

Perspective

Have Genome, Need Genotyping Tool

Dr. Jerry Taylor, cattle researcher, blazes a path for creating genotyping tools for non-human, non-model organisms.

Last year Americans consumed more than six million gallons of milk, courtesy of the dairy industry. But what did it take to make this glass of milk? It all begins with the right cow.

The cattle industry, including dairy and beef production, is a multi-hundred billion dollar a year global enterprise. Yet, almost all of the genes responsible for the traits valued by cattle breeders and consumers remain a mystery. Currently, breeders are using the “breed-wait-see” approach to select for cattle with desirable characteristics. This method is enormously expensive and time-consuming. Dr. Jerry Taylor, cattle researcher for 24 years, knew there had to be a better way.

Whole-genome genotyping technology, available for humans and model organisms, has rapidly advanced the understanding of the genetic basis of variation. Dr. Taylor imagined what whole-genome analysis could do for the cattle industry. If you could select for traits prior to breeding, the genetic merit of a cow or bull could be determined in days instead of generations. This would save time and money, and ultimately improve bovine health and the quality of beef and dairy products. So Dr. Taylor and the iBMAC Consortium¹ set out to build the first high-quality whole-genome genotyping array for bovine.

The development of this tool is more than a cow and bull story. Researchers working with non-human or non-model organisms can follow the same simple

and elegant strategy used by Dr. Taylor and colleagues to create genotyping tools of their own.

PARTNERING WITH ILLUMINA® FOR GENOTYPING TECHNOLOGY

Dr. Taylor and members of the iBMAC Consortium knew that they wanted to create a whole-genome genotyping array that would be useful across all economically important cattle breeds. After due diligence, they turned to Illumina’s iSelect® program to help them achieve their goal. The iSelect program is specifically designed to support collaborations between Illumina and external research groups to develop community-specific genotyping technology. “We partnered with Illumina because of the simplicity of the Infinium® Assay chemistry, the demonstrated data quality of their genotyping technology in the literature, and the company’s demonstrated commitment to being a real part of the community. We knew they were going to support this product and make it broadly available, at a reasonable cost, to people who would want to use it.”

A SIMPLE PLAN EXPANDED

The Consortium’s initial strategy was straightforward: design an Infinium DNA Analysis BeadChip with probes targeting 50,000 single-nucleotide polymorphisms (SNPs) uniformly distributed throughout the bovine genome. For content, they combed through the millions of publicly available SNPs identified during the assembly of the



Jerry Taylor, Ph.D., Wurdack Endowed Chair in Animal Genomics, Division of Animal Science, University of Missouri

“We partnered with Illumina because of the...demonstrated data quality of their genotyping technology... and the company’s demonstrated commitment to being a real part of the community.”

bovine reference genome and from shotgun sequences of six breeds of cattle^{2,3}. The Consortium had high standards for content and problems quickly surfaced. Like many reference genomes derived from a single inbred individual, the bovine genome lacked identifiable SNPs in large tracts of sequence (because the sequenced animal was homozygous by descent). For example, half of chromosome 5 was essentially SNP-free. Additionally, after genotyping a random subset of the SNPs, they discovered that many were not real polymorphisms, but artifacts likely introduced by sequencing errors. “We realized we were going to have to add on a SNP discovery experiment to this project. We looked at each other in shock and horror because it was something we had not anticipated we would need to do, and we certainly hadn’t budgeted on doing this,” recalls Dr. Taylor.

STAYING WITH ILLUMINA FOR SNP DISCOVERY

Now that SNP discovery was in the picture, Dr. Taylor and colleagues needed a way to generate the millions of bases of required data. Sanger sequencing would be too costly. Once again the Consortium turned to Illumina, this time for its massively parallel sequencing technology. The high-throughput Genome Analyzer offered the group an affordable solution.

The team predicted that they could discover enough SNPs to populate their genotyping array by selectively sequencing a subset of the genome from multiple individuals across many cattle breeds. After performing more than 60 *in silico* restriction digests of the bovine genome, they picked an enzyme that produced an ideal number of fragments representing a random subset of the genome that also lacked repetitive elements. “Shooting from the hip, we pooled together DNA from three populations of economically important cattle breeds: Holstein, Angus, and an additional group of six popular beef breeds. And with

our fingers crossed, we sent our digested and size-selected libraries off to Illumina for sequencing,” Dr. Taylor recalls. He and his colleagues were amazed with the data they received: more than 50 million high-quality short-read sequences providing 20X coverage of 0.5% of the bovine genome.

After crunching the data, Taylor reports, “the results exceeded our expectations. We thought we would discover 50,000 SNPs through this sequencing project. In the end, we found 62,000 SNPs that we could map uniquely to the bovine genome, and 92% of them turned out to be real SNPs⁴.” They also gained flanking sequence information, enabling easy design of probes for their iSelect Infinium BeadChip. Taylor adds, “With 20X coverage of the fragments—which randomly represented individuals from our pooled libraries—we could calculate the frequency of each allele and determine the utility of these SNPs across multiple cattle breeds. In this one experiment we were able to find novel SNPs, uniquely map their location, and simultaneously determine their allele frequencies. The results were mind-bogglingly sensational.”

BRINGING ISELECT TO THE BOVINE COMMUNITY

With their pool of newly discovered common SNPs in hand, Dr. Taylor and colleagues worked with Illumina scientists to design their genotyping array. The final product, the BovineSNP50 BeadChip, contains 51,386 SNPs validated in more than 20 economically important cattle breeds and provides uniformly spaced coverage of the bovine genome. Commercially released in late 2007, this BeadChip offers the cattle researcher community the first high-quality genotyping tool. “For the last 15 years, cattle researchers have mapped thousands of quantitative trait loci (QTLs), regions of the genome linked to

“In this one experiment we were able to find novel SNPs, uniquely map their location, and simultaneously determine their allele frequencies. The results were mind-bogglingly sensational.”



traits of interest, down to 10–50 megabases of sequence. At such low resolution, these QTLs don't tell us much about the genes underlying desired characteristics. Now, with the BovineSNP50 BeadChip, the community is moving at warp factor 10 to map these genes at significantly higher resolution," he says. By introducing this BeadChip, the iBMAC Consortium's efforts have brought the bovine community one giant step closer to identifying the genes behind the ideal pedigree.

SAME STRATEGY, DIFFERENT ANIMAL

Discovering common SNPs on Illumina's Genome Analyzer and building genotyping arrays through the iSelect program has proven so successful that other research groups are following suit. Dr. Taylor is co-principal investigator on the development of an iSelect BeadChip for swine and says, "I think you are going to see that the approach we used to develop genotyping technology for cattle will be embraced by the rest of the agricultural community. In the next 12 months the technology will be there for all agriculturally important plants and animals. Besides our swine project, I know that an Australian and New Zealand group are building one for sheep, and another group at Missouri is working on a soybean array."

A GENOME ANALYZER OF HIS OWN

Dr. Taylor was so impressed with the Genome Analyzer data that he recently acquired one for the DNA Core Lab at the University of Missouri. "The fact that Illumina has a head start in the market means they also have a head start in terms of data analysis and the availability of third-party data analysis platforms," cites Dr. Taylor as one reason for a Genome Analyzer purchase. "We also went with this system because we have a very large group here on campus that is working in human cancer epigenetics and they really need the ability to study methylation status. The Genome Analyzer is the only system with demonstrated utility for sequencing bisulfite-treated DNA, enabling methylation analysis." Now Dr. Taylor is using the Genome Analyzer to identify by "brute force" the causal mutations underlying QTLs in cattle. There are thousands to get through, but with his research and his legacy of the BovineSNP50 BeadChip, soon we may know what it takes to make a perfect glass of milk.

"I think you are going to see that the approach we used to develop genotyping technology for cattle will be embraced by the rest of the agricultural community."

REFERENCE

- (1) iBMAC Consortium includes researchers from Illumina; the United States Department of Agriculture, Agricultural Research Service (USDA ARS), Beltsville; University of Missouri; University of Alberta; and USDA ARS, Clay Center.
- (2) <ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/fasta>
- (3) ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/snp/Btau20070913_README
- (4) Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, et al. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods* 5(3):247-52.

We are committed to providing you with the content you want as a member of the Illumina community. Please email us with comments and suggestions for topics at icommunity@illumina.com.

FOR RESEARCH USE ONLY

© 2008 Illumina, Inc. All rights reserved.

Illumina, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, iScan, and GenomeStudio are registered trademarks or trademarks of Illumina. All other brands and names contained herein are the property of their respective owners.

Pub. No. 770-2008-008 Current as of 1 May 2008

ADDITIONAL INFORMATION

To learn more about the Genome Analyzer system and Illumina's iSelect program, visit www.illumina.com

