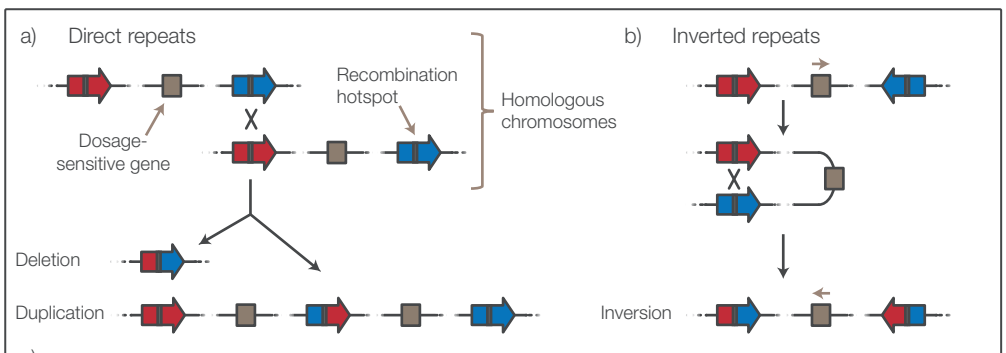
Resolving highly complex rearrangements of genomic architecture using long PromethION reads

Genomic disorders are diseases that result from chromosomal rearrangements, rather than from nucleotide-scale changes, and lead to the loss or gain of chromosomal material, or to inversions

Contact: **publications@nanoporetech.com** More information at: **www.nanoporetech.com** and **publications.nanoporetech.com**



c)

Hotspot	Location of repeats	Repeat length (kb)	Sequence similarity (%)	Inter-repeat distance (Mb)	Approximate frequency	Disorder
CMT1A-REP	17p11.2	24	98.7	1.5	1 x 10^{-5}	Charcot-Marie-Tooth disease (del.) Hereditary Neuropathy (dup.)
AZFa-HERV	Yq11.21	10	94.0	0.8	1 x 10^{-5}	Azoospermia (del.) Normal fertility (dup.)
WBS-LCR	7q11.23	143	99.6	1.5	1 x 10^{-5}	Williams-Beuren syndrome (del.) 7q11.23 duplication syndrome (dup.)
Int22h-	Xq28	9.5	99.7	0.6	1 x 10^{-4}	Haemophilia A (inv.)

Fig. 1 Non-allelic homologous recombination: a) and b) mechanisms c) examples of disorders

Deletion, duplication and inversion of dosage-sensitive genes

Incorrect pairing of direct homologous repeats during meiosis can result in non-allelic recombination. This reciprocal event leads to one gamete with duplication of the intervening sequence and one with deletion (Fig. 1a). Recombination between inverted homologous repeats results in inversion of the intervening sequence, which can disrupt gene function by inverting exons, or by disconnecting coding and regulatory regions (Fig. 1b). If the affected regions contain dosage-sensitive genes, the effect can range from mild to devastating, depending on the locus (Fig. 1c). Long reads enable assembly of repetitive regions and can also be mapped more unambiguously than short reads, enabling precise characterisation of

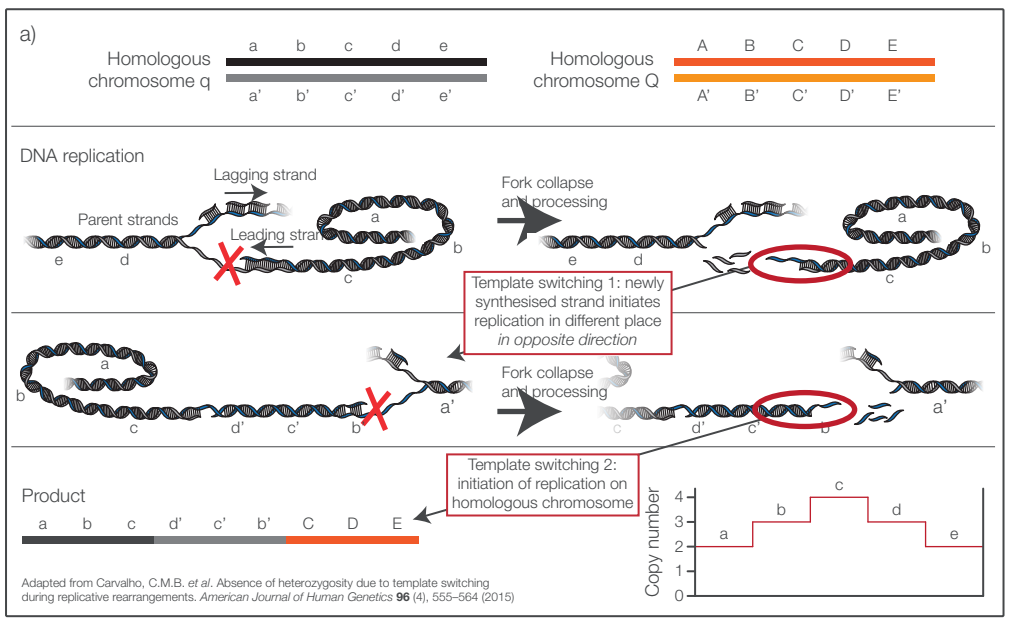


Fig. 2 Mechanism of the duplication-inverted triplication-duplication rearrangement

Elucidating the mechanism of complex genomic rearrangements with long reads

The duplication-inverted triplication-duplication structure (Fig. 2) is a class of complex genomic rearrangement that can lead to severe phenotypic consequences. Just two breakpoints can give rise to both duplicated and triplicated regions. To determine the mechanism that gives rise to the rearrangement it is necessary to i) distinguish between duplicated and triplicated regions, ii) identify breakpoint junctions and iii) identify any absence of heterozygosity (AOH) created by the event. We analysed a patient sample where the rearranged locus had previously been identified using array-CGH, but this method had given insufficient resolution to identify the duplicated regions and thus it had not been possible to ascertain the breakpoints (Fig. 3).

Fig. 3 PromethION sequence data showing AOH; copy number changes; structural variant calls

Fully characterising a duplication-inverted triplication-duplication on chromosome 14

We generated 16x of reads with an aligned read N50 of 15 kb. Reads were aligned against the hg38 reference using *ngmlr* and structural variants were detected by *Sniffles*. Copy number was estimated by comparing read depth at each position to control data and applying a correction factor for different amounts sequenced per sample. *Sniffles* detected 26,036 bp and 2,263 bp inversions (the latter as an inverted duplication) at the centromeric and telomeric ends of the locus respectively. In both cases one or more reads spanned the entire duplicated region. We determined both breakpoints and confirmed these by capillary sequencing. To identify AOH we called SNPs using *bcftools* and compared the density of heterozygous SNPs in the target sample against the controls – a moving average of 10 bins is plotted in the top panel of Fig. 3.

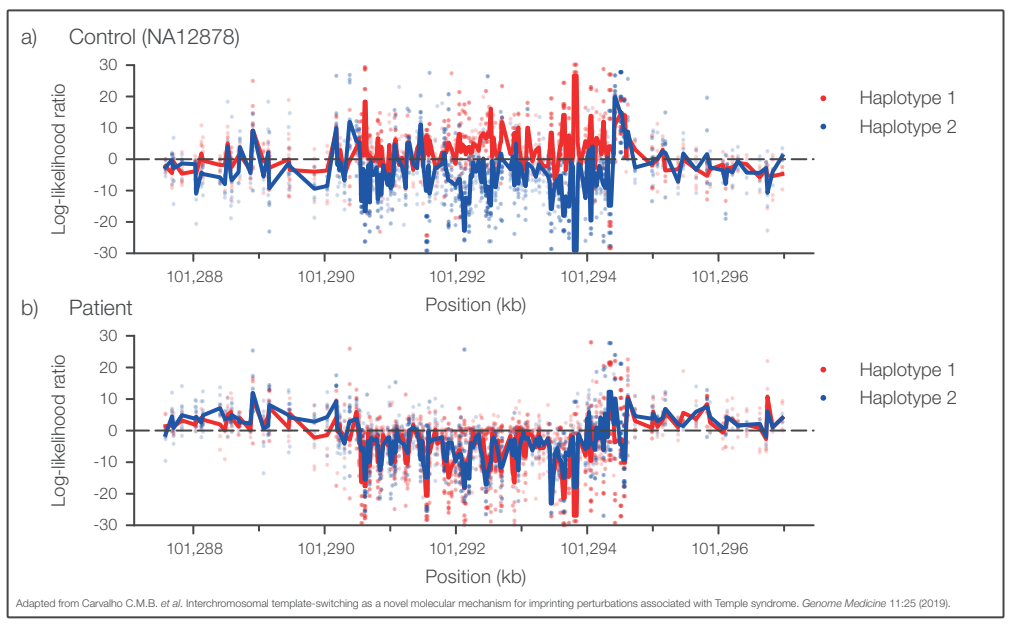


Fig. 4 Methylation of the MEG3 gene a) differential methylation in NA12878 b) AOH in the patient

Absence of the hyper-methylated paternal MEG3 allele in a patient genome

Over a 42 Mb region of chromosome 14, both of the patient's chromosomes contain maternally inherited DNA. AOH in this region can affect gene dosing through the disruption of normal imprinting mechanisms. Reads aligning to the differentially methylated region of the MEG3 gene were first assigned to haplotypes using *MarginPhase* based on the phasing of single nucleotide variants. Next, methylation of CpG sites for each haplotype was assessed using *Nanopolish* and the average of the resulting log-likelihood ratios plotted for each site. A positive LLR indicates CpG methylation, while a negative LLR indicates CpG non-methylation. The control sample (Fig. 4a) shows the normal parent-of-origin specific methylation pattern, while the patient sample (Fig. 4b) lacks any hyper-methylated haplotype, a phenomenon termed loss of imprinting.