A novel HLA-C*03:04:01:47 allele sequence identified using Pacific Biosciences SMRT sequencing

Angus G. Comerford[1] | Jonathan A. M. Lucas[1] | Thomas R. Turner[1,2] | Steven G. E. Marsh[1,2] | Neema P. Mayor[1,2]

1. Anthony Nolan Research Institute, Royal Free Hospital, London, UK
2. UCL Cancer Institute, Royal Free Campus, London, UK

Summary: A novel variant of HLA-C*03:04:01:47 was identified using Pacific Biosciences SMRT sequencing platform.

Keywords: HLA, HLA-C, genomic-sequence, novel allele, SMRT sequencing

The hyperpolymorphic HLA genes are located within the human major histocompatibility complex found on the short arm of chromosome 6.[1] To date, there are 19,586 HLA class I alleles, of which 5,842 are alleles of HLA-C, in the IPD-IMGT/HLA Database (Release 3.40.0, 2020-04-20).[2] HLA-C has 3,503 different protein structures, 255 null alleles and 2,084 alleles with only non-coding differences. Next generation sequencing platforms have allowed extended genomic DNA sequences of HLA class I and II genes to be routinely submitted to the IPD-IMGT/HLA Database. Submission of these extended regions, introns and untranslated regions (UTR), led to the realisation of the abundance of intronic variation.

The Anthony Nolan Research Institute has used Single Molecule Real-Time (SMRT) DNA sequencing to show improvements in haematopoietic stem cell transplants (HSCT) outcomes.[3] HSCT can be used to treat haematological malignancies. Before transplants the patient undergoes a conditioning routine to remove malignant cells. Donor stem cells are then engrafted into the patient to repopulate the patient's immune system. Finding a donor for a patient involves aiming to match the DNA sequence for six HLA genes: HLA-A, -B, -C, -DRB1, -DQB1 and -DPB1. Due to inheriting a copy of each gene from each parent it is known as a 12/12 match. Here at Anthony Nolan we aim for an ultra-high resolution (UHR) 12/12 match. UHR refers to typing the donor and patient to an allelic level by sequencing the full length of the gene, 5´ to 3´ UTR.

Here we describe a novel allele, HLA-C*03:04:01:47, identified in a healthy individual of northern European ethnicity (AN300334). Full-length amplicons of HLA-A, -B and -C were generated by polymerase chain reaction (PCR) and then sequenced using Pacific Biosciences' SMRT technology (Pacific Biosciences, Menlo Park, California).[4] Samples were barcoded with unique primer pairs which allowed multiple samples to be pooled together following amplification. Each sample was then identified retrospectively using the unique barcode within the sequence generated. Analyses were completed using in-house software to generate typing results (Anthony Nolan Research Institute, London, UK). The novel allele sequence was repeated from the PCR stage for in-house confirmation.

C*03:04:01:47 varies from C*03:04:01:01 by a single nucleotide polymorphism (SNP) within the 3´ UTR at genomic DNA position 2988 T>C (Figure 1). This SNP is not thought to affect primary structure, expression or function of the protein. The position 2988 is dimorphic with T seen most frequently.

The extended typing of AN300334 is: HLA-A*01:01:01:01, 31:01:02:01; -C*03:04:01:47, 07:01:01; -B*08:01:01:01, 40:01:02; -DRB1*03:01:01, 15:01:01; -DQB1*02:01:01, 06:02:01; -DPB1*04:01:01, 04:01:01:01. It is likely that this novel allele is present on the haplotype HLA-A*31:01:02:01~C*03:04:01:47~B*40:01:02~DRB1*15:01:01~DQB1*06:02:01 by virtue of the second haplotype being the ancestral AH8.1 haplotype.[5]

This sequence has been submitted to the European Nucleotide Archive (ENA) and has been assigned the accession number LR794125. The sequence was then submitted to IPD-IMGT/HLA Database and named HLA-C*03:04:01:47 by the World Health Organisation Nomenclature Committee for Factors of the HLA System in April 2020.[6]

References:

1. Horton R, Wilming L, Rand V, et al. Gene map of the extended human MHC. *Nat Rev Genet.* 2004;5(12):889-899.
2. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res.* 2020;48(D1):D948-D955.
3. Mayor NP, Hayhurst JD, Turner TR, et al. Recipients Receiving Better HLA-Matched Hematopoietic Cell Transplantation Grafts, Uncovered by a Novel HLA Typing Method, Have Superior Survival: A Retrospective Study. *Biol Blood Marrow Transplant.* 2019;25(3):443-450.
4. Mayor NP, Robinson J, McWhinnie AJ, et al. HLA Typing for the Next Generation. *PLoS One.* 2015;10(5):e0127153.
5. Degli-Esposti MA, Leaver AL, Christiansen FT, Witt CS, Abraham LJ, Dawkins RL. Ancestral haplotypes: conserved population MHC haplotypes. *Hum Immunol.* 1992;34(4):242-252.
6. Marsh SG, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens.* 2010;75(4):291-455.