A MULTIPLEXED TYPING STRATEGY FOR THE HLA CLASS II GENES HLA-DRB1, -DQB1 AND -DPB1 USING DNA BARCODES AND SMRT® DNA SEQUENCING

saving the lives of people with blood cancer

Neema P Mayor[1,2], Nicola Brosnan[1], William Midwinter[1], Alasdair JM McWhinnie[1], Thomas R Turner[1], James Robinson[1,2], Will P Bultitude[1,2], J Alejandro Madrigal[1,2], Katy Latham[1], Steven GE Marsh[1,2]

1 Anthony Nolan Research Institute, London, United Kingdom; **2** UCL Cancer Institute, University College London, United Kingdom

INTRODUCTION

- HLA typing to the highest possible resolution has important applications in transplantation, disease association studies and drug hypersensitivity testing
- Single Molecule Real-Time (SMRT) DNA sequencing generates ultra-high resolution HLA typing in a single experiment due to long-read length capabilities and the capacity to sequence amplicons in isolation
- Work to establish an HLA class I typing pipeline using SMRT DNA sequencing has been described previously
- The HLA class II genes are considerably larger than class I, are significantly less well characterised outside of the region encoding the peptide binding groove and there are limited numbers of genomic sequences available (Table 1). Coupled together, these factors make an HLA class II typing strategy more complicated
- In this study we determined the feasibility of using SMRT sequencing for HLA class II typing in a high through-put clinical laboratory

Table 1: The number of genomic sequences available for the classical HLA loci in comparison to the number of alleles of each locus. IMGT/HLA Database Release 3.24.0, April 2016.

	Number of alleles	Number of genomic sequences	% alleles with genomic sequences
HLA-A	3399	219	6.5
HLA-B	4242	337	7.9
HLA-C	2950	301	10.2
HLA-DRB1	1883	40	2.1
HLA-DQB1	911	27	3.0
HLA-DPB1	644	12	1.8

METHOD

- 151 DNA samples were selected for testing using the following inclusion criteria:
  - High molecular weight DNA was available
  - Previous HLA class II typing for HLA-DRB1, -DQB1 and DPB1 by sequence-specific oligonucleotide probing (SSOP), PCR using sequence-specific primers (PCR-SSP), sequencing-based typing (SBT), or a combination of these methods
  - Eight B-lymphoblastoid cell-lines (B-LCLs) that have previously been sequenced for their entire MHC and thus have genomic reference sequences available for all alleles present were included
- Samples underwent gene-specific PCR for each HLA gene. Regions targeted included:
  - HLA-DRB1: Exons 2-3
  - HLA-DQB1: Exons 2-5
  - HLA-DPB1: Exons 2-4
- PCR amplicons were labelled with DNA barcodes at the PCR stage to enable sample identification to be maintained
- Four 48-sample, three-gene libraries were prepared and sequenced according to in-house protocols
- HLA types were assigned using Anthony Nolan's in-house software, {AT} toolset (See Poster 90)
- HLA typing was analysed to a CDS level due to vast amount of variation in microsatellite and homopolymer regions observed, and the lack of available reference data for HLA class II intronic regions

Table 2: Results from HLA typing data analysis.

Library	No. alleles expected	No. alleles observed (%)	No. unresolved discrepancies (%)	No. allele drop out (%)
Library 1	247§	246 (99.6)	1 (0.4)	0 (0)
Library 2	258	255 (98.8)	0 (0)	3 (1.2)
Library 3	241§	240 (99.6)	0 (0)	1 (0.4)
Library 4	246	244 (99.2)	1 (0.4)	1 (0.4)
Total	992	985 (99.3)	2 (0.2)	5 (0.5)

§ Previous typing at intermediate resolution indicated two samples were homozygous at a single locus. These were subsequently identified as being heterozygous after SMRT DNA sequencing. These numbers have been corrected accordingly.

RESULTS

- All libraries were successfully sequenced with loading efficiencies between 56-62%
- Average number of reads for a heterozygous allele was 182 (range 29-500), with 95.9% of allele calls based on a read depth >60
- Average quality value (QV) scores across the four libraries was 41.96 (max. possible QV score: 42.00)
- 992 HLA class II alleles were expected across four libraries (Table 2)
- 98.7% of observed alleles immediately matched their expected types
- The eight B-LCLs were correctly identified and found to be homozygous at the three HLA class II loci in each of the libraries tested
- Of the 13 (1.3%) initial discrepancies:
  - Three allele calls corrected previous typing where there was incorrect SSOP bead assignment (one sample) or polymorphisms were in regions of the gene not tested by the technology at that time (two samples)
  - Three sequences showed differences to known CDS reference sequences; one where a nucleotide substitution suggests a novel allele and two further samples where differences were seen in homopolymer regions
  - Two samples previously typed as homozygous at high-resolution were found to be heterozygous at ultra-high resolution (previous type HLA-DPB1*04:02, *04:02, SMRT typing HLA-DPB1*04:02:01:01, *04:02:01:02; previous type HLA-DRB1*03:01, *03:01, SMRT typing HLA-DRB1*03:01:01:01, *03:01:01:03)
  - Five samples were found to have allele drop-out. Further investigation has shown this to be attributable to preferential amplification of certain alleles once DNA barcodes were added to primer sequences. All allelic combinations have been successfully sequenced using alternative DNA barcodes
- The overall success rate of this SMRT DNA sequencing method for HLA class II typing was 99.3%
- Ambiguous types resulting from this strategy based on currently known polymorphisms are listed in table 3

Table 3: HLA class II ambiguities using this strategy. Current ambiguities are limited to exon 1 (cleaved off and not part of the cell surface molecule), exon 4 (transmembrane region) and exon 5 (cytoplasmic tail).

Loci	Allele 1	Allele 2	Location of known polymorphism
HLA-DRB1	04:07:01	04:92	Exon 4
HLA-DRB1	04:10:01	04:10:03	Exon 4
HLA-DRB1	08:01:01	08:77	Exons 4 & 5
HLA-DRB1	09:01:02	09:21	Exon 4
HLA-DRB1	12:01:01	12:10	Exon 1
HLA-DQB1	06:01:01	06:01:15	Exon 1
HLA-DPB1	13:01:01	107:01	Exon 1
HLA-DPB1	02:01:02	02:01:19	Exon 5

CONCLUSIONS

- HLA class II typing using SMRT DNA sequencing generates highly accurate data and is a viable option for a high through-put laboratory
- Five cases of allele drop out could all be attributed to DNA barcode effects and have been resolved in further experiments
- Improvements in SMRT sequencing chemistry and refinement of the in-house bioinformatic settings have helped to significantly reduce sequencing discrepancies in homopolymer regions seen in previous studies
- This strategy allows for the majority of known ambiguous combinations to be resolved, with only a single ambiguous pairing causing an amino acid difference in the mature protein
- Ultra-high resolution HLA typing using this method has identified allelic differences in previously assumed homozygous samples that may have important clinical implications

Reg charity no 803716/SC038827 978PA/0416

None of the authors have any conflict of interest to declare