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

Philipp Rescheneder[1], Sergey Aganezov[2], Sean McKenzie[2], Keira Cheetham[1], Phill James[1] & Sissel Juul[2]

[1]Oxford Nanopore Technologies plc, Oxford, UK, [2]Oxford Nanopore Technologies Inc, New York, NY

Contact: philipp.rescheneder@nanoporetech.com

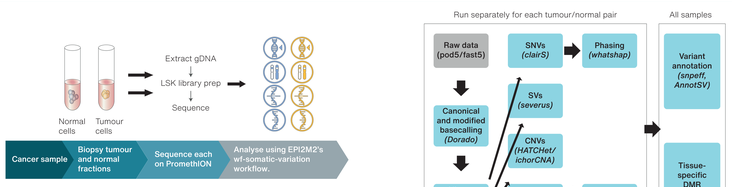
Download poster

PAN: 6218

Abstract

Cancer is a complex and dynamic disease driven by somatic genomic and epigenomic alterations that accumulate over time. These changes give rise to heterogeneous collections of cells or clones, each with distinct (epi)genomic profiles within a single tumour. Accurate characterisation of these changes is crucial for understanding the mechanisms driving the disease, identifying potential therapeutic targets, and personalising treatment strategies. Due to the technological constraints of short-read and array-based approaches, cancer research has historically had a strong focus on detecting small genomic changes like SNPs and small indels (SNVs) as well as the broad characterisation of large-scale copy number changes (CNVs), mostly ignoring other important variant classes and epigenetic modifications. Here we demonstrate how Oxford Nanopore native long-read sequencing enables the direct detection of not only SNVs, but also break point-resolved simple and complex structural variants (SVs and CNVs), haplotype phasing of all variant types, and identification of DNA modifications like 5-methylcytosine (5mC) and 5-hydroxy-methylcytosine (5hmC) from a single tumour-normal dataset. We illustrate how our end-to-end somatic workflow streamlines analysis for the different variant classes and benchmark somatic SNV and SV calling performance using well characterised cancer cell lines and in-silico benchmarking datasets at different sequencing depths. Finally, we use real-world tumour-normal pairs to showcase a comprehensive sample to answer tumour-normal analysis using Nanopore sequencing. This includes the characterisation of complex patterns of somatic SVs in the different cancer samples, the identification of unique 5mC methylation patterns, exploring characteristic differences in 5hmC levels between tumour and normal samples as well as showing how Nanopore long reads enable haplotype and clone specific CNV calling as well as phasing of genetic and epigenetic variation.

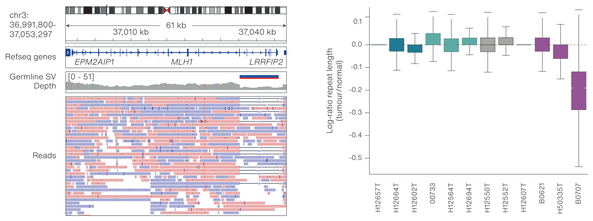
End-to-end tumour-normal workflow



End-to-end workflow

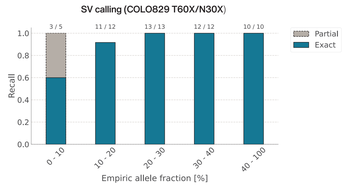
Cancer is a complex disease driven by somatic genomic and epigenomic alterations. Detection of these changes in cancer is crucial for understanding the disease, identifying potential therapeutic targets, and personalising treatment strategies[1]. A common way to analyse cancer data is tumour-normal sequencing[2]. Here we demonstrate Oxford Nanopore's end-to-end tumour-normal workflow, sequencing the well- characterised COLO829 cell-line as well as twelve matched tumour-normal pairs from four different tissues to a depth of up to 60X and 30X for tumour and normal samples respectively using the latest Kit14 chemistry. Read-length N50 ranged from 5 kilo base-pairs for more fragmented samples up to 30 kb. Analysis was carried out using a pre-release version of *wf-somatic-variation*[3].

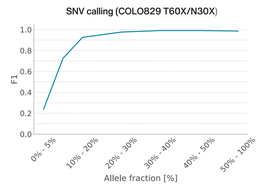
Detecting hereditary-cancer-related variation and MSI



First, we screened all samples for pathogenic germline mutations and identified a 10 kb deletion associated with Lynch syndrome in sample B0707 (left panel). Reduced DNA repair activity caused by Lynch syndrome can lead to a shortening of microsatellites in tumour cells (microsatellite instability or MSI)[4]. Thus, microsatellites are considered phenotypic markers in some cancers. To illustrate Oxford Nanopore's ability to detect MSI we compared microsatellite lengths between tumour and normal samples and confirmed a clear MSI signal only in B0707 (right panel).

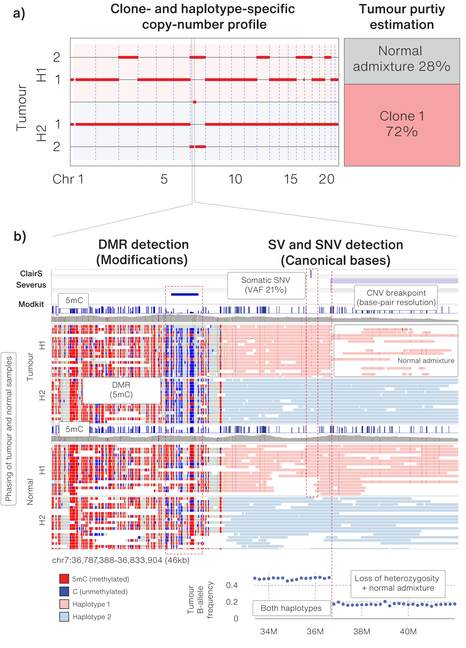
Somatic SV and SNV calling performance



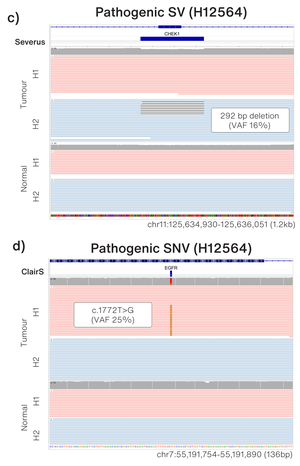


Next, we benchmarked somatic SV (Severus[5]) and SNV (ClairS[6]) calls against truth-sets from Valle-Inclan et al.[7] (SV) and Arora et al.[8] (SNV) for COLO892 (60X tumour and 30X normal). We found near-perfect recall for SV calling. 49 out of 52 SVs were called correctly with two more called but represented differently. For SNV calling we found high precision and recall (F1 > 0.9) down to 10% allele frequency.

Joint characterisation of genomic and epigenomic variants

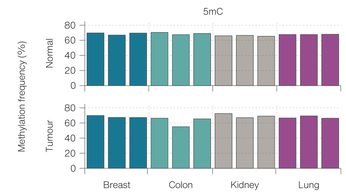


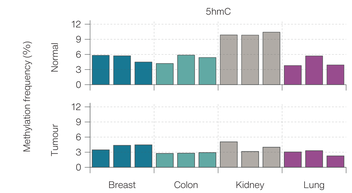
We then screened all tumour-normal pairs for somatic genetic and epi-genetic variation starting with genome-wide clone and haplotype-resolved CNV profiles and tumour purity estimations. Panel using HATCHet[9]. Panel (a) shows characteristic whole-chromosomal gains as well as a more complex amplification on chromosome 7 for one of the kidney cancer samples (H12550T).



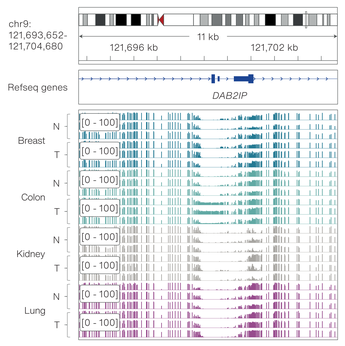
Using somatic SV calling we were able to pinpoint and phase the exact breakpoint of the complex CNV revealing a large region of loss-of-heterozygosity, another common event in cancer. Furthermore, we identified and phased a somatic SNV as well as a differentially methylated region in close proximity to the 5' break-point illustrating the benefits of Nanopore tumour-normal sequencing (b). To investigate smaller changes, we also performed SNV. Overall we found high variability in the number and types of variants including pathogenic SVs (c) and SNVs (d) across samples, highlighting the complexity present in cancer.

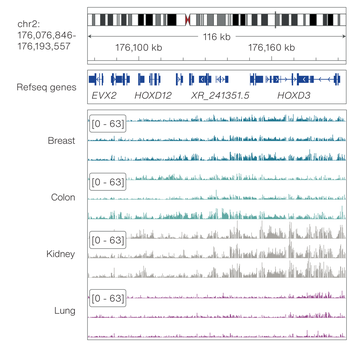
Genome-wide 5mC and 5hmC profiling in cancer





Regulation of methylation is a critical mechanism of cancer progression[10]. Nanopore sequencing can profile 5mC and 5hmC methylation without the need for bisulfite treatment. When comparing average 5mC levels we found genome-wide hypomethylation in cancer tissues alongside regional hypermethylation of promoters.





The left panel shows the DAB2IP promoter region. This gene plays a key role in colorectal cancer and is only methylated in our colon tumour samples. 5hmC is known to be highly tissue specific and strongly reduced in cancer tissue. This is confirmed by our observed average 5hmC frequencies across the genome (upper panel) and per-base frequencies across the HOXD3 gene (right panel). HOXD3 showed high levels of 5hmC in kidney and lower levels in lung, matching expected expression in these tissues.

Conclusion

A vast array of genomic aberrations has been linked to solid tumour formation and disease progression. Here we demonstrate that a single technology can provide valuable insight into the mutational landscape of solid tumours by identifying SNVs, SVs, CNVs as well as epi-genetic modifications using a single matched tumour-normal approach. Haplotype-resolved Copy-number analysis was used to detect large-scale chromosome rearrangements. Somatic SV calling helps refining beak-points of large CNVs and identifies other simple and complex structural changes while SNV calling provides insight into single-nucleotide driver-mutations affecting coding regions. Changes in methylation within the early tumour microenvironment are thought to be one of the first signals in tumour progression. Here we observed unique 5mC and 5hmC methylation patterns between both tissue types and tumour-normal pairs. The ability to gain this level of insight into a wide range of potential driver mutations, from a potentially diverse range of solid tumour samples using a single, non-targeted, whole-genome sequencing approach provides a powerful tool to help further the understanding, and hopefully prevention or treatment, of cancers in the future.

Open-data and workflow

References

[1] Hanahan D, Weinberg RA, Cell. 2001
[2] Mandelker D, et al., Trends Cancer. 2020
[3] https://github.com/epi2me-labs/wf-somatic-variation
[4] Boland CR, Goel A, Gastroenterology. 2010
[5] Keskus A., et al., medRxiv 2024.03.22.24304756
[6] Zheng Z., et al., bioRxiv 2023.08.17.553778
[7] Espejo Valle-Inclan J, et al., Cell Genom. 2022
[8] Arora, K., et al.,Sci Rep 9, 19123 (2019)
[9] Zaccaria S, et al., Nat Commun. 2020
[10] Lakshminarasimhan R, et al., Adv Exp Med Biol.