Oxford Nanopore Technologies

Comprehensive resolution of challenging genomic variants with Oxford Nanopore *de novo* genome assemblies

Telomere-to-telomere (T2T) genome assemblies generated using long and ultra-long reads, combined with Pore-C, provide the most complete picture of the human genome for platinum-standard references and complex variant analysis

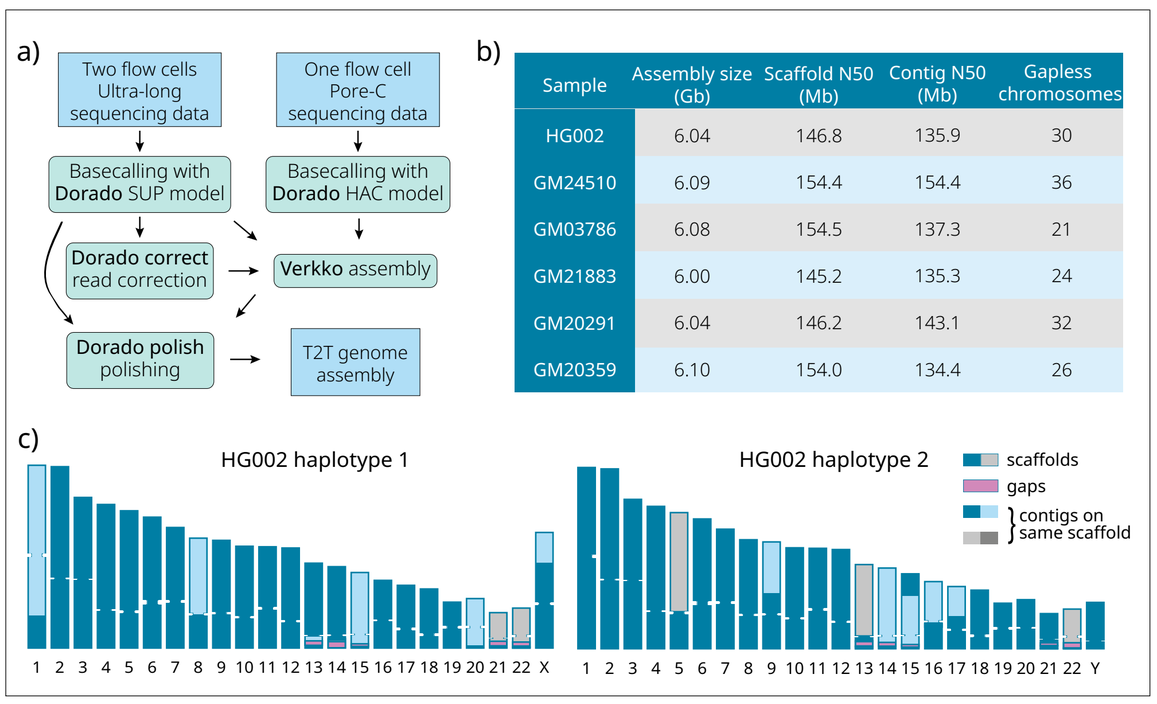

Fig. 1 a) Assembly methods, b) statistics, and c) ideogram of HG002.

Highly complete and contiguous haplotype-resolved T2T assemblies using ultra-long and Pore-C reads

Ultra-long and Pore-C data were generated and processed according to the workflow in **Fig. 1a**. All data was filtered for a mean Q score of 10 and a minimum read length of 1 kb. Dorado-corrected ultra-long, uncorrected ultra-long, and Pore-C reads were passed to Verkko with the "--hifi," "--nano," and "--porec" flags, respectively. Assemblies were subsequently polished with the original, un-corrected ultra-long reads and the Dorado polish tool, using the move-table-aware models. All samples gave highly complete and contiguous assemblies, with 21 to 36 chromosomes assembled as gapless T2Ts and contig N50s ranging from 135.3 Mb to 154.4 Mb (**Fig. 1b**). Alignment of our HG002 assembly to the Q100 T2T shows that most of the remaining gaps are in centromeres or acrocentric rDNA arrays (**Fig. 1c**).

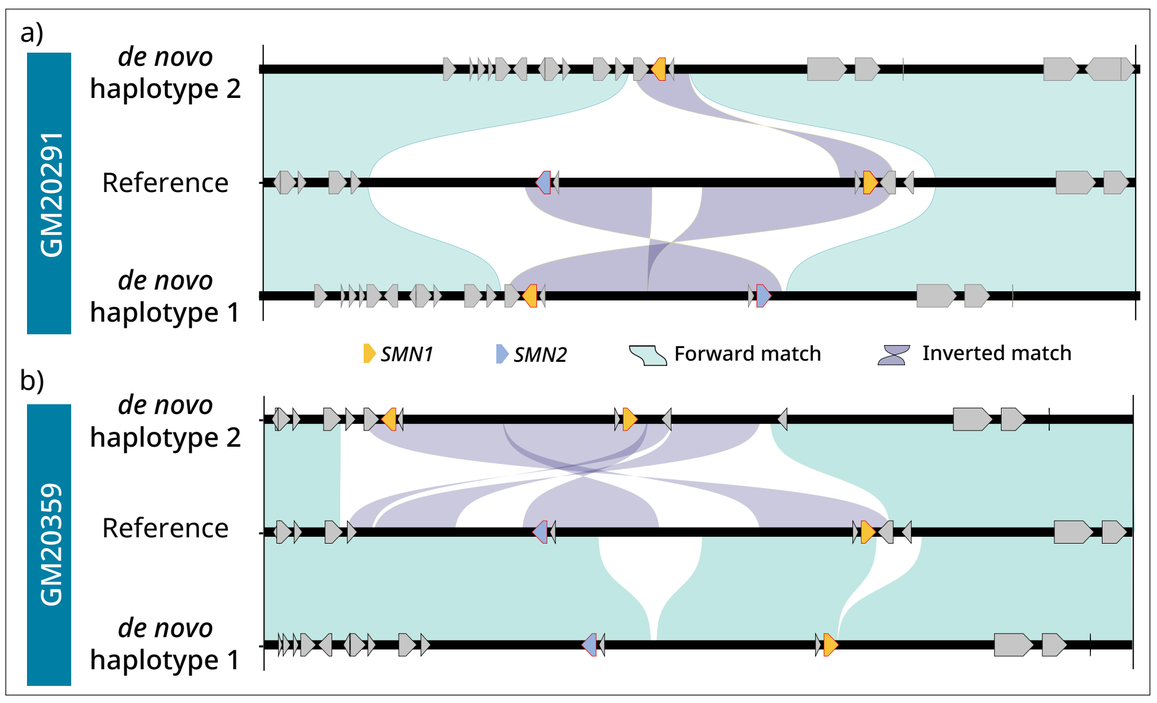

Fig. 3 Fully resolved haplotypes of positive 'silent carrier' samples a) GM20291 and b) GM20359.

T2T assembly fully resolves both haplotypes of spinal muscular atrophy (SMA) loci in putative silent carriers

SMA is a disorder caused by loss-of-function of the *SMN1* gene, which sits in a variable segmental duplication with a nearly identical pseudogene (*SMN2*). Read-alignment approaches resolve the copy number, but fail during allele phasing, resulting in "silent carriers" where a duplication of *SMN1* on one haplotype masks a deletion on the other. In two samples with silent carrier marker single nucleotide polymorphisms (SNPs), T2T assembly fully resolved the SMA loci in both haplotypes. GM20291 harboured inversions and large deletions in both haplotypes, with an *SMN2* deletion in haplotype 2; however, both haplotypes had functional *SMN1* copies (**Fig. 3a**). GM20359 harboured deletions in both haplotypes and an inversion in haplotype 2; however, all *SMN* copies were present (**Fig. 3b**). A gene conversion event appears to restore function to the *SMN2* gene, converting it into a second *SMN1* copy. Both samples would have received an incorrect carrier assignment using marker SNPs.

Fig. 2 a) Robertsonian translocation, b) reciprocal translocation (balanced), and c) ring chromosome.

Characterising complex chromosomal translocations and abnormalities with single-nucleotide resolution

Three samples with complex chromosomal rearrangements were assembled with the T2T protocol. For GM03786, we assembled the whole Robertsonian derivative as a single T2T contig in haplotype 2, resolving the exact breakpoint (**Fig. 2a**). This breakpoint was within a centromeric satellite with low homology to the reference, which was not resolvable by reference-based mapping approaches. In GM24510, one derivative assembled T2T was largely collinear to chr9p, terminating in distal chr21p acrocentric satellites (**Fig. 2b**). The second derivative was not scaffolded across the rDNA array by Verkko, but the breakpoint is localised to a chr9 α-satellite acrocentric p-arm satellite. In GM21883, the ring chromosome structure was clear in the assembly graph, even though Verkko lacks a circular contig detection module, and breakpoints can be elucidated by aligning the circularising edge to the chr15 reference (**Fig. 2c**).

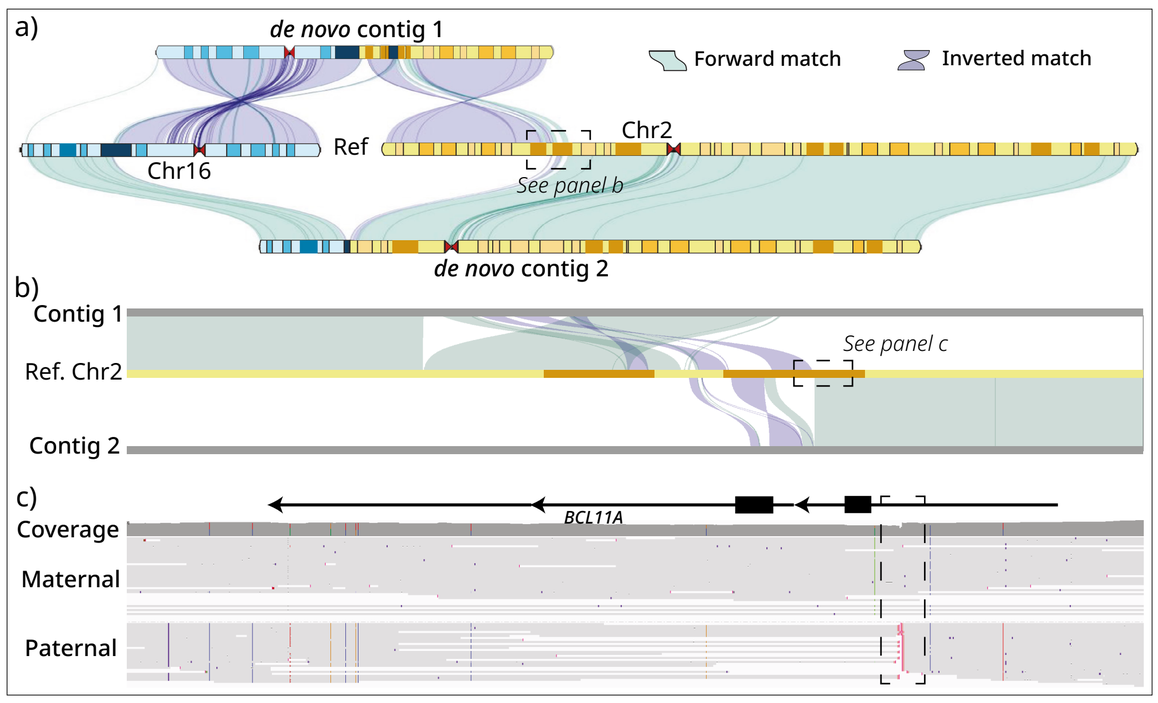

Fig. 4 a) Synteny plot of a large chr 2 and chr 16 translocation, b) chr 2 zoom, and c) *BCL11A* breakpoint.

Resolving structural variants in a rare disease case unsolved by cytogenetic and reference-based approaches

The *de novo* T2T genome assembly identified and fully resolved a highly complex translocation event between chromosome 2 (yellow) and chromosome 16 (blue), which had eluded prior analyses (**Fig. 4a**). The assembly yielded 33 out of 46 complete chromosomal contigs, including two contigs that confirmed the balanced t(2;16)(p;q) translocation event that had been previously identified by karyotyping, but had not been fully resolved. Additional inter- and intra-chromosomal translocation breakpoints were fully resolved at single-nucleotide resolution (n = 34). This paternally inherited event, plus the numerous smaller-scale adjacent rearrangements, reflects the type of complex structural variation associated with germline chromothripsis. Our analyses identified 17 candidate genes directly affected by these breakpoints. One disrupted gene, *BCL11A* on chr2, has been identified as a lead candidate for further investigation based on the individual's phenotype (**Fig. 4b, c**).

PO_1340(EN)_V1_19May2026