

Tag Team: Scientists Use Pippin to Optimize Nextera Libraries

In a prominent core facility at the Dana-Farber Cancer Institute, scientists have paired Pippin Prep size selection with Illumina's Nextera to build better, more reproducible libraries that yield improved genome assemblies.



Zach Herbert

Zach Herbert is a technology guy. As the associate director of the Molecular Biology Core Facilities at the Dana-Farber Cancer Institute, that's especially important — after all, his job includes considerable tech research and development to track down the equipment that will be most useful to lab clients.

"We have a few technology platforms in our lab, but in order to do my job well, I need to know about all the ones that could have been in our lab," he says. For the tools that pass Herbert's scrutiny, there will be more challenges once they are in the core lab. "I'm not afraid to take out some screws and look at what's inside these instruments," he adds. "We have to make sure that we're pushing these technologies as much as we can."

It was in that capacity of continually pushing technology forward that Herbert first started using Illumina's Nextera to build genomic libraries — and when he realized that those libraries could get a significant boost from size selection. The Nextera prep results in "a wide array of fragment sizes," Herbert says — a range that is certainly possible to sequence, but is not optimal for reproducibility, flow cell clustering, and downstream analysis. "It's really beneficial to have a narrower size range than the Nextera kit generates on its own," he says.

Herbert surveyed the size selection options and chose Pippin Prep, an automated platform from Sage Science. "It's been a good fit," he says. "Today, almost all the Nextera libraries we build are run on the Pippin."

"It's really beneficial to have a narrower size range than the Nextera kit generates on its own."

In the Lab

Herbert joined the Molecular Biology Core Facilities (MBCF) eight years ago as a research technician. The labs, led by director Paul Morrison, are highly regarded well beyond Dana-Farber — their sterling reputation in the scientific community keeps them on many vendors' short lists for new technology early access programs.

The labs are split between DNA and protein applications. Herbert oversees projects related to DNA sequencing, from sample prep through to the first-pass bioinformatics analysis. His clients hail from the Dana-Farber community, Harvard Medical School, and many other organizations, including some companies.

The MBCF sequencing lab has two Illumina MiSeqs — a third will be installed soon — and has kept two 3730s for ongoing Sanger work. But when it comes to next-gen sequencing, Herbert cut his teeth on an instrument that most people never even saw in action: the single-molecule HeliScope from Helicos BioSciences. "In 2009, we were a test site for Helicos. That was my first foray into next-gen sequencing," he says. The lab decommissioned the instrument two years later, but during that time Herbert racked up a record that virtually no one else in the world can match, running some 2,500 samples on the HeliScope — most of them for ChIP-seq experiments.

Herbert continues to do plenty of ChIP-seq work with the MiSeqs, which have come to serve as the preferred means for piloting large ChIP projects before they are fully launched. "A lot of what we do is assay development, where our clients are trying out a new protocol or a new antibody," he says. Clients choose the MiSeq for its rapid turnaround time — results are usually returned within a week, sometimes even within a day or two — compared to what might be a weeks-long wait to get on a HiSeq. The MBCF MiSeq service can provide quick feedback on whether a library is good before the client goes to the effort of pooling

and shipping a big project to a HiSeq, which could take five or six weeks to return results. “We see a lot of people who want that quick answer to make sure that it’s worth the wait for the HiSeq,” Herbert says.

Part of the beauty of the core lab setting is the breadth of scientific projects requested, and MBCF sees a lot more than ChIP-seq. Herbert handles small genome resequencing, BAC sequencing, and amplicon targeted sequencing. Another popular application for clients outside Dana-Farber is metagenomics, for which Herbert’s team relies on 16S ribosomal RNA sequencing. “We’ve sequenced 16S on dirt, air, water, air filters, and more,” he says. In more individualized projects, he sees uses as diverse as translocation mapping and polymerase error rate profiling.

Nextera and Pippin

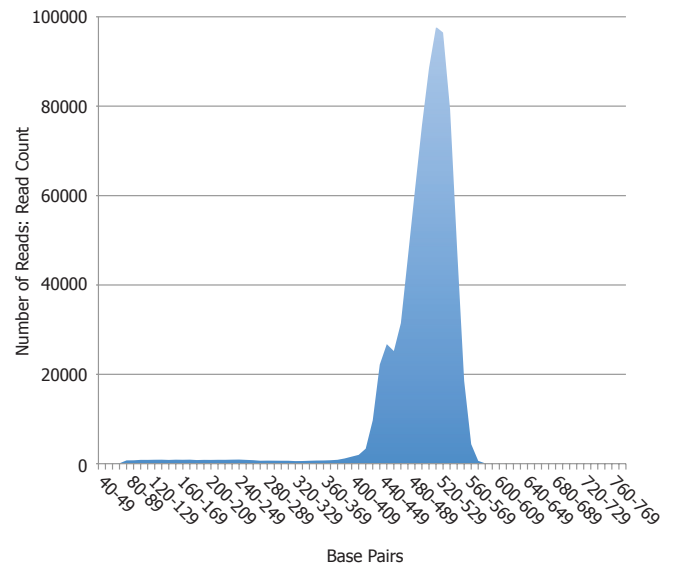
For much of the small genome work, as well as some of the larger amplicon projects, Herbert now generates genomic libraries with the speedy Nextera, which fragments DNA and adds adaptors in a single step known as “tagmentation.” The resulting libraries have a fairly wide insert size range — according to Illumina documentation, users can expect fragments to range from 300 base pairs to 1,000 base pairs — which Herbert found less efficient for sequencing.

His team has tested other size selection options, finding that manual gels are too time-consuming and laborious, while bead-based solutions show operator-to-operator variability. He selected the automated Pippin Prep and now uses that for virtually all Nextera-

Having average insert sizes around 450 or 475 base pairs, he notes, increases the yield on the assembly, resulting in fewer contigs and a larger N50.

built libraries to yield a far narrower fragment range. “You can go from the Nextera amplification directly onto the Pippin,” Herbert says. That combination of the two technologies has been valuable for de novo assemblies, reproducibility, and optimal clustering on the Illumina sequencer flow cell, he adds.

In the case of de novo assembly, a common application for clients whose work goes on the MiSeq, “it’s really challenging to do that if you have a wide range of insert sizes,” Herbert says. For example,



Insert sizes from alignment of a Pippin size-selected Nextera library.

having a lot of 300-base pair inserts while using the paired-end 250-base pair kit from Illumina means that “you end up sequencing a lot of fragments twice,” he says. He uses a 1.5% agarose cassette with Pippin to capture the larger fragments. Having average insert sizes around 450 or 475 base pairs, he notes, increases the yield on the assembly, resulting in fewer contigs and a larger N50. When you’ve got clients to keep happy, that’s a great asset. “Their assembly wouldn’t have been as good if we hadn’t used size selection,” Herbert says.

The Pippin/Nextera tag team also shows value beyond de novo assemblies. “Having a narrow and known size distribution makes calculating the molarity a lot easier so you can get a better cluster density and maximize the number of reads,” Herbert says. It’s also a boon for pooling samples. Attempting to pool samples with a broad size range in equimolar amounts is very tricky — “but if all those libraries are the same size, then we’re much more likely to get an even distribution of that pool.”

“For low- to medium-throughput size selection where you want to minimize sample loss and maximize reproducibility, I think the Pippin is a great instrument,” Herbert says.

Suite 3150
500 Cummings Center
Beverly, MA 01915
978.922.1832
www.sagescience.com