Cas9 targeted long-read nanopore sequencing of *ABCA4* and the *OPN1LW/OPN1MW* gene array for investigation of inherited retinal disease

Gavin Arno[1,2,3], Neringa Jurkutė[1,2], T. Christian Boles[4], Michel Michaelides[1,2], Omar A Mahroo[1,2], Andrew R Webster[1,2]

[1]Moorfields Eye Hospital NHS Foundation Trust, London, UK; [2]Institute of Ophthalmology, University College London, London, UK;
[3]North Thames Genomic Laboratory Hub, Great Ormond Street Hospital for Children, London, UK; [4]Sage Science, Inc., Beverly, MA, USA

Background and purpose

- Short, paired-end sequencing (next generation sequencing, NGS) enables interrogation of up to ~95% of the human genome.
- Several clinically relevant genes for inherited retinal disease (e.g.: *RPGR*-retinopathy*, OPN1LW/OPN1MW* associated cone dysfunction and blue-cone monochromacy) and variants (e.g.: complex rearrangements) cannot be resolved by NGS technology including whole genome sequencing (WGS).
- Long-read sequencing (LRS) has the capability to enable read-through of complex regions by anchoring reads to unique sequences flanking low-complexity regions facilitating confident assembly and annotation. LRS remains relatively high in cost per base compared to NGS, therefore, targeting of LRS to regions of interest is important to increase coverage and reduce cost.
- Nanopore LRS with the Oxford Nanopore Technologies MinION system (ONT-LRS) has the capability of sequencing megabase long molecules of DNA as single reads, however sample manipulation and library preparation lead to degradation of high molecular weight (HMW) DNA and limit read length.
- We sought to investigate the use of **Ca**s9 **t**argeting of **ch**romosomal segments (CATCH, **Figure 1**) with ONT-LRS for a WGS-intractable suspected *ABCA4*-retinopathy case (individual 1) and to sequence the NGS-intractable *OPN1LW/OPN1MW* gene array at chrXq28 from control 1 (female, heterozygous) and control 2 (male, hemizygous).

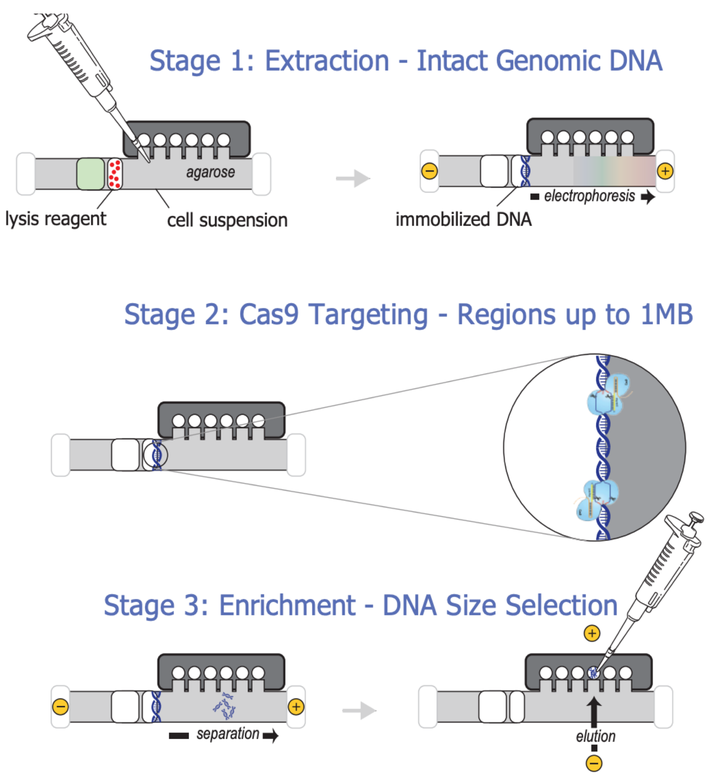

Figure 1: Schematic representation of SageHLS CATCH workflow (Courtesy of Sage Science).

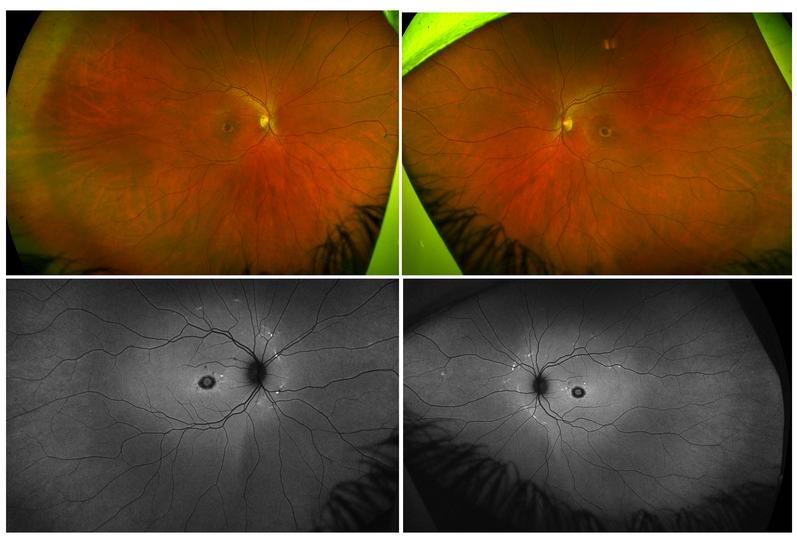

Figure 2: Ultra-widefield pseudocolour and autofluorescence fundus imaging (Optos, Dunfermline, UK) of individual 1 showing typical features of ABCA4-retinopathy.

Method

- Individual 1 underwent WGS as a singleton as part of the UK's 100,000 genomes project followed by clinical variant analysis.
- CATCH (**Figure 1**) was performed for *ABCA4* (individual 1) and the *OPN1LW/OPN1MW* array (control 1 and 2) using freshly purified white blood cells from ACD stabilised blood. CRISPR guide RNAs flanking the region of interest were optimised by cleaving PCR amplicons and agarose gel analysis. Efficient guide pairs were used for CATCH using the SageHLS system (SAGE science).
- Resultant eluates containing HMW DNA fragments were quantified by qPCR for the gene of interest (*ABCA4* or *TEX28*) to determine the enrichment and copy/uL (up to ~3-4000 copies/uL).
- Chosen elution fragments were concentrated for ONT-LRS using AMPure XP magnetic beads and used for gDNA library preps using the SQ-LSK109 or SQ-LSK110 kit followed by sequencing on a MinION flowcell (R9.4.1) for approximately 24 hours.
- In addition, simple Cas9 targeted ONT-LRS (non-CATCH) for *ABCA4* (individual 1) and the *OPN1LW/OPN1MW* array (control 1 and 2) was performed on freshly extracted gDNA from EDTA stabilised whole blood using the standard ONT Cas9 Sequencing Kit (SQK-CS9109) protocol.
- Fast5 data were basecalled, mapped to GRCh38 and sorted using Guppy, MiniMap2 and SAMtools to generate indexed BAM files for visualisation using the Integrative Genome Viewer 2.5.0 (IGV) **(Figure 3, 4, 5)**.

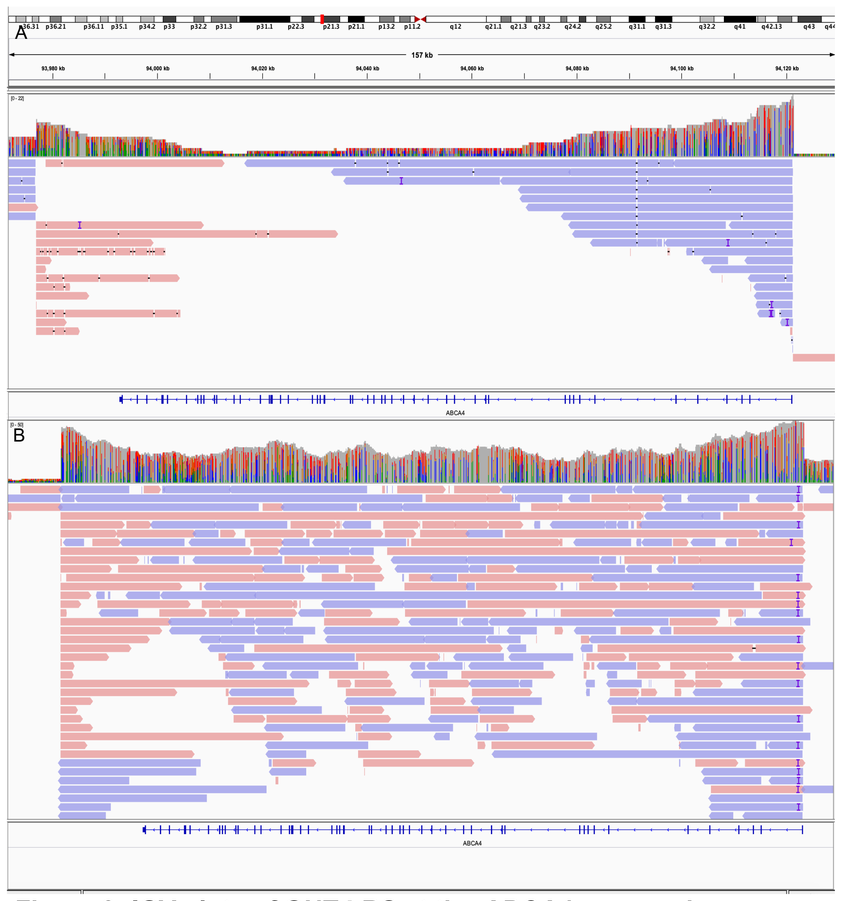

Figure 3: IGV plots of ONT-LRS at the *ABCA4* gene region. **A:** Simple Cas9 targeted ONT-LRS using CRISPR guides flanking *ABCA4* followed by standard library prep and sequencing with a MinION flowcell for 24h. Low coverage (max depth 22x at CRISPR guide sites) and read dropout in the central region (reduced to single read coverage) **B:** CATCH ONT-LRS showing even coverage (max 50x) throughout the region.

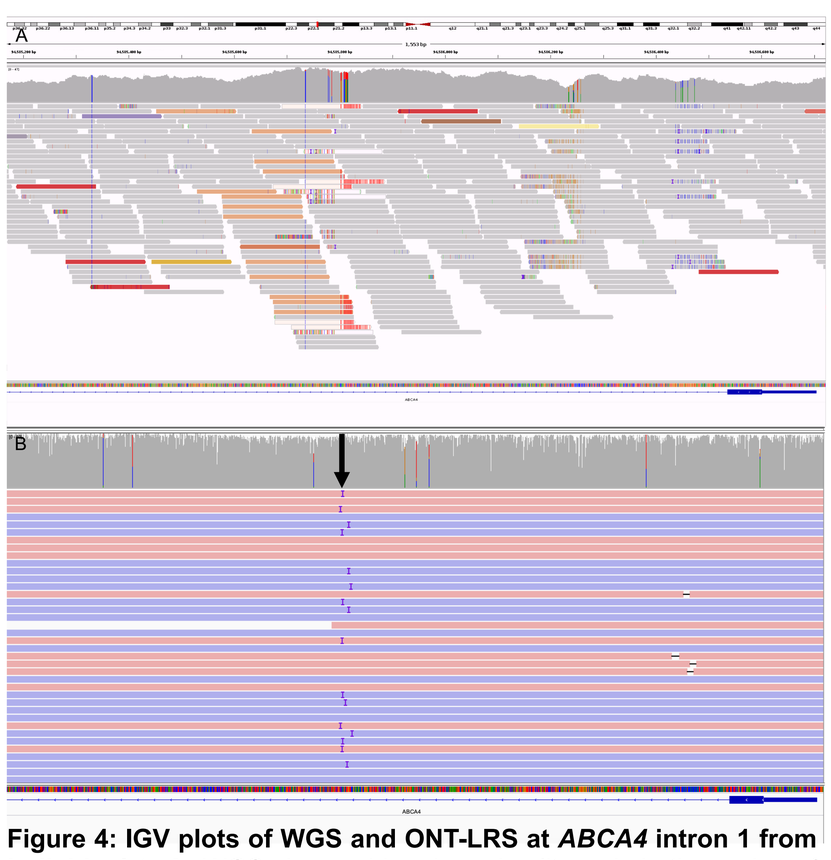

Figure 4: IGV plots of WGS and ONT-LRS at *ABCA4* intron 1 from individual 1. A: WGS shows chimeric and split reads in the region of 1:94120200-94120300. **B:** CATCH ONT-LRS showing effective detection of an insertion/translocation of approx. 1400bp into the *ABCA4* intron 1 (arrow).

Figure 5: IGV plots of WGS and ONT-LRS at the *OPN1LW/OPN1MW* array. A: Paired-end Illumina sequencing (WGS). Sequence dropout , poor mapping and no coverage of the opsin array. **B:** Simple Cas9 targeted ONT-LRS using CRISPR guides flanking the array with standard library prep and sequencing with a MinION flowcell for 24h shows poor coverage (0-6x depth). **C:** CATCH ONT-LRS showing even coverage throughout the region (max 36x depth). End to end reads generated by CATCH ONT-LRS enable haplotyping of the opsin array.

Results (*ABCA4*)

- Individual 1 was clinically diagnosed with Stargardt disease **(Figure 2).** WGS and clinical pipeline analysis revealed a single pathogenic variant: *ABCA4* c.5714+5G>A, Interrogation of the intronic and structural variant call data failed to identify a second pathogenic allele. Manual interrogation of the WGS BAM data showed a unique cluster of chimeric and split reads within intron 1 of *ABCA4 (***Figure 4**). Although these were ambiguous, it was suspected that they may represent an insertion or a translocation, but no conclusion could be drawn.
- Due to singleton WGS and unavailability of family members, no phasing was possible at that time.
- CATCH-nanopore sequencing using CRISPR-guides flanking *ABCA4* generated 337 on-target reads (1:93992834-94121148) and a maximal read depth of 50x compared to poor coverage with non-CATCH ONT-LRS **(Figure 3).**
- Interrogation of reads revealed a complex structural rearrangement (**Figure 4**) comprising insertion of 689bp from chr14q23.1 (14:59220385-59221073, and an ~800bp long interspersed nuclear element (LINE) into intron 1 of *ABCA4.*
- Splice prediction (nnsplice) showed a potential 14bp pseudoexon within the inserted sequence. The effect of this on the *ABCA4* transcript is yet to be demonstrated.
- Phasing of reads showed the two variants to be in trans.

Results (*OPN1LW/OPN1MW*)

- Sequencing of the *OPN1LW/OPN1MW* gene array following CATCH enrichment for a control female subject generated a maximal read depth of 36x **(Figure 5**), and for a control male, 25x read depth.
- Full-length reads of up to ~233kb spanning the entire region (end-to-end) were observed. These individual reads enable accurate haplotyping of the opsin arrays even in a female control subject, harbouring 1x*OPN1LW* followed by 2x*OPN1MW* on one allele and 1x*OPN1LW* followed by 4x*OPN1MW* on the trans allele.
- Comparison to WGS data for the region and non-CATCH simple Cas9 ONT-LRS showed effective coverage of the array (up to 36x depth) by CATCH ONT-LRS compared to low mapping quality (WGS) or low/absent coverage (0-6x depth) for non-CATCH Cas9 sequencing (**Figure 5**) .

Conclusions

- CATCH-nanopore sequencing is effective for cases intractable to NGS.
- We were able to generate significant enrichment for target regions compared to direct Cas9 Nanopore sequencing.
- Ultra-long reads spanning the entire targets were generated enabling phasing of distant variants *(ABCA4*) and haplotyping in a female carrying distinct opsin array haplotypes.
- This study resolved a complex structural rearrangement, missed by short-read WGS, in *ABCA4* that may lead to cryptic splicing.
- Effective sequencing of the NGS-intractable *OPN1LW/OPN1MW* gene array is possible for variant detection in males with blue cone monochromacy.

NIHR | Great Ormond Street Hospital Biomedical Research Centre
Moorfields Eye Charity
Moorfields Eye Hospital NHS Foundation Trust
NHS North Thames NHS Genomic Medicine Centre
Genomics england
FIGHT FOR SIGHT