# Development of a high-throughput NGS library preparation workflow with normalization adapters and inline barcode technology

Owen Smith, Kristin Butcher, Cibelle Nassif, Sean Tighe, Su Maw, Tong Liu, Danny Antaki, Esteban Toro, Elian Lee, Ramsey Zeitoun, Siyuan Chen

## Abstract

Advancements in next-generation sequencing (NGS) continue to decrease sequencing costs by increasing data generated per run. This reduction in sequencing cost enables population-level genomic experiments to help study and diagnose genetic disorders. Despite sequencer advancements, efficient NGS library preparation is hindered by the cost and effort needed to process individual samples. Libraries need to be quantified and pooled to equitably distribute sequencing reads, which becomes prohibitively tedious, costly, and time-consuming as sample numbers increase. We present a streamlined NGS library preparation technology that eliminates the need for sample-by-sample handling normalization, and pools samples early in the process, resulting in great simplification in workflow and up to a 48x increase in post-ligation sample processing throughput without compromising data quality.

This high-throughput workflow is built upon enzymatic fragmentation of samples that are converted to sequenceable libraries using novel normalization adapters with inline barcodes. Library normalization occurs from adapters during the ligation step, achieving library conversion independent from DNA input mass (20-200 ng, CV < 20%) on-par with qPCR-based pooling. Up to 48 inline barcodes are included on the adapters which uniformly ligate to samples for simple demultiplexing. The design of the inline barcodes allows for flexible pooling where up to 48 samples can be pooled in a modular manner. With normalization adapters and inline barcodes, pooling is performed immediately after adapter ligation, significantly reducing the number of clean-up and PCR reactions. This format can also support up to tens of thousands of libraries in a single sequencing run using dual unique index sequencing primers post-ligation. Finally, multiplexed PCR amplification artifacts can be reduced using this workflow. Library conversion is consistent and well-suited for both low-pass whole genome sequencing (WGS) and targeted sequencing for variant calling (>20x coverage).

This new library preparation process applies a novel technology to reduce hands-on time and increase efficiency. A process that once treated each sample individually can now be pooled while ensuring equal conversion independent of DNA mass input. This workflow significantly advances the experimental process of library preparation to complement throughput advancements made in sequencing instrumentation.

## High-Throughput Workflow

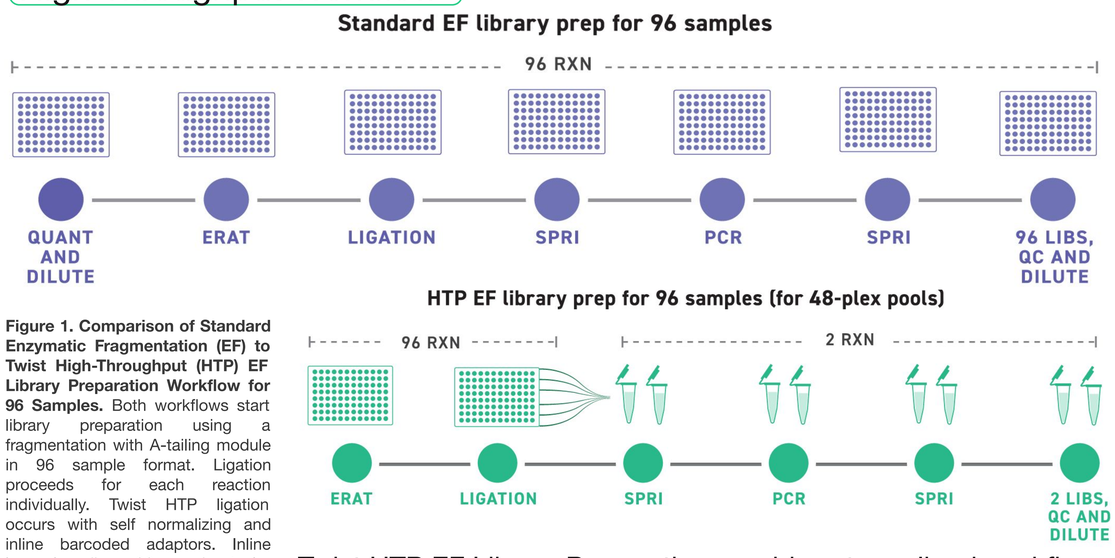


**Figure 1. Comparison of Standard Enzymatic Fragmentation (EF) to Twist High-Throughput (HTP) EF Library Preparation Workflow for 96 Samples.** Both workflows start library preparation using a fragmentation with A-tailing module in 96 sample format. Ligation proceeds for each reaction individually. Twist HTP ligation occurs with self normalizing and inline barcoded adaptors. Inline barcodes allows 96 samples to be pooled into fewer tubes simplifying downstream steps.

Twist HTP EF Library Preparation provides streamlined workflow that simplifies library preparation and reduces reagent usage.

## Inline Barcode & Normalization

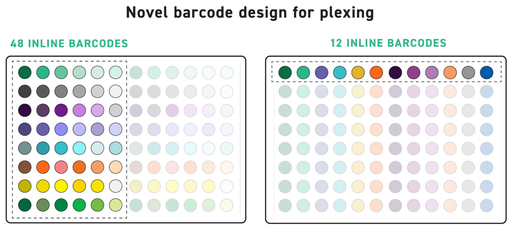


**Figure 2. Two Example Plate Layouts of Normalization Adapters.** The plate on the left depicts the layout of two identical sets of 48 inline barcodes. This plate of 96 samples can be pooled into two pools of 48. The plate on the right depicts the layout of eight identical sets of 12 inline barcodes. This plate of 96 samples can be pooled into eight pools of 12 in a automation friendly manner. Inline barcodes allow for sample demultiplexing within each pool. Individual pools can also be prepared with indices added by PCR for pool level demultiplexing.

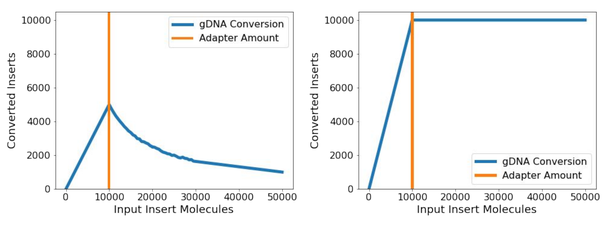


**Figure 3. Theory of Normalization Adapters.** Left Figure: The theoretical plot of DNA insert conversion when standard adapters are limited. Conversion requires two independent ligations and is dependent of insert concentration. Right Figure: The theoretical plot of DNA insert conversion when normalization adapters are limited. Conversion of a DNA insert requires a single normalization adapter. When adapter is limiting, conversion is independent of insert concentration.

- Inline barcodes on adapters allow for pooling of samples post-ligation.
- Normalization adapters convert independent of insert concentration, allowing for consistent conversion and even coverage when pooling a range of DNA masses.

## Enzymatic Fragmentation

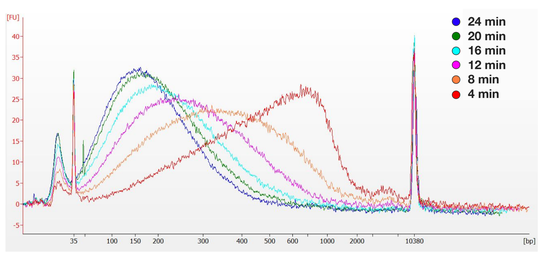


**Figure 4. Tunable Fragmentation Size with Time.** The fragmentation and A-tailing module can be manipulated by adjusting duration of fragmentation time to generate consistent DNA fragments.

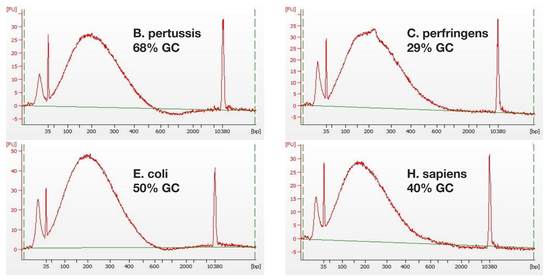


**Figure 5. Consistent Fragmentation using various GC Genomes.** The fragmentation and A-tailing module generates consistent fragment sizes regardless of GC content of DNA input. Four species with variable GC content are all fragmented with this module.

Twist HTP EF Library Preparation uses tunable EF that is robust to varying GC content.

## Whole Genome Sequencing Results

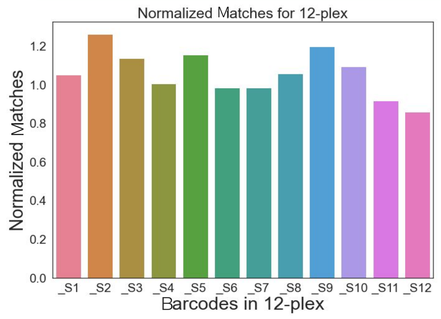


**Figure 6. Inline Barcode Uniformity with Equal Mass Input.** Bar plot of normalized reads with matching barcodes (Normalized Matches) after WGS. Twelve unique inline barcodes each with 60ng of input DNA were used in the Twist HTP EF Library Preparation workflow. CV for set of 12 is 10.6%.

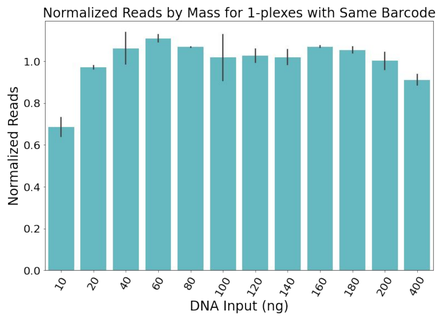


**Figure 7. Normalization Across Range of DNA Input Masses for Single Inline Barcode.** Bar plot of Normalized Reads after WGS. One inline barcode was used for individual library preparation with variable DNA mass inputs. CV for range of 20-200ng is 5.8%.

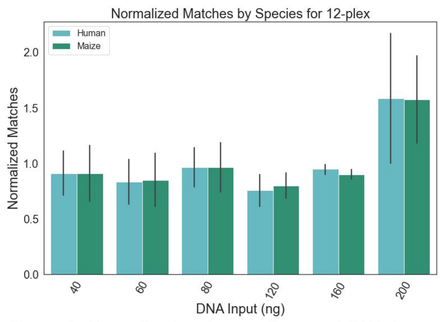


**Figure 8. Normalization Across Range of DNA Input Masses for Distinct Inline Barcode.** Bar plot of Normalized Matches after WGS with HTP EF Library Preparation workflow performed as 12-plex. Data shown from two independent 12-plexes, with 6 masses per plex for each species. CV for 12-plex is 39.3% for human and 35.7% for maize pools.

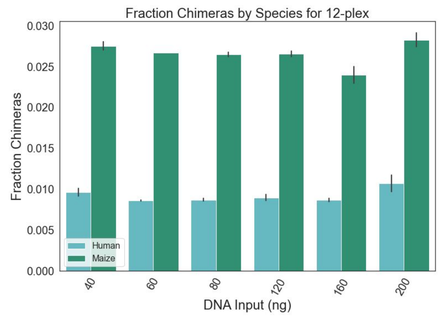


**Figure 9. Fraction Chimera for Range of DNA Input Masses in 12-plex for Human and Maize samples.** Bar plot of Fraction Chimeras after WGS with HTP EF Library Preparation workflow performed as 12-plex for each species. Data shown from two independent 12-plexes, with 6 masses per plex.

- Barcodes display uniformity when same mass used for library preparation.
- One inline adapter shows strong normalization across range of 20-200ng.
- Normalization achieved with the Twist HTP EF Library Preparation workflow when using distinct barcodes with a range of DNA mass inputs using both Human and Maize genomic DNA.
- Chimeras are not impacted by DNA mass inputs, but the genomes tested display differences in chimeric fraction.

## Target Enrichment Results

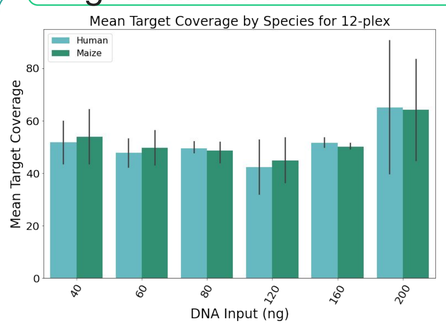


**Figure 10. Mean Target Coverage using the Twist HTP EF Library Preparation Workflow.** Bar plot of Mean Target Coverage after target enrichment (TE) with HTP EF Library Preparation workflow performed as 12-plex. Data shown from two independent 12-plexes, with 6 masses per plex for each species. Human capture performed using a 800kb panel and Maize with 1.25Mb panel, with read depth downsampled to 150x per library in pool. Duplication rate for human captures are ~15% and for maize captures ~32%.

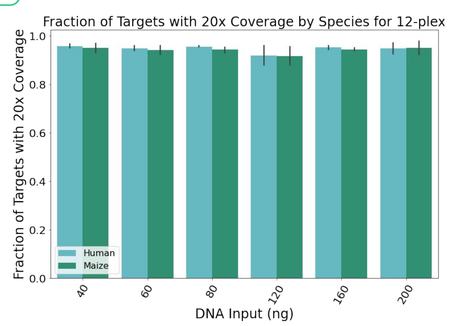


**Figure 11. Percent Target Bases at 20x Coverage using the Twist HTP EF Library Preparation Workflow.** Bar plot of Percent Target Bases at 20X Coverage after TE with HTP EF Library Preparation workflow performed as 12-plex. Data shown from two independent 12-plexes, with 6 masses per plex for each species. Human capture performed using a 800kb panel and Maize with 1.25Mb panel, with read depth downsampled to 150x per library in pool.

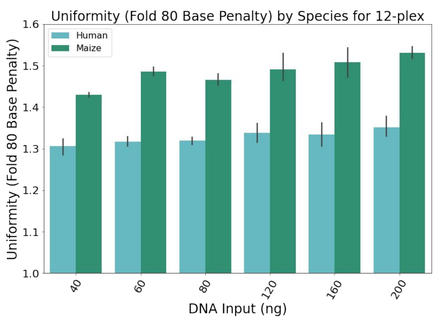


**Figure 12. Coverage Uniformity using the Twist HTP EF Library Preparation Workflow.** Bar plot of Uniformity (Fold-80 Base Penalty) after TE with HTP EF Library Preparation workflow performed as 12-plex. Data shown from two independent 12-plexes, with 6 masses per plex for each species. Human capture performed using a 800kb panel and Maize with 1.25Mb panel, with read depth downsampled to 150x per library in pool.

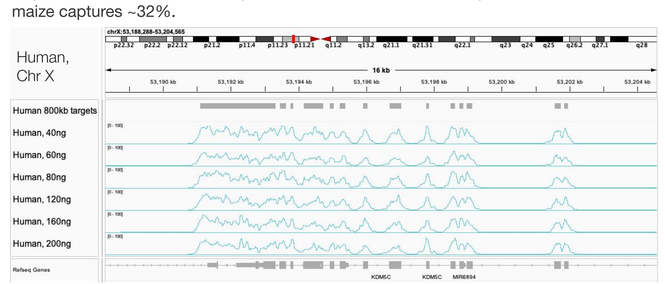


**Figure 13. Human Genome Coverage Plot After the Twist HTP EF Library Preparation Workflow.** Genome coverage plot for human KDM5C locus after TE using a 800kb panel without downsampling. Displaying one replicate for each mass input from a single 12-plex.

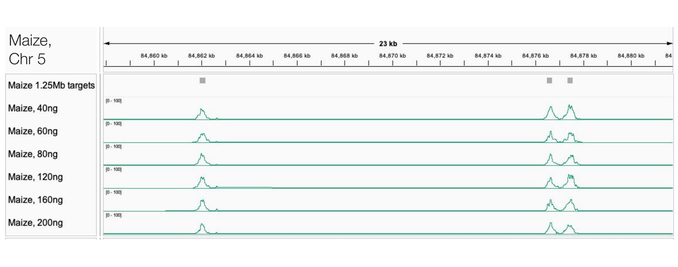


**Figure 14. Maize Coverage Plot After the Twist HTP EF Library Preparation Workflow.** Genome coverage plot for Maize Chr 5 locus after TE using a 1.25Mb panel without downsampling. Displaying one replicate for each mass input from a single 12-plex.

- Approximately 50x for human (800kb panel) and maize (1.25Mb panel) Mean Target Coverage is achieved after TE of 12-plex containing six different DNA mass inputs, when downsampling the pool to 150x coverage per library.
- Percent Target Bases at 20X Coverage and uniformity after TE is consistent for each sample in 12-plex despite different inline barcodes and variable DNA mass into library preparation.

## Conclusions

We describe the Twist HTP EF Library Preparation workflow. Utilizing novel normalization adapters allows the early pooling of samples and self-normalization to reduce the total number of steps, reagents, and consumables needed for the user. WGS and TE metrics generated using this HTP workflow establish normalization across a DNA mass input range of 20ng-200ng with a CV ~35%. TE allows mean target coverage of 50 for each sample in the workflow, enabling genotyping of samples robustly.

## Conflict of Interest Statement

All authors are employees and shareholders of Twist Bioscience.