Application Note

Scalable Transposase-Based Multiplexed Library Prep with Auto-Normalizing UDI Capability Using PurePlex™ on the Complete Genomics™ DNBSEQ-G400 Genetic Sequencer

- purePlex™ DNA Library Preparation offers a fast, flexible workflow with no requirement for full plate processing. Auto-normalization reduces QC burden and improves data consistency, while the early pooling step reduces the need for sample handling.
- DNBSEQ-G400 offers unparalleled accuracy and maximum versatility. The DNBSEQ-G400 provides a range of sequencing options, with read lengths up to SE400/PE300, generating up to 1440 Gb of data per run. It offers two types of flow cells (FCS and FCL), providing 300M-1800M reads per flow cell.

purePlex DNA Library Preparation Kits feature speed, batch flexibility, and data confidence with UDIs, with only 45-minute of hands-on time (and 2.5 hours of total time) for 96 samples. Robust performance simplifies implementation in high-throughput low-pass whole genome sequencing, whole small genome sequencing (<50 Mb), or metagenomics/ microbiome screening.

Introduction

Low-coverage WGS with imputation is a cost-effective alternative to microarrays that allows genotyping at orders of magnitude more positions.[1] In this application note, we demonstrate the integration of the seqWell purePlex DNA Library Preparation Kit and the Complete Genomics DNBSEQ-G400 Genetic Sequencer for low-pass human whole genome sequencing. The combined solution provides a highly flexible, cost-effective option for laboratories as they expand their sequencing capabilities.

Materials and Methods

Genomic DNA and DNA quantification – Eight individual human genomic DNA (hgDNA) preparations were obtained from the Coriell Institute for Medical Research. The set included the well-characterized CEPH/ Utah pedigree 1463 HapMap reference, the Ashkenazi cohort, as well as the Han Chinese cohort (Table 1). These samples were quantified using an Infinite® F200 PRO microplate reader (Tecan) and Quant-IT™ PicoGreen dsDNA Assay Kit (ThermoFisher Scientific) prior to library preparation.

Table 1. Summary of human genomic DNA used in this study.

Coriell ID	NIST ID	Ethnicity
NA12878	HG001	Utah/Mormon
NA12891	N/A	Utah/Mormon
NA12892	N/A	Utah/Mormon
NA24385	HG002	Ashkenazi
NA24149	HG003	Ashkenazi
NA24631	HG005	Chinese
NA24694	HG006	Chinese
NA24695	HG007	Chinese

Library construction with the purePlex™ DNA Library Preparation Kit and library conversion using DNBSEQ Universal Library Conversion Kit – The eight individual hgDNA were processed in triplicate with purePlex DNA Library Preparation Kit, resulting in a normalized 24-plex library pool. Samples were processed according to manufacturer's protocol (Figure 1A). DNA input ranged from 45.3 – 50.1 ng (median: 47.6 ng). The resultant library pool concentration and fragment size were determined by Qubit (1X HS kit) and Agilent TapeStation 2200 with HSD5000 ScreenTape. The library was further prepared for sequencing on the Complete Genomics DNBSEQ-G400 (Figure 1C) using the DNBSEQ Universal Library Conversion Kit (Figure 1B).

Sequencing and data analysis – Sequencing of purePlex libraries was performed with 2 x 150 bp paired end sequencing to a depth of ≥20 M read pairs per sample. Sequencing data were demultiplexed, and fastq files generated with the Complete Genomics base-calling pipeline. Paired end reads for each sample were aligned to the GRCh38 human reference genome using BWA MEM. Library QC metrics and sequencing statistics such as library insert size, library complexity, and genome coverage was calculated using standard tools from the Picard suite.

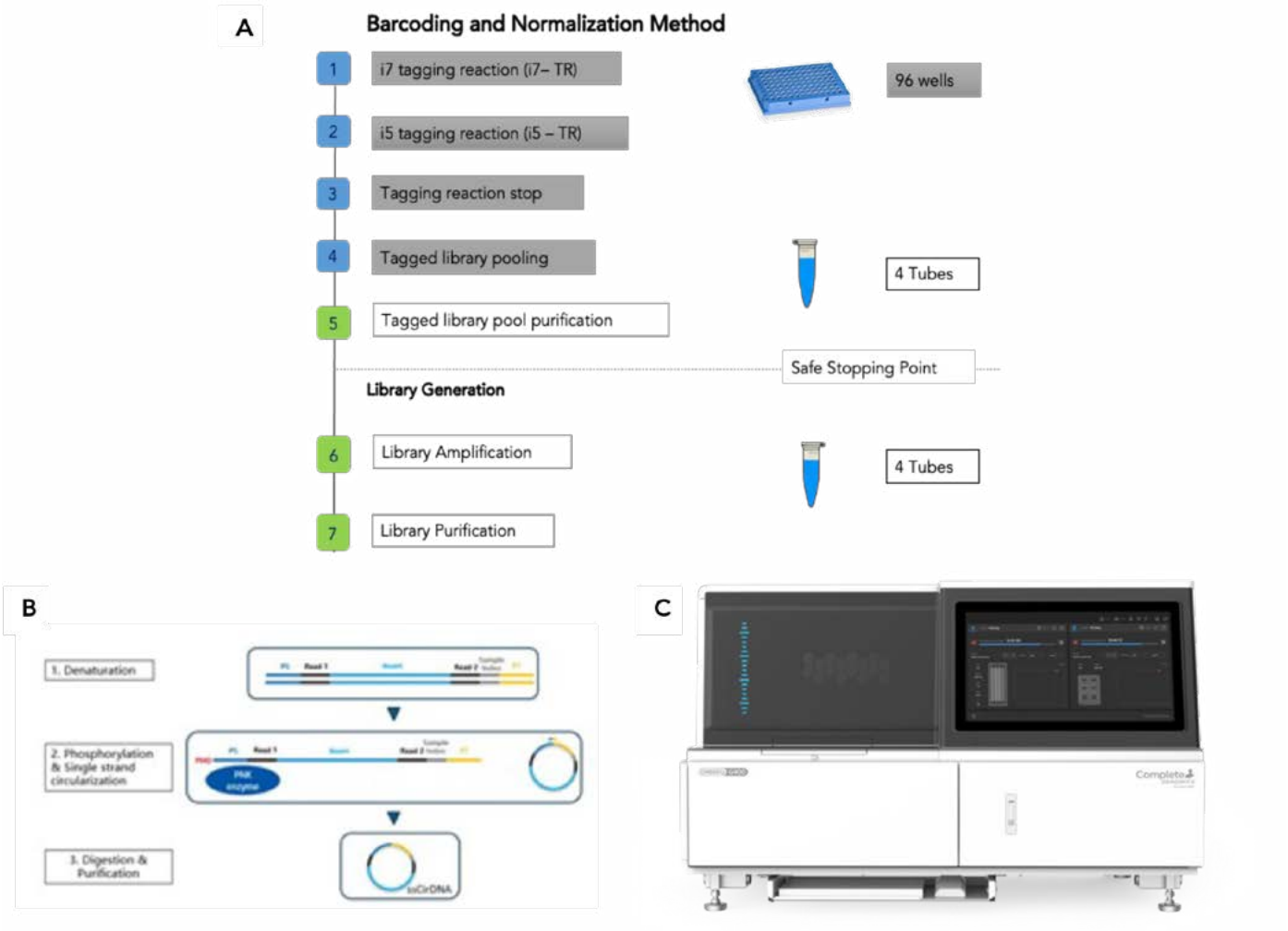

Figure 1. The wet lab workflow starts with the purePlex DNA Library Preparation Kit (A) and culminates with DNBSEQ-G400 sequencing (C). (B) Using a traditional PCR-free approach, the DNBSEQ Universal Library Conversion Kit converts sequencing libraries (linear double-stranded DNA libraries) into circular, single-stranded DNA libraries compatible with Complete Genomics' sequencers

Results and Discussion

purePlex DNA library preparation kit utilizes sequential transposase steps to incorporate full length indexed Illumina P7 and P5 adapter sequences. Each kit contains 96 unique i7 and i5 indices enabling processing of up to 96 samples. Additional index sets are available to achieve multiplex levels >96. Briefly, i7 tagging reactions are carried out on 5-50 ng of DNA input. Following the initial indexing reaction, a novel normalization reagent and i5 tagging reagent are added to each well for the 2nd tagging reaction. Samples are pooled volumetrically, undergo purification, and library generation to complete the 2.5-hour (45-minute hands on) workflow. Built-in auto-normalization alleviates the need to normalize sample input. Pooling during the workflow results in normalized pools of 24 samples greatly reducing the QC burden prior to further multiplexing for sequencing.

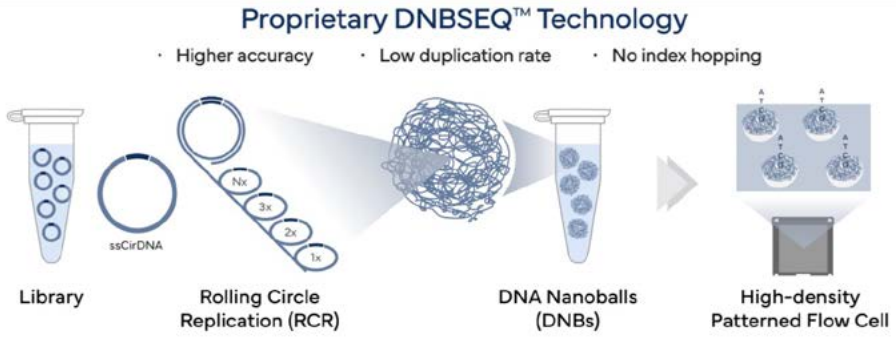

Figure 2. Complete Genomics proprietary DNBSEQ technology. Single-stranded circular DNA (sscirDNA) molecules are created. sscirDNA then serves as a template for rolling circle replication (RCR) to create billions of DNA nanoballs (DNBs). The DNBs are loaded onto flow cells and combinatorial probe-anchor synthesis (cPAS) chemistry hybridizes sequencing primers to the DNBs. Fluorescent probes are incorporated and imaged in successive sequencing cycles.

Table 2. Summary of 24-plex purePlex multiplexed library performance on DNBSEQ-G400 sequencer. Average duplication rate and genome coverage after samples were down sampled to 20M read pairs.

≥Q30	Average Median Insert Size	Reads Aligned	Duplication Rate	%Genome covered at ≥1X	Mean Coverage Depth (X)
90.7%	348nt	99.7%	4.52%	69.8%	1.51

The DNBSEQ Universal Library Conversion workflow preserves insert size and indexing of the purePlex library pool. Median insert sizes >300 nt (Table 2, Figure 3) paired with the low duplication rate of the DNBSEQ-G400 sequencer are well suited to multiplexing 96 samples per flow cell for 1X coverage using 2x150 sequencing.

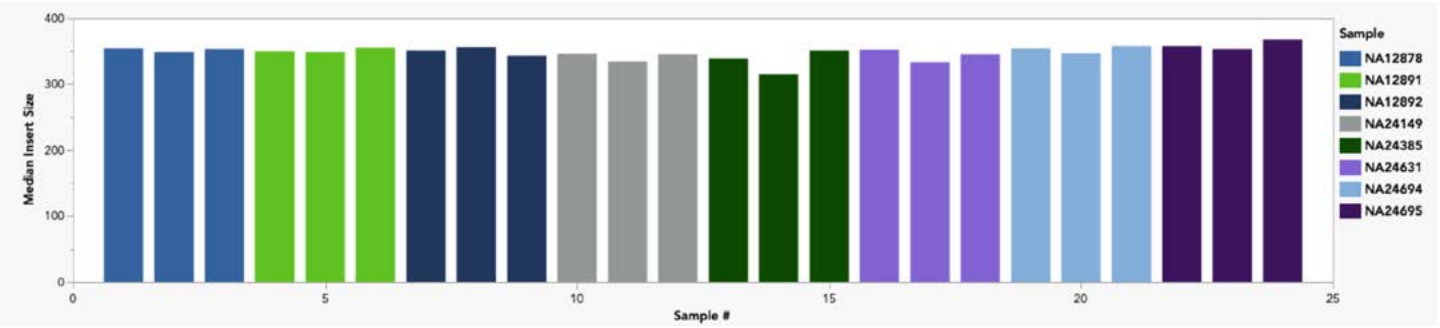

Figure 3. Insert size uniformity of 24-plex purePlex multiplexed library of 8 individual human genomic DNA. Insert size of 350-400 nt is optimal to generate unique data from 2x150 paired sequencing.

Summary

purePlex™ DNA library preparation kit supports truly multiplexed and highly scalable construction of library pools for low-pass whole genome sequencing. Together with DNBSEQ-G400 sequencing system, these two technologies offers an accurate, versatile, and flexible sequencing option. Specific benefits of the combination of purePlex DNA library prep and DNBSEQ-G400 sequencer include:

- A novel and streamlined unique dual indices library preparation workflow that enables the preparation of four auto-normalized 24-library pools in approximately 2.5-hour, with 45-minute of hands-on time.
- Plate-based reagents, flexible kit configurations, and up to 384 unique dual indexing barcodes that support different batch sizes and facilitate implementation in a wide range of laboratory settings.
- A versatile benchtop sequencer providing users with accurate, flexible, and efficient sequencing options.
- DNBSEQ sequencing technology produces high quality data and ensures exceptional accuracy, eliminating clonal errors and index hopping-generating >99% SNP/indel precision and sensitivity.
- Built with a dual flow cell system that can perform different types of flow cell individually in a single run, giving users a more flexible and streamlined sequencing.

[1]Application Note: Low-Pass Whole-Genome Sequencing Enabled by Scalable Library Preparation Offers a Competitive Alternative to Microarray-Based Genotyping. Download here.

Learn more about purePlex™ DNA Library Preparation Kit and DNBSEQ-G400 Sequencer at seqwell.com/products/pureplex-dna-library-prep-kit/ completegenomics.com/products/sequencing-platforms/dnbseq-g400/