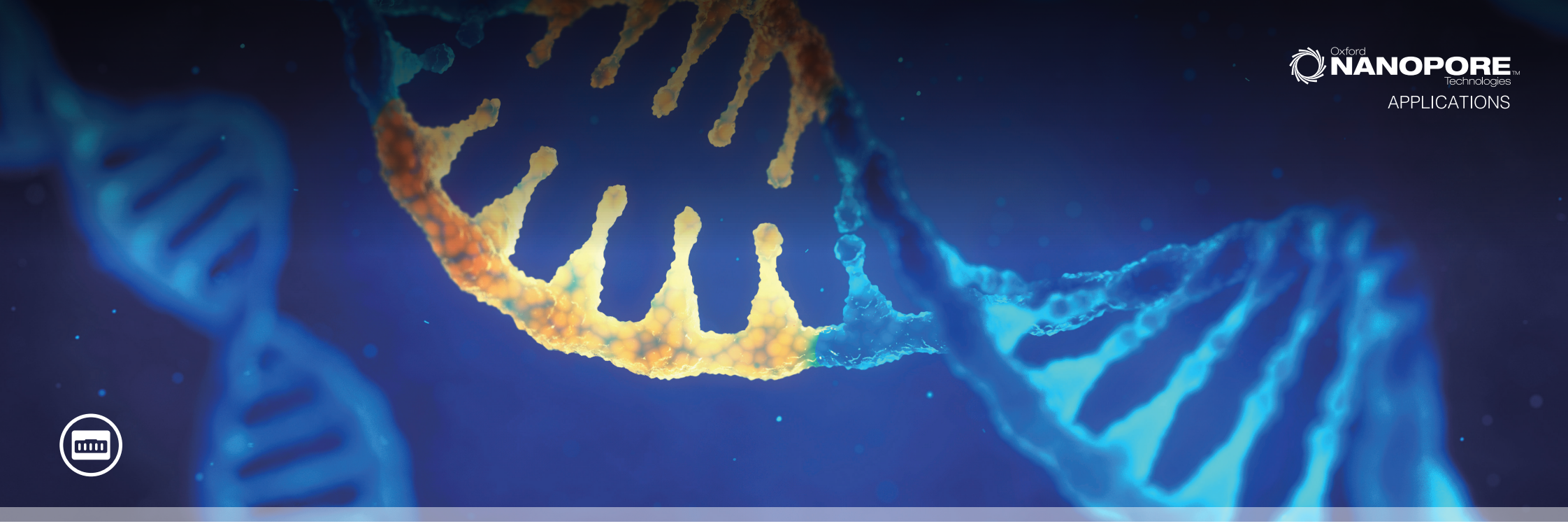

Amplification-free target enrichment to capture long genomic fragments and retain methylation status

Two approaches for PCR-free sequence capture allow users to make more efficient use of a sequencing run and to enrich long fragments of DNA without losing epigenetic modifications

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

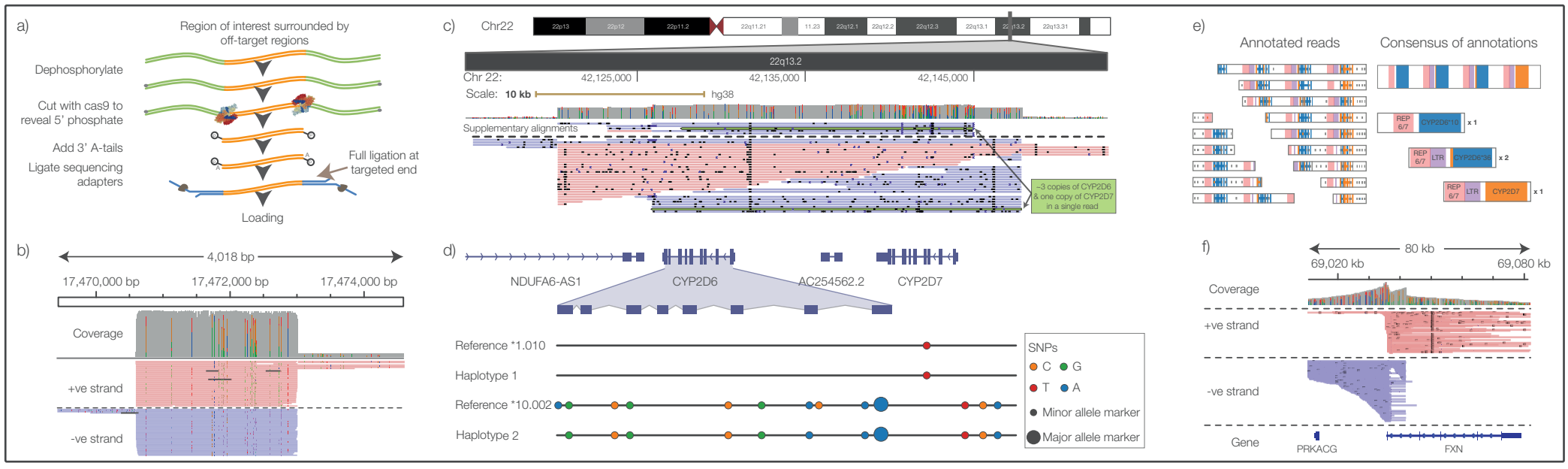

Fig. 1 CRISPR/Cas9 target enrichment a) overview of library prep b) example of mapped reads c) CYP2D6 locus and aligned reads d) called haplotypes e) gene conversion f) enriching the FXN locus

90-minute Cas9 library preparation for the PCR-free enrichment of target loci enables capture of long native genomic fragments while also retaining epigenetic modifications

Cas9 can be used for PCR-free enrichment of target loci (Fig. 1a). Unlike amplification-based methods, Cas9 allows native DNA strands to be sequenced, preserving fragment length and modifications. All DNA fragments in the sample are initially dephosphorylated and Cas9 is then used to cleave the DNA at predetermined sites, exposing ligatable ends. All 3' ends are then dA-tailed, before ligation of sequencing adapters. The adapters only ligate to cleaved ends because of the newly revealed phosphates. The entire library is then added to the flow cell. The region of interest is enriched several thousand-fold (Fig. 1b). We demonstrate the effectiveness of this method on the CYP2D6 locus (Fig. 1c). The enzyme CYP2D6 metabolises ~25% of common drugs and its activity between individuals varies widely, due to high levels of polymorphism. To determine an individual's optimal drug dosage, we need to identify all variants that influence the enzyme's activity. Long reads help to distinguish between CYP2D6 and its close paralogue, CYP2D7 (Fig. 1d) and between their gene conversion products. We enriched the region from NA18256, a sample described as having one copy of allele *1 and three copies of allele *10. Our analysis revealed one copy of *10 and two copies of *36, a gene-conversion product, in the haplotype (Fig. 1d). We have also used Cas9 to enrich the FXN gene on Chr 9 q21.12 in a Friedreich's Ataxia patient. Full characterisation of the locus requires methylation patterns to be retained. For more information please see our Friedreich's Ataxia poster.

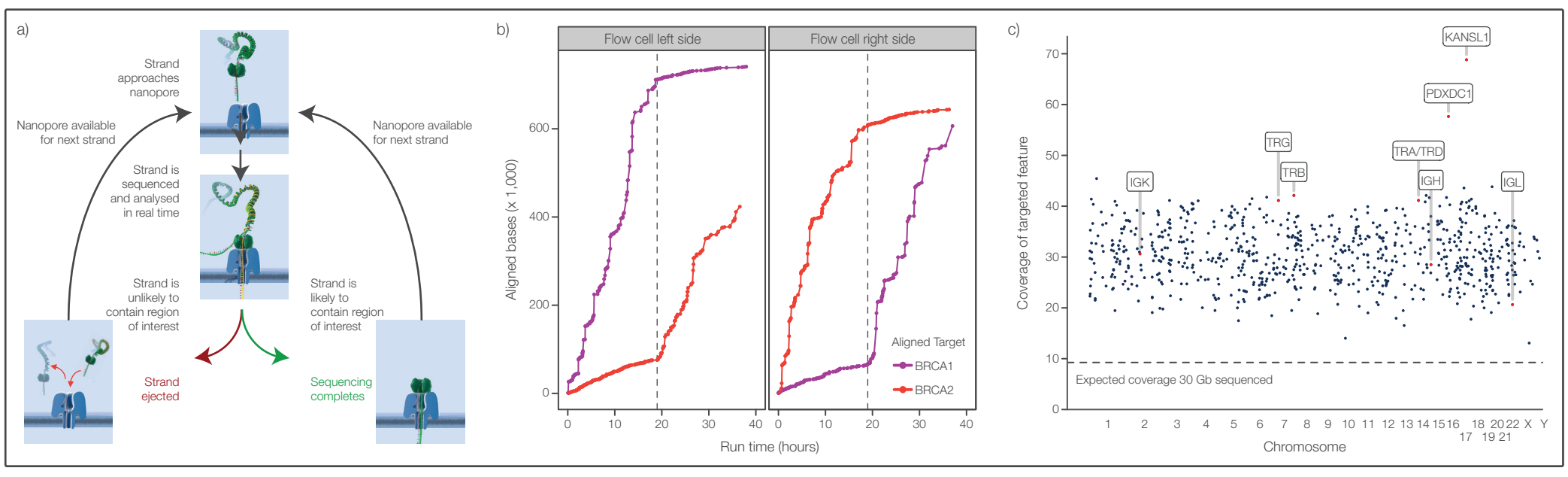

Fig. 2 Adaptive sampling a) selecting regions of interest b) targeting different regions at different times during a sequencing run c) coverage of >753 COVID-related targets in a 54.1 Mb search space

Adaptive sampling allows the automatic selection of an unlimited number of regions of interest during a sequencing run, without the need for additional library-preparation steps

If a project requires long reads or methylation analysis, it is typical to resort to whole-genome sequencing. This means generating more data than is ultimately required, increasing the experiment cost and prolonging analysis time. With adaptive sampling, many genomic regions of interest (ROI) can be targeted together, without needing additional steps during library preparation. As a strand of DNA passes through a nanopore, it is basecalled in real time, allowing the nascent sequencing read to be analysed while the strand is still passing through the pore. If the strand is likely to contain an ROI, the sequencing of that strand is allowed to continue to the end. If the strand is not likely to contain an ROI, the strand is ejected from the pore. In each case, the pore can then capture another strand and the selection process is repeated (Fig. 2a). Decisions are made for each pore individually, so it is possible to configure different pores to select for different target regions in the same sequencing run. To demonstrate this, for the first 20 hours of a sequencing run, one half of a MinION (TM) Flow Cell was set to capture the BRCA1 locus, and the other half captured BRCA2. After 20 hours of sequencing this was reversed (Fig. 2b). To demonstrate adaptive sampling on a larger panel of target regions, we captured >750 loci that have been reported to be associated with host response to COVID-19 including major immune-related genes, SNPs identified by GWAS, and genes identified by CRISPR knock-out. Sufficient coverage can be obtained from a single MinION Flow Cell for variant calling (Fig. 2c).