Imputation of Loss-of-Function Genotypes Using Low Coverage WGS in a Cattle Population

Warren M. Snelling[1], Tara McDaneld[1], Kristen Kuhn[1], Larry A. Kuehn[1], Jessica Smith[2], Ariele Hanek[2], Joseph Mellor[2]

(1) U.S. Department of Agriculture, Agricultural Research Service, U.S. Meat Animal Research Center, Clay Center, NE 68933, USA

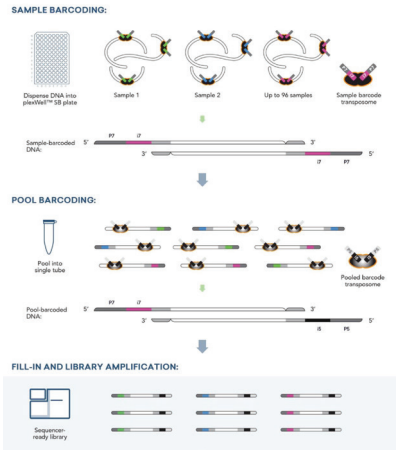
(2) seqWell, Inc. Beverly, MA USA

Introduction

- Obtaining genotype information for cattle breeding stocks helps to identify individuals within breeding populations that carry loss of function (LOF) alleles, which may contribute to genetic load.
- Low-coverage whole genome sequencing (lcWGS) is an increasingly attractive alternative to microarray genotyping, as the cost of sequencing data has dropped over the last decade.
- A key need for lcWGS is being able to generate multiplexed NGS libraries from genomic DNA in which individual sample libraries are sufficiently normalized to produce adequate coverage of each genome for accurate genotype imputation.
- This study evaluates the use of the plexWell Low-Pass 384 Library Preparation kit (seqWell, Inc) for lcWGS of a cattle breeding population.

Methods: Library Preparation and Sequencing

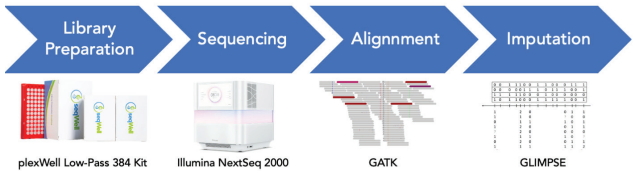
Figure 1: plexWell Low-Pass 384 Library Workflow



- Libraries were prepared with the plexWell Low-Pass 384 Library Prep kit
- Pooled transposase-based workflow allows for processing of large numbers of samples
- Simultaneously fragments & adds Illumina-compatible adapters
- Sequential transposition normalizes for variation in input DNA quantities to reduce need for individual library balancing

Methods: Analysis and Genotype Imputation

Figure 2: Library Preparation and Imputation Workflow



- FASTQ data was aligned with a GATK-based pipeline against the ARS-UCD1.2 reference genome.
- Imputation analysis was performed using GLIMPSE
- The imputation reference panel consisted of 946 previously identified representatives of Bos taurus and Bos indicus derived breeds.

Results

Library and Sequencing Performance

- 96-plex pooled libraries for 1074 samples were prepared and sequenced from DNA isolated from Bos taurus and Bos Indicus derived breeds, and 24 ancestors of the newly sequenced calves.
- Libraries were sequenced on NextSeq2000 with 2x150 read pair configuration.
- Resulting data were mapped using GATK to the ARS-UCD1.2 reference genome and characterized for uniformity of coverage and read-count per sample (Fig 4)

Coverage and Imputation Performance

- After mapping, imputation was performed using GLIMPSE and compared to previously obtained microarray-based genotyping results (Fig 5)
- A positive correlation was observed between coverage and genotyping concordance; at average coverage of 0.5X, concordance was observed to be 98%. At coverage values > 0.3X, the recovery of genotyping calls exceeds 95%.

Loss-of-Function Allele Analysis

- Calves were observed to have differing loads of loss-of-function alleles in selected vs control groups over three years in breeding populations derived from different ancestors (Table 1)

Figure 4: Representative Sequencing Performance of 96-plex plexWell Low-pass

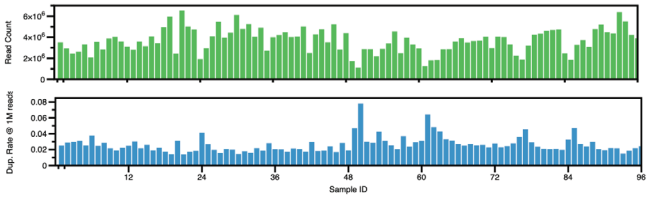


Figure 5: Genotype Imputation Results vs Array-based Genotyping

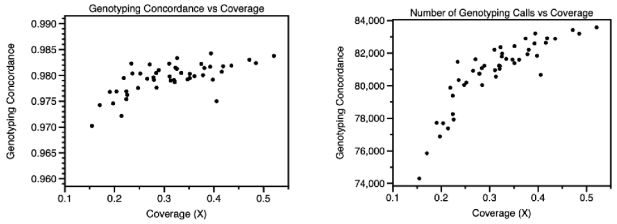


Table 1: Imputed Loss-of-Function Allele Analysis

	MARC I				MARC II				MARC III				Angus			
Birth year	select	sd	control	sd	select	sd	control	sd	select	sd	control	sd	select	sd	control	sd
2020	1182.0	44.7	1193.4	38.6	1152.2	42.4	1178.6	44.8	1108.5	40.5	1150.8	41.3	1084.5	39.6	1096.5	37.2
2021	1178.7	40.5	1189.7	45.3	1133.9	52.9	1177.8	52.1	1089.7	58.9	1141.6	48.6	1061.1	41.0	1087.3	46.9
2022	1173.8	42.3	1194.9	44.8	1124.7	49.5	1170.7	43.8	1098.1	36.1	1127.9	44.7	1045.2	39.4	1061.7	39.2

MARC I - 0.125 Angus, 0.125 Hereford, 0.25 Braunvieh, 0.25 Charolais, 0.25 Limousin
MARC II - 0.25 Angus, 0.25 Hereford, 0.25 Gelbvieh, 0.25 Simmental
MARC III - 0.25 Angus, 0.25 Hereford, 0.25 Red Poll, 0.25 Pinzgauer

Summary and Conclusions

- The results of this study support the routine use of lcWGS as a viable and accurate alternative to microarray-based genotyping.
- The plexWell Low-Pass 384 Kit in combination with standard sequencing instrumentation and open-source software tools can be used to generated lcWGS sequencing libraries that have sufficient genomic complexity for lcWGS-based genotyping applications.