Ultra High Throughput, High Quality Genotypes using the Applied Biosystems™ Eureka Genotyping Platform

Victor Missirian[1], John D. Curry[1], Mohini A. Patil[1], Ali Pirani[1]
[1]Thermo Fisher Scientific, 3450 Central Expressway, Santa Clara, California, USA, 95051

INTRODUCTION

Our Applied Biosystems™ Eureka™ Genotyping Solution is a low cost, ultra-high throughput (20k samples/week) targeted genotyping by sequencing platform that supports the detection of tens to thousands of genetic markers (SNPs as well as complex insertions/deletions). It has been successfully used for a variety of applications (including parentage, sex validation, genomic and trait evaluations) both in crops and animals.

The Eureka™ assay is highly resilient to variability in sample composition, including the presence of protein or other cellular debris. The robust nature of the assay enables consistent delivery of high quality genotypes across a wide range of sample types.

We demonstrate that the Eureka™ Genotyping platform consistently delivers high performance (high sample pass rate, call rate, and concordance to known genotypes) across multiple sample types including extracted DNA, dried blood spot lysates, and whole cell lysates from hair and ear punches. Lysates are used directly in the Eureka™ assay, removing the need for costly DNA purification steps. Eureka Genotyping Solution also enables genotypes in under 48 hours and high throughput processing to meet quick turnaround demands of the breeding industry.

METHODS

Eureka™ genotyping assay

Eureka Genotyping Solution utilizes NGS to enable genotyping of thousands of DNA samples for tens to thousands of markers. The Eureka genotyping assay is a ligation-dependent PCR reaction that uses interrogation site bar codes contained within the ligation probes as well as sample index bar codes added during the amplification step. NGS libraries can be created for thousands of DNA samples within 24 hours. Short-cycle sequence data is generated from the prepared libraries, and software is used to tabulate the number of reads that contain each combination of sample, locus, and allele bar code (as appropriate). The genotype of each sample for each locus is inferred from statistical analysis of the tabulated reads.

Genotyping analysis

Sample processing returns Eureka™ NGS read counts for each sample, at each genetic marker. Counts are normalized, scaled, and then transformed to expand the central call region according to a stretch constant K. Genotypes are then called in a cluster-based Bayesian framework (BRLMM-P) that adapts pre-positioned genotype cluster locations called “priors” to the sample data and computes three posterior cluster locations.

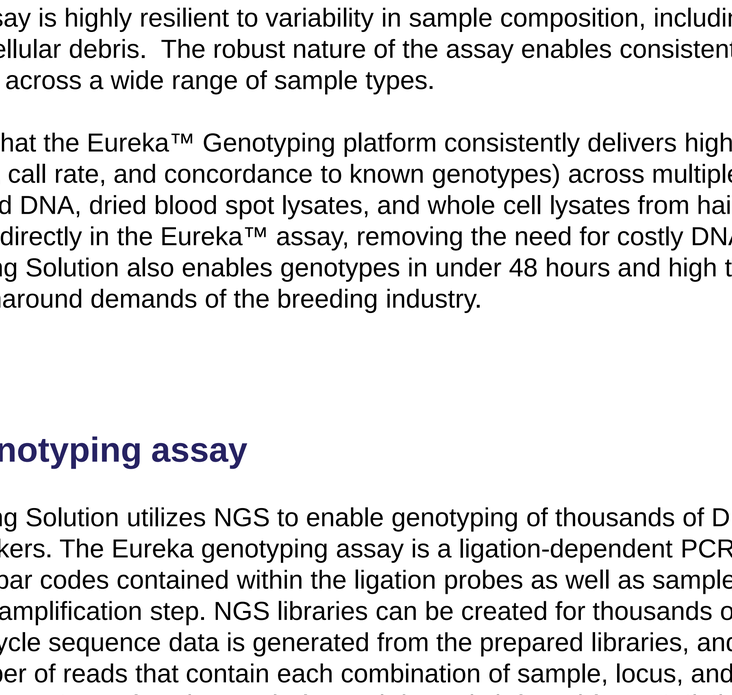

Figure 1. Example of a SNP cluster plot.
Eureka™ panels contain tens to thousands of markers. Performance of each can be visualized with a cluster plot. Each sample is represented by a single point, whose coordinates reflect its (scaled, normalized, transformed) read counts. Prior location (dashed oval) and observed samples (points) are then combined to get a posterior (solid oval), which is used to call genotypes.

Testing

Eureka™ genotyping was evaluated on dried blood spots, whole cell lysates (hair follicles and ear tags), and both purified and crude extracted DNA. For each sample type, we individually genotyped a single set of samples. Performance on each set of samples was measured in four ways: sample pass rate, call rate, concordance to known genotypes, and visual evaluation of cluster plots.

Sample pass rate: Samples are considered not passing if they have a mean call rate (see below) of less than 90%. Reported values of the next three measures were computed over passing samples.

Call rate: Genotyping uses a cluster-based Bayesian framework to compute a most likely genotype (AA, AB, or BB) and a confidence score that reflects the probability of that genotype. A (marker, sample) pair is called only if the confidence score of the most likely genotype exceeds a specific threshold, otherwise it is no-called (assigned no genotype). Call rate is the percentage of pairs that are called.

Concordance: Eureka™ genotyping calls (AA, AB, or BB) can be compared to reference calls from a second source. Concordance is the percent of (marker, sample) pairs with calls from both Eureka™ and reference for which the Eureka™ call matches the reference call.

Cluster plots: We visually evaluated cluster plots, taking into account clarity and cluster separation.

RESULTS

Performance

One way to measure performance is by computing sample pass rate, call rate, and concordance, as defined in the previous section. The table below displays sample pass rate for each sample type as well as mean call rate and concordance across all markers. These measures are consistently high across sample types.

Sample Type	Sample pass rate	Call rate (mean)	Concordance (mean)
Extracted DNA, Purified	100.0%	99.8%	99.8%
Extracted DNA, Crude	98.4%	98.9%	99.2%
Dried Blood Spot	100.0%	98.1%	NA
Ear Punch	97.0%	99.2%	NA
Hair Follicle	98.9%	99.5%	NA

Table 1. Sample pass rate, call rate, and concordance.
Performance of a Eureka™ panel can be measured in terms of sample pass rate, call rate, and concordance. We have calculated these measures on five different sample types.

Visual Evaluation

A second way to measure performance is the visual evaluation of cluster plots. A high performance SNP is expected to have a clear and well separated cluster pattern. The figure below displays cluster plots of one to three example markers for each tested sample type. Cluster patterns are clear and well resolved across all sample types.

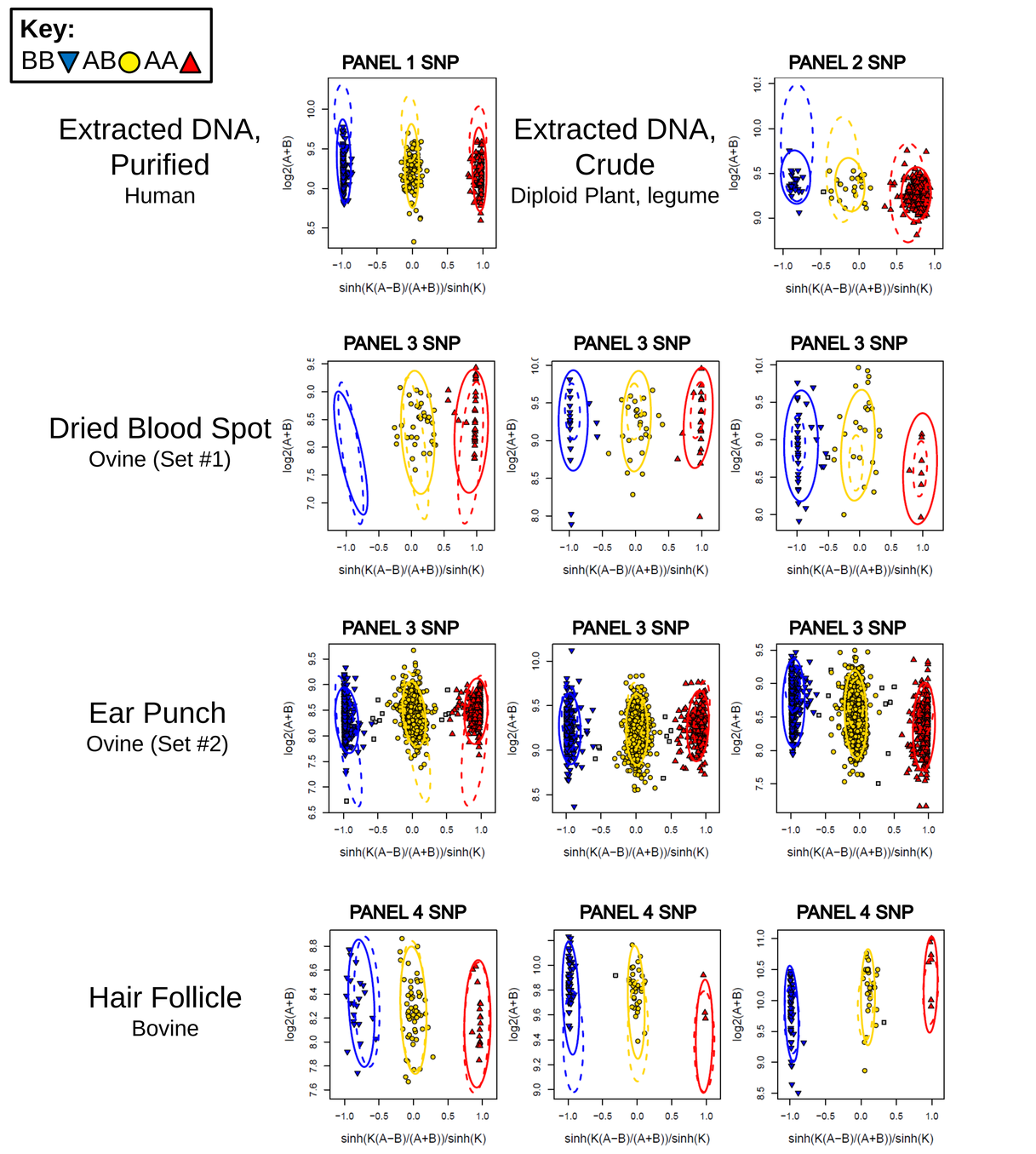

Figure 2. Cluster plots of example markers.
Performance of a Eureka™ sample set can be measured by the clarity and separation of cluster plots for each marker. We have cluster plots on one to three example markers for each tested sample type. Colors/shapes of points correspond to the three genotyping calls (AA, AB, or BB).

CONCLUSIONS

Eureka™ genotyping has consistently high performance across a wide range of sample types, as shown by sample pass rate, call rate, and concordance as well as by clear and well separated cluster plots. Eureka™ genotyping provides a high quality, versatile platform for ultra-high throughput genotyping.