Determination of Validated Kintelligence Thresholds for Application to Forensic Genetic Genealogy

Michelle A. Peck, Christina M. Neal, Stephen D. Turner, Jessica L. Bouchet, Erin M. Gorden, Mary C. Heaton, David A. Russell, V.P. Nagraj, Alexander F. Koeppel, Carmen R. Reedy.

Background

The power of high-density SNP profiles to infer distant kinship relationships to generate investigative leads in criminal casework and human identification efforts is rapidly advancing the forensic genomics field. Microarray analysis and whole genome sequencing are the most common approaches to generate millions of SNPs spanning the genome for Forensic Genetic Genealogy (FGG). However, microarray analysis with degraded and low input samples is particularly difficult and interrogating the whole genome brings up privacy concerns. To address these limitations, Verogen developed the ForenSeq® Kintelligence kit that targets 10,230 SNPs that span the genome, maintaining compatibility with direct-to-consumer kits that largely populate the publicly available databases, while excluding medically relevant SNPs. The targeted sequencing method is well-suited for low quantity samples, with the added feature of being able to add up to 25 µL of extract. Further, the kit is designed for degraded samples with an average amplicon size <150 bp. Kintelligence profiles are compatible with GEDmatch PRO, the law enforcement portal that can search GEDmatch data.

Objective

Our goal was to validate the Kintelligence kit according to the SWGDAM guidelines. We focused on optimizing the procedure for low input samples and evaluated nonprobative degraded samples. To establish analysis thresholds, we evaluated call rate, concordance, and matching within GEDmatch PRO.

Methods

We followed the ForenSeq Kintelligence Kit Reference Guide with some exceptions.

Library Preparation: At the purify libraries step of the procedure, the final resuspension volume of RSB was reduced in half to 27 µL, with a transfer supernatant volume of 25 µL. All samples except for those in the nonprobative study were processed in duplicate.

Sequencing: A molarity-based pooling approach was used instead of a volume-based approach. High input samples (0.25 ng and 1.0 ng) were sequenced in 4-plex sample pools, in addition to negative controls. Low input samples (0.1 ng and 0.05 ng) were sequenced in 2-plex sample pools.

Analysis: The sensitivity samples were analyzed with both the manufacturer's analysis method (20X threshold) and a modified method (10X threshold).

Conclusions

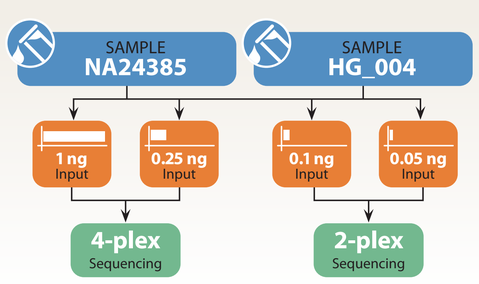
The ForenSeq Kintelligence kit is valid for generating accurate SNP profiles that are compatible with distant kinship matching in the GEDmatch PRO database. Sensitivity was demonstrated down to 50 pg of DNA with high quality samples. Utilizing an analysis threshold of 10X generated the best results. High performance was observed with nonprobative samples, notably with bone samples that included low inputs and degraded samples, from a range of conditions. Profile heterozygosity is useful to assess the quality of a sample, including the identification of contamination. Evaluating the sensitivity and nonprobative samples in GEDmatch PRO resulted in consistent results across replicates, DNA input, and sample type. This validated method is ideally suited for low input and degraded sample types eligible for FGG.

Acknowledgements:

Thank you to Rachel Houston and Jennifer Snedeker from Sam Houston State University for providing the bone extracts.

STUDY DESIGN | RESULTS | DISCUSSION

Sensitivity



DNA Input (ng)	Analysis Threshold	Call Rate Mean	Heterozygosity Mean	Concordance Rate Mean
0.05	10X	0.962	0.409	0.944
0.05	20X	0.934	0.383	0.921
0.1	10X	0.981	0.441	0.976
0.1	20X	0.966	0.428	0.965
0.25	10X	0.989	0.453	0.989
0.25	20X	0.978	0.444	0.982
1	10X	0.987	0.455	0.992
1	20X	0.975	0.449	0.988

Table 1: Sensitivity study call rate, heterozygosity, and concordance results per DNA input and analysis threshold (N=4 per condition). The mean call rate was >95% for all inputs at both analysis thresholds, except for the replicates of 0.05 ng at the 20X threshold. Concordance was evaluated against known GIAB profiles and was >96% for inputs > 0.1 ng. At all input levels, concordance was greater with the 10X threshold. At Kintelligence sites, the 1000 Genomes data set had a heterozygosity range of 0.316 – 0.546 (3SD). All samples fell within this heterozygosity range.

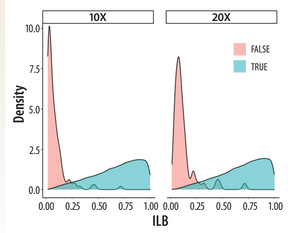
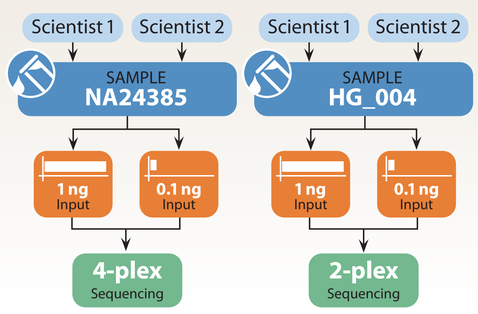


Figure 1: Intralocus balance (ILB) density of heterozygous (HT) calls split by concordance and analysis method. (*The count of true (N>62,000) and false calls (N<110) are quite imbalanced). Applying an ILB threshold based on the best F1 score, minimally impacted concordance at the cost of removing concordant calls. An ILB threshold was therefore not applied.

The sensitivity study demonstrated high performance across inputs down to 0.05 ng. Increasing the multiplexing level to four samples per pool for the high inputs maintained high success rate. Heterozygosity of all samples fell within the expected range. The evaluated metrics of call rate, concordance, and heterozygosity were always higher with the 10X analysis method indicating greater detection of accurate data. However, the rates of false HT calls were greater with 10X (N=109) than the 20X (N=53). While the false HT calls trended towards low ILB values, the occurrence of these false calls was so low there was minimal benefit to applying an ILB threshold. In all other studies, the 10X threshold was applied. Heterozygosity was used as a quality check of the data.

Precision and Accuracy



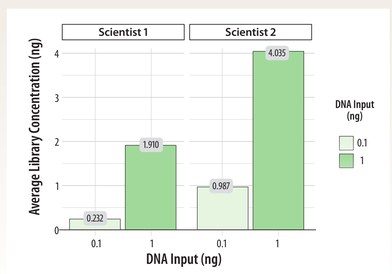


Figure 2: Average library concentration per scientist at 1 ng and 0.1 ng of DNA input were distinct in the Precision and Accuracy study.

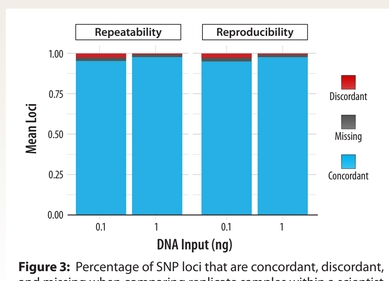
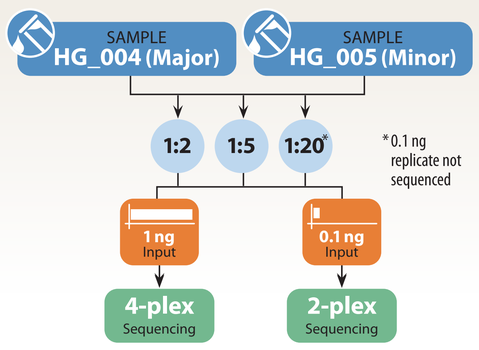


Figure 3: Percentage of SNP loci that are concordant, discordant, and missing when comparing replicate samples within a scientist (repeatability) and between scientists (reproducibility). All conditions resulted in high concordance.

Accuracy of the profiles was established in the Sensitivity Study, comparing Scientist 1 replicates to the known GIAB profiles. Precision was demonstrated with high rates of repeatability and reproducibility for both sets of replicates. The higher rate of discordant calls in the 0.1 ng samples was attributed to drop-out occurring in both samples. While the library concentrations were distinct between scientists, the high performance of both sets of profiles indicate that profile success correlates to the original DNA input.

Mixtures



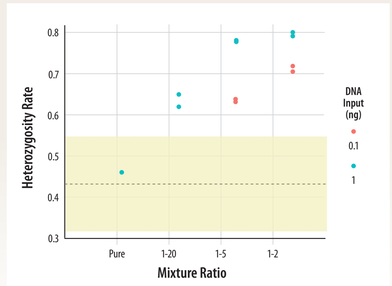


Figure 4: Heterozygosity rates of the mixture samples compared to the single source HG_004 sample, with the range from the 1000 Genomes samples highlighted. Extremely high heterozygosity values were observed for all mixture ratios and inputs.

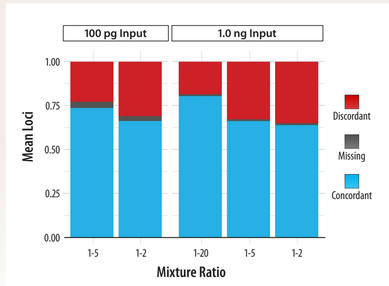
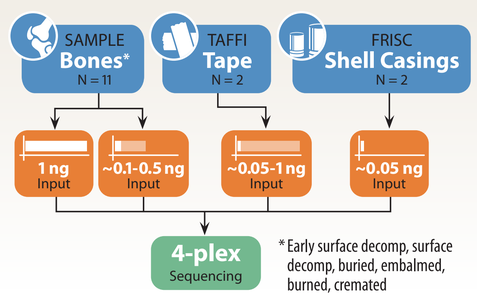


Figure 5: Percentage of SNP loci that are concordant, discordant, and missing when comparing the mixture samples to the single source HG_004 sample. High rates of discordance indicate detection of the minor contributor at all mixture ratios and inputs.

The mixture study was performed to establish detection of a mixture or contamination event, with no intention to deconvolute. All samples were flagged as mixtures in the UAS. High levels of heterozygous calls characterized mixtures, spanning the mixture ratios of 1:2, 1:5, and 1:20 at both 1.0 ng and 0.1 ng. The detection of the minor contributor corresponded to 0.05 ng in the 1:20 sample at 1.0 ng and 0.02 ng in the 1:5 sample at 0.1 ng, consistent with the results of the sensitivity study.

Nonprobative



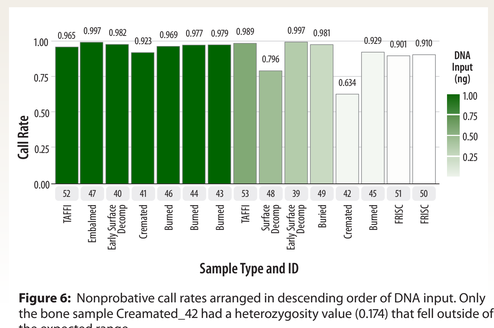
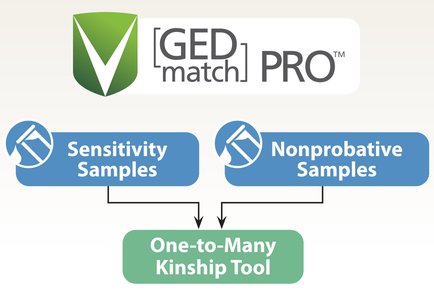


Figure 6: Nonprobative call rates arranged in descending order of DNA input. Only the bone sample Creamated_42 had a heterozygosity value (0.174) that fell outside of the expected range.

Figure 7: Purple channel 220bp – 400 bp of the Investigator 24plex QS profile for the bone sample Burned_46. This sample had a DI index of 26.1 and showed the ski slope effect, clearly indicating degradation. The Kintelligence profile had a 96.9% call rate.

High call rates were observed with the majority of nonprobative samples. The majority of bones at 1 ng resulted in greater than 95% call rate, including a highly degraded sample. The advantage of being able to add 25 µL was demonstrated in an extract at 0.042 ng/µL being able to reach the 1 ng target and resulting in a 96.9% call rate. High performance was also observed with low inputs. Call rate remained above 90% for a burned bone when tested at 0.1 ng input and was 90.7% concordant with the sample processed at 1 ng. Two FRISC samples were tested at approximately 50 pg and resulted in greater than 90% call rate.

GEDmatch Pro



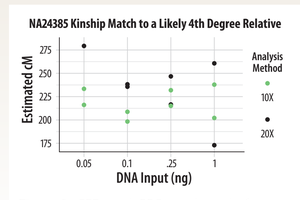


Figure 8: GEDmatch PRO match results for NA24385 samples from the sensitivity study with a likely 4th degree relative (based on One-to-Many Segment matching). One of the 0.05 ng replicates at 20X did not match to this individual. The two highest cM values were flagged as 4th degree while the rest were 5th degree. All matches were high confidence except for the two lowest cM values.

Donor	Sample	Input (ng)	Match	Shared cM	Degree	High Confidence
Bone 6	Burned_44	1	Alias X	1378.0	2nd	Yes
Bone 6	Burned_46	1	Alias X	1460.8	2nd	Yes
Bone 6	Burned_45	0.1	Alias X	1152.3	2nd	Yes
Bone 1	Early Surface Decomp_39	0.36	Alias Y	127.2	5th	No
Bone 1	Early Surface Decomp_40	1	Alias Y	129.3	5th	No
NA24385	TAFFI_52	1	Alias Z	253.9	5th	Yes
NA24385	TAFFI_52	0.57	Alias Z	242.7	5th	Yes

Table 2: GEDmatch PRO results for replicate nonprobative samples. Consistent results were observed with a range of inputs and sample types. For the TAFFI samples, the Alias Z match is the same individual as in Figure 8.

As proof of principle the sensitivity samples at 10X and 20X, along with the nonprobative samples at 10X were uploaded to GEDmatch PRO. Overall, reproducible match results were generated at all inputs for the sensitivity samples. The TAFFI samples (donor NA24385) resulted in consistent match results to a distant relative compared to the sensitivity study. The same high confidence second degree match resulted with Donor Bone 6 for two replicates at 1.0 ng and one replicate at 0.1 ng.

Michelle A. Peck
Signature Science, LLC
1670 Discovery Drive
Charlottesville, VA 22911
mpeck@signaturescience.com

signature® science LLC

www.signaturescience.com

Scan for more information about our FGG services