Versatile 96-SNP Genotyping Panel Enables DNA Fingerprint and Sample Integrity Assessments

Honrado Lopez, Jian Qin, Gang Sun and Christopher Kubu

Introduction

Biorepositories provide access to high-quality, curated samples for basic and clinical research purposes. Sample degradation, misidentification and contamination are significant risks to the integrity of banked samples. Distribution of such samples can waste time and laboratory resources and negatively impact the integrity of research studies.

Standard procedures for sample traceability and quality assessment have been employed by biorepositories for many years, including but not limited to barcode labeling, LIMS tracking and DNA quantification. Implementing a DNA fingerprinting method in the biorepository workflow provides more informative quality assessment tools and a direct assessment of sample molecular identity.

The Advanta™ Sample ID Genotyping Panel is a 96-SNP (single-nucleotide polymorphism) assay that generates a sample-specific genetic fingerprint and supports multiple quality assessments of research specimens throughout the sample journey. Developed for use with the Biomark™ HD system and based on Fluidigm microfluidics technology, the workflow uses integrated fluidic circuits (IFCs) to precisely combine multiple reactions at nanoliter volumes. In this poster, we demonstrate the utility of the Advanta Sample ID Genotyping Panel as a sample identity and quality assessment tool.

Advanta Sample ID Genotyping Panel content and workflow

The Advanta Sample ID Genotyping Panel facilitates DNA sample quality control (QC) and tracking of human samples through the use of 96 SNPs. 80 SNPs are located in exonic regions. The workflow uses Fluidigm microfluidics technology to complete profiling of up to 96 samples per run in four hours.

Quality (10)
- Located in regions susceptible to DNA degradation
- Call rates correlate with sample quality.

Gender (6)
- 3 X and 3 Y chromosome targets for gender ID

Population (40)
- Exonic SNPs located in housekeeping genes
- Population-specific SNPs* that enable ancestry prediction

Highly polymorphic (40)
- Exonic SNPs located in housekeeping genes
- Common SNPs** that provide high discriminatory power to differentiate even closely related individuals
- Duplicate genotype probability is 1 in 1.09 x 10^{17}

1 Prepare samples and assays.

2 Transfer samples and assays to the IFC.

3 Process the IFC using the automated Juno™ system.

4 Perform real-time PCR and data acquisition using Biomark HD.

5 Analyze results.

Figure 1. The Advanta Sample ID workflow, from sample preparation through data analysis. 96 samples and the 96-assay panel are loaded on an IFC and combined automatically in the Juno system. The IFC is subsequently moved to the Biomark HD system for thermal cycling and imaging. The genotyping run is then analyzed using Fluidigm SNP Genotyping Analysis software.

Advanta Sample ID Genotyping Panel provides an accurate, reproducible fingerprint

Call accuracy of Advanta Sample ID Genotyping Panel

Sample Name	Precision within and between IFC Runs (%)	Published Variant Calls from GIAB for Panel SNPs (out of 96)	Concordant Calls to GIAB	Call Accuracy (%)
NA12878	100	92	92	100
NA24143	100	87	87	100
NA24149	100	83	83	100
NA24385	100	83	83	100
NA24631	100	87	87	100

Figure 5a. Five well-characterized samples from Genome in a Bottle (GIAB) were run on two 96.96 Dynamic Array IFCs in triplicate and used to determine call accuracy to known variant calls. Call accuracy was determined by comparing IFC genotype calls to sample-specific variant information provided by GIAB. However, no sample had high-confidence variant calls for all 96 assays. High-confidence variant calls were determined by the GIAB Consortium by utilizing sequencing data generated by multiple technologies to generate variant calls and regions or use in benchmarking and validating variant calling pipelines. For example, NA12878 has high-confidence variant call information for 92 of the 96 SNPs in the panel. All 96 SNPs are represented across the five GIAB samples. All samples and replicates displayed 100% call accuracy when comparing the IFC genotype calls to the benchmark GIAB calls.

Reproducibility of Advanta Sample ID Genotyping Panel

IFC	Total Calls*	No Calls	Call Rate (%)
1	8,277	11	99.9
2	8,277	8	99.9

* Negative controls and Y chromosome targets were ignored.

Figure 5b. Data from both IFC runs were compared to test reproducibility. A 96.96 Dynamic Array can provide up to 9,216 datapoints. Removing the negative controls and Y chromosome assays from the analysis provides a more accurate assessment of call rate.

Results

SNP profiling uniquely identifies individuals

96-sample call map view of genotyping run

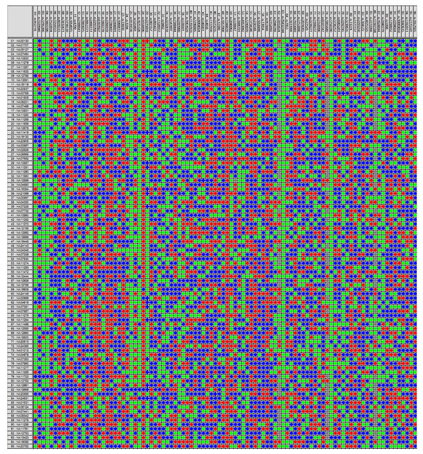

Figure 2a. 95 samples were run on a 96.96 Dynamic Array™ IFC and scanned on the Biomark HD system. Samples included one multigeneration family with 17 members. All samples generated a unique SNP profile (no duplicate genotype profiles). Shown here is a call map viewer of the 96.96 Dynamic Array run. Samples are represented in rows and individual SNP calls in columns.

CEPH/Utah Pedigree 1463

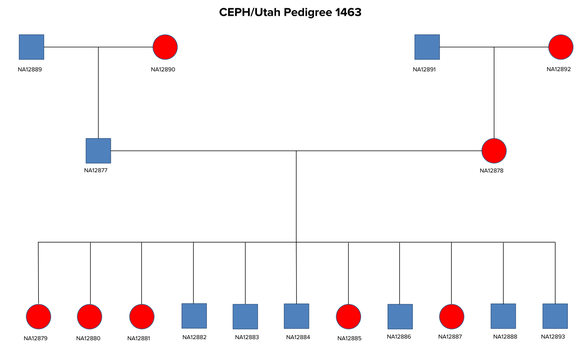

Figure 2b. Three generations of two families were analyzed, including 11 siblings. In the test detailed in Figure 2a, unique SNP profiles were established across 17 individuals of a three-generation family.

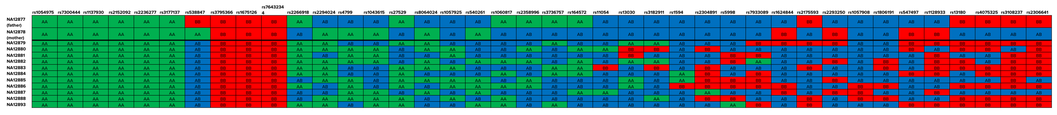

Figure 2c. Unique profiles with Advanta Sample ID Genotyping Panel differentiate parents and offspring. The utilization of 40 highly polymorphic SNPs [~0.5 minor allele frequency (MAF)] provides a high power of discrimination among even closely related individuals. According to the Hardy-Weinberg principle, the probability that two unrelated individuals will have the same genotype at a given biallelic locus with 0.5 MAF is ~1 in 3. Utilizing 40 SNPs with ~0.5 MAF, the probability for a duplicate SNP profile is 1 in 1.09 x 10^{17} individuals. For reference, the current estimated world population is 7.7 x 10^{9} individuals. This figure shows the unique SNP profiles generated for 13 individuals from the CEPH pedigree (father, mother and offspring) with 40 highly polymorphic SNPs (sample in rows and SNP call in columns). All 11 siblings shared the same alleles at 10 loci. In sibling-to-sibling comparison analyses, there were up to 32 shared loci (NA12884 vs. NA12893) with an average of 25 shared loci.

Sex chromosome analysis detects ≥5% contamination

Genetic sex discrepancy can signal a contamination event.

Sample	Gender Call by Software	Sample	Gender Call by Software
F1	Female	M1	Male
F1-1% M3	Female	M1-1% F3	Male
F1-5% M3	No call	M1-5% F3	Klinefelter male
F1-10% M3	No call/Klinefelter male	M1-10% F3	Klinefelter male
F2	Female	M2	Male
F2-1% M3	Female	M2-1% F3	Male
F2-5% M3	No call	M2-5% F3	Klinefelter male
F2-10% M3	Klinefelter male	M2-10% F3	Klinefelter male

Figure 3a. Two male (M1 & M2) and two female (F1 & F2) samples were tested on a 96.96 Dynamic Array IFC and Juno 96.96 Genotyping IFC using the Juno and Biomark HD workflow. Artificial mixes were created with 1%, 5% and 10% contamination levels with opposite gender samples. Mixes were run in triplicate with a negative control. The auto-assigned genetic sex from the Fluidigm SNP Genotyping Analysis software shows that as little as 5% contamination could be detected when comparing genetic sex calls with original sample, for example M1 vs. M1-5%F3.

Scatter plot view of contamination event: F1 vs. F1-5% M3

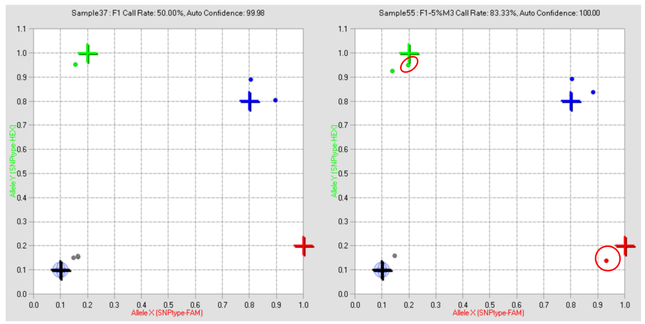

Figure 3b. Left scatter plot is the control sample (F1) and shows three genotype calls for the chromosome X targets (one homozygous AA and two heterozygous AB) and all three chromosome Y targets located near the origin, displaying no amplification. The right scatter plot shows the effect of 5% M3 contamination. Highlighted in red circles are genotype calls for two chromosome Y targets. In normal XX female samples, all results for chromosome Y assays should be No Call because there should be no amplification of those targets.

SNP genotyping can identify low-quality samples

SNP call rate at incubation time points

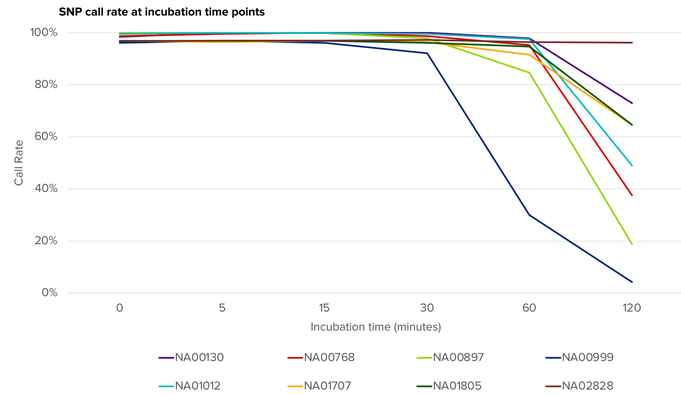

Figure 4a. Eight DNA samples were incubated over time at 95 °C in Tris pH 7 to induce DNA damage via heat and acid depurination. SNP call rates were near 100% in high-quality samples and generally decreased linearly with extended incubation time. Sample quality after heat-induced degradation also correlated with qPCR analysis (Figure 6b) of DNA quality (200 bp/60 bp concentration ratio). Concordance rate is a measure of the percentage of SNPs that are measured as identical, where in this case the control would be measured vs. time point samples. Although sample quality steadily decreased throughout time point study, SNP call rate does not drop below 85% for most samples until after the 60-minute mark. This shows that SNP genotyping is a robust method that can identify lower-quality samples.

qPCR concentration analysis of samples

Samples		Control	95 °C Incubation				
			5 min	15 min	30 min	60 min	120 min
200/60 ratio	Mean	0.92	0.66	0.33	0.10	0.02	0.00
	SD	0.11	0.11	0.14	0.09	0.03	0.00

Figure 4b. Quantities for 200 bp and 60 bp RNaseP assays were measured using a standard curve generated from a reference sample of known concentration. The 60 bp and 200 bp concentrations of each sample are used to calculate its 200/60 ratio, which is related to a sample's quality. Sample quality decreased linearly with extended incubation time.

Conclusion

The Advanta Sample ID Genotyping Panel generates a sample-specific DNA fingerprint that can be differentiated from even closely related samples. It can identify degraded samples or those that have been mixed up or contaminated. Early detection of poor-quality, contaminated or incorrectly curated samples can improve research quality and reduce costs of superfluous testing of poor-quality samples.

Fluidigm Corporation
7000 Shoreline Court, Suite 100
South San Francisco, CA 94080 USA
+1 650 266 6000 • Toll-free in the US and Canada: 866 359 4354
fluidigm.com