# Quality metrics evaluation of Ion S5, MiSeq and PacBio HLA typing data from Brazilian samples

**Stelet VN[1], Cita RF[2], Romero M[1], Mendes MAF[2], Abdelhay ESFW[1]**
1. Instituto Nacional de Cancer José Alencar Gomes da Silva / MS
2. Fundação Pio XII Barretos

*Author contact information: vinicius.stelet@inca.gov.br

## INTRODUCTION

Next generation sequencing HLA typing provides greater throughput, higher typing resolution and less ambiguity incidence when compared to other methodologies. One long- and two short-reads NGS platforms have been replacing Sanger sequencing on clinical high-resolution HLA typing routine.

Sanger typing agreement analysis has been used to validate Sanger to NGS transition. However, **quality metrics evaluation is the best way to check typing reliability, compare different NGS platforms and choose the best one to replace Sanger typing.**

## OBJECTIVE

To determine the best NGS platform for HLA typing service, we compared NGS data quality metrics from **three NGS platforms (IonS5, Thermo Fisher; MiSeq, Illumina; RS-II, Pacific Biosciences**) running six commercial HLA typing protocols (NxType, AllType; Holotype, NGS-Go, Trusight; Histogenetics).

## METHODS

**684 samples** were typed for at least six classical HLA genes at **Instituto Nacional de Cancer** (Ion S5, n=128), **Fundação Pio XII Barretos** (MiSeq; n=386) or **Histogenetics** (PacBio; n=170) HLA typing services. FastQ files were submitted to NGSEngine algorithm using manufacturer NGS specific thresholds.

**Mappability percentage, read count and read length were calculated** for six classical HLA genes (HLA-A, -B, -C, -DRB1, -DQB1 and -DPB1). **stimated second allele percentage (ESA), maximum noise level (MNL) and delta signal to noise (DSN) were also determined** for core regions (Class I: exons 2 and 3; Class II: exon 2).

## RESULTS

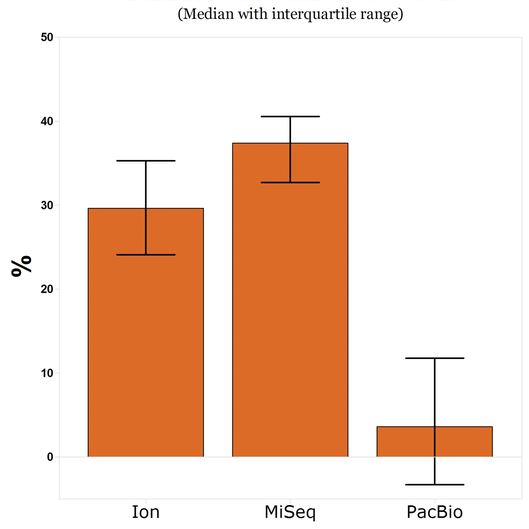
**MiSeq** data showed **better performance** followed by IonS5 and PacBio **for ESA** (44,79%; 44,04%; 40,18%), **MNL** (3,8%; 9,2%; 29,2%) and **DSN** (36,9%; 29,5%; 3,6%).

**IonS5 generated more reads (9666)** than MiSeq (9287) and PacBio(715);

**Short reads sequencers** (Ion S5 and MiSeq) showed **greater mappability (>95%)** than long reads (PacBio; 40%) ;

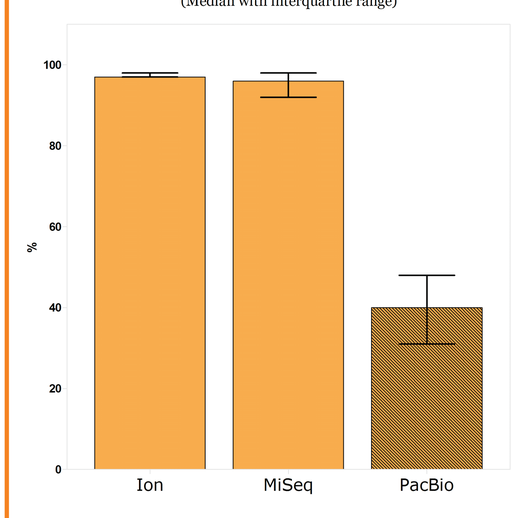
**PacBio showed the largest reads** (2713 bp) compared to Ion (252bp) and MiSeq (135bp).

**DELTA SIGNAL TO NOISE**
(Median with interquartile range)



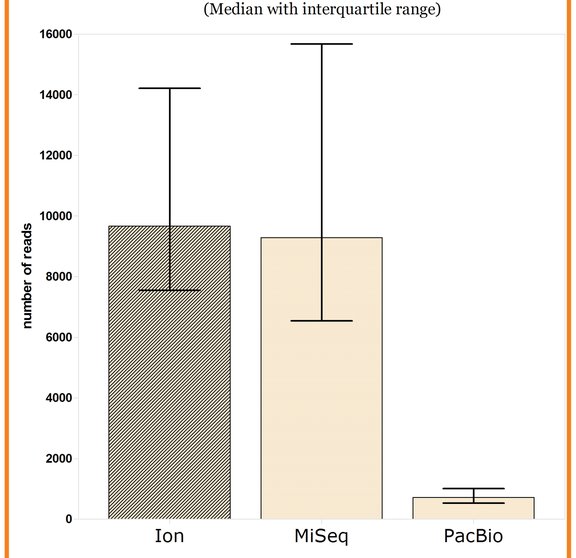

The difference between the highest noise position and the lowest heterozygous position in %.

**MAPPABILITY**
(Median with interquartile range)



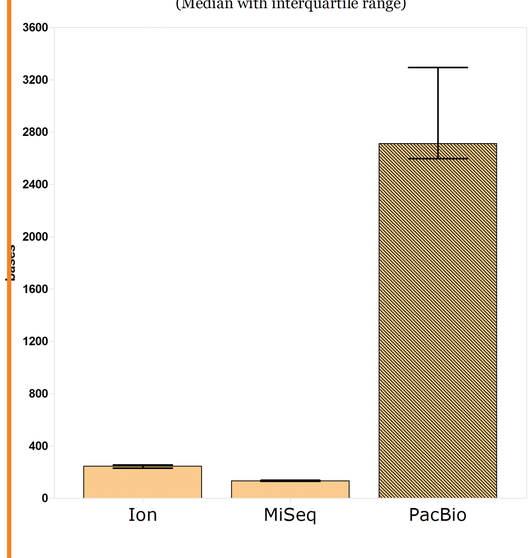

Percentage of reads that could be aligned to the reference.

**READ COUNT**
(Median with interquartile range)




The median of the number of reads covering every single position.

**READ LENGTH**
(Median with interquartile range)




The median read length of the accepted reads

## DISCUSSION

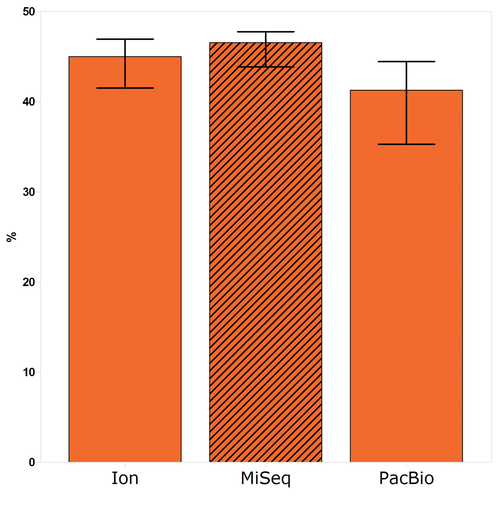
To reach **reliable HLA typing** is necessary high number of **high quality reads** to **lower MNL, elevates DSN levels**and, consequently,avoiding allele imbalance/dropout and false new allele calls.

## CONCLUSION

**Short reads** platforms presents **better quality metrics results** for HLA typing in comparison with long reads.

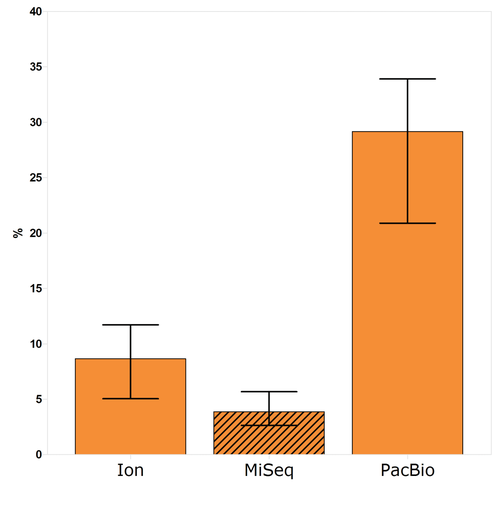
As quality metrics analysis describes how safely each platform determines HLA typing, our data suggest that **MiSeq commercial protocols showed the safer path to HLA typing** by generating high quality reads associated to lower sequencing error rate.

**ESTIMATED SECOND ALLELE**
(Median with interquartile range)




**ESA:** The median percentage of the heterozygous positions of the allele with the lowest abundance.

**MAXIMUM NOISE LEVEL**
(Median with interquartile range)




The difference between the highest noise position and the lowest heterozygous position in %.

Projeto Gráfico: Área de Edição e Produção de Materiais Técnico-Científicos / INCA

MINISTÉRIO DA SAÚDE