Scalable Multiplexed Library Preparation for Low-Depth Sequencing and Genotype Imputation

Ariele Hanek[1], Christopher Fragoso[2], Rebecca Feeley[1], Jack Leonard[1], Joseph Mellor[1]

(1) seqWell, Inc. 66 Cherry Hill Drive, Beverly MA 01915 USA
(2) Verinomics, Inc. 5 Science Park, New Haven CT 06511 USA

PAG XXVII Poster PO0195

Summary

The use of low-pass whole-genome sequencing (WGS) is a robost and cost-effective alternative approach to microarray-based genotyping for genotype imputation pipelines. With prior knowledge of known parental or population haplotypes, genotypes can be accurately imputed from 'shallow' genomic sequencing of individual plant or animal samples at raw coverage depths as low as 0.5x.

The economics of using sequencing in place of microarrays depends on the ability to reliably create balanced pools of libraries that efficiently multiplex as many genomic samples as possible on large high-output sequencing runs that achieve the lowest cost per unit of data.

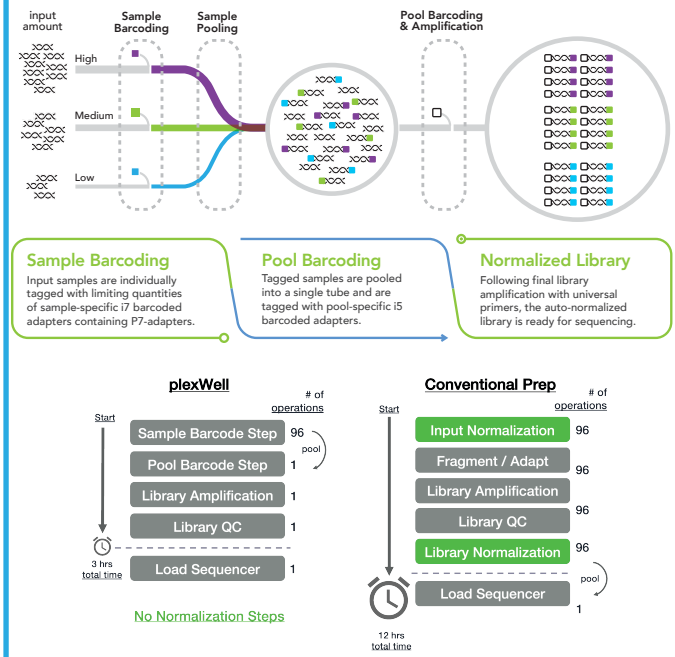
plexWell™ library preparation technology streamlines the ability to create normalized multiplexed libraries from multiple samples in batches of 96 to 384 samples without the need for time-consuming measurement or adjustment of input DNA concentrations, significantly simplifying the complex task of high-level multiplexing.

Methods

plexWell Library Preparation Technology

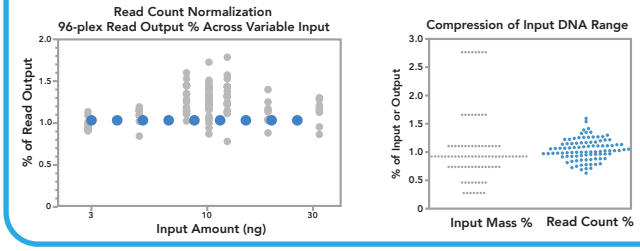
The technical foundation of the plexWell approach (Fig 1) is a reagent-limited initial transposition step performed on many samples in parallel, coupled to a subsequent pooled library generation step; when applied in conjunction, these two steps yield an approximately equal number of sequencer-ready library fragments from each of a potentially large collection of samples across a 10-fold input range.

Figure 1. plexWell Normalizing Iterative Barcoding Technology



The plexWell™ library workflow (above) reduces the time and effort associated with multiplexed library prep, while maintaining multiplex uniformity across variable input DNA concentrations (Fig 2, below)

Figure 2. plexWell Normalization and Multiplexing Performance



Results

139 individuals of a rice recombinant inbred line (RIL) population, derived from an *indica* x tropical *japonica* F_1 hybrid and advanced through single seed descent for 8 generations, were reanalyzed through a workflow of Verinomics DNA extraction, seqWell library preparation, and whole genome sequencing via Illumina NovaSeq 6000 2x150. Sequencing read alignment, variant calling, variant filtering, and imputation were performed by the Verinomics population genomics pipeline. The imputed WGS dataset contained an average of 940,897 SNPs per sample (Table 1., Fig. 3A) and were highly concordant (Fig. 3B) with the previously described GBS sequencing of the population (Fragoso et al. 2017). In addition to being highly accurate, the imputed WGS genotypes facilitate gene-level functional analysis for trait dissection (Fig. 4).

Table 1. Summary Statistics for N = 139 Rice RIL Sequencing

	Total mapped reads	Total genomic coverage (MB)	Coverage per base	Raw SNPs	Filtered SNPs	Imputed SNPs	Fraction of missing imputed calls
Average	9,510,767	1,427	3.82	1,360,634	572,211	940,897	0.08
SD	4,146,824	622	1.67	526,448	224,367	123,089	0.03

Figure 3. (A) Histograms of 139 Rice RIL Summary Statistics of Coverage and Imputed Genotype Calls and (B) Concordance with Genotyping by Sequencing

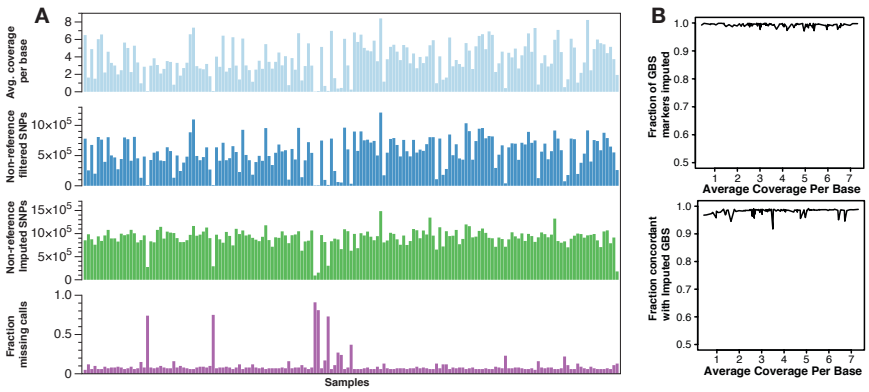
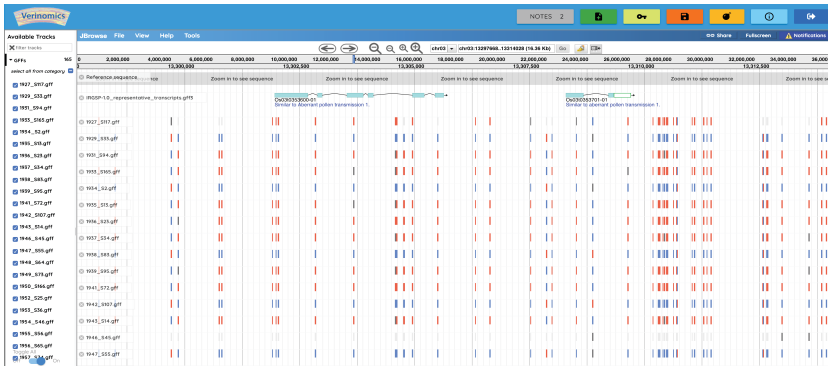


Figure 4. Low-Depth WGS Facilitates Gene-Level Functional Analysis



Conclusions

While low-pass WGS is an effective technical alternative to microarray-based genotyping, the economics of using sequencing in place of microarrays depends on the ability to reliably create balanced pools of libraries that efficiently multiplex as many genomic samples as possible on large high-output sequencing runs that achieve the lowest cost per unit of data.

Our results demonstrate the use of plexWell for routine low-pass WGS applications, where we show robust and scalable genotype imputation accuracy on large sample populations. The normalization capability of the plexWell workflow permits efficient generation of multiplexed libraries from hundreds to thousands of samples while avoiding sample dropouts due to input DNA variation.

References

Fragoso, Christopher, Maria Moreno, Zuoheng Wang, Christopher Heffelfinger, Lady J. Arbelaez, John A. Aguirre, Natalia Franco et al. "Genetic Architecture of a Rice Nested Association Mapping Population." G3: Genes, Genomes, Genetics (2017): g3-117.