INSTITUTE FOR GENOME SCIENCES EXPANDS LONG-READ SEQUENCING SERVICES WITH NEW SEQUEL™ SYSTEM

At the University of Maryland's Genomics Resource Center, SMRT® Sequencing has become an integral tool for generating complete microbial genomes, improving plant and animal genome assemblies, and exploring human genome variation.

Strength of Experience

The Genomics Resource Center (GRC) at the Institute for Genome Sciences (IGS) has a scientific pedigree and a sample-to-interpretation service commitment that place it in a league of its own. The team operates under a simple mantra: 'If it can be sequenced, we can do it.'

Both the GRC and IGS were founded in 2007 when a high-powered team of investigators formerly at The Institute for Genomic Research (TIGR), led by Claire Fraser, joined the University of Maryland School of Medicine. "The team of faculty and staff that came here to start the institute was heavily focused on infectious disease research," says Luke Tallon, scientific director and founding leader of the GRC. "Our primary goal in joining a medical school was to extend our pathogen genomics expertise into host-pathogen studies and direct clinical genomics applications."

In addition to its infectious disease and genomics expertise, TIGR was also renowned for its bioinformatics talent – a trait that continues with the group at the GRC. Their team of 15 staff members is evenly split between wet lab and bioinformatics, and more than half of the institute's 100-plus employees are bioinformaticians. "One of our strengths is that we go beyond generating efficient, high-quality sequence data. We have teams of bioinformatics analysts and software engineers who can assist investigators with downstream analysis and interpretation," Tallon says. Prior to project initiation, the GRC team consults with each investigator, recommending a custom solution for each scientist's particular goals and needs.

The GRC was formed both to serve the genomics institute and as a university core. "We serve investigators throughout the University of Maryland system as well as across the country and around the world," says Lisa Sadzewicz, administrative director of the facility. The GRC works with hundreds of investigators and has become more visible by presenting and exhibiting at conferences such as the annual meetings of the American Society for Microbiology and the American Society of Human Genetics, and engaging the community through social media, including the GRC blog.

"Our strength is not just our deep history and experience in sequencing and genomics, but our end-to-end service level from the initial project consultation through to publication, including all of the informatics," Sadzewicz says.

Over the past five years, the GRC has applied these strengths to the PacBio® platform. As early adopters of the technology, they have dedicated significant resources to the development of both laboratory and data analysis processes to leverage SMRT Sequencing. Since its adoption of the original PacBio *RS* in 2011, the GRC has steadily increased its utilization of the platform. In the past year alone, they have constructed more than 400 libraries and sequenced more than 1,200 SMRT Cells. These have spanned projects ranging from whole genome sequencing to metagenomics, Iso-Seq™ transcriptomes, and custom amplicons.

Microbial Genomics Expertise

Because of the institute's strong focus on infectious disease, the GRC conducts many sequencing projects for pathogen genomes, human microbiome samples, and other microbial genome applications. The PacBio platform has been a workhorse for generating finished or nearly complete genomes for the past five years. The inclusion of genome-wide methylation data provided in each run is an added bonus. "We're now routinely analyzing methylation patterns for our genomes," Tallon says.

Among the largest of the GRC's current microbial projects is the development of the FDA ARGOS database hosted at NCBI. Initiated in 2013, the database is supported by two contracts totaling over $4M to fund the sequencing, assembly, and analysis of more than 2,000 viral, bacterial, fungal, and parasite genomes. Designed to promote

University of Maryland School of Medicine
Institute for Genome Sciences

Facility name:	Genomics Resource Center
Institution:	Institute for Genome Sciences, University of Maryland School of Medicine
Staff size:	15 scientists, technicians, and bioinformaticians
Year founded:	2007
Investigators served:	More than 500 worldwide and growing
PacBio Systems Installed:	Sequel System (Feburary 2016), PacBio RS II (June 2011)
Website:	http://www.igs.umaryland.edu/grc
Email:	grc-info@som.umaryland.edu

A PacBio Certified Service Provider

a less burdensome regulatory approach for devices that incorporate infectious disease NGS diagnostic technology, the ARGOS database supplies a set of validated regulatory-grade microbial genomic sequences and associated metadata.

"We partnered with the GRC because of their long-standing experience and excellent end-to-end service to generate high-quality microbial genomes," says Heike Sichtig, principal investigator at FDA. "The pipeline was customized to fit our needs to produce and make publically available FDA-vetted microbial regulatory-grade genomes, including raw reads, assemblies, annotations, and methylation data. The PacBio System has been extremely helpful in this project to generate these near-complete regulatory-grade microbial genomes."

Exploring Larger Genomes and Human Variation

As read lengths and throughput have steadily increased, the PacBio System has expanded its application base. The GRC team is forging ahead with larger genomes, transcriptomes, and targeted sequencing. "We've sequenced a number of agriculturally and medically relevant plant and animal genomes, and we are rapidly expanding our sequencing of fungal and parasite genomes," Sadzewicz comments.

> "We partnered with the GRC because of their long-standing experience and excellent end-to end service …" Heike Sichtig, principal investigator at FDA.

Among these are the GRC's contributions to the cichlid fish genome sequencing projects led by Tom Kocher at the University of Maryland. Following up on a short-read-based assembly of *Metriaclima zebra* generated as part of the Assemblathon2 contest, Tom and his team used 16-fold coverage of SMRT Sequencing reads to close 68% of the assembly gaps and add more than 90 Mbp of sequence to the genome. Additional analysis and *de novo* assembly of 50-fold coverage data sets from this and other cichlid genomes is underway. "The high-quality PacBio sequencing performed by the IGS core has been critical to our efforts to improve the reference sequences for cichlid fish, including tilapia and species from Lake Malawi, Africa," notes Dr. Kocher.

They are also using PacBio long reads for haplotype phasing, structural variation detection, and variant validations in human genomes. While sequencing large-scale whole human genomes using PacBio remains expensive in the near term, it is an ideal platform for these targeted applications. "We have developed custom amplicon approaches to detect indels and other structural variations," says Tallon. "We are also using a barcoded, pooled amplicon approach on the PacBio RS II to validate SNVs from our cancer exome projects."

Luke Tallon and Lisa Sadzewicz are expanding their Iso-Seq transcriptome, amplicon, and whole genome sequencing services with the new Sequel System.

Early Adoption of the Sequel™ System

The GRC's newest PacBio sequencer arrived in February 2016, making them one of the first PacBio Certified Service Providers to take delivery of a Sequel System. "Given our history of early adoption and success with the PacBio *RS*, and the promise of increased and scalable throughput, we were excited to be among the first centers to acquire a Sequel instrument," says Dr. Fraser. The Sequel System represents the newest generation of SMRT Sequencing, providing 5-10 Gb of throughput per SMRT Cell, more scalability and lower sequencing project costs compared to the PacBio RS II.

Development of processes and applications for the Sequel System is well underway at the GRC. The team plans to use the increased throughput to expand their services for Iso-Seq transcriptome sequencing and amplicon projects. In collaboration with Dr. Jacques Ravel, associate director of IGS, they are also developing a full-length 16S sequencing pipeline to complement and expand their human microbiome and metagenomic research portfolios. As read lengths and throughput on the Sequel instrument improve, Tallon's team will shift whole genome sequencing projects onto the platform.

For other teams considering whether SMRT Sequencing is the right choice for them, Tallon says: "If you value complete genome sequences, *de novo* transcript discovery, and are looking at epigenetics in addition to the genome sequence, there's no better technology out there."

PN: PF105-052716