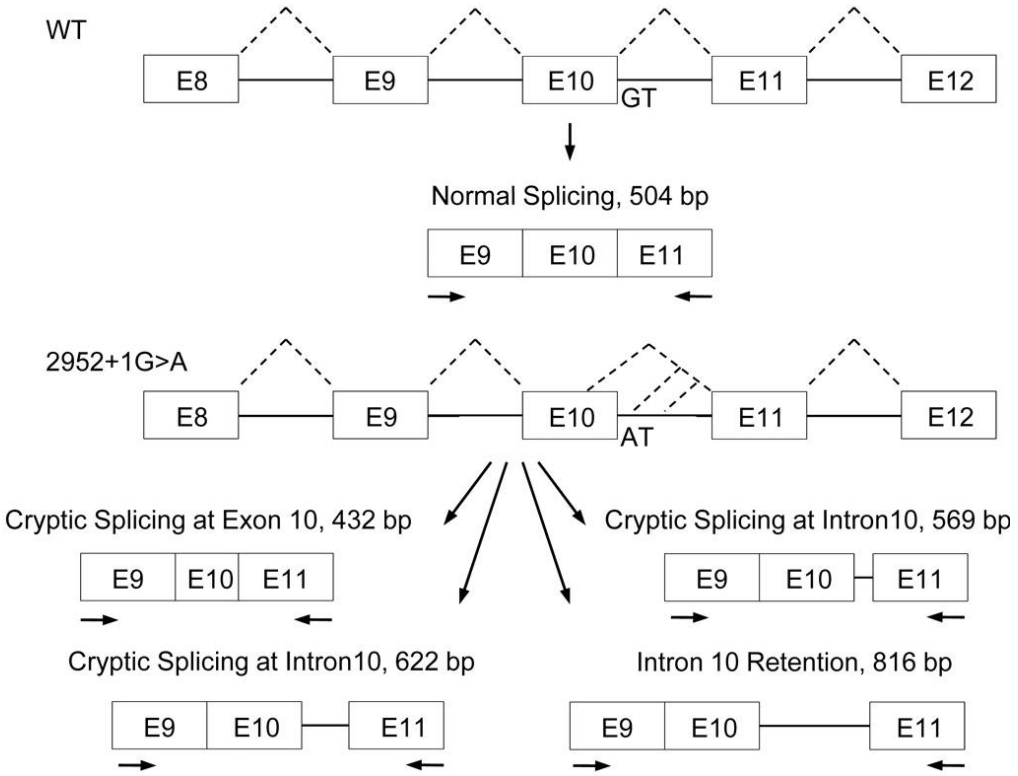
Diederichs et al.; 2016. EMBO Molecular Medicine 8(5):442-457

A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

Multiple splicing defects caused by hERG splice site mutation 2592+1G>A associated with long QT syndrome





Stump et al.; American Journal of Physiology - January 2011 Vol. 300 no. 1

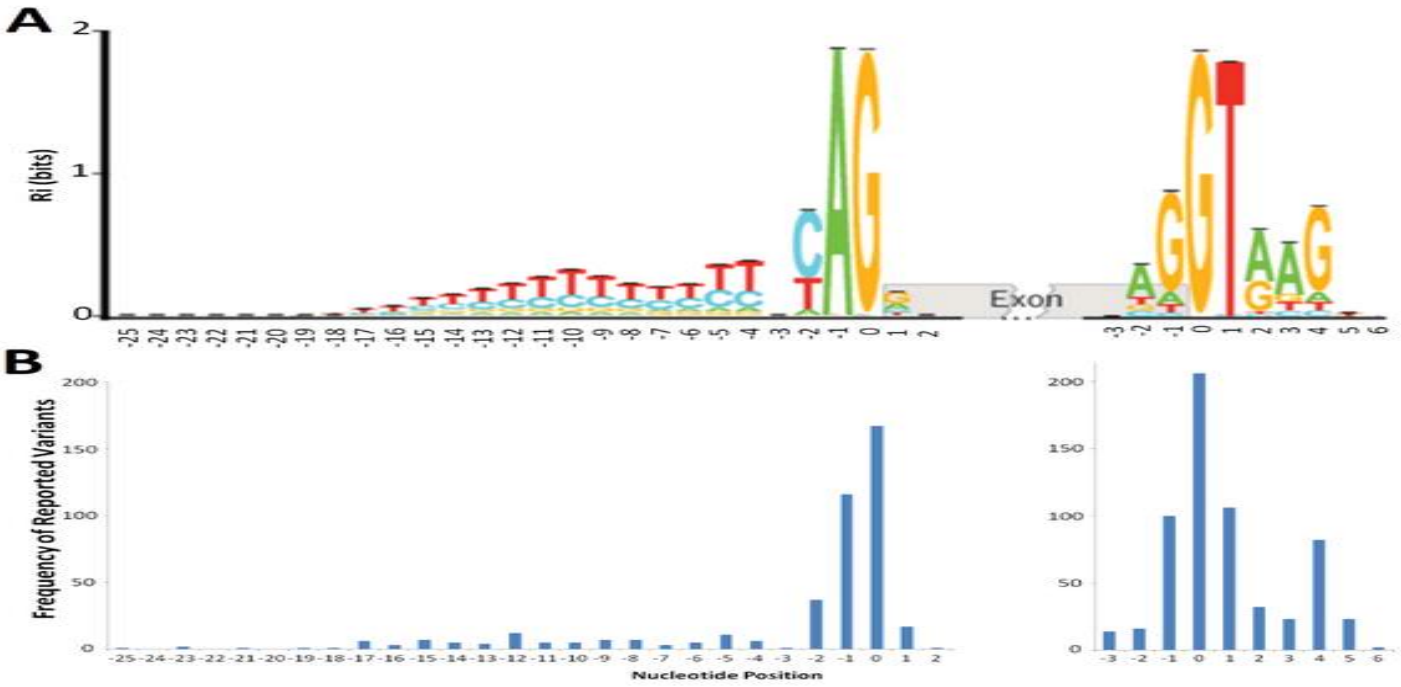
A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

Pathogenic variants that affect pre-mRNA splicing account for at least 15%-50% of disease-causing mutations (Wang, G.-S. & Cooper, T. A. *Splicing in disease: disruption of the splicing code and the decoding machinery*. *Nature Rev. Genet.* **8**, 749–761 (2007)).

With up to 50% of all pathogenic mutations described in some genes (NF1, ATM)
(Teraoka et al.; *Am J Hum Genet.* 1999; **64**(6): 1617–1631 / Ars et al.; *Hum Mol Genet.* 2000; **9**(2): 237–247.)

Most variants affect the canonical (+/- 1, 2) splice sites

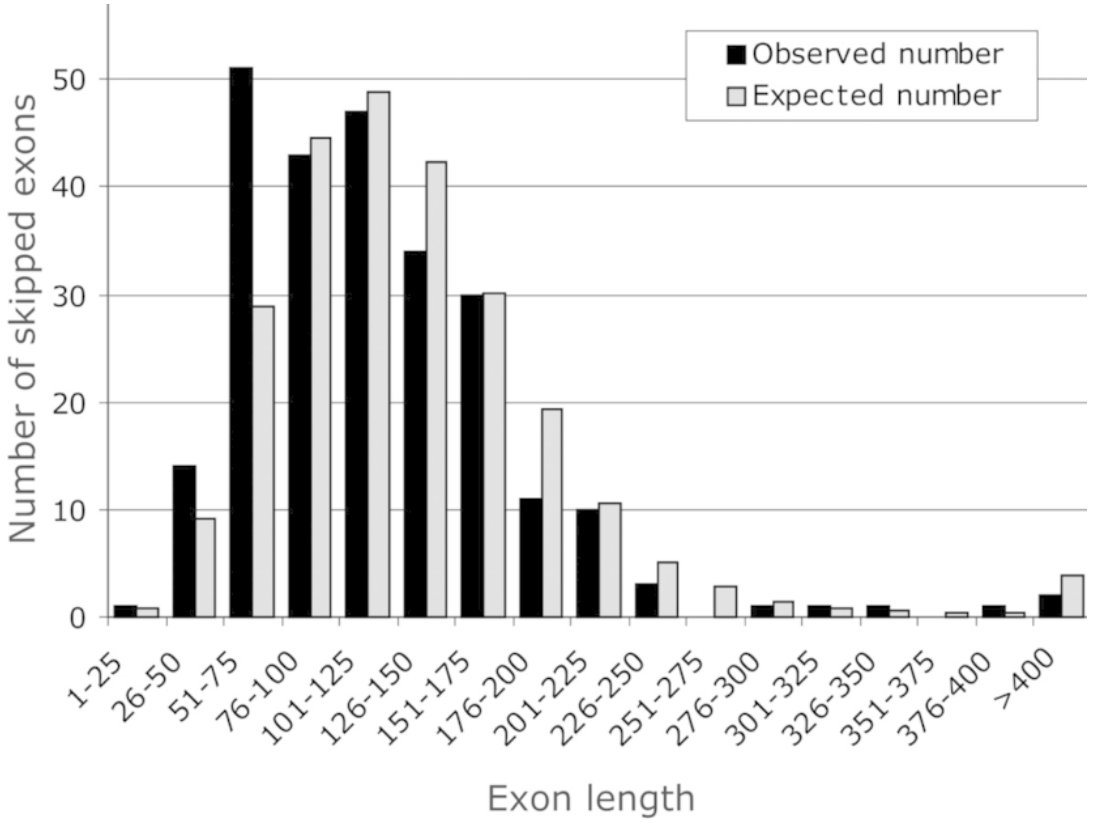


Caminsky et al.; *F1000Research* 2015

A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

Exons that were skipped as a result of splicing variants are shorter than average exons



Kralovicova et al.; Nucleic Acids Research 35(19):6399-413 · February 2007

A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

ACMG-AMP Classification Rules:

Criteria for classifying pathogenic variants (Tabelle 1)

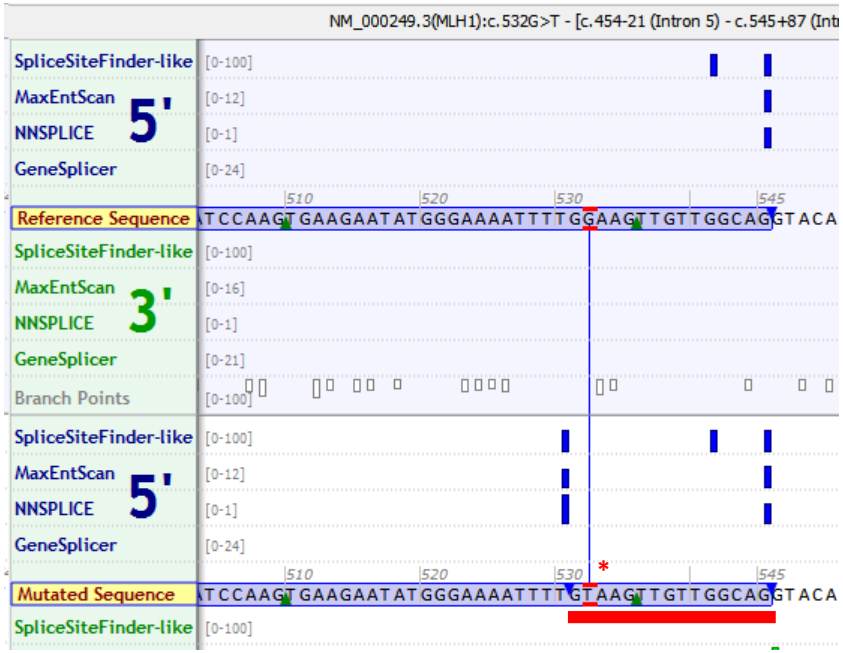
Evidence of pathogenicity		Category
Very strong	PVS1	Null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease. *Caveats:* • *Beware of genes where LOF is not a known disease mechanism (e.g., GFAP, MYH7)* • *Use caution interpreting LOF variants at the extreme 3' end of a gene* • *Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact* • *Use caution in the presence of multiple transcripts*

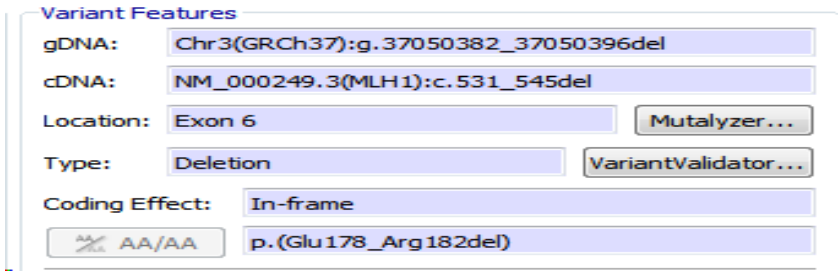
A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

***MLH1* c.532G>T (p.Glu178*)**

Creation of AG/ near the physiological SA site





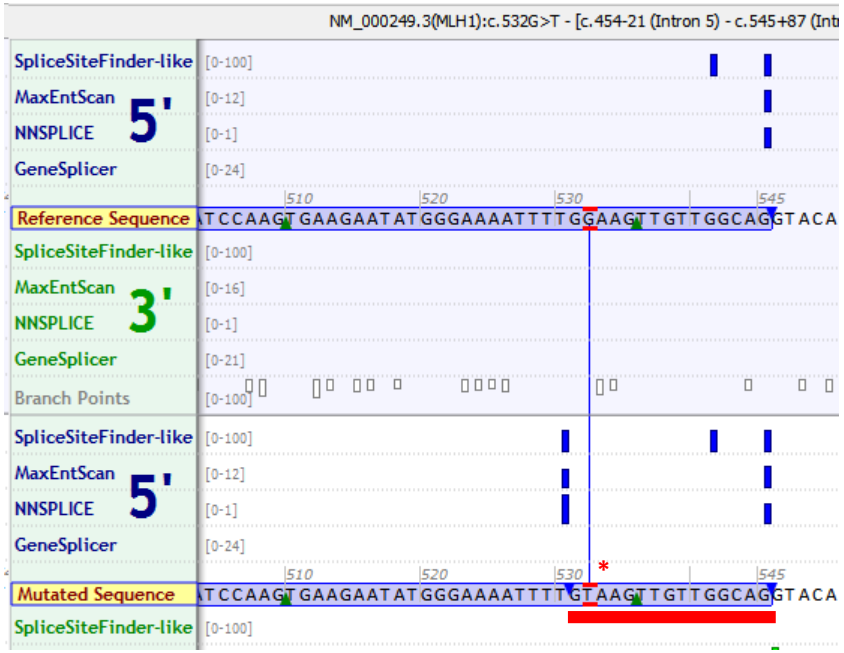
Blavier and Martins, unpublished

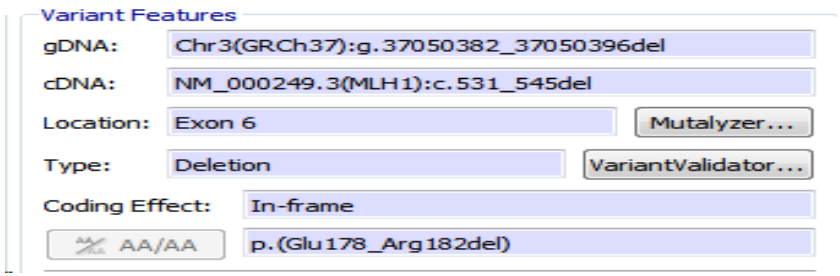
A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

***MLH1* c.532G>T (p.Glu178*)**

Creation of AG/ near the physiological SA site

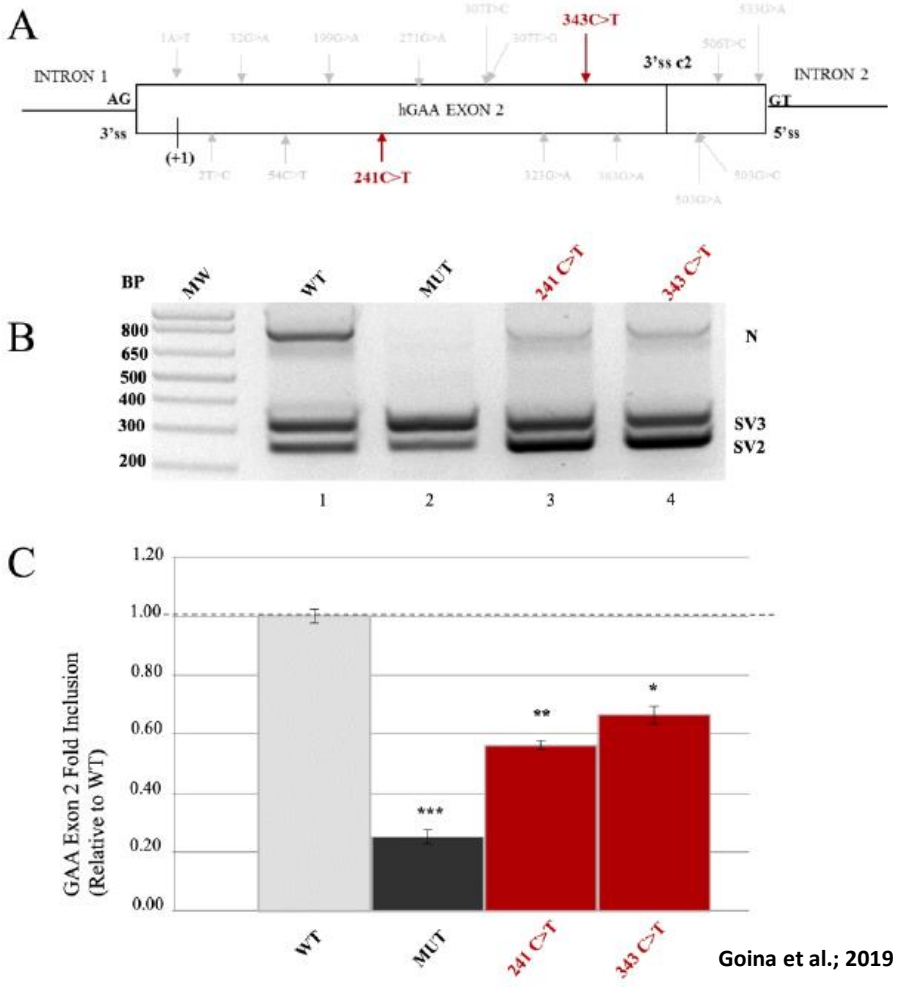




Blavier and Martins, unpublished

***GAA* c.241C>T (p.Gln81*) and c.343C>T (p.Gln115*)**

pTB minigene: 50% of the expression levels with respect



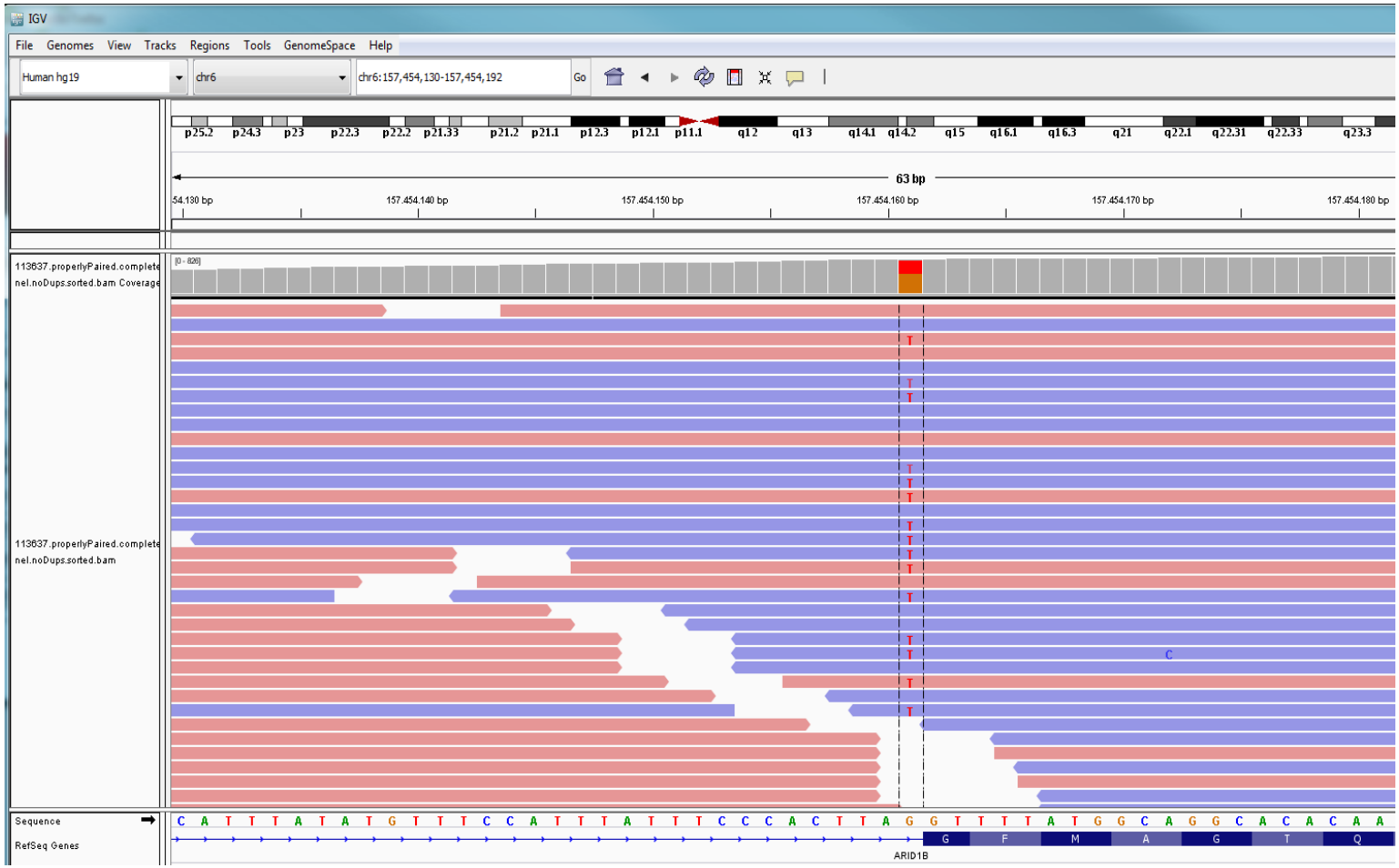
Goina et al.; 2019

A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

ARID1B c.2372-1G>T

Variant found in a patient without Coffin-Siris phenotype (IF)

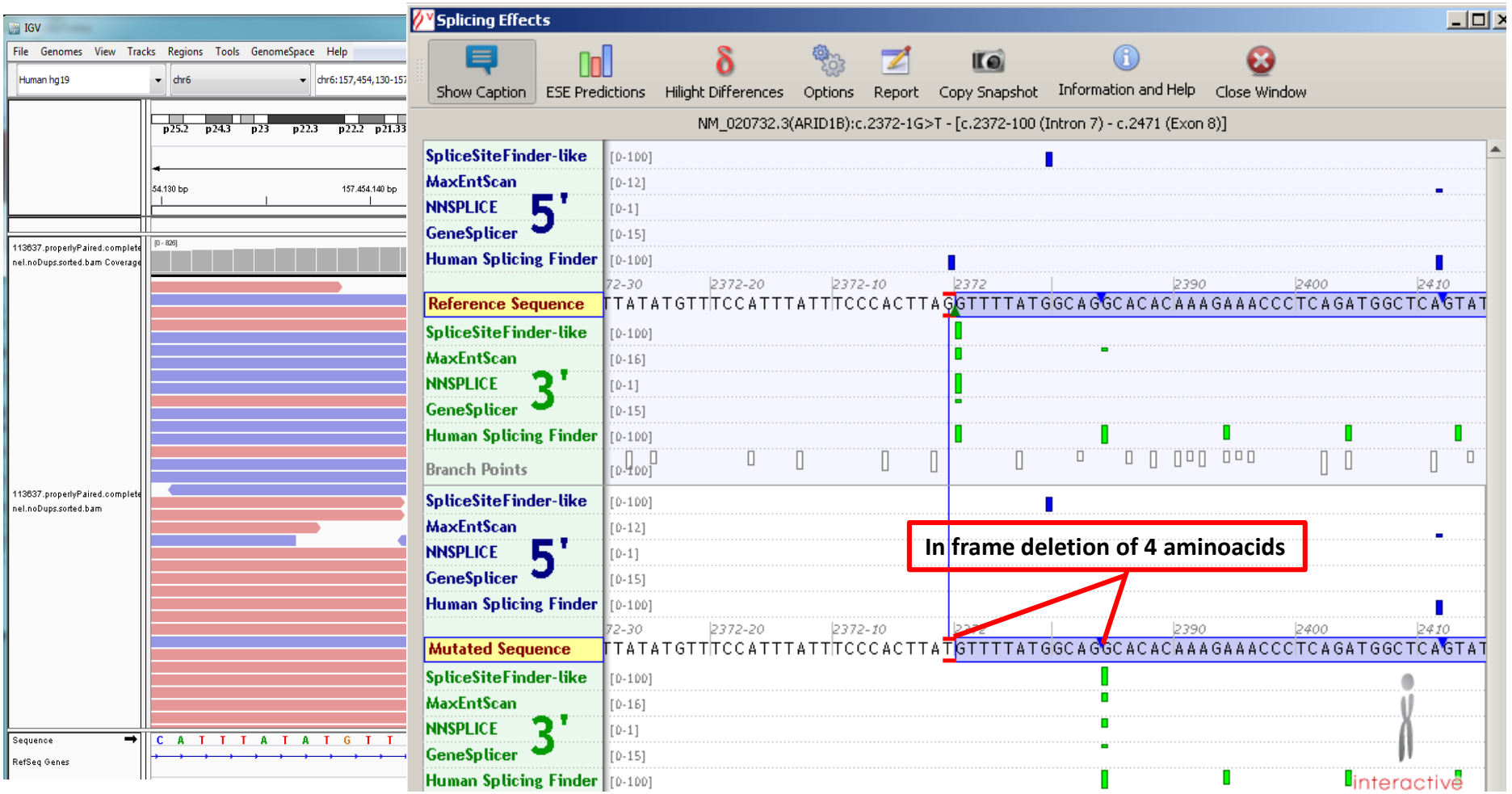


A. Variants altering the mRNA structure / integrity

pre-mRNA splicing

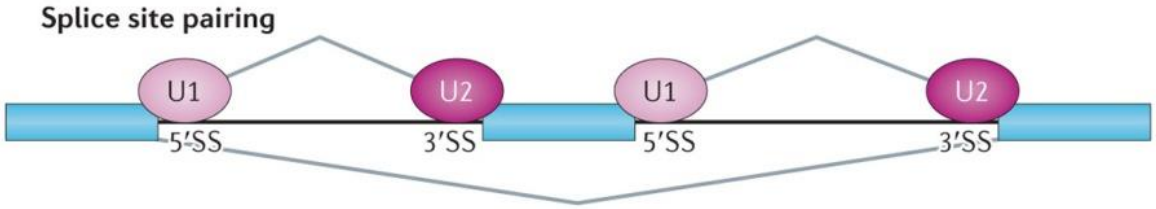
ARID1B c.2372-1G>T

Variant found in a patient without Coffin-Siris phenotype (IF)



A. Variants altering the mRNA structure / integrity

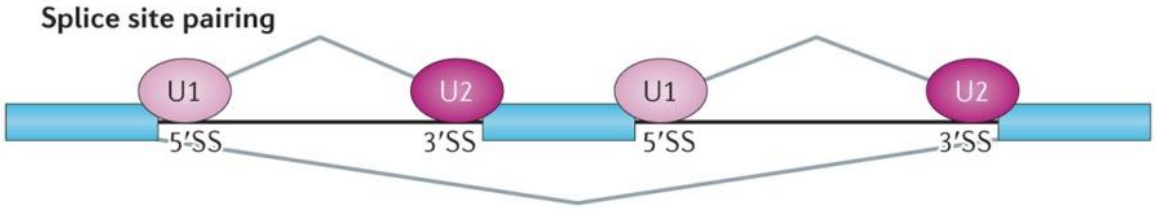
splicing regulatory elements: ESE ESS ISE ISS



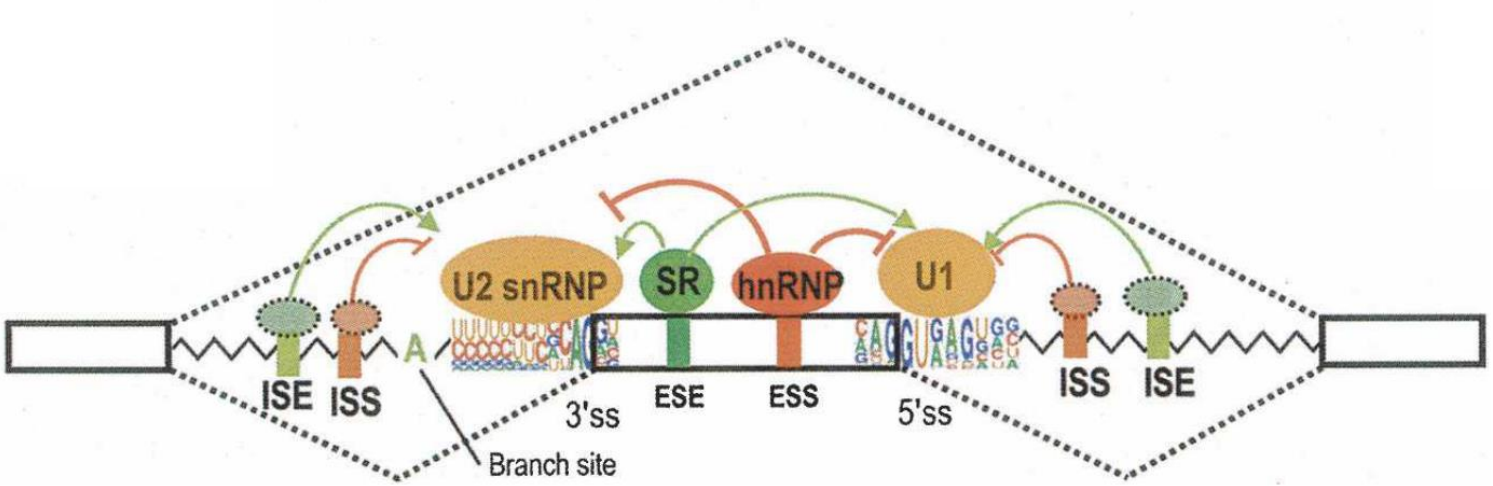
~ 20.000 protein coding genes (>90% multiple isoforms)
SRE are a major regulator of tissue specific alternative splicing

A. Variants altering the mRNA structure / integrity

splicing regulatory elements: ESE ESS ISE ISS



~ 20.000 protein coding genes (>90% multiple isoforms)
SRE are a major regulator of tissue specific alternative splicing



Wang, Z. & Burge, C. B. *Splicing regulation: from a parts list of regulatory elements to an integrated splicing code*. *RNA* **14**, 802–813 (2008)

A. Variants altering the mRNA structure / integrity

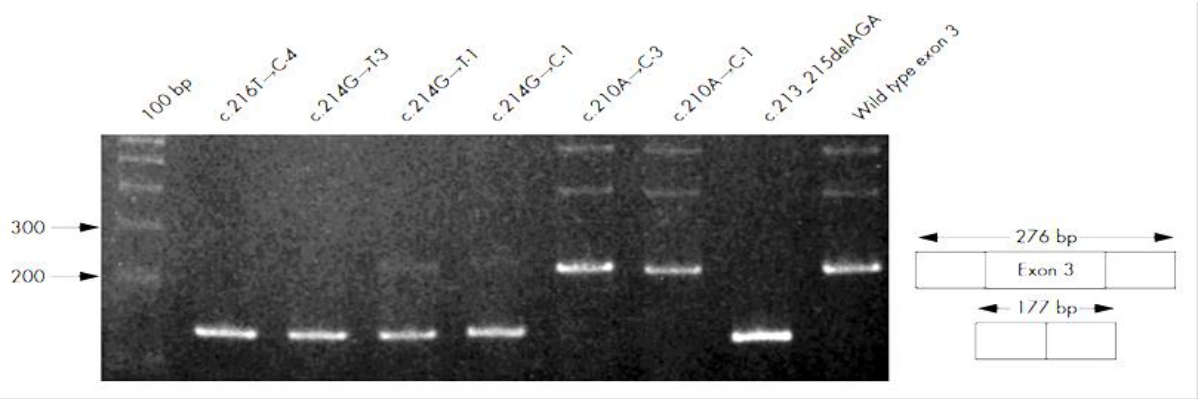
splicing regulatory elements: ESE ESS ISE ISS

LETTER TO JMG

Disruption of an exon splicing enhancer in exon 3 of *MLH1* is the cause of HNPCC in a Quebec family

S McVety, L Li, P H Gordon, G Chong, W D Foulkes

J Med Genet 2006;**43**:153–156. doi: 10.1136/jmg.2005.031997



- 3 bp deletion and several point mutations (silent, missense, nonsense) disrupt an exon splicing enhancer in exon 3 of MLH1 and is the cause of HNPCC in a Quebec family.
- All the mutations cause varying degrees of exon skipping, suggesting the presence of an ESE at the 5' end of exon 3.
- These mutations are situated in a GAAGAT sequence 3 bp downstream from the start of exon 3.

A. Variants altering the mRNA structure / integrity

splicing regulatory elements: ESE ESS ISE ISS

Gen	Variant	Disease	Effect	Ref.
GH 1	c.176A>G	Familial isolated GH deficiency type II	Disruption of ESE	Moseley et al.; J Clin Endocrinol Metab. 2002
CFTR	c.1966G>T	Cystic Fibrosis	Disruption of ESE	Aznarez et al.; Hum. Mol. Genet. 2003
ATP6AP2	c.345C>T	X-linked parkinsonism with spasticity (XPDS)	Novel ESS creation	Korvatska et al. Hum. Mol. Genet. 2013
DMD	c.4250T>A	Becker muscular dystrophy (BMD)	Novel ESS creation	Disset et al.; Hum. Mol. Genet. 2006
MAPT	c.892A>G	Frontotemporal dementia (FTDP-17)	Disruption of ESS	Iovino et al.; Acta Neuropathol. 2014
CYBB	c.389G>T	Chronic granulomatous disease (CGD)	Disruption of ESS	De Boer et al.; Blood Cells Mol Dis. 2017
OCRL	c.741G>T	Lowe syndrome / Dent-2 disease	Imbalance of ESE / ESS	Suarez-Artiles et al.; Genes. 2018
ACAT1	c.949G>A	Beta-ketothiolase deficiency (T2)	Disruption of ESS	Otsuka et al.; Mol Med Rep. 2016
ETFDH	c.158A>G	Multiple Acyl-CoA dehydrogen. deficiency	Imbalance of ESE / ESS	Olsen et al.; Hum Mutat. 2014

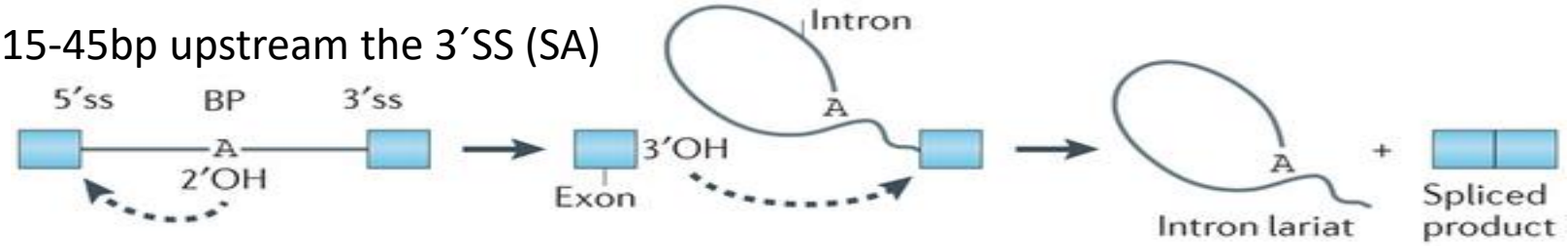
A. Variants altering the mRNA structure / integrity

pre-mRNA splicing / Branchpoint

A. Variants altering the mRNA structure / integrity

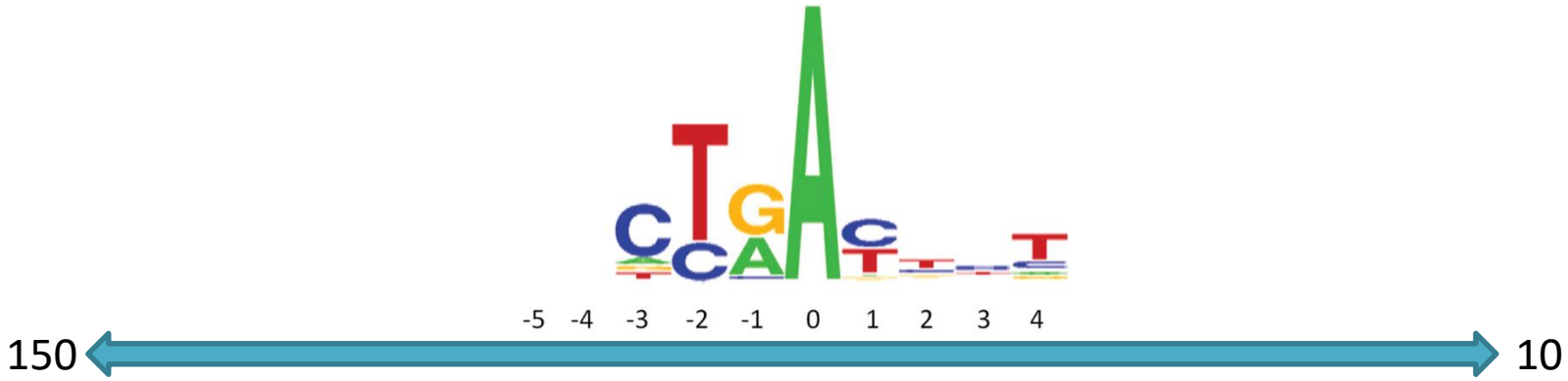
pre-mRNA splicing / Branchpoint

15-45bp upstream the 3´SS (SA)



A. Variants altering the mRNA structure / integrity

pre-mRNA splicing / Branchpoint



Pathogenic BP-Mutations are very rarely described (<20)

Possible explanations for the rarity of BPS mutations:

- Compensatory, alternative BPS sequences can be recognized and used
- The weak constraint on the precision of the distance between the BPS and the 3´ (acceptor) splice site further enables activation of these alternative sites.
- Bias due to technical limitations (primer location, NGS capture or bioinformatic filter criteria)

A. Variants altering the mRNA structure / integrity

pre-mRNA splicing / Branchpoint

Gene	Variant	Disease	Effect	Ref.
FBN2	c.3974-26T>G	Congenital contractural arachnodactyly	Skipping of Ex31 in 25% of transcripts	Maslen et al.; Am J Hum Genet 1997
COL5A1	c.2701-25T>G	Ehlers-Danlos syndrome (EDS)	45 bp of exon 33 are "skipped" in 60% of transcripts	Burrows et al.; Am J Hum Genet 1998
LCAT	c.524-22T>C	Fish-eye disease	Complete loss of function due to intron retention	Li et al.; Biochim Biophys Acta 1998
NPC1	c.882-28A>G	Niemann-Pick disease (NPC)	Shorter transcript lacking exon 7	Di Leo et al.; Hum Mutat 2004
KCNH2	c.2399-28A>G	Long QT (LQT)	Incorrect identification of the acceptor site of intron 9	Crotti et al.; Heart Rhythm 2009
UROS	c.661-31T>G	Congenital erythropoietic porphyria	100% intron retention without exon skipping (last exon)	Bishop et al.; Blood 2010
PTS	A>T substitution 9 nt upstream of its 3' splice site in a LINE-2 sequence	Tetrahydrobiopterin deficiency	Pseudoexon activation in a LINE-2 sequence	Meili et al.; Hum Mutat. 2009
NF2	c.516+232G>A	Neurofibromatosis 2	Creates a functional de novo BP sequence in intron 5	De Klein et al.; Hum Mol Genet 1998
ITGB4	c.1762-25T>A	Pyloric atresia-junctional epidermolysis bullosa	Resulted in two abnormal transcripts each with a PTC	Masunaga et al.; J Dermatol Sci 2015

Potential Consequences on the RNA Level and using prediction tools

A. Variants altering the structure/ integrity: pre-mRNA splicing

B. Variants altering the stability/ turnover: mRNA (UTRs, 3D, miRNA binding)

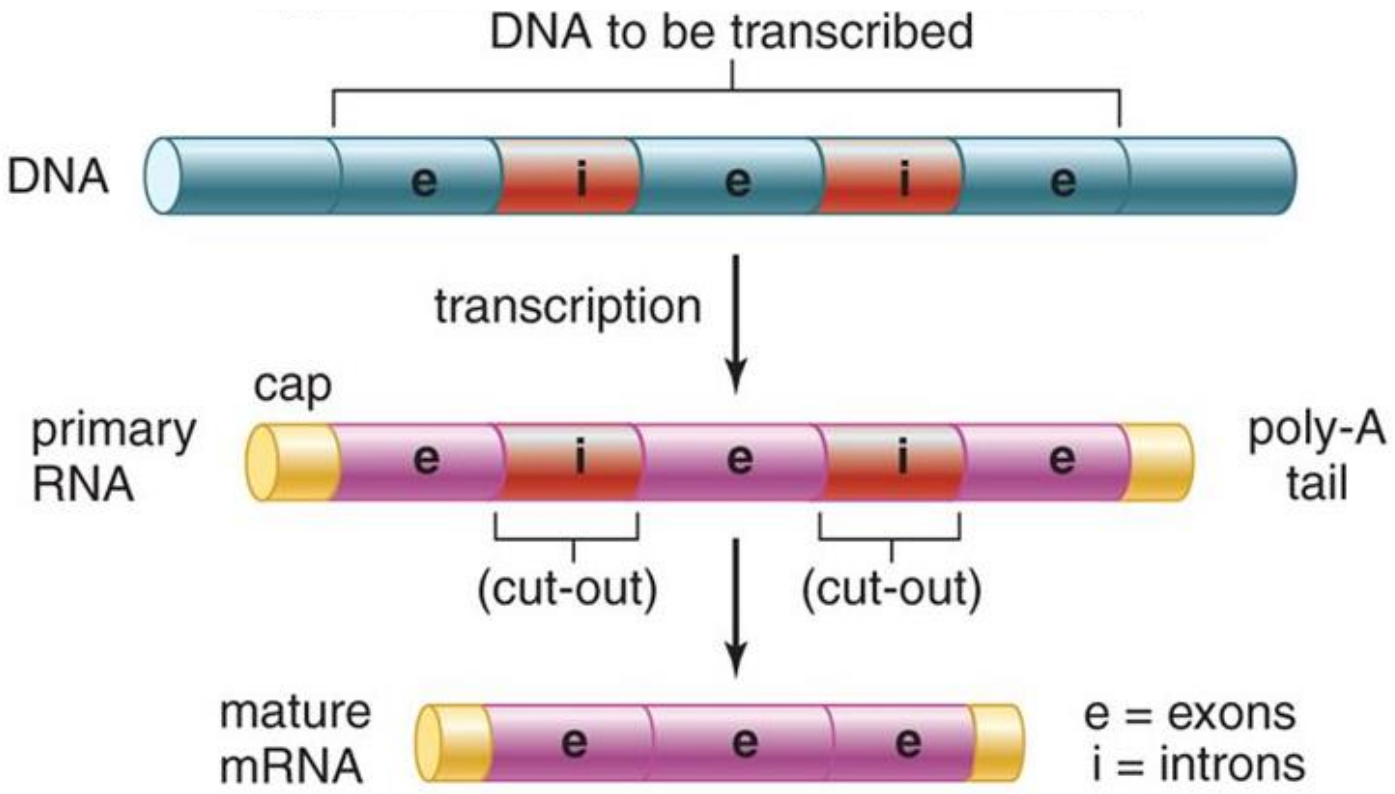
C. Variants altering the translation dynamics: mRNA (codon usage, +/- ribosomal PS)

D. Prediction Tools

E. Functional RNA studies

B. Variants altering the stability / turnover

mRNA



B. Variants altering the stability / turnover

mRNA / Possible mechanisms

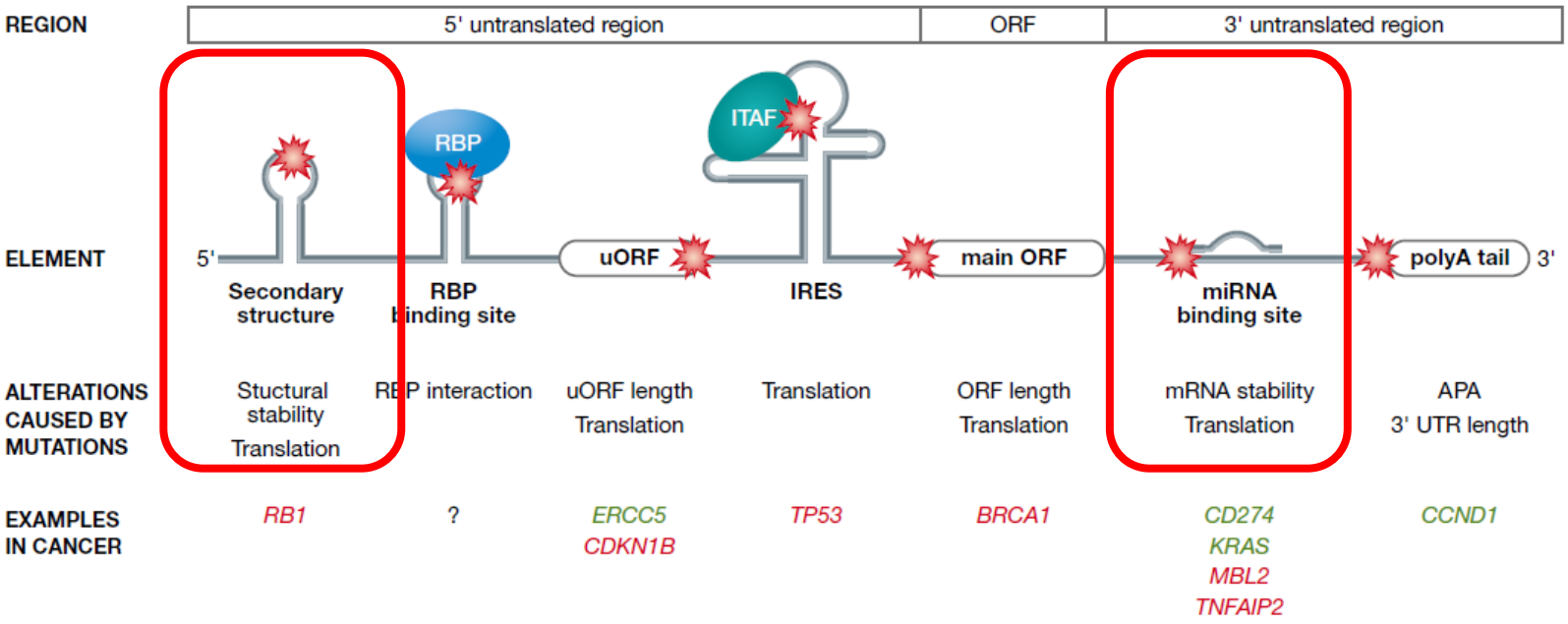
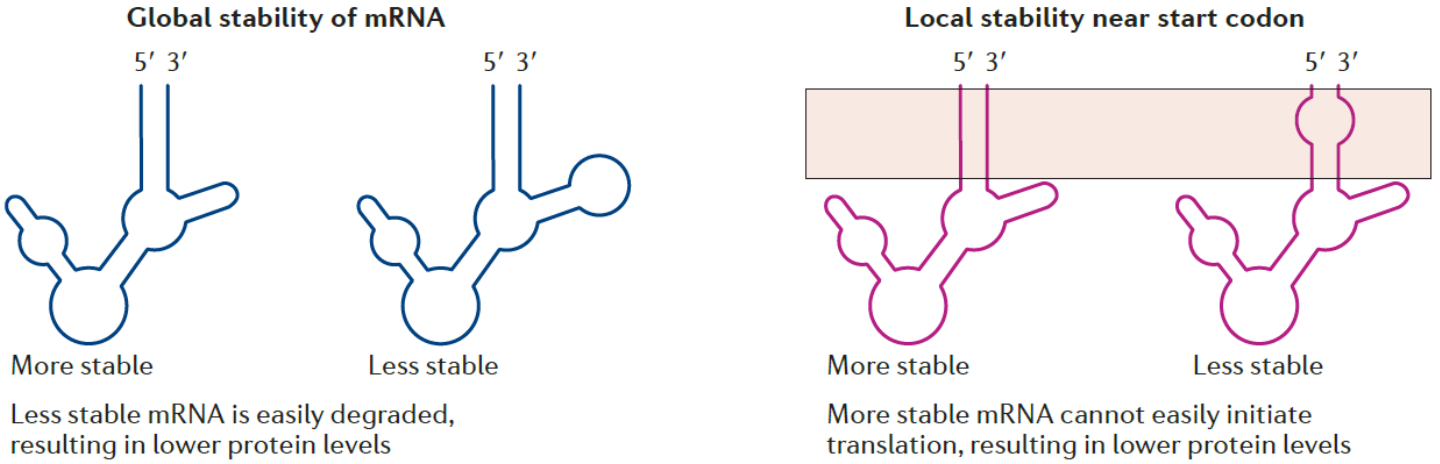


Figure 1. Schematic depiction of mutations within the 5′- and 3′-UTR.
Mutations can alter the secondary structure of the 5′- or 3′-UTR or occur in RNA binding protein (RBP) binding sites, upstream ORFs (uORF), internal ribosome entry sites (IRES; ITAF: IRES *trans*-acting factor), start codons of open reading frames (ORF), microRNA binding sites, or polyadenylation signals (polyA). These alterations can affect translation efficiency, mRNA stability, ORF length, or RBP interaction as well as cause alternative cleavage and polyadenylation (APA). Prominent examples of genes involved in tumorigenesis (green: induced, red: decreased) that exhibit mutations (red star) in such elements are illustrated.

Diederichs et al.; 2016. EMBO Molecular Medicine 8(5):442-457

B. Variants altering the stability / turnover

mRNA / stability



Sauna et al.; 2012. Nat Rev Genet 12: 683-691

GFP library: 154 copies of GFP with random changes at synonymous sites:

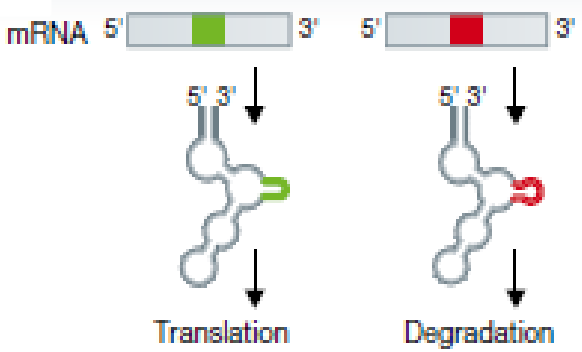
- >250-fold variation in protein levels
- stability of mRNA secondary structure near ATG explained > 50%

Kudla et al.; 2006. PLoS Biol. 4:933-942

B. Variants altering the stability / turnover

mRNA / stability

- CYP2D6: synonymous coding variant causes decreased mRNA expression by altering the secondary structure of the mRNA leading to its degradation (Toscano et al.; 2006)

Functional mechanisms caused by synonymous mutations	Examples in cancer	Examples in other diseases
mRNA stability		
	***CYP2D6*** (?)	***COMT*** Temporo-mandibular joint disorder

Diederichs et al.; 2016. EMBO Molecular Medicine 8(5):442-457

B. Variants altering the stability / turnover

mRNA / stability

Cowden Syndrome–Affected Patients with *PTEN* Promoter Mutations Demonstrate Abnormal Protein Translation

Rosemary E. Teresi, Kevin M. Zbuk, Marcus G. Pezzolesi, Kristin A. Waite, and Charis Eng

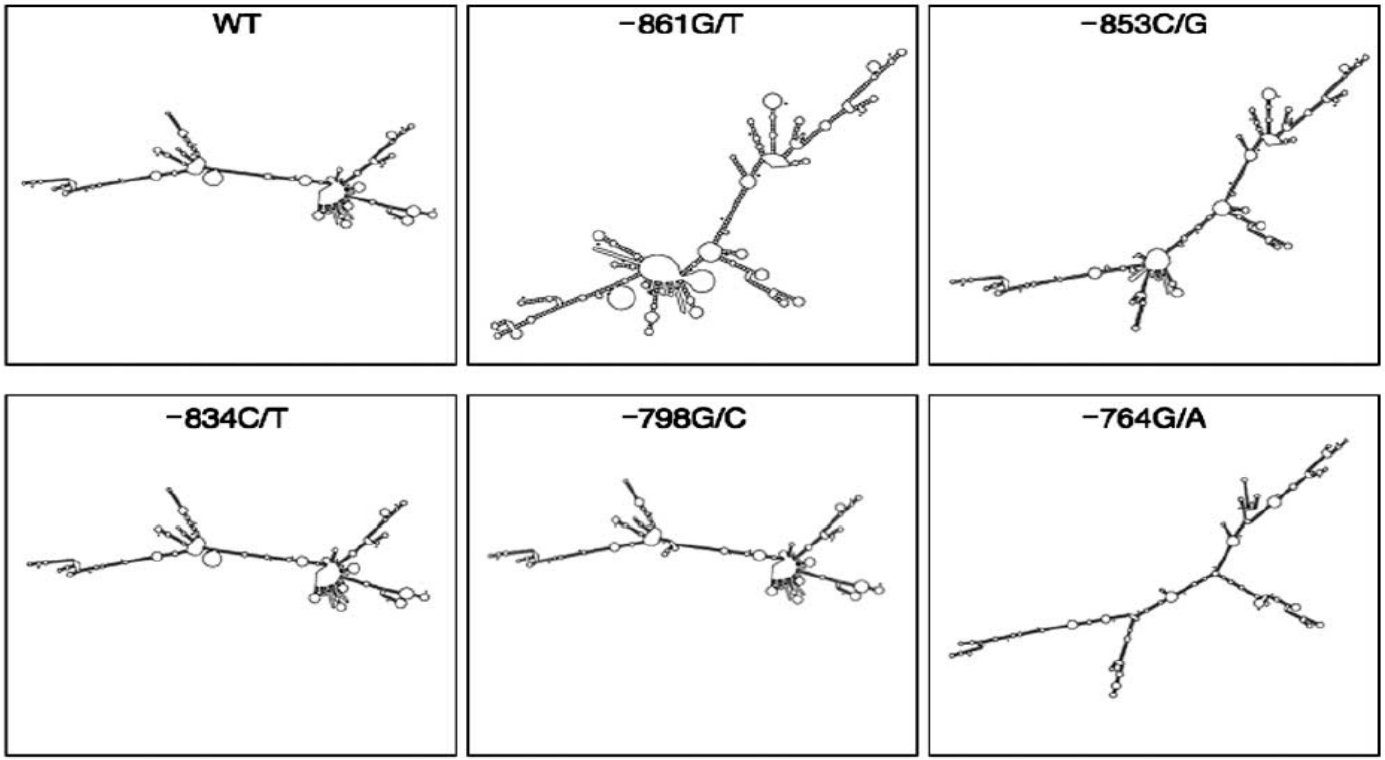
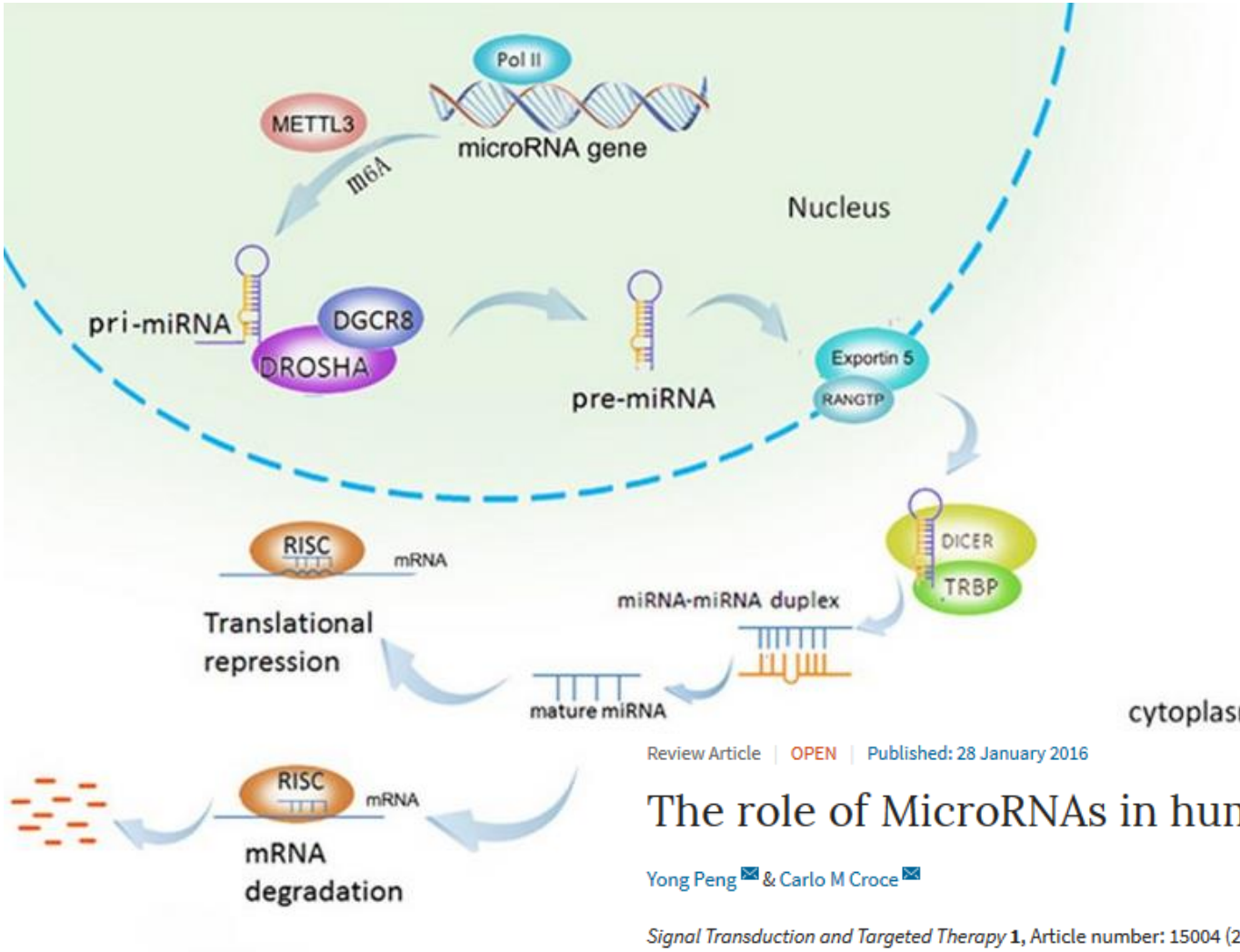


Figure 5. MFOLD-predicted secondary structures resulting from the five VUSs in patients with CS. The most-stable mRNA secondary structures predicted by MFOLD are illustrated here.

Teresi et al.; AJHG. Vol 81 October 2007

B. Variants altering the stability / turnover

mRNA / miRNA binding



Review Article | OPEN | Published: 28 January 2016

The role of MicroRNAs in human cancer

Yong Peng & Carlo M Croce

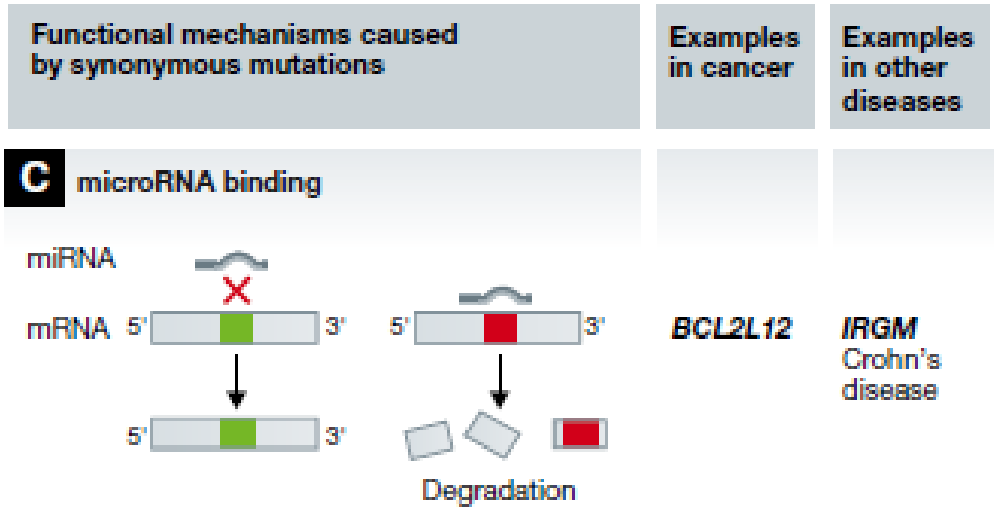
Signal Transduction and Targeted Therapy **1**, Article number: 15004 (2016) | Download Citation

B. Variants altering the stability / turnover

mRNA / miRNA binding

- 1.900 – 3.800 human miRNAs (Friedländer et al.; 2014 Genome Biology / Telonis et al.; 2015 Nucleic Acids Res.)
- ~ 60% (30-80%) of human genes are regulated by miRNAs (Friedmann et al.; 2008 Genome Res)
- 1 miRNA usually targets more than 100 human genes
- A gene may, in turn, be regulated by multiple miRNAs

Melanoma: synonymous coding variant causes increased mRNA stability of the oncogene *BCL2L12* due to loss of the mi-R-671-5p target site
(Gartner et al.; 2013)



Diederichs et al.; 2016. EMBO Molecular Medicine 8(5):442-457

B. Variants altering the stability / turnover

mRNA / miRNA binding

- 1.900 – 3.800 human miRNAs (Friedländer et al.; 2014 Genome Biology / Telonis et al.; 2015 Nucleic Acids Res.)
- ~ 60% (30-80%) of human genes are regulated by miRNAs (Friedmann et al.; 2008 Genome Res)
- 1 miRNA usually targets more than 100 human genes
- A gene may, in turn, be regulated by multiple miRNAs

Table 4. miRNAs in human diseases

Disease type	miRNA	Up/Down Regulation
Cardiac hypertrophy		
	miR-23a, miR-23b, miR-24, miR-195, miR-199a, and miR-214	Up
Down syndrome		
	miR-99a, let-7c, miR-125b-2, miR-155 and miR-802	Up
Alzheimer		
	miR-9, miR-128a, miR-125b	Up
Rheumatic arthritis		
	miR-155, miR-146	Up
Systemic lupus erythematosus		
	miR-189, miR-61, miR-78, miR-21, miR-142-3p, miR 342, miR-299-3p, miR-198 and miR-298	Up
	miR-196a, miR-17-5p, miR- 409-3p, miR-141, miR-383, miR- 112, and miR-184	Down
Psoriasis		
	miR-203	Up

Cancer type	miRNA	Up/Down Regulation
Breast		
	miR-21, miR-155, miR-23, and miR-191	Up
	miR-205, miR- 145, miR-10b, and miR-125b	Down
Ovary		
	miR-200a, miR-200c, and miR-141	Up
	miR-199a, miR-140, miR-145, and miR125bl	Down
Endometrioid adenocarcinoma		
	miR-205, miR155 miR 200a, 200b, 200c	Up
	miR-193a, 193b	Down

Avicenna Journal of Medical Biotechnology, Vol. 2, No. 4, October-December 2010

Potential Consequences on the RNA Level and using prediction tools

A. Variants altering the structure/ integrity: pre-mRNA splicing

B. Variants altering the stability/ turnover: mRNA (UTRs, 3D, miRNA binding)

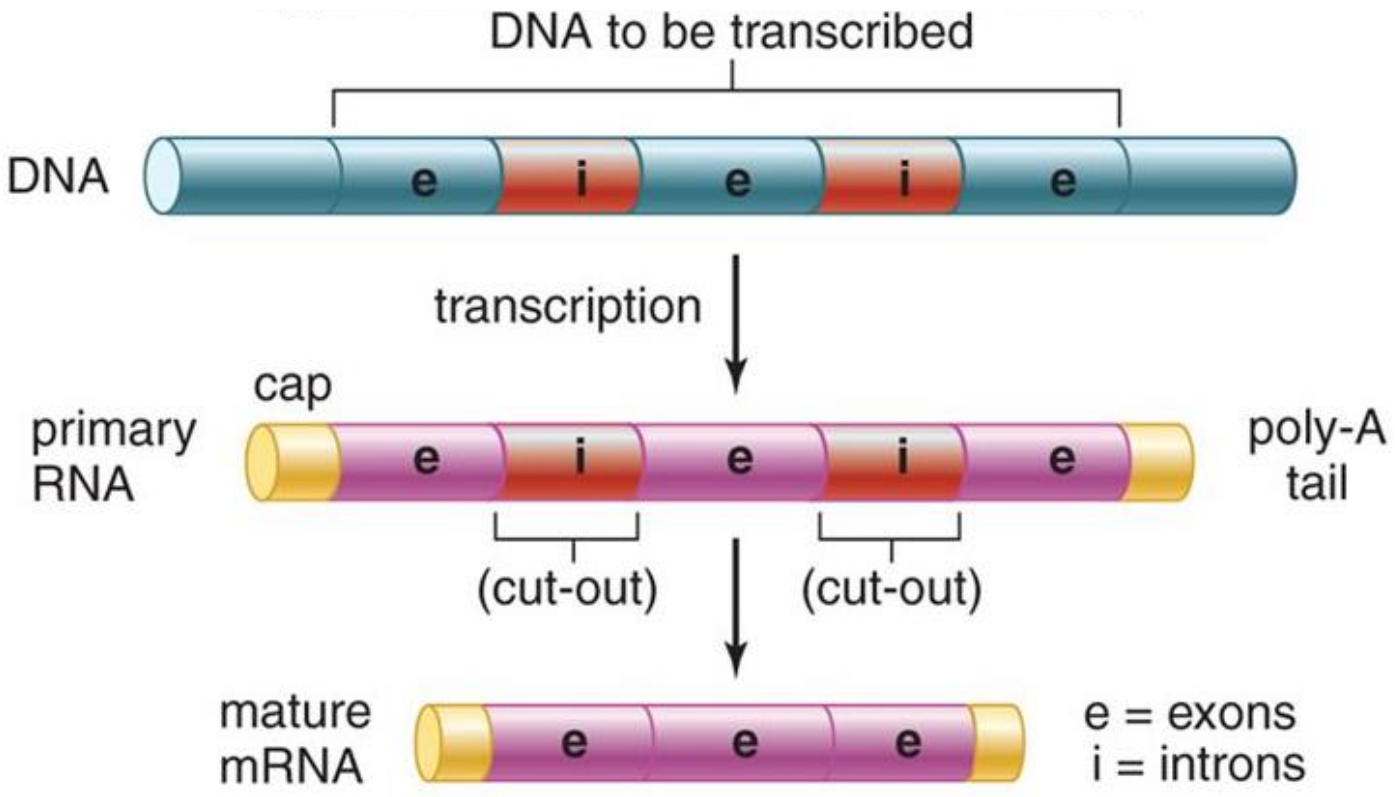
C. Variants altering the translation dynamics: mRNA (codon usage, +/- ribosomal PS)

D. Prediction Tools

E. Functional RNA studies

C. Variants altering the translation dynamics

mRNA / Codon usage



C. Variants altering the translation dynamics

mRNA / Codon usage

- *Codon Usage Bias*: although the genetic code is degenerate, synonymous codons are NOT used in equal frequencies

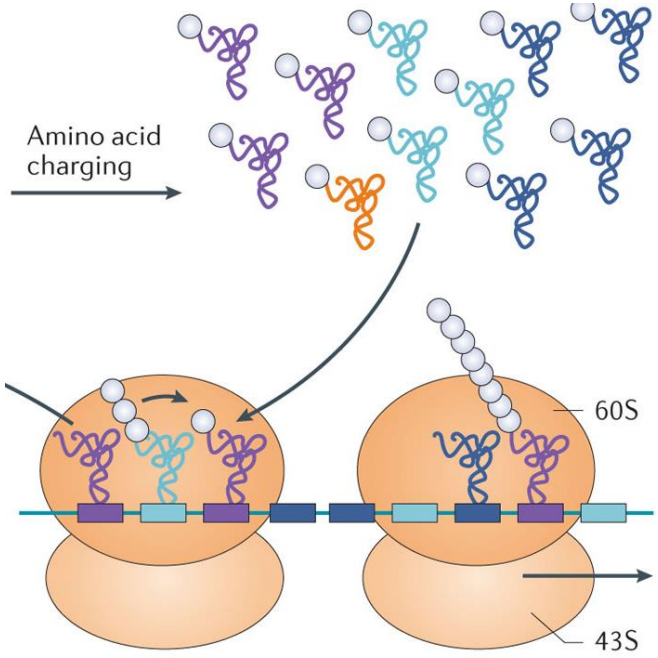
UUU F 0.46	UCU S 0.19	UAU Y 0.44	UGU C 0.46
UUC F 0.54	UCC S 0.22	UAC Y 0.56	UGC C 0.54
UUA L 0.08	UCA S 0.15	UAA * 0.30	UGA * 0.47
UUG L 0.13	UCG S 0.05	UAG * 0.24	UGG W 1.00
CUU L 0.13	CCU P 0.29	CAU H 0.42	CGU R 0.08
CUC L 0.20	CCC P 0.32	CAC H 0.58	CGC R 0.18
CUA L 0.07	CCA P 0.28	CAA Q 0.27	CGA R 0.11
CUG L 0.40	CCG P 0.11	CAG Q 0.73	CGG R 0.20
AUU I 0.36	ACU T 0.25	AAU N 0.47	AGU S 0.15
AUC I 0.47	ACC T 0.36	AAC N 0.53	AGC S 0.24
AUA I 0.17	ACA T 0.28	AAA K 0.43	AGA R 0.21
AUG M 1.00	ACG T 0.11	AAG K 0.57	AGG R 0.21
GUU V 0.18	GCU A 0.27	GAU D 0.46	GGU G 0.16
GUC V 0.24	GCC A 0.40	GAC D 0.54	GGC G 0.34
GUA V 0.12	GCA A 0.23	GAA E 0.42	GGA G 0.25
GUG V 0.46	GCG A 0.11	GAG E 0.58	GGG G 0.25

[Codon/a.a./fraction per codon per a.a.]
Homo sapiens data from the Codon Usage Database

C. Variants altering the translation dynamics

mRNA / Codon usage

- *Codon Usage Bias*: although the genetic code is degenerate, synonymous codons are NOT used in equal frequencies



UUU F 0.46	UCU S 0.19	UAU Y 0.44	UGU C 0.46
UUC F 0.54	UCC S 0.22	UAC Y 0.56	UGC C 0.54
UUA L 0.08	UCA S 0.15	UAA * 0.30	UGA * 0.47
UUG L 0.13	UCG S 0.05	UAG * 0.24	UGG W 1.00
CUU L 0.13	CCU P 0.29	CAU H 0.42	CGU R 0.08
CUC L 0.20	CCC P 0.32	CAC H 0.58	CGC R 0.18
CUA L 0.07	CCA P 0.28	CAA Q 0.27	CGA R 0.11
CUG L 0.40	CCG P 0.11	CAG Q 0.73	CGG R 0.20
AUU I 0.36	ACU T 0.25	AAU N 0.47	AGU S 0.15
AUC I 0.47	ACC T 0.36	AAC N 0.53	AGC S 0.24
AUA I 0.17	ACA T 0.28	AAA K 0.43	AGA R 0.21
AUG M 1.00	ACG T 0.11	AAG K 0.57	AGG R 0.21
GUU V 0.18	GCU A 0.27	GAU D 0.46	GGU G 0.16
GUC V 0.24	GCC A 0.40	GAC D 0.54	GGC G 0.34
GUA V 0.12	GCA A 0.23	GAA E 0.42	GGA G 0.25
GUG V 0.46	GCG A 0.11	GAG E 0.58	GGG G 0.25

[Codon/a.a./fraction per codon per a.a.]
Homo sapiens data from the Codon Usage Database

C. Variants altering the translation dynamics

mRNA / Codon usage

- *Codon Usage Bias*: although the genetic code is degenerate, synonymous codons are NOT used in equal frequencies

UUU F 0.46	UCU S 0.19	UAU Y 0.44	UGU C 0.46
UUC F 0.54	UCC S 0.22	UAC Y 0.56	UGC C 0.54
UUA L 0.08	UCA S 0.15	UAA * 0.30	UGA * 0.47
UUG L 0.13	UCG S 0.05	UAG * 0.24	UGG W 1.00
CUU L 0.13	CCU P 0.29	CAU H 0.42	CGU R 0.08
CUC L 0.20	CCC P 0.32	CAC H 0.58	CGC R 0.18
CUA L 0.07	CCA P 0.28	CAA Q 0.27	CGA R 0.11
CUG L 0.40	CCG P 0.11	CAG Q 0.73	CGG R 0.20

T 0.25 AAU N 0.47 AGU S 0.15
T 0.36 AAC N 0.53 AGC S 0.24
T 0.28 AAA K 0.43 AGA R 0.21
T 0.11 AAG K 0.57 AGG R 0.21

A 0.27 GAU D 0.46 GGU G 0.16
A 0.40 GAC D 0.54 GGC G 0.34
A 0.23 GAA E 0.42 GGA G 0.25
A 0.11 GAG E 0.58 GGG G 0.25

ion per codon per a.a.]
from the Codon Usage Database

Mutation Report

Report for mutation **NM_001009944.2(PKD1):c.8151C>A**

Warning: *This report is based on knowledge and data that are not firmly established. Consequently, medical decisions must not be made on the basis of this report.*

PKD1 Variation

Class 3-Unknown pathogenicity

Transversion from C to A in exon 22.
Synonymous substitution. Codon CTC changed to CTA.
Frequencies of Leu-codons in the human genome: CTC (0.197) / CTA (0.07)
This variant does not alter the protein sequence.

HGVS v2.0 Nomenclature

cDNA Level:	**NM_001009944.2:c.8151C>A**
gDNA Level:	**Chr16(GRCh37):g.2154509G>T**
Protein Level:	**p.=** *(p.Leu2717Leu)*

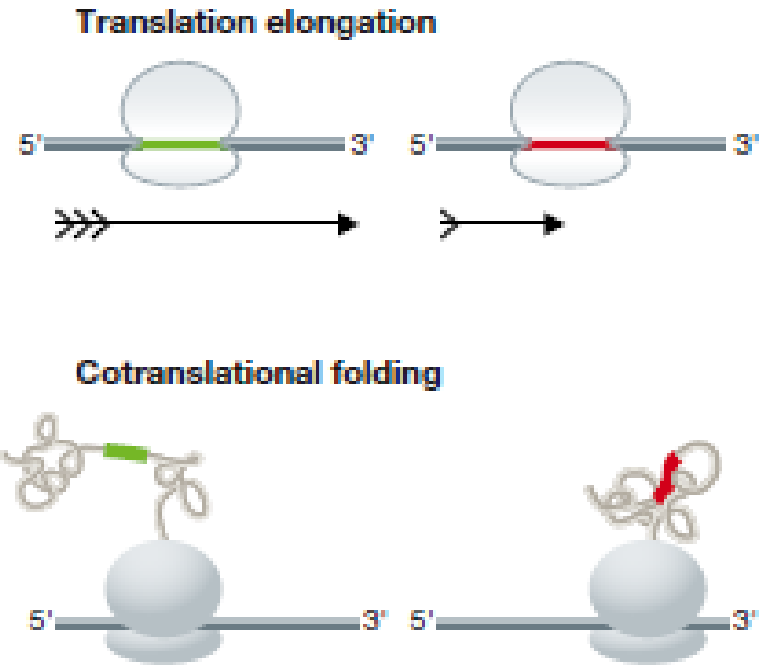
©Interactive Biosoftware - Created by Alamut Visual v.2.6 on 19.10.2016

C. Variants altering the translation dynamics

mRNA / translational speed

- *Codon Usage Bias*: although the genetic code is degenerate, synonymous codons are NOT used in equal frequencies

- Variants can alter translational speed (if a abundant codon is changed to a rare one) leading to a change in cotranslational protein folding (Yu et al, 2015)



Diederichs et al.; 2016. EMBO Molecular Medicine 8(5):442-457

Potential Consequences on the RNA Level and using prediction tools

A. Variants altering the structure/ integrity: pre-mRNA splicing

B. Variants altering the stability/ turnover: mRNA (UTRs, 3D, miRNA binding)

C. Variants altering the translation dynamics: mRNA (codon usage, +/- ribosomal PS)

D. Prediction Tools

E. Functional RNA studies

D. Prediction Tools

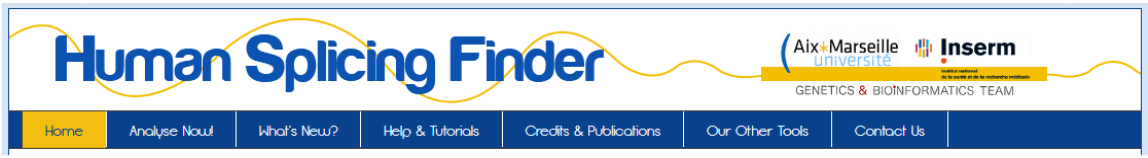
pre-mRNA splicing

Tool	Input	Output	Interpretation	Method
Splice-Site Analyzer Tool	Single/multiple sequences (5′: 9 bp (−3 to +6); 3′: 15 bp (−14 to +1))	S & S score (0–100)	Higher score implies a more similar ss sequence with the consensus sequence	
★ NetGene2	Single sequence (200 bp < length < 80,000 bp)	Confidence score (0–1)	Higher score implies a higher confidence of true site	Neural Networks
★ NNSplice	Single/multiple sequences	Score (0–1)	Higher score implies greater potential for splice site	Neural Networks
GENSCAN	Single sequence ≤1 million bp	Probability score (0–1)	Higher score implies a higher probability of correct exon	
SpliceView	Single sequence ≤31,000 bp	S & S score (0–100)	Higher score implies a more similar ss sequence with the consensus sequence	
Hbond	Single/multiple 11 bp sequences (−3 to +8) containing GT in +1/+2 or one genomic sequence	Hbond score	Higher score implies a stronger capability of forming H-bonds with U1 small nuclear RNA	
★ MaxEntScan	Single/multiple sequences (5′: 9 bp (−3 to +6); 3′: 23 bp (−20 to +3))	Maximum entropy score (log odds ratio)	Higher score implies a higher probability of the sequence being a true splice site	Maximum entropy principle
SplicePredictor	Single/multiple sequences	*-Value (3–15) determined by *P*, ρ, and γ values	Higher value implies greater reliability of the predicted splice site	
Automated splice site analyses	Mutation to be analyzed and the reference sequence	Information contents Ri	Color coded by direction and type of change in Ri	
SplicePort	Single/multiple sequences ≤30,000 bp	Feature generation algorithm score	Higher score implies a more precise prediction of splice site	
★ Human Splicing Finder	Single sequence ≤5,000 bp	S & S score (0–100)	Higher score implies greater potential for splice site	Position dependent logic
CRYP-SKIP	Single/multiple sequences ≤4,000 bp containing one exon in upper case and flanking intronic sequence ≥4 bp in lower case	Probability of cryptic ss activation (0–1)	Higher value implies a higher probability of cryptic ss activation as opposed to exon skipping	
SROOGLE	Target exon along with two flanking introns	Different scores with their percentile scores (0–1)	Higher percentile score implies a higher ranking of the ss within precalculated distributions	
AASsites	Single sequence containing the SNP(s) and the Ensembl gene ID to which the SNP(s) belong(s)	Classification of the probability for a change in splicing	Probable, likely, or unlikely	
Spliceman	Single/multiple sequences with one mutation and ≥5 bp in each side of the mutation	L1 distance and percentile rank	Higher percentile rank implies a higher likelihood the point mutation is to disrupt splicing	

Review of user orientated in silico tools for splicing: Xueqiu et al.; Genetics in Medicine (2013) 16

D. Prediction Tools

pre-mRNA splicing



Type of signal	Algorithm type	Prediction algorithm		CV threshold	Variation threshold	Comment
Donor or acceptor splice site	Position Weight Matrices	HSF		65	+/-10%	Consensus values go from 0 to 100 for HSF, -20 to +20 for MaxEnt. The threshold is defined at 65 for HSF, 3 for MaxEnt. This means that every signal with a score above the threshold is considered to be a splice site (donor or acceptor). When a mutation occurs, if the WT score is above the threshold and the score variation (between WT and Mutant) is under -10% for HSF (-30% for MaxEnt) we consider that the mutation breaks the splice site. In the other case, if the WT score is under the threshold and the score variation is above +10% for HSF (+30% for MaxEnt) we consider that the mutation creates a new splice site.
	Maximum Entropy	MaxEntScan		3	+/-30%	
Branch point site	Position Weight Matrices	HSF		67	+/-10%	Consensus values go from 0 to 100 and the threshold is defined at 67. This means that every signal with a score above 67 is considered to be a potential branch point. When a mutation occurs, if the WT score is above 67 and the score variation (between WT and Mutant) is under -10% we consider that the mutation breaks the branch point.
Exonic Splicing Enhancers (ESE)	Position Weight Matrices	HSF	9G8	59.24	Yes/No	Consensus values go from 0 to 100 and the threshold is defined differently for each algorithm. Every signal with a score above the defined threshold is considered to be a potential ESE. When a mutation occurs, if the WT score is above the threshold and the Mutant score is under it we consider that the mutation breaks the ESE.
			Tra2-β	75.96		
		ESE Finder	SF2/ASF	72.98		
			SF2/ASF(IgM)	70.51		
			SC35	75.05		
			SRp40	78.08		
			SRp55	73.86		
	Motif Comparison method	RESCUE ESE hexamers		Present/Absent		If the tested motif exists in the database, it is considered to be a potential ESE. When a mutation occurs, if the WT motif is present in the database and the Mutant one is absent we consider that the mutation breaks the ESE.
Exonic Splicing Silencers (ESS)	Position Weight Matrices	HSF hnRNP-A1		65.476		Consensus values go from 0 to 100 and the threshold is defined differently for each algorithm. Every signal with a score above the defined threshold is considered to be a potential ESS. When a mutation occurs, if the WT score is under the threshold and the Mutant score is above it we consider that the mutation creates a new ESS.
		Sironi motifs		60		
	Motif Comparison method	ESS decamers from Wang et al.		Present/Absent		If the tested motif exists in the database, it is considered to be a potential ESS. When a mutation occurs, if the WT motif is absent in the database and the Mutant one is present we consider that the mutation creates a new ESS.
Both ESEs and ESSs		PESE & PESS Octamers				If the tested motif exists in the database, it is considered to be a potential ESE or ESS. When a mutation occurs, if the WT motif is present in the database and the Mutant one is absent we consider that the mutation breaks the ESE. Else if the WT motif is absent in the database and the Mutant one is present we consider that the mutation creates a new ESS.
		ESR Sequences				
		ElEs & IlEs Hexamers				

HSF3 Pro takes both the U2 and U12 introns into account

http://www.umd.be/HSF3/

D. Prediction Tools

pre-mRNA splicing

RESEARCH ARTICLE

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

Guidelines for Splicing Analysis in Molecular Diagnosis Derived from a Set of 327 Combined *In Silico/In Vitro* Studies on *BRCA1* and *BRCA2* Variants

Claude Houdayer,[1]* Virginie Caux-Moncoutier,[1] Sophie Krieger,[2] Michel Barrois,[3] Françoise Bonnet,[4] Violaine Bourdon,[5] Myriam Bronner,[6] Monique Buisson,[7] Florence Coulet,[8] Pascaline Gaildrat,[9] Cédrick Lefol,[10] Mélanie Léone,[11] Sylvie Mazoyer,[7] Danielle Muller,[12] Audrey Remenieras,[3] Françoise Révillion,[13] Etienne Rouleau,[10] Joanna Sokolowska,[6] Jean-Philippe Vert,[14] Rosette Lidereau,[10] Florent Soubrier,[8] Hagay Sobol,[5] Nicolas Sevenet,[4] Brigitte Bressac-de Paillerets,[3,15] Agnès Hardouin,[2] Mario Tosi,[9] Olga M. Sinilnikova,[7,11] and Dominique Stoppa-Lyonnet[1,16]

Comprehensive *in silico* analysis (MES, SSF, NNsplice, HSF; ESEfinder, Rescue-ESE)

Comprehensive *in vitro* mRNA analysis (cDNA: PAX + cell culture; mini-gene)

Comparison of different *in silico* tools with regard to specificity and sensitivity

Houdayer et al. *Hum Mutat* 2013

D. Prediction Tools

pre-mRNA splicing

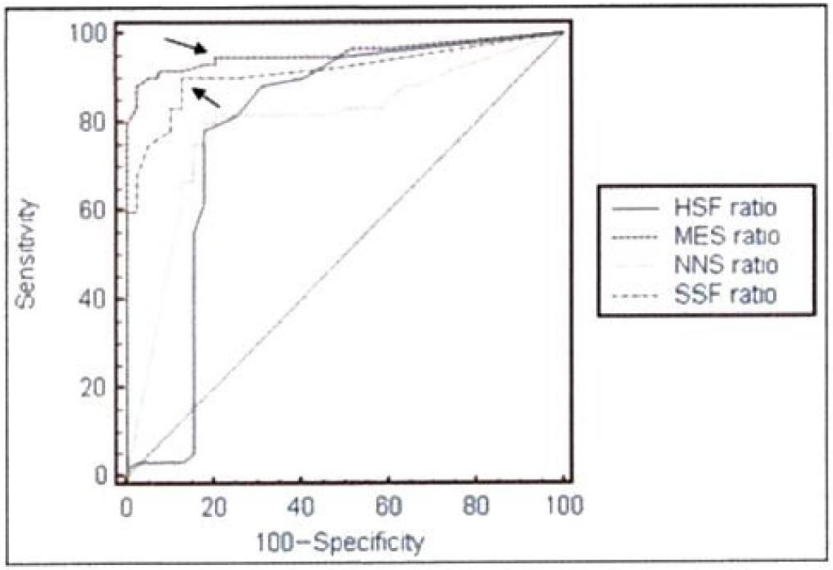
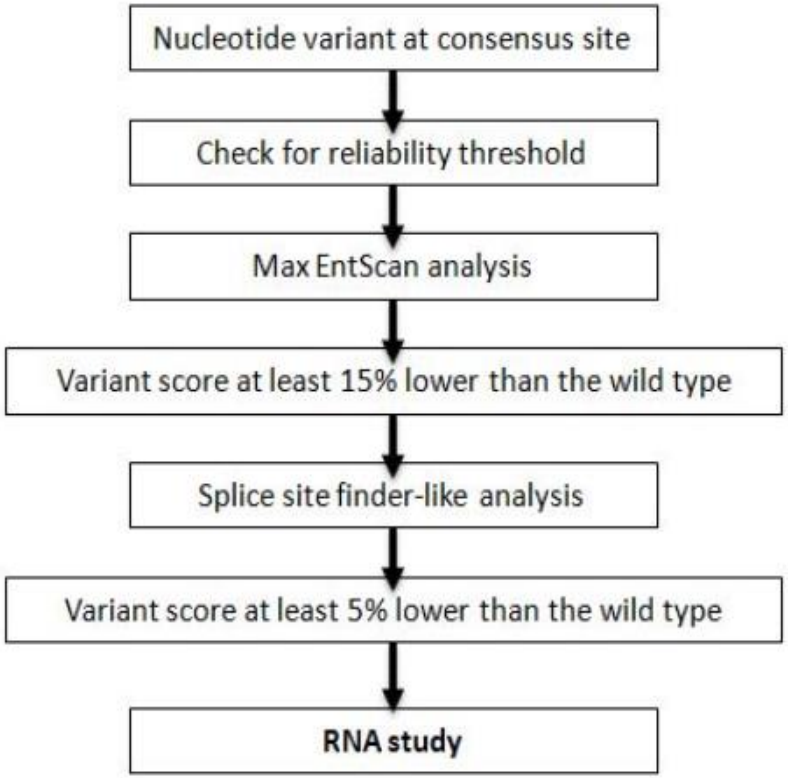


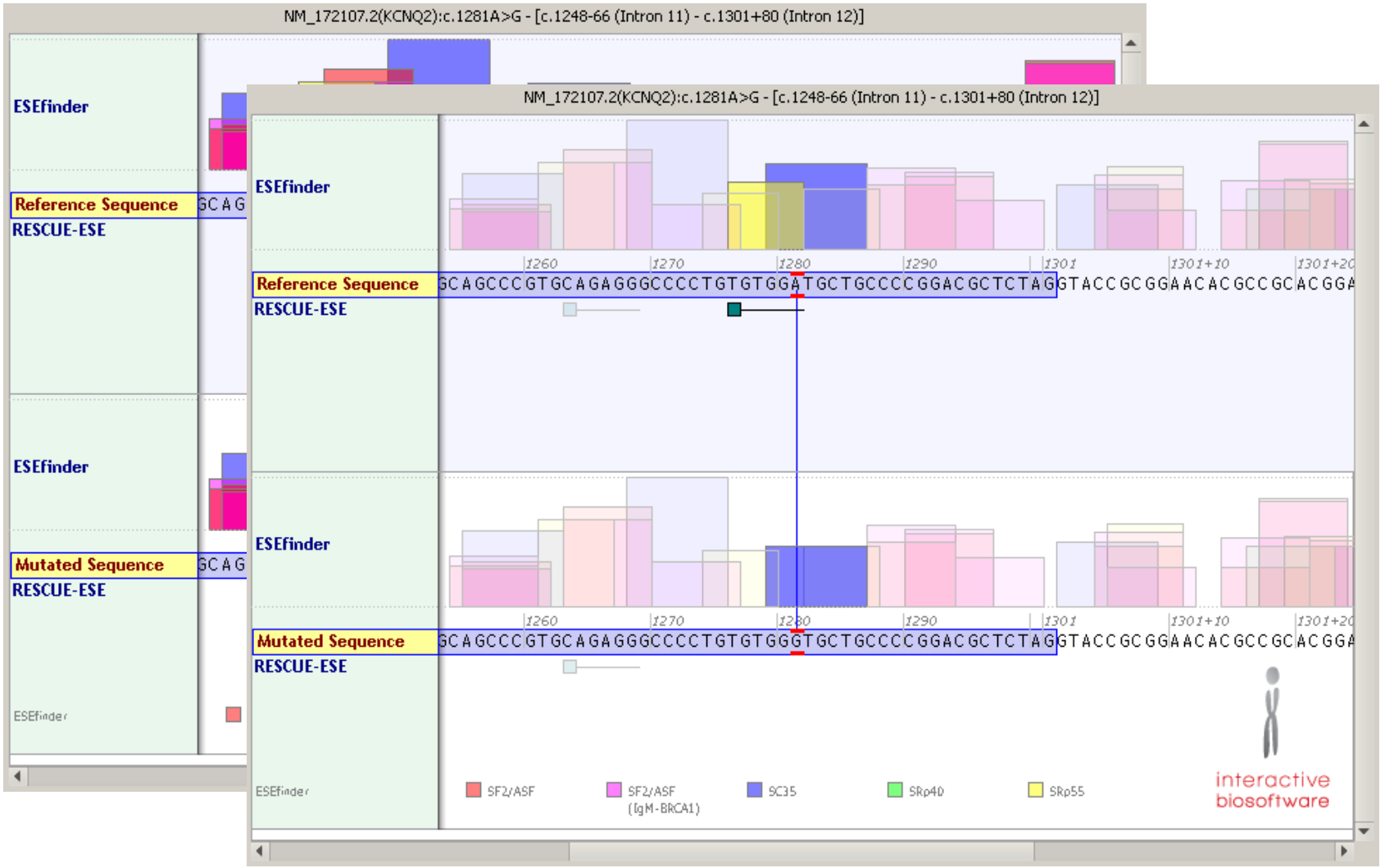
Figure 1. ROC plots for the four *in silico* tools. The ROC graph is a plot of all sensitivity/specificity pairs resulting from continuously varying the decision threshold over the entire set of results observed. For each prediction tool, the ROC curve plots sensitivity versus false-positive rate (100–specificity) for the complete range of decision thresholds. A perfect test has a ROC curve that passes through the upper left corner, where both sensitivity and specificity are 100%. Hence, MaxEntscan and Splice Site Finder-like provide the best ROC curves and also show excellent accuracy with areas under the curve of 0.956 and 0.914, respectively (see text for details). The points corresponding to decision thresholds of 15% and 5% are marked with arrowheads on the MES and the SSF curves, respectively.



MES+SSF: 96% sensitivity and 83% specificity

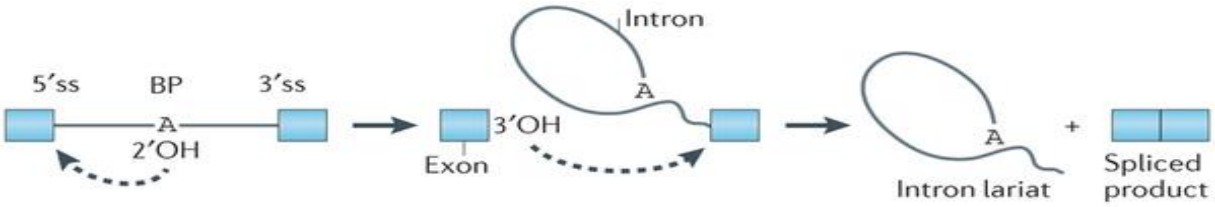
D. Prediction Tools

pre-mRNA splicing / ESE ESS ISE ISS



D. Prediction Tools

pre-mRNA splicing / Branchpoint



RNA SPLICING BRANCHPOINT DETECTION SOFTWARE TOOLS | GENOME ANNOTATION

GAEM / Genetic Algorithm-based weighted average Ensemble ...
Allows branchpoint (BP) determination. GAEM is an ensemble of learning method that integrates several features and multiple classifiers to construct BP p

LREM / Logistic Regression-based Ensemble Method
Allows branchpoint (BP) prediction. LREM is an ensemble of learning scheme that integrates different features and different classifiers to build BP predictio

LaBranchoR / Long short-term memory network Branchpoint Re...
Provides accurate genome-wide branchpoint annotations. LaBranchoR is a computational method that disregards noise in the experimental data lea

branchpointer
Allows users to identify branchpoints throughout the human genome thanks to gene annotations. Branchpointer generates branchpoint window regions

D. Prediction Tools

pre-mRNA splicing / Branchpoint

New Results

A sequence-based, deep learning model accurately predicts RNA splicing branchpoints

Joseph M. Paggi, Gill Bejerano

doi: https://doi.org/10.1101/185868

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract | Info/History | Metrics | Preview PDF

Abstract

Experimental detection of RNA splicing branchpoints, the nucleotide serving as the nucleophile in the first catalytic step of splicing, is difficult. To date, annotations exist for only 16-21% of 3' splice sites in the human genome and even these limited annotations have been shown to be plagued by noise. We develop a sequence-only, deep learning based branchpoint predictor, LaBranchoR, which we conclude predicts a correct branchpoint for over 90% of 3' splice sites genome-wide. Our predicted branchpoints show large agreement with trends observed in the raw data, but analysis of conservation signatures and overlap with pathogenic variants reveal that our predicted branchpoints are generally more reliable than the raw data itself. We use our

D. Prediction Tools

miRNA prediction

Gene: **SLC2A1** - Transcript: **NM_006516.2** - Variant: **c.*255T>C** - 3' UTR: **1666 bp**

□□ *Good mirSVR score, Conserved miRNA*
□□ *Good mirSVR score, Non-conserved miRNA*
□□ *Non-Good mirSVR score, Conserved miRNA*
□□ *Non-Good mirSVR score, Non-Conserved miRNA*
– *Not Target*, Source: microRNA.org

Removed (wild-type)

miRNA name	Targets	Position in 3' UTR	Alignment Length	Alignment Score	Identity	Similarity
hsa-miR-495	□□	c.*236_*257	6	120.00	100.00%	100.00%
hsa-miR-7-1*	□□	c.*236_*257	6	120.00	100.00%	100.00%
hsa-miR-7-2*	□□	c.*236_*257	6	120.00	100.00%	100.00%

Modified (wild-type □ mutant)

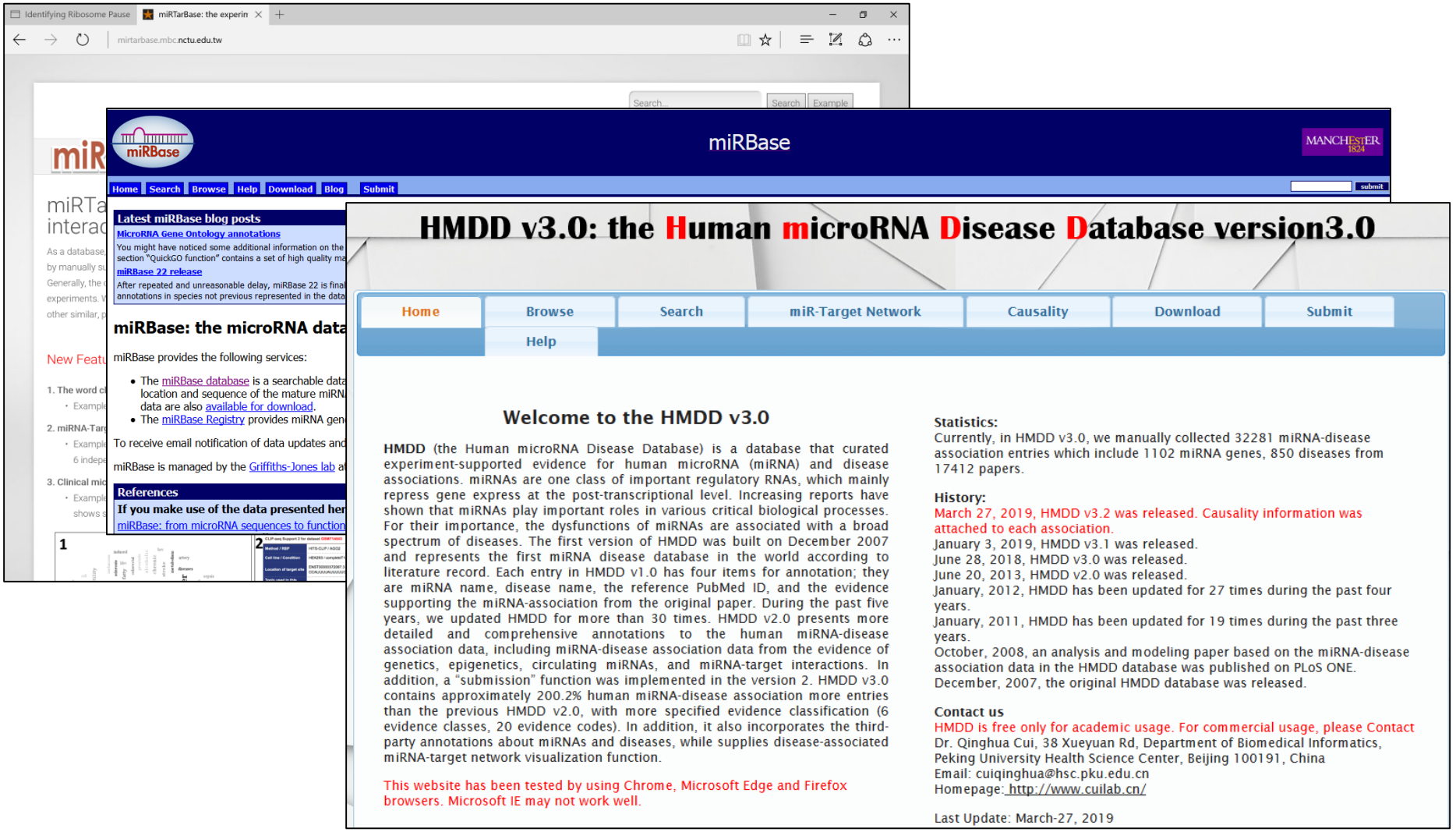
miRNA name	Targets	Position in 3' UTR	Alignment Length	Alignment Score	Identity	Similarity
hsa-miR-3065-5p	□□	c.*236_*258	16	154.00 □ 122.00	81.25% □ 75.00%	93.75% □ 87.50%
hsa-miR-548p	□□	c.*236_*257	17	138.00 □ 154.00	64.71% □ 70.59%	76.47%
hsa-miR-340 (1)	□□	c.*244_*266 □ c.*245_*266	16 □ 19	135.00 □ 132.00	62.50% □ 57.89%	87.50% □ 68.42%

Added (mutant)

miRNA name	Targets	Position in 3' UTR	Alignment Length	Alignment Score	Identity	Similarity
hsa-miR-31	–	c.*238_*257	18	121.00	61.11%	77.78%
hsa-miR-3121	–	c.*240_*261	6	120.00	100.00%	100.00%
hsa-miR-545	–	c.*235_*258	15	126.00	73.33%	73.33%

D. Prediction Tools

miRNA prediction



Potential Consequences on the RNA Level and using prediction tools

A. Variants altering the structure/ integrity: pre-mRNA splicing

B. Variants altering the stability/ turnover: mRNA (UTRs, 3D, miRNA binding)

C. Variants altering the translation dynamics: mRNA (codon usage, +/- ribosomal PS)

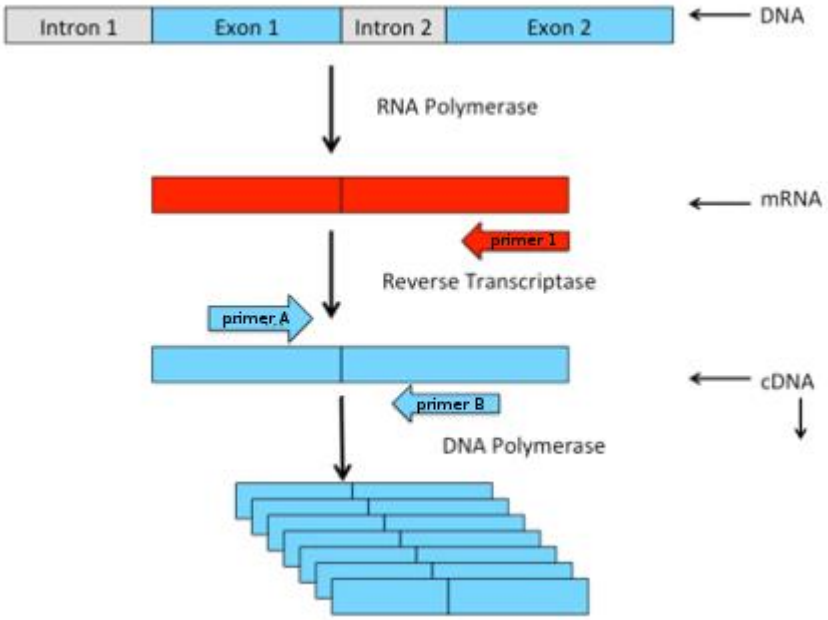
D. Prediction Tools

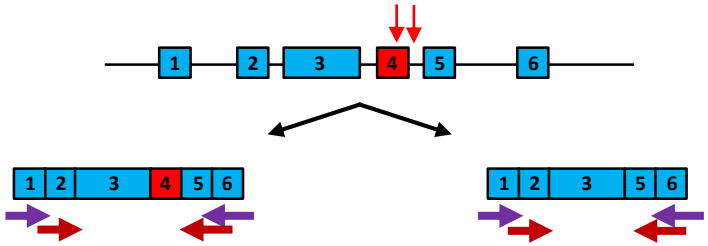
E. Functional RNA studies

E. Functional RNA studies

Strategies for RNA Analysis

- RT-PCR approach
- Assess allele-specific expression
- Quantify (alternative) transcripts
- Novel transcripts / gene fusions

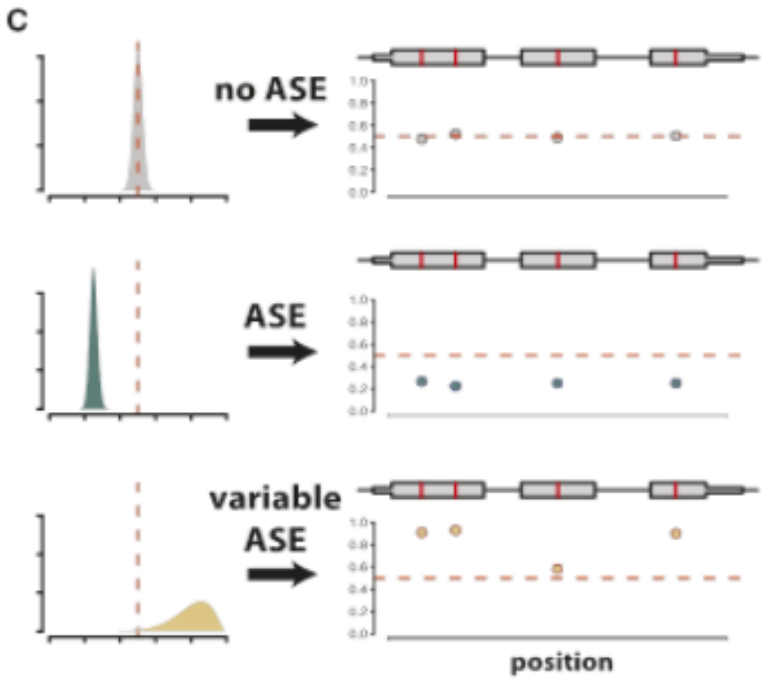




E. Functional RNA studies

Strategies for RNA Analysis

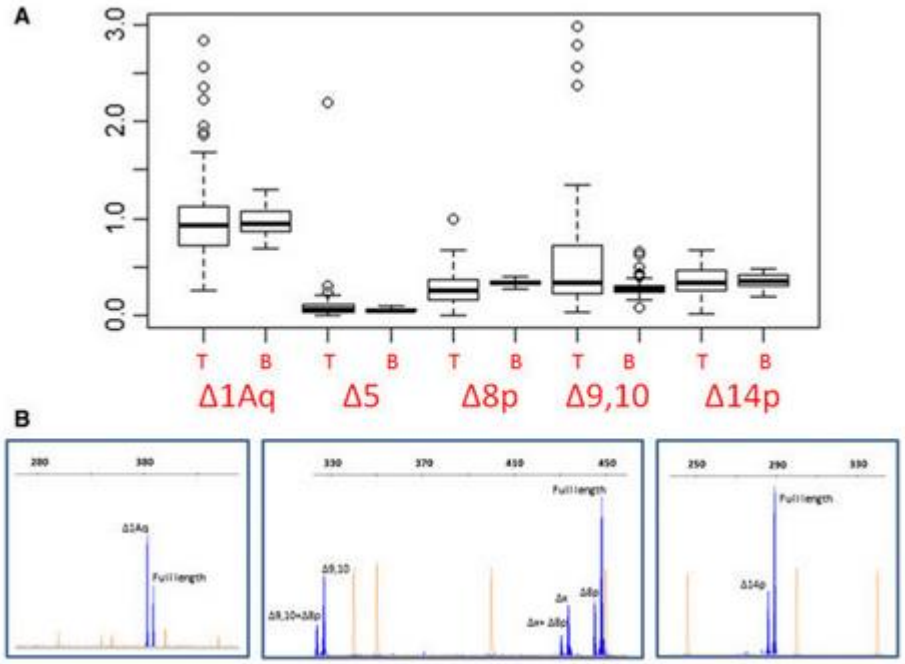
- RT-PCR approach
- Assess allele-specific expression
- Quantify (alternative) transcripts
- Novel transcripts / gene fusions



E. Functional RNA studies

Strategies for RNA Analysis

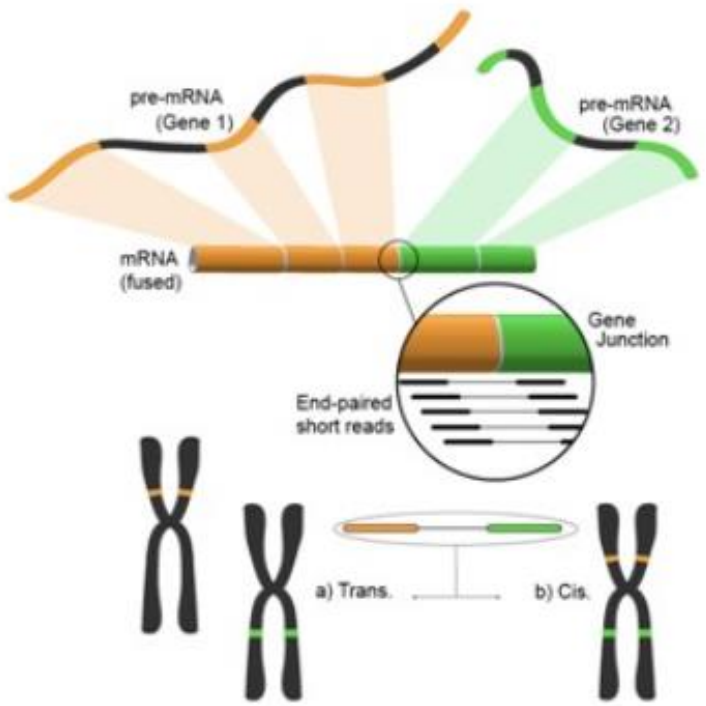
- RT-PCR approach
- Assess allele-specific expression
- Quantify (alternative) transcripts
- Novel transcripts / gene fusions



E. Functional RNA studies

Strategies for RNA Analysis

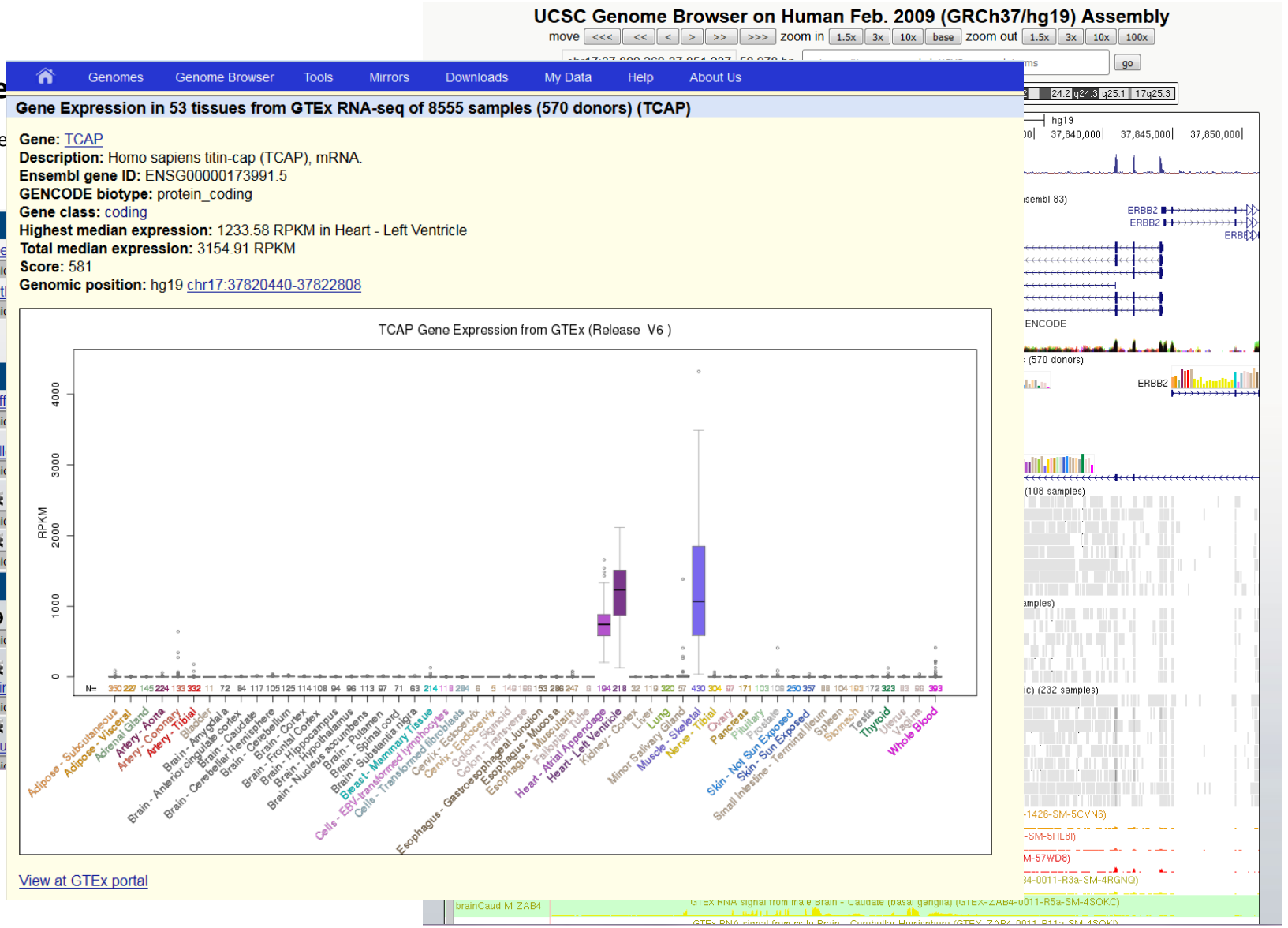
- RT-PCR approach
- Assess allele-specific expression
- Quantify (alternative) transcripts
- Novel transcripts / gene fusions



E. Functional RNA studies

Is blood the

Is the gene of inte



Potential Consequences on the RNA Level and using prediction tools