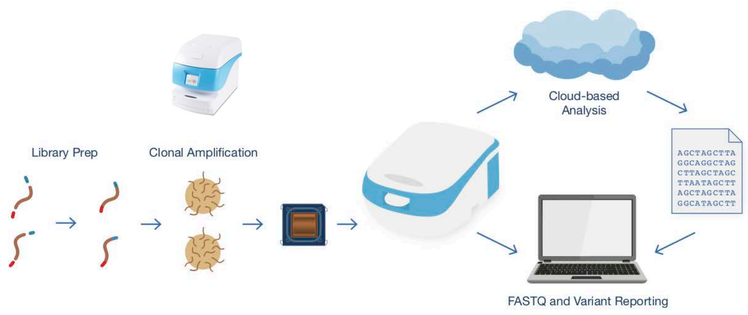
Applications of an Electrical-Based Sequencing Method

Ali Nabi, Tyson A. Clark, Saurabh Paliwal, Mohammad Fallahi, Hooman Nezamfar, Srijeeta Bagchi, Bin Dong, Eric LoPrete, Xavier Gomes, Maryam Jouzi, Meysam Razaei Barmi, Hannah Ritchie, Subra Sankar, and Hesaam Esfandyarpour

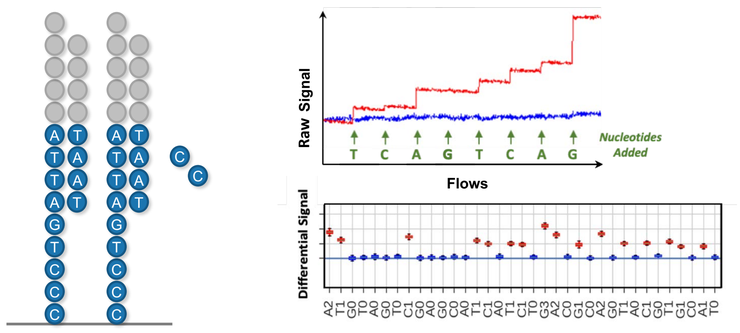
Abstract

The GenapSys Sequencing Platform uses an impedance-based, electronic detection modality on CMOS chips. A single run on a 16 million sensor chip produces 10-14 million individual sequences with read lengths that average 150 base pairs. With this chip, the platform is capable of generating ~2 Gb of sequence per run with more than 80% of calls >Q30 quality. The system is ideal for a broad range of genomic applications, as demonstrated here.

GenapSys Sequencing Workflow



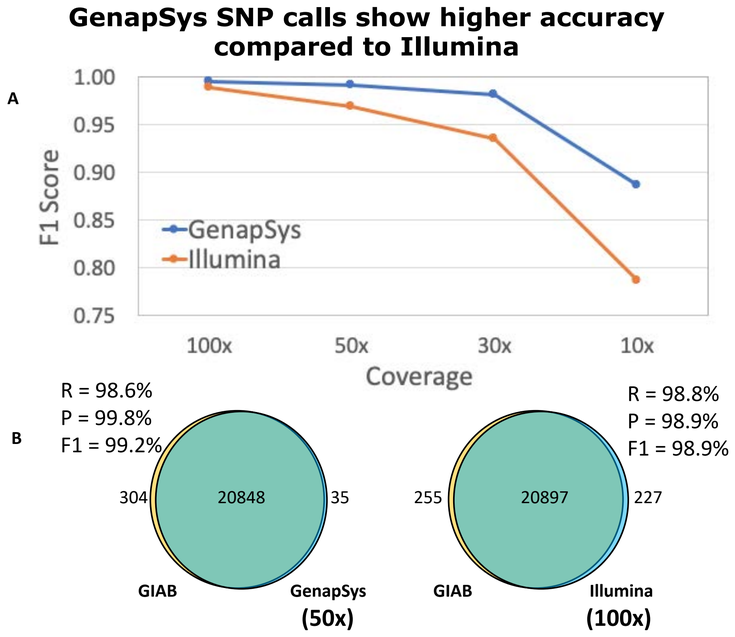
Genomic DNA is converted into a sequencing library via ligation of adapters using standard industry methods. Individual library molecules are clonally amplified onto beads and loaded into the sequencing chip. Automated sequencing is carried out on the GenapSys Sequencer which utilizes cloud-based analysis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



Nucleotides are injected one base at a time. When a nucleotide is incorporated, the measured impedance value of that sensor will jump, creating a graph that resembles a staircase. The magnitude of the differential signal correlates with the number of incorporated nucleotides. Due to the steady state nature of the impedance metric, multiple measurements can be taken to improve precision.

Whole Exome Sequencing

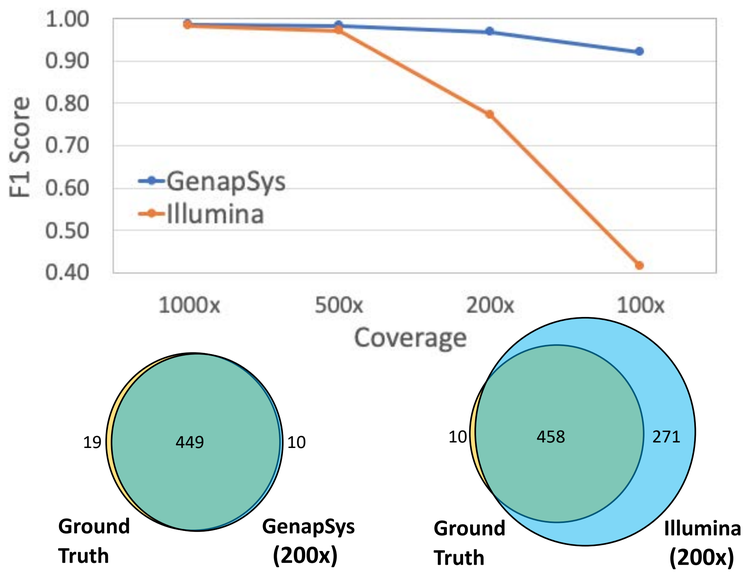
GenapSys SNP calls show higher accuracy compared to Illumina



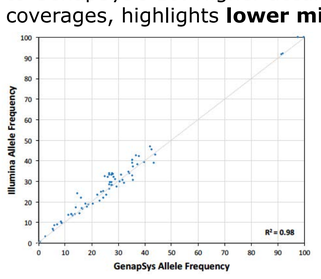
A The NA12878 WES library was sequenced on both the GenapSys Sequencer and Illumina NextSeq. Single Nucleotide Polymorphisms (SNPs) relative to the hg38 reference were called using DeepVariant trained on respective technologies' data. **GenapSys SNP calls show higher F1-score accuracy as compared to Illumina**, especially at lower coverages, highlighting lower mismatch error rates in GenapSys sequencing data. **B** GenapSys results show better concordance with the GIAB high-confidence variants (IDT xGen Exome panel) at half the coverage compared to Illumina.

Somatic Variant Calling

GenapSys SNV calls show higher accuracy compared to Illumina on Hybrid Capture Cancer Panel



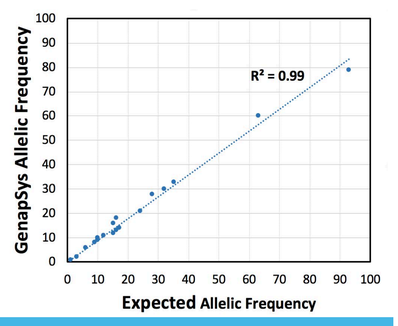
The Horizon HD827 Oncospan reference DNA standard was used to generate a library using the IDT xGen Pan Cancer panel v1.5 and sequenced on both the GenapSys Sequencer and Illumina. Single Nucleotide Variants (SNVs) down to 2% were called using VarDict. Ground truth vcf file was generated based on the common variants detected in high coverage Illumina and GenapSys sequencing data. At 200x, the Illumina data shows 271 false positive calls compared to GenapSys's 10. Higher accuracy in GenapSys data, especially at lower coverages, highlights **lower mismatch error**.



At 1000x coverage, high concordance of allele frequencies is seen between GenapSys and Illumina sequencing of the HD827 Oncospan reference DNA standard (R^2 = 0.98).

Amplicon PCR Cancer Panel

Predicted vs Expected allelic frequencies based on GenapSys sequencing of a library generated using the AmpliSeq Cancer Hotspot panel v2 for HD827. The expected allelic frequencies are quantified using ddPCR (R^2 = 0.99).



Clinical samples: Tumor FFPE, Fresh Frozen, Blood

Whole Exome Sequencing of Tumor FFPE samples

Statistic	FFPE
Average Depth (GenapSys)	201x
Fraction Covered at 10x (Gen)	98%
F1-score	95.5%
Precision	99.7%

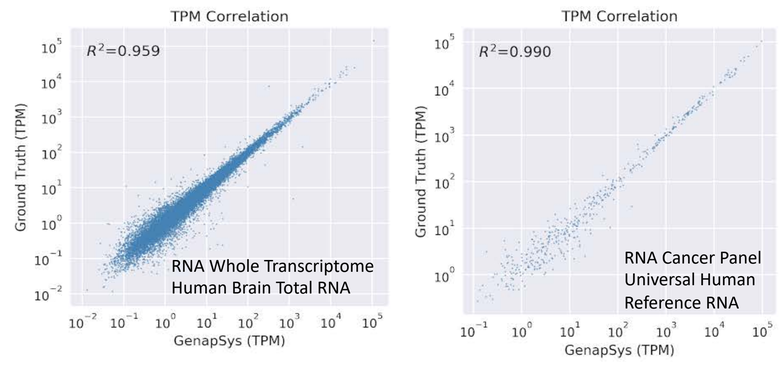
Concordance of SNVs: IDT xGen Pan Cancer Panel v1.5

Germline and Somatic Variant Calling (AF > 2%)

Concordance of SNV calls between the GenapSys and Illumina platforms, based on WES of a tumor FFPE sample (Top), Pan Cancer panel libraries of matched fresh frozen tumor and blood samples (Middle), and potential germline and somatic variant calls. GenapSys data shows 3 potential somatic SNVs (all 3 found in Illumina data, with only 1 SNV > 2% AF).

RNA-seq and sRNA-seq

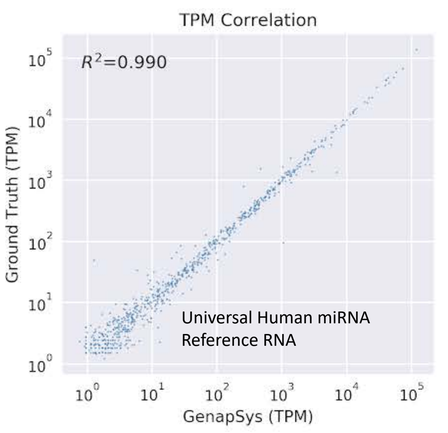
RNA-seq:
Full-transcriptome RNA-seq libraries (full transcriptome and targeted gene expression) show high correlation (R^2>95%) with Illumina sequencing



Full-transcriptome libraries were generated with the Human Brain Total RNA (left) using the Lexogen SENSE mRNA-Seq Library Prep Kit v2. Full transcriptome RNA-Seq libraries were were generated with the Universal Human Reference RNA and enriched for cancer-related genes using the IDT Pan Cancer v1.5 hybrid capture panel (right). Human transcriptome FASTA file (hg38, release 33) was downloaded from GenCode. Index files were generated using Salmon (v1.1.0) based on the transcriptome FASTA file. Adapter sequences in the fastq files were removed using cutadapt. Quantification of RNA was performed using Salmon (v1.1.0) based on the adapter-removed fastq files.

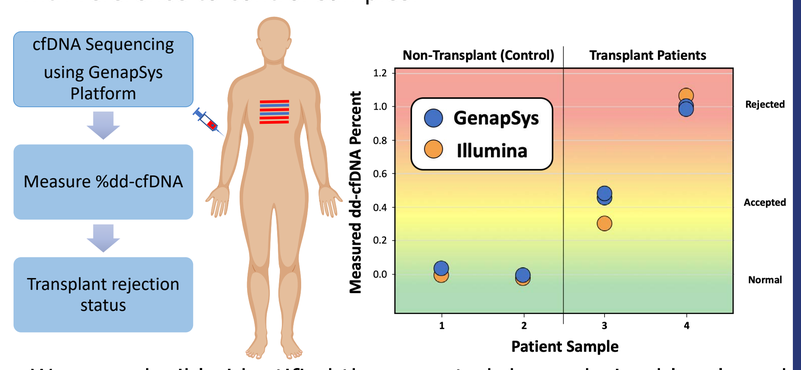
sRNA-seq:

Strong correlation between GenapSys and Illumina sequencing of a sRNA library. A library was generated with the Universal Human miRNA Reference RNA (Agilent) using the QiaSeq miRNA library prep kit. The final PCR step was performed with a GenapSys-customized PCR primer instead of the library PCR primer from the kit.



cfDNA Sequencing

Circulating cell-free DNA (cfDNA) can be used as a biomarker for rejection detection in organ transplant patients. Sequencing of cfDNA libraries from multiple samples was carried out with technical replicates. Amplicon libraries that target >200 single-nucleotide polymorphisms were generated from samples containing donor-derived DNA at a range of 0%, 0.5%, and 1%. Libraries were sequenced on the GenapSys Sequencer, adapter sequences were removed using cutadapt and trimmed reads were aligned to the hg38 reference genome using BWA-MEM. The allele frequencies of heterozygous and homozygous SNPs were calculated using BCFtools mpileup and calibrated with reference to control samples.



We reproducibly identified the expected donor derived levels and successfully recapitulated the standard Illumina data pipeline.

Summary

The performance of the GenapSys Sequencing Technology was evaluated across multiple applications. Notably, its superior variant calling capability was demonstrated by a higher F1 score at a lower coverage compared to the Illumina technology. The high accuracy, low capital and operational cost, and scalability of this technology will ensure its widespread use within the research and clinical genomics communities.

www.genapsys.com