Received: 11 March 2024 | Revised: 12 March 2024 | Accepted: 18 March 2024

DOI: 10.1111/tan.15459

NEW ALLELE ALERT

HLA Immune Response Genetics WILEY

Characterization of the novel *HLA-C*06:376N* allele by Pacific Biosciences HiFi sequencing in a Chinese individual

Zihang Li[1] | **Yankun Li**[1] | **Liying Wang**[2] | **Peng Gao**[1,3] | **Xiaoning Wang**[1]

[1]Department of Hematology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

[2]Department of Orthodontics, Stomatological Hospital, Xi'an Jiaotong University, Xi'an, China

[3]Genome Institute, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

Correspondence
Xiaoning Wang, Department of Hematology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China.
Email: wangxn99@163.com

Funding information
Ministry of Science and Technology of People's Republic of China, Grant/Award Number: 2022YFC3400302; Department of Science and Technology of Shaanxi Province, Grant/Award Number: KSJRC202202; National Key R&D Program, Grant/Award Number: 2022YFC2502701

*HLA-C*06:376N* differs from *HLA-C*06:02:01:01* by seven nucleotide changes in exon 2, intron 2, and exon 3.

KEYWORDS

*HLA-C*06:376N*, novel allele, third-generation sequencing

In this study, we identified a novel *HLA-C*06:376N* allele, which shows 7 nucleotide changes compared to the closely related allele *HLA-C*06:02:01:01*. These mutations include 5 base mutations and 2 base insertions, which result in a shift of the open reading frame (ORF) and an abnormal premature stop codon (Figure 1). Among them, base substitution at 3 positions (base 807, 1094, and 1160) and insertion at base 1279 result in the change of amino acid (Lysine 104 to Asparagine, Threonine 118 to Isoleucine, Serine 140 to Tyrosine, and Tryptophan 180 to Leucine, respectively). The insertion of C at the base 1279 also causes a shift of the ORF from then on. Downstream of this insertion, there is another base insertion with a T inserted at position 1303. In addition to these 5 nucleotide changes, 2 base substitutions at base 708 (C>G in exon 2) and base 879 (C>T in intron 2) do not result in the amino acid change. The frameshift mutations caused by the 2 base insertions results in premature termination of translation at base 1970 (codon 214 TGA) in exon 4.

This discovery was made during our investigation of a cohort of healthy Chinese individuals, with the novel *HLA-C*06:376N* observed in a healthy 34-year-old female participant who is a native of Weinan City, Shaanxi Province, China. The complete HLA genotyping obtained for this individual was: *HLA-A*02:01:01, 24:02:01*; *-B*13:02:01, 35:01:01*; *-C*03:03:01, 06:376N*; *-DRB1*07:01:01, 15:01:01*; *-DRB4*01:03:01*; *-DRB5*01:01:01*; *-DQA1*01:02:01, 02:01:01*; *-DQB1*02:02:01, 06:02:01*; *-DPA1*02:01:01, 02:02:02*; *-DPB1*02:01:02, 17:01:01*.

```
gDNA                      |Exon2                                                                                         660
C*06:02:01:01     GCCCCCAG|GC TCCCACTCCA TGAGGTATTT CGACACCGCC GTGTCCCGGC CCGGCCGCGG AGAGCCCCGC TTCATCTCAG TGGGCTACGT GGACGACACG
C*06:376N         --------|-- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

gDNA                                                                                                                    760
C*06:02:01:01     CAGTTCGTGC GGTTCGACAG CGACGCCGCG AGTCCGAGAG GGGAGCCCCG GGCGCCGTGG GTGGAGCAGG AGGGGCCGGA GTATTGGGAC CGGGAGACAC
C*06:376N         ---------- ---------- ---------- ---------- -------G-- ---------- ---------- ---------- ---------- ----------

gDNA                                                                                                   |Intron2         860
C*06:02:01:01     AGAAGTACAA GCGCCAGGCA CAGGCTGACC GAGTGAACCT GCGGAAACTG CGCGGCTACT ACAACCAGAG CGAGGACG|GT GAGTGACCCC GGCCCGGGGC
C*06:376N         ---------- ---------- ---------- ---------- ------C--- ---------- ---------- --------|-- ---------- ----------

gDNA                                                                                                                    960
C*06:02:01:01     GCAGGTCACG ACCCCTCCCC ATCCCCCACG GACGGCCCGG GTCGCCCCGA GTCTCCCGGT CTGAGATCCA CCCCGAGGCT GCGGAACCCG CCCAGACCCT
C*06:376N         ---------- --------T- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

gDNA                                                                                                                   1060
C*06:02:01:01     CGACCGGAGA GAGCCCCAGT CACCTTTACC CGGTTTCATT TTCAGTTTAG GCCAAAATCC CCGCGGGTTG GTCGGGGCTG GGGCGGGGCT CGGGGGACGG
C*06:376N         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

gDNA                          |Exon3                                                                                   1160
C*06:02:01:01     GGCTGACCAC GGGGGCGGGG CCAG|GGTCTC ACACCCTCCA GTGGATGTAT GGCTGCGACC TGGGGCCCGA CGGGCGCCTC CTCCGCGGGT ATGACCAGTC
C*06:376N         ---------- ---------- ----|------ ---T------ ---------- ---------- ---------- ---------- ---------- ---------A

gDNA                                                                                                                   1260
C*06:02:01:01     CGCCTACGAC GGCAAGGATT ACATCGCCCT GAACGAGGAC CTGCGCTCCT GGACCGCCGC GGACACGGCG GCTCAGATCA CCCAGCGCAA GTGGGAGGCG
C*06:376N         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

gDNA                                                                                                                   1360
C*06:02:01:01     GCCCGTGAGG CGGAGCAG.T GGAGAGCCTA CCTGGAGGGC AC.GTGCGTG GAGTGGCTCC GCAGATACCT GGAGAACGGG AAGGAGACGC TGCAGCGCGC
C*06:376N         ---------- --------C- ---------- ---------- --T------- ---------- ---------- ---------- ---------- ----------

gDNA                |Intron3                                                                                           1460
C*06:02:01:01     GG|GTACCAGG GGCAGTGGGG AGCCTTCCCC ATCTCCTGTA GATCTCCCGG GATGGCCTCC CACGAGGAGG GGAGGAAAAT GGGATCAGCG CTAGAATATC
C*06:376N         --|-------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
```

FIGURE 1 Sequence alignment of *HLA-C* gene, partial intron 1, exon 2, intron 2, exon 3 and partial intron 3 nucleotide sequences of the novel *HLA-C*06:376N* allele and the closely related *HLA-C*06:02:01:01* allele. Dashes (–) indicate nucleotide identity between the two alleles. The pipes (|) indicate the exon and intron boundaries. The numbers above the sequence indicate the *HLA-C*06:376N* allele gDNA position.

The high molecular weight (HMW) genomic DNA from peripheral blood mononuclear cells (PBMCs) was extracted using the Monarch HMW DNA Extraction Kit for Cell & Blood (NEB, T3050L) following the vendor's manual. Each subject's classical HLA genes (*HLA-A*, *-B*, *-C*, *-DRB1*, *-DRB3*, *-DRB4*, *-DRB5*, *-DQA1*, *-DQB1*, *-DPA1*, and *-DPB1*) were amplified through multiplex PCR for HiFi library preparation. Subsequently, the libraries from different subjects were combined and sequenced using the third-generation sequencing platform Pacific Biosciences Sequel II. The genotypes of the HLA genes were determined utilizing the IPD-IMGT/HLA Database (version 3.50.0) and various software packages, including pbmm2, CCS, lima, and WhatsHap.

The sequence data of the recipient have been submitted to the GenBank and the IPD-IMGT/HLA Database.[1] The Accession Numbers assigned to them were OR972663 and HWS10068731, respectively. The name *HLA-C*06:376N* has been officially assigned by the WHO Nomenclature Committee for Factors of the HLA System in February 2024. This follows the agreed policy that, subject to the conditions stated in the most recent Nomenclature Report,[2] names will be assigned to new sequences as they are identified. Lists of such new names will be published in the following WHO Nomenclature Report.

AUTHOR CONTRIBUTIONS

Zihang Li, Yankun Li, and Liying Wang collected samples, analyzed the data, submitted sequence data to GenBank and IPD-IMGT/HLA Database and wrote the manuscript. Zihang Li did the sequencing. Peng Gao reviewed the manuscript. Xiaoning Wang designed the study and reviewed the manuscript. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors would like to offer sincere gratitude to the volunteer donors for participation in their bone marrow donor registry. This work was supported by High-level

Talent Introduction Plan of Shaanxi (Youth Program) Department of Science and Technology of Shaanxi Province (grant no. KSJRC202202), Ministry of Science and Technology of People's Republic of China (grant no. 2022YFC3400302) and National Key R&D Program (grant no. 2022YFC2502701).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Information of this study is freely available from the IPD-IMGT/HLA Database.

ORCID

Zihang Li https://orcid.org/0000-0003-3800-2493
Liying Wang https://orcid.org/0000-0002-2472-4076
Peng Gao https://orcid.org/0000-0002-7785-2389
Xiaoning Wang https://orcid.org/0000-0002-7925-2774

REFERENCES

1. Barker DJ, Maccari G, Georgiou X, et al. IPD-IMGT/HLA Database. *Nucleic Acids Res*. 2023;51:D1053-D1060.
2. Marsh SGE, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*. 2010;75:291-455.

How to cite this article: Li Z, Li Y, Wang L, Gao P, Wang X. Characterization of the novel *HLA-C*06:376N* allele by Pacific Biosciences HiFi sequencing in a Chinese individual. *HLA*. 2024;103(4):e15459. doi:10.1111/tan.15459