One-step Library Preparation Method Supporting Skim-seq, Low Pass Sequencing Approach for Genotype Imputation

Michelle Rahardja, Ken Tenan, Stella Huang, Yanyan Liu, Jack Leonard, Rebecca Feeley, and Jenna Couture
seqWell, Inc. Beverly, MA USA

Introduction

Recent advancements in next-generation sequencing (NGS) and bioinformatics have enabled whole-genome sequencing (WGS) to become a routine tool in both research and clinical applications. Low-coverage WGS, also known as skim-sequencing or "low-pass" sequencing, combined with imputation is a highly effective and cost-conscious alternative approach to microarray-based genotyping. Here, we demonstrate as a proof of concept, the use of ExpressPlex™ 2.0 Custom High Strength, a one-step library preparation method, for skim-seq approach to genotype imputation on human samples. The ExpressPlex 2.0 Custom High Strength was used to processed two individual human genomic DNA at three different total mass inputs in duplicate, generating a normalized 6-plex library pool. A 2 x 150 bp run on a NovaSeq X Plus sequenced the ExpressPlex libraries to ≥ 20 million paired-end reads per sample. Paired-end reads for each sample were aligned to the GRCh38 human reference genome, the precision and accuracy of SNP calls were determined for chromosome 1. The open-source GLIMPSE pipeline performed imputation using default settings. Our results show the utility of the ExpressPlex 2.0 Custom High Strength for routine low-pass WGS applications, where we characterize multiplexing uniformity and genotype imputation accuracy on a collection of reference samples.

ExpressPlex Library Preparation Workflow

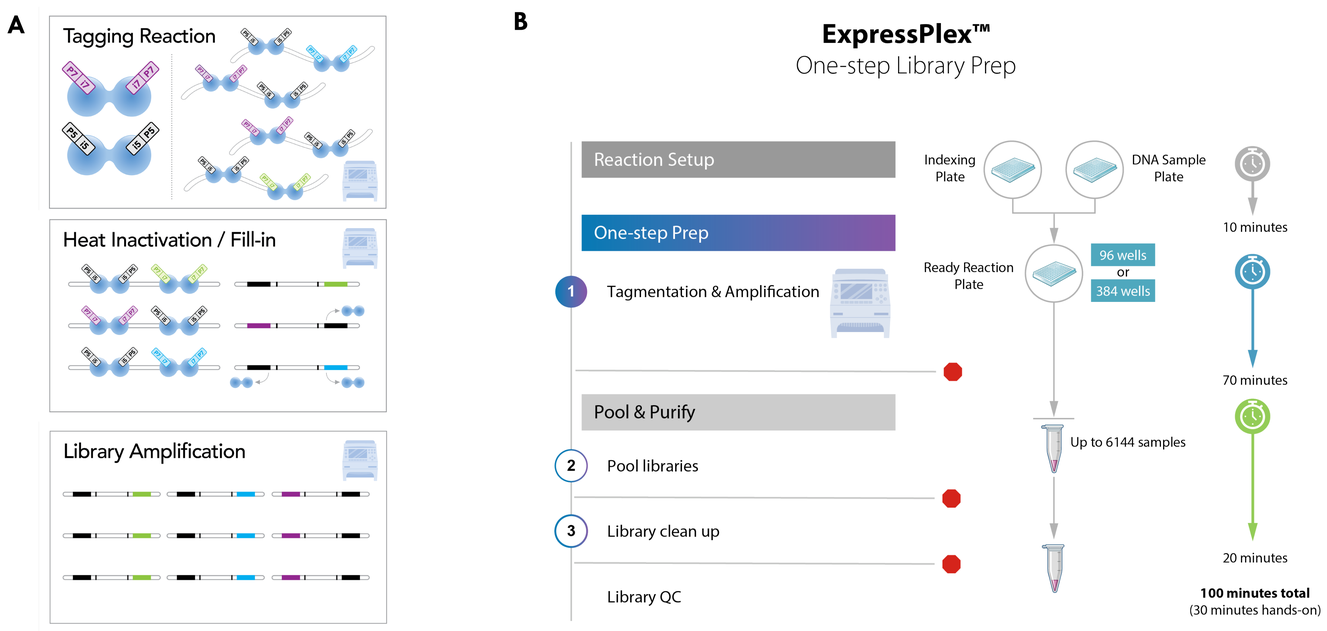

Figure 1. ExpressPlex 2.0 uses seqWell's high performance TnX™ transposase that was specifically engineered for NGS library preparation. (A) ExpressPlex 2.0 library preparation kits tag input DNA with indexed adapters and amplify libraries all in a single reaction. Different full-length i7 indexed adapters tag the 96 DNA samples and barcoded libraries are amplified in separate wells, making for a highly efficient, one-step multiplexed library prep workflow. (B) Using the ExpressPlex 2.0 (96-well) kit, a 96-plex library can be prepared for library QC and sequencing in under 120 minutes, with less than 30 minutes of hands-on time. ***The ExpressPlex 2.0 Custom High Strength used in the experiments and reported in this poster were performed using the same workflow as our standard ExpressPlex 2.0 Library Preparation Kit.***

Methods

- Two individual human genomic DNA from the reference cohorts of CEPH/Utah and Han Chinese cohorts (Table 1) varying in mass input (10, 50, and 100 ng) were processed using the ExpressPlex 2.0 Custom High Strength Library Preparation Kit.
- The 6-plex ExpressPlex library was sequenced on the NovaSeq X Plus. Sequencing data were individually down-sampled to 1M, 2M, 5M, 10M, and 20M random paired-end reads (0.1X, 0.2X, 0.5X, 1X, and 2X coverage, respectively) before variant calling and imputation.
- The open-source GLIMPSE pipeline (v2.0.0) performed imputation using default settings.

Table 1. Summary of hgDNA samples assessed in the study.

Coriell ID	NIST ID	Reference Cohort	DNA Input Amount (ng)
NA12878	HG001	CEPH/ Utah	10
			50
			100
NA24631	HG005	Chinese	10
			50
			100

Genotyping Imputation from Various Depth of Coverage

Figure 2. The F1-score of low-pass WGS genotyping for two individual human genomic samples with ExpressPlex 2.0 Custom High Strength library prep on an Illumina NovaSeq X Plus System at various depth of coverage across three different DNA inputs.

The precision and accuracy of SNP calls were determined for chromosome 1. Prior analysis of a genetically diverse group of human samples sequenced at a low depth indicated that summary statistics of chromosome 1 statistically agreed with the imputation results from all autosomes. Figure 2 and Table 2 show the F1-score, recall, and precision of correct imputed genotype remains high despite various DNA input used at various coverage depth.

Table 2. Summary of genotype imputation accuracy for chromosome 1 in HG001 and HG005 at various depth of coverage

Coriell ID (NIST ID)	Depth of Coverage	Recall	Precision	F1-score
NA12878 (HG001)	0.1X	0.84	0.80	0.82
	0.2X	0.88	0.83	0.86
	0.5X	0.93	0.84	0.88
	1X	0.94	0.85	0.89
	2X	0.95	0.85	0.90
NA24631 (HG005)	0.1X	0.84	0.79	0.81
	0.2X	0.87	0.81	0.84
	0.5X	0.91	0.83	0.87
	1X	0.93	0.84	0.88
	2X	0.94	0.84	0.89

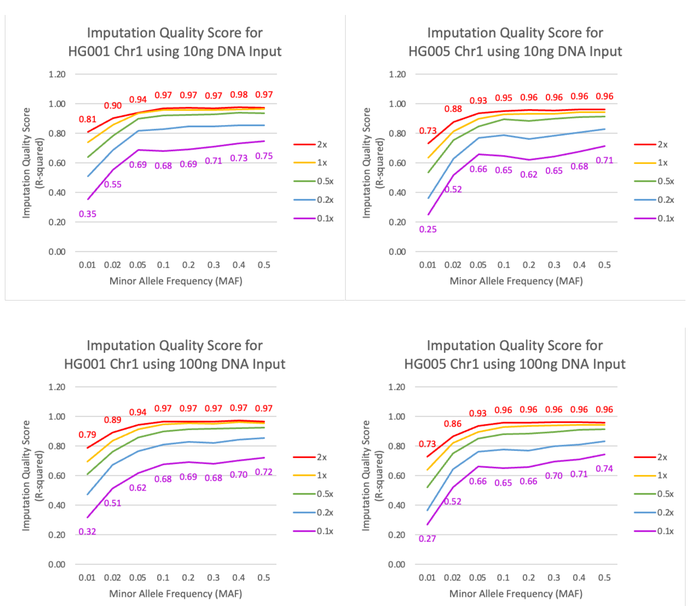

Figure 3. Imputation quality score for all genetic variants at different minor allele frequency (MAF) for chromosome 1 in the GIAB Consortium HG001 and HG005 at various depth of coverage across low (10 ng) and high (100 ng) DNA inputs. The imputation quality score is an estimate of imputation quality on a scale of 0 to 1, where 1 indicates that a genotype has been imputed with high certainty.

Sequencing Metrics

Table 3. Sequencing metrics summary of 6-plex ExpressPlex 2.0 Custom High Strength on the NovaSeq X Plus.

Avg PF Reads	Reads Aligned	Avg Mean Read Length	Avg Median Insert Size	Avg Duplication Rate	Mean Coverage Depth (X)
5.84×10^7	99.6%	140.3	216 nt	19.8%	2.78

Summary and Conclusions

- The ExpressPlex 2.0 Custom High Strength supports truly multiplexed and highly scalable construction of library pools for low-pass WGS through its one-step workflow (Figure 1).
- A high proportion of imputed calls (>98%) were identical to those in the truth data set (Figure 2, Table 2) despite various DNA input used at various coverage depth, confirming that high accuracy of genotyping can be achieved from a minimal sequencing data.
- Increasing depth of coverage from 0.1X to 1X significantly improves the imputation quality score for genetic variant MAF (Figure 3).

The ExpressPlex 2.0 Custom High Strength is currently in Alpha Testing. For more information about this product or interest in joining the alpha program, please contact **alpha@seqwell.com**