viapath

Guy's and St Thomas' NHS
NHS Foundation Trust

Nanopore long read sequencing for detection of point mutations and structural variants

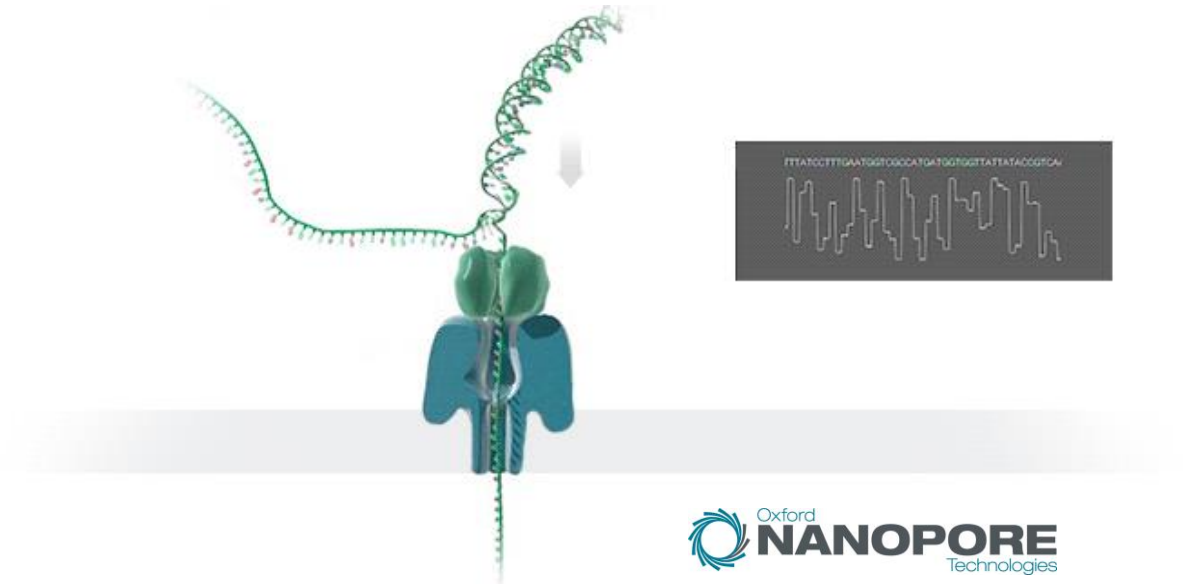
Viapath Genetics, Guy's Hospital, London

Presented by: Kezia Brown

viapath

Nanopore for long read sequencing

- Why do we need long read sequencing?
- Amplification Targets
- Laboratory work
- Analysis
- Confirmations
- Summary
- Questions



viapath

Image: NHBLI WGS Project

WGS is Not "Whole Genome" Sequencing

Four types of known unassembled regions:

1) Telomeres
2) Centromeres
3) short-arms of acrocentric chromosomes (chr13,14,15,21,22, Y)
4) large heterochromatic regions (in chr1, 9, 16, Y).

All these regions involve repeat sequences

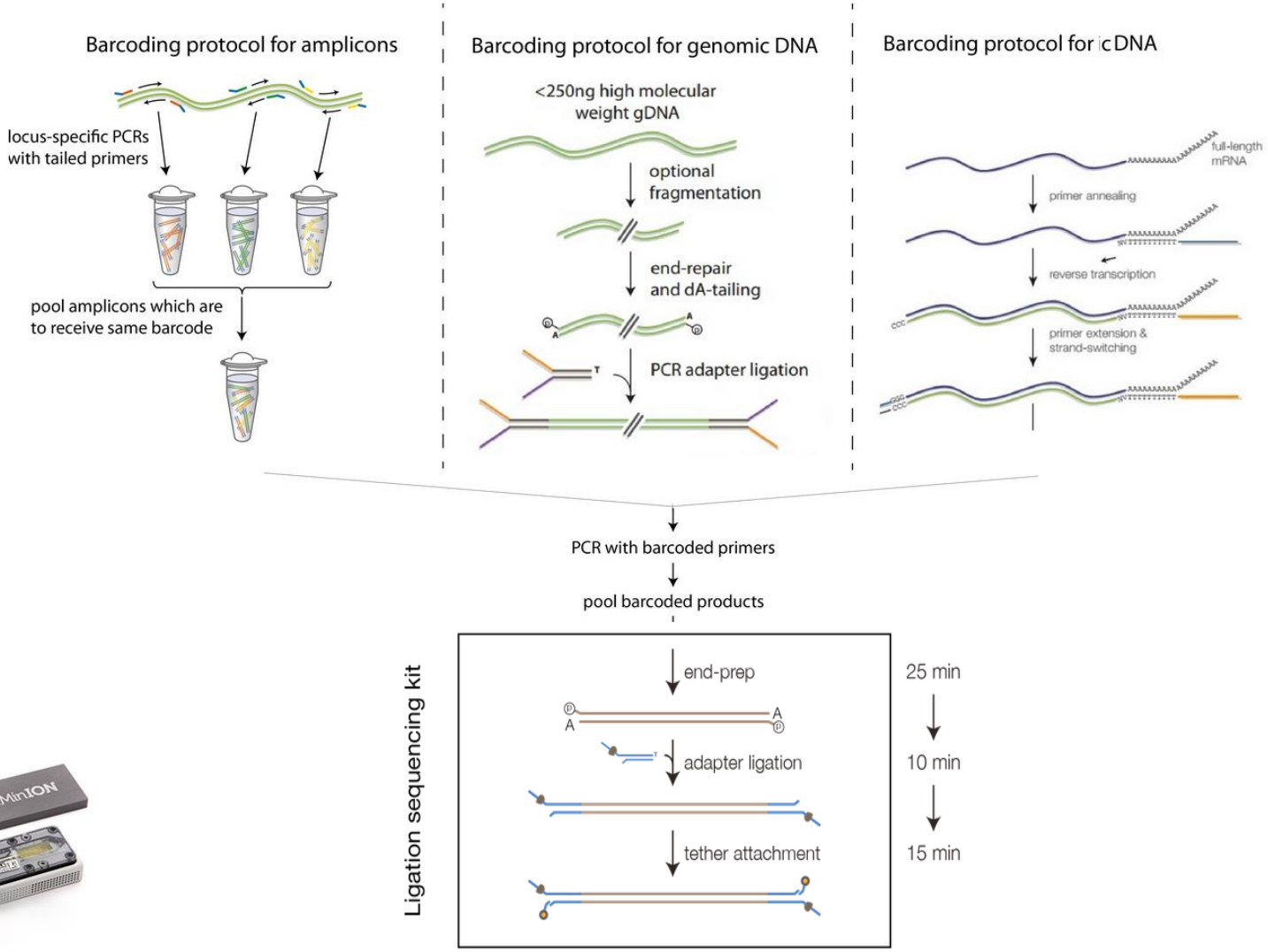
Other regions are challenging for current short read technologies:

1. Translocations in cancer
2. HLA typing
3. Trinucleotide repeat expansion
4. Structural variants
5. Pseudogenes gene paralogues
   - *CYP2D6, MYH6/7, CYP21B, PMS2 etc*.

viapath

MinION Nanopore Sequencing

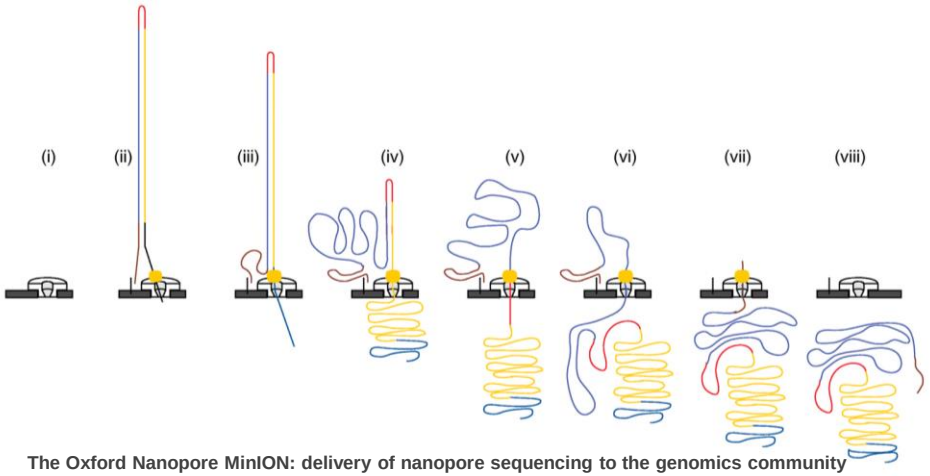
Barcoded amplicons for multiplexed processing



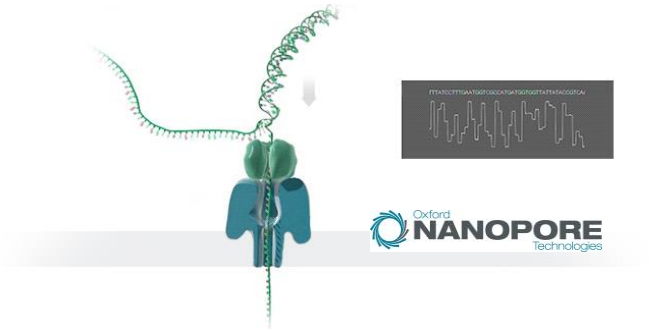
viapath

MinION Nanopore Sequencing

- Oxford Nanopore MinION uses strand sequencing
- 1D, 2D, 1D^2 kits
- Detects changes in current as different bases pass through the pore
- Basecalling performed by MinKNOW software
- Real time sequencing



The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community
Miten Jain, Hugh E. Olsen, Benedict Paten and Mark Akeson. Genome Biology 2016 17:239



Fast5 → EPI2ME → FastQ → BWA MEM → SAMtools mpileup

viapath

MinION Nanopore Sequencing

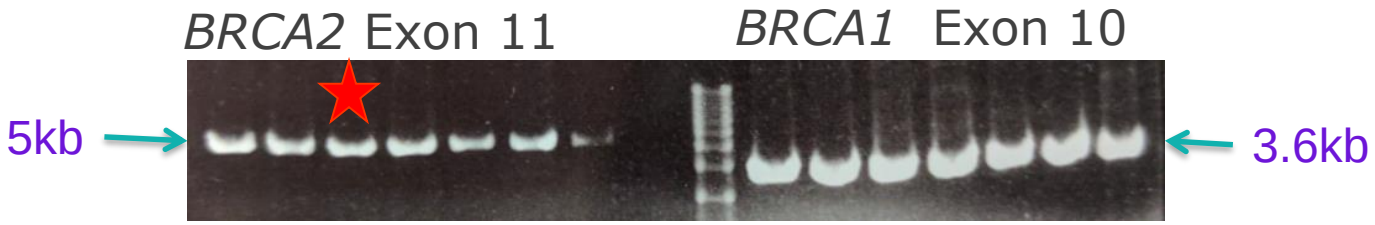
Our experiments – Long read amplicon sequencing

- We performed multiplex analysis of barcoded *BRCA1*, *BRCA2*, *SMN1*, *HLA* and *LDLR* amplicons (3.6 to 16kb)

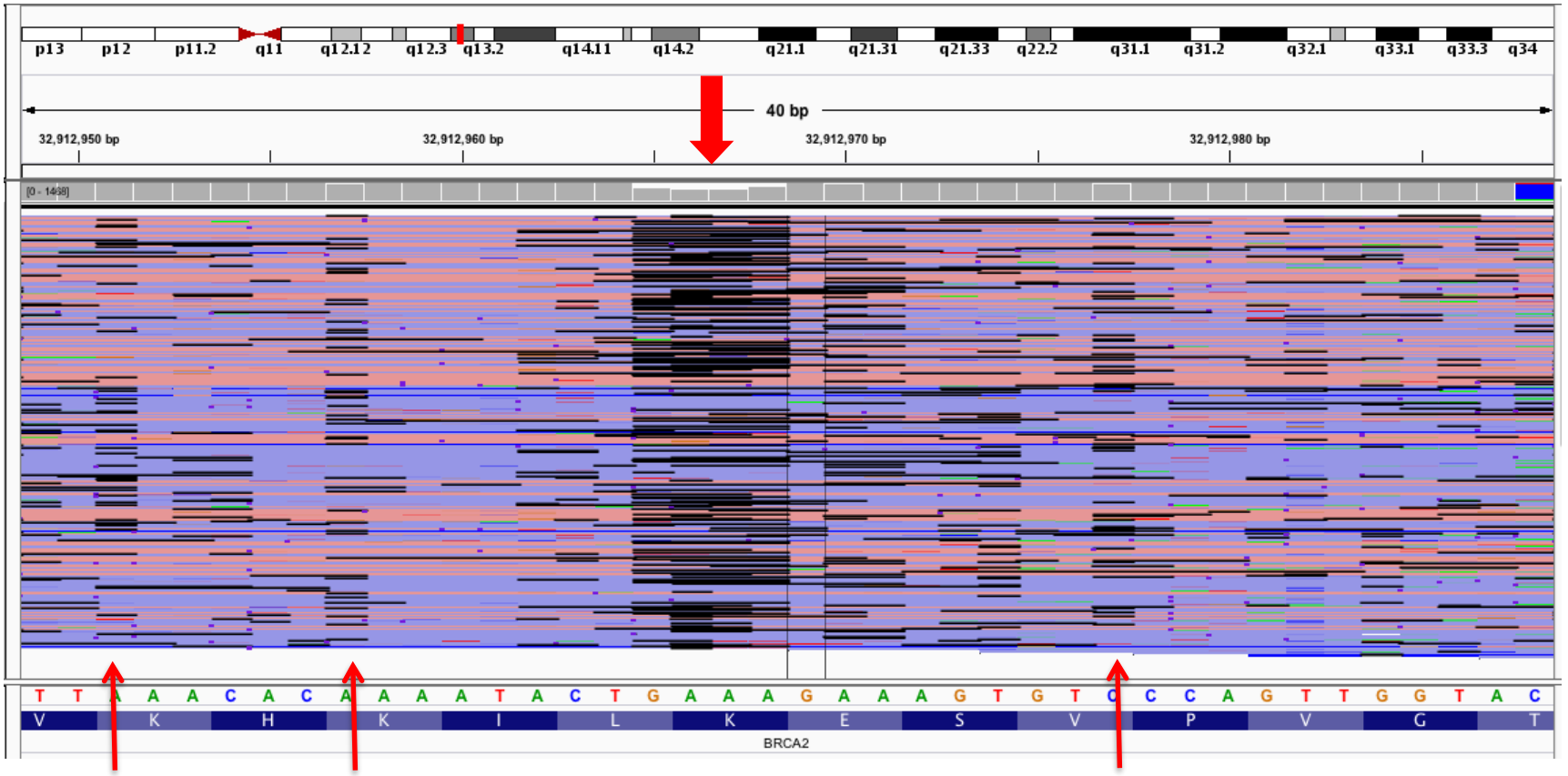
Gene	Exon(s)	Size (bp)	Variants
BRCA1	10	3688	1 point mutation, 3 deletions, 1 duplication, NA12878
BRCA2	11	5101	2 point mutations, 3 deletions, NA12878
SMN1	2 - 8	16434	2 point mutations, 2 duplications, NA12878
LDLR	2 - 8	11943	1 deletion, 2 duplications, NA12878
LDLR	8 - 15	12664	1 deletion, 1 duplication, NA12878
HLA Region		~5000	

viapath

BRCA1 and *BRCA2*



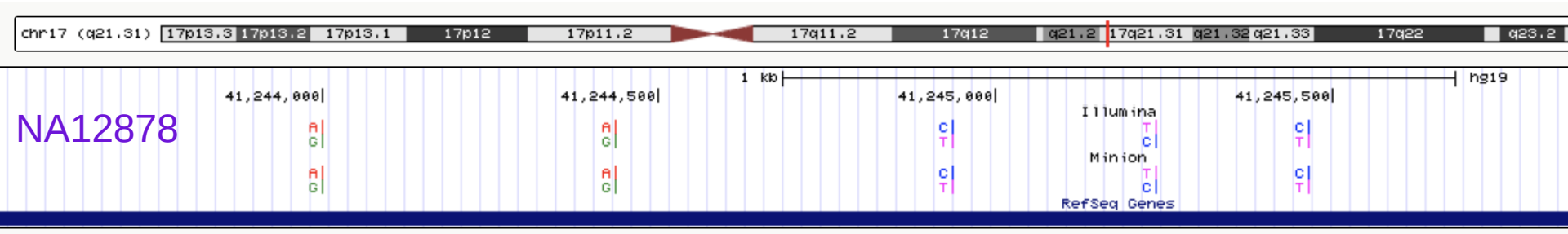
Heterozygous c.4478_4481delAAAG p.(Glu1493fs)



viapath

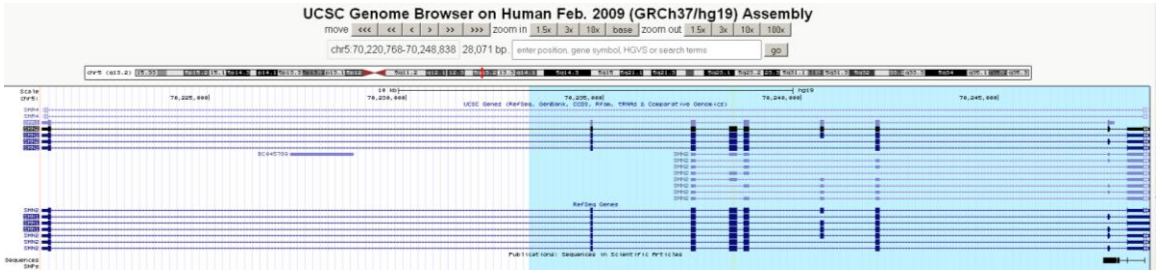
BRCA1 and *BRCA2*

Variant	Seen on Nanopore?
BRCA2	
Heterozygous c.4478_4481delAAAG p.(Glu1493fs)	Yes
Heterozygous c.6275_6276delTT p.(Leu2092fs)	Yes
Heterozygous c.5350_5351delAA p.(Asn1784fs)	Yes
Heterozygous c.4576dupA p.(Thr1526fs)	Yes
Heterozygous c.5682C>A p.(Tyr1894Ter)	Yes
BRCA1	
Heterozygous c.1961delA p.(Lys654fs)	Yes
Heterozygous c.2475delC p.(Asp825fs)	Yes
Heterozygous c.3607C>T p.(Arg1203Ter)	Yes
Heterozygous c.3400G>T p.(Glu1134Ter)	Yes
Heterozygous c.3358_3359delGT p.(Val1120Ter)	Yes

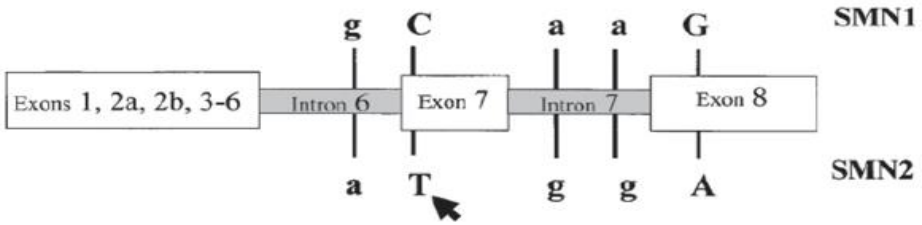


SMN1: Ex2-8

SMN1 gene = ~28kb



Amplicon Exon 2-8 size: ~16kb

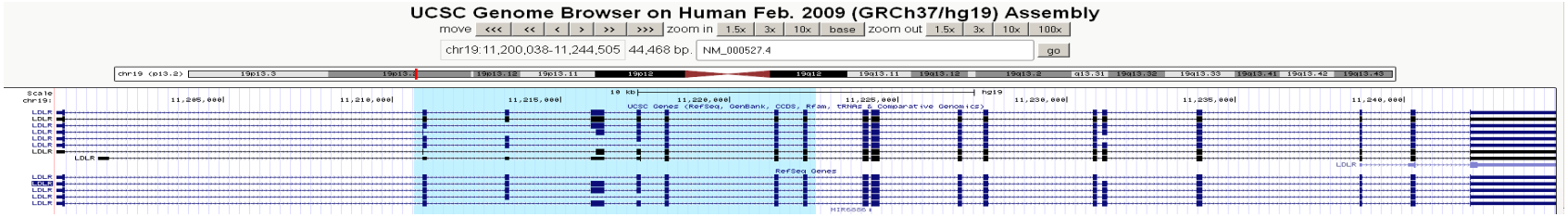


SNPs:	SNP 1		SNP 2		SNP 3		SNP 4		SNP 5	
Reads aligning to:	SMN1	SMN2	SMN1	SMN2	SMN1	SMN2	SMN1	SMN2	SMN1	SMN2
Read Depth	97	246	87	249	84	217	92	193	90	255
A	14%	93%	0%	2%	89%	1%	87%	6%	16%	85%
C	2%	2%	80%	2%	1%	0%	0%	9%	4%	0%
G	81%	6%	9%	5%	10%	98%	13%	82%	80%	14%
T	0%	0%	10%	91%	0%	1%	0%	3%	0%	0%

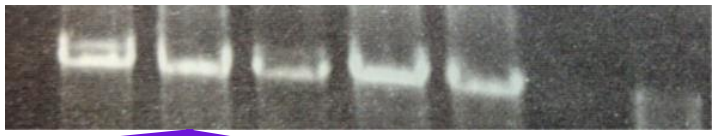
viapath

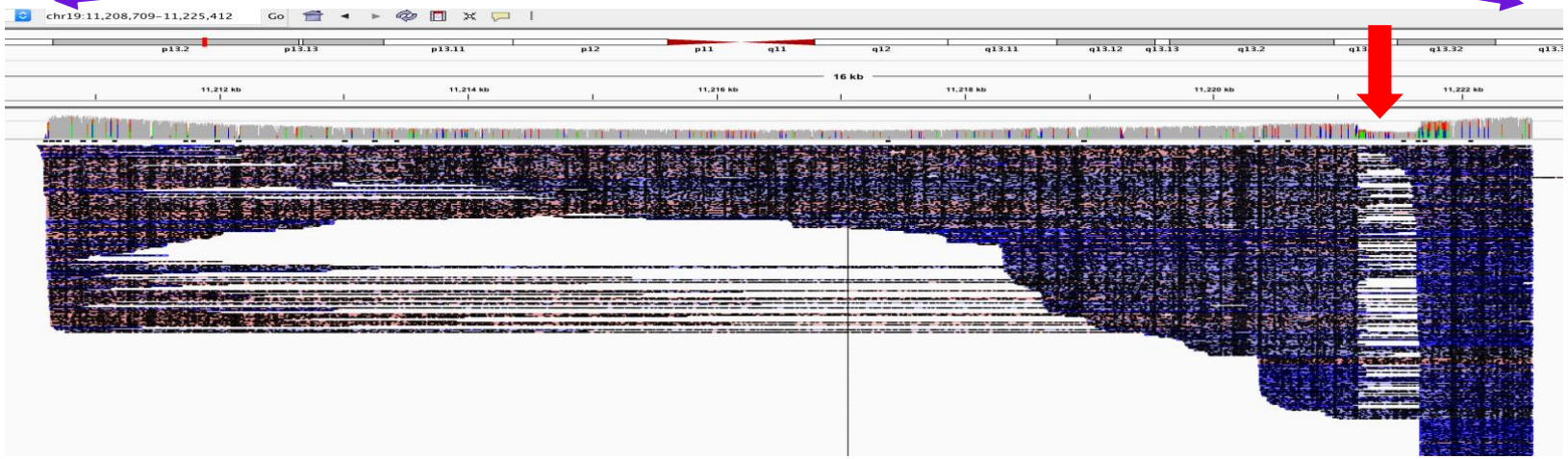
LDLR Exon 7 deletion

LDLR gene = ~44kb



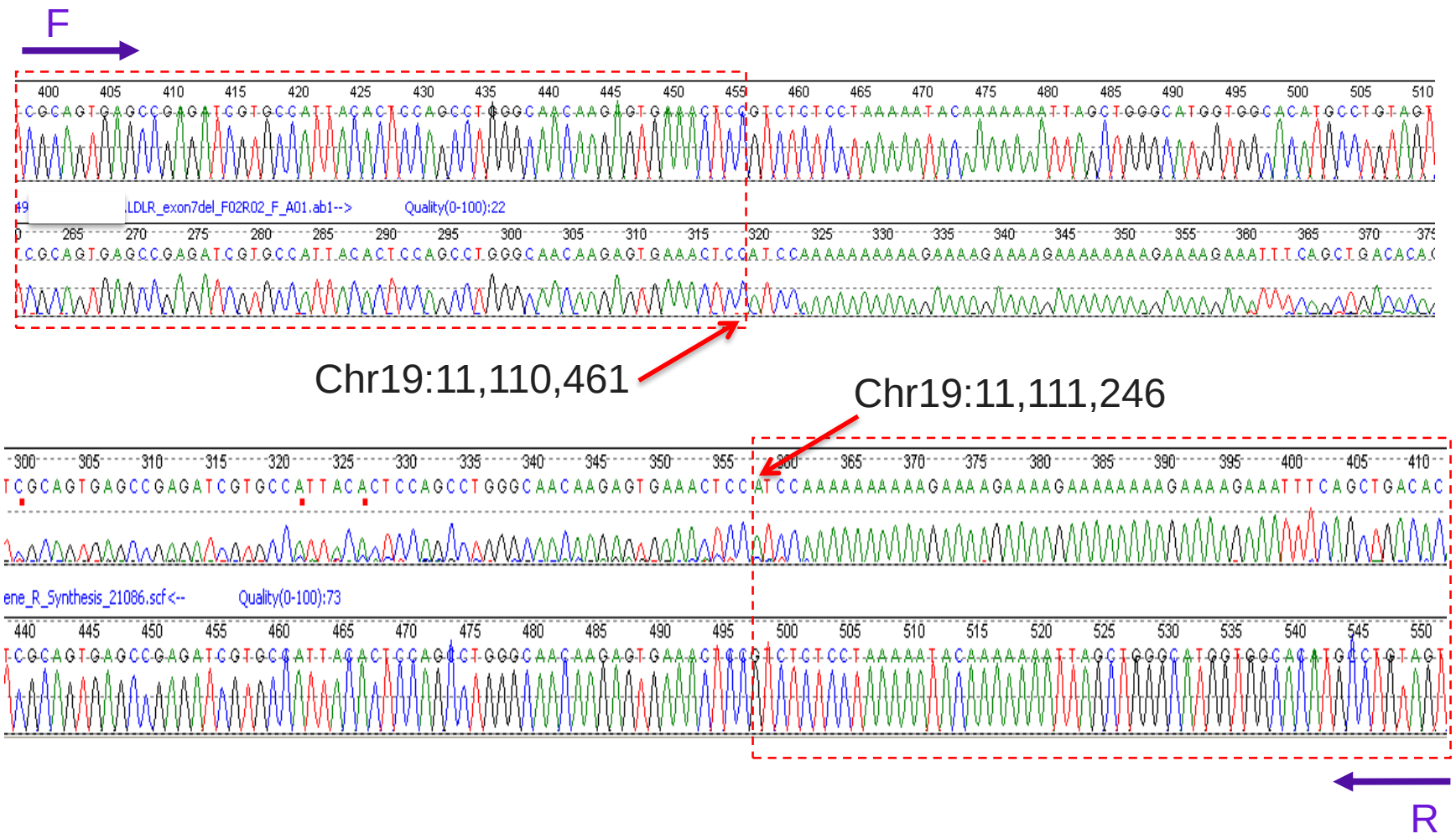
Amplicon Exon 2-8: ~12kb





viapath

LDLR exon 7 deletion breakpoint confirmation (784bp)

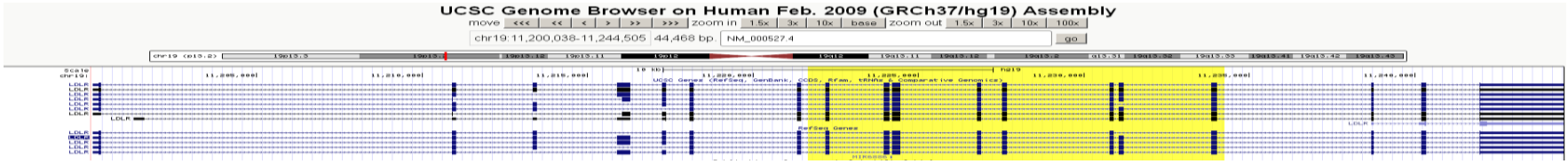


Alu-mediated homologous recombination

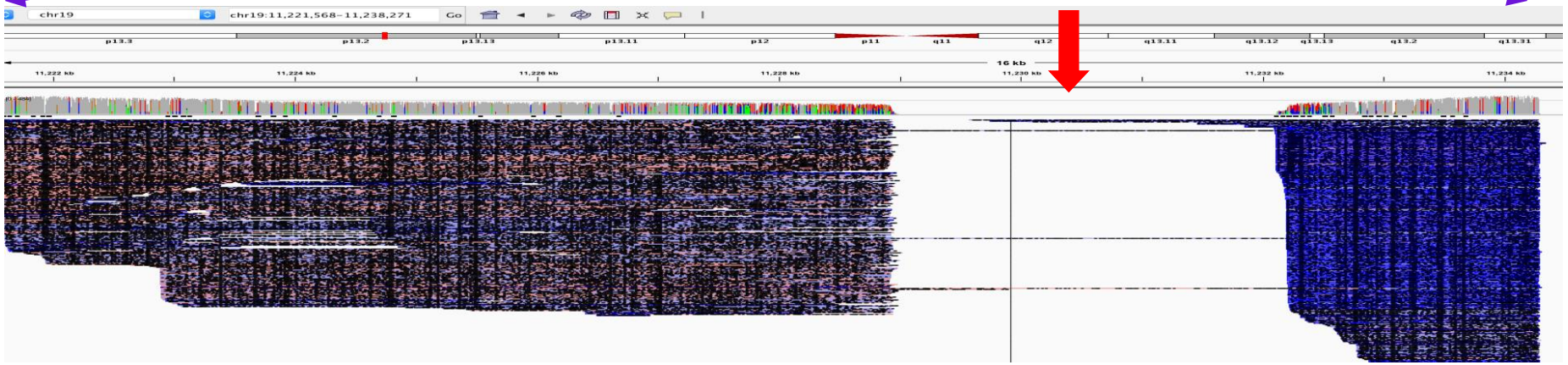
viapath

LDLR Exon 13-14 deletion

LDLR gene = ~44kb



Amplicon Ex8-15 : 12kb



viapath

LDLR Exon 13-14 deletion

Highly repetitive region with stretches of PolyTs

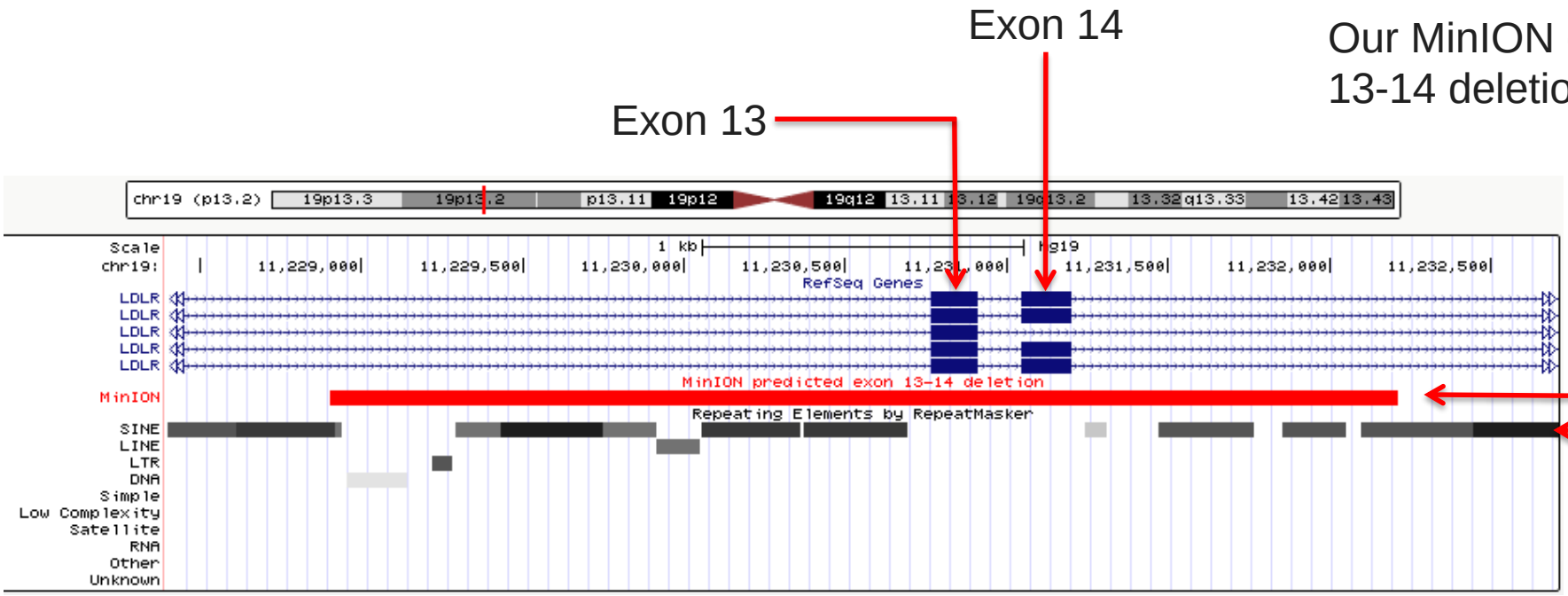
F

viapath

LDLR Exon 13-14 deletion

UCSC output of our deletion

The only large gap in Alu elements is at the exons themselves

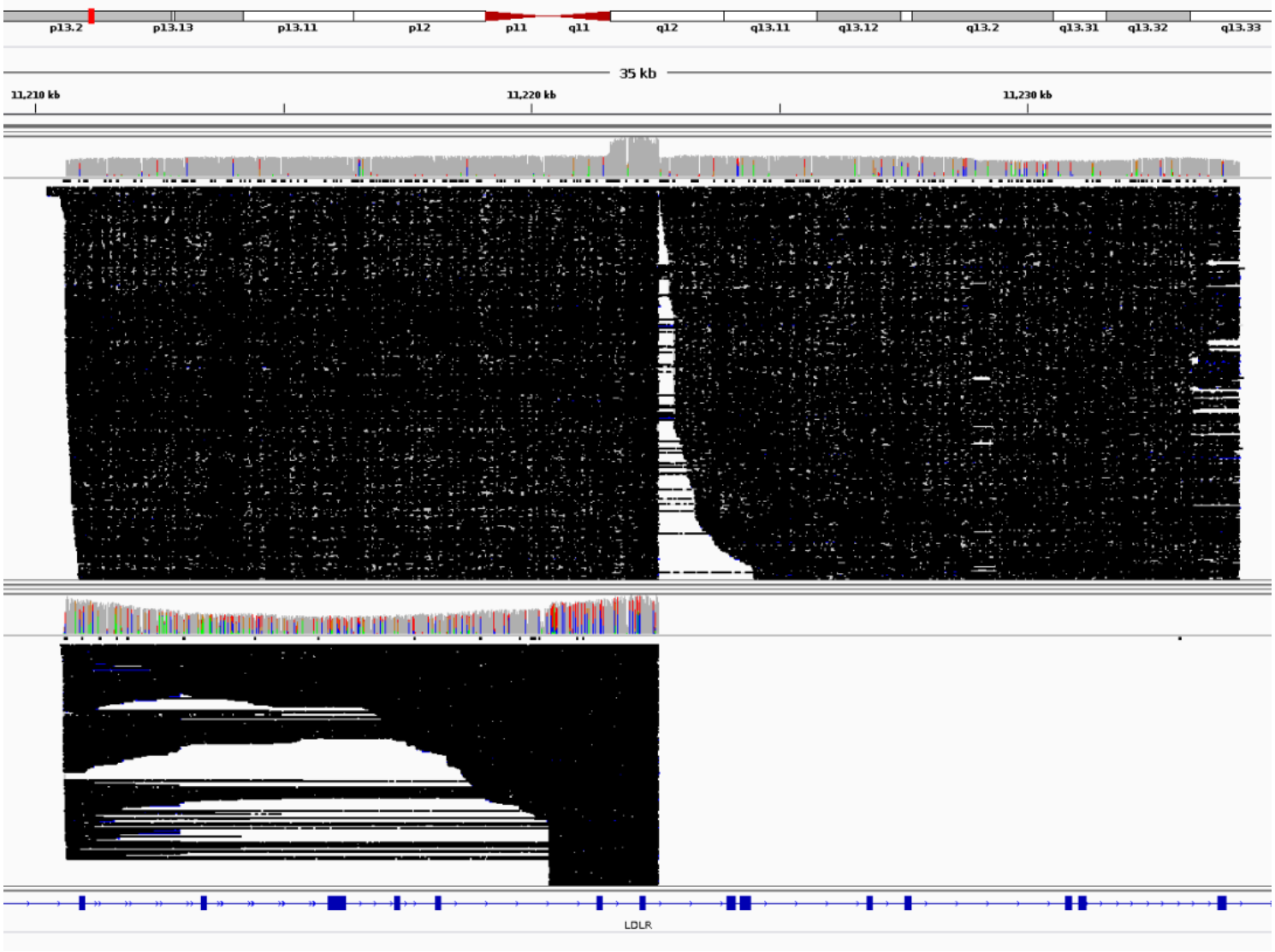


viapath

Preliminary 1D² data – NA12878

1D²

1D

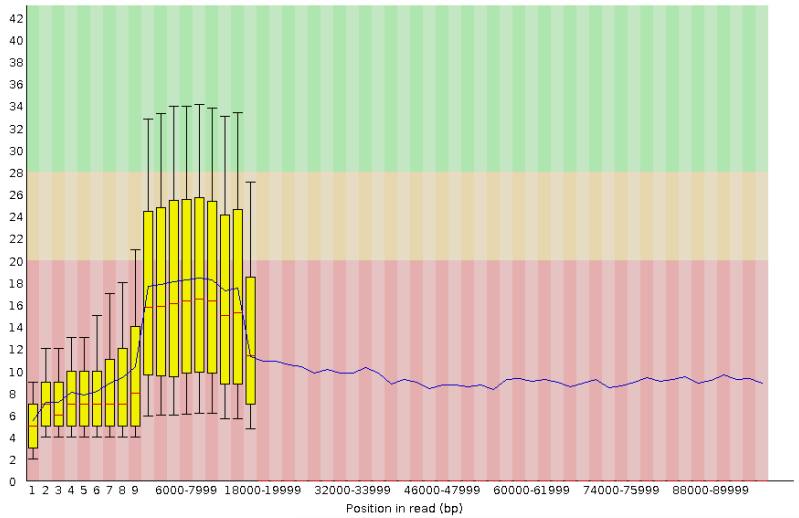


1D vs 1D²

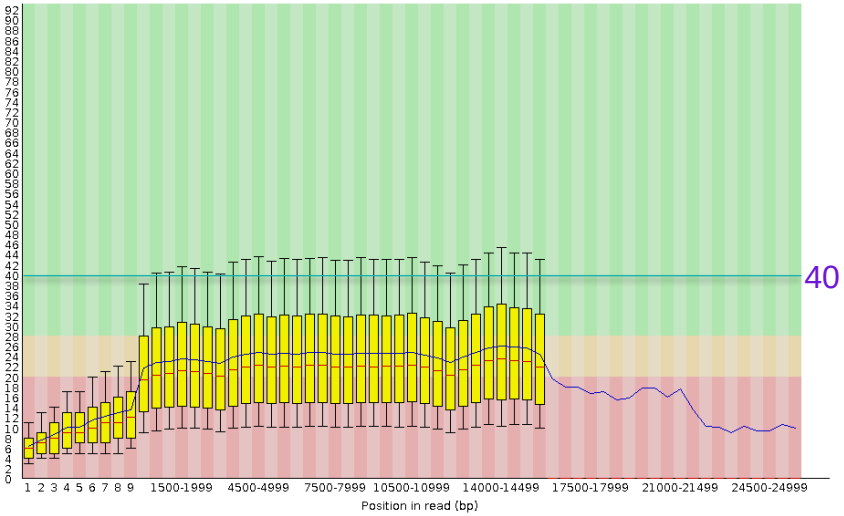
- Increased mean phred-scaled base quality score
- Some Q scores now reaching 40+
- Increased mapping quality
- Reduction in observed error rate compared to reference sequence

FastQC:

1D



1D²



	1D	1D²
Average base Q score (Phred)	Q17	Q23
% reads MAPQ > 20	47%	55%
Observed error rate (in reads MAPQ > 20)	7.7%	3.8%

viapath

Conclusions

- All 10 variants in *BRCA1* and *BRCA2* were correctly identified, however there were also false positives detected due to systematic (non-random) errors in 1D data.

- Accurate identification of two large deletions (3.3kb and 784bp) at a base pair level.
  - Sanger analysis identified highly repetitive regions and RepeatMasker identified Alu repeats flanking the 784bp deletion, indicating Alu-mediated homologous recombination
  - Confirmations of the 3.3kb deletion are still ongoing

- It is possible to differentiate between SMN1 and it's pseudogene SMN2 using long reads and alignment in BWA MEM.

viapath

Conclusions

- Achieved consensus accuracy with 2D reads
- Data analysis is still in progress for HLA

- Random error rates are tractable by consensus alignment and over-sequencing
- Providing systematic errors are avoided, Nanopore sequencing can deliver unique tools for clinical use and point of care testing

- Currently testing the newly released $1D^2$ kit on our previous NA12878 fragments, analysis still ongoing however preliminary data is looking very promising.

viapath

Acknowledgements

Viapath Genetics Laboratory, Guy's Hospital

- Andrew Bond
- Graham Taylor
- Michael Yau

Clinical Transplantation Laboratory, Guy's Hospital

- Robert Collins

- Prof. Jonathan Edgeworth, Medical Director

- Viapath Innovation Fund