Haplotype-resolved analysis of cancer genomes and epigenomes using Oxford Nanopore sequencing

Comprehensive identification of somatic and germline single nucleotide variants, structural and copy number variants as well as epigenetic modifications (5mC and 5hmC) from a single dataset

More information at: **www.nanoporetech.com** and **nanoporetech.com/resource-centre**

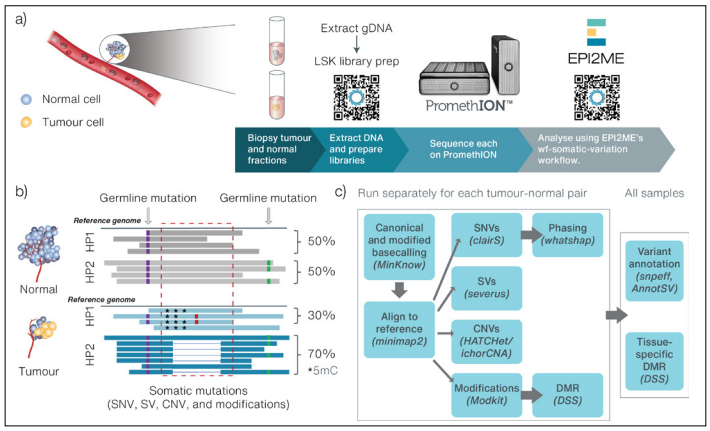

Fig. 1 a) Tumour-normal sequencing b) germline vs somatic variants c) analysis workflow

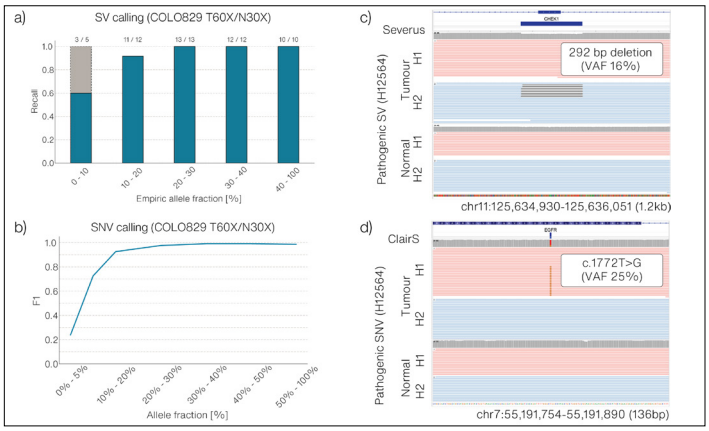

Fig. 2 a) SV and b) SNV calling performance, c) pathogenic SV, d) pathogenic SVN

End-to-end tumour-normal nanopore sequencing workflow

Cancer is a complex disease driven by somatic genomic and epigenomic alterations that accumulate over time. Detection of these alterations in cancer is crucial for understanding the disease, identifying potential therapeutic targets, and personalising treatment strategies. A common way to analyse cancer data is tumour-normal sequencing (Fig. 1a). This enables detection of both germline and somatic mutations (Fig. 1b). Native long reads enable direct detection of SNVs, complex structural variants (SVs and CNVs), haplotype phasing, and identification of DNA modifications like 5-methylcytosine (5mC) and 5-hydroxy-methylcytosine (5hmC) from a single dataset using open-source analysis tools (Fig. 1c).

Accurate identification of somatic single-nucleotide and structural variation

We benchmarked somatic SV and SNV calls against truth-sets from Valle-Inclan *et al.* and Arora *et al.* for COLO829. We found near-perfect recall for SV calling. 49 out of 52 SVs were called correctly with two more complex SVs called but their break points represented differently compared with the truth-set (Fig. 2a). For SNV calling we found high precision and recall (F1 > 0.9) down to 10% allele frequency (Fig. 2b). Next, we sequenced 12 tumour-normal pairs from four different tissues and performed somatic genetic and epigenetic variant calling. We found high variability in the number and types of variants including pathogenic SVs (Fig. 2c) and SNVs (Fig. 2d) across samples, highlighting the complexity present in cancer.

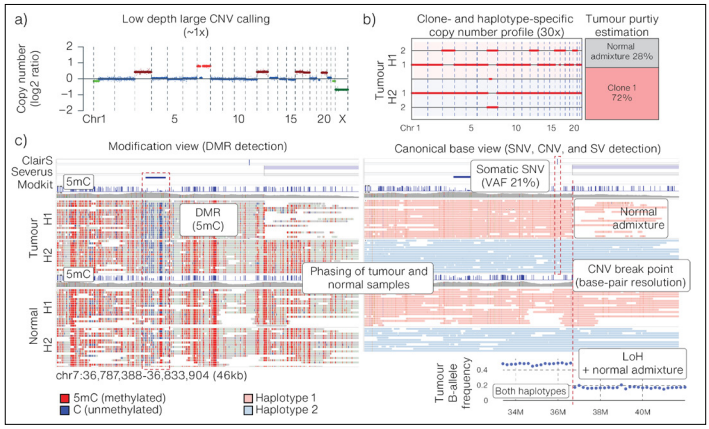

Fig. 3 a) CNV profile b) phased CNV with break points on chr7 c) zoomed-in region

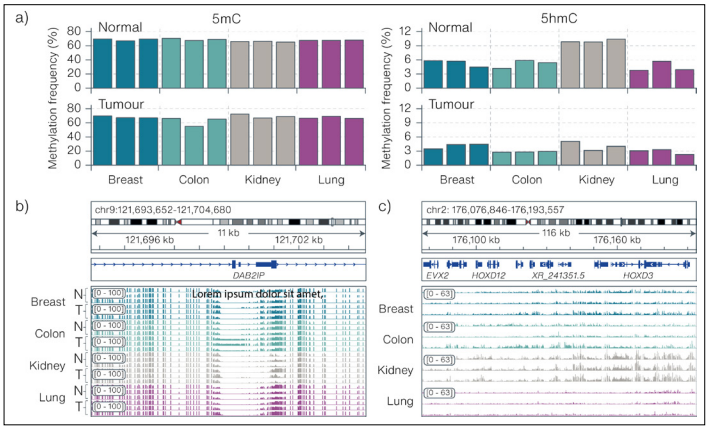

Fig. 4 a) Average 5mC/5hmC levels b) 5mC in *DAB2IP* promoter c) 5hmC in *HOXD3* gene

Joint characterisation of genomic and epigenomic variants using a single dataset

We performed shallow (~1x) nanopore sequencing on a kidney cancer sample and identified characteristic whole chromosomal gains, as well as a more complex amplification on chromosome 7 (Fig. 3a). Using the full high-depth dataset of the same sample, we inferred genome-wide clone- and haplotype-resolved CNV profiles (Fig. 3b). Using somatic SV calling we were able to pinpoint and phase the exact breakpoint of the complex CNV, revealing a large region of loss-of-heterozygosity, another common event in cancer (Fig. 3c). Furthermore, we identified and phased a somatic SNV as well as a differentially methylated region in close proximity to the 5' break point, illustrating the benefits of nanopore tumour-normal sequencing.

Genome-wide base-pair resolution 5mC and 5hmC profiling in cancer

Nanopore sequencing profiles 5mC and 5hmC methylation without bisulfite treatment. We found genome-wide hypomethylation in cancer tissues (Fig. 4a) alongside regional hypermethylation of promoters. *DAB2IP* plays a key role in colorectal cancer and the promoter region is only methylated in the colon tumour samples (Fig. 4b). 5hmC is known to be highly tissue specific and strongly reduced in cancer tissue. This is confirmed by average 5hmC frequencies across the genome (Fig. 4a) and per-base frequencies across the ***HOXD3*** gene (Fig. 4c) in our data. *HOXD3* showed high levels of 5hmC in kidney and lower levels in lung, matching expected expression in these tissues.