Whole Exome Sequencing of Clinical Samples on the GenapSys NGS Platform and Performance Comparison to Common Industry Technology

Hesaam Esfandyarpour, Ali Nabi, Hooman Nezamfar, Kosar B. Parizi, Mohammad Fallahi, Narin Tangprasertchai, Bin Dong, Srijeeta Bagchi, Eric LoPrete, Maryam Jouzi, Anthony Thomas, Meysam Rezaei Barmi, Ali Mazouchi, Lisen Wang, HamidReza Golnabi, and Saurabh Paliwal

Abstract

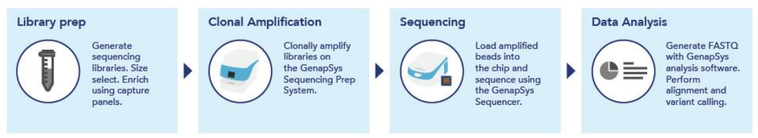
Introduction: Advances in Next Generation Sequencing (NGS) technologies and NGS-based diagnostic applications have ushered in a new age of clinical testing. These clinical diagnostic applications include germline and somatic mutation testing, non-invasive prenatal testing etc. Clinical testing is typically performed at centralized laboratories, partly due to the large capital costs of NGS platforms and high run costs, which necessitates pooling of large numbers of samples. There is a strong need for decentralized NGS-based clinical testing, e.g. at hospitals, enabling cheaper and faster results for doctors and patients. The GenapSys Sequencing Platform offers an accurate solution with low capital equipment and run costs, making it ideal for point-of-care testing. In this study, the GenapSys™ system performance was characterized by sequencing clinical samples from collaborators such as Novogene, and compared to Illumina technology.

Methods: GenapSys NGS libraries were generated from patient samples, and sequencing metrics, as well as germline and somatic mutation calling, were compared with Illumina sequencing. Genomic DNA was extracted by Novogene from patient FFPE & Blood samples, and by another collaborator from patient fresh frozen tumor and blood samples. Genomic libraries were generated following mechanical fragmentation, adapter ligation, size selection and PCR. Hybrid capture-based enrichment on GenapSys libraries was done using the IDT Exome Research panel (39 Mb region, 19,396 genes) or the IDT Pan Cancer ver1.5 (800 Kb region, 127 genes) panel to generate Whole Exome Sequencing (WES) or Pan Cancer libraries respectively. WES was performed for both FFPE and blood sample libraries on the GenapSys and Illumina systems, with typical coverage >100x. Cancer panel sequencing has typical coverage >600x. Sequencing reads were aligned to the hg38 reference genome using BWA-MEM. Variant calling analysis involved additional training of the Google DeepVariant model based on GenapSys sequencing data.

Results: Comparison of GenapSys and Illumina sequencing demonstrated high concordance, with the F1 score of ~95% on SNV detection in high confidence regions of the Exome libraries. Average read length of GenapSys sequencing was 150 bp and the WES library on-target rate was > 85%.

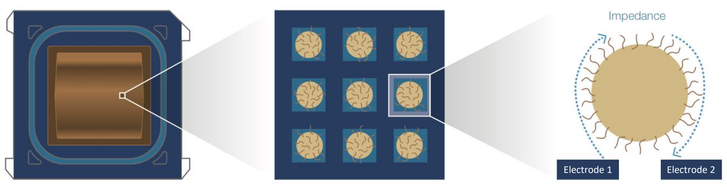
GenapSys Sequencing Workflow

Sequencing Workflow Overview



Genomic DNA is randomly sheared and converted into a sequencing library via ligation of adapters using standard industry methods. Individual library molecules are clonally amplified onto beads and loaded into the sequencing chip. Automated sequencing is carried out on the GenapSys Sequencer which utilizes cloud-based analysis methods that ultimately deliver FASTQ and variant reporting files.

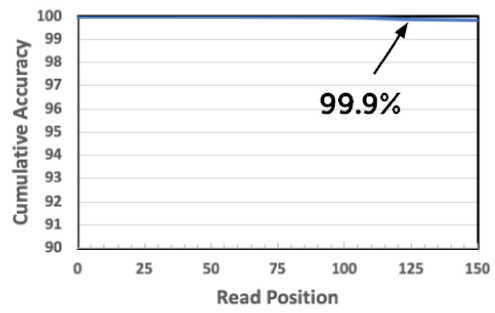
Detection methodology



The CMOS sequencing chip contains millions of individual sensors, which are each loaded with a clonally amplified bead. The electrodes in each sensor are capable of measuring minute changes in impedance when nucleotides are incorporated opposite the bead-bound templates.

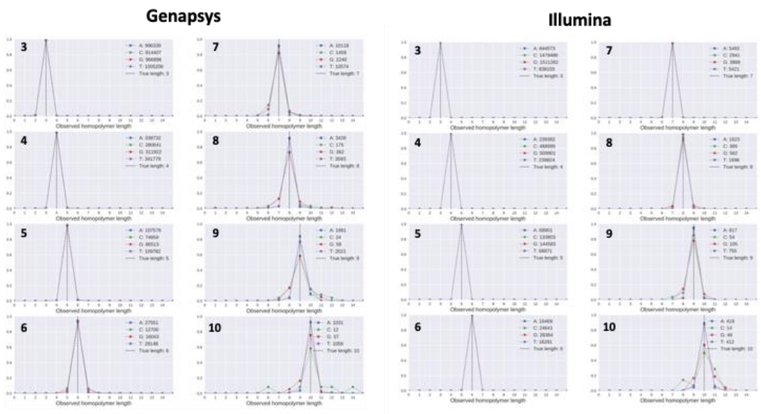
Sequencing Performance

Accuracy



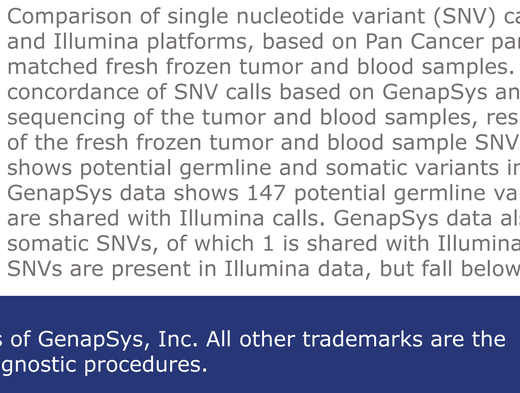
Cumulative raw accuracy profile of a typical sequencing run

Homopolymer Performance



Using data from a human exome sequencing run, base calls from homopolymer regions (3-10 nt) were compared to the hg38 reference for each individual nucleotide. Performance is shown relative to the same library sequenced with Illumina.

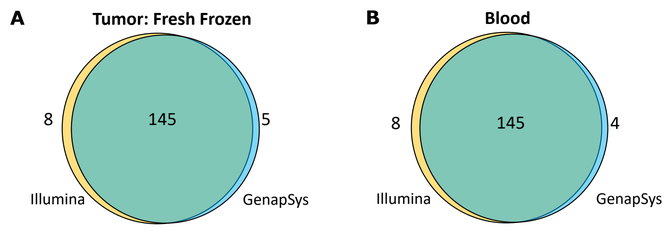
Read Length



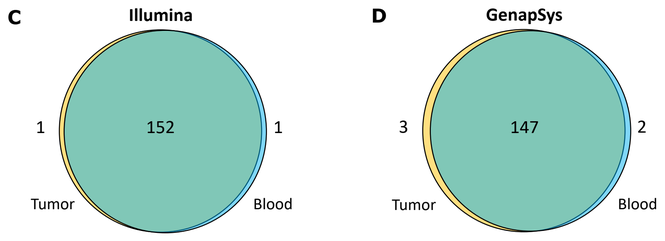
Distribution of read length for a typical sequencing run.

Cancer Panel Sequencing of Tumor Fresh Frozen & Blood Samples

Concordance of SNV calls: Pan Cancer Panel



Germline and Somatic Variant Calling



Comparison of single nucleotide variant (SNV) calls with the GenapSys and Illumina platforms, based on Pan Cancer panel sequencing of matched fresh frozen tumor and blood samples. **(A) & (B)** show high concordance of SNV calls based on GenapSys and Illumina NextSeq sequencing of the tumor and blood samples, respectively. Comparison of the fresh frozen tumor and blood sample SNV calls (AF > 2%) shows potential germline and somatic variants in **(C) & (D)**. GenapSys data shows 147 potential germline variants, of which 144 are shared with Illumina calls. GenapSys data also shows 3 potential somatic SNVs, of which 1 is shared with Illumina and the other two SNVs are present in Illumina data, but fall below the 2% threshold

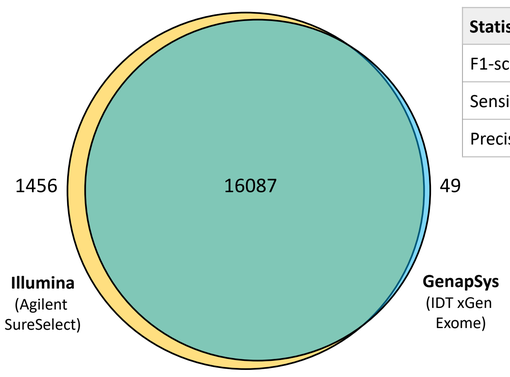
Whole Exome Sequencing Metrics

Statistic	FFPE	Blood
Q20 Bases	99.5%	99.4%
Q30 Bases	84.9%	85.2%
Mapping Rate	99.9%	99.9%
Fraction Covered at 10x	97.1%	94.9%

Whole exome sequencing was carried out on multiple cancer patient samples. The cancer results are extracted from an FFPE tumor biopsy and normal control comes from a blood draw. The GenapSys sequencing data demonstrated high accuracy on libraries made from both FFPE and blood, typical sequencing metrics are shown.

Variant Calling Comparison: WES of Tumor FFPE sample

Tumor FFPE – Whole Exome Sequencing



Statistic	FFPE
F1-score	95.5%
Sensitivity	91.7%
Precision	99.7%

Comparison of Single nucleotide variant (SNV) calls with the GenapSys and Illumina platforms, based on whole exome sequencing of a tumor FFPE sample. The Genapsys library was generated using the IDT Exome Research panel and the Illumina library is generated using an Agilent SureSelect exome panel. SNV calls in the regions of overlap are reported here, and show high concordance with an F1-score of 95.5%, Sensitivity of 91.7%, and Precision of 99.7%.

Conclusions

This study demonstrated that the GenapSys Sequencing Platform generated comparable sequencing performance to a common industry NGS technology. It identified germline and somatic mutations in clinically relevant WES libraries and Pan Cancer libraries generated from patient Tumor FFPE, Fresh Frozen Tumor and Blood samples. Thus, our results show promise for use of the GenapSys Sequencing Platform in decentralized clinical testing applications.

The GenapSys Sequencing Platform offers:

- Low price per run and low price per sample
- Up to 2 Gb of highly accurate DNA sequence per run
- Highly accurate sequencing data yields high confidence variant calling
- Excellent concordance for variant calling relative to established sequencing technologies