AlloSeq™ Assign®

Addition of splice sites to the AlloSeq Assign v1.0 references

Margot Duane
Stem Cell Transplant Recipient

CareDx Technical Support

TEC746-S_AlloSeq Assign 1.0_Splicing Version 1.0 Effective 14 Jun 22

CareDx®

Table of Contents

1. What is splicing?
2. The Spliceosome
3. What is alternate splicing?
4. Why add splice sites to the core layer in Assign?
5. IMGT/HLA characterized splicing variants
6. Examples of unexpressed and alternatively expressed HLA alleles
   - DRB4*01:03:01N: Example of a null allele
   - DQB1 and alternative splicing
   - Impacts of DQB1 non-expressed exon 5 in Assign
     G*01:01:01:14Q: An Example of an allele that is alternatively expressed due to splice site mutation
   - B*44:02:01:02S[4]: an example of an allele with alternative splicing due to splice site mutation
7. Bibliography

CareDx®

1. What is Splicing?

- Splicing is the process by which the noncoding regions of genes, termed introns, are excised out of the primary messenger RNA (mRNA) transcript. The coding regions, termed exons, are subsequently joined together to form mature mRNA.
- Splicing usually takes place in the cell nucleus after or during transcription of the DNA template.

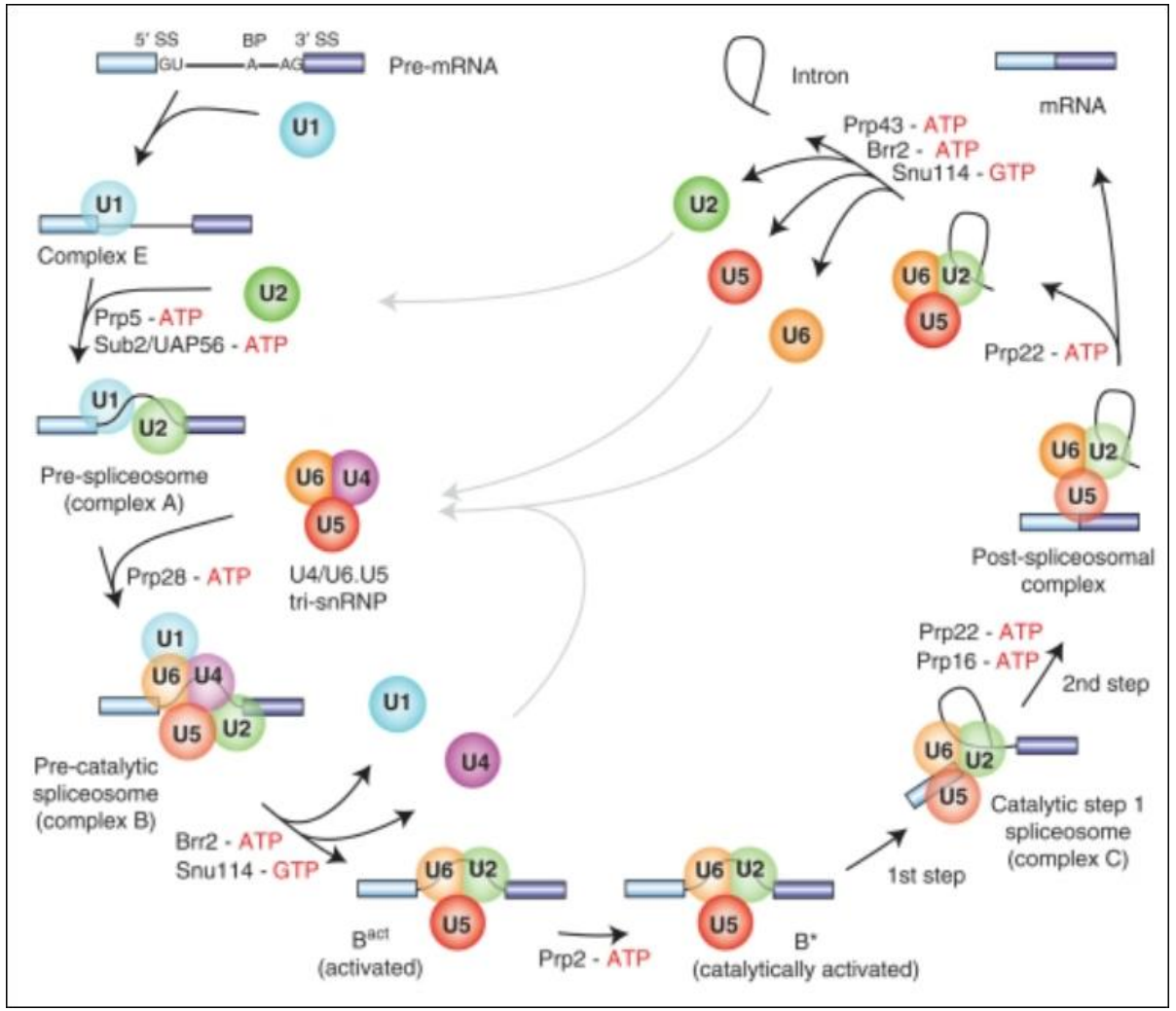
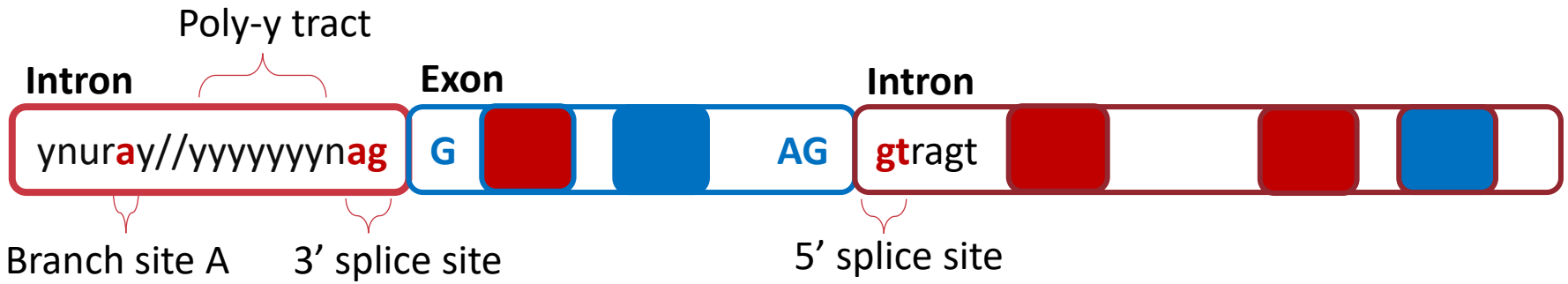


Image credit: Spliceosome Structure and Function by Max Planck Institute for Biophysical Chemistry, Department of Cellular Biochemistry(7)

CareDx®

2. The Spliceosome

- The splicing machinery, known as the spliceosome, identifies splicing signals and initiates the “cut and paste” splicing reactions
- The spliceosome consists of :
  - 5 small nuclear ribonucleoproteins (snRNPs) and more than 100 proteins
  - Conserved short sequences (**ag**/**gt**) at the 3’ and 5’ sites called splice sites
  - Polypyrimidine (Poly-Y) tract of variable length upstream of the 3’site
  - Branch site A made of 18-40 nucleotides upstream of the Poly-Y tract



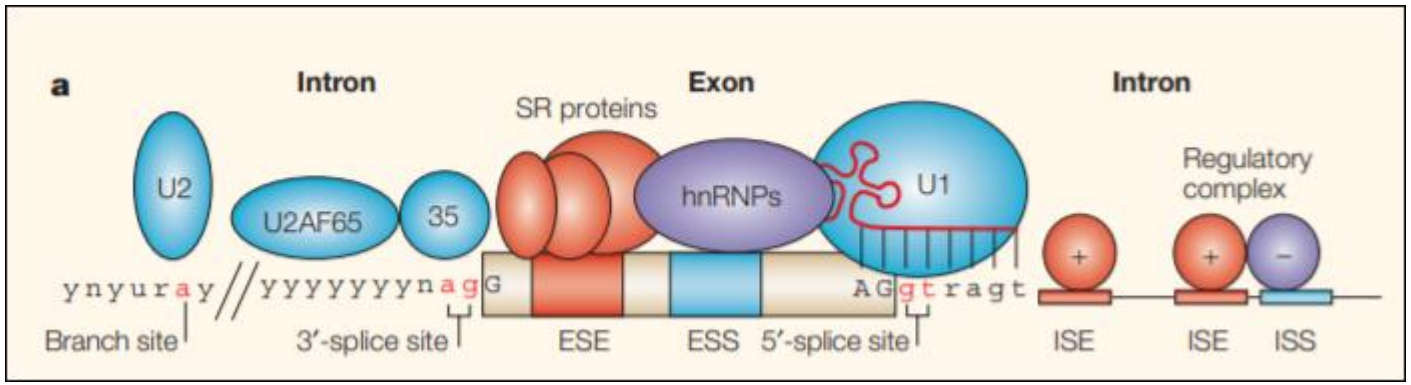
CareDx®

2.1 Additional Components of The Spliceosome

- Additional Exonic/Intronic Splicing Enhancers **(ESE, ISE**) and Exonic/Intronic Splicing Silencers (**ESS, ISS**) may be present within the spliceosome. These elements allow the correct splice sites to be distinguished from the many cryptic splice sites that have identical signal sequences.
- The enhancers are generally considered to be binding sites for serine arginine (SR) proteins, which often function to activate splicing.

CareDx®

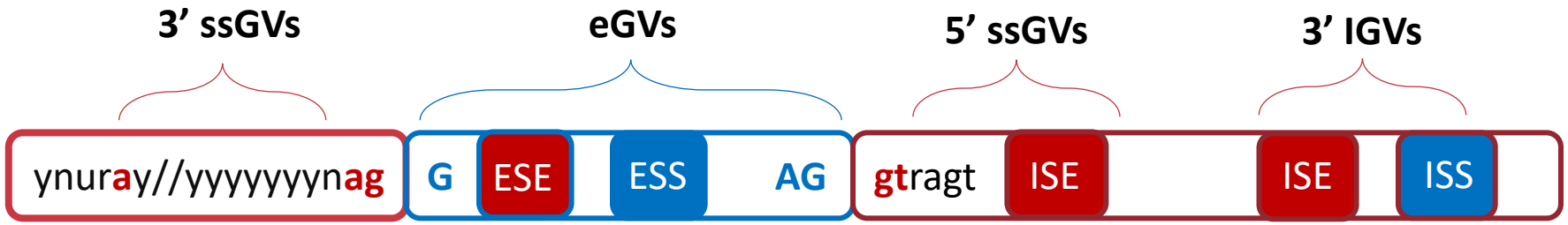
2.2 Protein Binding



- Protein–protein interactions in the spliceosome that modulate the recognition of the splice sites are the probable cause of splicing inhibition or activation.
- The U1 snRNP binds to the 5′ splice site while the U2 snRNP binds the branch site through RNA-RNA interactions.
- Trans-acting splicing factors can interact with enhancers and silencers.
- Trans-acting splicing factors fall into 2 groups:
    1) The serine arginine (SR) family of proteins (SR protein binding at ESE facilitates exon recognition)
    2) Inhibitory heterogeneous nuclear ribonucleoprotein particles (hnRNPs)
- The U2AF65 protein binds to the Poly-Y tract during spliceosome assembly.

Image credit: Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet*

2.3 Genomic Variants



- Splice site Genomic Variants (**ssGVs**), exonic Genomic Variants (**eGVs**) and Intronic Genomic Variants (**IGVs**) such as single nucleotide substitutions, small insertions or deletions may result in alternative splicing, loss of expression or changes in the level of protein expression.

CareDx®

3. What is alternative splicing?

- Alternative splicing is the process by which the production of different mature mRNA molecules occurs from the same initial transcript.
- Alternative splicing can occur as a result of a splice site mutation (genomic variant).

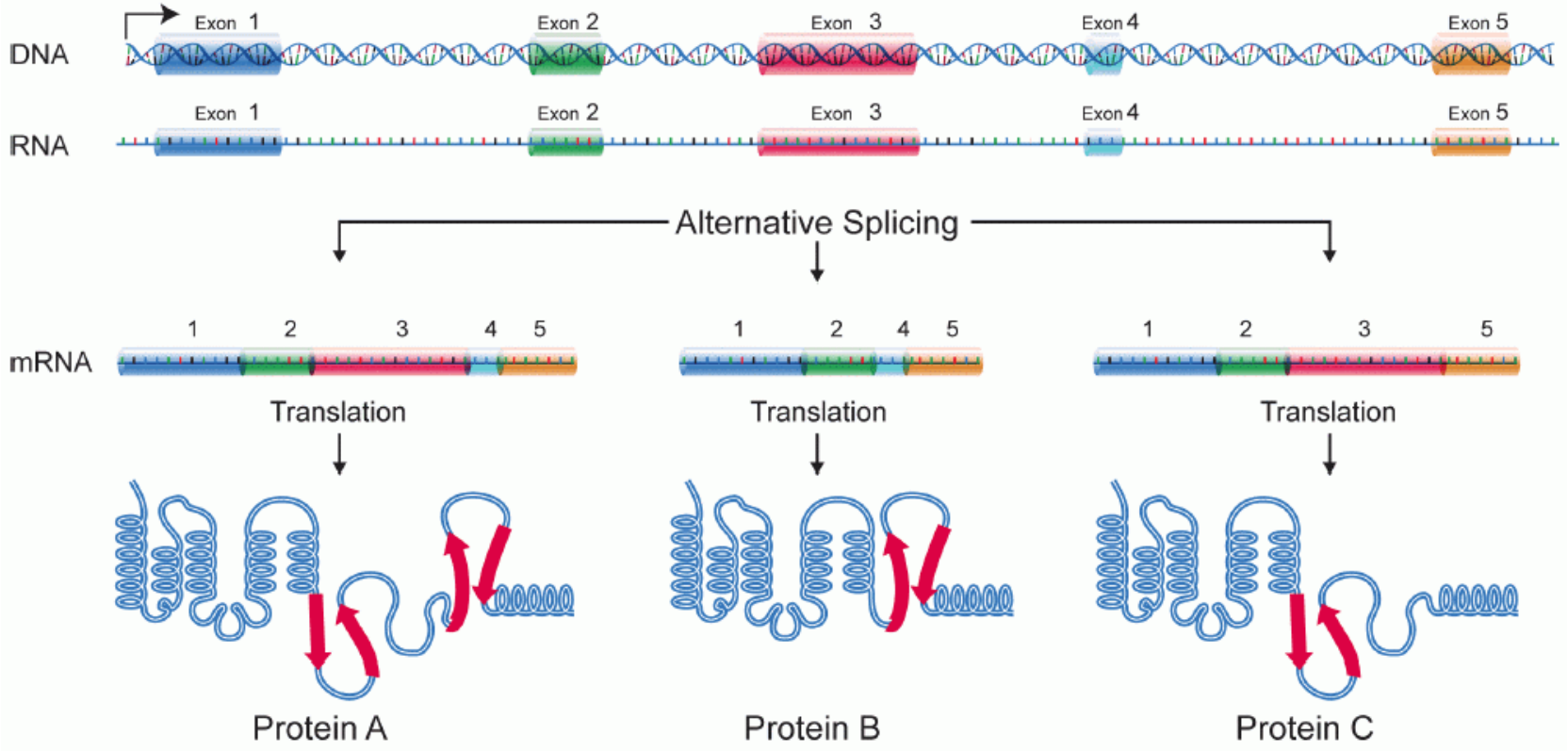
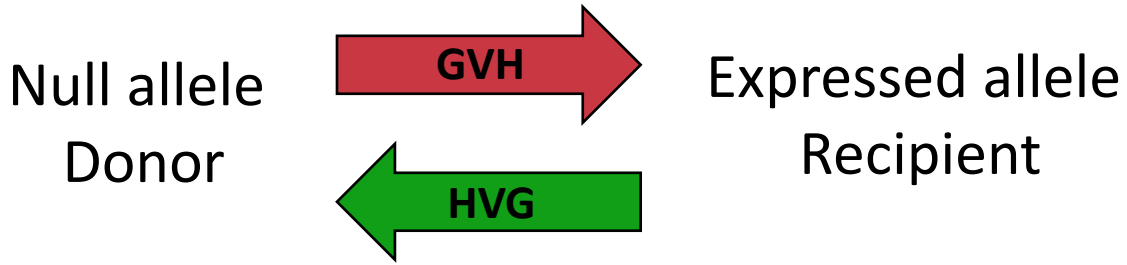


Image credit: "DNA, alternative splicing," by the National Human Genome Research Institute (public domain)

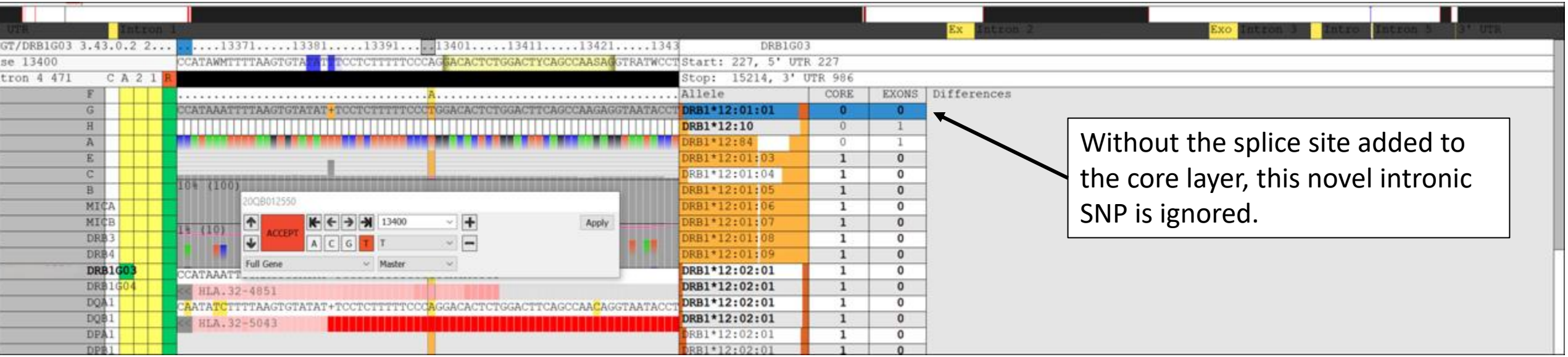
4. Why add splice sites to the core layer analysis in Assign?

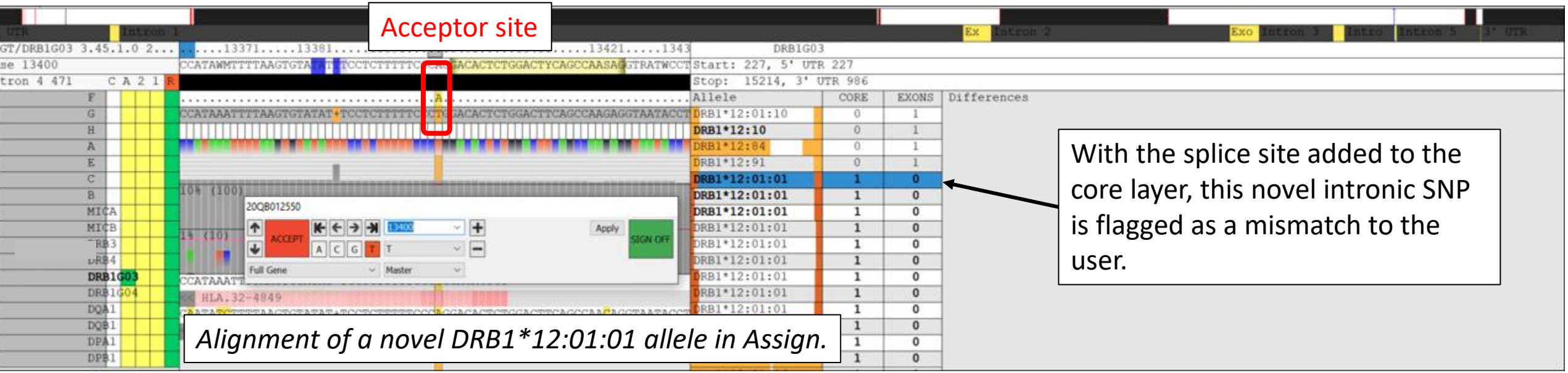
- Non-expressed (null) HLA alleles are relevant in transplant medicine, as they may lead to missing antigen in donor or recipient, resulting in disease.



CareDx®

4.1. Adding the splice sites to the core layer in Assign flags novel polymorphisms that may impact splicing.





*Alignment of a novel DRB1*12:01:01 allele in Assign.*

5. IMGT/HLA characterized splicing variants

- The IMGT/HLA database lists many characterized alleles that have genomic variants in the splice sites resulting in alternative splicing of the mRNA transcript, or complete loss of gene expression.
- A complete list of unexpressed and alternatively expressed alleles characterised in IMGT can be found at: http://hla.alleles.org/alleles/nulls.html

Locus	Unexpressed (N)*	Alternatively expressed (Q, L, S, A)*
HLA-A	5	12
HLA-B	2	8
HLA-C	4	16
HLA-DRB4	3	
HLA-DQA1		1
HLA-DQB1		2
HLA-DPB1		1
HLA-DPA1		1
HLA-G		2
MICA		1

*Counts obtained from IMGT/HLA database version 3.45.1

6. Examples of unexpressed and alternatively expressed HLA alleles

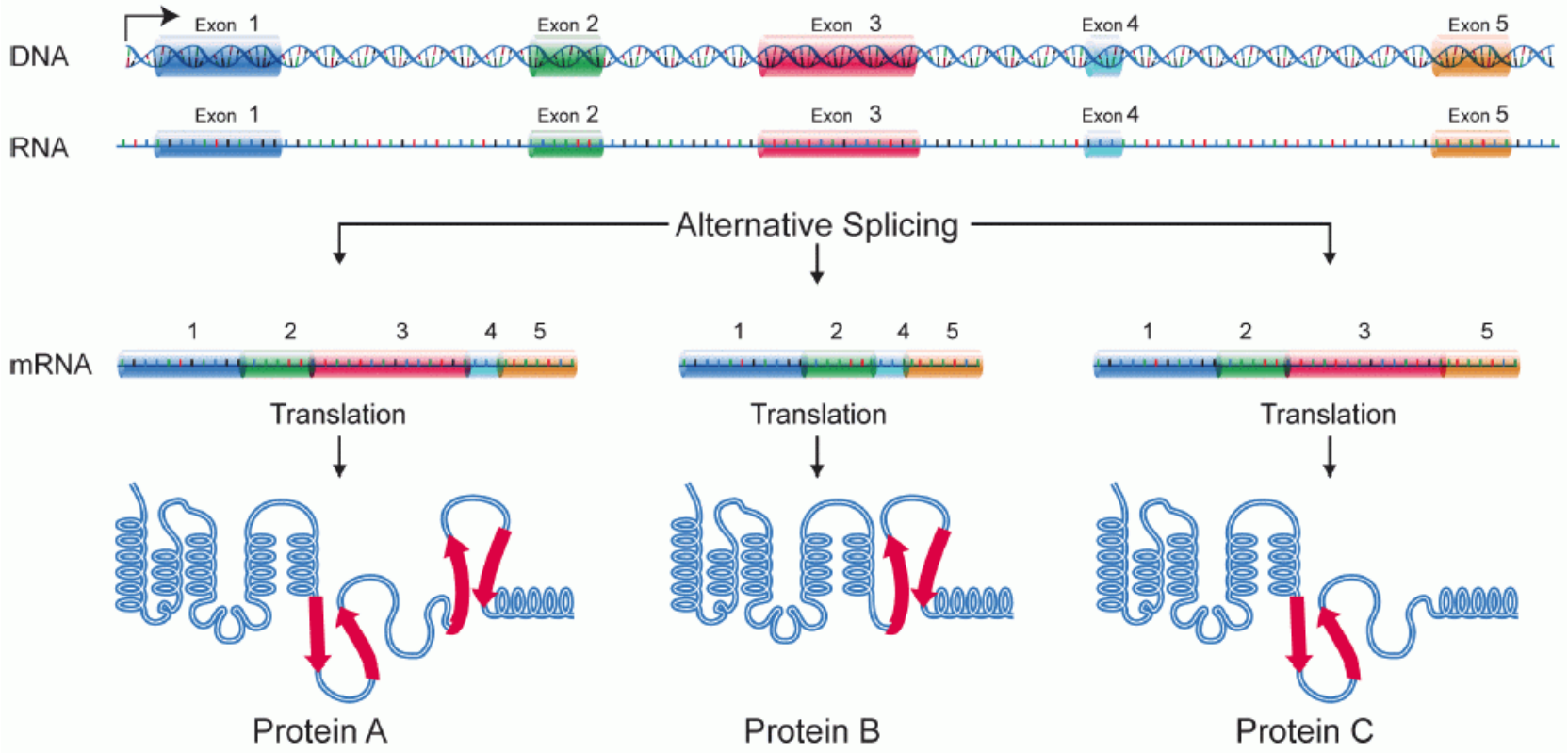
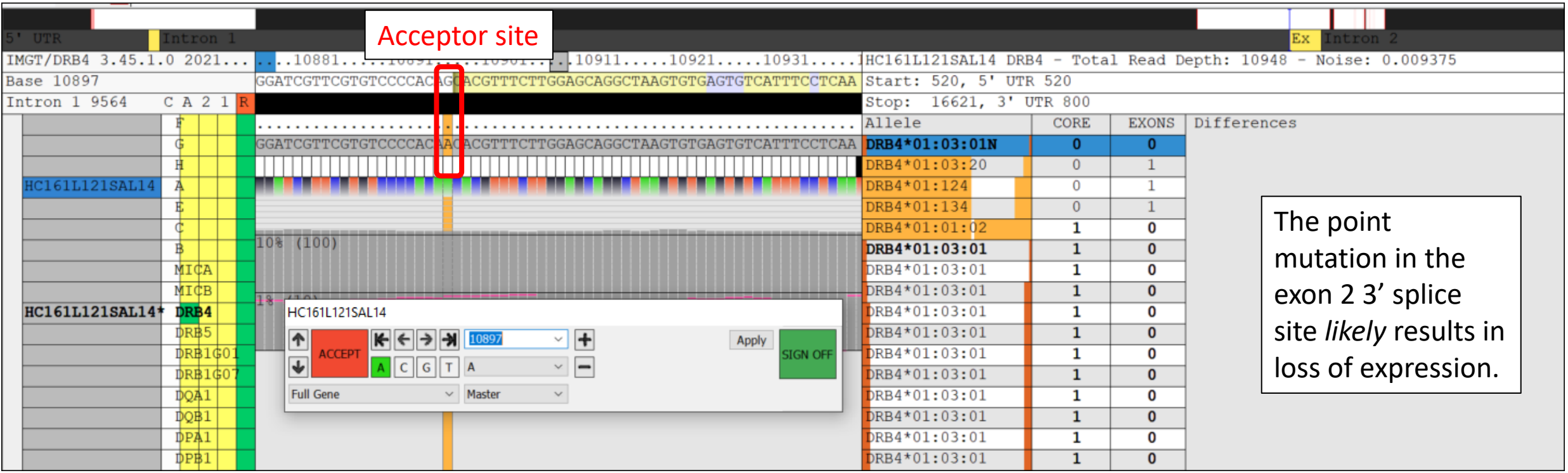


Image credit: "DNA, alternative splicing," by the National Human Genome Research Institute (public domain

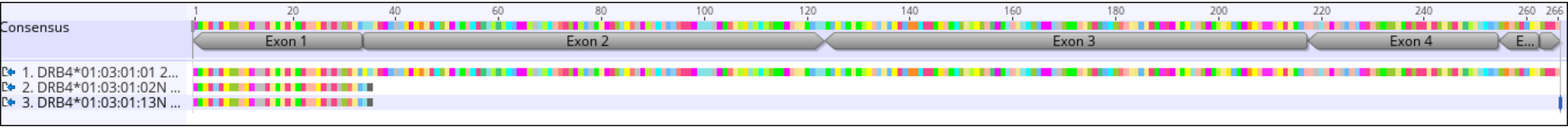
6.1. DRB4*01:03:01N: Example of a null allele

*DRB4*01:03:01:02N + DRB4*01:03:01:13N alleles have this genomic variant.*



The point mutation in the exon 2 3' splice site *likely* results in loss of expression.

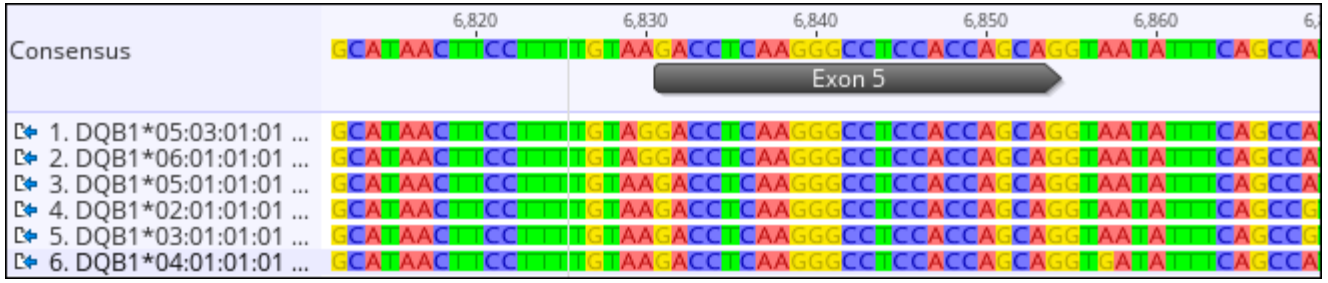
*Protein sequence alignment of DRB4*01:03:01N Geneious Prime® 2022.0.1 highlighting the result of the splice site mutation*



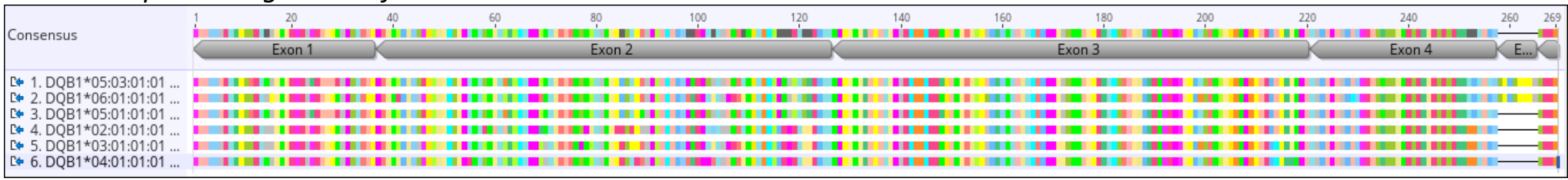
6.2. DQB1 and alternative splicing[5]

- A single base substitution from G>A in the exon 5 3’ splice site is present in the majority of characterized DQB1 alleles.
- This substitution results in the elimination of exon 5 and a protein sequence that is shorter than other class II beta chains by 8 amino acids.
- DQB1*05:03 and DQB1*06:01 are among the alleles that do not have this substitution as shown in the genomic and protein sequence alignments below.

Genomic sequence alignment of DQB1 alleles in Geneious Prime® 2022.0.1



Protein sequence alignment of DQB1 alleles in Geneious Prime® 2022.0.1

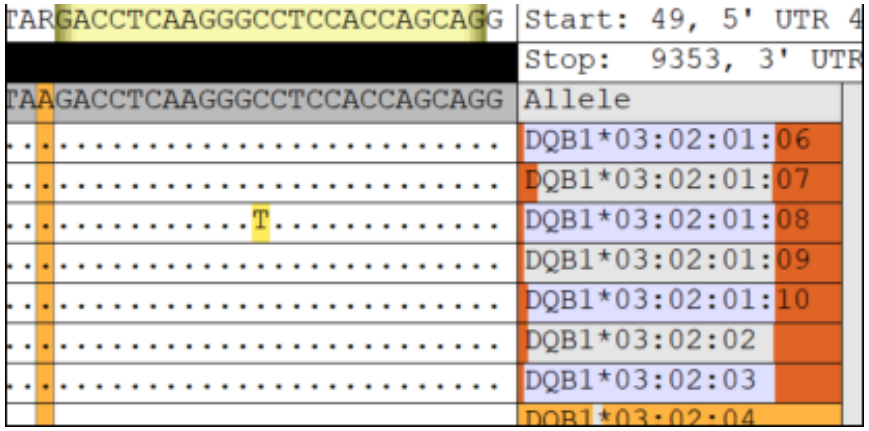


CareDx®

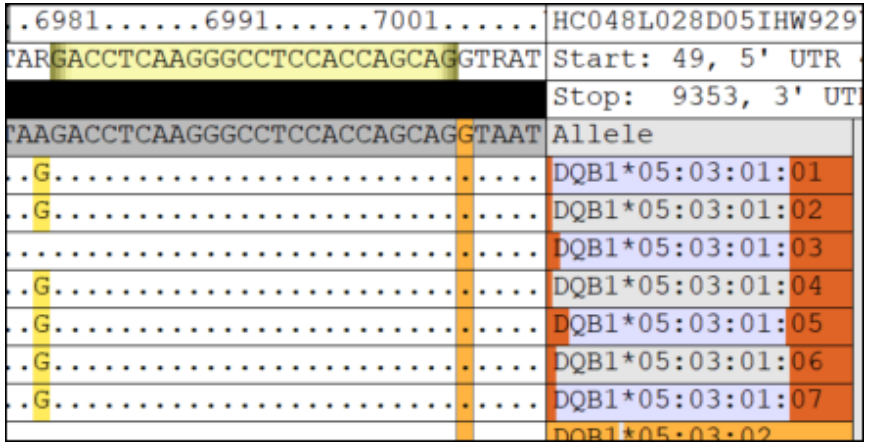
6.3. Impacts of DQB1 non-expressed exon 5 in Assign

- As the nucleotide substitution G>A is present in most alleles, SNPs in exon 5 won't impact expression in many cases.
- SNPs in the exon 5 5' splice site wont impact expression

DQB1*03:02:01:08 differs from the other DQB1*03:02:01 alleles by a single SNP in exon 5. Due to how Assign filters and sorts alleles, this leads to peculiar looking (but correct) results!



Allele 1	Allele 2	CORE	EXONS
DQB1*03:02:01	X	0	0
DQB1*03:02:01	DQB1*03:02:01	0	1
DQB1*03:02:01	DQB1*03:02:01	0	1
DQB1*03:02:01	DQB1*03:02:01	0	1
DQB1*03:02:01	DQB1*03:02:01	0	1
DQB1*03:02:01	DQB1*03:02:01	0	1
DQB1*03:02:01	DQB1*03:02:01	0	1
DQB1*03:02:01	DQB1*03:02:01	0	1
DQB1*03:02:01	X	0	1
DQB1*03:02:01	DQB1*03:02:01	0	1
DQB1*03:02:01	DQB1*03:02:01	0	1



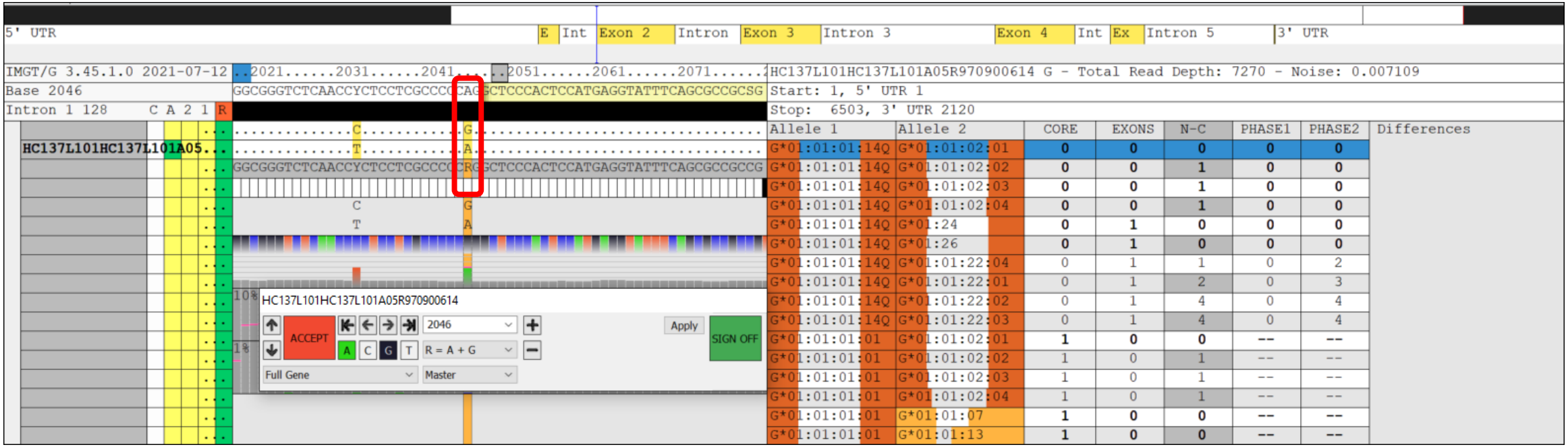
DQB1*05:03:01:03 differs from the other DQB1*05:03:01 alleles by a single SNP in the exon 5 3' splice site. This results in a core layer mismatch in Assign for this allele.

Allele 1	Allele 2	CORE	EXONS
DQB1*05:03:01	X	0	0
DQB1*05:03:01	DQB1*05:03:01	1	0

CareDx®

6.4. G*01:01:01:14Q: alternatively expressed allele due to splice site mutation

*Alignment of G*01:0101:14Q in Assign highlighting the splice site mutation*

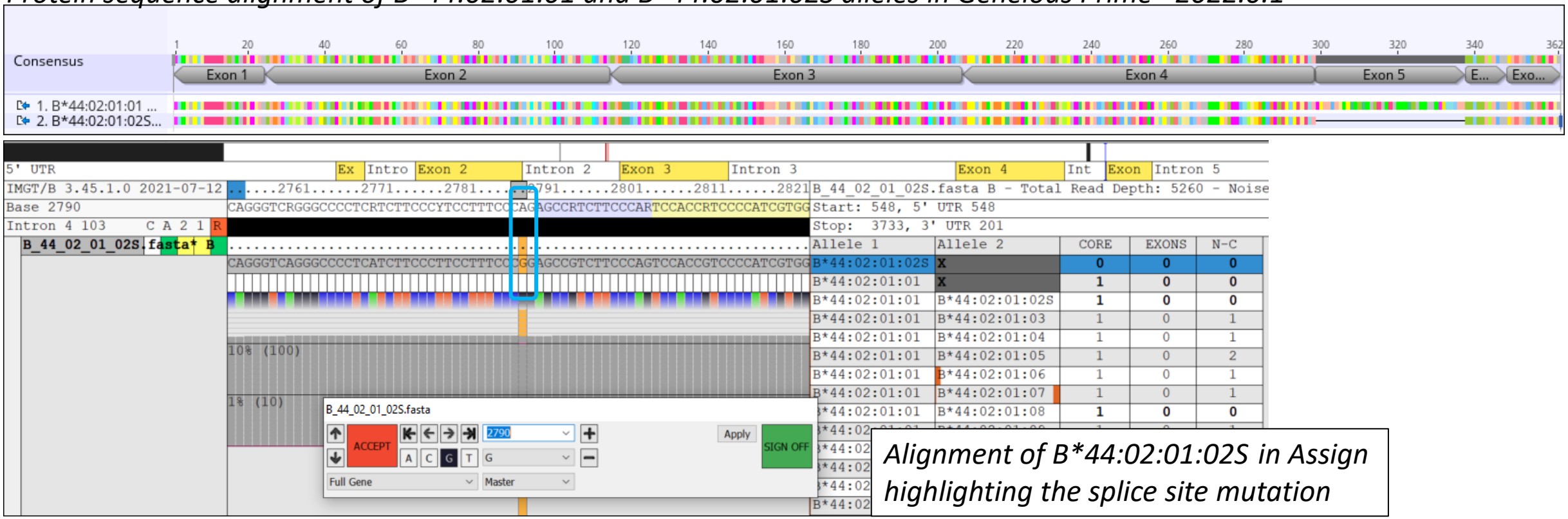


CareDx®

6.5. B*44:02:01:02S[4]: allele with alternative splicing due to splice site mutation.

- B*44:02:01:02S is a soluble form of B44 without any detectable cell-surface expression
- Identified in a volunteer bone marrow donor when it typed as blank by microlymphocytotoxicity
- Absence of B44 protein on the cell surface confirmed by Fluorescence-activated cell sorter analysis
- A>G point mutation in the exon 5 3’ splice site that results in the deletion of exon 5 that encodes the transmembrane domain of the HLA antigen.

*Protein sequence alignment of B*44:02:01:01 and B*44:02:01:02S alleles in Geneious Prime® 2022.0.1*



*Alignment of B*44:02:01:02S in Assign highlighting the splice site mutation*

7. Bibliography

1. The role of gene polymorphism in HLA class I splicing C. E. M. Voorter, K. E. H. Gerritsen, M. Groeneweg, L. Wieten & M. G. J. Tilanus

2. Fuerst, D, Tsamadou, C, Gowdavally, S, Schrezenmeier, H, Mytilineos, J. HLA-DRB4*01:14 is a null allele and renamed HLA-DRB4*01:14N. HLA. 2020; 95: 73– 75. https://doi.org/10.1111/tan.13701

3. Elsner, H.-A. and Blasczyk, R. (2004), Immunogenetics of HLA null alleles: implications for blood stem cell transplantation. Tissue Antigens, 64: 687-695. https://doi.org/10.1111/j.1399-0039.2004.00322.x

4. Dubois, V., Tiercy, J., Labonne, M., Dormoy, A. and Gebuhrer, L. (2004), A new HLA-B44 allele (B*44020102S) with a splicing mutation leading to a complete deletion of exon 5. Tissue Antigens, 63: 173-180. https://doi.org/10.1111/j.1399-0039.2004.00134.x

5. Senju, S., Kimura, A., Yasunami, M. et al. Allele-specific expression of the cytoplasmic exon of HLA-DQB1 gene. Immunogenetics 36, 319–325 (1992). https://doi.org/10.1007/BF00215661

6. Pagani, F., Baralle, F. Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5, 389–396 (2004). https://doi.org/10.1038/nrg1327

7. Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor perspectives in biology*, *3*(7), a003707. https://doi.org/10.1101/cshperspect.a003707

8. Immunogenetics 31: 112-117, 1990 Brief communications I muno- gene 'cs © Springer-Verlag 1990 An aberrant DRB4 null gene transcript is found that could encode a novel HLA-DR/ chain Vivien R. Sutton* and Robert W. Knowles Cell Surface Immunochemistry Laboratory, Sloan-Kettering Institute for Cancer Research, New York, NY 10021, USA Received September 28, 1989; revised version received November 2, 1989

CareDx®

Appendix

A.1. Unexpressed Alleles with Splice site mutations characterized in IMGT (3.45.1)

Allele	Mutation	Location	Consequence
A*01:11N	Point	Exon 3, 968G>T	results in an alternative splice site at the end of exon 3 which prevents translation into a correct and stable class I molecule expressed on the cell surface
A*03:01:01:02N	point	in intron 4 g1846G>A	causes incorrect splicing and premature stop codon.
A*23:19N	point	Exon 3, 619G>A	The mutation occurs at exon boundary, potentially affecting the splice site. This allele shown to be non-expressed.
A*26:01:01:03N	point	Intron 4, g1846G>A,	causes incorrect splicing and premature stop codon.
A*29:01:01:02N	point	Intron 4, g1846G>T	causes incorrect splicing and premature stop codon.
A*31:01:02:03N	deletion	Intron 1, g176-185delGCGGATCTCA	affects splice site for intron 1.
B*07:44N	point	Exon 4, 852T>G	causes an alternative splice site at the end of exon 4 which causes translation of intron 4 sequence and an abnormal truncated peptide
B*15:01:01:02N	deletion	Intron 1, g175-184delCGGGTCTCAG	affects splice site for exon 2.

CareDx®

A.2. Unexpressed Alleles with Splice site mutations characterized in IMGT (3.45.1)

Allele	Mutation	Location	Consequence
C*03:03:01:50N	Point	Intron 2, 718A>G	causes a mutation in the splice site preceding exon 3
C*03:03:01:52N	Point	Intron 1, 203G>A	affects the splice site for intron 2
C*03:23N	Point	Exon 3, 406G>A	Affects the splice site for intron 2, which causes a frameshift and premature stop codon
C*07:02:01:17N	Point	Intron 3, g710T>A	Causes incorrect splicing and premature stop codon
C*15:02:01:08N	Point	Intron 2, g431A>T	causes an incorrect splicing leading to the deletion of part of exon 3
DPB1*04:01:01:24N	Point	Intron 2, g5163G>T	Affects splicing site for exon 2
DRB4*01:03:01:02N	Point	Intron 1, g9656G>A	results in Incorrect splicing and lack of protein sequence
DRB4*01:03:01:13N	Point	Intron 1, g9656G>A	Causes a mutation in the splice site prior to exon 2, which may affect expression
DRB4*01:14N	Point	Intron 1, g9656G>A	causes incorrect splicing which results in lack of protein sequence

CareDx®

A.3. Alternatively expressed Alleles with Splice site mutations characterized in IMGT (3.45.1)

Allele	Mutation	Location	Consequence
A*01:01:38L	Point	Exon 4, 703-705GCG>GCA	causes an aberrant dominant splice site which results in low expression
A*01:301Q	Point	Intron 2, g708G>A	causes a mutation in the splice site prior to exon 3
A*02:01:01:134Q	Point	Intron 2, g475T>C	Causes a mutation in the splice site proceeding exon 2, which may affect expression.
A*02:01:14Q	Point	Exon 4, 703-705GCG>GCA	Causes an aberrant dominant splice site, which may affect expression
A*24:02:01:02L	Point	Intron 2, g708G>A,	Causes a mutation in the splice site prior to exon 3
A*24:02:01:17Q	Point	Intron 7, g2897A>C	causes a mutation in the splice site prior to exon 8, which may affect expression
A*24:02:03Q	Point	Exon 4, 703-705GCG>GCA	Causes an aberrant dominant splice site, which may affect expression
A*24:447Q	Point	Intron 2, g475T>C	Causes a mutation in the splice site proceeding exon 2
A*24:450Q	Point	Intron 2, g708G>A	Causes a mutation in the splice site prior to exon 3
A*31:01:02:30Q	Point	Intron 7, g2730G>T	Causes a mutation in the splice site proceeding exon 7, which may affect expression
A*33:03:03Q	Point	Exon 4, 703-705GCG>GCA	Causes an aberrant dominant splice site, which may affect expression

CareDx®

A.4. Alternatively expressed Alleles with Splice site mutations characterized in IMGT (3.45.1)

Allele	Mutation	Location	Consequence
B*15:01:01:39Q	Point	Intron 1, g201G>C	causes a mutation in the splice site preceding exon 2, which may affect expression
B*18:01:01:12Q	Point	Intron 2, g716G>C	causes a mutation in the splice site prior to exon 3, which may affect expression
B*18:01:01:42Q	Point	Intron 2, g715A>G,	causes a mutation in the splice site prior to exon 3, which may affect expression.
B*27:05:02:04Q	Point	Intron 2, g716G>A	affects splice site for exon 3, which may affect expression
B*38:01:01:03Q	Deletion	Intron 2, g472delG	affects splice site for exon 2, which may affect expression
B*44:02:01:02S	Point	Intron 4, g1934A>G	causes an incorrect splicing leading to the deletion of exon 5
B*44:02:01:13Q	Point	Intron 3, g994T>C	causes a mutation in the splice site proceeding exon 3, which may affect expression.
B*56:01:01:05S	Point	Intron 4, g1935A>G	causes an incorrect splicing leading to the deletion of exon 5

A.5. Alternatively expressed Alleles with Splice site mutations characterized in IMGT (3.45.1)

Allele	Mutation	Location	Consequence
C*02:02:02:34Q	Point	Intron 3, g996G>T	causes a mutation in the splice site after exon 3
C*03:04:01:35Q	Point	Intron 3, g996G>A	causes a mutation in the splice site after exon 3
C*04:01:01:47Q	Point	Intron 4, g1860T>G	causes a mutation in the splice site after exon 4
C*04:01:01:84Q	Point	Intron 5, g2538G>C	causes a mutation in the splice site prior to exon 6, which may affect expression
C*05:01:01:53Q	Point	Intron 1, g203G>A	causes a mutation in the splice site prior to exon 2, which may affect expression
C*07:01:01:100Q	Point	Intron 2, 718A>G	causes a mutation in the splice site preceding exon 3, which may affect expression
C*07:01:01:14Q	Point	Intron 1, g203C>G	affecting the splice site for exon 2, which may affect expression
C*07:02:01:124Q	Point	Intron 6, g2679G>A	causes a mutation in the splice sites at the end of intron 6, which may affect expression

CareDx®

A.6. Alternatively expressed Alleles with Splice site mutations characterized in IMGT (3.45.1)

Allele	Mutation	Location	Consequence
C*07:02:01:125Q	Point	Intron 3, 1002T>C	causes a mutation in splice site proceeding exon 2, which may affect expression.
C*07:02:01:74Q	Point	Intron 2, g2727G>A	causes a mutation in the splice site after to exon 7
C*07:04:01:15Q	Point	Intron 2, 718A>G	causes a mutation in the splice site preceding exon 3, which may affect expression
C*07:04:01:16Q	Point	Intron 1, 203G>C	causes a mutation in the splice site prior to exon 2
C*07:06:01:05Q	Point	Intron 5, g2537G>A	causes a mutation in the splice site prior to exon 6
C*15:02:01:30Q	Point	Intron 1, g75T>C	causes a mutation in the splice site proceeding exon 1, which may affect expression
C*15:02:01:35Q	Point	Intron 2, 474G>C	causes mutation in splice site proceeding exon 2, which may affect expression
C*17:01:01:16Q	Point	Intron 2, 718A>G	causes a mutation in the splice site preceding exon 3, which may affect expression

CareDx®

A.7. Alternatively expressed Alleles with Splice site mutations characterized in IMGT (3.45.1)

Allele	Mutation	Location	Consequence
G*01:01:01:14Q	Point	in Intron 1, g201A>G	causes a mutation in the splice site preceding exon 2, which may affect expression
G*01:04:01:04Q	Point	Intron 4, g1969A>G	causes a mutation in the splice site preceding exon 5, which may affect expression
MICA*002:01:13Q	Point	Intron 4, g8696	causes a mutation in the splice site, which may affect splicing