Development of an Optimized Hybrid Capture System for Target Enrichment of Bovine Samples

Ryan O'Donnell, Owen Smith, Cibelle Nassif, Paul Doran, Michael Bocek, Elian Lee, Esteban Toro, Siyuan Chen
Twist Bioscience, Research and Development, South San Francisco, California, USA

Abstract

High-throughput next-generation sequencing workflows are required for population-level bovine genomic research that aid SNP detection and discovery for genetic verification, breeding, and microbiome analysis. With the rapid decline in sequencing costs and driven by the superior data quality as well as the ability to use a single platform for both variant discovery and genotyping, sequencing-based methods are increasingly replacing traditional genotyping arrays.

Targeted sequencing is achieved using a DNA probe target enrichment panel that specifically hybridizes to predefined genomic regions of interest. Effective target enrichment requires a hybrid capture solution that maximizes both the uniformity and coverage of the desired targets, of which, blockers play a critical role in this process by preventing non-specific hybridization between target enrichment probes and off-target sequences. By minimizing unwanted probe and target sequence interactions, blockers enhance capture specificity and reduce the amount of sequencing required per sample, ultimately improving data quality and cost efficiency.

Human-specific genomic blockers have previously been used to reduce the non-specific hybridization of repetitive regions of non-human genomes, but their utility is extremely limited. Recently, Twist Bioscience developed a comprehensive suite of products, including the Bovine Blocker Solution and a Genotyping Panel - Bovine 100k, to deliver an optimized, end-to-end target enrichment workflow for bovine genomic and agricultural studies. The Bovine Blocker Solution is specifically designed to block repetitive regions of the ARS-UCD1.2/bosTau9 genome, reducing off-target binding during target enrichment of NGS libraries and thus improving sequencing efficiency. The Twist Genotyping Panel - Bovine 100k is designed to target approximately 100k SNP markers including common markers found in SNP arrays such as the Illumina Bovine SNP50k, as well as additional SNPs crucial to parentage verification and discovery.

In this study, to demonstrate the effectiveness of these bovine-specific target enrichment products, we leveraged Twist's high-throughput, self-normalizing FlexPrep™ library preparation and 96-plex FlexPrep™ target enrichment workflow, which efficiently captured genomic regions corresponding to those represented in current bovine SNP arrays.

As target enrichment sequencing applications expand, optimized bovine panels and species-specific genomic blockers offer significant potential to enhance breeding program efficiency, accelerate genetic gain, and advance the precision and sustainability of modern agricultural practices. The Twist Bovine Blocker Solution and Twist Genotyping Panel - Bovine 100k are well positioned to reduce off-target reads, enabling researchers to lower sequencing costs per sample while maximizing usable data for bovine genetic testing.

FlexPrep Ultra High-Throughput Library Prep Workflow

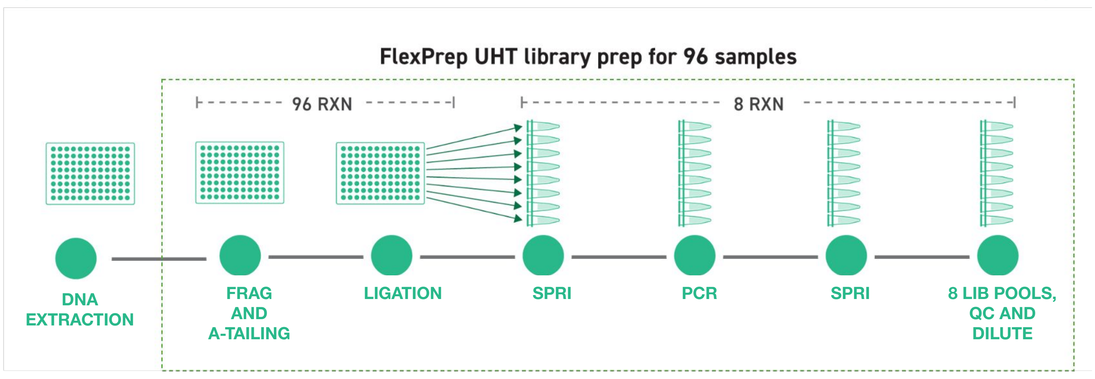

Figure 1. FlexPrep Ultra High-Throughput (UHT) Library Preparation Workflow for 96 Samples. Workflow begins with a fragmentation and A-tailing module in 96-sample format. Next, ligation proceeds for each reaction individually using self-normalizing and inline barcoded adaptors. Inline barcodes allow 96 samples to be pooled into eight library pools.

Twist FlexPrep UHT Library Preparation Kit provides a streamlined workflow that simplifies library preparation and reduces reagent usage.

Inline Barcode Adapters & Input Normalization

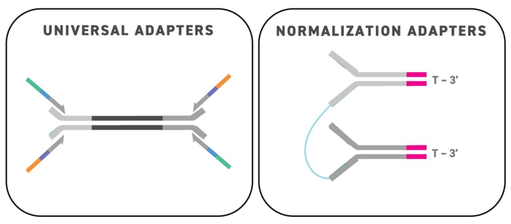

Figure 2. Structure of Normalization Adapters in Comparison to Universal Adapters. Twist FlexPrep Normalization Adapters feature a linked adapters and introduce an inline barcode (pink) to allow for multiplexed NGS library pooling for cleanup and PCR. Twist FlexPrep Normalization Adapters are compatible with Twist UDI primer system.

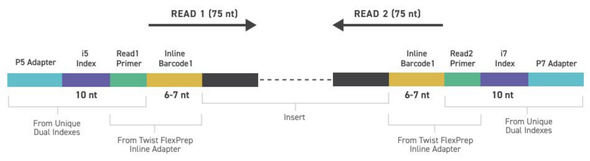

Figure 3. Twist FlexPrep UHT Library Structure. After ligation with Twist FlexPrep Normalization Adapters and indexing with UDI primers by PCR, library molecules will be flanked by hybridization adapters (P5/P7). Molecules will contain standard i5 and i7 indices used for pool-level demultiplexing, and inline with read 1 and read 2 will contain 6-7 bp inline barcodes for sample-level demultiplexing within each pool.

- Inline barcodes on adapters allow for pooling of samples post-ligation.
- Normalization adapters convert independent of insert concentration.

FlexPrep Target Enrichment Capture with Twist Genotyping Panel - Bovine 100k & Twist Bovine Blocker Solution

- Target enrichment is performed by combining the 8 library pools from each plate of FlexPrep libraries together with equal mass inputs of 800 ng each for 6.4 ug total and performing an overnight hybridization reaction to separate the target sequences of interest from the genomic material that makes up the input library.

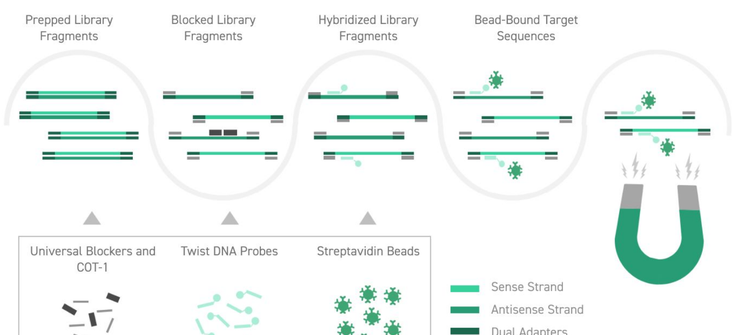

Figure 4. Twist FlexPrep Target Enrichment. Library pools are QC'ed after FlexPrep to determine size and concentration before normalizing the mass input into a 96-plex capture with target enrichment panel(s) and blockers. Blockers reduce non-specific binding and daisy-chaining during hybrid capture to help enrich for sequences of interest in the final capture library solution. 120 base-pair biotinylated probe sequences complementary to the sequences of interest bind during the hybridization reaction and are separated from the remaining library fragments using streptavidin binding beads and subsequent washes.

- All 96 samples from each plate were captured together in a single TE reaction with the Twist Bovine 100k SNP Panel and Twist Universal Blockers. Each 96-plex capture reaction underwent different blocker solution conditions and included no blocker solution, human blocker solution, or bovine blocker solution.
- The Twist Genotyping Panel - Bovine 100k is a targeted sequencing panel designed to target approximately 100k SNP markers, including targets found in the most commonly used SNP arrays such as the Illumina BovineSNP50k and from a variety of research groups, councils & societies that specialize in bovine genomic research.

Panel Name	TE ID	Panel Size (Mb)	Targets Sources	Reference Genome
Twist Genotyping Panel - Bovine 100k	TE-95409079	12.1	ILMN BovineSNP50k, iSAG 195, ICAR, GGP, CDCB	BosTau9 (GCF_002263795.3 aka ARS_UCD_2.0)

- Twist Bovine Blocker Solution is a bovine-specific genomic blocker designed against the BosTau9 genome assembly and tested against the most commonly commercially utilized breeds for beef production in the US. This product reduces the non-specific binding during hybridization to perform

NGS TE Sequencing vs Genomic Arrays

- DNA Arrays are an older approach for genomic testing that predates Next Generation Sequencing (NGS) and are still used by some bovine-focused genetic researchers to this day.
- Current NGS assays provide numerous advantages over array-based assays:

Assay	Cost Per Sample	Input Mass Required	Utility	Flexibility	Coverage Uniformity
DNA Arrays	+	++	Genotyping known variants	Designed for a fixed number of targets, not easy to modify	Highly variable coverage
NGS TE	+	+	Genotyping and Discovery (Known and Unknown Variants)	Panels can be easily modified to add or remove targets	Difficult targets can be boosted to provide more even sequencing coverage

- DNA libraries generated for NGS can be subsequently re-assayed with different target enrichment panels
- Twist Bioscience offers a comprehensive integrated workflow for sample extraction, library preparation, & target enrichment

Target Enrichment Capture Performance - Panel & Blockers

- 96-Plex TE Capture libraries generated with the FlexPrep target enrichment protocol were sequenced on an Illumina NextSeq 2000 with a P4 300-cycle kit at 2x151 read lengths and downsampled to 100x raw coverage prior to analysis using Twist's standard TE data analysis pipeline.
- Picard HsMetrics were generated to assess the performance of each of the 96-plex capture libraries generated with each of the blocker solution conditions tested

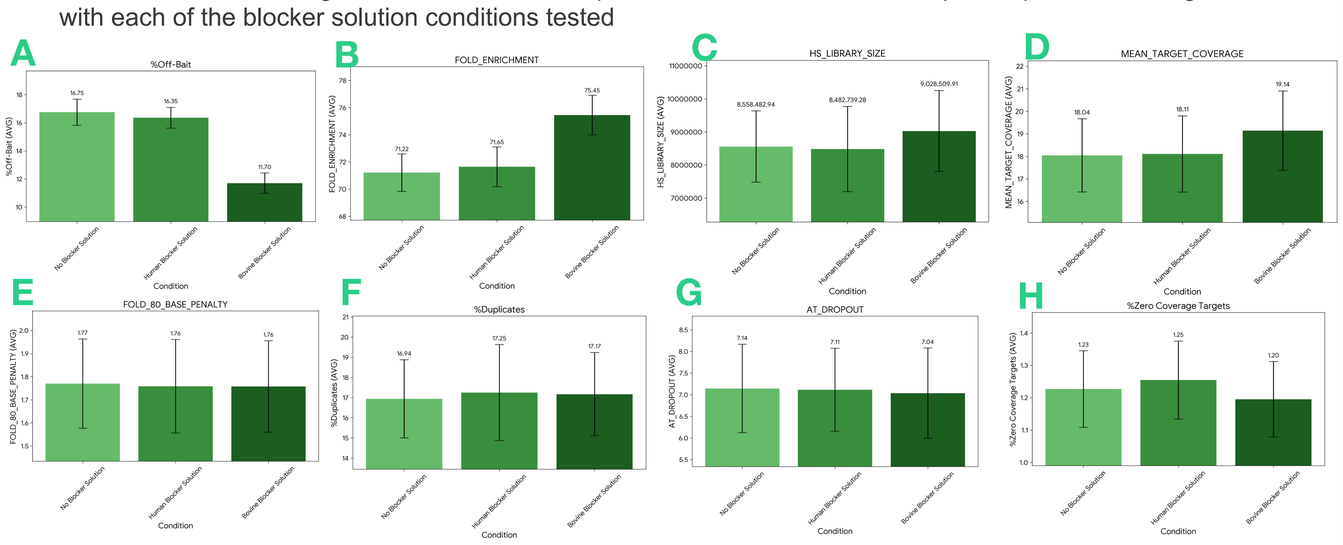

Figure 5. Target Enrichment Sequencing Picard Metrics. The capture libraries generated with the Twist Bovine Blocker Solution showed improved %Off-Bait (A), Fold Enrichment (B), Hs Library Size (C), & Mean Target Coverage (D); while maintaining performance for Fold 80 (E), %Duplicates (F), AT/GC Dropout (AT Dropout shown) (G), % Zero Coverage Targets (H)

- Improvements in Mean Target Coverage averaged across all samples for the capture library that included the Twist Bovine Blocker Solution allow for ~5% more samples to be loaded on the same sequencing run and achieve the same level of mean target coverage for the given level of downsampling. This increase in the number of samples able to be loaded per sequencing run can lead to a ~5% reduction in cost per sample.

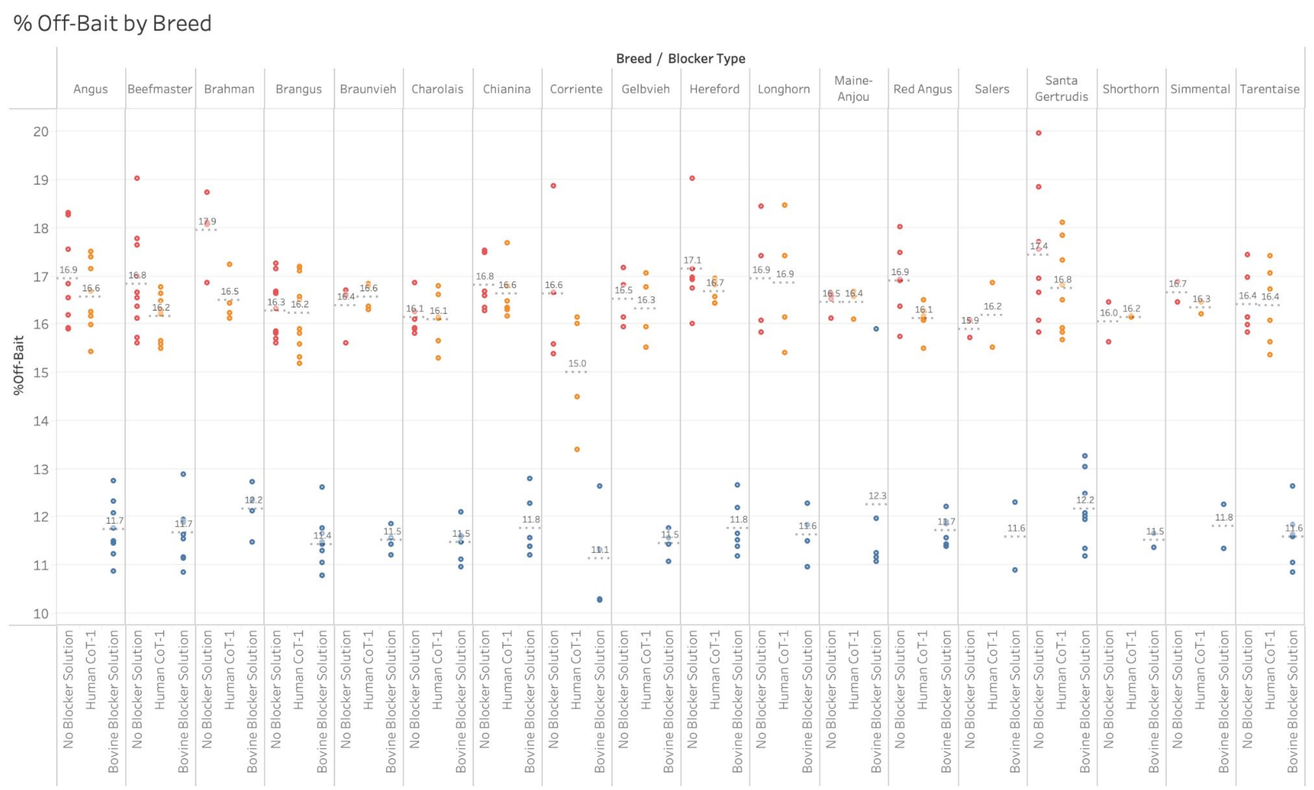

Figure 6. TE Sequencing - %Off-Bait by Breed. The inclusion of the Twist Bovine Blocker solution in the TE capture reaction reduces the relative % Off-Bait picard metric by ~30-40% across the 18 breeds tested in the study captured with the Twist Genotyping Panel - Bovine 100k. This leads to a relative improvement in %On-Bait which translates to improvements in the TE sequencing metrics of interest for customers.

Conclusions

We describe the Twist FlexPrep UHT Library Preparation and Target Enrichment workflow performed on gDNA from 48 individuals across 18 different breeds representing the vast majority of germplasm used for beef production in the US. Utilizing novel normalization adapters with inline barcodes allows the early pooling of samples and self-normalization to reduce the total number of steps, reagents, and consumables needed for the user. Specifically, the multiplexing technology allows for 12 samples to be pooled after ligation and for 96 samples to be pooled into a single target enrichment reaction, which reduces the total number of workflow steps, reagents, and consumables needed. Here we demonstrate the utility of the FlexPrep UHT and Target Enrichment workflows to generate NGS libraries utilizing the Twist Genotyping Panel - Bovine 100k and the Twist Bovine Blocker Solution to maximize the efficacy of the data generated and allow breeders and researchers to move beyond array-based workflows.

Conflict of Interest Statement

All authors are employees and shareholders of Twist Bioscience.