Faster Microbial Genomics Studies with Small-Scale, High-Speed, Next-Generation Sequencing

Using the MiSeq® personal sequencer, Professor Byung-Kwan Cho can quickly develop data-driven hypotheses, enabling even more discoveries.

Byung-Kwan Cho, Ph.D. is an assistant professor in the Department of Biological Sciences at the Korea Advanced Institute of Science and Technology (KAIST). He runs the Laboratory of Systems and Synthetic Biology where he is working on decoding microbial genomes and transcriptomes. By understanding the genotype–phenotype relationship of microbes, he hopes to ultimately be able to manipulate production of biochemicals, biofuels, and other valuable materials. Professor Cho uses high-throughput data generation platforms, such as next-generation sequencing, to help elucidate nodes and edges in transcriptional regulatory networks.

Q: Why were you interested in testing the MiSeq system?

Byung-Kwan Cho (BK): We have been using Illumina sequencing for some time now, including the work done for our 2009 *Nature Biotechnology* paper where we found a tremendous number of new genomic features that have changed the view of the bacterial genome[1]. We've always liked Illumina sequencing technology because it is very powerful and gives us the freedom to obtain unusual transcriptomics and genomics information that we were unable to get using classical methods. Offering a single sample format, MiSeq gives us speed. It's one of the system's best features. That's why I became interested in it.

Q: How does Illumina sequencing improve upon classical methods?

BK: Microarrays were developed a decade ago, but I already consider them a classical method. When I used microarrays to find novel genomic features, such as entire transcripts, transcription start sites, or promoter regions, the problem was resolution. I wanted single base pair resolution, but microarrays don't provide that level of detail. They also require a lot of bioinformatics tools for data processing, giving us bias or errors. Those are the biggest limitations of microarrays. The Illumina sequencing platform provides that single base pair resolution and a tremendous amount of data. Those two factors are really important. They make a huge difference in our research.

> "Offering a single sample format, MiSeq gives us speed. It's one of the systems's best features."

Professor Byung-Kwan Cho runs the Laboratory of Systems and Synthetic Biology at KAIST. He is using next-generation sequencing to decode the microbial genome.

Q: How will the MiSeq system's single sample format impact the time lines of your experiments?

BK: With a high-throughput platform, I sometimes need to wait for additional samples to be ready to go before I can run my samples. The timing can be unpredictable. With the single sample format of the MiSeq, I don't have to wait. This will allow me to obtain the data quicker, saving a lot of time. Now that we can multiplex bacterial samples on MiSeq, we'll save even more time.

Q: How did you test the MiSeq system?

BK: We sequenced a bacterial genome that was genetically modified. We made a list of regions we thought were not necessary for this bacterial cell, and then removed them—we cut out about 1 Mb from the bacterial strain. Then we tested for physiological effects. Our assumption was that any phenotypic changes would be a result of removing this region. However, after sequencing, we were surprised when we actually saw mutations in other regions that were causing some of the physical changes we observed. This is the first time we were able to accurately find these SNP mutations.

Q: How will these results impact your next steps?

BK: Our data analysis and the future direction of our research will change based on these results. In the MiSeq test, we reduced the genome size using a genetic tool. Based on previous work, we didn't expect to see any mutations in the remaining genomic regions, but the sequencing data proved otherwise. These mutations actually caused a lot of phenotypic changes. With the sequencing data, we can make more discoveries. We'll be able to go in many different directions.

What's most important is that we can make a hypothesis from this data set. This is data-driven hypothesis generation. It's a funny statement, but it's really a nice thing. These SNPs can cause some phenotypic changes and we need to test them. This is a very important factor of this sequencing platform. Because MiSeq has speed, it can give us lots of data to produce hypotheses that we will test using the other methods.

Q: Can you tell us about your experience preparing libraries for the MiSeq system?

BK: Library preparation for the MiSeq is not that difficult. Usually we make libraries based on our own protocols. We normally never use kits because we sequence very specialized libraries, for example, transcription start sites or terminator sites. We design the primers and the protocols. But this time we had a resequencing sample and used the TruSeq® kit that Illumina provided. We followed Illumina's protocol and we realized, "Wow, the TruSeq kit is really simple to use." Just add the reagent, wait 10 minutes, add more reagent, run the gel, and select the size. That's it. We made that library within a day. The good thing is that the TruSeq kit provides the index. Most importantly, it decreased the cost of library preparation.

Also, we don't need experts to run MiSeq. It will be really nice for other people to use as well. I think it is very easy to operate, very simple, very fast, and overall very satisfactory.

> "...we don't need experts to run MiSeq."

> We followed Illumina's protocol and we realized, 'Wow, the TruSeq kit is really simple to use.'

Q: Does the MiSeq data meet your expectations?

BK: The coverage and the amount of data are better than I expected. Illumina indicates that MiSeq will give us 3 Gb level of information. This is what we obtained. I was impressed by the level of the data. The data quality was also very nice with satisfactory coverage and Q scores.

Q: How will the MiSeq system impact projects in your lab?

BK: MiSeq has only one channel, giving us more freedom. Whenever I want to sequence one sample, I can just start sequencing. The Genome Analyzer™$_{IIx}$ and HiSeq® 2000 systems provide large amounts of sequencing information, but they have multiple channels. I need to wait until all eight channels are fully packed before running. With MiSeq I can run today or tomorrow. That is a very nice feature.

Q: What makes the MiSeq system amenable to bacterial sequencing?

BK: The output level of MiSeq is perfect for microbial applications. Many people in the microbial field are using next-generation sequencing to study very novel genomic features such as entire transcripts, transcription start sites, or promoter regions. Based on their genome size, the Genome Analyzer$_{IIx}$ and HiSeq 2000 systems' output is too much for microbial applications. Those researchers will want MiSeq.

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

Reference

1. Cho BK, Zengler Q, Qiu Y, Park YS, Knight EM, et al. (2009) The transcription unit architecture of the *Escherichia coli* genome. Nat Biotechnol. 27: 1043–9.

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

FOR RESEARCH USE ONLY

Pub. No. 770-2012-004 Current as of 19 November 2014