A PCR free, whole gene, long read next generation sequencing method for human neutrophil genotyping.

Tom Browne[1], Pearl E. Hazell[2, 3], Leigh Keen[1] and Anthony Poles[1].
[1]Histocompatibility and Immunogenetics, NHSBT-Filton, UK. [2]Bristol Institute for Transfusion Sciences, NHSBT, UK. [3]School of Biochemistry, University of Bristol, UK.
Correspondence: tom.browne@nhsbt.nhs.uk

Background

Clustered regularly interspaced short palindromic repeats (CRISPR) associated (*Cas*) genes can be used to cleave specific regions of interest from genomic DNA sequences for targeted next generation sequencing (NGS). A method to sequence the Human Neutrophil Antigen (HNA) genes, *FCGR3B, CD177, SLC44A2, ITGAM* and *ITGAL* utilising the CRISPR-Cas9 and Oxford nanopore technologies (ONT) MinION was designed and evaluated for use in the laboratory for HNA genotyping.

Methods

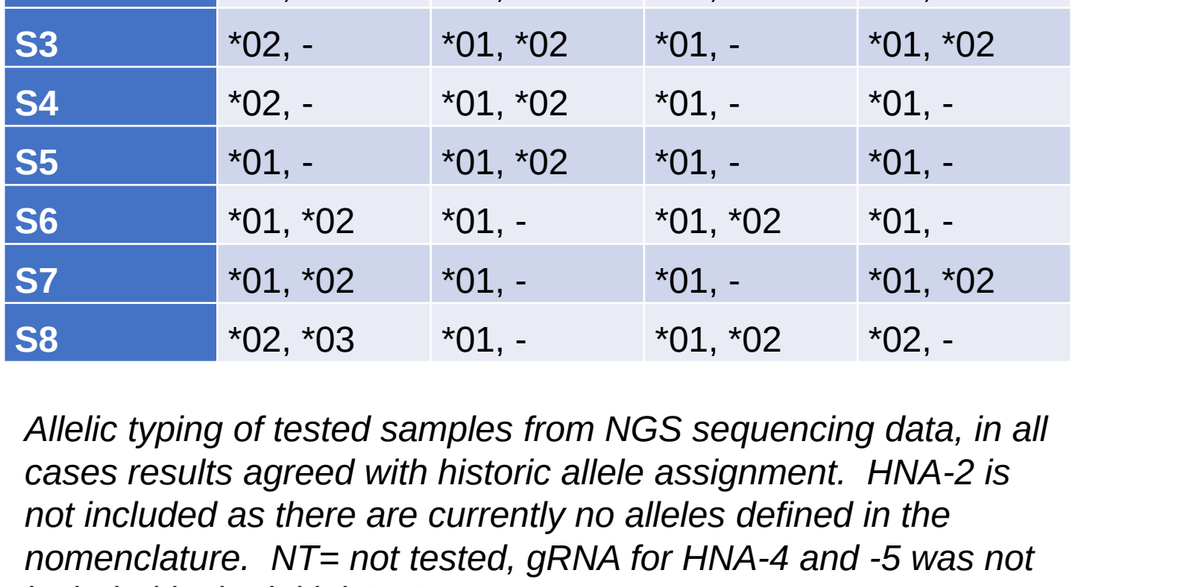
Suitable targets for CRISPR-Cas9 cleavage approximately 10kb apart were identified for each HNA gene, using the on-line tool CHOPCHOP v3 (https://chopchop.cbu.uib.no/), covering the known region(s) of interest for HNA genotyping. DNA was extracted from eight blood donor samples (MagnaPure, Roche, Switzerland) with a concentration >200ng/µL. Sequencing adapters were ligated to the cleaved ends and the targeted regions sequenced on a R9.4 MinION flow cell (ONT, UK). Basecalling was performed by MinION software (v21.06.0), aligned to hg38 with minimap2 and visualised with the Integrated Genomics Viewer (Broad Institute, USA).

Defined alleles by CRISPR-Cas9 NGS

Sample	HNA-1 (FCGR3B)	HNA-3 (SLC44A2)	HNA-4 (ITGAM)	HNA-5 (ITGAL)
S1	*02, -	*01, -	NT	NT
S2	*02, -	*01, *02	*01, -	*01, *02
S3	*02, -	*01, *02	*01, -	*01, *02
S4	*02, -	*01, *02	*01, -	*01, -
S5	*01, -	*01, *02	*01, -	*01, -
S6	*01, *02	*01, -	*01, *02	*01, -
S7	*01, *02	*01, -	*01, -	*01, *02
S8	*02, *03	*01, -	*01, *02	*02, -

Allelic typing of tested samples from NGS sequencing data, in all cases results agreed with historic allele assignment. HNA-2 is not included as there are currently no alleles defined in the nomenclature. NT= not tested, gRNA for HNA-4 and -5 was not included in the initial RNP.

Aligned sequencing against the human genome



Sequencing "reads" obtained by the Cas9 HNA protocol are indicated by the horizontal black lines and show both targeted and 'off-target' sequencing reads. Chromosome identity (number, letter) is indicated on the right. The HNA genes 1-5 are located on chromosomes 1, 16 and 19, indicated by the green circles. All other vertical lines are 'off target' sequencing.

The CRISPR-Cas9 Ribonucleoprotein (RNP)

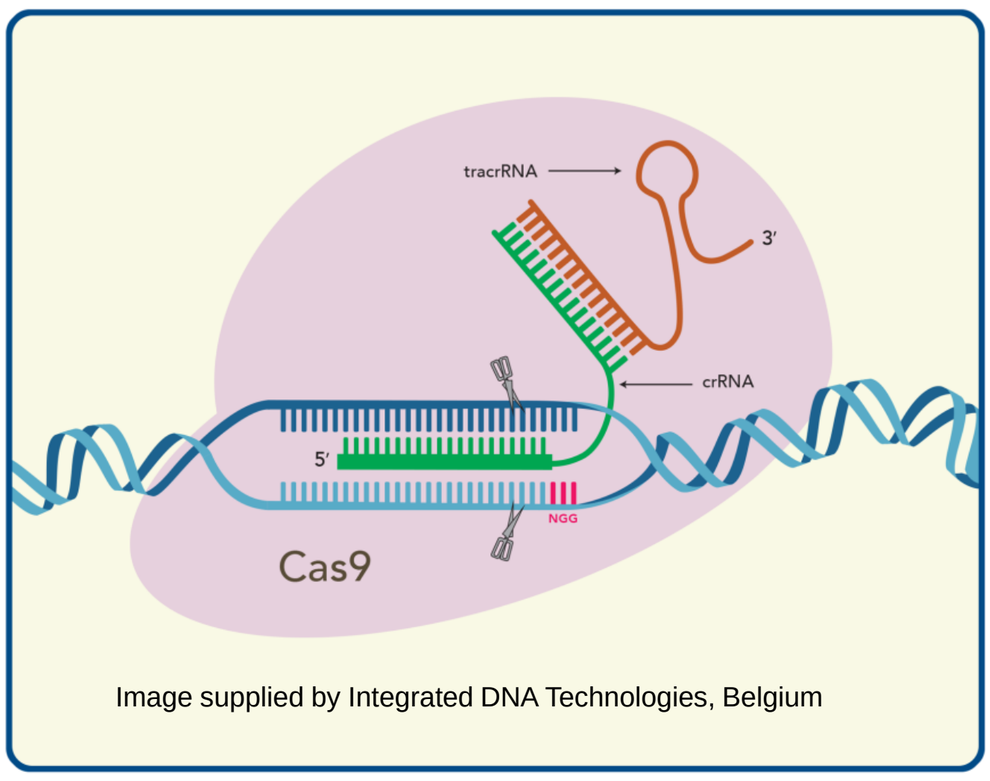


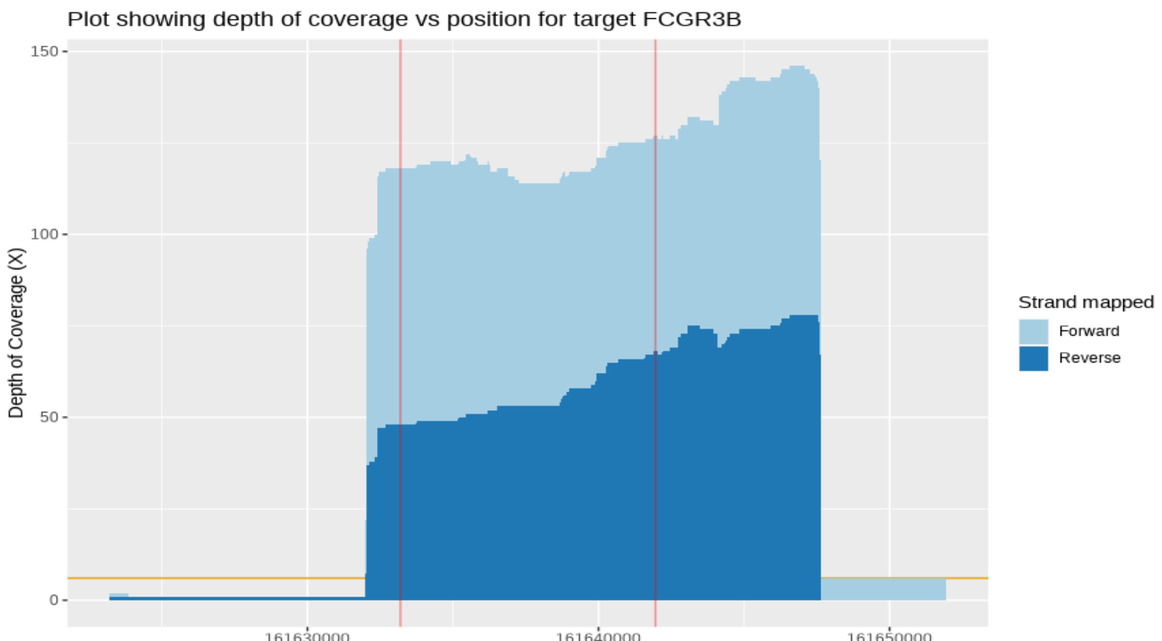
Image supplied by Integrated DNA Technologies, Belgium

For each gene 20nt crRNA's were identified as 'target sites' for Cas9 cleavage to 'cut out' the region of interest

Results

Alleles for HNA-1, -3, -4 and -5 were assigned as defined in the most recent ISBT nomenclature. Allelic definition for all eight samples tested were identical to the historic genotyping results. A significant variation in sequencing depth (both genes and samples) was observed (5x-185x). Mean sequencing depth for each gene was 64x (*FCGR3B*), 30x (*CD177*), 70x (*SLC44A2*), 23x (*ITGAM*) and 56x (*ITGAL*).

Example of sequence coverage for HNA-1



Coverage and depth of the HNA-1 (FCGR3B) gene which is indicated by the vertical red lines. The cleavage sites either side of the gene can be clearly seen as well as the distribution of sense (light blue) and anti-sense (dark blue) sequencing data.

Conclusion

The method offers many potential benefits. It is relatively quick ('hands-on' time <2 hours and results in two working days), and requires standard equipment (e.g. pipettes, thermal cycler) used in the molecular laboratory. The method doesn't require an initial PCR amplification step as it is performed on 'native' genomic DNA and therefore does not incorporate any associated bias. This would be particularly useful in analysis of HNA-1alleles where significant copy number variation has been observed. Parallel testing of multiple samples would increase throughput and reduce cost per sample (single sample >£200/sample, two days) making it more suitable for routine laboratory genotyping. Further work to optimise the DNA template and reduce 'off-target' sequencing is ongoing.