# Targeted Nanopore sequencing for phased variant detection

**Isac Lee[1], Rachael Workman[1], Josh Wang[2], Yunfan Fan[1], Winston Timp[1]**

[1]Johns Hopkins University, Department of Biomedical Engineering

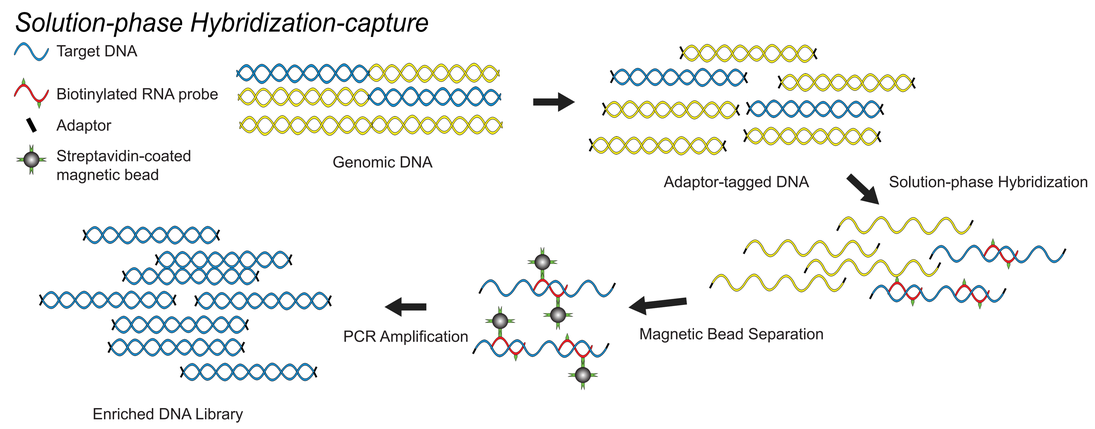
[2]Agilent Technologies

## Introduction

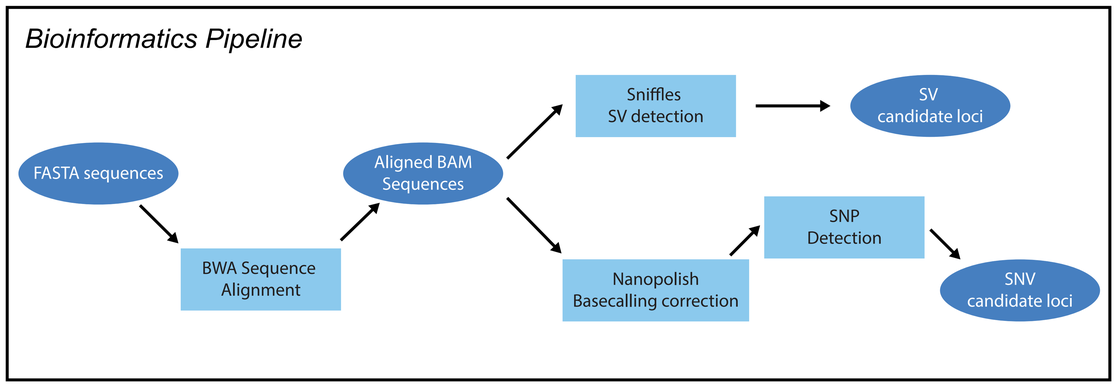
Large-scale genomic anomalies – structural variations (SVs) – are pervasive in cancer. Due to the scale of the SVs and the repetitive nature of the sequences frequently flanking them, short-read sequencing is limited in its ability to detect SVs. In contrast, long read sequencing, e.g. nanopore sequencing, enables sequencing of unfragmented or otherwise long strands of DNA, improving detection of SVs. Recently, nanopore sequencing has been commercialized in the MinION, a sequencer from Oxford Nanopore (ONT). However, the technology has relatively low yield (~2Gb) making it difficult to employ on whole human genome samples at sufficient coverage, especially for heterogeneous samples or rare event detection.

To increase the sequencing depth at "hotspot" regions of SVs, we have adapted solution-phase hybridization capture (SureSelect) to nanopore sequencing, enriching for the p16/CDKN2A and SMAD4 tumor suppressor gene regions, known to be frequently disrupted in pancreatic ductal adenocarcinoma (PDAC). Here we present data from this method on control NA12878 DNA and a patient-derived PDAC cell line DNA.

## Methods




Hybridization probes were designed to capture a 2.5Mb region encompassing the p16/*CDKN2A* and *SMAD4* tumore suppressor genes. DNA was sheared to ~3kb, then adapters ligated and PCR amplified, followed by Blue Pippin size selection before hybridization and streptavidin bead capture. The resulting library was prepared for ONT nanopore sequencing on the Mk1 MinION with R9 pore chemistry. Nanopore sequencing yielded current signals basecalled using the Metriochor cloud basecalling system (ONT).




Sequenced reads were aligned to the human reference genome (hg38) using *bwa mem -x ont2d*. The aligned BAM sequences were used to 1) detect structural variations using sniffles (https://github.com/fritzsedlazeck/Sniffles) and 2) corect the sequence reads using raw signal via nanopolish[1] (https://github.com/jts/nanopolish), resulting in calls for both SVs and single nucleotide variations from nanopore data.

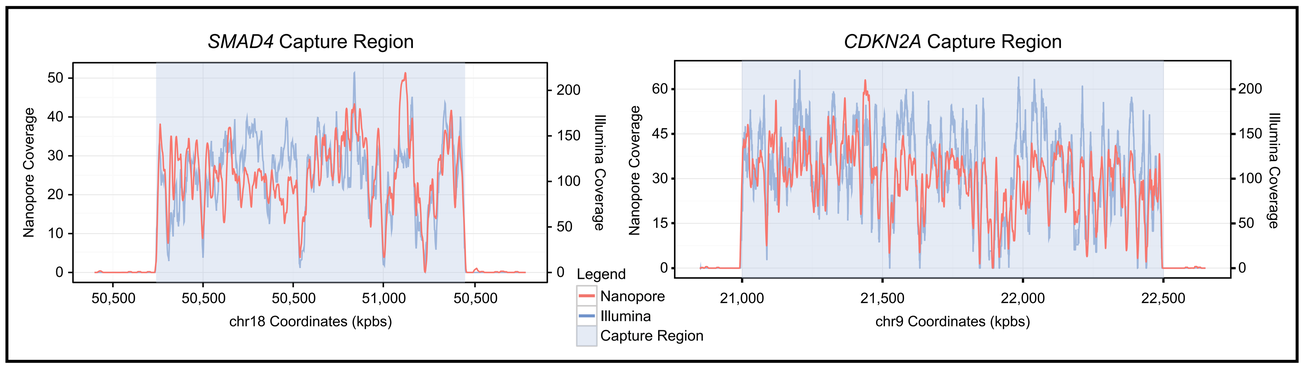
## Acknowledgments

Thanks to Oxford Nanopore for outstanding technical support, Agilent Technologies for working on the targeted enrichment optimization, Jared Simpson for assistance with nanopolish, and Dr. James Eshleman's group for providing the PDAC cell line DNA.

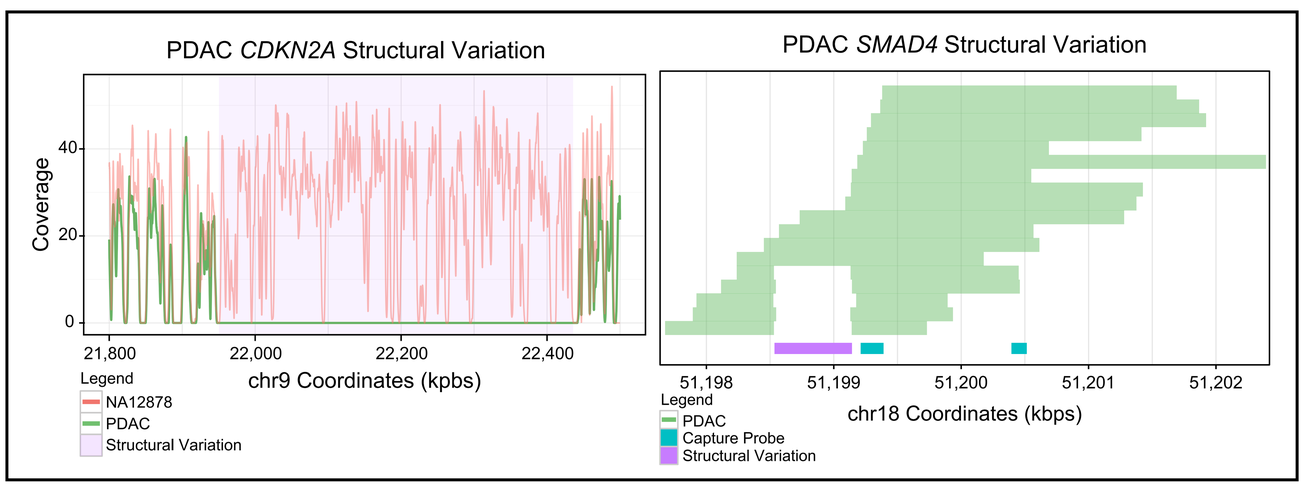
*References:*

1. Loman, NJ, et al. 2015. "A Complete Bacterial Genome Assembled de Novo Using Only Nanopore Sequencing Data." bioRxiv. doi:10.1101/015552.
2. Norris, AL., et al. 2015. "Transflip Mutations Produce Deletions in Pancreatic Cancer." Genes, Chromosomes & Cancer, May. doi:10.1002/gcc.22258.
3. Eberle, MA., et al. 2016. "A Reference Dataset of 5.4 Million Phased Human Variants Validated by Genetic Inheritance from Sequencing a Three-Generation 17-Member Pedigree." bioRxiv. doi:10.1101/055541.

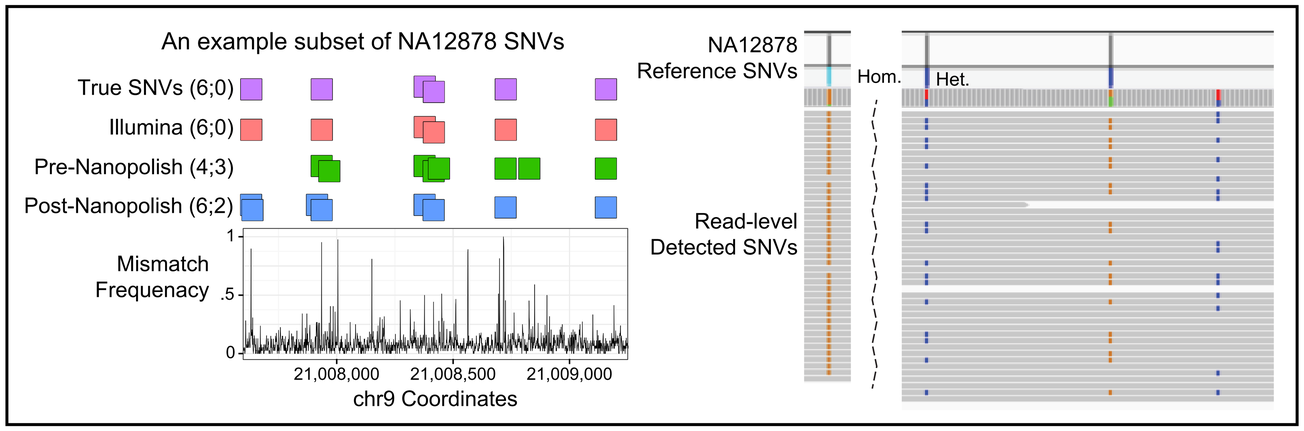
## Results




**Coverage of sequencing on long-read nanopore (minION) and short-read illumina sequencing (MiSeq) in the target regions**. Our targeted sequencing shows >300x fold-enrichment, >30% on-target percentage, resulting in >20x average coverage on both platforms.




**Structural variations detected using sniffles algorithm on the PDAC nanopore sequence reads.** SVs detected on PDAC DNA included the previously discovered SV in *CDKN2A* gene[2] and a novel putative SV in *SMAD4* gene.




**SNV detection at read-level resolution using nanopolish algorithm on the NA12878 nanopore sequence reads.** Nanopolish successfully reduces false positive SNV detection (numbers in parentheses indicate correct;incorrect) in comparison with the annotated NA12878 SNVs[3] and enables phased SNV analysis as shown by the distinction of homozygous versus heterozygous SNVs.

## Discussion

We found that our targeted sequencing method is effective in generating ~20x average coverage in our regions on a single MinION nanopore sequencing flowcell, sufficient to perform SV and SNV analysis. With recent improvements in yield, we should be able to multiplex and run multiple samples on one flowcell. Comparison of our SVs and SNVs detected on the control NA12878 DNA with control data confirms the validity of this approach, though sensitivity will likely be improved with increased sequencing depth. In application of this technique to the PDAC DNA, we confirmed a previously discovered SV and detected a new SV.

Nanopolish basecalling correction enables SNV detection with nanopore sequence reads - comparable to sequence reads from conventional platforms. With the ability to phase the variants using nanopolish-corrected reads, we identify patterns of mutation and allele-specificity of mutation along kilobase scale stretches of the genome. We believe this will be useful in enhancing the variant calling accuracy, detecting rare mutations, and/or deciphering novel relationships between SNVs. We are currently working to extend the read length and apply our method to more clinical samples.