APPLICATION NOTE

Embryo Identification, tracking and contamination monitoring using Mitochondrial SNVs.

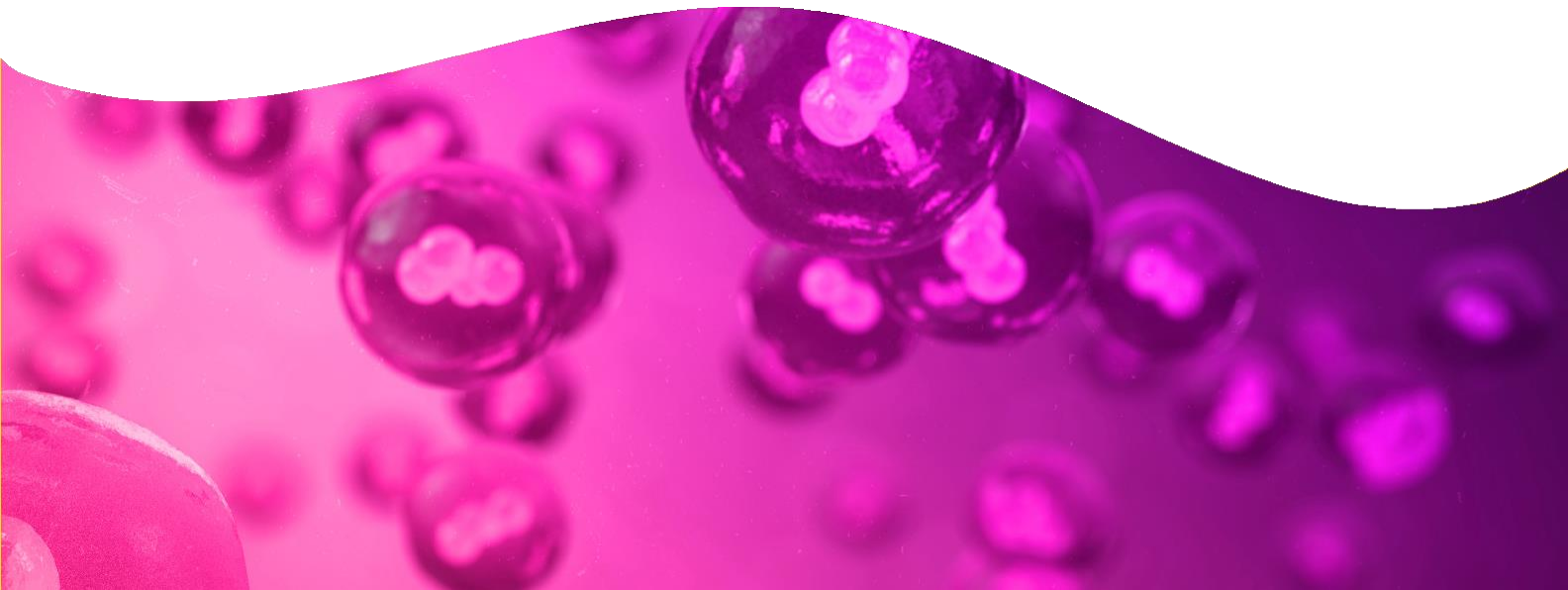
For research use only. Not for use in diagnostic procedures.

Introduction

The mitochondrial genome is maternally derived and contains single nucleotide variants (SNVs) that can be used to distinguish between individuals. In preimplantation genetic testing for aneuploidy (PGT-A), the mitochondrial DNA can be amplified, sequenced and analyzed to allow confirmation that sibling embryos are of the same maternal origin, while simultaneously checking for external human mtDNA contamination. A custom research analysis workflow and software was designed and developed for embryo biopsy samples based on each sample's mitochondrial DNA SNVs. Here, we detail the performance and accuracy of the analysis software on embryos analyzed with the PG-Seq™ Rapid v2 PGT-A kit.

Methods

Sibling and individual embryo biopsy samples were processed using the PG-Seq™ Rapid v2 kit workflow. Sequencing was performed by multiplexing 96 samples per run using a 75 cycle high output kit on the Illumina® NextSeq® Instrument. After demultiplexing, fastq files were uploaded and analyzed with the mtDetect™ analysis software. This workflow aligned reads to the hg38 mtDNA genome then performed variant calling. Known pathogenic variants along with low depth SNVs (depth <2) were excluded from the analysis. Total SNV and SNVs with ambiguous frequencies (FRQ <0.8) were counted. SNV profiles (Figure 2) for each run of 96 samples were compared and an identity matrix (Figure 3) was created and used to group



samples with 100% identical SNV profiles. Sample groupings (Figure 4) were analyzed for accuracy against known patient and sample information. To simulate lower reads numbers that would be obtained from a MiSeq® or MiniSeq® instrument, fastq reads were down sampled to 500,000 then re-run through the workflow to determine accuracy.

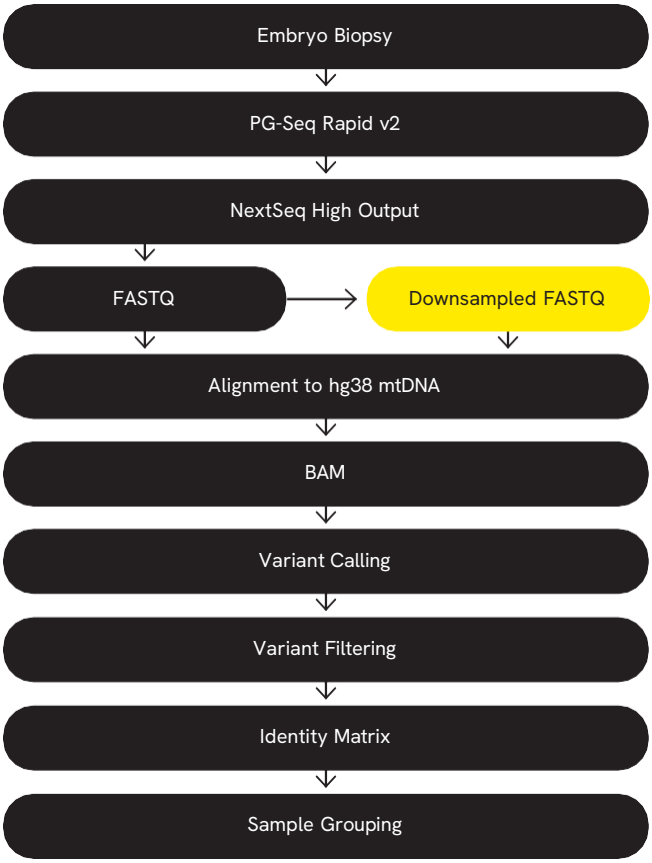


Figure 1: Flow diagram of the sample preparation and analysis workflow for embryo group determination.

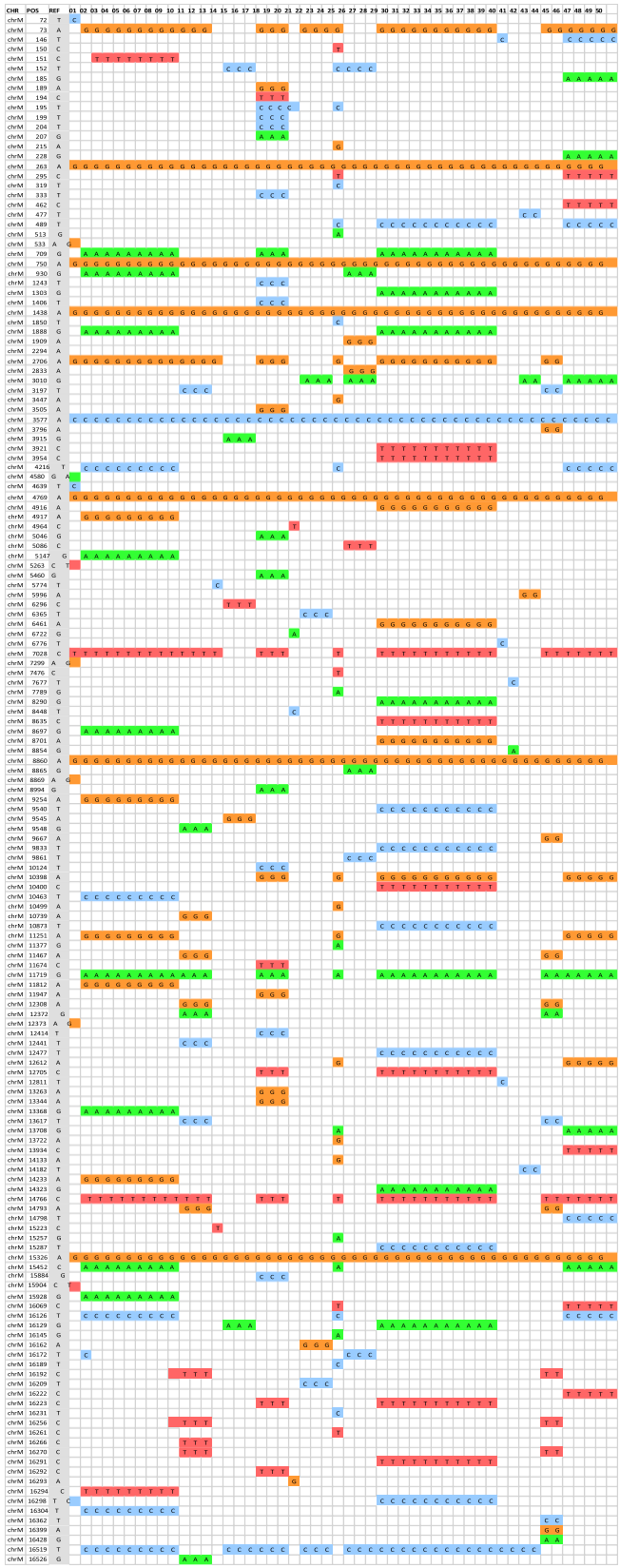


Figure 2: SNV profile example of 50 samples. Sample IDs are listed in the first row. Chromosome locations and reference bases for passed filter SNVs are shown in the first 3 columns. SNVs detected per sample listed in colored cells.

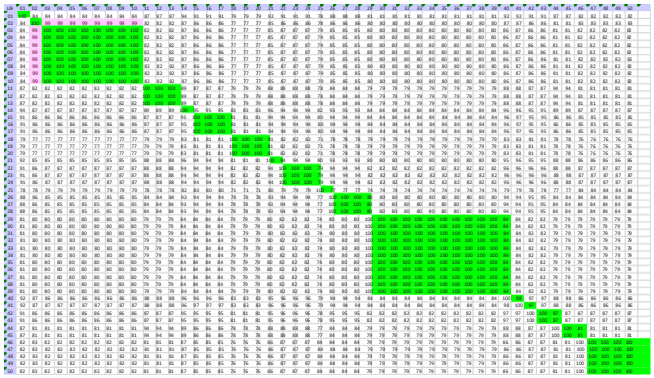


Figure 3: Matrix differentiation example of 50 samples showing the percentage of similarity to each other. Samples with 100% similarity are identified and highlighted in green.

Group	Samples									
G1	01									
G2	02									
G3	03	04	05	06	07	08	09	10		
G4	11	12	13							
G5	14									
G6	15	16	17							
G7	18	19	20							
G8	21									
G9	22	23	24							
G10	25									
G11	26	27	28							
G12	29	30	31	32	33	34	35	36	37	38	39
G13	40									
G14	41									
G15	42	43								
G16	44	45								
G17	46	47	48	49	50					

Figure 4: Determined groupings of 50 samples based on the matrix differentiation.

Results

A total of 1140 samples had WGA and sequencing performed, with each sample averaging 5.4M total reads (Figure 5). The percentage of reads aligning to the mitochondrial genome ranged from 0.4% to 20.2%, with an average of 4.8% (Figure 6).

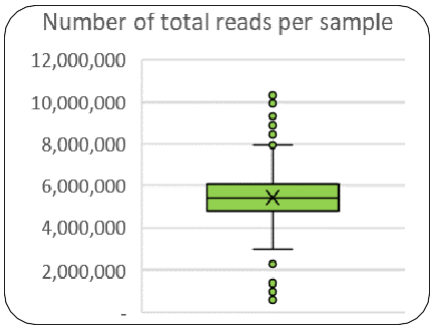


Figure 5: Plot displaying the distribution of total number of reads per sample.

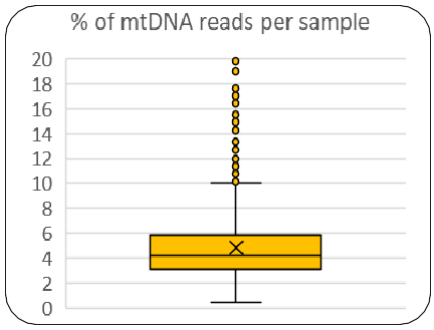


Figure 6: Plot displaying the distribution of reads per sample that align to the mitochondrial chromosome, as a percentage of the total number of reads.

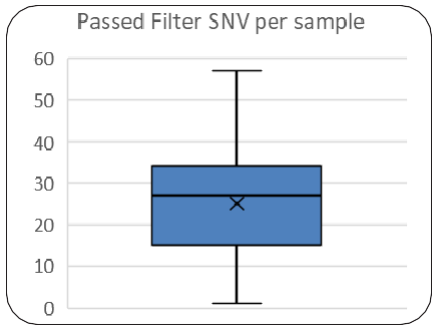


Figure 7: Plot displaying the number of SNV per sample that passed filter and was used for matrix differentiation.

Of the samples processed, 1138 (99.82%) had mtDNA aligned reads above the specified threshold of 20,000 and proceeded to SNV analysis. The number of SNV detected per sample after low depth, ambiguous frequency, and pathogenic filtering ranged from 1-57, with an average of 25.1 (Figure 7).

Matrix differentiation of mitochondrial IDs successfully grouped 1137/1138 individual and sibling embryos (99.9% accuracy) into 375 total groups from 12 sequencing runs. The single incorrectly grouped sample totally matched another group except for one SNV that was removed during filtering due to poor base quality. The difference was confirmed by review of the pre-filtered variants.

After down sampling to 500,000 fastq reads, only 3 additional samples were excluded from analysis due to having too few (< 2000) mtDNA reads. The mtDetect™ software correctly grouped 368 groups (98.13%), with 5 incorrectly grouped together when they should have been separate and 2 incorrectly grouped separately when they should have been a single group (table 1).

Table 1: Accuracy of sample groupings for high and low read number data, based on 1138 individual samples being grouped into 375 total groups.

	5.4M Average Reads	0.5M reads
Correctly grouped	374 (99.73%)	368 (98.13%)
Incorrectly grouped together	1 (0.26%)	5 (1.33%)
Incorrectly grouped separately	0 (0%)	2 (0.53%)

Case Study 1: Identification of External DNA Contamination

External DNA contamination, or the mixing of two different mtDNA profiles can be suspected when an embryo biopsy sample has a higher number of total mitochondrial SNV and SNV with ambiguous frequencies (FRQ <0.8) compared to sibling embryos and expected values. A higher number of total SNV is expected in contaminated samples as the contaminating DNAs mitochondrial SNV are added to the samples total mitochondrial SNV. Due to natural variation, the contaminating SNVs are likely to occur in positions where the embryo biopsy sample does not contain a SNV. This will show as a mtDNA base location having 2 different base frequencies and these are counted as an AMB SNV when the frequency is less than 0.8.

The expected number of AMB SNV in an embryo biopsy sample was determined by analysis of over 1000 samples. This value ranged from 3 to 14 with an average value of 8.9 (Figure 8). It is expected that samples with number of AMB SNV significantly greater than 14 are likely to be contaminated.

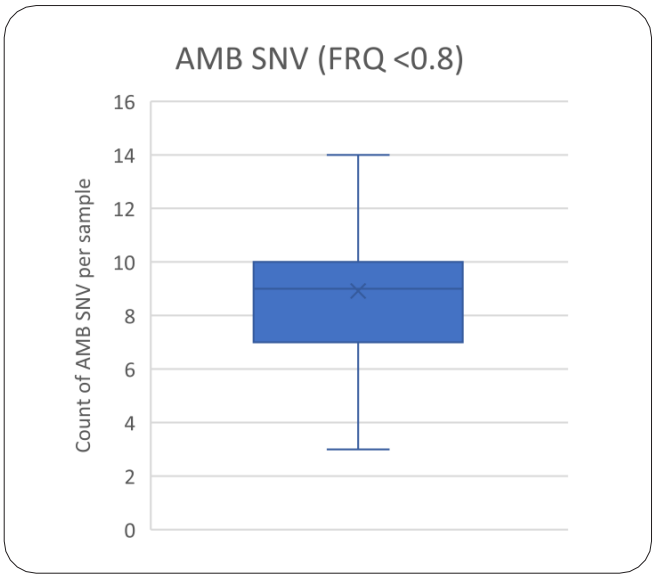


Figure 8: Plot displaying the distribution of AMB SNV per sample.

Analysis of total SNV and ambiguous SNV with FRQ <0.8 from an alternate sample set detected 2 samples with external human DNA contamination. These samples had a higher number of SNV and AMB SNV compared with their sibling embryos and the expected number from review of data from a large sample set. The two cases are described in tables 1 and 2 below. Both cases were originally suspected of being triploid 69,XXY based on the copy number ratio of the X and Y chromosomes and whole genome SNP analysis. Reviewing the mitochondrial SNV counts enabled these samples to be designated as potentially contaminated and allowed the option of re-biopsy to achieve a reliable result.

Table 1: Counts of SNV and AMB SNV for a suspected contaminated sample in comparison to expected values. Sample 1 produced double the number of total SNV and 2.5x more AMB SNV compared to its uncontaminated sibling embryos and the expected number of AMB SNV.

	# of SNV detected	AMB SNV with FRQ <0.8
Sample 1	44	34
Sibling Embryo average	22	13
Expected from large sample set analysis		3-14 (average 8.9)

Table 2: Sample 2 showing external mitochondrial ID contamination. This sample had an increased number of total SNV detected and 5x more AMB SNV detected compared to its uncontaminated sibling embryos and the expected number of AMB SNV.

	# of SNV detected	AMB SNV with FRQ <0.8
Sample 2	62	45
Sibling Embryo average	40	7
Expected from large sample set analysis		3-14 (average 8.9)

Considerations:

Mitochondrial SNV are maternally inherited and as such, this method of contamination testing cannot detect maternal DNA contamination.

Case Study 2: Identification of Sample Preparation Indexing Error

Bioinformatic grouping of sibling embryos based on their unique mtDNA profiles and comparison of results against the expected sample information allows sample tracking throughout the PGT process to be performed.

Through use of the mtDetect™ software and the Summary ID table produced, an instance was identified in which 24 samples in a 96 sample run mistakenly had inversed indexes added. The figures below highlight the groupings generated by the mtDetect™ software (figure 9) compared to the expected sample information (figure 10) and also the index sequence plate layout (figure 11).

Group	Samples								
G1	1	2							
G2	3								
G3	4	5	6						
G4	7								
G5	8	16							
G6	9								
G7	10	11	12	13	14	15			
G8	17	18	19	20					
G9	21	22	23						
G10	24								
G11	25	26	27	33	34	35			
G12	28	29	30						
G13	31	32							
G14	36								
G15	37	38	39	40					
G16	41	42	43						
G17	44	45	46	47	48	49	50		
G18	51	52	53						
G19	54	55							
G20	56	57	58	59	60	61			
G21	62	63	64						
G22	65								
G23	66	67	68	69	70	71	76	77	78
G24	72	73	74	75					
G25	79	80	81	82					
G26	83	84							
G27	85	86	87	88					
G28	89	90	91	92					
G29	93	94	95	96					

Figure 9: Summary ID table generated from the mtDetect™ analysis. Green highlighted samples indicate where the groupings match the expected results.

Through analysis of the obtained results compared to the expected results, it was identified that a number of samples had swapped results. For example: samples 9 and 16, samples 17 and 24, samples 18 and 23. Analysis of all incorrect sample groupings led to the conclusion that inverse indexes were added to samples 9-32. Detection of

Group	Samples								
G1	1	2							
G2	3								
G3	4	5	6						
G4	7								
G5	8	16							
G6	9								
G7	10	11	12	13	14	15			
G8	17	18	19	20					
G9	21	22	23						
G10	24								
G11	25	26	27	33	34	35			
G12	28	29	30						
G13	31	32							
G14	36								
G15	37	38	39	40					
G16	41	42	43						
G17	44	45	46	47	48	49	50		
G18	51	52	53						
G19	54	55							
G20	56	57	58	59	60	61			
G21	62	63	64						
G22	65								
G23	66	67	68	69	70	71	76	77	78
G24	72	73	74	75					
G25	79	80	81	82					
G26	83	84							
G27	85	86	87	88					
G28	89	90	91	92					
G29	93	94	95	96					

Figure 10: Expected results from provided sample information.

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	95
H	8	16	24	32	40	48	56	64	72	80	88	96

Figure 11: Index plate layout

this sample preparation error enabled the sample indexing to be repeated and prevented 24 samples from having the incorrect results reported to the patients. Without analysis of the mitochondrial sample groupings, this error would not have been detected.

Analysis of Mitochondrial SNV for the purposes of sample grouping is an extremely valuable tool for sample tracking and tracing. Comparison of mtDetect™ groups against the expected results can enable sample preparation errors to be detected at any stage of the PGT-A workflow.

Conclusions

Embryo biopsies processed with the PG-Seq™ Rapid v2 kit and analyzed with the mtDetect™ software can successfully be identified and grouped based on their mitochondrial DNA SNV profiles. External DNA contamination can be detected by analysis of SNV's an alternate allele frequency <0.8. The mtDetect™ tool may have the future potential to improve reproductive health research by providing accurate mtDNA based identity confirmation of sibling embryos. Identity determination allows sample tracking to be performed and can assist in the identification and monitoring of possible sample contamination along with mislabeled samples.

Revvity, Inc.
940 Winter Street
Waltham, MA 02451 USA

(800) 762-4000
www.revvity.com