Comprehensive detection of homologous recombination deficiency by Nanopore sequencing

Thissen R.[1*], Artesi M.[1], Durkin K.[1], Josse C.[2], Palmeira L.[2], Bours V.[1,2]

[1] Human Genetics Laboratory, GIGA Institute, University of Liège, Liège, Belgium

[2] Department of Human Genetics, Unilab, CHU of Liège, Liège, Belgium

Corresponding author*: rthissen@uliege.be

1. Introduction

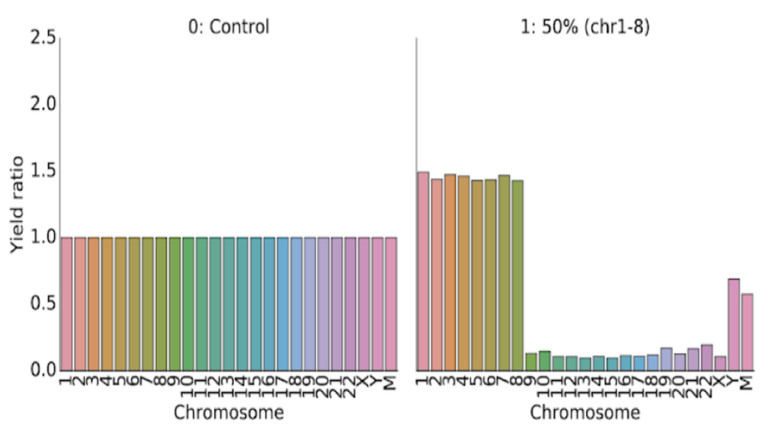
New treatments against cancers are emerging, one of them uses PARP inhibitors (PARPi). These PARPi have been FDA-approved in ovarian and partially in breast cancers treatment for patients carrying BRCA1 and/or BRCA2 mutations. In addition, some ovarian and breast cancers with no mutations on BRCA1 and BRCA2 show also a good response to PARPi treatment. Indeed, these PARP inhibitors are also effective against ovarian and breast cancers presenting a homologous recombination repair deficiency (HRD). It is likely that other types of cancers with a HRD could also benefit from these new treatments.

2. Goals

Here, we are trying to assess multiple **HRD biomarkers** in one straightforward experiment:

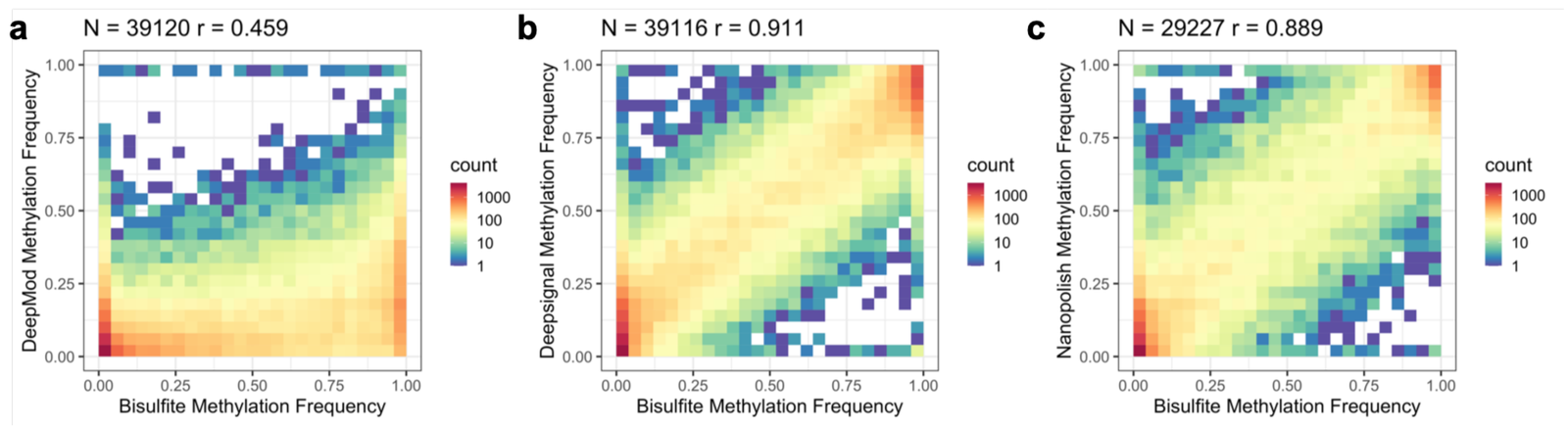
- **HRD pathway genes methylation**
- **Mutational signatures**
- **HRD pathway genes mutations** (SNPs)
- **Genomic scars** (CNVs)

We use the **Nanopore adaptive sequencing** which allow to choose the targets without the need of a dedicated library preparation. The **adaptive sequencing** ejects reads mapped outside the target regions to sequence only those mapped in the target regions [1].



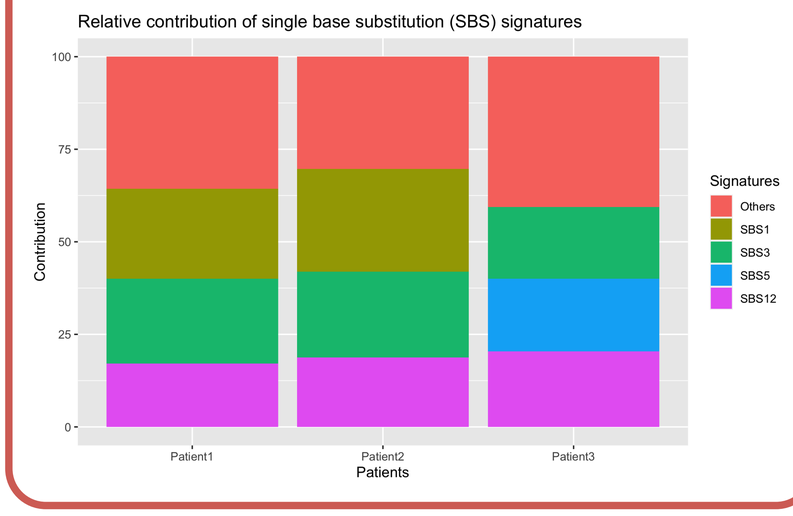
3. Methylation

We have compared three different tools (**Deepmod, Deepsignal and Nanopolish**). This has been done on the chromosome 20 from 30X coverage public data; the GM12878 GIAB sample. Results from each tool have been compared with gold standard bisulfite sequencing data for the same region. **N** is the number of **CpG sites** and **r** is the **Pearson correlation**.



4. Mutational signatures

We managed to bring to light the presence of the signature 3 in whole exome sequencing from three colorectal cancers with a pathogenic BRCA1 mutation. The **signature 3** is known to be linked with **HRD** [2]. In each sample, the **signature 3** is in the top 3 of the **largest contribution** signatures. Signatures 1 and 5 are known to be the **"clock-like" signatures** because they are related to the age of the patient. These "clock-like" signatures are almost always present in samples. The signature 12 has no aetiology linked to it yet.



5. SNPs and CNVs

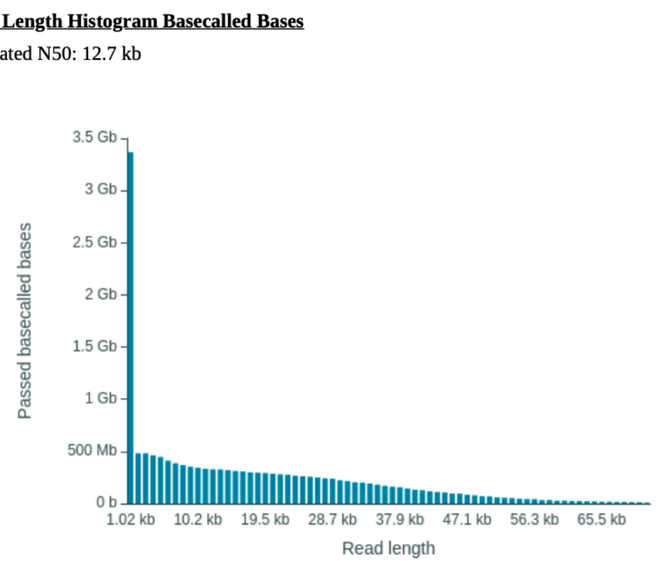
We used the **adaptive sequencing** to target the **whole exome** during the sequencing of the GM24385 GIAB sample with one Nanopore flow cell. We obtained a **13.5X** coverage on targets and **4X** on the rest of the genome. We used the **pipeline PEPPER** to call SNPs [3]. Despite the low coverage, we managed to obtain 92% of precision and 84% of recall.

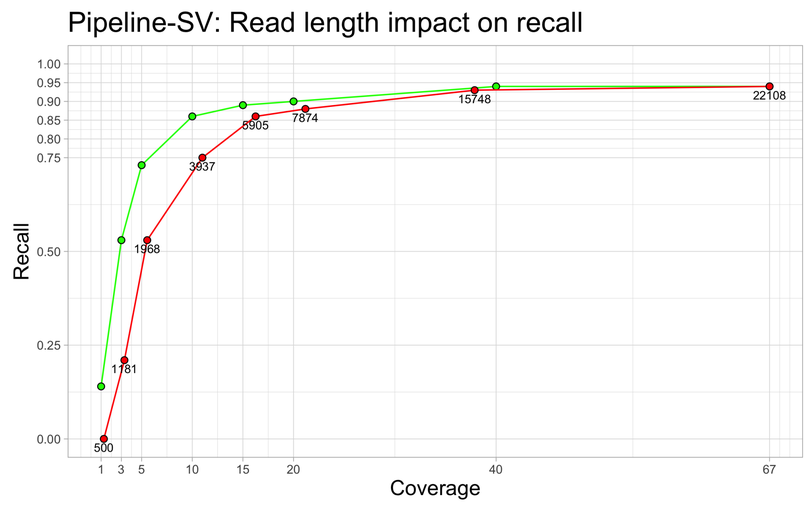
We have used the pipeline **pipeline-structural-variation (Pipeline-SV)** from Oxford Nanopore to call CNVs on the 4X whole genome sequencing. Unfortunately, the recall obtained was lower than expected. Reads from the whole genome are smaller (400bps) than those in target regions, this is inherent to the adaptive sequencing method.

Thus, we assessed if the length of reads has got an impact on Pipeline-SV efficiency.

- Data used are a public Nanopore benchmark from the GM24385 GIAB sample. Raw data has a 67X coverage and a mean read length of 22108bps.
- **In green**, we ran Pipeline-SV on these data at different coverage by downsampling the raw data.
- **In red**, we ran Pipeline-SV on the same data but at different length of read by cutting all reads in order to have the same different coverage than before.

We didn't see any impact of read length on the precision of Pipeline-SV but there was a clear impact on recall/sensitivity, see below.





6. Conclusions

We showed that we are able to detect **methylation, mutational signatures and SNPs** efficiently. Nonetheless, the small length of reads outside of target regions is an issue for the CNVs detection. A solution could be to use a small read adapted tool relying on the computation of **read depth in bin**.

In the future, we will work on the detection of **genomic scars** which goes hand in hand with that of CNVs. In addition, we are going to begin the sequencing of **fresh tumor samples** to perform these analyses on real cancer data.

7. References

[1] Loose M. Nanopore adaptive sequencing for mixed sample, whole exome capture and targeted panel. *bioRxiv:926956*, 2020.

[2] Stratton M. R. The repertoire of mutational signatures in human cancer. *Nature*, 2020.

[3] Shafin K. Haplotype-aware variant calling enables high accuracy in nanopore long-reads using deep neural networks. *bioRxiv:433952*, 2021.