A novel NGS-based sample tracking system for robust sample chain of custody and quality control in fresh frozen and FFPE tumor samples

Katherine Milligan, Yu Qiu, Andrea O'Hara, Carrie Ziemniak*, David Corney, Haythem Latif

*Presenting Author

GENEWIZ from Azenta Life Sciences, South Plainfield, NJ 07080

Abstract

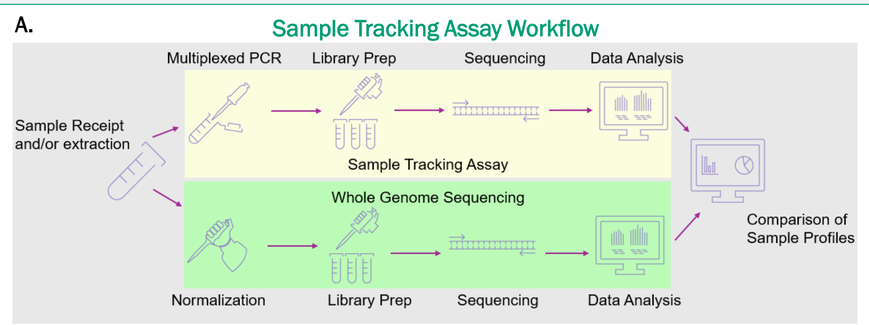
When processing large cohorts of samples, sample identity and functional quality are key to ensure the derived data is accurate and attributed to the appropriate sample. This is especially critical when working with precious and degraded material such as formalin-fixed paraffin-embedded (FFPE) samples. A potential sample swap or poor sample quality may result in incorrect data or delays due to re-processing. While traditional methods such as qPCR can be used for quality control, they may not be an effective predictor of sample success in Next-Generation Sequencing (NGS) based assays due to differential susceptibility to inhibitors and interfering substances. Performing a functional quality control assessment using the same methodology for data generation allows for a more accurate prediction of performance as it would be similarly affected by interfering substances or sample degradation. We have created a scalable, cost-effective sample tracking and functional quality control assay that may be assessed prior to performing more expensive downstream sequencing assays. This assay was designed to confirm consistency in sample identity during processing and confirm identity when multiple samples are submitted from the same subject.

This amplicon panel targets autosomal SNPs with varying minor allele frequencies in different ethnic populations, limiting a bias towards variation in European ethnic populations. The average minor allele frequency for the panel developed for the general population is 0.45, more specifically 0.46 for European background, 0.44 for African background and 0.46 for Asian background based on data from the Allele Frequency Aggregator (ALFA) database. Most amplicons were designed to be less than 150 base pairs to increase the likelihood of success for samples with degraded DNA. Targets were included on the X and Y chromosome for sex determination.

Tumor-normal paired samples were processed on the sample tracking assay, targeting regions throughout the exome to allow tracking of sample identity. When comparing "normal" samples extracted from blood to their paired FFPE tumor tissue, known matches had an average of 98.0% concordance, while known mismatches had an average of 40.4% concordance, with the highest known mismatch showing 61.9% concordance. Additionally, when comparing well-characterized samples to their corresponding data from the 1000 Genomes Project, samples had 99.4% concordance. This exemplifies a robust sample tracking assay that can ensure sample identity is correct. A strong quality control assay enables NGS laboratories to identify poor sample quality or degradation, allowing for re-extraction or new sample collection before processing on a more expensive assay. This newly developed sample tracking assay ensures both sample integrity and identity, leading to an overall increase in data quality for cancer research studies.

ASSAY	ONBOARDING COST	PER SAMPLE COST	FUNCTIONAL ASSAY FOR NGS	ABILITY TO DETECT CONTAMINATION	ABILITY TO ID SAMPLES	FLEXIBILITY TO UPDATE PANEL
NGS	N/A	$	✓✓✓	✓✓✓	✓✓✓	✓✓✓
Sanger Sequencing	$$	$$$	✓✓	✓	✓✓✓	✓✓
Fragment Analysis	$$	$$$	✓✓	✓✓	✓✓✓	✓✓
High Throughput qPCR	$$$	$$	✓	✓	✓✓✓	✓✓
Commercial Panel	$$$	$$	✓	✓	✓✓✓	X

Figure 1. There are different considerations when choosing an appropriate sample identification panel for a high throughput sequencing laboratory, in addition to its ability to identify samples. The cost of onboarding an assay, such as equipment or infrastructure, will vary depending on what equipment is in the laboratory and the cost to run it will vary based on the kit used and the design. A functional assessment of a sample to perform on a different assay is best done by a scaled down version of the same assay. NGS provides a high level of multiplexing and quantitative data that can allow for thorough contamination detection. There is also a high degree of flexibility with an amplicon-based NGS panel to update the panel to include additional targets.



B.

PANEL CHARACTERISTICS	
Number of Autosomal SNPs Targeted	42
Number of Sex Chromosome Regions Targeted	14
Chromosomes Included	21
Average Overall SNP MAF	0.451
Average European SNP MAF	0.459
Average African SNP MAF	0.436
Average Asian SNP MAF	0.463

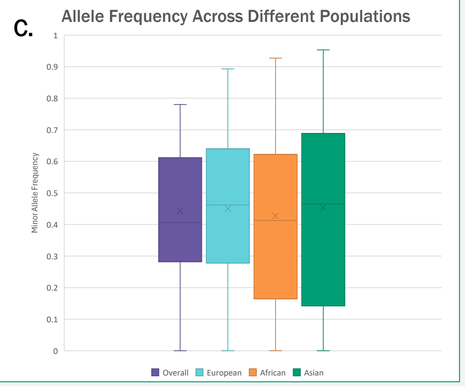


Figure 2. Sample tracking workflow and panel details. (A) A schematic of the NGS amplicon-based sample tracking workflow. Samples are received and DNA is extracted. The DNA is tested on the sample tracking workflow, including a multiplexed PCR that targets the amplicons, library prep, sequencing and data analysis to generate a variant file. Concurrently, samples are processed on the workflow of interest, such as WGS or WES. Finally, the profiles are compared between the assays. (B) The characteristics of the panel include SNPs from both the autosomal and sex chromosomes, targeting most of them to avoid potential linkage disequilibrium. The average MAF is also identified for the overall population as well as different ethnic groups. (C) The distribution of the MAF across different ethnic groups is displayed.

METRIC	CORIELL SAMPLES: SAMPLE TRACKING VS 1000 GENOMES PUBLIC DATA	EXTRACTED DONOR DNA: SAMPLE TRACKING VS LP-WGS
Accuracy	99.40%	97.04%
Sensitivity	99.40%	97.01%
Specificity	99.39%	97.06%
Positive Predictive Value	99.40%	96.81%
Negative Predictive Value	99.39%	97.24%

Figure 3. Summary of analytical performance of the sample tracking assay. All metrics were obtained by comparing SNV calls obtained from the assay to 1000 genome data or to Azenta's Low-Pass Whole Genome Sequencing (LP-WGS) data.

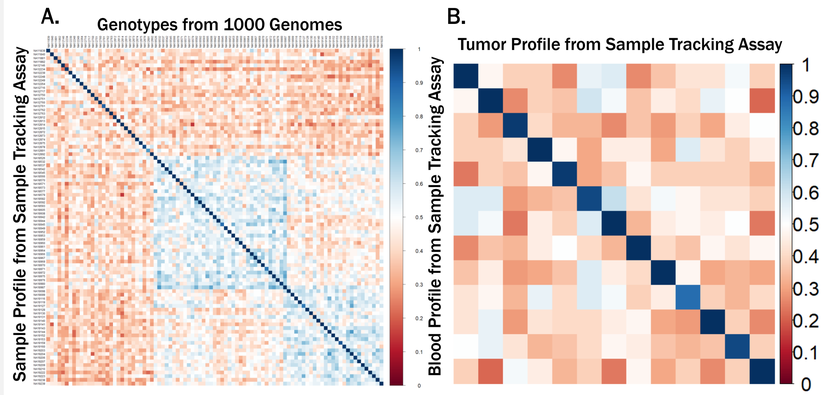


Figure 4. Graphic display of concordance between known matching samples and all other samples in the cohort. The color of the square is linked to the percent concordance of the SNPs between the associated data set. A: Concordance of Coriell samples to 1000 Genomes data set. B: Concordance of sample profiles from DNA derived from matched tumor and blood.

Conclusions

- GENEWIZ from Azenta Life Sciences has created a robust, cost-effective sample tracking assay that allows functional quality control, as well as sample identity tracking.
- The sample tracking panel includes varied SNPs across the majority of chromosomes with minor allele frequencies that are varied between different ethnic groups.
- The sample tracking assay is robust with consistently high accuracy and precision for both blood and tumor samples.
- The sample tracking assay provides profiles that are unique to the sample of interest to allow for identification of swapped samples.

Acknowledgements: We are grateful to Tina Cheung and Mila Paiva for library preparation support and Alex Guzzo, Drasti Patel, and Rupal Jain for sequencing assistance.