Switching from Arrays to Sequence-Based-Genotyping in Cattle Breeding Operation

Shawn Quinn (Curio Genomics), Paul Doran (Twist Bioscience)

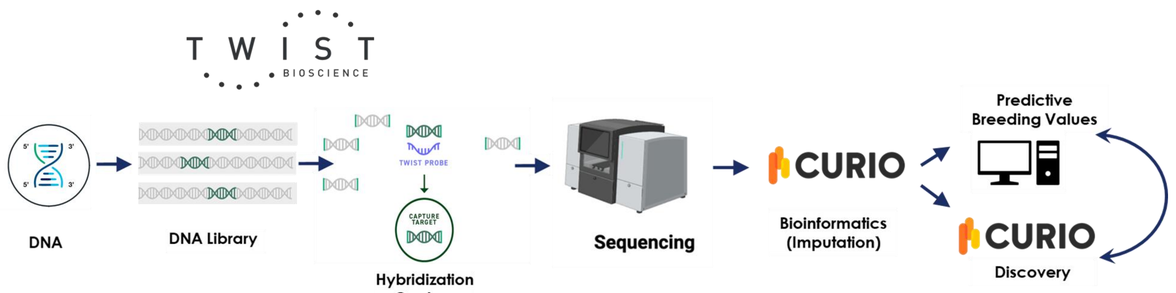
Abstract

As sequencing costs continue to drop dramatically, sequencing-based approaches are in the process of replacing costly genotyping arrays. This shift to sequencing is driven by better data, lower cost, and the capability to run a single platform for both discovery and genotyping. This simplification of lab capabilities has profound implications for breeding program efficiency, genetic gain, and the sustainability and precision of agricultural practices.

On the other hand, breeders have trusted array data for a long time, and its use is fully integrated into their breeding and discovery operations. Therefore, any sequencing-based approach to genotyping must provide breeders with the same level of confidence in the results while also allowing them to continue to leverage their wealth of legacy array data.

Here we present an example of a 28,000-head cow-calf operation that has made the switch working in collaboration with Curio Genomics and Twist Bioscience. First, we mapped the existing array data from the 70k bovine array into the CURIO™ agrigenomics data analysis platform. Next, we leveraged Twist's new HTP self-normalizing FlexPrep™ library prep and multiplex capture to target genomic regions previously captured by arrays. Finally, using CURIO, we evaluated the concordance of results from arrays and targeted sequencing in the same animals. The results were 99%+ concordant with each other providing the confidence to complete the switch from arrays to sequencing, both increasing efficiency and accelerating best practice development.

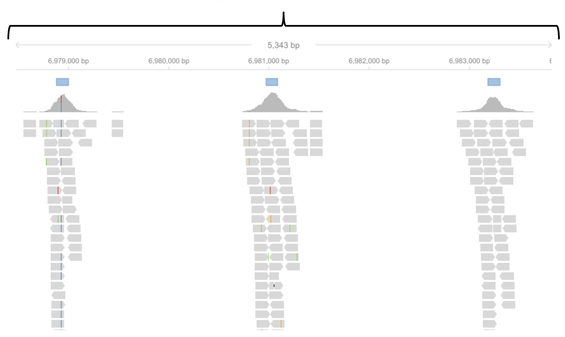
NGS-Based Genotyping Workflow



Probed Regions from FlexPrep + 70K Bovine Panel

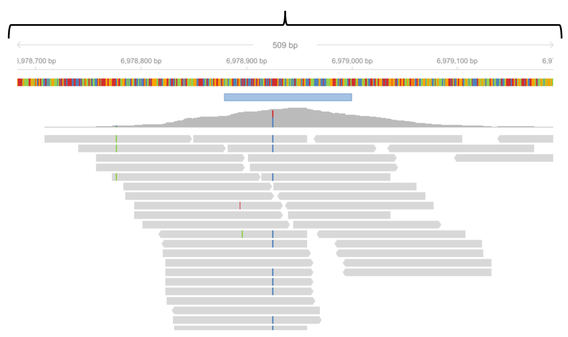
Targeted regions visually reviewed within the CURIO platform to see how the bovine panel performed.

5k Region, 3 Twist Probes



chr4:6,978,293-6,983,894

Reads Captured by 120 bp Probe

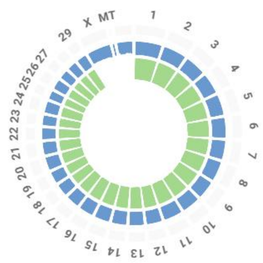


chr4:6,978,685-6,979,192

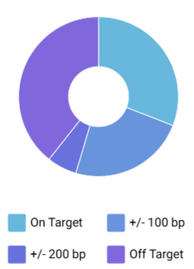
Capture Efficiency for FlexPrep + 70K Bovine Panel (w/o Blockers)

The ARS UCD 1.2 bovine reference assembly was used as the basis for the analysis, and a coverage analysis of the aligned data was performed within the platform to measure how well the reads covered the 70k targeted regions. The results of a representative sample are shown here. Note that these libraries were prepared without using any bovine blockers.

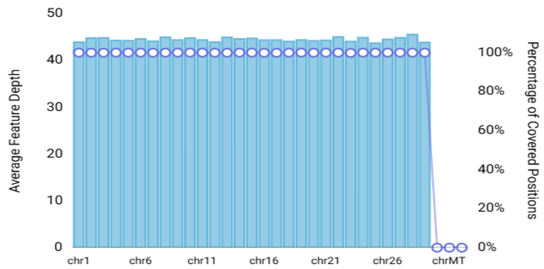
Consistent Coverage Across Genome



55% of Reads Overlap Targeted Regions

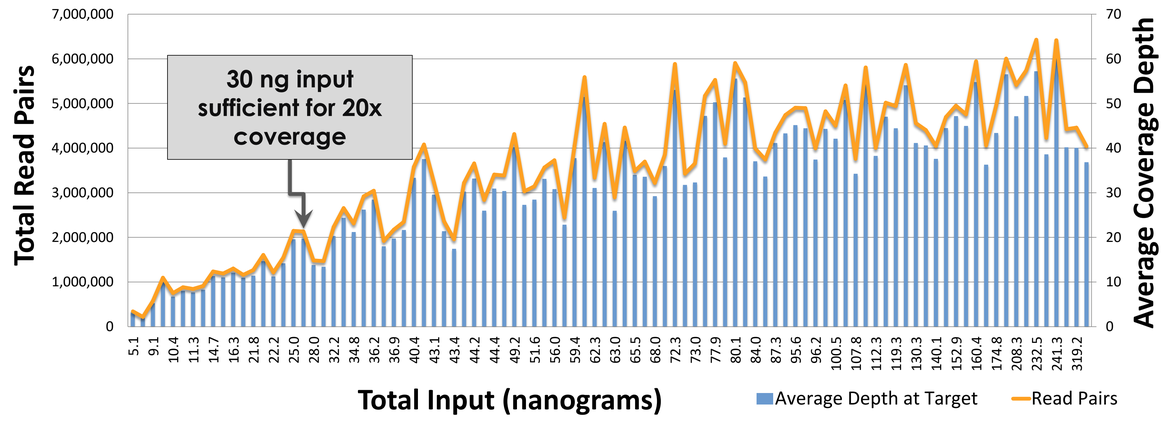


100% of Targeted Regions Have Good Coverage



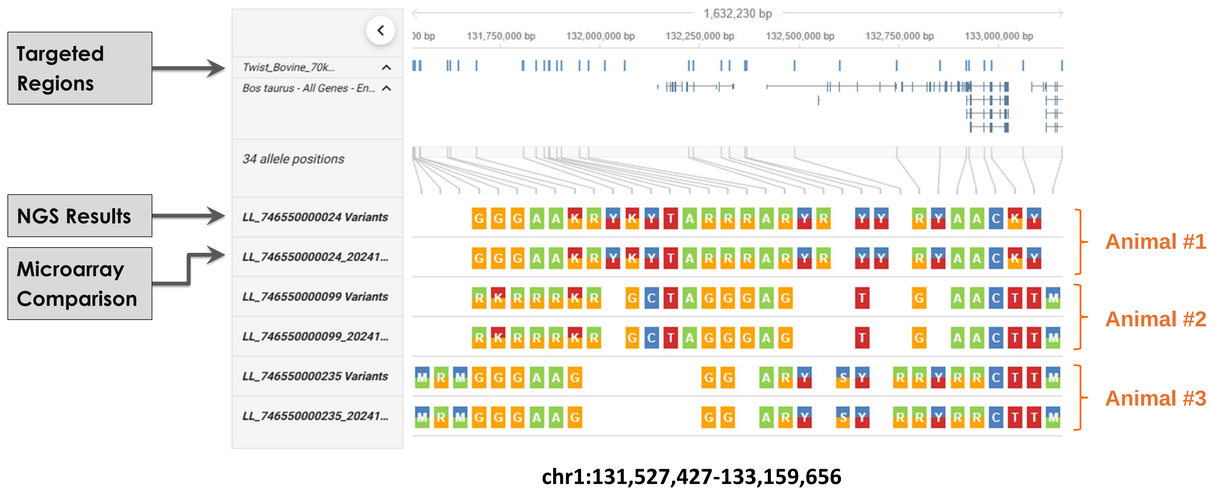
DNA Input – VS – Depth at Target for 70K Bovine Panel

96 bovine samples were prepared using different amounts of DNA input, and then the average coverage depth of the targeted regions (i.e. of the 70K bovine panel) was measured. A minimum of 30 nanograms consistently provided for 20x+ coverage at the targeted regions.



NGS Capture Panel vs 70K Bovine Array

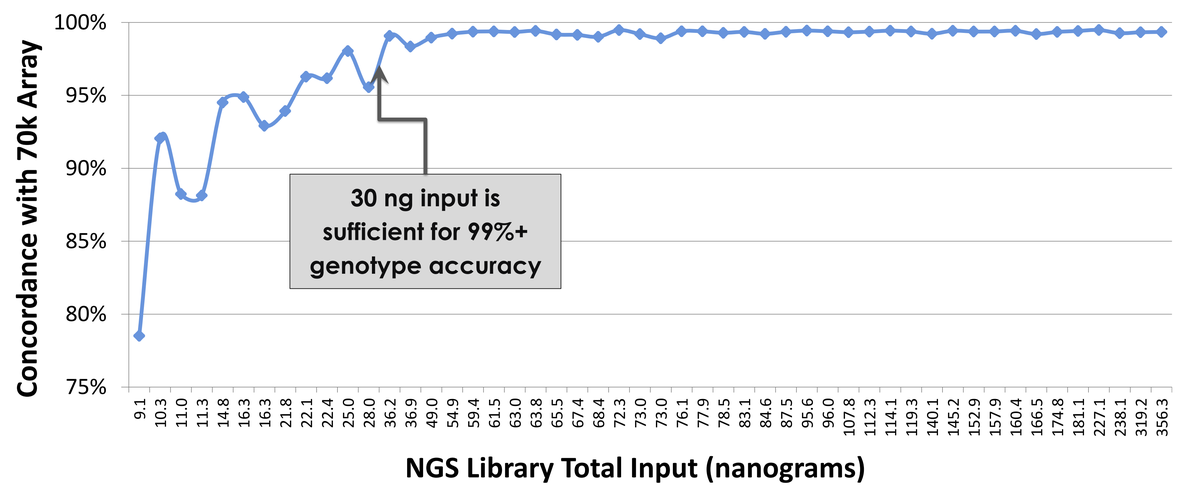
Microarray data for many of the samples was also available which can be simultaneously visualized alongside the NGS-based genotype results within the CURIO platform. Here the genotype data for three different animals is shown, comparing the microarray-based results to the NGS-based results.



chr1:131,527,427-133,159,656

DNA Input – VS – Array Concordance

Finally, a concordance analysis was performed within the platform to compare the microarray-based results to the NGS-based results. The plot here shows the overall concordance results in relationship to the amount of DNA input that was used when preparing each library. A minimum of 30 nanograms consistently provided for 99%+ genotype accuracy across the sample population. Upon deeper inspection, the remaining incongruent sites (< 1%) appear to be the result of microarray manifest inconsistencies.



** (All screenshots, results, and charts shown are taken from the Curio Genomics platform)*