Cancer genomics: analysing structural variation, aneuploidy, and heterogeneity with long nanopore reads

Cancer is a disease of the genome in which there can be thousands of large-scale deletions, duplications and translocations, in addition to abnormal numbers of chromosomes

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

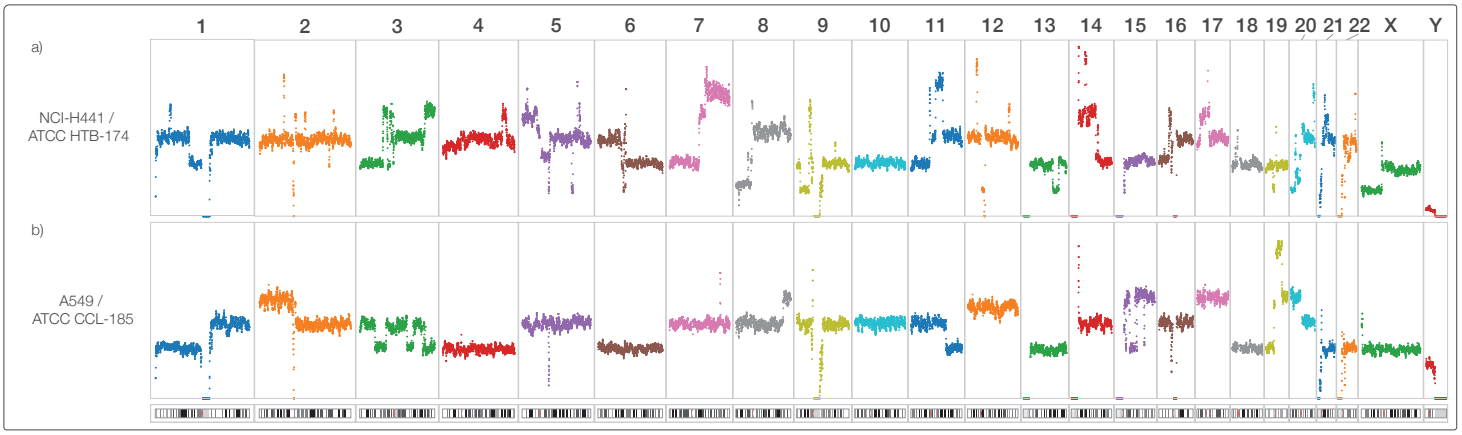

Fig. 1 Whole-genome sequencing of two lung cancer cell lines. Each has a different pattern of duplications, deletions and translocations a) cell line H441 b) cell line A549

Whole-genome sequencing of two lung cancer cell lines reveals different and extensive patterns of deletion, duplication and translocation, in addition to aneuploidy

DNA rearrangements that are not found in normal cells are frequently found in cancer cells from the same patient. Such rearrangements include all types of structural variation (SV): copy number changes, intra- and inter-chromosomal rearrangements, and the insertion of transposable elements and other exogenous sequences. Many of these rearrangements do not contribute directly to the propagation of the cancer, but some alter key genes that drive oncogenesis forward. Analysis of SV in cancer can be used to characterise particular tumour types, and this can have ramifications for treatment. Cancer SVs are often of a complexity or size that can only be resolved with long reads that fully traverse the rearrangement event. To investigate the utility of long nanopore reads in delineating SVs in cancer we took genomic DNA from two lung cancer cell lines, NCI-H441, and A549, prepared libraries and sequenced both samples on a GridION. Both of these cell lines are known to carry extensive structural variation. We generated 19.5x read coverage for H441 and 15x read coverage for A549. We aligned reads to the human reference sequence using ngmlr, and plotted coverage along the reference (Fig. 1). The coverage values are in close agreement with the published karyograms for these samples.

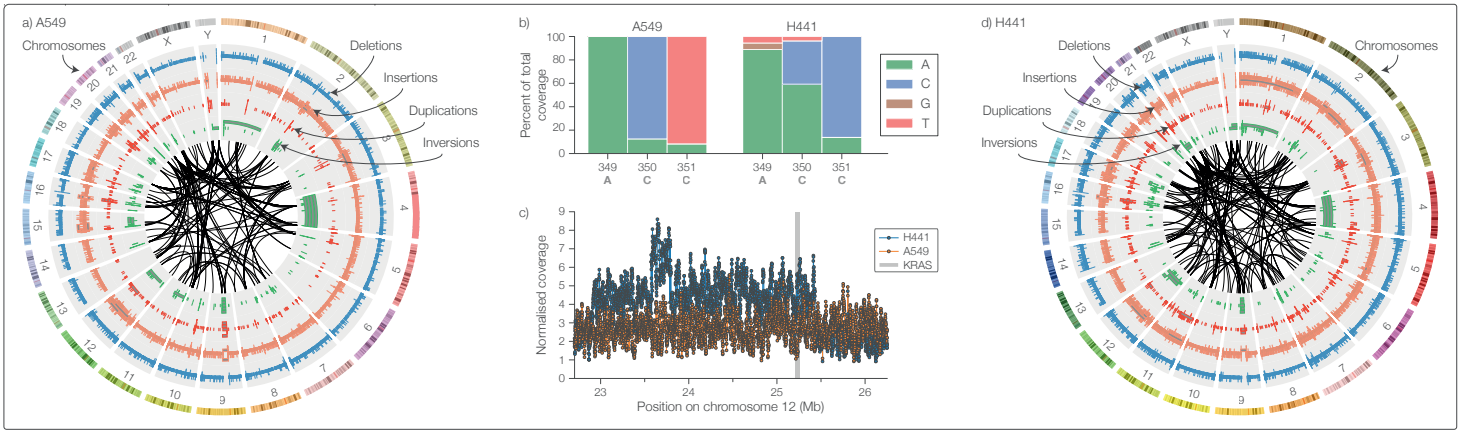

Fig. 2 Further analysis of inter-and intra-chromosomal rearrangements in the two cell lines a) Circos plot of A549 b) KRAS SNPs c) KRAS duplication in H441 d) Circos plot of H441

Both lung cancer cell lines harbour point mutations in the KRAS signal transduction GTPase gene in addition to multiple inter- and intra-chromosomal rearrangements

To detect SVs we processed the alignments with the software tool sniffles, discarding any SVs that were not supported by five or more reads. We created Circos plots for each cell line (Fig 2). A large number of inter- and intra chromosomal rearrangements can be seen. Additionally, both cell lines are known to have non-synonymous mutations in the KRAS gene on chromosome 12, in adjacent positions. A549 carries a C -> T transition in position 25245351 (G12S) and H441 carries a C -> A transversion in position 25245350 (hg38 coordinates). These mutations are clearly visible in the sequence data (Fig. 2b). The C -> A mutation in H441 was detected at a ratio of approximately 2:1 in our reads. Consistent with this, chromosome 12 appears to be triploid in the published H441 karyogram. Fig 2d shows normalised coverage in 1 kb bins for both cell lines for the 12p12.1 region of chromosome 12. The KRAS gene and its surrounding area has higher normalised coverage in cell line H441 compared to A549, supporting this interpretation. The full extent of the duplication can be seen clearly, and a further, smaller, duplication is visible within that duplication. These results show the power of long nanopore sequence reads for the analysis of tumour genomic DNA, and the identification of SVs with far higher resolution than karyotyping.