Using Next-Generation Sequencing to Address Genomic Challenges from Microbes to Cancer

Dr. David Buck, Ph.D. searches out cutting-edge technologies, such as the MiSeq® personal sequencer, to ensure that his customers get accurate answers quickly.

David Buck, Ph.D. is Head of High-Throughput Genomics at The Wellcome Trust Centre for Human Genetics at the University of Oxford. The Centre is composed of scientific research groups that work to understand the genetic foundations of human variation and disease. Dr. Buck runs the High-Throughput Genomics Core, whose mission is to provide researchers with rapid access to cutting-edge technologies through outstanding service. They strive to be the world's leader in generating and handling next-generation sequencing data.

Q: How do you help your customers identify the appropriate technology?

David Buck (DB): We're pivotal in guiding our users in experimental design. A lot of researchers who come to the Centre are new to next-generation sequencing technologies. It's only recently that the cost has come down to the point where people can really think seriously about using it for a lot of experiments. We provide advice on the appropriate platform, the amount of DNA to use, the number of replicates, and the depth of coverage.

Q: Why have you partnered with Illumina?

DB: Since we provide access to cutting-edge technology, we have to evaluate all of the platforms. We partnered with Illumina because we had a strong belief that Illumina technology was the best on the market at the time, and it continues to be the best on the market. It's really difficult to support multiple platforms because they each have different pipelines for handling sample input and data analysis. In our world, where we're supporting medical research, speed and lowest cost point are critical. The HiSeq® system is the leader in that field. The ability to use the same libraries on MiSeq and HiSeq is obviously an advantage for us because the preparation is the same. We would need a completely parallel system of library preparation if we used other technologies.

> "I think we can start referring to the data that's generated on MiSeq or HiSeq as the new gold standard."

Q: What projects have you run on the MiSeq system?

DB: We're part of a major project here called the Modernizing Medical Microbiology Consortium, where we're looking at using sequence data to trace infection points within hospital environments to help understand how to manage, explain, and prevent an outbreak. We're sequencing and multiplexing tens of thousands of bacterial genomes. Unexpectedly, right in the middle of participating in the MiSeq early access program, there was a suspected outbreak in a local hospital. We had an urgent need to obtain genomic data on the bacteria very quickly so we fired up three runs of four-plex sequencing on MiSeq, generating data in less than five days. This enabled the hospital to determine if there was a common source of bacteria. It was pretty exciting and showed the power of this machine. You can get to results very, very quickly.

David Buck, Ph.D. runs the High-Throughput Genomics Core at The Wellcome Trust Centre for Human Genetics at the University of Oxford.

Q: What applications benefit from the MiSeq system's reduced time to results?

DB: If you're in a diagnostic-type mode where you need data quickly, or you need an answer very quickly as we did to track the recent hospital outbreak, then I think MiSeq is a wonderful instrument to use. Routinely, we sequence 96 multiplexed bacterial genomes on a single HiSeq lane. That's fine. It's doable. On MiSeq, we do four genomes in a lane, but we can sequence them a lot quicker. MiSeq is allowing us to get faster access to the bacterial genome sequence.

Q: What projects are you transitioning to the MiSeq system?

DB: We will move projects with thousands of samples that involve resequencing individual targets, exons, or genes to the platform. We have one project with 3,000 samples. To do it using classical Sanger sequencing would cost about £42,000. The sequencing cost alone on MiSeq is only going to be £2,000, if you heavily load it. When we start using TSCA [TruSeq® Custom Amplicon], it will be a lot cheaper.

Q: What types of MiSeq libraries are you currently preparing?

DB: We have used the MiSeq system to evaluate the Illumina TSCA system, the targeted PCR-based multiplex amplification system that generates sequencing-ready libraries. We have also prepared and run a number of whole bacterial genome libraries but, depending upon the amount of data required, any type of library can be run on MiSeq. We are currently evaluating the Nextera® system as a fast-track library preparation alternative to couple with the rapid turnaround time of MiSeq.

Q: Has the MiSeq data met your expectations?

DB: My expectations were that MiSeq would only generate 1 Gb of data on a paired-end 150-bp run. We found that MiSeq generates way more data than that. On average, we are getting 1.5–2 Gb of data. In terms of the amount of data, MiSeq has far exceeded my initial expectation. In terms of the quality of the data, it is what I expected, being comparable to HiSeq, up to 100 base pairs. We haven't run HiSeq beyond 100 bp. I think MiSeq data quality out at 150 bp is very good. Base quality is not an issue. The analytical power in it is better than I expected.

Q: Will you still offer capillary sequencing in your lab?

DB: Some customers who run a screen on MiSeq and find potential SNPs in an exon, may want go back and validate using capillary sequencing. It's not a factor of the MiSeq system. I think it's a hangover to the old days when Sanger sequencing was regarded as the gold standard. I think we can start referring to the data that's generated on MiSeq or HiSeq as the new gold standard. For instance, if you've got a somatic change in an amplicon, you're only going to see it at a very low level. I don't think classical Sanger sequencing is sensitive enough to detect it, but next-generation sequencing could.

Q: What do you see as the main benefits of the MiSeq system?

DB: We were pleased that it worked the first time right out of the box. One of the things I really love about MiSeq is the fact that you put a library on and it does the initial clustering, the sequencing, and the reclustering. You put your sample onto the machine, press 'GO', and generate data. We like the ability to work on new protocols, and the ability to do the rapid follow-up sequencing of pools of samples tied to a smaller number of specific regions. There is significant interest in having access to this platform, which is very rewarding. It's just pure speed and convenience.

"One of the things I really love about the MiSeq workflow is the fact that you put a library on and it does the initial clustering, the sequencing, and the reclustering...It's just pure speed and convenience."

Q: What future projects do you expect to run on the MiSeq system?

DB: A lot of bacterial and viral whole-genome studies, targeted follow-up sequencing, and ChIP-based studies where the data output of MiSeq is ideal. The platform is designed for projects where speed is critical. It will be a great tool for process development, or when you want to look at the quality of a library or balance of libraries in a multiplex prior to committing a full lane on a HiSeq.

Q: With the MiSeq system, are you able to take on projects that you couldn't do before?

DB: MiSeq uses the same sequencing by synthesis chemistry employed by HiSeq. As such, any project that can be done on HiSeq can also be done on MiSeq. Where the system really gains, is the speed with which data can be generated and the longer read lengths that allow full sequencing of the 250-bp amplicons generated in the TSCA system for targeted re-sequencing.

Currently, we are involved in a key collaborative project with Illumina to sequence 500 whole genomes, focusing on a mixture of life-threatening diseases. We have already identified, and will continue to identify, a number of associated SNPs. The follow up and validation of these could be done on MiSeq.

Learn more about the MiSeq system at www.illumina.com/miseq

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 770-2012-042 Current as of 15 February 2012

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA