Received: 10 November 2021 | Revised: 6 April 2022 | Accepted: 28 May 2022

DOI: 10.1111/tan.14695

ORIGINAL ARTICLE

HLA Immune Response Genetics WILEY

Polymorphic differences within HLA-C alleles contribute to alternatively spliced transcripts lacking exon 5

Femke A. I. Ehlers[1,2,3] | **Timo I. Olieslagers**[1,3] | **Mathijs Groeneweg**[1,3] | **Gerard M. J. Bos**[2,3] | **Marcel G. J. Tilanus**[1,3] | **Christina E. M. Voorter**[1,3] | **Lotte Wieten**[1,3]

[1]Department of Transplantation Immunology, Tissue Typing Laboratory, Maastricht University Medical Center+, Maastricht, The Netherlands

[2]Department of Internal Medicine, Division of Tumor Immunology, Maastricht University Medical Center+, Maastricht, The Netherlands

[3]GROW—School for Oncology and Developmental Biology, Maastricht University Medical Center+, Maastricht, The Netherlands

Correspondence
Lotte Wieten, Department of Transplantation Immunology, Tissue Typing Laboratory, Maastricht University Medical Center+, Maastricht, The Netherlands.
Email: l.wieten@mumc.nl

The HLA genes are amongst the most polymorphic in the human genome. Alternative splicing could add an extra layer of complexity, but has not been studied extensively. Here, we applied an RNA based approach to study the influence of allele polymorphism on alternative splicing of HLA-C in peripheral blood. RNA was isolated from these peripheral cells, converted into cDNA and amplified specifically for 12 common HLA-C allele groups. Through subsequent sequencing of HLA-C, we observed alternative splicing variants of *HLA-C*04* and **16* that resulted in exon 5 skipping and were co-expressed with the mature transcript. Investigation of intron 4 sequences of *HLA-C*04* and **16* compared with other HLA-C alleles demonstrated no effect on predicted splice sites and branch point. To further investigate if the unique polymorphic positions in exon 5 of *HLA-C*04* or **16* may facilitate alternative splicing by acting on splicing regulatory elements (SRE), in-silico splicing analysis was performed. While the *HLA-C*04* specific SNP in exon 5 had no effect on predicted exonic SRE, the *HLA-C*16* specific exon 5 SNP did alter exonic SRE. Our findings provide experimental and theoretical support for the concept that polymorphisms within the HLA-C alleles influence the alternative splicing of HLA-C.

KEYWORDS
alternative splicing, exon skipping, HLA-C, splicing regulatory elements

Abbreviations: ESE, exonic splicing enhancer; ESRE, exonic splicing regulatory element; ESS, exonic splicing silencer; HSF, human splicing finder; ISE, intronic splicing enhancers; ISRE, intronic splicing regulatory element; ISS, intronic splicing silencers; KIR, killer-cell immunoglobulin-like receptor; NI, neighborhood inference; NK cell, natural killer cell; PBMC, peripheral blood mononuclear cell; pre-mRNA, precusor mRNA; SNP, single nucleotide polymorphism; snRNP, small nuclear ribonucleoprotein; SRE, splicing regulatory element.

Femke A. I. Ehlers and Timo I. Olieslagers are shared first authors.

1 | INTRODUCTION

HLA-C belongs to the classical HLA class I molecules and its cell-surface expression is generally about 10-fold lower than expression of HLA-A and HLA-B molecules.[1] Despite its relatively low expression, HLA-C surface expression is critical for immunological reactions, for example in the setting of kidney- or stem cell

transplantation where an HLA-C mismatch between patient and donor is associated with acute rejection and graft-versus-host disease, respectively.[2–5]

The genes encoding classical HLA class I molecules including HLA-C are notoriously polymorphic and single nucleotide polymorphism (SNP) in the HLA-C genes have been shown to impact HLA-C expression levels: For example, polymorphisms in the transcription factor binding sites have been associated with differences in HLA-C promoter activity and cell type- or tissue specific expression.[6] In addition, a SNP located 35kB upstream of the HLA-C gene (-35C/T) has been associated with high HLA-C mRNA and cell-surface expression levels.[7] In a follow up study, evidence was acquired showing that this -35 SNP, presumably, marks a functional 3′UTR insertion/deletion variant influencing HLA-C expression levels through its effects on binding of microRNA-148.[8]

Alternative splicing may be another mechanism leading to differences in HLA cell surface expression as it may impact protein expression both quantitatively and qualitatively. During alternative splicing of precursor mRNA (pre-mRNA), several alternative mRNA isoforms can be formed and the resulting proteins can be expressed at different levels on the cell surface or appear as soluble proteins.[9] Alternative splicing can take place in parallel to mature splicing and the alternative variants can be co-expressed with the normal variant.[10] The splicing process is catalyzed by small nuclear ribonucleoproteins (snRNPs), which together form the spliceosome, and splicing is initiated when snRNPs attach to the four regions of the pre-mRNA involved in the splicing process: the 3′ splice site, the branch point, the 5′ splice site and the polypyrimidine tract.[11] Splicing activity can further be regulated by SREs. These SREs are found in both intron and exon regions and are classified as splicing enhancers or splicing silencers. Splicing enhancer elements are recognized by specific splicing activators that enhance exon inclusion, while splicing silencer elements are recognized by specific repressors that inhibit exon inclusion.[11] Nucleotide variation in these cis elements may result in alternative splicing such as exon skipping, intron retention, alternative 5′ splice sites, alternative 3′ splice sites and mutually exclusive exons.[12]

Alternative splicing has been estimated to occur in more than 90% of the human genes.[13] For HLA genes, the number of alternative transcripts is probably strongly underestimated because the genes have been primarily studied by analysis of genomic DNA and not by mRNA analysis. So far, alternative splicing of HLA is best known for the non-classical HLA-G molecule. Seven variants of HLA-G have been identified (HLA-G1 to -G7), of which only HLA-G1 has the typical beta-2 microglobulin associated cell membrane appearance. HLA-G5 is the soluble counterpart of HLA-G1 as a result of a stop codon leading to exclusion of exon 5 encoding the transmembrane region.[10,14,15] Both HLA-G1 and -G5 are functional proteins that can exert inhibitory functions on T and NK cells.[16]

For classical HLA class I molecules, a limited number of alternatively spliced variants have been observed. Two studies identified soluble HLA-A and HLA-B molecules in human plasma and cell line supernatant, which was suspected to occur because of alternative splicing of exon 5.[17,18] Other studies also reported alternative splicing in HLA-A and -B, one of them showing that a common alternative splice site is contained in all alleles of *HLA-A*11*, resulting in co-expression of a mature and alternative transcript.[19] This alternative transcript is the result of a unique polymorphism in intron 5 and encodes a protein with 6 additional amino acids in the cytoplasmic tail.[19] The *HLA-B*44:02:01:02S* allele has been described as a soluble form of HLA-B44 because of a substitution of a single basepair (A>G) at the end of intron 4 that led to skipping of the complete exon 5.[20]

Also for HLA-C, alternative splicing has been observed: Already in 1989, Cianetti et al observed in two different clones from a cDNA library from a SV40 transformed human fibroblast line (GM637) two alternative ways of splicing, since both clones had identical HLA-C sequences (later identified as *HLA-C*16:01:01:01*), but one clone was devoid of exon 5.[21] Moreover, Vilches et al have described the occurrence of *HLA-C*12* and **15* variants lacking exon 5 in HLA-C specific clones prepared from cDNA of B-LCL from two individuals from a cohort of Spanish gypsies.[22] Yang et al identified an *HLA-C*02* variant lacking exon 5 in RPMI1788 cells secreting soluble HLA molecules in their culture supernatants.[23] They used a targeted approach to specifically amplify the alternative splice product since the quantities of alternatively spliced products may be much lower than that of the mature mRNA variants which may lead to poor amplification in the PCR reaction of the rare variant.[23] More recently by using an RNA-based sequencing approach, we identified an alternative transcript that includes 18 additional nucleotides from intron 5 for the *HLA-C*03* allele group which may result from a SNP unique to *HLA-C*03* at position 2499 within intron 5, that may cause an alternative branch point site and thus an alternative transcript.[10] In addition, a NK cell-specific promoter in the HLA-C gene has been identified, which resulted in many alternatively spliced HLA-C mRNAs in NK cells but not in other cells.[24] The alternative transcripts had varying ability to be translated into proteins and allele-specific differences in the 5′ untranslated region could further modulate expression levels of HLA-C in NK cells, which were functionally relevant for regulating NK cell activity.[24,25]

In the present study, we used a full-length sequencing approach to study possible additional alternative splice

variants of HLA-C in peripheral blood cells covering all the common allele groups, except for *HLA-C*17* and **18*. By using this methodology, we observed an alternatively spliced variant of HLA-C lacking exon 5 in samples with *HLA-C*04* or **16*, that was co-expressed with the normal mature variant. We followed up on this observation by an in-silico analysis to study the impact of the SNPs distinguishing *HLA-C*04* or **16* from the other HLA-C alleles on the strength of the intronic and exonic SREs.

2 | MATERIALS AND METHODS

2.1 | RNA panel

A panel of RNA samples with known HLA-C high resolution typing was used for alternative splicing analysis. The panel covered 12 of the 14 HLA-C allele groups with one to six samples per allele group (Table 1). Samples with *HLA-C*17* and **18* alleles were not available and therefore not included in the panel.

TABLE 1 RNA panel covering the 12 common HLA-C allele groups

HLA-C allele group	Allele	Number of samples	Alternative splicing of exon 5
**01*	01:02:01	1	No
**02*	02:02:02	3	No
**03*	03:03:01	3	No
	03:04:01	1	No
**04*	04:01:01	6	Yes (6 of 6)
**05*	05:01:01	5	No
**06*	06:02:01	2	No
**07*	07:01:01	6	No
	07:02:01	5	No
	07:04:01	1	No
**08*	08:01:01	3	No
	08:02:01	1	No
**12*	12:03:01	3	No
**14*	14:02:01	2	No
**15*	15:02:01	4	No
**16*	16:01:01	3	Yes (3 of 3)
**17*	not available	0	unknown
**18*	not available	0	unknown

Note: cDNA was amplified from RNA samples with HLA-C allele-specific primers to investigate alternative splicing of exon 5 in all allele groups.

2.2 | cDNA synthesis and PCR amplification

RNA was purified and isolated from peripheral blood samples with the RNeasy mini kit (Qiagen, Hilden, Germany), followed by a Dnase treatment (Thermo Fisher Scientific, Waltham, USA). cDNA was synthesized from 8 µl RNA using the Superscript III first-strand synthesis system for reverse transcriptase PCR according to the manufacturer's protocol (Thermo Fisher Scientific). Amplification of cDNA was performed using gene-specific primers for HLA-C or allele-group specific primers for each HLA-C allele group (Table 2). For PCR reactions, 3 µl of cDNA were used in a total volume of 30 µl. The PCR reaction mix consisted of 16.6 mM ammonium sulphate (Merck, Darmstadt, Germany), 0.1 µg/µl cresol red (Sigma-Aldrich, St. Louis, USA), 0.2 mM of each dNTP (GE Healthcare, Diegem, Belgium), 1.4 U Expand High Fidelity Enzyme Mix (Roche, Basel, Switzerland), 5% Glycerol (Alfa-Aesar, Karlsruhe, Germany), 1.5 mM $MgCl_2$ (Thermo Fisher Scientific), 67 mM Tris–HCl with pH 8.8, 0.01% Tween 20 (both from Merck, Darmstadt, Germany), and 15 pmol of each primer (Sigma-Adrich, St. Louis, USA).

The initial denaturation step was performed at 94°C for 2 min, followed by 10 cycles of 15 sec at 94°C, 30 sec at 63°C, 4 min at 68°C; 10 cycles of 15 sec at 94°C, 30 sec at 60°C, 6 min at 68°C; 10 cycles of 15 sec at 94°C, 30 sec at 60°C, 10 min at 68°C, and the final elongation step was performed for 7 min at 68°C. After the PCR products were checked by electrophoresis on a 1.5% agarose gel, the amplicons were purified using ExoSAP-IT according to the manufacturer's protocol (Affymetrix, Santa Clara, USA).

2.3 | Low melting point agarose gel

The PCR products were separated on a 0.6% low melting point agarose gel (Thermo Fisher Scientific) at 50 V for 6.5 h. The PCR fragments were purified from the low melting gel using the MinElute gel extraction kit (Qiagen) and were used for sequencing.

2.4 | RNA-based Sanger sequencing

Samples, either PCR products or gel extracted PCR fragments, were sequenced with forward and reverse primers as described in Table 2 using the ABI BigDye Terminator Chemistry (Thermo Fisher Scientific). The sequencing reaction contained 1 µl purified amplification product, 0.5 µl sequencing primer (5 pmol, Sigma-Aldrich), 1 µl of BigDye Terminator v1.1 mix, 1.5 µl 5x BigDye Terminator sequencing buffer and 6 µl water. Cycle sequencing was

TABLE 2 Overview of primers used for allele-specific amplification (A), gene-specific amplification (B) and primers used for sequencing of allele-specific products (C)

Allele group	Direction	Primer sequence (5′-3′)	Location	Position cDNA	Position gDNA
(A)					
01	FW	GGGTCTCACACCCTCCAGT	Exon 3	I2-361	719-737
02	FW	AGCACGAGGGGCTGCCG	Exon 4	857-873	1820-1836
03	FW	AGGACCAAACTCAGGACACT	Exon 4	737-756	1700-1719
04	FW	CTCCGCGGGTATAACCAGTT	Exon 3	400-419	776-795
05/06/08	FW	AGCACGAGGGGCTGCCA	Exon 4	857-873	1820-1836
07	FW	GCAGAGATACACGTGCCATA	Exon 4	834-853	1797-1816
12	FW	CGTGAGGCGGAGCAGTG	Exon 3	523-539	900-915
14	FW	GGGTCTCACACCCTCCAGT	Exon 4	I2-361	719-737
15	FW	CGCCTCCTCCGCGGGC	Exon 3	394-409	770-785
16	FW	CGTGCGGCGGAGCAGCA	Exon 3	523-539	900-915
all	REV	CACAGGCAGCTGTCTCAGG	3′UTR	1097-1115	2891-2909
(B)					
all	FW	GAGCTTGTGGAGACCAGG	Exon 4	757-774	1720-1737
all	REV	GCATCTCAGTCCCACACAG	3′UTR	1111-1128	2905-2923
(C)					
15	FW	CGCCTCCTCCGCGGGC	Exon 3	394-409	770-785
01/04/14	FW	GACGGCAAGGATTACA	Exon 3	427-442	803-818
12/16	FW	CGTGCGGCGGAGCAGCA	Exon 3	523-539	900-915
03	FW	AGGACCAAACTCAGGACACT	Exon 4	737-756	1700-1719
07	FW	GCAGAGATACACGTGCCATA	Exon 4	834-853	1797-1816
05/06/08	FW	AGCACGAGGGGCTGCCA	Exon 4	857-873	1820-1836
02	FW	AGCACGAGGGGCTGCCG	Exon 4	857-873	1820-1836
all	REV	AGAGCAGCTCCCTCCTTTTC	Exon 6	1019-1038	2543-2561

performed as follows: 1 min at 96°C; followed by 25 cycles of 10 s at 96°C, 5 s at 50°C and 4 min at 60°C. Sequencing reactions were purified by Sephadex G-50 Fine (GE Healthcare Life Sciences, Little Chalfont, UK) and analyzed on the ABI3730 DNA analyzer (Applied Biosystems, Foster City, USA). DNASTAR Lasergene software was used for sequence analysis (DNASTAR Lasergene 12 Core Suite, Madison, USA).

2.5 | In-silico splicing analysis

The online available Human Splicing Finder (HSF) tool was used (HSF pro system, https://www.genomnis.com/, version 2.23.6[26]) for splicing analysis. This tool combines multiple algorithms to identify and predict sequence effects on splicing motifs, including splice sites and branch point sites. In this study, the HSF and the MaxEnt prediction algorithms were used to calculate the branch point and splice site strength of the different HLA-C alleles.

Additional to the HSF tool, the EX-SKIP tool (http://ex-skip.img.cas.cz/[27]) was used to predict exon skipping of HLA-C alleles by comparing ESE/ESS motif differences. EX-SKIP analyses exon sequences using integrated algorithms to identify the number of 8-mer ESEs and ESSs (PESE/PESS[28]), ESS decamers (FAS-ESS[29]), exon- and intron-identity elements (EIE/IIE[30]), RESCUE-ESE hexamers (RESCUE-ESE[31]) and neighborhood inference (NI-ESE/NI-ESS[32]). The ESS/ESE ratio is determined based on the total ESE and ESS count; a higher ratio is predicted to increase chance to skip an exon.

3 | RESULTS

3.1 | Alternative splicing of exon 5 in HLA-C alleles

To investigate alternative splicing in HLA-C, mRNA was transcribed into cDNA which was subsequently amplified

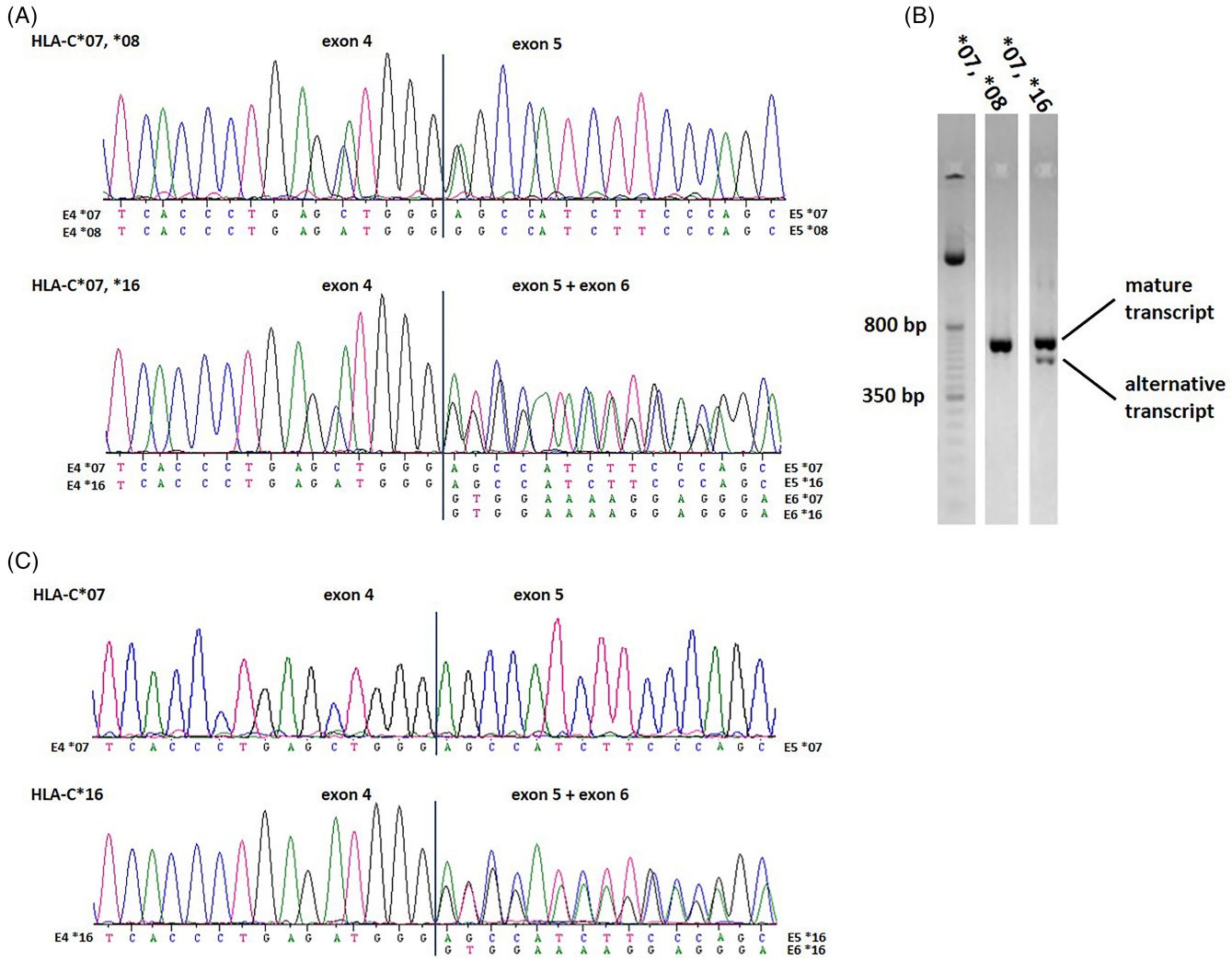

FIGURE 1 Identification of alternative HLA-C transcript missing exon 5. Two cDNA samples either with the *HLA-C*07:02:01* and **08:01:01* alleles or *HLA-C*07:01:01* and **16:01:01* was amplified with HLA-C gene- and allele-specific primers. A) Gene-specific sequence results of both samples. (B) Electrophoresis result of gene-specific PCR product of both samples. (C) Allele-specific sequence results of a sample with *HLA-C*07:01:01* and **16:01:01*

using HLA-C gene-specific primers. By sequencing the amplified product covering exon 4 until 3′UTR, we observed an alternatively spliced variant of HLA-C in some samples and this alternative transcript lacked exon 5. For example, a sample with an *HLA-C*07:02:01* and **08:01:01* typing did not express the alternatively spliced variant and the sequences matched the expected sequences of *HLA-C*07:02:01* and **08:01:01* (Figure 1A). In contrast, a sample with an *HLA-C*07:01:01* and **16:01:01* typing expressed both the mature and the alternative transcript, revealed by the presence of two PCR products of 716 bp and 596 bp (Figure 1B) and the overlapping peaks in the sequences after exon 4, with one peak representing the exon 5 sequence and the other peak the exon 6 sequence (Figure 1A). By using allele-specific amplification primers we determined in which of the two alleles exon 5 skipping was detected. In the sequence of *HLA-C*07:01:01* no alternative transcript was observed, but in the sequence of *HLA-C*16:01:01* both the mature and the alternative transcript were present, represented by the double peaks after exon 4 (Figure 1C).

3.2 | Evidence for the role of HLA polymorphism in alternative splicing of HLA-C

To examine whether alternative splicing differs between HLA-C allele groups, we used the samples from our RNA

panel, which covered 12 of the 14 HLA-C allele groups (Table 1). cDNA from each sample was amplified with HLA-C allele-specific primers to detect alternative splicing in both alleles individually. We observed the alternatively spliced variant by gel electrophoresis and/or sequencing in all samples with *HLA-C*04* and **16* alleles, but not in any of the other HLA-C allele groups (Table 1). In all samples where the alternative variant was present, it was co-expressed with the mature full-length transcript (data not shown).

To confirm that exon 5 was skipped in the alternative variants, the amplified products of HLA-*C*04:01:01* and **16:01:01* were run on a low melting gel to separate the alternatively spliced product from the mature product. The alternative product was isolated from the gel and sequenced. For both *HLA-C*04:01:01* and **16:01:01*, the sequencing results of the alternative transcript showed a single sequence, in which exon 4 was directly followed by the exon 6 sequence (Figure 2). These results confirmed that the alternative product was the result of exon 5 skipping. The reading frame was maintained in the alternative transcript since exon 5 is 120 nucleotides long.

3.3 | In-silico analysis of the effect of intronic HLA-C variation on the branch point and splice site strength

Splicing specificity is mainly determined by the 5′ splice site (donor), the branch point and 3′ splice site (acceptor) sequences in the intron regions.[10] SNPs in the splice sites downstream and upstream exon 5 could potentially affect the splicing process by alteration of the complementary U1 snRNA binding sequence at the 5′ splice site or the complementary U2AF1 snRNA binding sequence at the 3′ splice site. SNPs introducing potentially new branch point sequences or branch point breaks might affect splicing via the attachment of splicing factor 1. We therefore hypothesized that intronic differences between *HLA-C*04* or **16* and the other HLA-C allele groups may influence the branch point strength and/or the 5′- or 3′ splice sites resulting in differences in exon 5 splicing specificity or efficacy.

To test this hypothesis on an in-silico basis we used the human splicing finder (HSF), an online tool from genomnis (https://www.genomnis.com/,[26]). First, we compared the branch point strengths within the intron 4 region of all HLA-C reference alleles. The tool predicted multiple branch point motifs in the intron 4 sequence of the different HLA-C alleles, but the branch point with the highest score was identical for all alleles (Table 3), indicating that exon 5 skipping in *HLA-C*04* and **16* alleles is probably not the result of branch point differences.

Secondly, the variation in and strength of the splice sites downstream and upstream of exon 5 were compared for all the different HLA-C reference alleles using both the HSF algorithm as well as the MaxEnt algorithm. These regions were analyzed since exon 5 skipping could occur in case of acceptor- or donor splice site loss or when the acceptor splice site of exon 6 is preferred over the acceptor splice site of exon 5. For the acceptor splice site of exon 5, two (HSF) or three (MaxEnt) different splice site motifs were calculated (Table 3). Both algorithms predicted a different splice site for *HLA-C*05* and **08* compared with the other HLA-C alleles and the MaxEnt algorithm additionally predicted a different splice site for *HLA-C*07*. Importantly, *HLA-C*04* and **16* had splice sites that were comparable to the alleles that did not show exon 5 skipping in our assays (i.e., *HLA-C*01*, *-*02*, *-*03*, *-*06*, *-*12*, *-*14*, *-*15*). The splice site motif for the donor splice site of exon 5 and the acceptor splice site of exon 6 were identical for all HLA-C allele groups (Table 3). Based on this in-silico analysis, the calculated splice strength of *HLA-C*04* and **16* was comparable to the other HLA-C alleles, in which no alternative splicing was detected. Therefore, these results do not support the hypothesis that variation in the splice site sequences is responsible for the observed exon 5 skipping in *HLA-C*04* and **16* alleles.

3.4 | In-silico analysis of the effect of exonic HLA-C variation on splicing regulatory elements

SREs are short stretches (4–18 nucleotides) of DNA that act as cis-regulating elements by recruiting regulatory proteins that interact with components of the splicing machinery and enhance or suppress splicing.[33] SNPs in SRE may influence binding of those regulatory proteins consequently leading to alternative splicing. Four types of SRE exist: (1) exonic splicing enhancers (ESEs), (2) exonic splicing silencers (ESSs), (3) intronic splicing enhancers (ISE) and (4) intronic splicing silencers (ISS).

To provide a theoretical basis for the potential influence of exon SNPs in *HLA-C*04* or **16* alleles on exon 5 skipping, we compared the polymorphic content of exon 5 of the different HLA-C alleles and studied the effect of exonic HLA-C polymorphism on ESE and ESS motifs in-silico. Unfortunately, it was not possible to test the effect on intronic motifs, since there are no reliable algorithms available to perform this in-silico analysis. The IPD-IMGT/HLA database (V3.44.0) describes 741 different *HLA-C*04* alleles, of which 475 have a known sequence for exon 5.[34] With the exception of *HLA-C*04:03*, **04:06*, and **04:82*, all common and well-

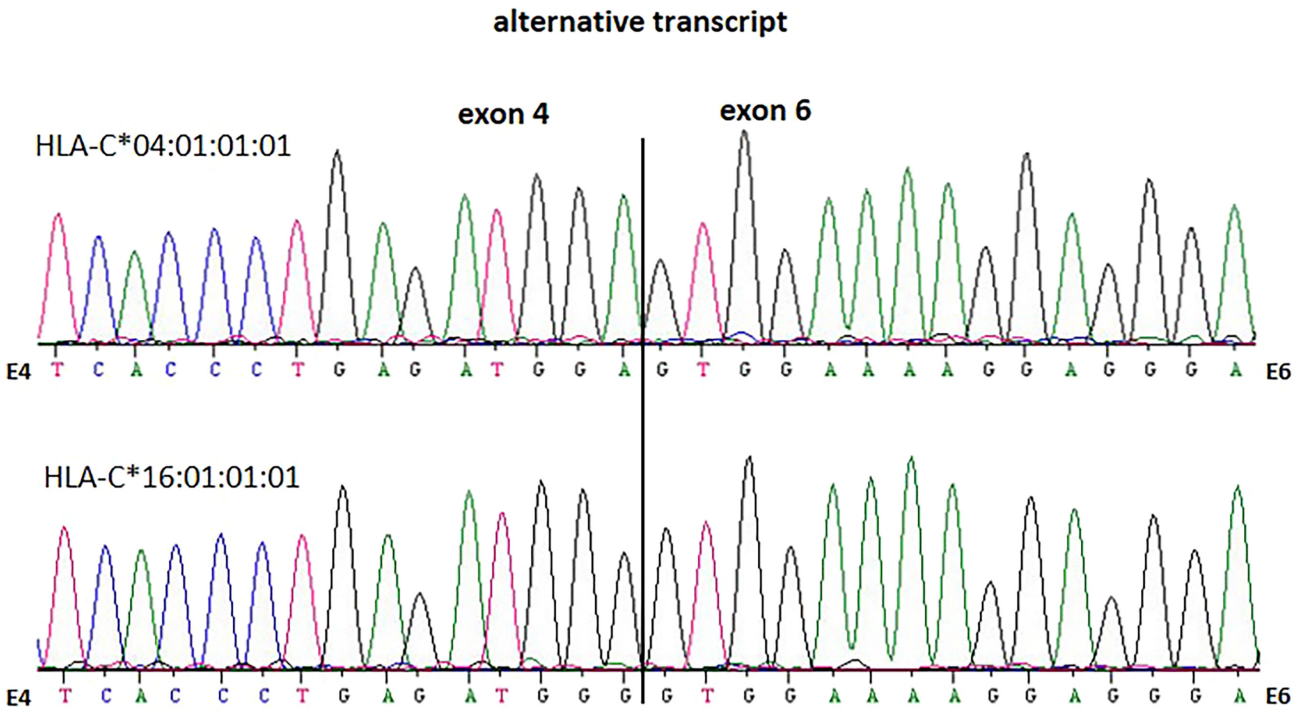

FIGURE 2 Exon 5 skipping in *HLA-C*04* and *HLA-C*16*. The alternatively spliced products of *C*04:01:01:01* and **16:01:01:01* were separated from the mature product by using a low melting gel. The isolated alternative products were sequenced and the resulting nucleotide sequences are shown

TABLE 3 Analysis of intronic branch points and splice site motifs in HLA-C

Branch point analysis	**HLA-C allele group**	**Strongest branch point motif**	**Strength**
Intron 4	All HLA-C groups	CCCTCAC	99.76
HSF	**HLA-C allele group**	**Splice site motif**	**Strength**
Acceptor splice site exon 5	**01*, **02*, **03*, **04*, **06*, **07*, **12*, **14*, **15*, **16*, **17*, **18*	CTCCTTTCCCAGAG	87.86
	**05*, **08*	CTCCTTTCCCAG**GG**	89.77
Donor splice site exon 5	All HLA-C groups	CAGGTAGGG	90.21
Acceptor splice site exon 6	All HLA-C groups	TTCTTCCCACAGGT	92.21
MaxEnt	**HLA-C allele group**	**Splice site motif**	**Strength**
Acceptor splice site exon 5	**01*, **02*, **03*, **04*, **06*, **12*, **14*, **15*, **16*, **17*, **18*	ACCTTCCCCTCCTTTCCCAGAGC	11.26
	**05*, **08*	ACCTTCCCCTCCTTTCCCAG**G**GC	11.91
	**07*	ACCTTC**T**CCTCCTTTCCCAGAGC	10.77
Donor splice site exon 5	All HLA-C groups	CAGGTAGGG	9.46
Acceptor splice site exon 6	All HLA-C groups	AGGGCATTTTCTTCCCACAGGTG	9.73

Note: The Human Splicing Finder (HSF) tool from Genomnis was used to analyze branch point- and splice site motifs. For the branch point analysis of intron 4, the branch point with the highest score is depicted. To assess splice sites of exon 5 and 6, the HSF tool used the HSF and the MaxEnt algorithm to determine the splice site motifs and strength of the different HLA-C reference alleles: *HLA-C*01:02:01:01*, **02:02:02:01*, **03:02:01*, **04:01:01:01*, **05:01:01:01*, **06:02:01:01*, **07:01:01:01*, **08:01:01:01*, **12:02:02:01*, **14:02:01:01*, **15:02:01:01*, **16:01:01:01*, **17:01:01:02*, and **18:01:01:01*.

documented *HLA-C*04* allele sequences have identical exon 5 sequences with an *HLA-C*04* specific non-synonymous nucleotide substitution (G2063A) that results in an amino acid change from Valine to Methionine (Figure 3).[35] For the *HLA-C*16* allele group, the IPD-IMGT/HLA database describes 273 different alleles, of which 156 have a known sequence for exon 5. All the common and well-documented *HLA-C*16* allele sequences are identical in this region and

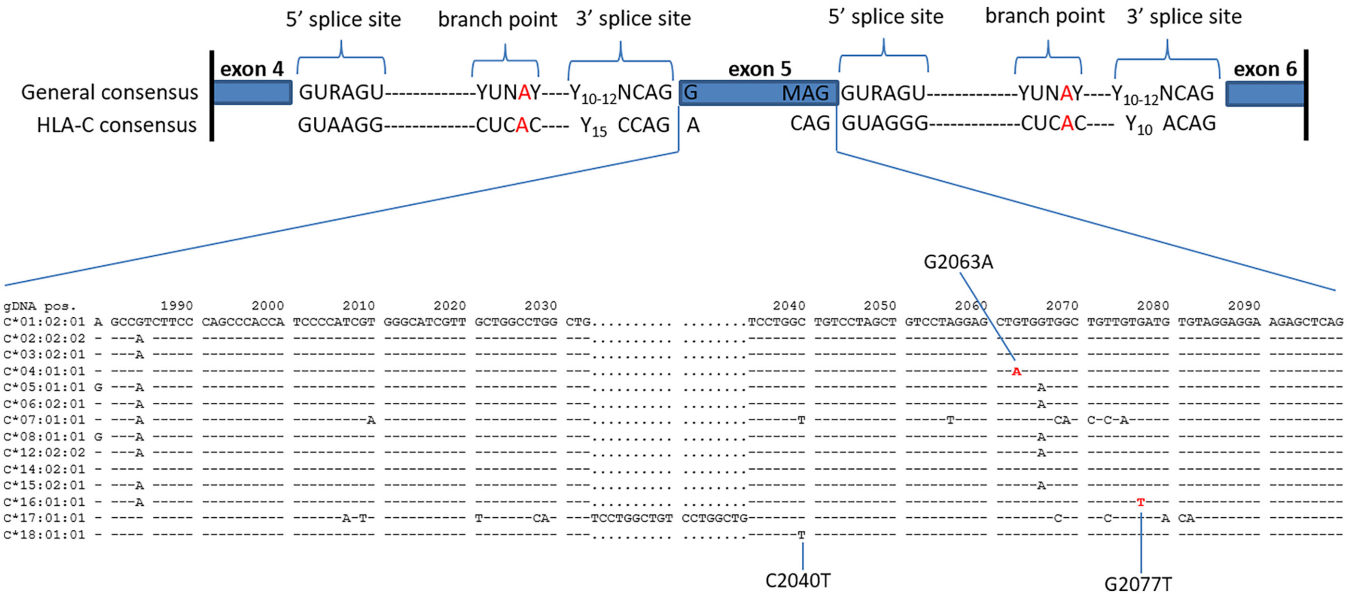

FIGURE 3 HLA-C polymorphism analysis. The consensus HLA-C sequence of the splice site and branch site regions surrounding exon 5 are shown on top. The HLA-C sequence variation within the exon 5 region is shown for each HLA-C allele group and single nucleotide differences in *C*04*, *C*16*, and *C*18* are highlighted

display an *HLA-C*16* specific synonymous nucleotide substitution in exon 5 (G2077T, Figure 3).

To test whether these exonic polymorphisms could alter sequence motifs of ESE/ESS and affect exon 5 skipping, the prediction tool EX-SKIP was used. The tool determines the total number of ESE and ESS motifs using the PESE/PESS-,[28] FAS-ESS-,[29] EIE/IIE-,[30] RESCUE-ESE-[31] and the neighborhood inference (NI)[32] algorithms. Based on the calculated ESS/ESE ratio it predicts which of two exon sequences has a higher chance of exon skipping. To determine the effect of the *HLA-C*04* and **16* specific substitutions, we compared the ESS/ESE prediction results of the exon 5 sequences of *HLA-C*04:01:01:01* and **16:01:01:01* to the exon 5 sequences of **01:02:01:01* and **02:02:02:01* respectively, because in this region the sequences are identical, except for substitutions G2063A and G2077T. EX-SKIP predicted that *C*04* has 1 ESS and 5 ESE's less than *C*01* changing the ESS/ESE ratio from 0.58 to 0.60 (3%) (Table 4). Although the tool predicted a higher change of exon skipping for *HLA-C*04*, the effect of G2063A on ESS/ESE ratio is minimal. For HLA-*C*16*, EX-SKIP predicted 10 ESS and 1 ESE more than C*02 and the ESS/ESE ratio changed from 0.56 to 0.66 (18%) (Table 4). By comparing the ESS/ESE ratio of all the HLA-C reference alleles, we observed that the HLA-C alleles for which we did not find exon 5 skipping had lower ESS/ESE ratios than *HLA-C*16*. Interestingly, the ratio for *HLA-C*18* was identical to **16*, but this HLA-C allele was not included in our sequencing panel and therefore we could not relate this ratio to the actual presence of the alternatively spliced variant.

Overall, exon analysis by EX-SKIP predicted ESE and ESS motif changes in *HLA-C*04* and **16* alleles. But, since the effect was only very small for *HLA-C*04*, we assume that only for **16* these changes may contribute to altered splicing processes and eventually result in the exon 5 skipping we observed.

4 | DISCUSSION

HLA class I genes are highly polymorphic, but only a limited number of alternatively spliced transcripts has been identified.[10] In this study, sequencing of mRNA revealed an alternatively spliced transcript missing exon 5 in *HLA-C*04* and **16* alleles. Alternative splicing of exon 5 has been show before in cell lines.[21–23] The strength of our study is that we studied alternative exon splicing in peripheral blood samples covering the most common HLA-C allele groups that were processed and analyzed with the same methodology. Hence, our data suggest that in peripheral blood samples splicing of exon 5 occurs more frequently in *HLA-C*04* and **16* alleles than in the other allele groups. Variability in the nucleotide sequence can affect exon and intron recognition during the splicing process by altering the splice sites, branch point or SREs and this may eventually result in the formation of an alternatively spliced transcript. We therefore assessed how the specific nucleotide variability in *HLA-C*04* and **16* alleles affected the splice sites, branch point and SREs by using an in-silico approach with several existing prediction algorithms.[36,37] This suggested that *HLA-C*04* and **16* specific nucleotide variants in intron 4 did not affect splice site and branch site sequences. Furthermore,

TABLE 4 EX-SKIP analysis of the exon 5 sequence of HLA-C alleles

HLA-C allele group	PESS (count)	FAS-ESS hex2 (count)	FAS-ESS hex3 (count)	IIE (count)	NI-ESS trusted (count)	PESE (count)	RESCUE-ESE (count)	EIE (count)	NI-ESE trusted (count)	ESS (total)	ESE (total)	ESS/ESE (ratio)
**01*	0	4	3	42	4	12	5	32	42	53	91	0.58
**02*	0	4	3	41	4	12	6	32	43	52	93	0.56
**03*	0	4	3	41	4	12	6	32	43	52	93	0.56
**04*	0	4	3	40	5	10	5	32	39	52	86	0.60
**05*	1	3	2	39	3	12	6	30	42	48	90	0.53
**06*	1	3	2	39	3	12	6	30	42	48	90	0.53
**07*	1	3	2	36	6	12	7	34	40	48	93	0.52
**08*	1	3	2	39	3	12	6	30	42	48	90	0.53
**12*	1	3	2	39	3	12	6	30	42	48	90	0.53
**14*	0	4	3	42	4	12	5	32	42	53	91	0.58
**15*	1	3	2	39	3	12	6	30	42	48	90	0.53
**16*	1	6	4	41	10	12	6	33	43	62	94	0.66
**17*	0	5	3	41	6	16	6	30	61	55	113	0.49
**18*	0	4	3	44	6	12	5	32	38	57	87	0.66

Note: The exon 5 sequences of *HLA-C*01:02:01:01*, **02:02:02:01*, **03:02:01*, **04:01:01:01*, **05:01:01:01*, **06:02:01:01*, **07:01:01:01*, **08:01:01:01*, **12:02:02:01*, **14:02:01:01*, **15:02:01:01*, **16:01:01:01*, **17:01:01:02*, *and *18:01:01:01* alleles were entered into the online EX-SKIP tool and the ESE/ESS results of the different algorithms are shown: 8-mer putative ESSs and ESEs (PESS/PESE), fluorescence-activated screen for ESS (FAS-ESS hex2, FAS-ESS hex3), intron- and exon-identity elements (IIE/EIE), Relative Enhancer and Silencer Classification by Unanimous Enrichment ESEs (RESCUE-ESE), neighborhood interference ESS and ESEs (NI-ESS/NI-ESE).

the specific SNP in exon 5 observed in *HLA-C*04* compared with other HLA-C allele groups did not seem to have an effect on exonic SRE, while the *HLA-C*16* specific SNP in exon 5 did alter exonic SRE. Although the results of these in-silico analysis are only predictive, they may provide a theoretical basis for the alternative splicing observed for the *HLA-C*16:01:01* allele. For the alternative splicing of *HLA-C*04:01:01* it is not clear why this occurs, although a role for intronic SREs could not be excluded based on our data.

Splice sites have been extensively characterized and in-silico tools like HSF are able to reliably predict alterations in splice site strength.[38,39] SREs are less well understood and are more difficult to predict by using in-silico approaches. The best way to accurately predict the effect of nucleotide variability on SREs and alternative splicing is the use of a combination of algorithms. In this study, we used the EX-SKIP tool to predict SRE alterations, since several of the available prediction algorithms have been combined in this tool. Multiple studies have evaluated EX-SKIP and showed promising results with a sensitivity of around 75%.[40,41] Although no tool can ensure 100% accuracy and experimental validation is necessary to draw a clinically relevant conclusion, these tools are very useful and commonly used to predict and understand the mechanisms of alternative splicing.

In our study, we investigated, on an in-silico basis, whether SNPs in the introns of the different HLA-C allele groups influenced branch point or splice site sequences and we observed no major differences between the groups. Intronic SNPs could, at least in theory, also influence intronic silencing and enhancer motifs by creating novel regulatory elements that in case of intron splicing silencers might promote exon skipping.[37] Although studies have identified more than 100 potential intronic splicing regulatory elements (ISREs) in other genes, the ISREs are less well understood[42,43] and good prediction algorithms for ISREs are limited. Hence, we have not included the intron regions in our in-silico analysis of intronic regulatory elements and more research on the identification of ISREs is required to predict the effect of intronic variants on exon skipping.

When using the EX-SKIP tool to study the impact of exon 5 SNPs on alternative splicing of HLA-C, EX-SKIP predicted the same ESE/ESS ratio for *HLA-C*16* and **18* alleles (Table 4), which suggests that exon 5 skipping might also occur in the reference allele *HLA-C*18:01*. Unfortunately, we were not able to test this because the *HLA-C*18* genotype is rarely present in the Caucasian population (0.03%,[44]) and therefore a sample with *HLA-C*18* could not be included in our test panel. In comparison, *HLA-C*04* and **16* alleles are found in the Caucasian population with a frequency of 10.6% and 3.4%, respectively, and were thus present in our RNA panel. Next to *HLA-C*18:01*, **18:02* is the only other well-documented HLA-C allele with known exon 5 sequence in the IPD-IMGT/HLA database. In the exon 5 region, the sequence of HLA-*C*18:02* differs from **18:01* by one nucleotide (T2040C) and is identical to the exon 5 sequence of *HLA-C*01*, with an EX-SKIP ESS/ESE ratio of 0.58 (Table 4). This observation suggests that exon 5 skipping might occur more frequently in *HLA-C*18:01* than in **18:02* alleles. The C2040T substitution in *HLA-C*18:01* is also present in the *HLA-C*07* alleles (Figure 3), but in *HLA-C*07* alleles it is not predicted to have an impact on exon 5 skipping since additional variation in exon 5 of *HLA-C*07* seems to result in more enhancing and less suppressing elements, as is shown by the EX-SKIP analysis with an ESS/ESE ratio of 0.52 for *HLA-C*07* (Table 4).

We found that the complete exon 5 region was spliced out in the alternative transcript of *HLA-C*04* and **16* alleles. It remains to be determined whether the alternative transcripts of *HLA-C*04* and **16* are translated into a mature and functional protein. However, because the reading frame was maintained, no premature stop codon was created, and therefore it is expected that the alternative transcripts can be fully translated into a protein. To analyze the protein expression, cells lacking HLA class I expression could be transfected with constructs encoding the mature and alternative sequence of *HLA-C*04* and **16* and subsequently be analyzed for HLA-C protein expression for instance with western blot or flow cytometry.

Here, we studied alternative splicing of HLA-C in peripheral blood samples and we demonstrated that allele polymorphism in *HLA-C*04* and **16* alleles correlate with occurrence of exon 5 skipping. We did not observe alternative splicing in the other allele groups in peripheral blood cells in a standardized setting measuring all HLA-C groups in the same experimental conditions. Also, our in-silico models support the hypothesis that polymorphism in the HLA-C alleles has an influence on alternative splicing leading to exon 5 skipping. Since the process of alternative splicing can be influenced by cell intrinsic and cell extrinsic factors, it could be relevant to assess whether these allele-related differences in exon 5 skipping we observed in peripheral blood cells are also relevant under other conditions. This could be especially interesting because two studies described transcripts lacking exon 5 for *HLA-C*02*, **12:02* and **15:02* in transformed B-cell lines,[22,23] whereas a third study showed exon 5 skipping of *HLA-C*16* in transformed B cell lines.[21] Since transformation of cells or other environmental conditions could influence HLA expression levels and alternative splicing processes it would be interesting to structurally study the impact of polymorphisms in these settings as well.[45]

Our study provides new perspectives for the analysis of alternative splicing products, not restricted to exon

5 skipping in HLA-C, for instance by using RNAseq data sets to ultimately understand their occurrence and functional role. Additionally, it will be interesting to further examine if the alternative variants occur in a tissue-specific or cell type-specific manner. One research group discovered alternative splicing of HLA-C specifically in NK cells.[24] Interestingly, alternative splicing occurred in the promotor region of HLA-C, resulting in different promoters and differential HLA-C expression on NK cells.[24,25] The differential expression of HLA-C on NK cells themselves regulated NK cell functions and the authors suggest that NK cell-intrinsic regulation of HLA-C evolved to fine-tune NK cell responses.

In summary, we reported alternative splicing of HLA-C in peripheral blood cells, which resulted in a transcript without the exon 5 region. The alternative transcript was only detected in *HLA-C*04* and *16 alleles. In-silico splicing analysis predicted that nucleotide variability in *HLA-C*16* alleles altered ESREs. Hence, our collective data support the concept that specific SNPs in the HLA-C genes can influence the process of alternative exon 5 splicing.

AUTHOR CONTRIBUTIONS

Femke A. I. Ehlers and Timo I. Olieslagers performed the experiments, analyzed the data and wrote the paper. Mathijs Groeneweg participated in data visualization. Mathijs Groeneweg, Gerard M. J. Bos, Marcel G. J. Tilanus, Christina E.M. Voorter and Lotte Wieten reviewed and edited the paper. Timo I. Olieslagers, Marcel G. J. Tilanus, Christina E. M. Voorter and Lotte Wieten conceptualized the study.

ACKNOWLEDGMENT

The Authors thank Tom Jansen for his assistance in sequencing and data analysis.

CONFLICT OF INTEREST

The authors have nothing to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Femke A. I. Ehlers https://orcid.org/0000-0003-2272-6286
Timo I. Olieslagers https://orcid.org/0000-0001-9419-945X
Mathijs Groeneweg https://orcid.org/0000-0002-6615-9239
Marcel G. J. Tilanus https://orcid.org/0000-0001-9214-392X

REFERENCES

1. Apps R, Meng Z, Del Prete GQ, Lifson JD, Zhou M, Carrington M. Relative expression levels of the HLA class-I proteins in normal and HIV-infected cells. *J Immunol*. 2015;194(8):3594-3600.
2. Flomenberg N, Baxter-Lowe LA, Confer D, et al. Impact of HLA class I and class II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplantation outcome. *Blood*. 2004;104(7):1923-1930.
3. Pidala J, Wang T, Haagenson M, et al. Amino acid substitution at peptide-binding pockets of HLA class I molecules increases risk of severe acute GVHD and mortality. *Blood*. 2013;2005:501510.
4. Woolfrey A, Klein JP, Haagenson M, et al. HLA-C antigen mismatch is associated with worse outcome in unrelated donor peripheral blood stem cell transplantation. *Biol Blood Marrow Transplant*. 2011;17(6):885-892.
5. Frohn C, Fricke L, Puchta JC, Kirchner H. The effect of HLA-C matching on acute renal transplant rejection. *Neprol Dial Transplant*. 2001;16(2):355-360.
6. Vince N, Li H, Ramsuran V, et al. HLA-C level is regulated by a polymorphic Oct1 binding site in the HLA-C promoter region. *Am J Hum Genet*. 2016;99(6):1353-1358.
7. Thomas R, Apps R, Qi Y, et al. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nat Genet*. 2009;41(12):1290-1294.
8. Kulkarni S, Savan R, Qi Y, et al. Differential microRNA regulation of HLA-C expression and its association with HIV control. *Nature*. 2011;472(7344):495-498.
9. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu rev Biochem*. 2003;72(1):291-336.
10. Voorter CE, Gerritsen KE, Groeneweg M, Wieten L, Tilanus MG. The role of gene polymorphism in HLA class I splicing. *Int J Immunogenet*. 2016;43(2):65-78.
11. Matera AG, Wang Z. A day in the life of the spliceosome. *Nat rev Mol Cell Biol*. 2014;15:108-121.
12. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat rev Genet*. 2002;3(4):285-298.
13. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 2008;40(12):1413-1415.
14. Moreau P, Carosella E, Teyssier M, et al. Soluble HLA-G molecule: an alternatively spliced HLA-G mRNA form candidate to encode it in peripheral blood mononuclear cells and human trophoblasts. *Hum Immunol*. 1995;43(3):231-236.
15. Paul P, Cabestre FA, Lefebvre S, et al. Identification of HLA-G7 as a new splice variant of the HLA-G mRNA and expression of soluble HLA-G5,-G6, and-G7 transcripts in human transfected cells. *Hum Immunol*. 2000;61(11):1138-1149.
16. Carosella ED, Rouas-Freiss N, Tronik-Le Roux D, Moreau P, LeMaoult J. HLA-G: an immune checkpoint molecule. *Adv Immunol*. 2015;127:33-144.
17. Krangel MS. Two forms of HLA class I molecules in human plasma. *Hum Immunol*. 1987;20(2):155-165.
18. Krangel MS. Secretion of HLA-A and-B antigens via an alternative RNA splicing pathway. *J Exp Med*. 1986;163(5):1173-1190.

19. Tijssen H, Sistermans E, Joosten I. A unique second donor splice site in the intron 5 sequence of the *HLA-A* 11* alleles results in a class I transcript encoding a molecule with an elongated cytoplasmic domain. *Tissue Antigens*. 2000;55(5):422-428.
20. Dubois V, Tiercy J, Labonne M, Dormoy A, Gebuhrer L. A new HLA-B44 allele (*B*44020102S*) with a splicing mutation leading to a complete deletion of exon 5. *Tissue Antigens*. 2004;63(2):173-180.
21. Cianetti L, Testa U, Scotto L, et al. Three new class I HLA alleles: structure of mRNAs and alternative mechanisms of processing. *Immunogenetics*. 1989;29(2):80-91.
22. Vilches C, de Pablo R, Herrero MJ, Moreno ME, Kreisler M. Molecular cloning and polymerase chain reaction-sequence-specific oligonucleotide detection of the allele encoding the novel allospecificity HLA-Cw6. 2 (Cw* 1502) in Spanish gypsies. *Hum Immunol*. 1993;37(4):259-263.
23. Yang D, Le J. Targeted amplification of alternatively spliced transcripts of major histocompatibility complex class I heavy chain. *J Immunol Methods*. 1994;176(2):265-270.
24. Li H, Ivarsson MA, Walker-Sperling VE, et al. Identification of an elaborate NK-specific system regulating HLA-C expression. *PLoS Genet*. 2018;14(1):e1007163.
25. Goodson-Gregg FJ, Rothbard B, Zhang A, et al. Tuning of NK-specific HLA-C expression by alternative mRNA splicing. *Front Immunol*. 2019;10:3034.
26. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37(9):e67.
27. Raponi M, Kralovicova J, Copson E, et al. Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. *Hum Mutat*. 2011;32(4):436-444.
28. Zhang XH, Chasin LA. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev*. 2004;18(11):1241-1250.
29. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. *Cell*. 2004;119(6):831-845.
30. Zhang C, Li W-H, Krainer AR, Zhang MQ. RNA landscape of evolution for optimal exon and intron discrimination. *Proc Natl Acad Sci USA*. 2008;105(15):5797-5802.
31. Fairbrother WG, Yeo GW, Yeh R, et al. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res*. 2004;32(suppl_2):W187-W190.
32. Stadler MB, Shomron N, Yeo GW, Schneider A, Xiao X, Burge CB. Inference of splicing regulatory activities by sequence neighborhood analysis. *PLoS Genet*. 2006;2(11):e191.
33. Lee Y, Rio DC. Mechanisms and regulation of alternative pre-mRNA splicing. *Annu rev Biochem*. 2015;84:291-323.
34. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res*. 2020;48(D1):D948-D955.
35. Mack SJ, Cano P, Hollenbach JA, et al. Common and well-documented HLA alleles: 2012 update to the CWD catalogue. *Tissue Antigens*. 2013;81(4):194-203.
36. Badr E, ElHefnawi M, Heath LS. Computational identification of tissue-specific splicing regulatory elements in human genes from RNA-Seq data. *PloS One*. 2016;11(11):e0166978.
37. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*. 2018;59(3):253-268.
38. Moles-Fernández A, Duran-Lozano L, Montalban G, et al. Computational tools for splicing defect prediction in breast/ovarian cancer genes: how efficient are they at predicting RNA alterations? *Front Genet*. 2018;9:366.
39. Jian X, Boerwinkle E, Liu X. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet Med*. 2014;16(7):497-503.
40. Grodecká L, Buratti E, Freiberger T. Mutations of pre-mRNA splicing regulatory elements: are predictions moving forward to clinical diagnostics? *Int J Mol Sci*. 2017;18(8):1668.
41. Soukarieh O, Gaildrat P, Hamieh M, et al. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools. *PLoS Genet*. 2016;12(1):e1005756.
42. Wang Y, Ma M, Xiao X, Wang Z. Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat Struct Mol Biol*. 2012;19(10):1044-1052.
43. Voelker RB, Berglund JA. A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Res*. 2007;17(7):1023-1033.
44. Gragert L, Madbouly A, Freeman J, Maiers M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Hum Immunol*. 2013;74(10):1313-1320.
45. Stamm S, Ben-Ari S, Rafalska I, et al. Function of alternative splicing. *Gene*. 2005;344:1-20.

How to cite this article: Ehlers FAI, Olieslagers TI, Groeneweg M, et al. Polymorphic differences within HLA-C alleles contribute to alternatively spliced transcripts lacking exon 5. *HLA*. 2022;100(3):232-243. doi:10.1111/tan.14695