Analysis of the Elements Involved in the Enrichment of a Panel of Genomic Regions by Nanopore Sequencing Using Adaptive Sampling

Llamas-López M[1], Garrido-Rodríguez P[1], Cuenca-Guardiola J[2], Navarro Manzano E[1], Padilla Ruiz J[1], Ayala F[1], Macías JA[1], de la Morena Barrio ME[1], Fernández-Breis JT[2], Lozano ML[1], de la Morena Barrio B[1], Corral J[1]

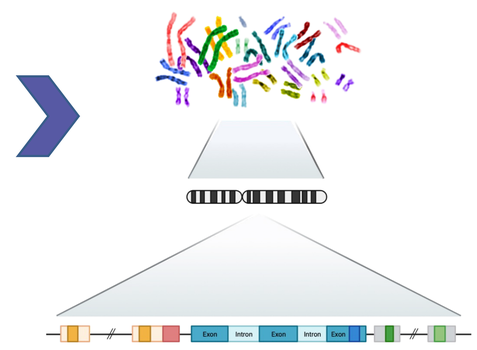
[1] Servicio de Hematología, Hospital Universitario Morales Meseguer, Centro Regional de Hemodonación, Universidad de Murcia, IMIB Pascual Parrilla, CIBERER-ISCIII, Murcia, Spain
[2] Departamento de Informática y Sistemas, Universidad de Murcia, IMIB Pascual Parrilla, Murcia, Spain

*Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

Background

One of the main limitations of nanopore sequencing is the **low coverage** obtained.

Different methods may increase coverage, but **adaptive sampling** is the only one which does not require manipulation and can be used in MinION.



Real-time targeted sequencing

AIM: To dissect the factors involved in the **enrichment** process using adaptive sampling

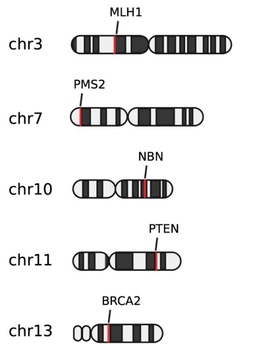
Materials & Methods

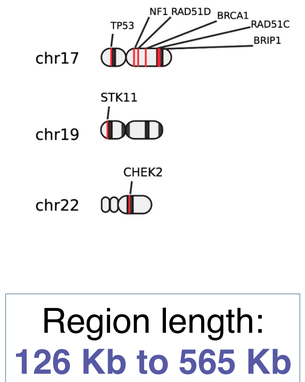
Samples

Sequencing: 18 regions involving 20 genes

Negative for NGS screening

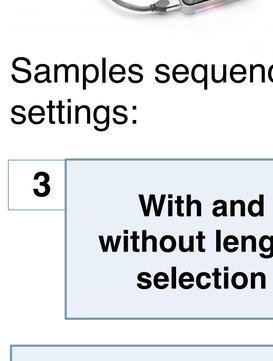
16 cases with familial breast cancer





Region length: **126 Kb to 565 Kb**

Nanopore sequencing



8	**Length selection (>1,000 kb)**
5	**Without length selection**

Samples sequenced **twice** with **two different** settings:

3	**With and without length selection**	1	**DNA fragmentation (g-TUBE Covaris® for 6 Kb fragments)**

Different sequencing times

Results

Coverage reached with this method:	**Region of Interest 5.1x**	Up to **13.0x**	**Rest of the Genome 2.0x**	Up to **3.2x**

We have analyzed the **parameters** and **elements** potentially involved in the enrichment process by using two different approaches:

1) Analysis of samples:

Long Read Selection

$$Ratio = \frac{Coverage\ area}{Coverage\ genome}$$

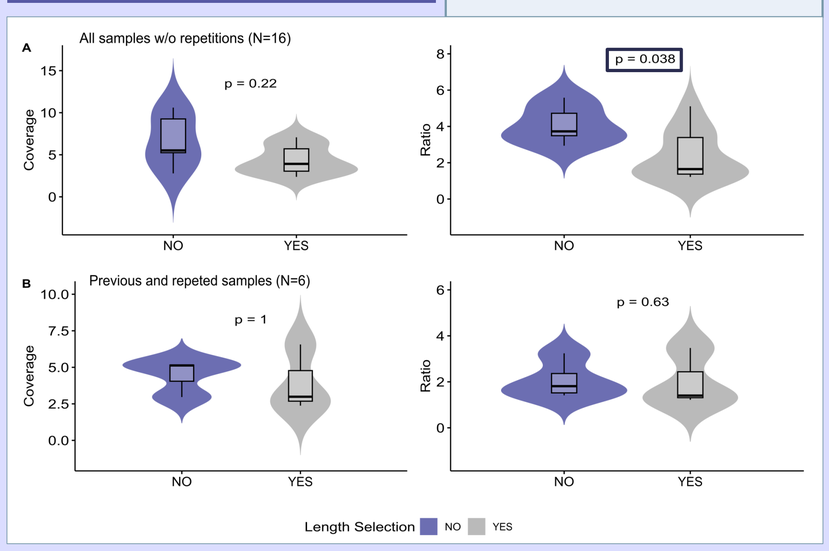


Figure 1. Average coverage and enrichment (ratio) with or without long read selection. **A)** Full cohort of samples **B)** Three Paired samples

Sequencing time

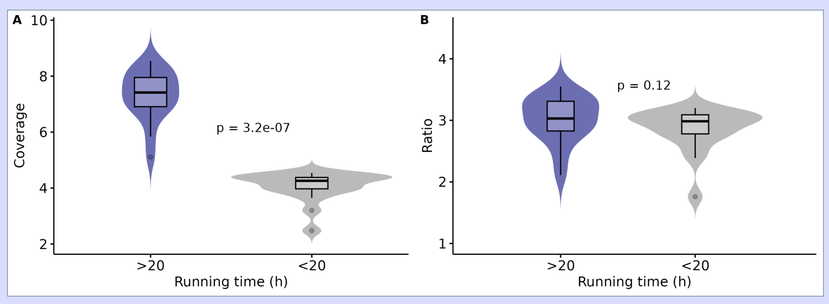


Figure 3. Average coverage **A)** and enrichment **B)** as a function of the running time

Sequencing time (>20 h) is the main element improving the coverage of selected regions. Likewise, it also improves coverage in the rest of the genome.

Selection of long reads (>1,000 Kb) and **DNA fragmentation** did not improve the enrichment yield.

2) Analysis of regions:

One region was **less covered: *PMS2.***

	PMS2	**VS**	**Rest of the regions**
Average	**3.3x**		**4.5x**

Fragmentation

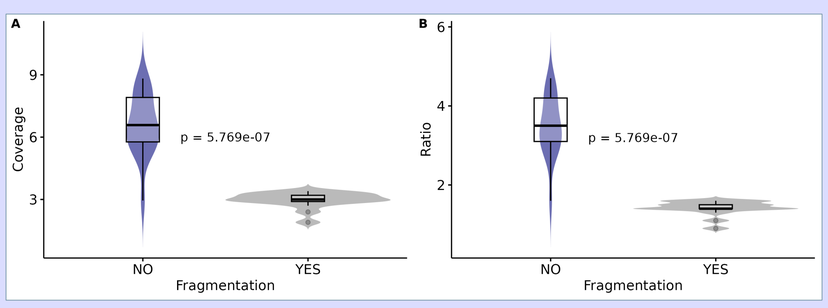


Figure 2. Average coverage **A)** and enrichment **B)** of regions of interest according to DNA fragmentation

Elements evaluated

Repetitive elements | **Coding regions** | **Homopolymers** → Did **not affect** the process

Pseudogenes

NF1	*CHEK2*	***PMS2***
12	7	**14**

Table 1. Regions with a high number of pseudogenes

Coverage of other pseudogenes: Similar or higher than the gene of interest

High homology between genes and pseudogenes

Conclusions

Analysis of factors that could influence the effectiveness of adaptive sampling in a panel of 18 genomic regions using the accessible MinION device revealed these **two factors as the most relevant** for providing **higher coverage** in the areas of interest:

1. **Sequencing time**
2. **Presence of pseudogenes**

Contact: maria.llamasl@um.es; pedro.garridor@outlook.es; belendelamorenabarrio@gmail.com

Funding:

PMP21/00052 21642/PDC/21