EVALUATION OF AUTOMATIC CLASS III DESIGNATION FOR MISEQDX PLATFORM

DECISION SUMMARY
INSTRUMENT ONLY TEMPLATE
Correction Date: February 24, 2017
This Decision Summary contains corrections to the November 19, 2013 Decision Summary.

A. 510(k) Number: k123989

B. Purpose for Submission: De novo request for evaluation of automatic class III designation for the Illumina MiSeqDx Platform

C. Type of Test or Tests Performed: High-throughput DNA sequencing

D. Applicant: Illumina Inc.

E. Device Name: MiSeqDx Platform

F. Regulatory Information:

FDA identifies this type of device as:

1. New regulation number: 21 CFR 862.2265
2. Classification: Class II.
3. Product code: PFF - High throughput DNA sequence analyzer
4. Panel: Toxicology (91)

G. Intended Use:

1. Intended uses(s):

   The MiSeqDx Platform is a sequencing instrument that measures fluorescence signals of labeled nucleotides through the use of instrument specific reagents and flow cells (MiSeqDx Universal Kit 1.0), imaging hardware, and data analysis software. The MiSeqDx Platform is intended for targeted sequencing of human genomic DNA from peripheral whole blood samples. The MiSeqDx Platform is not intended for whole genome or de novo sequencing.

2. Indication for uses(s):

   Same as intended use above.

3. Special conditions for use statement(s):

    1. This product is limited to delivering:
        - Sequencing output >1 Gb
        - Reads >3 million
        - Read length (in paired end run) 2 x 150 bp
        - Bases higher than Q30 >75% (Greater than 75% of bases have Phred scale quality score greater than 30, indicating base call accuracy greater than 99.9%)
    2. Variants in homopolymer runs exceeding eight bases will be filtered out in the VCF files (R8 filter).
    3. The system has been validated for the detection of SNVs and up to 3 base deletions. Evaluation of 1 base insertions was been limited to 3 different insertions on 3 separate chromosomes.
    4. The system has problems detecting 1 base insertions or deletions in homopolymer tracts (e.g., polyA).
    5. This MiSeqDx system is designed to deliver qualitative (i.e. genotype) results.
    6. As with any hybridization-based workflow, underlying polymorphisms or mutations in oligonucleotide-binding regions can affect the alleles being probed and, consequently, the calls made.
    7. Recommended minimal coverage per amplicon needed for accurate variant calling (Q(max_gt | poly_site) >= 100) is 75x.

H. System Descriptions:

1. Device Description:

    The MiSeqDx Platform is a high throughput DNA sequence analyzer for clinical use.

    The MiSeqDx Platform consists of the MiSeqDx instrument and data analysis software. It is for use with the MiSeqDx Universal Kit 1.0 [MiSeqDx reagent cartridge, MiSeqDx flow cell, SBS Solution (PR2 buffer)] for library preparation and sample indexing (k133136). The end-user inputs extracted genomic DNA to be sequenced and provides the Analyte Specific Reagents (ASRs) to develop a sequencing assay that targets their sequence of interest.

2. Principles of Operation:

    Testing begins with genomic DNA extracted from a peripheral whole blood sample. The genomic DNA is processed through library preparation, which specifically amplifies the intended genomic regions of each sample while also adding the indexes and flow cell capture sequences to the amplified products. The resulting sample libraries are then transferred into a MiSeqDx reagent cartridge which contains all of the reagents required for cluster generation and sequencing (Sequencing By Synthesis - SBS). The MiSeqDx cartridge, MiSeqDx flow cell, and MiSeqDx SBS Solution (PR2 buffer) are then inserted

into the MiSeqDx instrument which performs cluster generation, sequencing and data analysis.

The instrument uses cluster generation on the flow cell surface followed by sequencing using the Sequencing by Synthesis (SBS) process.

After the flow cell images are captured by the MiSeqDx instrument following each sequencing cycle, primary analysis is performed without user intervention. Primary analysis is performed by the RTA (Real Time Analysis) software, and consists of base calling of each cluster at each cycle. In addition to calling the bases, RTA assigns an analytical quality score (Q-score) to each base call. Calculations of Q-scores are based on the ratio of the signal intensity of the highest base in a given cluster during a given cycle to the signal intensity of the three other bases. The quality score Q is calculated as -10 $\log_{10}$ P, where P is the probability that base call is incorrect.

Secondary analysis is performed by the MiSeqDx Reporter software. It also occurs without user intervention and consists of de-multiplexing and FASTQ file generation. De-multiplexing is the process of using the index sequences to assign clusters to the sample from which they originated.

After base calling and de−multiplexing, the software generates FASTQ files that contain sequence and quality information. Due to the massively parallel nature of the SBS biochemistry, hundreds of independent sequencing reads, each with their own quality score, are generated for each amplicon in each sample. The FASTQ file which is a widely accepted text based format for storing both a nucleotide sequence and its corresponding quality score. FASTQ files serve as input files for various sequence alignment and subsequent variant calling algorithms.

The MiSeqDx has a sequence alignment and variant calling program available for use.

3. Modes of Operation:

The MiSeqDx is a high throughput nucleic acid analyzer.

4. Specimen Identification:

Samples up to 96 unique specimens can be analyzed. Eight unique index primer sequences (forward), named i5 primers, and 12 unique index primer (reverse) sequences, named i7 primers, are provided. These 8 unique forward index primers and 12 unique reverse index primers, when combined in a pair wise manner, produce 96 unique index combinations allowing for up to 96 samples to be processed in parallel during the library preparation process. These are added during the library preparation process. The sample sheet, a file that the user provides the software, contains the link between each of the sample names and their associated index sequences.

After completion of the sequence run, MiSeq Reporter software de-multiplexes the samples using the index sequences and creates FASTQ files as the data analysis output. The user can also utilize the MiSeq Reporter Software for sequence alignment and variant calling.

5. Specimen Sampling and Handling:

   The MiSeqDx specimen is a pooled library (or libraries) derived from genomic DNA extracted from peripheral whole blood that then undergoes the following steps to create the pooled library: the genomic DNA is quantified and qualified and then used to make a library; the library sample is processed to remove remaining library preparation reagents (e.g. unused primers), normalized, and then pooled for input on the analyzer. Library normalization is used to ensure that each library is equally represented in the pooled sample.

   At a minimum, eight samples must be present. If six unique samples (excluding the positive and negative controls) are not available, it is acceptable to fill the run with sample replicates or any human genomic DNA sample.

6. Calibration:

   There is no end-user calibration of the system. During installation of the platform, a company representative (Field Applications Scientist) begins a series of tests to validate the performance of the instrument subsystems, which include optical alignment, fluidic delivery, and thermal calibration, among others. In the case of a test failure, the MiSeqDx company representative uses a set of instrument-specific tools to adjust and/or repair the instrument to meet operational specifications. Re-calibration occurs during the preventive maintenance visit.

7. Quality Control:

   A PhiX internal control (i.e. genomic DNA from the bacteriophage ΦX174) is added to each pooled library prior to placement on the instrument. Successful sequencing of the PhiX genome indicates that the sequencing chemistry worked as expected. A negative control, or no template control, (not provided by the sponsor) should be included in every run in order to detect the presence of contamination in the environment or run.

8. Software:

   FDA has reviewed applicant's Hazard Analysis and Software Development processes for this line of product types:

   Yes___X____ or No_____

H. Substantial Equivalence Information:

1. Predicate Device Name(s) and 510(k) numbers:

   Not applicable.

2. Comparison with Predicate Device:

   Not applicable.

I. Special Control/Guidance Document Referenced (if applicable):

Not applicable.

J. Performance Characteristics*:*

1. Analytical Performance:

   a. Accuracy:

   Three accuracy studies were conducted.

   Study 1: This accuracy study used a representative assay designed to query a variety of genes covering 24,434 bases across 19 different chromosomes, and containing potentially clinically relevant exons. The 13 unique samples used in this study are from two parents and 11 children that have been frequently sequenced by multiple laboratories and sequencing methodologies. There are six samples from females and seven from males.
   Accuracy was determined for single nucleotide variants (SNVs) by comparing the study data to well-characterized composite reference information. The reference database sequence was derived from the combination of multiple sequencing methodologies, publicly available data, and hereditary information. The following table to evaluate accuracy of the system was compiled based on data from the first run in the study. No repeat testing was done for this study.

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of unique samples	total # of samples analyzed[2]	# calls/ sample that could be made[3]	# of no calls[4]	# of correct calls/ sample[5]	# in-correct calls[6]	% correct calls[7]
1	1	132	Poly C (5), 63% GC	13	15	132	0	132	0	100.00
2	1	128	Poly T (5)	13	15	128	0	128	0	100.00
3	2	133	-	13	15	133	0	133	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of unique samples	total # of samples analyzed[2]	# calls/ sample that could be made[3]	# of no calls[4]	# of correct calls/ sample[5]	# in-correct calls[6]	% correct calls[7]
4	2	119	-	13	15	119	0	119	0	100.00
5	2	127	Poly T (5)	13	15	127	0	127	0	100.00
6	2	135	Poly A (6)	13	15	135	0	135	0	100.00
7	2	122	Poly T (5), Poly C (5)	13	15	122	0	122	0	100.00
8	2	110	Poly T (5)	13	15	110	0	110	0	100.00
9[8]	2	131	Poly A (14)	13	15	130-131	0	130-131	9	99.54
10	2	117	-	13	15	117	0	117	0	100.00
11	2	121	-	13	15	121	0	121	0	100.00
12	2	114	-	13	15	114	0	114	0	100.00
13	2	129	Poly A (5)	13	15	129	0	129	0	100.00
14	3	131	Poly A (5), Poly T (5)	13	15	131	0	131	0	100.00
15	3	130	-	13	15	130	0	130	0	100.00
16	3	130	-	13	15	130	0	130	0	100.00
17	3	117	-	13	15	117	0	117	0	100.00
18	3	136	Poly T (5)	13	15	136	0	136	0	100.00
19	3	131	Poly T (5), SNV	13	15	131	0	131	0	100.00
20	3	123	Poly A (5)	13	15	123	0	123	0	100.00
21	3	117	Poly A (6), Poly T (5), Homologous region on a different chromosome	13	15	117	0	117	0	100.00
22	3	119	Homologous region on a different chromosome	13	15	119	0	119	0	100.00
23	3	120	-	13	15	120	0	120	0	100.00
24	3	129	Poly T (5)	13	15	129	0	129	0	100.00
25	4	133	Poly C (7), 66% GC	13	15	133	0	133	0	100.00
26	4	135	Poly C (5), 69% GC	13	15	135	0	135	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of unique samples	total # of samples analyzed[2]	# calls/ sample that could be made[3]	# of no calls[4]	# of correct calls/ sample[5]	# in-correct calls[6]	% correct calls[7]
27	4	123	SNV	13	15	123	0	123	0	100.00
28	4	134	-	13	15	134	0	134	0	100.00
29	4	132	-	13	15	132	0	132	0	100.00
30	4	121	Poly A (5), SNV	13	15	121	0	121	0	100.00
31	4	125	-	13	15	125	0	125	0	100.00
32	4	134	Poly T (5)	13	15	134	0	134	0	100.00
33	4	118	-	13	15	118	0	118	0	100.00
34	4	122	Poly A (5)	13	15	122	0	122	0	100.00
35	4	131	-	13	15	131	0	131	0	100.00
36	4	133	-	13	15	133	0	133	0	100.00
37	4	128	Poly T (6)	13	15	128	0	128	0	100.00
38	4	131	-	13	15	131	0	131	0	100.00
39	4	129	Poly A (5), Poly T (5), SNV	13	15	129	0	129	0	100.00
40	4	133	Poly T (5), SNV	13	15	133	0	133	0	100.00
41	4	112	SNV	13	15	112	0	112	0	100.00
42	4	133	-	13	15	133	0	133	0	100.00
43	4	135	-	13	15	135	0	135	0	100.00
44	4	122	-	13	15	122	0	122	0	100.00
45	4	117	-	13	15	117	0	117	0	100.00
46[9]	4	124	-	13	15	125	0	125	0	100.00
47	4	117	Poly T (5)	13	15	117	0	117	0	100.00
48	4	128	Poly A (7)	13	15	128	0	128	0	100.00
49	4	123	Poly A (6)	13	15	123	0	123	0	100.00
50	4	133	-	13	15	133	0	133	0	100.00
51	4	112	-	13	15	112	0	112	0	100.00
52	4	129	-	13	15	129	0	129	0	100.00
53	4	126	-	13	15	126	0	126	0	100.00
54	4	132	-	13	15	132	0	132	0	100.00
55	5	131	-	13	15	131	0	131	0	100.00
56	5	119	-	13	15	119	0	119	0	100.00
57	5	120	Poly A (5)	13	15	120	0	120	0	100.00
58	5	119	-	13	15	119	0	119	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of unique samples	total # of samples analyzed[2]	# calls/ sample that could be made[3]	# of no calls[4]	# of correct calls/ sample[5]	# in-correct calls[6]	% correct calls[7]
59	5	118	-	13	15	118	0	118	0	100.00
60	5	112	-	13	15	112	0	112	0	100.00
61	5	120	-	13	15	120	0	120	0	100.00
62	5	120	Poly A (5)	13	15	120	0	120	0	100.00
63	5	115	CT(5)	13	15	115	0	115	0	100.00
64	5	112	SNV	13	15	112	0	112	0	100.00
65	5	135	Poly T (6)	13	15	135	0	135	0	100.00
66	5	131	63% GC	13	15	131	0	131	0	100.00
67	5	121	-	13	15	121	0	121	0	100.00
68	5	132	Poly A (6), Poly T (8)	13	15	132	0	132	0	100.00
69	7	133	-	13	15	133	0	133	0	100.00
70	7	120	60% GC	13	15	120	0	120	0	100.00
71	7	135	-	13	15	135	0	135	0	100.00
72	7	126	Poly A (5), 59% GC	13	15	126	0	126	0	100.00
73	7	134	-	13	15	134	0	134	0	100.00
74	7	122	Poly C (5), 63% GC	13	15	122	0	122	0	100.00
75	7	127	59% GC; SNV	13	15	127	0	127	0	100.00
76	7	123	-	13	15	123	0	123	0	100.00
77	7	125	-	13	15	125	0	125	0	100.00
78	7	133	Poly A (5), Poly T (5)	13	15	133	0	133	0	100.00
79	7	116	-	13	15	116	0	116	0	100.00
80	7	135	-	13	15	135	0	135	0	100.00
81	7	118	-	13	15	118	0	118	0	100.00
82	7	136	67% GC	13	15	136	0	136	0	100.00
83	7	131	58% GC	13	15	131	0	131	0	100.00
84	7	119	Poly G (6), 61% GC	13	15	119	0	119	0	100.00
85	7	122	Poly T (5)	13	15	122	0	122	0	100.00
86	7	123	Poly A (6)	13	15	123	0	123	0	100.00
87	8	127	60% GC	13	15	127	0	127	0	100.00
88	8	129	57% GC	13	15	129	0	129	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of unique samples	total # of samples analyzed[2]	# calls/ sample that could be made[3]	# of no calls[4]	# of correct calls/ sample[5]	# in-correct calls[6]	% correct calls[7]
89	9	130	Poly T (5)	13	15	130	0	130	0	100.00
90	9	116	-	13	15	116	0	116	0	100.00
91	9	119	Homologous region on a different chromosome	13	15	119	0	119	0	100.00
92	9	121	-	13	15	121	0	121	0	100.00
93	9	117	Homologous region on a different chromosome	13	15	117	0	117	0	100.00
94	9	114	-	13	15	114	0	114	0	100.00
95[10]	9	129	Poly A (14)	13	15	130	0	129 (of 130)	15	99.23
96	9	114	Homologous region on a different chromosome; SNV	13	15	114	0	114	0	100.00
97	9	122	-	13	15	122	0	122	0	100.00
98	9	127	Poly A (5), Poly C (5)	13	15	127	0	127	0	100.00
99	9	133	-	13	15	133	0	133	0	100.00
100	9	138	64% GC	13	15	138	0	138	0	100.00
101	9	139	-	13	15	139	0	139	0	100.00
102	9	116	-	13	15	116	0	116	0	100.00
103	9	133	Poly A (5), 57% GC	13	15	133	0	133	0	100.00
104	9	138	57% GC	13	15	138	0	138	0	100.00
105	9	136	Poly C (5), 67% GC	13	15	136	0	136	0	100.00
106	9	118	70% GC	13	15	118	0	118	0	100.00
107	10	128	62% GC	13	15	128	0	128	0	100.00
108	10	120	60% GC	13	15	120	0	120	0	100.00
109	10	139	58% GC; SNV	13	15	139	0	139	0	100.00
110	10	118	57% GC	13	15	118	0	118	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of unique samples	total # of samples analyzed[2]	# calls/ sample that could be made[3]	# of no calls[4]	# of correct calls/ sample[5]	# in-correct calls[6]	% correct calls[7]
111	10	123	Poly T (5)	13	15	123	0	123	0	100.00
112	10	121	-	13	15	121	0	121	0	100.00
113	10	129	26% GC	13	15	129	0	129	0	100.00
114	10	122	-	13	15	122	0	122	0	100.00
115	10	124	Poly T (5); Homologous region on a different chromosome	13	15	124	0	124	0	100.00
116	10	135	CA(4)	13	15	135	0	135	0	100.00
117	10	135	Poly A (6); Homologous region on a different chromosome	13	15	135	0	135	0	100.00
118	10	119	Poly C (5); SNV	13	15	119	0	119	0	100.00
119	10	125	-	13	15	125	0	125	0	100.00
120	10	131	-	13	15	131	0	131	0	100.00
121	10	117	-	13	15	117	0	117	0	100.00
122	10	116	-	13	15	116	0	116	0	100.00
123	10	129	58% GC	13	15	129	0	129	0	100.00
124	11	117	Poly T (10)	13	15	117	0	117	0	100.00
125	11	117	Poly T (5)	13	15	117	0	117	0	100.00
126	11	113	Poly A (5)	13	15	113	0	113	0	100.00
127	11	129	-	13	15	129	0	129	0	100.00
128	11	121	Poly T (5)	13	15	121	0	121	0	100.00
129	11	123	-	13	15	123	0	123	0	100.00
130	11	127	Poly A (6)	13	15	127	0	127	0	100.00
131	11	136	Poly T (6)	13	15	136	0	136	0	100.00
132	11	132	Poly T (5)	13	15	132	0	132	0	100.00
133	11	115	-	13	15	115	0	115	0	100.00
134	11	117	Poly T (8); 19% GC	13	15	117	0	117	0	100.00
135	11	134	Poly A (5); Poly T (5)	13	15	134	0	134	0	100.00
136	11	131	Poly A (5)	13	15	131	0	131	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of unique samples	total # of samples analyzed[2]	# calls/ sample that could be made[3]	# of no calls[4]	# of correct calls/ sample[5]	# in-correct calls[6]	% correct calls[7]
137	11	133	26% GC; SNV	13	15	133	0	133	0	100.00
138	11	137	Poly T (8); SNV	13	15	137	0	137	0	100.00
139	11	131	Poly A (5)	13	15	131	0	131	0	100.00
140	12	131	-	13	15	131	0	131	0	100.00
141	12	128	-	13	15	128	0	128	0	100.00
142	12	133	Poly A (5)	13	15	133	0	133	0	100.00
143	12	136	-	13	15	136	0	136	0	100.00
144	12	124	-	13	15	124	0	124	0	100.00
145	12	122	59% GC	13	15	122	0	122	0	100.00
146	13	122	-	13	15	122	0	122	0	100.00
147	13	116	Poly C (5)	13	15	116	0	116	0	100.00
148	13	133	-	13	15	133	0	133	0	100.00
149	13	117	SNV	13	15	117	0	117	0	100.00
150	13	124	Poly T (6)	13	15	124	0	124	0	100.00
151	13	123	Poly T (5); 26% GC	13	15	123	0	123	0	100.00
152	13	115	Poly A (5)	13	15	115	0	115	0	100.00
153	13	125	-	13	15	125	0	125	0	100.00
154	13	121	-	13	15	121	0	121	0	100.00
155	13	123	-	13	15	123	0	123	0	100.00
156	13	114	-	13	15	114	0	114	0	100.00
157	13	119	-	13	15	119	0	119	0	100.00
158	14	122	58% GC	13	15	122	0	122	0	100.00
159	16	122	-	13	15	122	0	122	0	100.00
160	16	121	-	13	15	121	0	121	0	100.00
161	16	123	Poly C (5)	13	15	123	0	123	0	100.00
162	17	119	-	13	15	119	0	119	0	100.00
163	17	119	61% GC	13	15	119	0	119	0	100.00
164	17	135	-	13	15	135	0	135	0	100.00
165	17	116	Poly C (6); 60% GC; SNV	13	15	116	0	116	0	100.00
166	17	123	-	13	15	123	0	123	0	100.00
167	17	116	62% GC	13	15	116	0	116	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of unique samples	total # of samples analyzed[2]	# calls/ sample that could be made[3]	# of no calls[4]	# of correct calls/ sample[5]	# in-correct calls[6]	% correct calls[7]
168	17	118	Poly C (5); 65% GC	13	15	118	0	118	0	100.00
169	17	129	-	13	15	129	0	129	0	100.00
170	17	131	Poly G (6); 67% GC; SNV	13	15	131	0	131	0	100.00
171	17	127	61% GC	13	15	127	0	127	0	100.00
172	17	118	Poly C (5)	13	15	118	0	118	0	100.00
173	17	138	61% GC	13	15	138	0	138	0	100.00
174	17	131	58% GC	13	15	131	0	131	0	100.00
175	18	112	-	13	15	112	0	112	0	100.00
176	18	124	-	13	15	124	0	124	0	100.00
177	18	134	Poly A (6)	13	15	134	0	134	0	100.00
178	18	129	-	13	15	129	0	129	0	100.00
179	18	133	-	13	15	133	0	133	0	100.00
180	18	118	-	13	15	118	0	118	0	100.00
181	18	114	60% GC	13	15	114	0	114	0	100.00
182	18	118	-	13	15	118	0	118	0	100.00
183	19	122	Poly G (6); 66% GC	13	15	122	0	122	0	100.00
184	19	139	64% GC	13	15	139	0	139	0	100.00
185	19	131	67% GC	13	15	131	0	131	0	100.00
186	19	141	59% GC; Homologous region on a different chromosome	13	15	141	0	141	0	100.00
187	19	121	Poly C (5); 72% GC; Homologous region on a different chromosome	13	15	121	0	121	0	100.00
188	19	138	58% GC	13	15	138	0	138	0	100.00
189	19	123	64% GC	13	15	123	0	123	0	100.00
190	19	138	-	13	15	138	0	138	0	100.00
191	20	117	Poly T (5)	13	15	117	0	117	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of unique samples	total # of samples analyzed[2]	# calls/ sample that could be made[3]	# of no calls[4]	# of correct calls/ sample[5]	# in-correct calls[6]	% correct calls[7]
192	22	136	Poly A (7)	13	15	136	0	136	0	100.00
193	22	122	Poly A (5); Poly C (5)	13	15	122	0	122	0	100.00
194	22	122	62% GC; SNV	13	15	122	0	122	0	100.00
195	22	119	66% GC	13	15	119	0	119	0	100.00

[1]. Analyzed fragment means the size of the sequenced genomic region in bases, not including target-specific primers.

[2] Total # of samples analyzed is 15 because two of the 13 unique samples were run in two independent replicates.

[3] # calls/sample that could be made is the number of bases that had adequate quality to be called by the system

[4] # of no calls is the number of bases in an amplicon that results in a no call in the run

[5] # correct calls per sample is number of bases in the amplicon that were called that had results that matched the human genome reference sequence build 19[1] and the well characterized composite reference.

[6] # incorrect calls were the total number of incorrect calls for the SNV or indel in that amplicon; addition details on incorrect calls are presented in footnotes below.

[7] % correct calls equals the correct call rate for all of the bases in the amplicon, where the correct call for the SNV or indel is based on the well characterized composite reference information and the correct call for the bases in the remainder of the amplicon sequence is based on comparison to human genome reference sequence build 19. This column may have more than one expected result for a given amplicon if some samples contain an indel while some do not, e.g., amplicon 9.

[8] Amplicon 9 has a homopolymer run of 14 A's according to the human genome reference sequence build 19. However, the well characterized composite reference information for 7 out of 13 samples have 13 A's in this homopolymer run. In these 7 samples, this one base pair deletion represents a false negative in the MiSeqDx sequencing accuracy study.

[1] Human Feb. 2009 (GRCh37/hg19) assembly available from NCBI

[9] Amplicon 46 has a one base insertion which is reported in 9 samples in the well characterized reference database and is correctly detected in all analyzed samples.

[10] Amplicon 95 has a homopolymer run of 14 A's according to human genome reference sequence build 19. However, the well characterized composite reference sequences for 13 out of 13 samples have 15 A's in this homopolymer run. In these 13 samples, this one base pair insertion is a false negative in the MiSeqDx sequencing accuracy study.

The following table contains data from study 1 presented with positive and negative percent agreement (PPA and NPA, respectively), where the variant results are compared to the well characterized composite reference information for PPA calculations. Since the composite reference information only provides results for the single nucleotide variants and insertions/deletions, non-variant base results are compared to human genome reference sequence build 19 for NPA calculations. All non-variant bases had 100% agreement with the reference sequence. All SNVs had 100% agreement with the reference sequence. Variants that were missed were either 1 base insertions or 1 base deletions in the homopolymer regions.

Sample	# amplicons	% Amplicon Coverage[1]	Variants expected per sample[2]	Variants Correctly Called	Variants Missed[3]	Non-variant bases called correctly	PPA[4]	NPA[5]
NA12877	195	100	19	17	2	24418	89.5	100
NA12878	195	100	19	17	2	24417	89.5	100
NA12879	195	100	20	19	1	24416	95	100
NA12880	195	100	20	18	2	24417	90	100
NA12881	195	100	22	20	2	24415	90.9	100
NA12882	195	100	16	15	1	24419	93.8	100
NA12883	195	100	24	23	1	24412	95.8	100
NA12884	195	100	21	20	1	24415	95.2	100
NA12885	195	100	19	17	2	24417	89.5	100
NA12886	195	100	22	20	2	24415	90.9	100
NA12887	195	100	19	18	1	24416	94.7	100
NA12888	195	100	24	23	1	24412	95.8	100
NA12893	195	100	20	18	2	24417	90	100

[1] % Amplicon coverage is number of bases in the amplicons sequenced with confidence

[2] Variants expected per sample includes both SNVs and indels

[3] For the variants missed, please see the first table for study 1 and the footnotes 8-10.

[4] Positive percent agreement (PPA) = 100xTP/(TP+FN) where the true positives (TP) are the number of positive variant calls at genomic coordinates where variants are

present according to the reference sequence and mutant allele called is concordant with reference sequence (column named "Variants called correctly") and the false negatives (FN) are the number of negative variant calls at genomic coordinates where variants are present according to the reference sequence (column named "Variants missed).

[5] Negative percent agreement (NPA) = 100xTN/(FP+TN) where the false positives (FP) are the number of positive variant calls at genomic coordinates where variants are absent according to the reference sequence, or if mutant allele called is discordant with reference sequence (not in the table; no false positive variants calls were made in this study) and true negatives (TN) are the number of negative variant calls at genomic coordinates where variants are absent according to the reference standard (column named "non-variant bases called correctly").

Study 2: The sequencing results for the amplicon panel above were compared to a highly confident genotype established for NA12878 by the National Institutes of Standards and Technology (NIST) (v.2.15[2]). Out of the 195 amplicons, 184 amplicons lied within highly confident reference calls in the NIST sequence and were compared. Non-variant base calls were compared to human genome reference sequence build 19.

Sample	# Amplicons	% Amplicon Coverage[1]	Variants expected	Variants Correctly Called	Variants Missed	Non-variant bases called correctly	PPA[2] (%)	NPA[3] (%)
NA12878	184	100	17	16	1[4]	23066	94.1	100

[1] % Amplicon coverage is number of bases in the amplicons sequenced with confidence

[2] Positive percent agreement (PPA) = 100xTP/(TP+FN)

[3] Negative percent agreement (NPA) = 100xTN/(FP+TN)

[4] The missed variant is the one base pair deletion in amplicon 9 in the homopolymer run of 14 A's not called by the MiSeqDx that is present in the NIST sequence. Note that the NIST sequence does not include the one base pair insertion in the other homopolymer of A's that was present in the other reference database used above in study 1.

[2] Zook, JM et al. Integrating sequencing datasets to form highly confident SNP and indel genotype calls for a whole human genome. arXiv:1307.4661 [q-bio.GN]

Study 3: An additional accuracy study was performed to assess the performance of small insertions and deletions within a representative assay, the Illumina MiSeqDx Cystic Fibrosis 139 Variant Assay, that included a subset of CFTR clinically significant genetic variations analyzed with the MiSeq Reporter (MSR) software v2.2.29.2 using the MiSeqDx Platform targeted DNA sequencing workflow. The queried insertions and deletions were detected where expected with high confidence. These samples were characterized by bidirectional Sanger sequencing as a reference method to establish the expected sequence.

Amplicon	Analyzed fragment size[1]	Amplicon Genomic Content	# calls/ sample that could be made	# of bases called/ sample	# of no calls	# of correct calls/ sample	# of incorrect calls	% correct calls
1	129	1 base insertion	130	130	0	130	0	100.00
2	154	3 base deletion	151	151	0	151	0	100.00
3	167	2 base deletion	165	165	0	165	0	100.00
4	134	1 base deletion	133	133	0	133	0	100.00
5	132	1 base deletion	131	131	0	131	0	100.00
6	129	1 base deletion	128	128	0	128	0	100.00

The data provided by these three accuracy studies supports the claim that the MiSeqDx Instrument can accurately sequence:

- GC content ≥ 19% (all bases in 135 out of 135 sequenced amplicons with 19% GC content called correctly)
- GC content ≤ 72% (all bases in 135 out of 135 sequenced amplicons with 72% GC content called correctly)
- PolyA lengths ≤ 7 (PolyA repeat of 7 nucleotides was called correctly in 270 out of 270 sequenced amplicons containing PolyA =7)
- PolyT lengths ≤ 8 (PolyT repeat of 8 nucleotides was called correctly in 270 out of 270 sequenced amplicons containing PolyT =8)
- PolyG lengths ≤ 6 (PolyG repeat of 6 nucleotides was called correctly in 405 out of 405 sequenced amplicons containing PolyG =6)
- PolyC lengths ≤ 7 (PolyC repeat of 7 nucleotides was called correctly in 135 out of 135 sequenced amplicons containing PolyC =7))
- Dinucleotide repeat lengths ≤ 5x (all bases in 135 out of 135 sequenced amplicons with 5x dinucleotide repeat were called correctly)
- Trinucleotide repeat lengths ≤ 4x (all bases in 810 out of 810 sequenced amplicons with 4x trinucleotide repeats were called correctly)
- 1 base insertions and 3 or fewer base deletions

- 2 out of 3 1-base insertions tested were called correctly. Correct calls were made for two 1-base insertions in non-homopolymer regions in 82 amplicons. One 1-base insertion was not called in a homopolymer run of 14 A's on chromosome 2.in 135 amplicons.
- 3 out of 4 1-base deletions called correctly. All correct calls were made in non-homopolymer regions in 4 amplicons. One 1-base deletion was not called in a homopolymer run of 14 A's on chromosome 9 in 63 amplicons.
- 2-base deletion called correctly in one sample
- 3-base deletions called correctly in 21 samples

c. *Precision/Reproducibility*:

Two reproducibility studies were conducted.

Study 1: This reproducibility study used a representative assay designed to query a variety of genes covering 24,434 bases across 19 different chromosomes, and containing potentially clinically relevant exons. The study examined 13 samples over nine runs using three different MiSeqDx instruments and three different operators. The 13 samples are from two parents and 11 children that have been frequently sequenced by multiple laboratories and sequencing methodologies. The samples were run in duplicate, so each run generated results for 15 samples.

For the evaluation of lot-to-lot reproducibility, 94 samples and two non-template controls were tested across three lots. Each lot was split into two 48-sample runs to test all reagents and possible index primer combinations. All sequencing runs were completed by a single operator and on a single MiSeqDx instrument to remove any potential variance contributed from operator or instrument.

Correct calls were determined for single nucleotide variants (SNVs) by comparing the study data to well characterized reference information. No repeat testing was done for the reproducibility study. The following tables show the results of the study to evaluate reproducibility of the system.

Results of the study by instrument:

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of samples run[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					total # of no calls[3]	total # of incorrect calls[4]	% correct calls[5]	total # of no calls	total # of incorrect calls	% correct calls	total # of no calls	total # of incorrect calls	% correct calls
1	1	132	Poly C (5); 63% GC	135	0	0	100.00	23[6]	0	99.61[7]	39[6]	0	99.34[7]
2	1	128	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
3	2	133	-	135	0	0	100.00	0	0	100.00	0	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of samples run[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					total # of no calls[3]	total # of incorrect calls[4]	% correct calls[5]	total # of no calls	total # of incorrect calls	% correct calls	total # of no calls	total # of incorrect calls	% correct calls
4	2	119	-	135	0	0	100.00	0	0	100.00	0	0	100.00
5	2	127	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
6	2	135	Poly A (6)	135	0	0	100.00	0	0	100.00	0	0	100.00
7	2	122	Poly T (5); Poly C (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
8	2	110	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
9	2	131	Poly A (14)	135	0	27[8]	99.54	0	27[8]	99.54	0	27[8]	99.54
10	2	117	-	135	0	0	100.00	0	0	100.00	0	0	100.00
11	2	121	-	135	0	0	100.00	0	0	100.00	0	0	100.00
12	2	114	-	135	0	0	100.00	0	0	100.00	0	0	100.00
13	2	129	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
14	3	131	Poly A (5); Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
15	3	130	-	135	0	0	100.00	0	0	100.00	0	0	100.00
16	3	130	-	135	0	0	100.00	0	0	100.00	0	0	100.00
17	3	117	-	135	0	0	100.00	0	0	100.00	0	0	100.00
18	3	136	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
19	3	131	Poly T (5); SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
20	3	123	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
21	3	117	Poly A (6); Poly T (5); Homologous region on a different chromosome	135	0	0	100.00	0	0	100.00	0	0	100.00
22	3	119	Homologous region on a different chromosome	135	0	0	100.00	0	0	100.00	0	0	100.00
23	3	120	-	135	0	0	100.00	0	0	100.00	0	0	100.00
24	3	129	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
25	4	133	Poly C (7); 66% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
26	4	135	Poly C (5); 69% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
27	4	123	SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
28	4	134	-	135	0	0	100.00	0	0	100.00	0	0	100.00
29	4	132	-	135	0	0	100.00	0	0	100.00	0	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of samples run[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					total # of no calls[3]	total # of incorrect calls[4]	% correct calls[5]	total # of no calls	total # of incorrect calls	% correct calls	total # of no calls	total # of incorrect calls	% correct calls
30	4	121	Poly A (5); SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
31	4	125	-	135	0	0	100.00	0	0	100.00	0	0	100.00
32	4	134	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
33	4	118	-	135	0	0	100.00	0	0	100.00	0	0	100.00
34	4	122	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
35	4	131	-	135	0	0	100.00	0	0	100.00	0	0	100.00
36	4	133	-	135	0	0	100.00	0	0	100.00	0	0	100.00
37	4	128	Poly T (6)	135	0	0	100.00	0	0	100.00	0	0	100.00
38	4	131	-	135	0	0	100.00	0	0	100.00	0	0	100.00
39	4	129	Poly A (5); Poly T (5); SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
40	4	133	Poly T (5); SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
41	4	112	SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
42	4	133	-	135	0	0	100.00	0	0	100.00	0	0	100.00
43	4	135	-	135	0	0	100.00	0	0	100.00	0	0	100.00
44	4	122	-	135	0	0	100.00	0	0	100.00	0	0	100.00
45	4	117	-	135	0	0	100.00	0	0	100.00	0	0	100.00
46	4	124	-	135	0	0	100.00	0	0	100.00	0	0	100.00
47	4	117	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
48	4	128	Poly A (7)	135	0	0	100.00	0	0	100.00	0	0	100.00
49	4	123	Poly A (6)	135	0	0	100.00	0	0	100.00	0	0	100.00
50	4	133	-	135	0	0	100.00	0	0	100.00	0	0	100.00
51	4	112	-	135	0	0	100.00	0	0	100.00	0	0	100.00
52	4	129	-	135	0	0	100.00	0	0	100.00	0	0	100.00
53	4	126	-	135	0	0	100.00	0	0	100.00	0	0	100.00
54	4	132	-	135	0	0	100.00	0	0	100.00	0	0	100.00
55	5	131	-	135	0	0	100.00	0	0	100.00	0	0	100.00
56	5	119	-	135	0	0	100.00	0	0	100.00	0	0	100.00
57	5	120	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
58	5	119	-	135	0	0	100.00	0	0	100.00	0	0	100.00
59	5	118	-	135	0	0	100.00	0	0	100.00	0	0	100.00
60	5	112	-	135	0	0	100.00	0	0	100.00	0	0	100.00
61	5	120	-	135	0	0	100.00	0	0	100.00	0	0	100.00
62	5	120	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
63	5	115	CT(5)	135	0	0	100.00	0	0	100.00	0	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of samples run[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					total # of no calls[3]	total # of incorrect calls[4]	% correct calls[5]	total # of no calls	total # of incorrect calls	% correct calls	total # of no calls	total # of incorrect calls	% correct calls
64	5	112	SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
65	5	135	Poly T (6)	135	0	0	100.00	0	0	100.00	0	0	100.00
66	5	131	63% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
67	5	121	-	135	0	0	100.00	0	0	100.00	0	0	100.00
68	5	132	Poly A (6); Poly T (8)	135	0	0	100.00	0	0	100.00	0	0	100.00
69	7	133	-	135	0	0	100.00	0	0	100.00	0	0	100.00
70	7	120	60% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
71	7	135	-	135	0	0	100.00	0	0	100.00	0	0	100.00
72	7	126	Poly A (5); 59% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
73	7	134	-	135	0	0	100.00	0	0	100.00	0	0	100.00
74	7	122	Poly C (5); 63% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
75	7	127	59% GC; SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
76	7	123	-	135	0	0	100.00	0	0	100.00	0	0	100.00
77	7	125	-	135	0	0	100.00	0	0	100.00	0	0	100.00
78	7	133	Poly A (5); Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
79	7	116	-	135	0	0	100.00	0	0	100.00	0	0	100.00
80	7	135	-	135	0	0	100.00	0	0	100.00	0	0	100.00
81	7	118	-	135	0	0	100.00	0	0	100.00	0	0	100.00
82	7	136	67% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
83	7	131	58% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
84	7	119	Poly G (6); 61 GC	135	0	0	100.00	0	0	100.00	0	0	100.00
85	7	122	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
86	7	123	Poly A (6)	135	0	0	100.00	0	0	100.00	0	0	100.00
87	8	127	60% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
88	8	129	57% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
89	9	130	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
90	9	116	-	135	0	0	100.00	0	0	100.00	0	0	100.00
91	9	119	Homologous region on a different chromosome	135	0	0	100.00	0	0	100.00	0	0	100.00
92	9	121	-	135	0	0	100.00	0	0	100.00	0	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of samples run[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					total # of no calls[3]	total # of incorrect calls[4]	% correct calls[5]	total # of no calls	total # of incorrect calls	% correct calls	total # of no calls	total # of incorrect calls	% correct calls
93	9	117	Homologous region on a different chromosome	135	0	0	100.00	0	0	100.00	0	0	100.00
94	9	114	-	135	0	0	100.00	0	0	100.00	0	0	100.00
95	9	129	Poly A (14)	135	0	45[9]	99.22	0	45[9]	99.22	0	45[9]	99.22
96	9	114	Homologous region on a different chromosome; SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
97	9	122	-	135	0	0	100.00	0	0	100.00	0	0	100.00
98	9	127	Poly A (5); Poly C (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
99	9	133	-	135	0	0	100.00	0	0	100.00	0	0	100.00
100	9	138	64% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
101	9	139	-	135	0	0	100.00	0	0	100.00	0	0	100.00
102	9	116	-	135	0	0	100.00	0	0	100.00	0	0	100.00
103	9	133	Poly A (5); 57% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
104	9	138	57 GC	135	0	0	100.00	0	0	100.00	0	0	100.00
105	9	136	Poly C (5); 67% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
106	9	118	70% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
107	10	128	62% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
108	10	120	60% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
109	10	139	58% GC; SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
110	10	118	57% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
111	10	123	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
112	10	121	-	135	0	0	100.00	0	0	100.00	0	0	100.00
113	10	129	26% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
114	10	122	-	135	0	0	100.00	0	0	100.00	0	0	100.00
115	10	124	Poly T (5); Homologous region on a different chromosome	135	0	0	100.00	0	0	100.00	0	0	100.00
116	10	135	CA(4)	135	0	0	100.00	0	0	100.00	0	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of samples run[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					total # of no calls[3]	total # of incorrect calls[4]	% correct calls[5]	total # of no calls	total # of incorrect calls	% correct calls	total # of no calls	total # of incorrect calls	% correct calls
117	10	135	Poly A (6); Homologous region on a different chromosome	135	0	0	100.00	0	0	100.00	0	0	100.00
118	10	119	Poly C (5); SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
119	10	125	-	135	0	0	100.00	0	0	100.00	0	0	100.00
120	10	131	-	135	0	0	100.00	0	0	100.00	0	0	100.00
121	10	117	-	135	0	0	100.00	0	0	100.00	0	0	100.00
122	10	116	-	135	0	0	100.00	0	0	100.00	0	0	100.00
123	10	129	58% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
124	11	117	Poly T (10)	135	0	0	100.00	0	0	100.00	0	0	100.00
125	11	117	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
126	11	113	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
127	11	129	-	135	0	0	100.00	0	0	100.00	0	0	100.00
128	11	121	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
129	11	123	-	135	0	0	100.00	0	0	100.00	0	0	100.00
130	11	127	Poly A (6)	135	0	0	100.00	0	0	100.00	0	0	100.00
131	11	136	Poly T (6)	135	0	0	100.00	0	0	100.00	0	0	100.00
132	11	132	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
133	11	115	-	135	0	0	100.00	0	0	100.00	0	0	100.00
134	11	117	Poly T (8); 19% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
135	11	134	Poly A (5); Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
136	11	131	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
137	11	133	SNV; 26% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
138	11	137	Poly T (8); SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
139	11	131	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
140	12	131	-	135	0	0	100.00	0	0	100.00	0	0	100.00
141	12	128	-	135	0	0	100.00	0	0	100.00	0	0	100.00
142	12	133	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
143	12	136	-	135	0	0	100.00	0	0	100.00	0	0	100.00
144	12	124	-	135	0	0	100.00	0	0	100.00	0	0	100.00
145	12	122	59% GC	135	0	0	100.00	0	0	100.00	0	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of samples run[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					total # of no calls[3]	total # of incorrect calls[4]	% correct calls[5]	total # of no calls	total # of incorrect calls	% correct calls	total # of no calls	total # of incorrect calls	% correct calls
146	13	122	-	135	0	0	100.00	0	0	100.00	0	0	100.00
147	13	116	Poly C (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
148	13	133	-	135	0	0	100.00	0	0	100.00	0	0	100.00
149	13	117	SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
150	13	124	Poly T (6)	135	0	0	100.00	0	0	100.00	0	0	100.00
151	13	123	Poly T (5); 26% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
152	13	115	Poly A (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
153	13	125	-	135	0	0	100.00	0	0	100.00	0	0	100.00
154	13	121	-	135	0	0	100.00	0	0	100.00	0	0	100.00
155	13	123	-	135	0	0	100.00	0	0	100.00	0	0	100.00
156	13	114	-	135	0	0	100.00	0	0	100.00	0	0	100.00
157	13	119	-	135	0	0	100.00	0	0	100.00	0	0	100.00
158	14	122	58% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
159	16	122	-	135	0	0	100.00	0	0	100.00	0	0	100.00
160	16	121	-	135	0	0	100.00	0	0	100.00	0	0	100.00
161	16	123	Poly C (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
162	17	119	-	135	0	0	100.00	0	0	100.00	0	0	100.00
163	17	119	61% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
164	17	135	-	135	0	0	100.00	0	0	100.00	0	0	100.00
165	17	116	Poly C (6); 60% GC; SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
166	17	123	-	135	0	0	100.00	0	0	100.00	0	0	100.00
167	17	116	62% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
168	17	118	Poly C (5); 65% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
169	17	129	-	135	0	0	100.00	0	0	100.00	0	0	100.00
170	17	131	Poly G (6); 67% GC; SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
171	17	127	61% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
172	17	118	Poly C (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
173	17	138	61% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
174	17	131	58% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
175	18	112	-	135	0	0	100.00	0	0	100.00	0	0	100.00
176	18	124	-	135	0	0	100.00	0	0	100.00	0	0	100.00
177	18	134	Poly A (6)	135	0	0	100.00	0	0	100.00	0	0	100.00

Amp-licon	Chr.	Ana-lyzed frag-ment size[1]	Amplicon Genomic Content	# of samples run[2]	MiSeqDx 1			MiSeqDx 2			MiSeqDx 3		
					total # of no calls[3]	total # of incorrect calls[4]	% correct calls[5]	total # of no calls	total # of incorrect calls	% correct calls	total # of no calls	total # of incorrect calls	% correct calls
178	18	129	-	135	0	0	100.00	0	0	100.00	0	0	100.00
179	18	133	-	135	0	0	100.00	0	0	100.00	0	0	100.00
180	18	118	-	135	0	0	100.00	0	0	100.00	0	0	100.00
181	18	114	60% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
182	18	118	-	135	0	0	100.00	0	0	100.00	0	0	100.00
183	19	122	Poly G (6); 66% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
184	19	139	64% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
185	19	131	67% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
186	19	141	59% GC; Homologous region on a different chromosome	135	0	0	100.00	0	0	100.00	0	0	100.00
187	19	121	Poly C (5); 72% GC; Homologous region on a different chromosome	135	0	0	100.00	0	0	100.00	0	0	100.00
188	19	138	58% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
189	19	123	64% GC	135	0	0	100.00	0	0	100.00	0	0	100.00
190	19	138	-	135	0	0	100.00	0	0	100.00	0	0	100.00
191	20	117	Poly T (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
192	22	136	Poly A (7)	135	0	0	100.00	0	0	100.00	0	0	100.00
193	22	122	Poly A (5); Poly C (5)	135	0	0	100.00	0	0	100.00	0	0	100.00
194	22	122	62% GC; SNV	135	0	0	100.00	0	0	100.00	0	0	100.00
195	22	119	66% GC	135	0	0	100.00	0	0	100.00	0	0	100.00

[1] Analyzed fragment size means the size of the sequenced genomic region in bases, not including target-specific primers.

[2] Number of samples is calculated from 9 runs of 15 samples (11 samples run once and 2 samples run twice)

[3] Total number of no calls is the combined number of no calls obtained for all 135 runs at that amplicon

[4] Total number of incorrect calls is the combined number of incorrect calls obtained

for all 135 runs at that amplicon

[5] % correct calls equals the correct call rate for all of the bases in the amplicon, where the correct call for the SNV or indel is based on the well characterized reference database and the correct call for the bases in the remainder of the amplicon sequence is based on comparison to . This column may have more than one expected result for a given amplicon if some samples are expected to have an indel and some are not, e.g., amplicon 9.

[6] Amplicon 1 had a number of bases whose genotype could not be called: 12 bases in 1/9 runs in NA12881; 1 base in 2/9 runs and 3 bases in 1/9 runs in NA12886; 20 bases in 1/9 runs and 26 bases in 1/9 runs in NA12888. This is due to low coverage at no call bases in those runs, where the average sequencing depth was 33.2, with a minimum of 21 and maximum of 52.

[7] When no-calls are not included in the calculation, the correct call rate is 100%.

[8] Amplicon 9 has a homopolymer run of 14 A's according to the human genome reference sequence build 19. However, the well characterized reference information for 7 out of 13 samples have 13 A's in this homopolymer run. In these 7 samples, this one base pair deletion is called a false negative, and is called as false negative reproducibly in all nine runs..

[9] Amplicon 95 has a homopolymer run of 14 A's according to human genome reference sequence build 19. . However, the well characterized reference information sequences for 13 out of 13 samples have 15 A's in this homopolymer run. In these 13 samples, this one base pair insertion is 100% reproducibly not called (i.e., it is false negative).

Study 1 results below present results each sample compounded from all nine runs into one column. The results displayed are solely for the single nucleotide variants and insertions/deletions results versus the reference database sequence. This analysis demonstrated that the results for the variants were reproducible across nine runs for these samples.

Sample panel indicating sample number for the study and associated sample name/ID:

DNA #	DNA Sample ID
1	NA12877
2	NA12878
3	NA12879
4	NA12880
5	NA12881
6	NA12882
7	NA12883

DNA #	DNA Sample ID
8	NA12884
9	NA12885
10	NA12886
11	NA12887
12	NA12888
13	NA12893

Reproducibility results for SNVs and Indels per sample:

DNA #	# Runs per sample	Single Nucleotide Variants (SNVs)				Insertions\Deletions (Indels)			
		# of SNVs	# Called Correctly	# of False Positives[1]	# of False Negatives[2]	# of Indels	# Called Correctly	# of False Positives[1]	# of False Negatives[2]
1[3]	18	16	16	0	0	3	1	0	2
2[3]	18	17	17	0	0	2	0	0	2
3	9	18	18	0	0	3	1	0	1
4	9	17	17	0	0	3	1	0	2
5	9	19	19	0	0	3	1	0	2
6	9	15	15	0	0	1	0	0	1
7	9	22	22	0	0	2	1	0	1
8	9	19	19	0	0	2	1	0	1
9	9	17	17	0	0	2	0	0	2
10	9	19	19	0	0	3	1	0	2
11	9	18	18	0	0	1	0	0	1
12	9	22	22	0	0	2	1	0	1
13	9	17	17	0	0	3	1	0	2

[1] False Positive = Variant called by MiSeqDx sequencing run but not in reference database

[2] False Negative = Variant in reference database but not called in MiSeqDx sequencing run.

[3] Samples NA12877 and NA12878 were run in duplicate. Replicate samples generated identical results

Study2: A reproducibility study performed with a representative assay, the Illumina MiSeqDx Cystic Fibrosis 139 Variant Assay, included a subset of CFTR clinically significant genetic variations analyzed with the MiSeq Reporter (MSR) software v2.2.29.2 using the MiSeqDx Platform targeted DNA sequencing workflow. The blinded study used 3 trial sites and 2 operators at each site. Two well-characterized panels of 46 samples each were tested by each of the operators at each site for a total

of 810 calls per site. The panels contained a mix of genomic DNA from cell lines with known variants in the *CFTR* gene, as well as leukocyte-depleted blood spiked with cell lines with known variants in the *CFTR* gene. The blood samples were provided to allow incorporation of the extraction steps used to prepare gDNA that serves as the primary input for the assay workflow. The sample pass rate, defined as the number of samples passing QC metrics on the first attempt, was 99.9%. All test results are based on initial testing. No repeat testing was done for the reproducibility study.

Sample Genotype	Total calls per site	Positive Agreeing Calls (Variants)			Negative Agreeing Calls (Wild type)			# Miscalls	# No Calls	Positive Agreement (%)	Negative Agreement (%)	Overall Agreement (%)
		Site 1	Site 2	Site 3	Site 1	Site 2	Site 3					
S549N (HET)	810	6	6	6	804	804	804	0	0	100	100	100
1812-1 G->A (HET)	810	6	6	6	804	804	804	0	0	100	100	100
Q493X/F508del (HET)	810	12	12	12	798	798	798	0	0	100	100	100
F508del/2184delA (HET)[1]	810	12	12	12	797	798	798	0	1	100	100	100
Y122X/R1158X (HET)[2]	810	12	11	12	798	664	798	0	135	97.22	94.40	94.44
F508del/2183AA>G	810	12	12	12	798	798	798	0	0	100	100	100
R75X (HET)	810	6	6	6	804	804	804	0	0	100	100	100
I507del/F508del (HET)	810	12	12	12	798	798	798	0	0	100	100	100
F508del/W1282X (HET)[3]	810	12	11	12	798	797	798	2	0	97.22	99.96	99.9
F508del/3272-26A>G (HET) [3]	810	12	11	12	798	797	798	2	0	97.22	99.96	99.9
F508del/3849+10kbC>T (HET)	810	12	12	12	798	798	798	0	0	100	100	100
621+1G>T/3120+1G>A (HET)	810	12	12	12	798	798	798	0	0	100	100	100
E60X/F508del (HET)	810	12	12	12	798	798	798	0	0	100	100	100
M1101K (HET)	810	6	6	6	804	804	804	0	0	100	100	100
M1101K (HOM)	810	6	6	6	804	804	804	0	0	100	100	100
F508del (HOM) [4]	828	6	6	6	822	822	822	0	0	100	100	100
F508del/3659delC (HET)	810	12	12	12	798	798	798	0	0	100	100	100

Sample Genotype	Total calls per site	Positive Agreeing Calls (Variants)			Negative Agreeing Calls (Wild type)			# Miscalls	# No Calls	Positive Agreement (%)	Negative Agreement (%)	Overall Agreement (%)
		Site 1	Site 2	Site 3	Site 1	Site 2	Site 3					
R117H/F508del (HET) [5]	816	18	18	18	798	798	798	0	0	100	100	100
621+1G>T/711+1G>T (HET)	810	12	12	12	798	798	798	0	0	100	100	100
G85E/621+1G>T (HET)	810	12	12	12	798	798	798	0	0	100	100	100
A455E/F508del (HET)	810	12	12	12	798	798	798	0	0	100	100	100
F508del/R560T (HET)	810	12	12	12	798	798	798	0	0	100	100	100
F508del/Y1092X (C>A) (HET)	810	12	12	12	798	798	798	0	0	100	100	100
N1303K (HET)	810	6	6	6	804	804	804	0	0	100	100	100
G542X (HOM)	810	6	6	6	804	804	804	0	0	100	100	100
G542X (HET)	810	6	6	6	804	804	804	0	0	100	100	100
G551D/R553X (HET)	810	12	12	12	798	798	798	0	0	100	100	100
3849+10kbC>T (HOM)	810	6	6	6	804	804	804	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
F508del (HET)	810	6	6	6	804	804	804	0	0	100	100	100
1717-1G>A (HET)	810	6	6	6	804	804	804	0	0	100	100	100
R1162X (HET)	810	6	6	6	804	804	804	0	0	100	100	100
R347P/G551D (HET)	810	12	12	12	798	798	798	0	0	100	100	100
R334W (HET)	810	6	6	6	804	804	804	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
G85E (HET)	810	6	6	6	804	804	804	0	0	100	100	100
I336K (HET)	810	6	6	6	804	804	804	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
F508del/3849+10kbC>T (HET)	810	12	12	12	798	798	798	0	0	100	100	100
621+1G>T/3120+1G>A (HET)	810	12	12	12	798	798	798	0	0	100	100	100
F508del/3659delC (HET)	810	12	12	12	798	798	798	0	0	100	100	100
R117H/F508del (HET)[5]	816	18	18	18	798	798	798	0	0	100	100	100
G85E/621+1G>T (HET)	810	12	12	12	798	798	798	0	0	100	100	100
A455E/F508del (HET)	810	12	12	12	798	798	798	0	0	100	100	100

Sample Genotype	Total calls per site	Positive Agreeing Calls (Variants)			Negative Agreeing Calls (Wild type)			# Miscalls	# No Calls	Positive Agreement (%)	Negative Agreement (%)	Overall Agreement (%)
		Site 1	Site 2	Site 3	Site 1	Site 2	Site 3					
N1303K (HET)	810	6	6	6	804	804	804	0	0	100	100	100
G551D/R553X (HET)	810	12	12	12	798	798	798	0	0	100	100	100
2789+5G>A (HOM)	810	6	6	6	804	804	804	0	0	100	100	100
F508del/1898+1G>A (HET)	810	12	12	12	798	798	798	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
F508del/2143delT (HET)	810	12	12	12	798	798	798	0	0	100	100	100
3876delA (HET)	810	6	6	6	804	804	804	0	0	100	100	100
3905insT (HET)	810	6	6	6	804	804	804	0	0	100	100	100
394delTT (HET)	810	6	6	6	804	804	804	0	0	100	100	100
F508del (HET)	810	6	6	6	804	804	804	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
F508del (HET)	810	6	6	6	804	804	804	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
L206W (HET)	810	6	6	6	804	804	804	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
G330X (HET)	810	6	6	6	804	804	804	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
R347H (HET)	810	6	6	6	804	804	804	0	0	100	100	100
1078delT (HET)	810	6	6	6	804	804	804	0	0	100	100	100
G178R/F508del (HET)	810	12	12	12	798	798	798	0	0	100	100	100
S549R (c.1647T>G) (HET)	810	6	6	6	804	804	804	0	0	100	100	100
S549N (HET)	810	6	6	6	804	804	804	0	0	100	100	100
W846X (HET)	810	6	6	6	804	804	804	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
E92X/F508del (HET)	810	12	12	12	798	798	798	0	0	100	100	100
621+1G>T/1154insTC (HET)[6]	810	12	12	12	798	798	797	0	1	100	99.96	99.96
G542X (HET)	810	6	6	6	804	804	804	0	0	100	100	100

Sample Genotype	Total calls per site	Positive Agreeing Calls (Variants)			Negative Agreeing Calls (Wild type)			# Miscalls	# No Calls	Positive Agreement (%)	Negative Agreement (%)	Overall Agreement (%)
		Site 1	Site 2	Site 3	Site 1	Site 2	Site 3					
F508del (HET)	810	6	6	6	804	804	804	0	0	100	100	100
F508del (HET)[2]	810	6	5	6	804	670	804	0	135	94.44	94.44	94.44
F508del (HET)	810	6	6	6	804	804	804	0	0	100	100	100
621+1G>T/A455E (HET)	810	12	12	12	798	798	798	0	0	100	100	100
1812-1 G->A (HET)	810	6	6	6	804	804	804	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
F508del/R553X (HET)	810	12	12	12	798	798	798	0	0	100	100	100
F508del/G551D (HET)	810	12	12	12	798	798	798	0	0	100	100	100
R347P/F508del (HET)	810	12	12	12	798	798	798	0	0	100	100	100
R117H/F508del (HET)	816	18	18	18	798	798	798	0	0	100	100	100
I507del (HET)	810	6	6	6	804	804	804	0	0	100	100	100
2789+5G>A (HOM)	810	6	6	6	804	804	804	0	0	100	100	100
F508del/1898+1G>A (HET)	810	12	12	12	798	798	798	0	0	100	100	100
WT	810	0	0	0	810	810	810	0	0	N/A	100	100
F508del/2143delT (HET)	810	12	12	12	798	798	798	0	0	100	100	100
3905insT (HET)	810	6	6	6	804	804	804	0	0	100	100	100
394delTT (HET)	810	6	6	6	804	804	804	0	0	100	100	100
F508del (HET)	810	6	6	6	804	804	804	0	0	100	100	100

[1] Variant N1303K was not called in this sample

[2] One replicate of samples 5 and 75 had a 0% call rate. Further investigation indicates that samples may not have been added to the sample plate prior to library preparation

[3] Evidence indicates that samples 9 and 10 were likely switched by the operator prior to library preparation.

[4] I506V, I507V, F508C not present in sample

[5] Sample also has the (TG)10 (T)9/(TG) 12(T)5 variant

[6] Variant M1V was not called in these samples

c. Linearity:

Not applicable.

d. Carryover:

A study was performed to evaluate the potential for inter-run and intra-run sample carryover. Intra-run sample carryover tested the system in the most challenging scenario for sample carryover within a single sequencing run. One 48-sample library composed of two samples with unique variants was setup in a checkerboard matrix pattern at alternating high (500 ng) and low (100 ng) concentrations, along with 4 NTC's (no template control samples).

Inter-run sample carryover tested the system for sample carryover between successive sequencing runs. Two libraries were prepared; each library was composed of 47 replicates of a single genomic DNA sample and one NTC. Each library used a different sample from the other.

For both inter- and intra-run, sample carryover was determined by measuring the error rate at the position of variant calls for all samples used in the study. Acceptance criteria for this study were reviewed and deemed acceptable. This study met the sponsor's acceptance criteria and demonstrated that there is minimal to no carryover on the MiSeqDx.

e. Interfering Substances:

To assess the impact of interfering substances on the MiSeqDx Platform, a representative assay (Assay 2) was evaluated in the presence and absence of potential interferents. Eight whole blood samples representing eight unique genotypes were utilized in the study. Four endogenous interfering substances (bilirubin, cholesterol, hemoglobin, and triglycerides) were tested by spiking them into the blood specimens prior to DNA extraction. To assess interference resulting from blood collection (short draw), potassium EDTA was spiked into blood samples at two concentrations. The concentration limits for each substance is shown in the following table. Additionally, to assess interference resulting from sample preparation, 15% wash buffer was added to 8 purified genomic DNA with 100% correct calls.

Test Substance	Total Number of Replicates	Concentration Tested in Blood (Upper Limit)	Concentration Tested in Blood (Lower Limit)	Call Rate
Bilirubin	48	684 µmol/L	137 µmol/L	100%
Cholesterol	48	13 mmol/L	2.6 mmol/L	100%
Hemoglobin	48	2 g/L	0.4 g/L	100%
Triglyceride	48	37 mmol/L	7.4 mmol/L	100%
EDTA	48	7.0 mg/mL	2.8 mg/mL	100%

2. Other Supportive Instrument Performance Data Not Covered Above:

DNA extraction study: Three different extraction methods, magnetic bead extraction, alcohol precipitation and silica filter column isolation were evaluated using K_2EDTA anticoagulated whole blood. Fourteen unique blood samples were used in the study representing a range of genotypes from one representative gene. The three DNA extraction methods were tested independently by 2 different operators who each performed 3 runs per extraction method. Each extraction was performed by each operator on different days. The DNA concentration and A260/A280 ratio of the extracted gDNA samples was determined using spectrophotometry. The total sample size for each extraction method in this study was 168 (14 samples x 2 operators/extraction method x 3 runs/operator x 2 replicates/extracted gDNA sample).

Extraction Method	Number of samples tested	Call Rate	Accuracy	Sample First Pass Rate*
Alcohol Precipitation (Qiagen Gentra PureGene)	168	100%	100%	100%
Silica Filter Column Isolation (Qiagen Blood Mini)	168	100%	100%	100%
Magnetic Bead extraction (Biomerieux Easy Mag)	168	100%	100%	100%

DNA input study: The DNA input range for the MiSeqDx Platform was evaluated by performing a serial dilution study using 14 representative DNA samples containing 16 unique single gene variants. Each sample was tested in duplicate at 9 DNA input levels ranging from 1250 ng to 1 ng (1250 ng, 500 ng, 250 ng, 100 ng, 50 ng, 25 ng, 10 ng, 5 ng, and 1 ng). For determination of accuracy, sample genotypes were compared to bidirectional Sanger sequencing data and the deletions were compared to PCR assay. 1250 ng and 25 ng were identified as the upper and lower bound for DNA input respectively as they had ≥95% sample first pass rate with no incorrect calls (100% accuracy and call rate).

DNA inputs of 1250 ng, 250 ng, and 100 ng were further tested with 4 representative DNA samples and 20 replicates per DNA input level for each sample (n=4*20=80 samples), while the lower bound of 25 ng was tested with 14 samples, 20 replicates for each sample (n=14*20=280 samples). The accuracy and sample first pass rate was 100% at all DNA input levels. There were 2 no calls overall observed at the 25 ng DNA input level, with sample call rates of 99.26%.

Thermal cycler study: Three different commercially available thermal cyclers were evaluated using the representative assay, the Illumina MiSeqDx Cystic Fibrosis 139 Variant Assay. Thermal cycles are used in the library preparation. Three unique sample sets were processed through all three thermal cyclers across 3 days. This enabled performance assessment across different thermal cyclers on different days. Each sample set was processed in triplicate each day (i.e. one replicate per thermal cycler). Acceptance criteria for this study were reviewed and deemed acceptable. This study met the sponsor's acceptance criteria and demonstrated that any commercially available thermal cycler would be adequate for library preparation for use with the MiSeqDx.

Sample indexing study: Sample index primers are used in the kit to assign a unique barcode to each sample DNA, allowing the ability to pool multiple samples together into a single sequencing run.

A total of 96 samples indexes were tested with Assay 2 using 8 unique DNA samples to verify the ability of the assay to consistently make a genotyping call for a given sample across different indexing primer combinations. Each sample was tested with 12 different indexing primer combinations. Sample results were compared against bidirectional Sanger sequencing data for all positions/variants. Reproducibility and accuracy were 100% for all sample/index primer combinations.

Specimen Storage: To verify the storage conditions and handling of blood samples for use with the MiSeqDx test system, six K_2EDTA anti-coagulated blood samples were divided to six aliquots, one aliquot of each blood sample were stored under 6 different conditions: 2°C to 8°C for 1 day; -15°C to -25°C for 1 day; 2°C to 8°C for 30 days; -15°C to -25°C for 30 days; room temperature (20-25°C) for 7 days; and controlled room temperature (30°C) for 7 days. Genomic DNA was isolated from each aliquot using a commonly used commercial DNA extraction kit. All extractions were performed by a single operator. The extracted gDNA samples were stored at -15°C to -25°C until the libraries were prepared and sequenced.

The impact of repeated freeze-thaws on gDNA samples were tested by subjecting 15 DNA samples to 6 freeze thaw cycles.

Library preparations for both the samples from both the specimen storage and gDNA freeze-thaw studies were performed at the same time point. The samples from a single library preparation were pooled into one run of 48 samples and a second run of 32 samples prior to sequencing. Impact on call rate, reproducibility, and sample first pass rate were determined for each sample as compared to a respective control sample. No miscalls or no calls were observed for any of the specimens and demonstrated that the blood and gDNA storage conditions tested did not affect assay results.

K. Proposed Labeling:

Labeling satisfies the requirements of 21 CFR 809.10, 21 CFR 801.109, including an appropriate prescription statement as required by 21 CFR 801.109(b), and the special controls for this type of device.

L. Other Supportive Instrument Characteristics Data Not Covered In The "Performance Characteristics" Section above:

None.

M. Identified Potential Risks and Required Mitigation Measures:

Identified Potential Risk	Required Mitigation Measure
Inaccurate test results due to unavailability of necessary components of the instrument system	The labeling for the instrument system must reference pre-analytical and analytical reagents to be used with the instrument system and include or reference legally marketed analytical software that includes sequence alignment and variant calling functions, to be used with the instrument system.
Inaccurate results due to unknown performance of the instrument system	The labeling for the instrument system must include a description of the following information: i) The specimen type(s) validated as an appropriate source of nucleic acid for this instrument. ii) The type(s) of nucleic acids (e.g., germline DNA, tumor DNA) validated with this instrument. iii) The type(s) of sequence variations (e.g. single nucleotide variants, insertions, deletions) validated with this instrument.

Identified Potential Risk	Required Mitigation Measure
	iv) The type(s) of sequencing (e.g., targeted sequencing) validated with this instrument. v) The appropriate read depth for the sensitivity claimed and validation information supporting those claims. vi) The nucleic acid extraction method(s) validated for use with the instrument system. vii) Limitations must specify the types of sequence variations that the instrument cannot detect with the claimed accuracy and precision (e.g., insertions or deletions larger than a certain size, translocations). viii) Performance characteristics of the instrument system must include: *A)* Reproducibility data generated using multiple instruments and multiple operators, and at multiple sites. Samples tested must include all claimed specimen types, nucleic acid types, sequence variation types, and types of sequencing. Variants queried shall be located in varying sequence context (e.g., different chromosomes, GC-rich regions). Device results shall be compared to reference sequence data with high confidence. *B)* Accuracy data for all claimed specimen types and nucleic acid types generated by testing a panel of well-characterized samples to query all claimed sequence variation types, types of sequencing, and sequences located in varying

Identified Potential Risk	Required Mitigation Measure
	sequence context (e.g., different chromosomes, GC-rich regions). The well-characterized sample panel shall include samples from at least two sources that have highly confident sequence based on well-validated sequencing methods. At least one reference source shall have sequence generated independently of the manufacturer with respect to technology and analysis. Percent agreement and percent disagreement with the reference sequences must be described for all regions queried by the instrument. *C)* If applicable, data describing endogenous or exogenous substances that may interfere with the instrument system. *D)* If applicable, data demonstrating the ability of the system to consistently generate an accurate result for a given sample across different indexing primer combinations. ix) The upper and lower limit of input nucleic acid that will achieve the claimed accuracy and reproducibility. Data supporting such claims must also be summarized.

N. Benefit/Risk Analysis:

Summary

Summary of the Benefit(s)	• This is a tool for clinical laboratories that can provide accurate and reproducible high throughput genomic sequencing of genomic regions of interest at greater sequencing depth than current sequencing technology. • No other instruments are available for high throughput genomic sequence analysis. There is an unmet medical and public health need for a well-validated IVD labelled high throughput genomic sequence analyzer.
Summary of the Risk(s)	• Patients are subject to blood specimen collection, which is a standard procedure in clinical care and carries minimal risk. • Risk is related to inaccurate test results as follows: False positive: The risks to the individual of a false positive result could include unnecessary testing or treatment related to an inaccurate test result. Often, the result from this test would be used with results from other diagnostic tests and clinical signs and symptoms to identify the genetic cause or contribution for a patient's disease or condition. False negative: The risks to the individual of a false negative result due to an inaccurate test result could delay further evaluation and appropriate therapy which will vary depending on the disease or condition. Public Health Risk from Incorrect Test Results: • The consequences to public health for both false positive and false negative results are similar.
Summary of Other Factors	Not applicable.
Conclusions Do the probable benefits outweigh the probable risks?	
Given robust analytical performance characteristics and risk mitigation (i.e. extensive performance data provided in the labeling), the probable benefits to both the individual and public health outweigh the probable risks of this device.	

O. Conclusion:

The information provided in this de novo submission is sufficient to classify this device into class II under regulation 21 CFR 862.2265. FDA believes that special controls, along with the applicable general controls, provide reasonable assurance of the safety and

effectiveness of the device type. This device, and similar devices, is classified under the following:

Product Code: PFF

Device Type: High throughput genomic sequence analyzer for clinical use

Class: II (special controls)

Regulation: 21 CFR 862.2265

(a)*Identification.* A high throughput genetic sequence analyzer for clinical use is an analytical instrument system intended to generate, measure and sort signals in order to analyze nucleic acid sequences in a clinical sample. The device may include a signal reader unit; reagent handling, dedicated instrument control, and other hardware components; raw data storage mechanisms; data acquisition software; and software to process detected signals.

(b)*Classification.* Class II (special controls). A high throughput genetic sequence analyzer for clinical use must comply with the following special controls:

1) The labeling for the instrument system must reference legally marketed pre-analytical and analytical reagents to be used with the instrument system and include or reference legally marketed analytical software that includes sequence alignment and variant calling functions, to be used with the instrument system.
2) The labeling for the instrument system must Include a description of the following information:
    i) The specimen type(s) validated as an appropriate source of nucleic acid for this instrument.
    ii) The type(s) of nucleic acids (e.g., germline DNA, tumor DNA) validated with this instrument.
    iii) The type(s) of sequence variations (e.g. single nucleotide variants, insertions, deletions) validated with this instrument.
    iv) The type(s) of sequencing (e.g., targeted sequencing) validated with this instrument.
    v) The appropriate read depth for the sensitivity claimed and validation information supporting those claims.
    vi) The nucleic acid extraction method(s) validated for use with the instrument system.

vii) Limitations must specify the types of sequence variations that the instrument cannot detect with the claimed accuracy and precision (e.g., insertions or deletions larger than a certain size, translocations).

viii) Performance characteristics of the instrument system must include:

A) Reproducibility data generated using multiple instruments and multiple operators, and at multiple sites. Samples tested must include all claimed specimen types, nucleic acid types, sequence variation types, and types of sequencing. Variants queried shall be located in varying sequence context (e.g., different chromosomes, GC-rich regions). Device results shall be compared to reference sequence data with high confidence.

B) Accuracy data for all claimed specimen types and nucleic acid types generated by testing a panel of well-characterized samples to query all claimed sequence variation types, types of sequencing, and sequences located in varying sequence context (e.g., different chromosomes, GC-rich regions). The well-characterized sample panel shall include samples from at least two sources that have highly confident sequence based on well-validated sequencing methods. At least one reference source shall have sequence generated independently of the manufacturer with respect to technology and analysis. Percent agreement and percent disagreement with the reference sequences must be described for all regions queried by the instrument.

C) If applicable, data describing endogenous or exogenous substances that may interfere with the instrument system.

D) If applicable, data demonstrating the ability of the system to consistently generate an accurate result for a given sample across different indexing primer combinations.

ix) The upper and lower limit of input nucleic acid that will achieve the claimed accuracy and reproducibility. Data supporting such claims must also be summarized.