High-Throughput Library Preparation for Low-pass Sequence-based Genotyping Pipelines

Ariele Hanek[1], Rebecca Feeley[1], Jacquelyn Wilson[1], James Smith[1], Jack Leonard[1], Joseph Pickrell[2], Kaja Wasik[2], Chase Mazur[2], Joseph Mellor[1]

(1) seqWell, Inc. 376 Hale Street, Beverly MA 01915 USA, (2) Gencove, Inc. 430 E. 29th Street, New York NY 10016 USA

AGBT 2019, poster 510

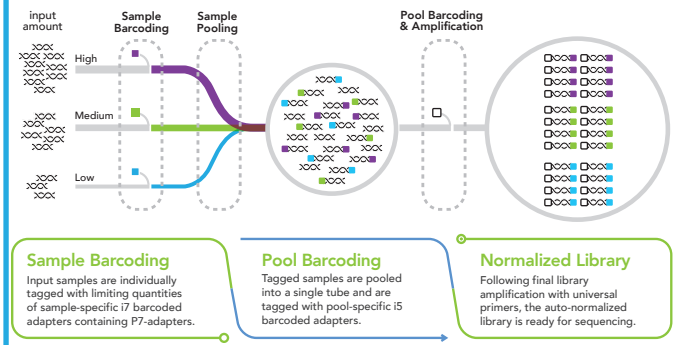
Summary

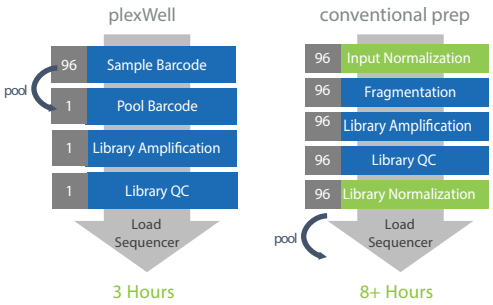
The use of low-pass whole-genome sequencing (WGS) represents a cost-effective and sample-type-agnostic alternative approach to microarray based genotyping for robust genotype imputation pipelines as well as accurate ascertainment of CNV and aneuploidy. As established in previous studies, raw NGS coverage depths from human genomic DNA as low as 0.4x, allow accurate common SNP imputation in human studies comparable to that of widely used commercial microarrays. Widespread use of low-pass WGS as an effective alternative to microarray-based technology, depends strongly on an ability to efficiently and equally prepare and pool large numbers of samples in a single sequencing lane.

Here, we describe the use of a high-throughput library preparation technology (plexWell™) for application to low-pass whole human genome sequencing. plexWell library preparation kits create normalized multiplexed libraries from batches of 96 samples without the need for time-consuming measurement or adjustment of input DNA concentrations, significantly simplifying the complex task of high-level multiplexing. The technical foundation of the plexWell approach 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. Our results show the utility of plexWell for routine low-pass WGS applications, where we characterize multiplexing uniformity and genotype imputation accuracy on a collection of reference samples.

Methods

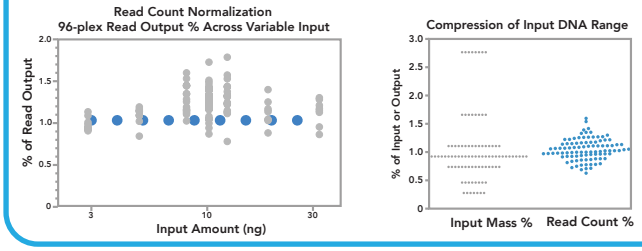
Figure 1. plexWell Normalizing Iterative Barcoding Technology





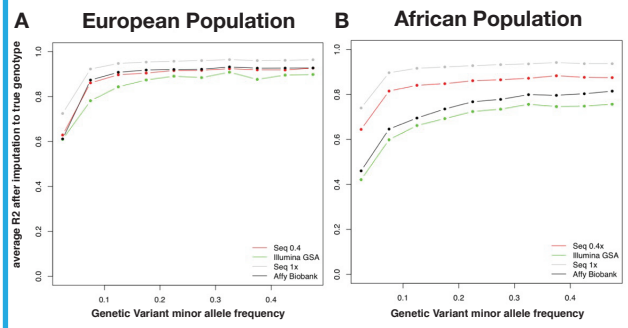
The plexWell™ library workflow (above) reduces the time and effort associated with multiplexed library prep, while maintaining multiplex uniformity across variable input DNA concentrations (Fig 2, below)

Figure 2. plexWell Normalization and Multiplexing Performance



Results

Figure 1. Low Pass WGS performance for genotype imputation outperforms microarrays by 10-20% in well characterized populations (A) and 50-100% in poorly characterized populations (B)



Coverage	SNP Concordance
4X	99.50%
2X	99.30%
1X	99.00%
0.4X	98.10%

Table 1. High SNP Concordance for imputed low-pass genome sequences and gold standard genotypes

Figure 2. plexWell library preparation and sequencing outcome for PW96 kit. (A) Example read/sample profile using non-normalized DNA input and negative controls. (B) Expected coverage/sample based on Illumina flow cell output. All wells containing DNA equal or exceed 0.4X average coverage at 2 billion reads/PW96 library.

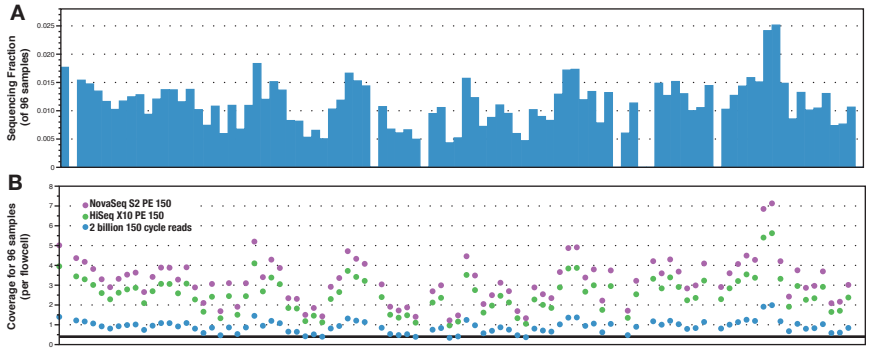
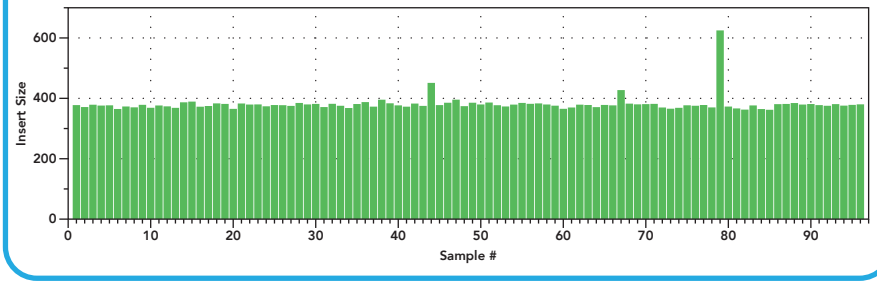


Figure 3. Insert Size Uniformity of 96-plex human low-pass WGS multiplexed library



Conclusions

Low-pass WGS is an effective technical alternative to microarray-based genotyping. The economics of using sequencing in place of microarrays is further increased by 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 human WGS applications, where we show robust and scalable genotype imputation accuracy. 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

Pasaniuc, et al Extremely low-coverage sequencing and imputation increases power for genome-wide association studies; Nature Genetics volume 44, pages 631–635 (2012)