Comparison of ForenSeq® Kintelligence and Whole Genome Sequencing in Searching for Relatives in GEDmatch

S.M. Radecke[1], J. Antunes[1], G. Padmabandu[1], K.M. Stephens[1]

[1]Verogen, A QIAGEN Company, San Diego, CA USA

Introduction

- Whole genome sequencing is a popular alternative to microarrays for highly degraded and/or low input samples for forensic genetic genealogy[1]
- ForenSeq Kintelligence 10,230 SNP panel was developed concurrently with GEDmatch's One-To-Many Kinship Tool that maximizes kinship SNPs to identify up to 5th degree relationships[2-3]
- ForenSeq Kintelligence also performed well on degraded samples such as DNA from skeletal remains

Materials and Methods

- WGS and ForenSeq Kintelligence libraries were generated with V024 Samples: 1ng intact, 1ng with DI 6, 100pg intact, or 100pg with DI 6 (Table 1)
- Sequencing was done on the NovaSeq at 2x151 with a target coverage of 30x (WGS) or 3-plex on MiSeq FGx 2x151 (ForenSeq Kintelligence)
- WGS Analysis
  - Aligned with bwa mem (v0.7.17-r1188), PCR duplicates removed with picard (v2.18.29), and SNPs called using bcftools (v1.9) mpileup and bcftools call
  - Average coverage calculated using samtools (v1.18) coverage command
  - All samples first filtered for MQ>20, BQ>30, VQ>40, and DP>10
  - A custom utility developed to type SNPs in loci that are accepted in GEDmatch (GM) and type SNPs in a custom set of loci and applied to all samples (Intact samples performed best with this workflow)
  - Because the Degraded 100pg sample had too few SNPs to upload, all filters were removed
  - WGS VCFs were filtered for specific loci: loci in Ancestry/23AndMe (Ancestry), Global Screening Array (GSA), all loci that are accepted by GEDmatch (GM)
  - Genotypes were uploaded to GEDmatch PRO
  - For matchy kits, matchy segments were removed by estimating matchy ranges using the DNA Kit Evaluation tool and were removed
  - Heterozygosity for WGS was calculated by summing the number of heterozygous variants and dividing that by the number of nonreference homozygous loci
- ForenSeq Kintelligence Analysis
  - UAS automatic analysis
  - GEDmatchPRO reports were uploaded to GEDmatchPRO
  - Heterozygosity for ForenSeq Kintelligence was calculated by sum of heterozygous loci divided by the total number of SNPs typed (UAS output)

Results

- Bioanalyzer traces show degraded DNA fragments (<400bp in length)
- WGS
  - Coverage remained close to 30x for all samples with the degraded samples having the lowest average coverage (Table 1)
  - The degraded 100pg sample had the lowest number of SNPs called (Fig. 4A & 4B), which required removal of all variant calling filters
- All ForenSeq Kintelligence results had >9000 SNPs typed (Fig. 4C)
- WGS heterozygosity ratio was close to expected heterozygosity ratio for Europeans (~1.6[4]) for most samples except Degraded 100pg (Table 1)
- WGS with segment matching (Table 2)
  - The Intact and Degraded 1ng samples were too matchy, even after matchy segments were removed using the Filtered VCF (Fig. 3)
  - The custom utility performed best for the 1ng samples (Fig. 3) and was able to identify true matches but only up to 1st degree for Degraded using GM loci and up to 4th for Intact using the custom loci set (Table 2)
  - For Degraded 100pg, applying no filters and uploading Ancestry or GSA loci returned highest number of true positives
- ForenSeq Kintelligence with kinship (Table 2)
  - 1ng samples matched all expected 1st-5th relationships and one 6th degree
  - Degraded 100pg matched up to 4th degree
  - False positives remained low (Fig. 5)

Figure 1. Pedigree of a family with various kit types available in GEDmatch

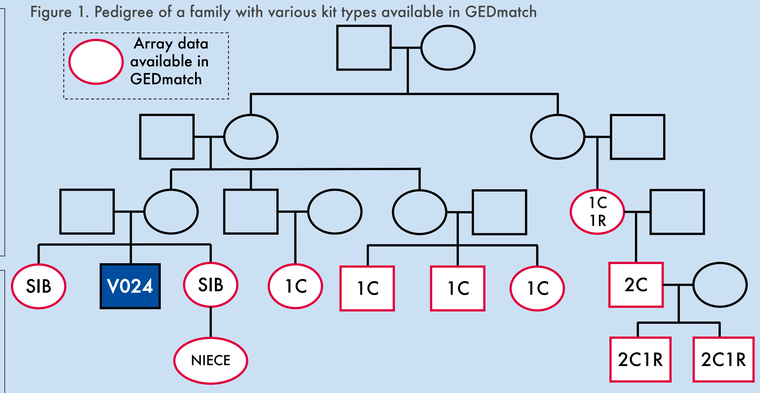


Figure 2. WGS bioanalyzer traces of Degraded 1ng (left) and Degraded 100pg (right)

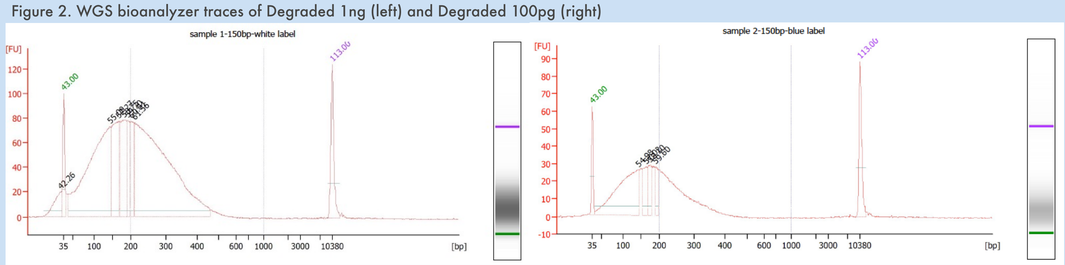


Figure 3. Bioinformatics workflows implemented for each type of sequencing (WGS [left] or ForenSeq Kintelligence [right]). ("Ancestry" includes Ancestry and 23AndMe loci; GSA=Global Screening Array; GM = Loci accepted in GEDmatch). *One-To-Many Kinship Tool was used

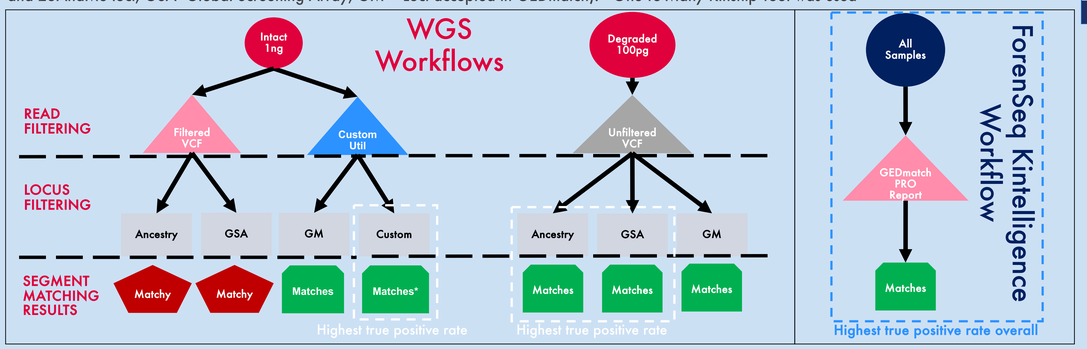


Table 1. Average coverage and heterozygosity of four samples processed with whole-genome sequencing (WGS) or ForenSeq Kintelligence on the MiSeq FGx

Sample	Degradation Index (DI)	DNA Input Amount (ng)	WGS Average Coverage	WGS Heterozygosity Ratio[4]	ForenSeq Kintelligence Heterozygosity (%)
V024	1	1.0	30.0	1.79	47.5
		0.1	37.2	2.18	39.9
	6	1.0	21.0	1.84	46.9
		0.1	28.2	4.66	33.4

Figure 4. SNP counts per sequencing platform per sample (WGS is pink, ForenSeq Kintelligence is blue)

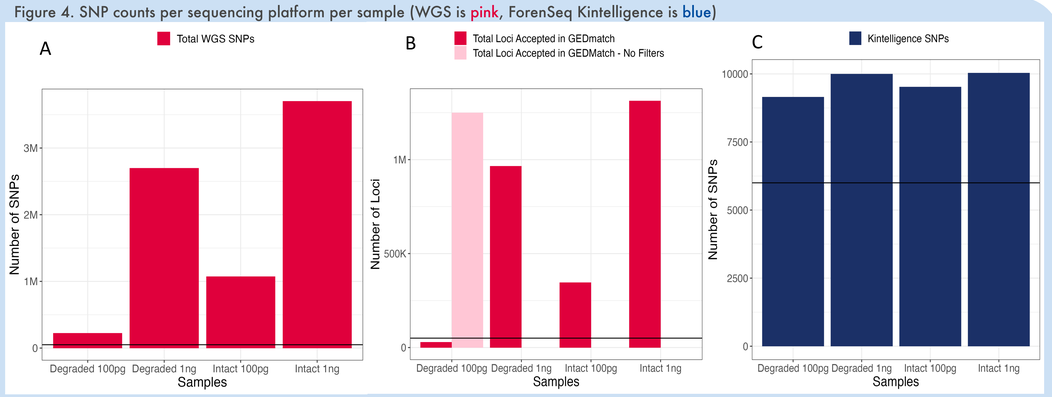
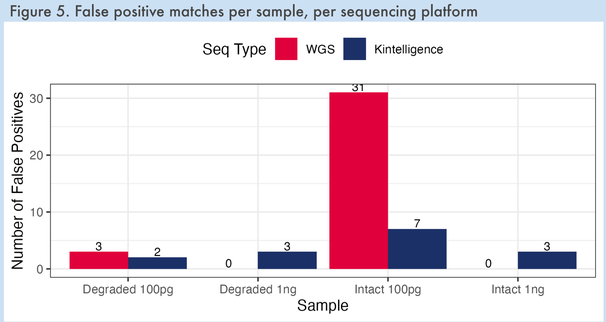


Table 2. True positive match matrix: WGS (pink) or ForenSeq Kintelligence (blue) sample per row vs expected matches per column. Green is a match, white is no match

Sample	SELF (V024)										1st	2nd	3rd					4th	5th	6th	
	Self 1	Self 2	Self 3	Self 4	Self 5	Self 6	Self 7	Self 8	Self 9	Self 10	SIB	SIB	NIECE	1C	1C	1C	1C	1C1R	2C	2C1R	2C1R
Degraded 100pg																					
Degraded 1ng																					
Intact 1ng																					
Intact 100pg																					

Sample	SELF (V024)										1st	2nd	3rd					4th	5th	6th	
	Self 1	Self 2	Self 3	Self 4	Self 5	Self 6	Self 7	Self 8	Self 9	Self 10	SIB	SIB	NIECE	1C	1C	1C	1C	1C1R	2C	2C1R	2C1R
Degraded 100pg																					
Degraded 1ng																					
Intact 1ng																					
Intact 100pg																					

Figure 5. False positive matches per sample, per sequencing platform



Conclusions

- Increased computation time and resources to process WGS data compared to ForenSeq Kintelligence. Lack of standards with WGS data processing. Unknown expectations for identifying high order relationship degrees
- ForenSeq Kintelligence workflow is simple and easy to use and is reproducible across samples. Expectations for matches are known (e.g., up to 5th degree). Consistent results across sample types: all samples matched up to 4th degree; 1ng samples matched one 6th degree
- Disease SNPs are excluded from the ForenSeq Kintelligence assay, but must be removed bioinformatically with WGS
- Method to resolve matchy kits is subjective. ForenSeq Kintelligence kits can never be matchy

Sample to Insight

REFERENCES

1. Tillmar A, et al. Whole-genome sequencing of human remains to enable genealogy DNA database searches - A case report. Forensic Sci Int Genet. 2020 May;46:102233. doi: 10.1016/j.fsigen.2020.102233. Epub 2020 Jan 17. PMID: 31981902.
2. Snedecor J, et al. Fast and accurate kinship estimation using sparse SNPs in relatively large database searches. Forensic Sci Int Genet. 2022 Nov;61:102769. doi: 10.1016/j.fsigen.2022.102769. Epub 2022 Aug 27. PMID: 36087514.
3. Verogen. SNP Typing in Universal Analysis Software and Kinship Estimation with GEDmatch PRO. Technical Note.March 2021.
4. Samuels DC, et al. Heterozygosity Ratio, a Robust Global Genomic Measure of Autozygosity and Its Association with Height and Disease Risk. Genetics. 2016 Nov;204(3):893-904. doi: 10.1534/genetics.116.189936.