

GENOMIC DNA SEQUENCES OF HLA CLASS I ALLELES GENERATED USING MULTIPLEXED BARCODES AND SMRT® DNA SEQUENCING TECHNOLOGY

Neema P Mayor^{1,2}, James Robinson^{1,2}, Alasdair McWhinnie¹, Swati Ranade³, Kevin Eng³, Will P Bultitude¹, William Midwinter¹, Brett Bowman³, Patrick Marks³, Henny Braund¹, J Alejandro Madrigal^{1,2}, Katy Latham¹, Steven GE Marsh^{1,2}

¹ Anthony Nolan, Royal Free Hospital, London UK; ² UCL Cancer Institute, Royal Free Campus, London, UK; ³ Pacific Biosciences, Menlo Park, CA, USA.

INTRODUCTION

- Allelic-level resolution HLA typing is known to improve survival prognoses post Unrelated Donor (UD) Haematopoietic Stem Cell Transplantation (HSCT).
- Currently, many commonly used HLA typing methodologies are limited either due to the fact that ambiguity cannot be resolved or that they are not amenable to high-throughput laboratories.
- Pacific Biosciences' Single Molecule Real-Time (SMRT®) DNA sequencing technology (Figure 1) enables sequencing of single molecules in isolation and has read-length capabilities to enable whole gene sequencing for HLA.
- DNA barcode technology labels samples with unique identifiers that can be traced throughout the sequencing process.
- The use of DNA barcodes means that multiple samples can be sequenced in a single experiment but data can still be attributed to the correct sample.
- Here we describe the results of experiments that use DNA barcodes to facilitate sequencing of multiple samples for full-length HLA class I genes (known as multiplexing).

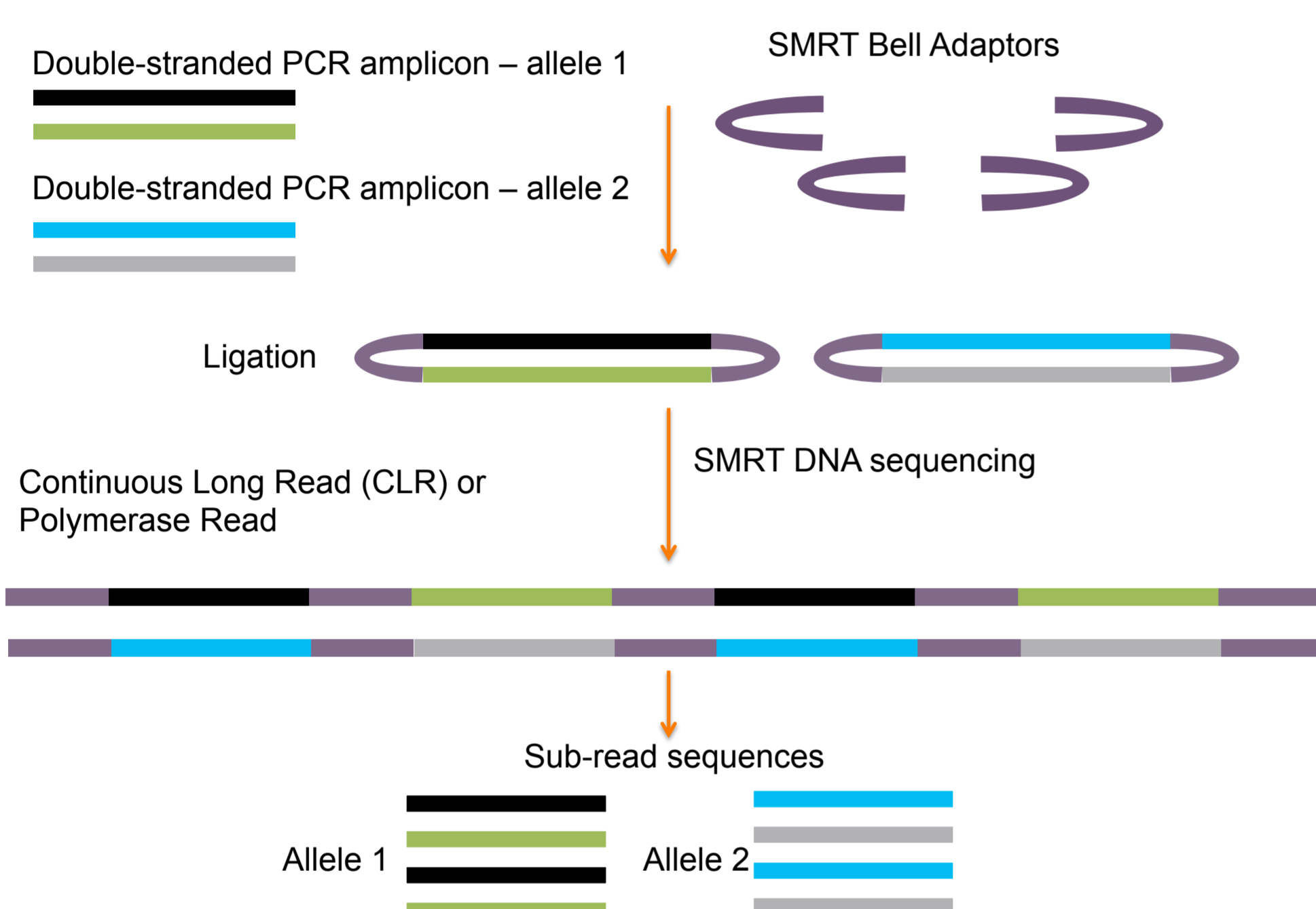


Figure 1: Single Molecule Real-Time (SMRT) DNA sequencing method. SMRTbell adaptors are ligated onto the ends of a blunt-ended PCR amplicon to facilitate continuous sequencing of both strands of the amplicon. The entire sequence generated may include multiple copies of the sense and anti-sense strands of the PCR amplicon in a single read known as the Continuous Long Read (CLR). The post-sequencing bioinformatic post-processes are able to break down the CLR into shorter sub-reads, which encompass the sequence of one strand of the amplicon. These sub-reads can then be compared and used to create a consensus sequence.

MATERIALS & METHODOLOGY

- Two experiments were set up to test the multiplexing capabilities of SMRT DNA sequencing; each experiment were sequenced in a single sequencing reaction:
 - **Experiment 1:** 20 DNA samples amplified for HLA-A only;
 - **Experiment 2:** Eight DNA samples amplified for HLA-A, -B and -C
- DNA samples were selected so as to include as many HLA class I alleles as possible that are commonly seen in the UK (North-West European Caucasoid).
- Samples with two alleles for a particular locus that differed only by one or two nucleotides, and homozygous samples were included to establish accuracy.
- PCR amplification strategies targeted whole gene amplification using gene-specific primers tagged with unique DNA barcodes (Figure 2).
- SMRT DNA sequencing was performed according to Pacific Biosciences' recommended protocols on a PacBio RS II sequencing system.
- HLA allele assignment was performed using in-house methods.
- Genomic sequences were used as references where possible. CDS sequences were used for those alleles without genomic references and extended comparisons made to their closest related allele with a genomic sequence.
- Reference sequences used for comparisons are available on the IMGT/HLA database.

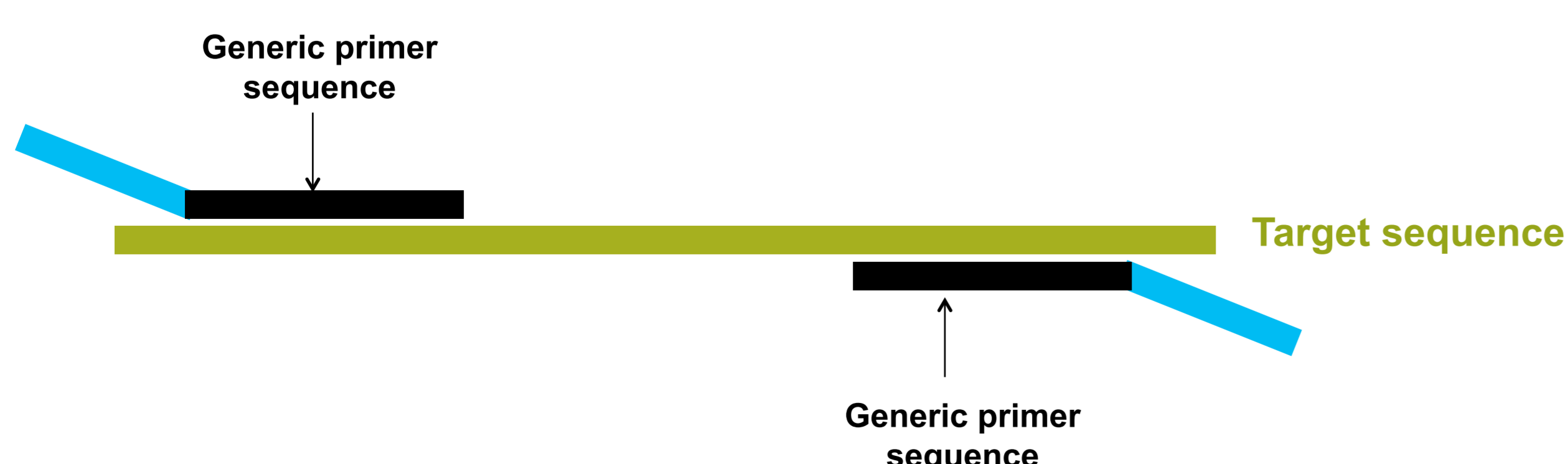


Figure 2: The principle behind DNA barcode technology. Generic primers (black) are designed to enable amplification of a specific target region (green). A specific, unique DNA sequence (the DNA barcode, blue) is then added onto the 5' end of the primer prior to manufacture. The unique DNA barcode is added on to each amplicon during each round of PCR cycling. The number of barcode-labelled primers required will be dependent on the degree of multiplexing.

RESULTS

EXPERIMENT 1:

- The mean number of reads was 155 (range 34-433).
- 39/40 possible HLA-A alleles were assigned HLA types that immediately matched our expectations.
- One sample was reported to have a single nucleotide substitution in intron 7 (gDNA 2770A>G) for the HLA-A*30:02 allele – This novel polymorphism was confirmed using Sanger sequencing and the new allele has been designated the name HLA-A*30:02:02 (Figure 3).
- The overall success rate for this experiment was 100%.

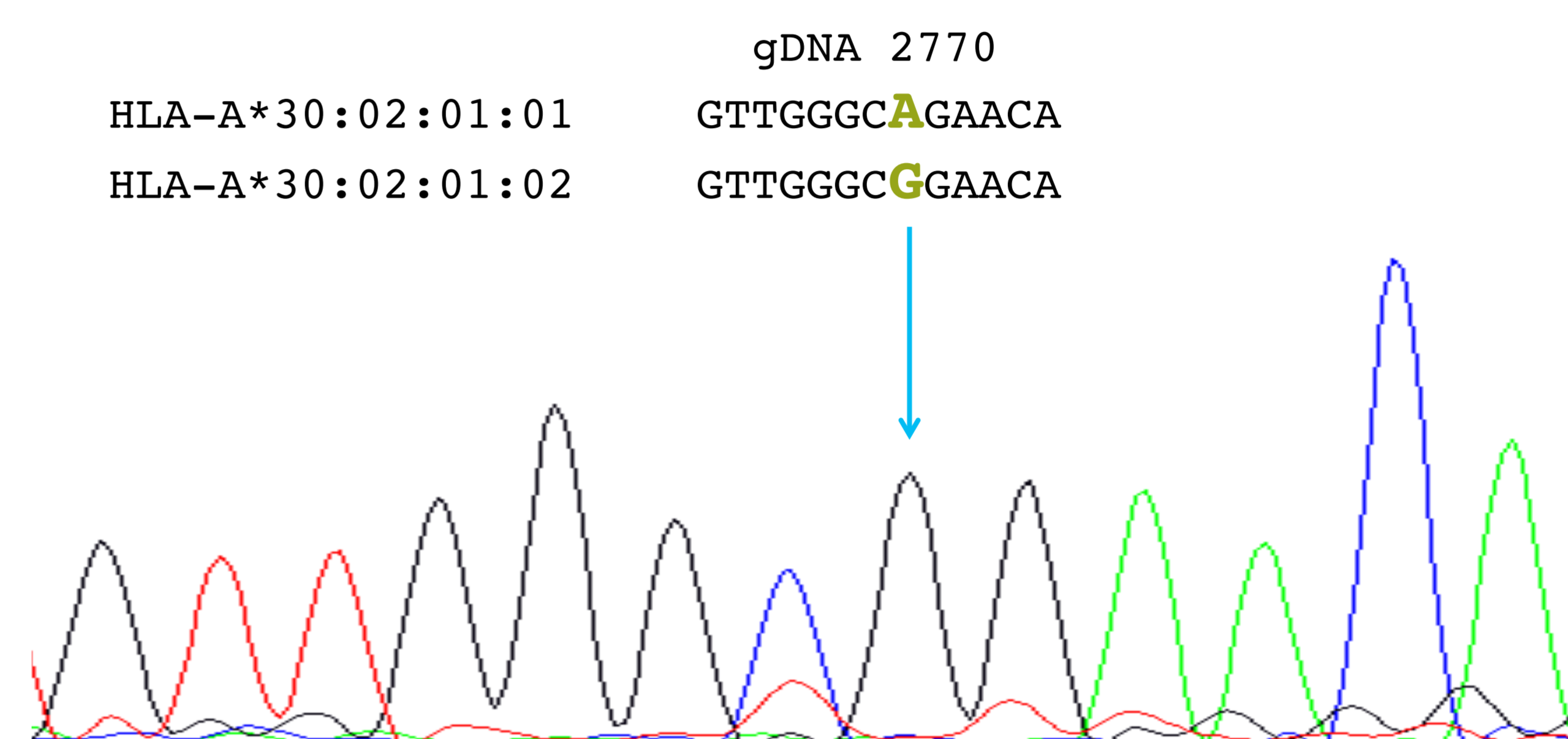


Figure 3: Sanger sequencing electropherogram confirming the novel intron 7 variant gDNA 2770A>G in the new allele HLA-A*30:02:01:02

EXPERIMENT 2:

- The mean number of reads was 149 (range 50-360).
- 46/48 possible HLA-A alleles were assigned HLA types that immediately matched our expectations.
- One sample was reported to have two intronic single nucleotide substitutions within an HLA-C*07:04:02 allele; intron 3 (gDNA 1333T>C) and intron 6 (gDNA 2586T>C) when compared to the reference sequence.
- These novel polymorphisms were confirmed in-house using Sanger sequencing as well as by the original submitter of this allele sequence to the IMGT/HLA Database on the original source material (Figure 4).
- As these errors have been confirmed on the original DNA sample from which the genomic sequence was generated, the reference sequence for HLA-C*07:04:02 has been corrected on the IMGT/HLA database.
- A second sample showed a discrepancy in a G homopolymer region of the 3' UTR of HLA-B*55:01:01 as compared to the genomic reference sequence of this allele.
- Sanger sequencing of this sample confirmed the presence of a G 9-mer as per the reference sequence.
- Further analysis of several different DNA samples containing the B*55:01:01 allele showed correctly identified the 9-mer and assigned HLA types.
- Further optimisation of the bioinformatics used in generating the consensus sequence is required to accurately report the correct number of reports in this region.
- The overall success rate for this experiment was 98%.

CONCLUSIONS

- Pacific Biosciences' SMRT DNA sequencing methodology in combination with DNA barcode technology is amenable for use with full length HLA class I gene sequencing.
- Both multiplexing strategies produced sufficient numbers of reads to enable HLA allele calling and Bioinformatic processes were able to correctly attribute the HLA types to the correct sample based on their DNA barcode.
- The number of reads achieved for certain alleles suggest that multiplexing a greater number possible with both strategies.
- The ability to provide definitive allele level resolution HLA typing in a single method could result in significant clinical implications, improving survival and lowering the risk of post transplant complications.

Poster courtesy of Anthony Nolan.

For Research Use Only. Not for use in diagnostic procedures.

Pacific Biosciences assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of Pacific Biosciences products and/or third party products. Please refer to the applicable Pacific Biosciences Terms and Conditions of Sale and to the applicable license terms at <http://www.pacificbiosciences.com/licenses.html>. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, and Iso-Seq are trademarks of Pacific Biosciences in the United States and/or certain other countries. All other trademarks are the sole property of their respective owners.