Electrical Detection of Single Base Incorporations Enables High Accuracy Sequencing on a Small, Scalable Platform

Tyson A. Clark, Mohammad Fallahi, Saurabh Paliwal, Meysam Rezaei Barmi, Ali Nabi, Xavier Gomes, Kosar B. Parizi, Maryam Jouzi, Seth Stern, Paul Kenney, Linda Hsie, Pallavi Choudhary, Ramya Akula Suresh Babu, Hamid Rategh, Subra Sankar, and Hesaam Esfandyarpour

Abstract

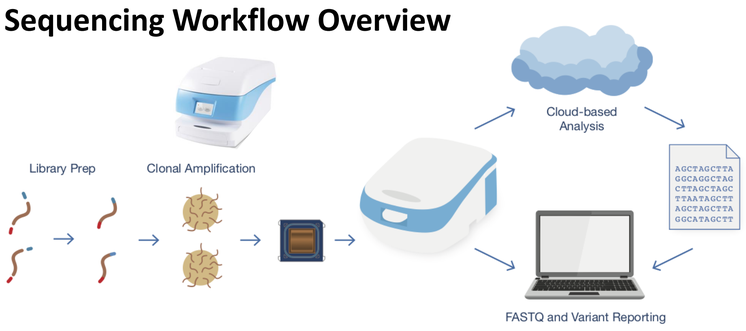
DNA sequencing technologies have undergone tremendous development over the past decade. Though optical-based sequencing is responsible for the majority of data generated, it requires a large capital investment and aggregation of samples to achieve optimal cost per sample. Accurate DNA sequencing on an accessible platform that is small, scalable, modestly priced, and inexpensive to operate, will allow for a more distributed model where the power of genomics is put back into the hands of individual researchers.

GenapSys™ has developed a novel electronic-based platform capable of accurately detecting single base incorporations. The detection method, which utilizes CMOS chips, enables the system to be compact, accessible, and affordable. The platform is capable of generating up to 2 Gb of high-quality nucleic acid sequence in a single run, and we routinely generate sequence data that exceeds 99% raw accuracy (85% of bases >Q30) with average read lengths of 150 bp.

Here, we demonstrate the functionality of the novel impedance-based GenapSys sequencing technology and highlight the utility of this platform for variant detection in human samples. Performance was evaluated on several well characterized cell lines from the Genome in a Bottle (GIAB) Consortium including NA12878. Exome libraries were generated using the IDT xGen probe-based capture method and sequenced on the GenapSys Sequencer to greater than 50-fold average coverage. We evaluated BCFtools and Google DeepVariant for variant calling. DeepVariant is an analysis pipeline that uses a deep neural network to call genetic variants from next-generation DNA sequencing data. The SNVs detected using the GenapSys Sequencing Platform correlated extremely well with those detected from Illumina sequencing data generated from the same samples and with the high confidence calls from the GIAB consortium.

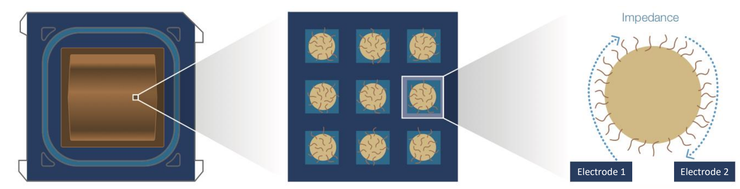
GenapSys Sequencing Workflow

Sequencing Workflow Overview



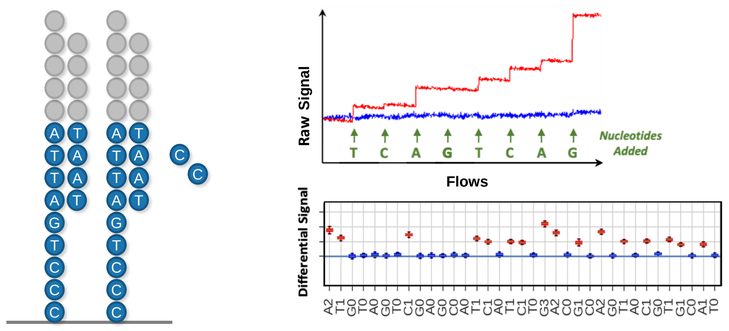
Genomic DNA is randomly sheared and converted into a sequencing library via ligation of adapters using industry standard methods. Individual library molecules are clonally amplified onto beads and loaded into the sequencing chip. Automated sequencing is carried out on the GenapSys Sequencer which utilizes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

GenapSys Sequencing Chip



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.

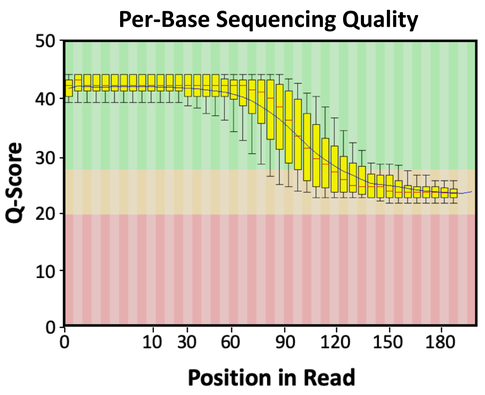
Electronic Sequencing by Synthesis



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, the measured values do not significantly change over time. Thus, if desired, multiple measurements can be taken to improve precision.

Sequencing Performance

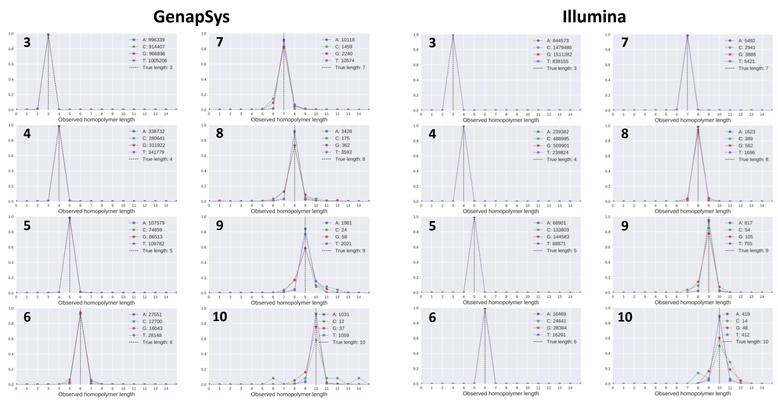
Accuracy



Error Type	Rate
Substitution	0.016%
Deletion	0.089%
Insertion	0.044%

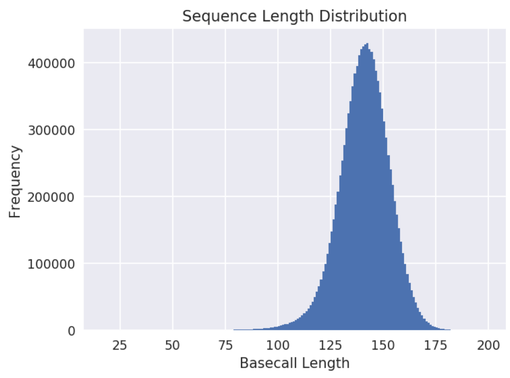
Low substitution error rates are key for accurate variant detection.

Homopolymer Performance



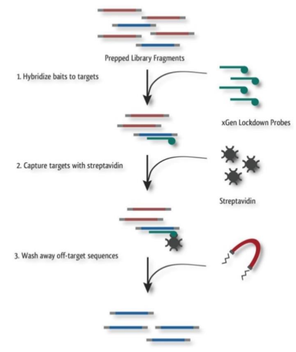
Using data from a human exome sequencing run, base calls from homopolymer regions (3-10 nt) were compared to the hg38 reference for each of the four different bases. Performance is shown relative to the same library sequenced with Illumina.

Read Length



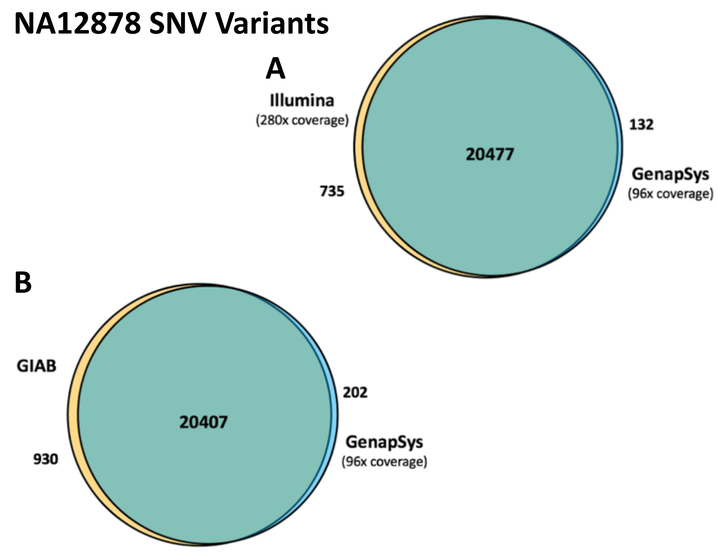
Distribution of filtered read length for a typical sequencing run.

Exome Sequencing



Genomic libraries were enriched for exome regions using the IDT xGEN Exome Research Panel v1.0 which employs a probe-based capture methodology. The panel was made from 429,826 individually synthesized probes and spans 19,396 human genes (39 Mb target region).

Variant Calling



As a demonstration of the technology applied to human samples, we carried out whole exome sequencing of DNA from the well characterized NA12878 cell line that has served as a human reference standard for the Genome-in-a-Bottle (GIAB) Consortium. Variants were called relative to the hg38 reference using BCFtools and DeepVariant. The same exome-enriched libraries were also subjected to Illumina sequence as a comparison. As shown above, the overlap between the two platforms was extremely high (A). If we consider the GIAB high confidence call set as ground truth (B), we maintain high concordance.

Summary

We have demonstrated that the GenapSys Sequencing Platform is capable of generating very accurate DNA sequence data. The system's architecture, including CMOS-based electronic detection, an absence of moving parts and optics, minimal computational requirements, and simple fluidic controls, allows for an instrument that is compact, scalable, and affordable.

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 (85% >Q30)
- Highly accurate sequencing data yields high confidence variant calling
- Excellent concordance for variant calling relative to established sequencing technologies

© 2019 GenapSys, Inc. "GenapSys" and the GenapSys logo are trademarks of GenapSys, Inc. All other trademarks are the property of their respective owners.For research use only. Not for use in diagnostic procedures.