AGBT 2020

plexWell Low Pass 384 Library Preparation Kit: Cost-effective High-throughput Low Pass Genotyping

Jessica Smith, Ariele Hanek and Kerri Stellato

seqWell, Inc. 376 Hale Street, Beverly MA 01915 USA

Introduction

Here, we describe the use of a high-throughput library preparation technology (plexWell™) for 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 plexWell 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 5-fold input range. Our results show the utility of plexWell for routine low-pass WGS genotyping.

Methods

Figure 1. plexWell Workflow

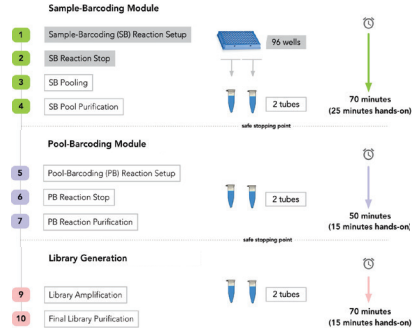
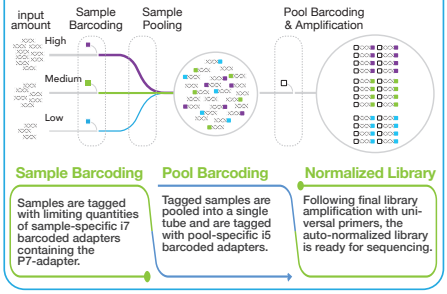


Figure 2. plexWell Normalizing Iterative Barcoding Technology



Design

- **25 individual DNA samples consisting of Coriell Polymorphism Discovery Panel set of 24 and NA12878**
- **8 in triplicate, 7 in duplicate, 10 as single replicate**
- **Sequence to >20M paired reads/sample on NovaSeq**
- **High Level Library Performance metrics**
  - **Read Count CV, Duplication Rate, insert size**
- **Reproducibility between replicates**
- **Coverage across the genome**
- **Imputation results with NA12878**

	7	8	9	10	11	12
A	PD01	PD09	PD17	PD17	PD17	PD01
B	PD02	PD10	PD18	PD18	PD18	PD02
C	PD03	PD11	PD19	PD19	PD19	PD03
D	PD04	PD12	PD20	PD20	PD20	PD04
E	PD05	PD13	PD21	PD21	PD21	PD05
F	PD06	PD14	PD22	PD22	PD22	PD06
G	PD07	PD15	PD23	PD23	PD23	PD07
H	PD08	PD16	NA12878	NA12878	NA12878	PD24

Characteristic Low Pass Library Perfromance

Figure 3. Fractional representation of each i7 index in sequencing results is agnostic of sample input amount.

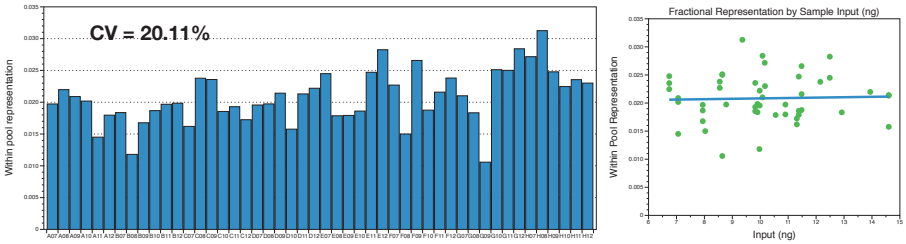


Table 1. Characteristic plexWell™Low Pass 384 Library Performance

Data set	Read Count CV	Average Median Insert	Range Median Insert	Average Mean insert	Range Mean Insert
48 samples	20.1%	322	29	346	29
PD17	18.9%	321	3	342	4
PD18	8.1%	321	3	343	3
PD19	13.2%	322	4	345	3
PD20	16.6%	325	4	348	5
PD21	18.3%	324	2	346	2
PD22	17.7%	324	3	346	4
NA12878	4.9%	317	2	338	2

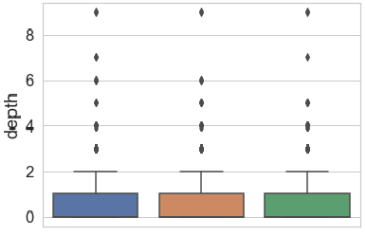
plexWell™Low Pass 384 library preparation kits give reproducible library characteristics (Table 1). Both inter-sample and intra-sample analysis of read count and insert size are consistent. Inserts of 300-350 cluster well on Illumina Sequencers while maximizing the unique data generated from 2x150 paired sequencing, thereby increasing the usable data per dollar of sequencing.

Coverage of the human genome at 10 M and 20 M paired reads is consistent within replicates of the same sample as well as across all samples in the pool (Table 2, figure 4).

Table 2. plexWell™Low Pass 384 Coverage

Data set	% Reads Aligned	% Read Duplication (10M)	Mean Coverage (10M)	% ≥1X	Mean Coverage (20 M)	% ≥1X
48 samples	99.5%	6.8%	0.70	48.2%	1.32	68.4%
PD17	99.4%	7.9%	0.70	47.9%	1.28	67.9%
PD18	99.4%	7.1%	0.70	47.7%	1.29	67.9%
PD19	99.7%	6.3%	0.71	48.4%	1.33	68.7%
PD20	99.6%	6.9%	0.71	48.4%	1.32	68.6%
PD21	99.6%	6.4%	0.71	48.2%	1.33	68.5%
PD22	99.6%	6.9%	0.70	48.1%	1.31	68.2%
NA12878	99.5%	5.9%	0.71	48.8%	1.35	69.3%

Figure 4. Average Coverage across 3 replicates of NA12878



Imputation Results

Table 3. Correctly imputed percentage of SNVs on Chr22 of NA12878 is ~constant from 1-5X coverage

	10M	20M	30M	40M	50M	Rep 1	Rep 2	Rep 3
Homozygous	97.8%	98.3%	97.9%	98.1%	98.3%	97.9%	97.9%	98.0%
Heterozygous	97.3%	97.3%	97.5%	97.6%	97.8%	96.5%	97.0%	97.0%

Figure 5. Percentage of SNVs called with low confidence for NA12878 at 5 depths of coverage

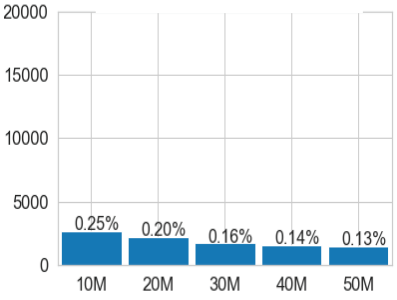
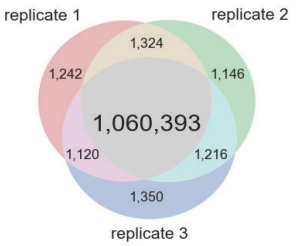


Figure 6. Chromosome 22 SNV concordance for 3 replicates of NA12878 at 1X coverage



A single replicate of NA12878 randomly downsampled to 50M, 40M, 30M, 20M, and 10M read pairs then chromosome 22 was analyzed using Gencove's imputation pipeline. Figure 5 & Table 3 demonstrate the low percentage of low confidence calls and incorrectly called SNVs utilizing plexWell technology with Gencove analysis. 1X coverage is sufficient for imputation as coverage at 5x does not siginicantly improve the low confidence or incorrect calls.

In a second analysis 3 replicates of NA12878 were randomly downsampled to 10M read pairs and run through the same pipeline. Table 3 demonstrates the percentage of called SNVs at 1X coverage is consistent, while Figure 6 shows >99% concordanceof called SNVs between replicates

Conclusions

plexWell™ Low Pass 384 Library Preparation Kit provides a streamlined scalable reproducible solution for Genotyping by sequencing. Nominal 1X coverage across the human genome (10 M clusters at 2x150 paired end sequencing) detects over 99% of the variants detected at 5X coverage thus eliminating the need for sequencing at grater depths of coverage. The data at 1X coverage is consistent between replicates with concordance of >99% of detected SNVs. Commercially available kits support multiplexing of 1152 sample sin a single sequencing run.