HLA Typing by Next Generation Sequencing (NGS)

Validation using MGI DNBSEQ-G99

Silva, LR; Pereira, LCG; Mion, M
Biometrix Diagnóstica, Curitiba, Paraná, BRAZIL

INTRODUCTION

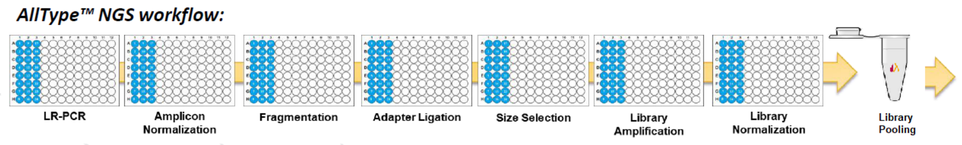
MGI, part of the BGI Group (Beijing Genomic Institute), was established in 2016 and has emerged as a significant force in the field of genomics, offering innovative solutions for DNA and RNA analysis. Like other sequencing platforms, MGI has its own sequencing technology based on DNA nanoballs (DNBs), capable of generating a high volume of data. By combining DNBseq with advanced bioinformatics technologies, MGI has reached new levels of efficiency and precision in genomic analysis [1].

The introduction of the G99 equipment in 2021 to the MGI portfolio represents a significant leap forward in genomics. This equipment has the capability to generate a large amount of data in record time, unprecedented for a low to medium throughput sequencer, capable of performing a PE150 sequencing in just 12 hours and generating approximately 80 million of reads [2].
The objective of this study was to evaluate the DNBseq sequencing chemistry using the G99 equipment and allele calling for HLA typing, aiming to understand better how this can impact histocompatibility laboratories in Brazil.

METHODOLOGY

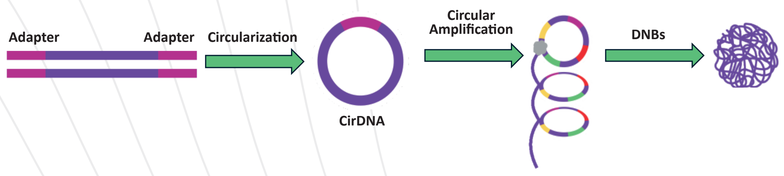
The samples used in this evaluation contained know HLA genotypes obtained by a high-resolution technology, previously validated. Samples were chosen from the Quality Control Brazilian Program provided by the Brazilian Association of Histocompatibility and Immunogenetics (ABHI). Additionally, in collaboration with some Brazilian laboratories, samples containing rare and null alleles were included. Samples containing allelic groups with a tendency for dropout, such as DQB1*02 and DQB1*03, were also selected. All DNA samples were isolated using the Biopur Isolation Kit and normalized to the concentrations between 20 and 30 ng/µL, based on Qubit quantification.
For the amplification of target genes and library preparation, the AllType™ NGS 11 loci kit from One Lambda, Thermo Fisher Scientific (Anvisa: 80298490139) was used. The loci evaluated were HLA-A, B, C, DRB1, DRB345, DQB1, DPB1, DQA1, and DPA1. The library preparation process followed the manufacturer's instructions accordingly:

AllType™ NGS workflow:



In total, 4 libraries were prepared, which were combined in different ways to generate different throughputs. After preparation, the libraries were measured using the TapeStation System to check the distribution profile of fragment sizes.

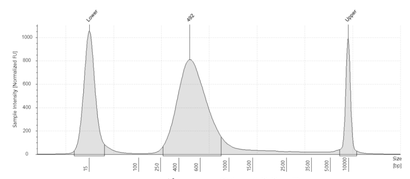
Since the adapters provided with the library preparation kit are structured for sequencing on the Illumina platform, it was necessary to perform the conversion before preparing the DNBs and ligation. The required conversions were carried out using the Universal Library Conversion Kit (App-A).



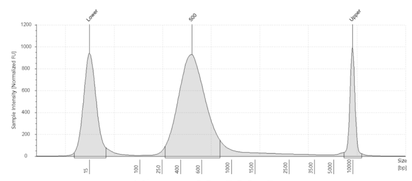
The results obtained were analyzed using the TypeStream™ Visual NGS Analysis Software (Anvisa: 80298490140) version 3.0, with IMGT 3.52 used as reference for the alignment.

RESULTS

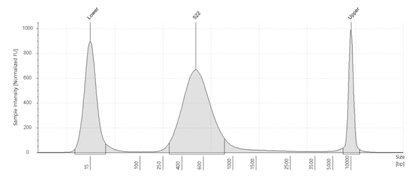
All libraries after completion were measured on the TapeStation System, the distribution profile of the 4 libraries were as follows:



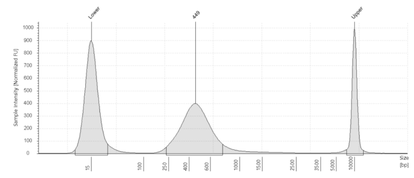
Library 01: 492pB



Library 02: 500pB



Library 03: 522pB



Library 04: 449pB

Each library contained 96 samples and they were combined equimolarly in 4 different ways to generate different throughputs.

- Run 1 (96 samples): Library 01
- Run 2 (192 samples): Library 01 + Library 02
- Run 3 (288 samples): Library 01 + Library 02 + Library 03
- Run 4 (384 samples): Library 01 + Library 02 + Library 03 + Library 04

For the run with 96 samples, a final %Q30 of 91.3% and 104.3 million reads were obtained. In the run with 192 samples, a %Q30 of 93.82% and 119.90 million reads were obtained. In the run with 288 samples, the sequencing metrics were %Q30 at 86.19% and 75.16 million reads. Finally, in the last run of the validation process containing 384 samples, a %Q30 of 93.57% and a production of 108.87 million reads were obtained. Considering that the minimum percentage promised by the MGI manufacturer for the flow cell/cartridge in use is 85% and the maximum reads is 80 million, we consider the quality and quantity of generated data were satisfactory. Regarding the number of reads, it was possible to obtain almost 50% more than the maximum promised. Additionally, it can be observed that the distribution of the Q score throughout the cycles tends to remain quite stable.

CONCLUSION

After analyzing 1,228 samples, 13,508 genotypes and 27,016 alleles were identified. It is possible to state that there was a concordance rate of 99.97% with the pre-existing HLA typing data of the samples used.
It can be concluded that the runs performed with the G99 were able to obtain satisfactory and high-quality results, using 96, 192, 288, and 384 samples on a single flow cell. The ability to generate a high volume of data in such a short sequencing time can be a more of sequencer option.

REFERENCES

[1] Jeon, S.A., Park, J.L., Park, SJ. et al. Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. Genes Genom 43, 713–724 (2021).
[2] Wu, Changcheng et al. Rapid identification of full-length genome and tracing variations of monkeypox virus in clinical specimens based on mNGS and amplicon sequencing. Virologica Sinica 39, 134-143 (2024)

www.biometrix.com.br