EVALUATION OF AUTOMATIC CLASS III DESIGNATION FOR

Helix Laboratory Platform

DECISION SUMMARY

A. DEN Number:

DEN190035

B. Purpose for Submission:

De Novo request for evaluation of automatic class III designation for the Helix Laboratory Platform

C. Measurands:

Single nucleotide variants, insertions, and deletions in whole exome sequence in human genomic DNA

D. Type of Test:

Qualitative whole exome sequencing

E. Applicant:

Helix OpCo, LLC

F. Proprietary and Established Names:

Helix Laboratory Platform

G. Regulatory Information:

1. Regulation section:

   21 CFR 866.6000

2. Classification:

   Class II

3. Product code(s):

   QNC

4. Panel:

   88- Pathology

H. Indications for use:

1. Indications for use:

   The Helix Laboratory Platform is a qualitative in vitro diagnostic device intended for exome sequencing and detection of single nucleotide variants (SNVs) and small insertions and deletions (indels) in human genomic DNA extracted from saliva samples collected with Oragene® Dx OGD-610. The Helix Laboratory Platform is only intended for use with other devices that are germline assays authorized by FDA for use with this device. The device is performed at the Helix laboratory in San Diego, CA.

2. Special conditions for use statement(s):

   For prescription use

   For *in vitro* diagnostic use

3. Special instrument requirements:

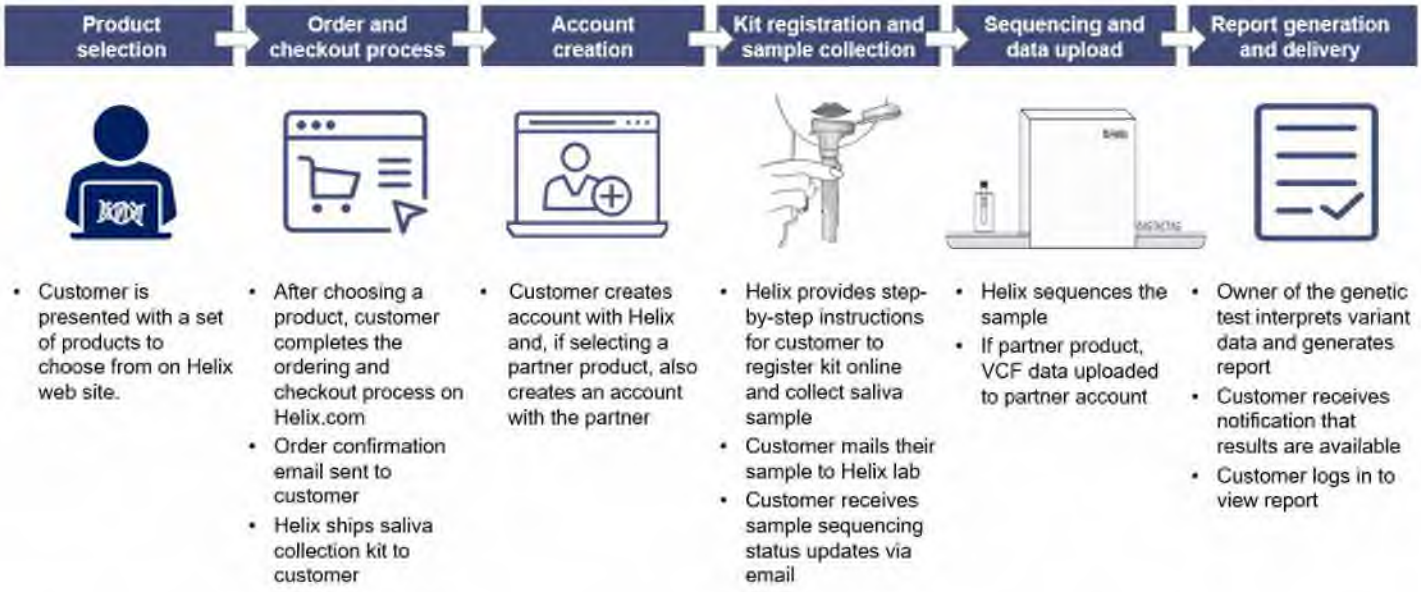
   Illumina HiSeq X Sequencer (qualified by Helix)

   Oragene®·Dx OGD-610 (DNA Genotek, Inc; k192920)

I. Device Description:

The Helix Laboratory Platform (HLP) is a high throughput DNA sequencing platform for targeted sequencing of an individual's whole exome. It is intended for targeted sequencing of an individual's whole exome for use with a genetic test application. Genetic test applications may be third party partner ("Partner") genetic test applications or a Helix genetic test application such as the Helix Genetic Health Risk App (HRA; k192073). The DNA sequence generated by this device is intended as input to clinical germline DNA assays intended for use with this device that have FDA marketing authorization. A brief overview of the commercialized workflow is shown in **Figure 1** (refer to the section titled "Test Principle" for more specific information regarding the commercialized workflow.) HLP consists of a HiSeq sequencing instrument, cBot system, library preparation reagents, sequencing reagents, and data analysis software. The Helix Laboratory Platform also interacts with the Helix Laboratory Automation Systems and Content Mapping Systems which serve as repositories for the data and do not perform data analysis. The test detects single nucleotide variants (SNVs) insertions and deletions (indels) up to 20 base pairs (bp) and is limited to making high-confidence variant calls that meet prespecified quality metrics (i.e., the analytical range) within the reportable range. Sequencing is performed at the Helix clinical laboratory in San Diego, CA.

Figure 1. Workflow Relationship of Helix and "Partner" Genetic Tests Overview



The HLP instruments, reagents and software are qualified by Helix and are comprised of the following:

1. Specimen Collection and DNA Preparation

Saliva is collected in the Oragene®·Dx OGD-610 collection device. At least 1 mL saliva sample must be collected. DNA is extracted using Maxwell HT Saliva DNA Prep. Extracted DNA concentration must be equal or greater than 3.5 ng/µL. A total of 50 to 70ng of DNA input is used for the library preparation. DNA may be stored at -20°C for up to two weeks.

2. Library Preparation

Library preparation consists of five key steps: Fragmentation, Adapter addition by polymerase chain reaction (PCR), Hybridization capture, PCR amplification, and Library normalization. In the first step, purified genomic DNA is broken into small pieces by enzymatic fragmentation, generating overlapping small DNA fragments tiling the entire human genome. Library fragments are size-selected to enable paired-end sequencing with (b)(4) between read pairs. Sequencing adapters are then added by PCR. DNA samples are barcoded using a dual-indexing strategy with barcode sequences that require multiple sequencing errors to become ambiguous. Quality Control (QC) is performed on the purified libraries by DNA quantitation. Purified DNA libraries are combined into a multiplex for enrichment by hybrid-capture. This process includes enrichment using probes to capture the genomic regions of interest. The DNA concentration of the enriched libraries is determined and normalized to allow equal loading of libraries on the flow cell along with clustering reagents into a cBot instrument for cluster generation. The enriched libraries are then clustered on the sequencing flow cell and loaded into the DNA sequencing instrument. Sample concentration of equal or greater than (b)(4) will pass quality control. The cBot System is an automated system that creates clonal clusters from single molecule DNA templates, preparing them for sequencing by synthesis (SBS) on the HiSeq instrument. The cBot isothermally amplifies cDNA fragments that

have been captured by complementary adapter oligonucleotides covalently bound to the surface of flow cells (this is called the Cluster Generation Process).

Different sets of reagents are used for this process. HiSeq DNA sequencing reagents, cluster generation reagents and a flow cell are obtained from Illumina and qualified by Helix. The Helix *Exome+* ® reagents includes library preparation, sample indexing reagents and the capture probe reagents that targets the whole exome.

3. Sequencing and Data Analysis

When cluster generation is complete, the flow cell is inserted with SBS reagents into a HiSeq instrument to perform paired-end DNA sequencing with (b)(4) read lengths. SBS technology uses four fluorescently labeled nucleotides to sequence the (b)(4) of clusters on the flow cell surface in parallel. During sequencing, images are captured from the flow cell by the HiSeq instrument and processed through primary analysis after each sequencing cycle. Primary analysis is performed by the RTA (Real Time Analysis) software without user intervention. This analysis consists of base calling of each cluster at each cycle. Reads are filtered to require a minimum rate of high-quality base calls. Flow cells that do not meet a minimum yield of filtered reads will undergo additional sequencing. The bcl2fastq software de-multiplexes the data and generates sample specific FASTQ files. Next-generation sequencing (NGS) secondary analysis is performed by the Helix bioinformatics pipeline to process base calls into genomic variant calls, starting with FASTQ files. The Helix bioinformatics pipeline is hosted in the cloud and analyzes the targeted human genome sequencing data. The pipeline consists of software analytical tools for short read alignment (Aligner), variant calling (Variant Caller), variant refinement (Variant Refinement), and quality control (Quality Control). The Helix bioinformatics pipeline is a suite of both proprietary and open-source software programs for high-throughput processing and analysis of sequence data. Individual major components of the Helix bioinformatics pipeline are described in **Figure 2**. Quality control metrics of varying depth and resolution are incorporated into several checkpoints along the data processing and analysis pipeline. The Helix bioinformatics inputs are as follows:

- **Metadata:** De-identified sample identifier, self-reported sex and age, and the molecular barcode assigned to the library

- **Sequencing reads:** The set of sequencing reads generated for the sample by the Helix Laboratory Platform

- **Annotation data:** BED files specifying the reportable range, short tandem repeats (STRs), and known variant truth data for control sample concordance analysis

- **Reference genome:** The "Helix Reference Genome" based on the Human Reference Genome Consortium Build 38 assembly with additional alternative contigs as described in detail in the Helix Reference Genome section in the Helix Laboratory Platform User Manual.

Figure 2. High-level modular and functional overview of the bioinformatics pipeline.

The analysis steps include read alignment, variant caller, variant refinement and quality control.

a) *Alignment:* A software aligner aligns short-read nucleotide sequences from FASTQ files to the Helix Reference Genome to generate aligned nucleotide sequences and associated mapping quality data. The nucleotide sequences in the FASTQ files are aligned to the human reference genome GRCh38 as (b)(4) (b)(4) (b)(4) The human reference genome, GRCh38, is a human genome reference generated by the Genome Reference Consortium. Reference for the human oral microbiome is obtained from the Human Oral Microbiome Database (HOMD) and (b)(4) (b)(4), not present within HOMD, from the ATCC® MSA-1002™ panel genomes. Aligned reads are then processed to flag any reads that may be polymerase chain reaction (PCR) or optical (hardware limitation of the sequencer) duplicates. Duplicate marking is necessary to ensure that duplicates are not treated as independent evidence of a genomic sequence. The aligned sequence reads, with the associated base and mapping quality information, are stored in a standard file format, such as a binary alignment mapping file (BAM) or the compressed columnar file format CRAM. These formats are community accepted formats for storing the aligned nucleotide sequences and their corresponding quality scores. They also serve as input files for the Variant Caller.

b) *Variant Calling:* The variant calling software uses existing OTS software. The Variant Caller

is able to detect single nucleotide variants (SNVs) and small (≤ 20 nucleotides) insertion or deletion variants (indels). Genotype calls are stored in an industry standard format such as variant call format (VCF) or its binary format BCF and genomic VCF (gVCF).

The VCF includes the following fields:

- Chromosome
- Start Coordinate
- End Coordinate, if not equal to start + length of reference allele - 1 (1-based inclusive numbering, following the gVCF standard).
- Reference Allele: Reference base
- Alternate Allele: Non-reference base(s) called in the sample, if any
- Genotype: This field indicates the alleles carried by the sample.
- Additional information:
    - Genotype Likelihood: The likelihood of each of the possible genotypes given observed data
    - Genotype Quality: Difference in genotype likelihood between the most likely and second-most likely genotypes. Higher numbers represent higher-confidence genotype calls.
    - Read Depth: The number of reads at a position
    - Allele Depth: The number of reads observed for each allele

Each genomic locus will have a genotype call with an associated genotype likelihood. The genotype likelihood is a standard probability-based score used to evaluate the likelihood of a genotype call at a given locus conditioned on observed data. The observed data includes the aligned sequence data with associated quality data. More specifically, at each locus, the algorithm will count the number of occurrences of each distinct nucleotide from the aligned nucleotide sequences. The number of aligned sequencing reads represents the "coverage" at a given locus. This coverage data is combined with associated mapping quality, base call quality, and other prior information to generate a genotype call with an associated genotype likelihood. The genotype likelihood can be used to determine if there is insufficient information for a confident call (resulting in a no-call). Two separate genotype call files are created by the variant caller. The first file provides reference calls and the second file provides finalized posterior genotype likelihoods.

c) *Variant Refinement:* This software analytic tool uses existing open-source libraries and internal code to perform additional processing on variant and reference calls produced during the Variant Calling step to generate the final observed variant call output. Variant refinement first merges the two variant data files generated by the Variant Caller for the sample into a single variant data file by merging records in the file that represent adjacent reference calls and records in the file that represent overlapping variant calls. This is followed by a sex inference using a statistical model based on the counts of fragments mapping to chromosomes X and Y compared to fragments mapping to the autosomes. This model also accounts for possible chromosome Y loss using the self-reported age range of the individual when it is available. Furthermore, it performs ploidy correction as the Variant Caller assumes that all contigs are diploid. The ploidy correction process will retain variant calls that are inconsistent with the expected ploidy with sufficiently high quality.

d) *Variant Scrubbing:* Metadata for individual files are collapsed. Additional curation against empirically determined "blacklist" and "whitelist" genomic regions with intrinsically poor mapping quality, known polymorphism (e.g. HLA loci), and significant violation of Hardy-Weinberg equilibrium (i.e., non-neutral selective pressures). Haplotype (i.e., phase set) information is calculated.

4. Sequencing Quality Control methods:

Quality Control (QC) methods are incorporated into the Helix bioinformatics pipeline at every step. Quality control methods include Sample-level QC for each region. In addition, the metrics are monitored and used for root cause analysis if samples have a QC failure. These include: (b)(4) (b)(4).

A set of QC metrics are applied to distinct regions of the reportable range. These are collectively referred to as "callability". Callability = 1-No-call rate, is calculated base by base and ensures that the no call rate is acceptable.The sample must pass all thresholds listed in **Table 1**. For example, the Priority white list has a callability metric of (b)(4) which is a no call-rate of (b)(4). If any of the thresholds is not met, the sample(s) are routed back to the appropriate restarting point - library preparation, pooling for enrichment or cluster generation - to be re-tested. Samples may be retested up to (b) times.

Table 1. Secondary Analysis Quality Control (QC) Metrics for Sample Evaluation

Metric	Threshold
(b)(4)	

5. Content Mapping Systems

a) *The Helix Interpretation Module (HIM)* is a mapping tool that does not perform data analytics. Rather, it maps the end user's variant data against the reporting rules (b)(4) (b)(4) to identify and match the correct variant call data with the correct clinical result for that customer (also known as the user). In the case of the Helix Genetic Health Risk (k192073), (b)(4) (b)(4) (b)(4).

The Helix Laboratory Platform provides data to the partner in one of two ways (1) variant call data (VCD) on the genetic variants for a predefined set of genetic coordinates for their genetic test indication. The VCD includes information that is typically contained in a Variant Call

Format (VCF or BCF) file but is transferred securely to the partner via an API. The partner app will review the VCD and make the interpretation and generate the report or, (2) The Helix Laboratory Platform provides the final genotype (b)(4) (b)(4) (b)(4) The partner generates their final report using the content from the JSON file.

In both scenarios, Helix provides data only to the owner of the genetic test. As the sequencing laboratory, the Helix Laboratory Platform will provide data to partners as described under scenario 2.

b) (b)(4)

c) (b)(4)

6. Software (Laboratory Automation Sub Systems and User Interfaces)

The Helix Laboratory Automation sub-systems are used to automate physical sample and sample data tracking. The Helix Laboratory Automation system's purpose is to track samples from the point at which they enter the clinical laboratory, through sample processing and sequencing workflow steps, and to results delivery (either to a customer for the Helix Genetic Health Risk App or to a Helix third party partner). The various sub-systems assess the quality of the runs and provide information about sample queuing to laboratory personnel and partners. The sub-systems include the Accessioning Sub-system, Laboratory Information Management System (LIMS), Pre-and Post-review sub-systems, Data Delivery Review sub-system, Helix Partner API sub-system and Genomic Data Service sub-system.

7. Controls

A Process Control is defined as the DNA reference material with known sequence that can be used to determine the success of a sequencing run. The NA12877 and NA12878 process controls used by HLP are cell lines from the Coriell Institute for Medical Research. These two cell lines are required in the

Helix workflow and are introduced prior to library prep and carried through enrichment and sequencing. The Process Control workflow is as follows:

(b)(4)

(b)(4)

(b)(4)

Table 2. Secondary analysis metrics for process control evaluation

Metric	Threshold
(b)(4)	

8. Definitions of the Genomic Regions Reported by the HLP:

   a) *Reportable Range:* Helix has defined the Reportable Range as all coding exons ("coding region" also referred to as "white list") minus prespecified regions that are not reported (referred to as "black list") which is defined as a list of regions in the HLP that are either not covered by the reagents or are excluded from reporting because they have empirically observed

elevated false positive rates as defined by deviations from Hardy-Weinberg equilibrium. The reportable range excludes the following:

- Regions outside the coding regions.
- Regions that are difficult to map or have poor mapping quality, defined as regions where fewer than 20% of reads have a mapping quality greater than 20.
- Regions with highly polymorphic gene clusters (portions of the HLA loci, immunoglobulin-like receptors on chr19 and Golgin family on chr15)
- Sites with variants called that deviate significantly from Hardy-Weinberg equilibrium
- Regions with ≥ 25% of bases with quality score < 30.
- Regions with high sequence similarity (i.e. pseudogenes)

DNA sequence is aligned to the Helix Reference Genome as previously described (reference genome GRCh38 and select contigs) and variant calling occurs using the Helix bioinformatics pipeline. The Helix *Exome+* reagents are designed to cover the whole exome ("coding region"), but with enhanced sequence coverage over clinically relevant genes in the human genome. In addition to the probes covering overall exon regions, more probes (gap-filling) targeting genomic locations of clinical significance are designed and added to the capture panel. As a result, coverage has been optimized for two subsets of the reportable range. The distinct reporting sets are as follows (**Figure 3**).

i. **Coding:** All targets within the exome; The Helix Laboratory Platform reportable range is limited to coding regions minus the regions excluded as described above.

ii. **Mendeliome:** A subset of the coding region consisting of (b)(4) genes that are associated with human disease with enhanced coverage

iii. **Priority**: A subset of Mendeliome consisting of (b)(4) genes identified as clinically relevant and which has the highest coverage.

Figure 3. Reportable ranges for Coding region and subsets

Region	Region Size* (bp)	Percentage of the Reportable Range
(b)(4)		

The reportable range excludes (b)(4) variants of clinical significance as curated in ClinVar. These variants are excluded because they are either not targeted by the HLP or reside in regions that are difficult to sequence or known to not meet performance criteria.

Analytical Range: All variants and reference records called within the genomic regions specified above are then further filtered based on variant type, sequence context, and variant call quality metrics. The analytical range refers to single nucleotide variants and insertions and deletions up to 20 bp in length that pass quality metrics for reporting. Additionally, each variant record is evaluated based on its proximity to short tandem repeat elements (STR). Specific rules are applied for the location of variants and the sizes of the STRs which will determine whether the variant is reportable. All variant calls are required to meet Helix's variant level QC metrics applied to every locus within the reportable range (**Table 3**).

Table 3. Quality Metric Thresholds

Metric	Threshold Required for Reporting
Depth of Coverage (DP)	(b)(4)
Genotype quality (GQ)	
Allele fraction of heterozygous calls	
Fisher's strand bias score (FS)	
Root mean square of mapping quality (MQ)	

Only loci that meet the above criteria are reported. There is variation on which loci are reportable on a run to run basis.

Callability: Callability refers to the proportion of the coding region that is reportable to ensure that the no-call rate within the likely reporting windows is acceptable. Callability incorporates base quality, alignment/mapping quality, and minimum coverage levels and determines the number (%) of base calls that pass the QC metrics for a genotype call. The callability thresholds for the reportable regions and subregions are reported in **Table 1**.

The Helix Laboratory Platform is intended for use as part of a genetic test that has been validated for specific clinical indications. Each genetic test that uses the HLP will have indication specific callability assessments prior to commercialization so that the expected no-call rate is acceptable for the specific indication.

J. Standard/Guidance Document Referenced (if applicable):

- Guidance for Industry and FDA Staff: Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices; May 11, 2005
- General Principles of Software Validation; Final Guidance for Industry and FDA Staff; January 11, 2002
- Guidance for Industry – Cybersecurity for Networked Medical Devices Containing Off-the-Shelf (OTS) Software; January 2005
- Guidance for Industry, FDA Reviewers and Compliance on Off-the-Shelf Software Use in Medical Devices; September 9, 1999

K. Test Principle:

The Helix Laboratory Platform (HLP) is a specimen to variant call file platform that is intended for use with sponsors of legally marketed germline genetic assays validated for use with the HLP. The HLP provides results only to sponsors and does not provide results to ordering physicians or patients (collectively referred to as "customers"). Customers interact with the Helix Laboratory Platform through a web-based application. The customer places an order for a genetic test (that has been cleared for use on the platform), and an FDA cleared specimen collection device, the Oragene®·Dx OGD-610 is shipped to the ordering customer. The customer's saliva sample is returned by the customer to the Helix clinical laboratory for processing. Upon receipt by the Helix clinical laboratory, the sample is accessioned. Samples with insufficient saliva amount will be rejected and a new sample collection kit will be sent to the ordering customer. Samples that pass accessioning enter the extraction process where purified genomic DNA is extracted from saliva samples. Sequencing libraries are prepared, pooled, captured and sequenced. Sequencing reads are aligned to a reference human genome. FASTQ files serve as input files for the Aligner (Helix bioinformatics pipeline) which aligns nucleotide sequences from the FASTQ file to the Helix Reference Genome to generate aligned nucleotide sequences and associated mapping quality data. Aligned sequence data is analyzed to identify samples that do not meet quality thresholds and minimum sequencing coverage on the target regions. Samples that do not pass quality thresholds will undergo additional sequencing. This data is typically stored in a compressed file format and serves as input for the Variant Caller.

As the sequencing lab, Helix provides variant call data (VCD) on the genetic variants for the Genetic Health Risk App (or a future third party partner genetic test validated for use on the platform), but does not provide variant data directly to the ordering customer. The VCD includes information that is typically contained in a Variant Call Format (VCF) file but is transferred securely to the partner via an Application Programming Interface (API). For partner products with clinically relevant intended uses, the reporting lab will review the VCD and make the final decision on interpretation and labeling and generate the final report.

The Helix Laboratory Platform interacts with the [genetic test app] in two primary ways as shown in Figure 1 (Section I device description). First, a [genetic test app] software module called the Helix Report Module (HRM) interacts directly with HLP to obtain a customer's first name, last name, email address, and Helix User Id. This allows the [genetic test app] to input the customer's name in the final report as well as use the email and Helix User Id for secondary confirmation purposes to address customer service inquiries. The second interaction is through the [genetic test app] software module called the Helix Interpretation Module (HIM), which maps the customer's genetic information with the associated clinical interpretation for the final report. (b)(4) (b)(4). The final end user results are displayed as a PDF on the product webpage (also called the Helix Report Module User Interface) upon end user secure login.

For third party partner genetic tests, the partner-defined genotype/interpretation rules will be reviewed by Helix through an internal process to ensure the variants and indications have been cleared by the FDA. The HIM associates the interpretations provided by the partner via the Helix Health Traits API for the specific application and the future third party partner creates their final report.

In summary, the single clinical laboratory provides the following next generation sequencing (NGS)

services to customers who order genetic tests validated for use on the Helix Laboratory Platform: (i) Specimen collection, handling and storage, (ii) *Exome+* sequencing; (iii) storage of the customer's genetic information in a secure cloud database, and (iv) an initial and subsequent in-silico query of the customer's Exome+ data for a predefined set of genetic variants as offered by genetic tests validated for use on the Helix Laboratory Platform. The ordering customer's Exome+ data will be generated with their initial order and additional queries in the future will not require additional sequencing.

Refer to Section I for more details on the device description.

L. Performance Characteristics:

1. Device Optimization and Quality Metrics

   a) *Optimization of variant read depth, allele fraction and callability thresholds:* Parts of the HLP was optimized based on (b)(4) historical reference sample runs to develop per-variant QC metrics. Accuracy (positive predictive value; PPV) was assessed for optimal read depth (**Figure 4**). The data demonstrated that the pre-specified acceptance criteria for PPV was acceptable at a minimum read depth of (b)(with no significant improvement above (b)(. The minimum reads coverage for variant calling was set at 20X

   Figure 4. Optimization of Variant Read Depth.

   Optimal accuracy was assessed for allele fraction (**Figure 5**). Based on the data, the allele fraction range was set between (b)(4)

Figure 5. Optimization and Selection of Allele fraction

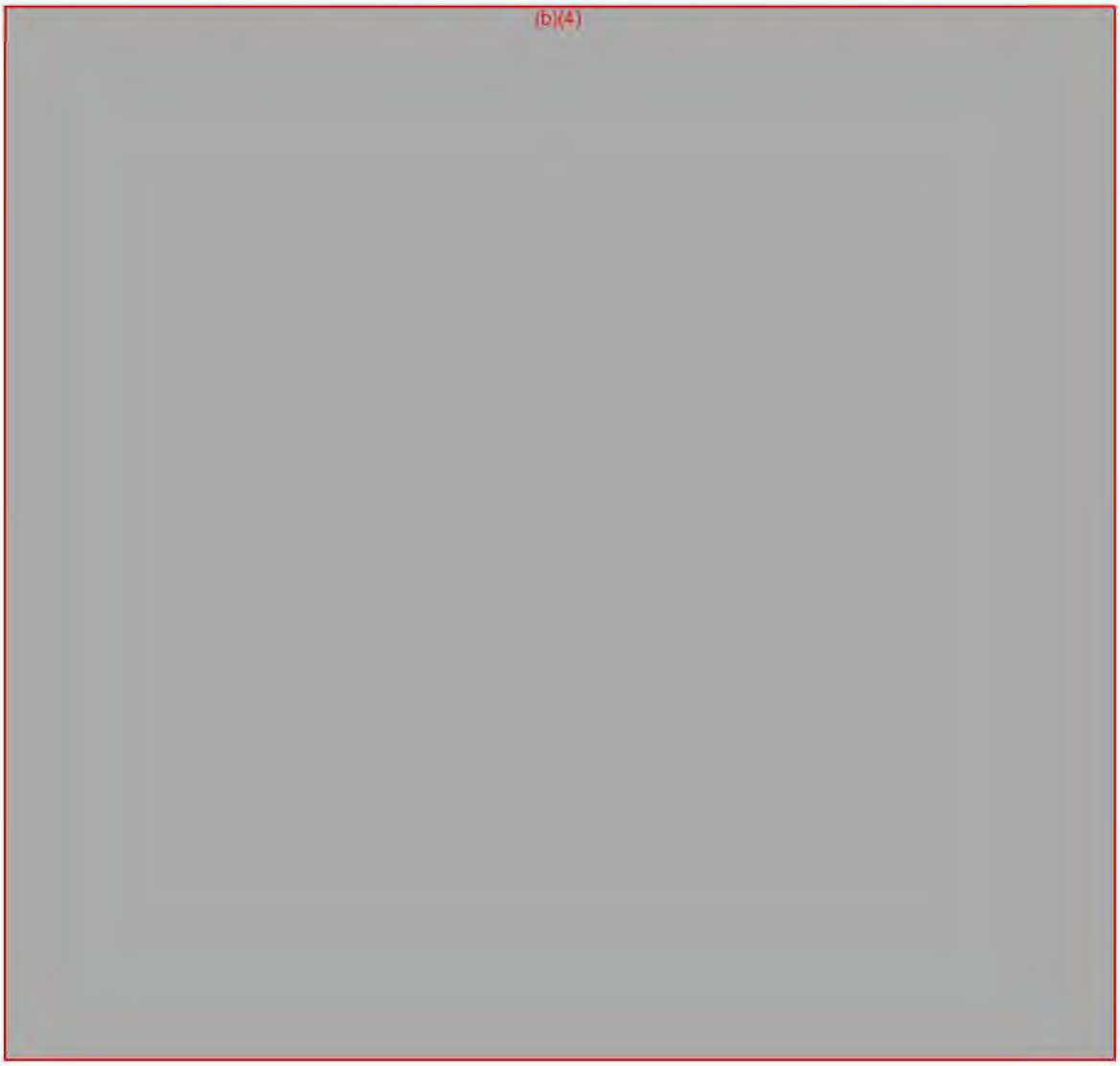
Callability thresholds for (b)(4) (b)(4), respectively. On the Y axis is the number of samples having the callability values within the ranges defined on the x-axis.

Figure 6. Selection of Callability Thresholds

b) *Establishment of filter and QC threshold for variant calling*

i. Genotype quality score (GQ) and filter lists: Reference samples were analyzed with (b)(4) (b)(4) of the bioinformatics pipeline representing different conditions relative to the quality metric criteria (see Section quality metrics above). The control condition retained all of the criteria above and the other (b)(4) were each removed. Figure 7 shows the PPA, PPV, and NRC performance metrics for SNV, insertion, and deletion variant types in each of the conditions compared to control. The results in **Figure 7** show that the removal of any one of the criteria results in reduced accuracy and that retaining all criteria (the control condition) results in the highest accuracy. The (b)(4) has the greatest effect on PPA and NRC across all variant types while the DP threshold has the greatest effect on PPV. (Refer to section below for definition of PPA, PPV and NRC).

Figure 7. Accuracy compared to reference datasets under various filtering conditions



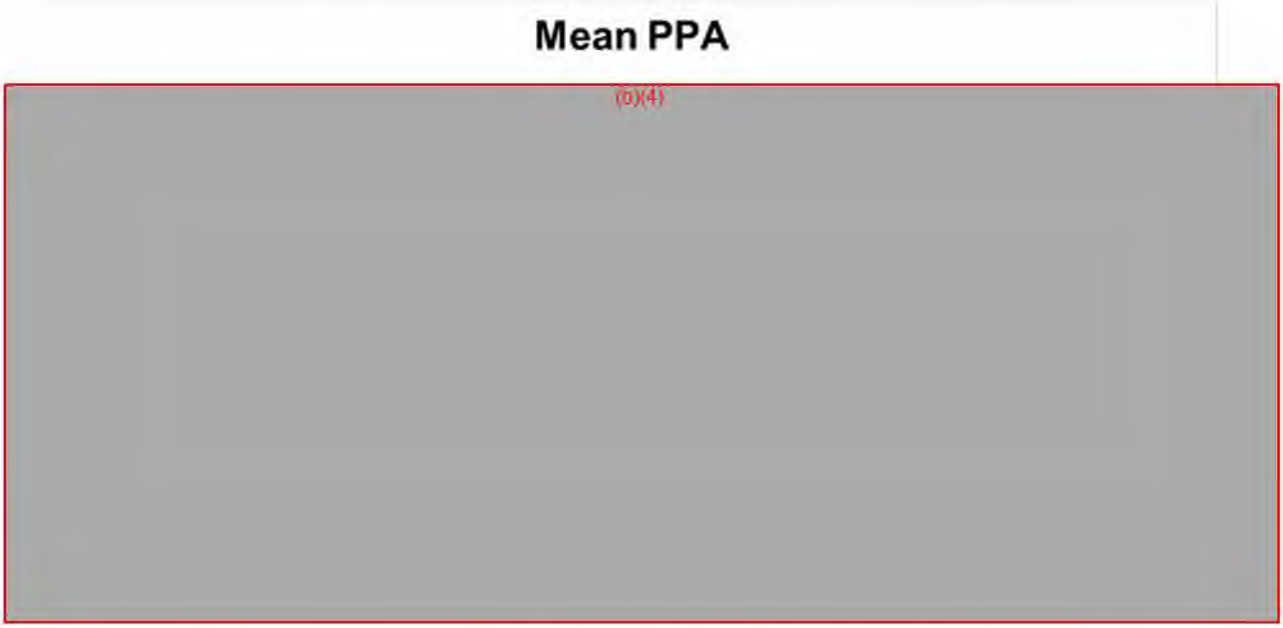
ii. Fisher strand bias score (FS): True positive (TP) and false positive (FP) variant calls from the Accuracy study 1 were grouped based on FS score ranges, (b)(4) (b)(4) While calling accuracy (TPPV) for SNVs was achieved at 99.91% for FS (b)(4), more FPs were detected than TPs in with FS (b)(4) (**Table 4**). Therefore, FS threshold for SNV calling was set at (b)(4)

Table 4. Performance SNV calling in Accuracy study 1 based on Fisher strand bias score ranges

(b)(4)

iii. Root mean square mapping quality score (MQ): Performance metrics (PPA) from samples in all analytical studies were grouped based on MQ score ranges, (b)(4) and (b)(4). The mean PPA in MQ (b)(4) was significantly lower than mean PPA for variants with MQ (b)(4). Therefore, the evidence did not provide adequate support that MQ scores (b)(4) would be sufficient for variant calling with high accurate results (**Figure 8**). The data supported high confidence results when the MQ score is (b)(4) The threshold for MQ threshold was set (b)(4)

Figure 8. Performance of variant calling (PPA) from all samples in analytical validation studies based on the range of MQ score



iv. Indel size: Performance metrics (PPA) of indel calling from samples in all analytical studies were grouped based on size ranges, 1-2bp, 3-5bp, and 6-20bp. Although the

median PPA for indel size of 6-20bp was achieved at (b)(4) (**Figure 9**), it was significantly lower than size 1-2bp and 3-5bp (this conclusion was supported through FDA analysis of p-values, data not shown). Thus, caution should be taken for indel size 6-20bp.

Figure 9. Performance of indel calling (PPA) from all samples in analytical validation studies based on the range of size

b

v. Removal of sequencing regions with low quality bases: The (b)(4) cell line samples from Accuracy 1 and one replicate of (b)(4) unique clinical samples from the DNA Input (b)(4), Endogenous (b)(4), Exogenous (b)(4), Smoking (b)(4), Precision (b)(4), Between-Lot Reproducibility (b)(4), and the Microbial (b)(4) studies were used to identify regions with a high percentage of bases with quality score (b)(4). Base quality scores within windows of (b)(4) base pairs were aggregated across all samples and then the percentage of bases with

quality score (b)(4) within each window were calculated. Based on the empirical data provided in **Table 5**, this product is not intended to report variants in regions with (b)(4) of bases with quality score (b)(4), total of (b)(4) in the coding region with high potential for sequencing errors were excluded. The data in the table below demonstrates that when greater than or equal to (b)(4) of the bases have base quality score (b)(4) the (b)(4). Therefore, this data supports the criterion that (b)(4) % of the bases need to have a base quality score ≥ (b)(4) With this criterion the risk of false positives is minimized.

Table 5. Summary of regions with base quality scores (b)(4) across 6 cell line samples and 102 clinical samples.

(b)(4)

2. Analytical performance:

The reportable range includes the coding range minus distinct regions that the HLP does not report due to unreliable resolution of the genetic sequence. Specific regions of the coding region have increased targeted coverage and different quality metrics for reporting. For this reason, performance was evaluated for each region based on the pre-specified quality metrics and thresholds described above. The definitions and formulas used for each study is described in **Table 6**. Performance data generated with cell lines were not overinflated with respect to saliva samples (data not shown).

Table 6. Definitions and Formulas

Term	Definition
Number of Variants Expected	The total number of known truth variants in the intersection of the Helix *Exome+* reportable range and the reference sequence high confidence regions.
Coding	Targets representing coding regions (the exome); the callability threshold for the coding region is (b)(4)
Mendeliome	Targets representing (b)(4) genes; (b)(4) (b)(4)
Priority	Coding regions of 615 genes (a subset of Mendeliome) which were tested to support k19206; (b)(4) (b)(4)
STR	Short tandem repeats identified that are nucleotide repeats (b)(4) long. This includes strings of homopolymer runs and di/tri nucleotide repeat regions. STRs are not reported. Variant detection was determined near STR, i.e., variants that are located within or immediately next to a STR.
Non-STR	All other variants that are not "near STR".
True Negative (TN)	Bases consistent with reference dataset (i.e., wild-type) which are expected variants that were correctly called with the expected genotype.
True Positive (TP)	Bases that differ from the reference dataset (i.e., variant) which are expected variants that were correctly called with the expected genotype.
False negative (FN)	Expected variants that did not receive a no-call and were not correctly called.

Term	Definition
False positive (FP)	HLP called variants that did not match an expected variant genotype such as wild type
Non-Reference Concordance (NRC)	Agreement when ignoring reference calls. TP/(TP+FP+FN)
Positive Percent Agreement (PPA)	Ability of the test to correctly identify variants that are present in a sample. TP/(TP+FN)
Negative Percent Agreement (NPA)	Ability of the test to correctly identify wild-type bases (the probability that the test will not call a variant). TN/(TN+FP)
Technical Positive Predictive Value (TPPV)	Defined as the probability of a variant calling being a true positive. TP/(TP+FP)
Concordance assessment for precision	The degree of agreement between test samples defined as calculating the accuracy for a sample to the reference dataset or gold standard dataset or majority call rate of the same sample. (TP+TN)/(TP+TN+FP+FN)
Acceptance Criteria	SNVS (PPA and TPPV ≥ 99.5%, NPA ≥99.99%) INDELS (PPA and TPPV ≥ 99.0%)
Majority Call	The majority (consensus) calls are generated from at least two replicates where more than 50% of replicates have the same (as defined by the location and genotype) passing variant or reference record.
Negative Predictive Value (NPV)	Defined as the probability of a variant call being a true negative TN/(TN +FN)
Positive Predictive Value (PPV)	Defined as the probability of a variant call being a true positive (TP/(TP+FP)
Adjusted Negative Percent Agreement (aNPA)	The adjusted Negative Percent Agreement (aNPA) is defined as the ability of the test to correctly identify wild-type bases (the probability that the test will not call a variant). aNPA = (NPV*(1-p)) / (NPV*(1-p) + (1-PPV)*p) where p is the sampling prevalence, i.e. p = percentage of people positive for at least one variant in the set of tested variants. This value was calculated to adjust for ascertainment bias given clinical samples are previously ascertained using HLP.

Term	Definition
Adjusted Positive Percent Agreement (aPPA)	The adjusted Positive Percent Agreement (aPPA) is defined as the ability of the test to correctly identify polymorphic variants that are present in a sample. aPPA =(PPV*p) / (PPV*p+(1-NPV)*(1-p)) where p is the sampling prevalence, i.e. p = percentage of people positive for at least one variant in the set of tested variants. This value was calculated to adjust for ascertainment bias given clinical samples are previously ascertained using HLP.

Reference Samples used in these studies were obtained from various sources as listed in **Table 7**:

Table 7. Reference Samples used in the studies.

Sample Name (Source)	Reference Data	Ethnicity	Gender	Variants
NA12877	Platinum Genomes	Northern European from Utah	Male	All high confidence variants present in the HLP reportable and analytical ranges
NA12878	GIAB	Northern European from Utah	Female	
NA24385	GIAB	Ashkenazim Jewish	Male (son)	
NA24149	GIAB	Ashkenazim Jewish	Male (father)	
NA24143	GIAB	Ashkenazim Jewish	Female (mother)	
NA24695	GIAB	Asian Chinese	Female (mother)	

a. Precision

A study was conducted to assess the precision of the Helix Laboratory Platform. A total of 24 unique samples (DNA from 6 cell lines with known variants and 18 saliva samples) were tested across 3 different operators using3 separate cBots at the clustering step; and across 3 different HiSeq instruments at the sequencing steps. Each unique sample started with 9 replicates in library prep (3 replicates/sample x 3 plates of library prep = 9 replicates/sample). Each library prep was processed twice through independent enrichments, increasing the number of replicates to 18/sample. Each of the enriched libraries was sequenced on 4 independent runs of the cBot and HiSeq instruments, resulting in a total of 72 replicates/sample (18 enriched libraries/sample x 4 sequencing events = 72 replicates/sample). A total of 1728 samples were included in this study design (24 unique samples x 72 replicates/ sample = 1728).

Positive percent agreement (PPA), negative percent agreement (NPA) and technical positive predictive value (TPPV) were calculated for each sample replicate. The mean of these values for each sample per condition and for all samples for each condition was determined. This study passed all acceptance criteria thresholds for mean PPA, NPA and TPPV at all conditions evaluated (library prep operator, enrichment operator, cBot operator, cBot instrument, HiSeq operator and HiSeq instrument – data not shown) when taken in aggregate. Mean quality metrics were evaluated for the coding regions and subsets (Mendeliome and Priority). Precision is described below for cell lines and clinical samples independently. The invalid rate in the precision studies was approximately (b)(4).

i. Precision - Reference Samples (Cell lines)

Precision using reference cell lines was evaluated. Results of this studyusing six reference cell line samples each with 72 replicates processed on the Helix Laboratory Platform were found to pass all acceptance criteria with PPA 99.91% and TPPV 99.93% for SNVs > 99.5%, and PPA 99.52% and TPPV 99.29% for insertions and PPA 99.59% and TPPV 99.18% for deletions when considered in aggregate, however PPA for indels ≥ 6bp could be <99% (**Figure 11**). No call rates for each variant type are shown and ranged from 5% for SNVs to 29% for indels (**Figure 10**). Calling performance based on variant type and size in each sample were broken down by Coding regions (**Table 8**), Mendeliome regions (**Table 9**), and Priority regions (**Table 10**). The SNV was further stratified by zygosity (heterozygous, homozygous, hemizygous) and indels were stratified by size (1-2bp, 3-5 bp and 6-20 bp). Acceptance criteria were not applied to the stratified data, however data that was less than the acceptance criteria are shaded and shown in bold in the tables. **The instructions for use indicate that insertions ≥ 6bp will be validated independently.**

For all tables below (cell lines and clinical specimens) the data is shown for all calls combined across the 72 replicates. The mean of per sample values was also assessed and showed slightly improved (data not shown).

Figure 10. Percentage of No Call Rate for SNV and Indels in Six Reference Cell Lines

Figure 11. Concordance of Variant Calling for SNV and Indels in Six Reference Cell lines

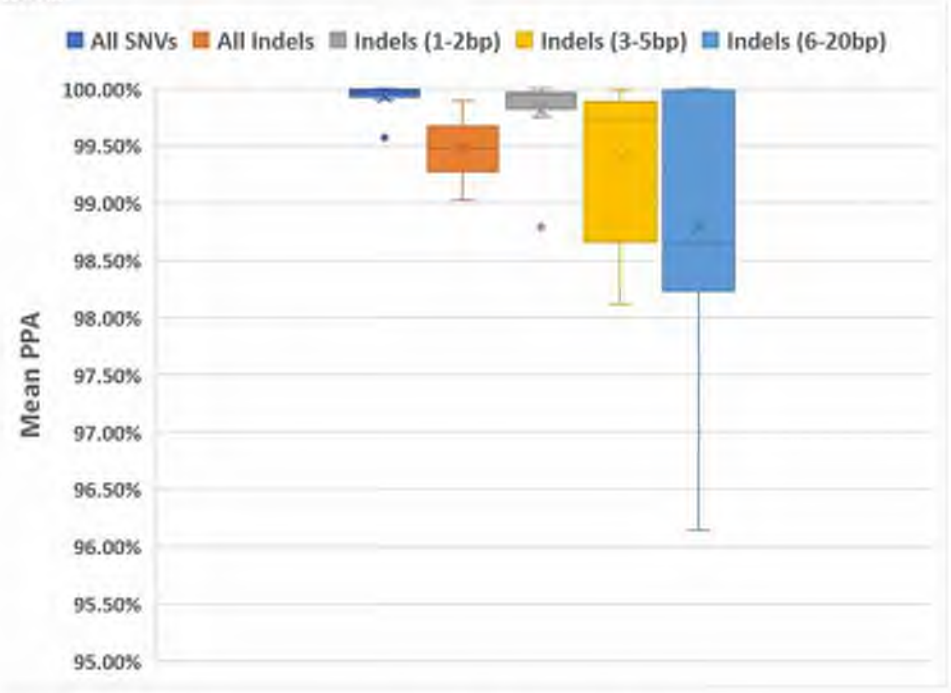


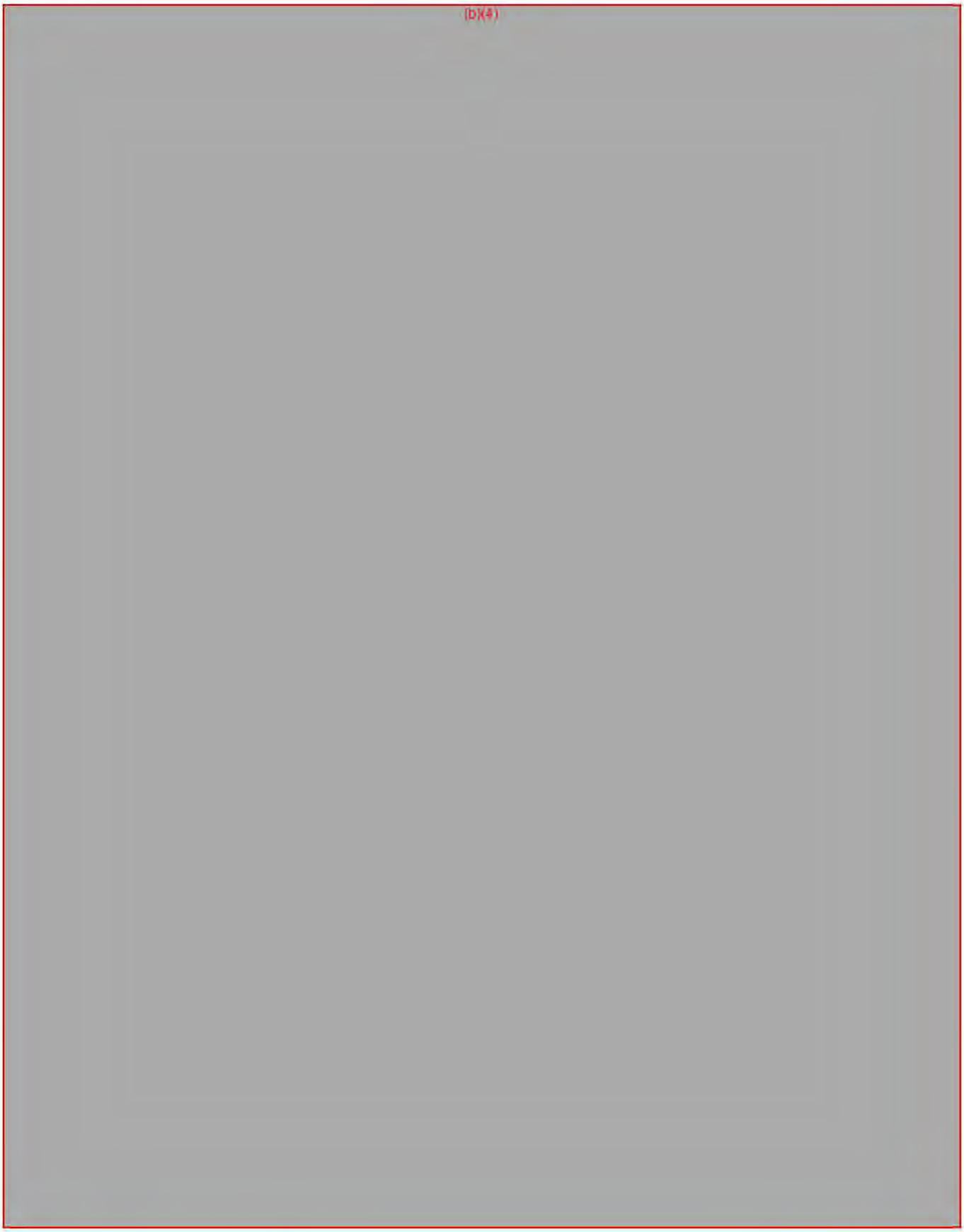
Table 8: Precision Using Reference Cell lines – Coding Region

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
NA12877	SNV (het)	All	(b)(4)		953346	957	4083	99.57%	99.90%
	SNV (hom)	All			571420	68	137	99.98%	99.99%
	SNV (hemi)	All			18512	0	0	100.00%	100.00%
	Insertion	1-2			5579	28	1	99.98%	99.50%
		3-5			3641	17	91	**97.56%**	99.54%
		6-20			1359	88	17	**98.76%**	**93.92%**
		All			10579	133	109	98.98%	**98.76%**
	Deletion	1-2			6634	41	0	100.00%	99.39%
		3-5			4933	8	1	99.98%	99.84%
		6-20			1300	1	2	99.85%	99.92%
		All			12867	50	3	99.98%	99.61%
NA12878	SNV (het)	All			783200	532	369	99.95%	99.93%
	SNV (hom)	All			487707	3	140	99.97%	100.00%
	Insertion	1-2			4295	8	1	99.98%	99.81%
		3-5			3128	13	8	99.74%	99.59%
		6-20			842	74	72	**92.12%**	**91.92%**
		All			8265	95	81	99.03%	**98.86%**
	Deletion	1-2			5621	86	78	**98.63%**	**98.49%**
		3-5			3598	87	2	99.94%	**97.64%**
		6-20			1070	0	0	100.00%	100.00%
		All			10289	173	80	99.23%	**98.35%**
NA24143	SNV (het)	All			768755	538	64	99.99%	99.93%
	SNV (hom)	All			459671	4	17	100.00%	100.00%
	Insertion	1-2			4381	14	1	99.98%	99.68%
		3-5			2577	5	4	99.85%	99.81%
		6-20			905	18	27	**97.10%**	**98.05%**
		All			7863	37	32	99.59%	99.53%
	Deletion	1-2			5485	26	4	99.93%	99.53%
		3-5			3624	18	2	99.94%	99.51%
		6-20			1082	0	0	100.00%	100.00%
		All			10191	44	6	99.94%	99.57%
NA24149	SNV (het)	All			768869	2008	570	99.93%	99.74%
	SNV (hom)	All			459159	4	18	100.00%	100.00%
	Insertion	1-2			5173	17	4	99.92%	99.67%
		3-5			2714	18	0	100.00%	99.34%
		6-20			991	14	2	99.80%	**98.61%**
		All			8878	49	6	99.93%	99.45%

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	Deletion	1-2	(b)(4)		5109	147	3	99.94%	**97.20%**
		3-5			3469	10	1	99.97%	99.71%
		6-20			836	3	1	99.88%	99.64%
		All			9414	160	5	99.95%	**98.33%**
NA24385	SNV (het)	All			801423	506	815	99.90%	99.94%
	SNV (hom)	All			477860	4	13	100.00%	100.00%
	Insertion	1-2			4669	9	2	99.96%	99.81%
		3-5			2919	4	1	99.97%	99.86%
		6-20			850	13	0	100.00%	**98.49%**
		All			8438	26	3	99.96%	99.69%
	Deletion	1-2			5135	21	3	99.94%	99.59%
		3-5			3658	12	72	98.07%	99.67%
		6-20			869	0	0	100.00%	100.00%
		All			9662	33	75	99.23%	99.66%
NA24631	SNV (het)	All			744482	872	561	99.92%	99.88%
	SNV (hom)	All			518002	4	105	99.98%	100.00%
	Insertion	1-2			5429	14	4	99.93%	99.74%
		3-5			2774	15	14	99.50%	99.46%
		6-20			969	13	7	99.28%	98.68%
		All			9172	42	25	99.73%	99.54%
	Deletion	1-2			5143	40	15	99.71%	99.23%
		3-5			3483	8	74	**97.92%**	99.77%
		6-20			1140	4	0	100.00%	99.65%
		All			9766	52	89	99.10%	99.47%

Table 9. Precision Using Reference Cell Lines - Mendeliome

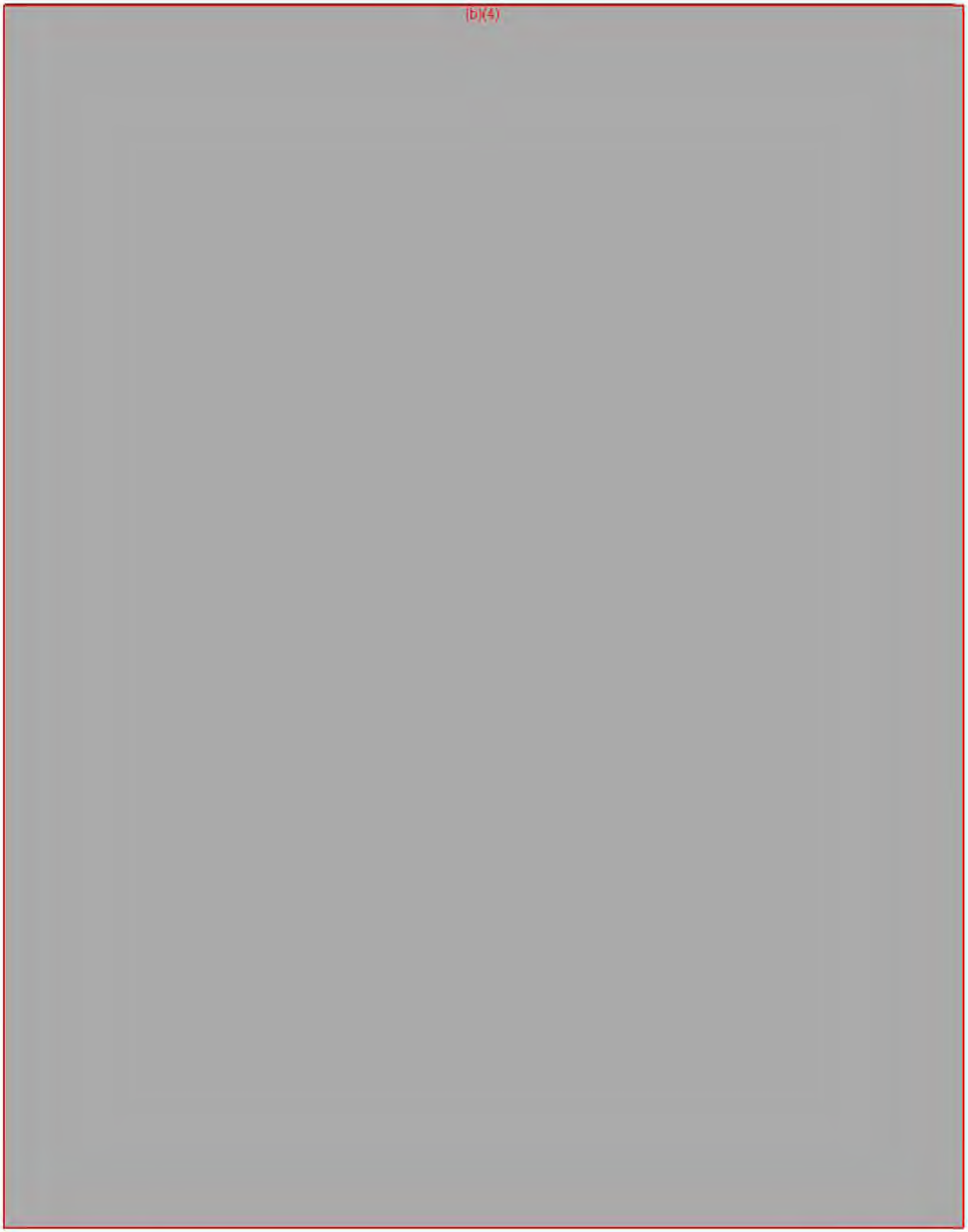
(b)(4)


(b)(4)

(b)(4)

Table 10. Precision Using Reference Cell Lines -Priority

(b)(4)


(b)(4)

ii. Precision - Clinical Specimens

The reproducibility of 18 saliva (clinical) specimens was evaluated as described above. Of 18 saliva samples, one sample (SF-6203-80058) had all 9 replicates fail at the library preparation step, which were not processed further. This failure resulted in all downstream replicate samples being removed from this study (9 library preparations x 2 enrichments/library preparations x 4 sequencing events/enrichment = 72 replicates not sequenced). The sample was not replaced, as per protocol. Two samples (SF-5744-44790, SF-8763-29117) each had one library preparation replicate, of 9 total per sample, failed at library preparation quantitation, and were not replaced as per protocol. Failure of these two samples removed 16 sample replicates from the study (2 library preparations x 2 enrichments/library preparation x 4 sequencing events/enrichment = 16 replicates not sequenced). One sample (SF-2856-29005) had a total of 9 replicates fail and were not replaced as per protocol.

Additionally, 21 sample replicates failed to pass the DNA sequencing QC metric thresholds and were not included in the analysis for this study, as per protocol. In summary, 118 sample replicates were not evaluable and were not included in the final data analyses presented below (9.1% of total replicates). The total number of replicates for the 17 saliva specimens is shown in **Table 11**.

Table 11. Post-Sequencing Clinical Sample Replicate Number

Group Name	Number of Sample Replicates Available for Analysis
SF-3312-24549	69
SF-5744-44790	63
SF-8763-29117	64
SF-2856-29005	60
SF-8085-16718	72
SF-3607-46036	69
SF-5221-11824	72
SF-3002-89172	72
SF-7804-33749	72
SF-1901-15675	69
SF-6203-80058	0
SF-8038-20902	72
SF-1070-81401	72
SF-2757-45105	72
SF-2711-11665	72
SF-5730-56578	64
SF-5044-81848	72

Group Name	Number of Sample Replicates Available for Analysis
SF-7430-13589	70
Total	1178*

*1178 of 1296 possible replicates

Performance based on variant type and size in each saliva sample were broken down by Coding regions (**Table 12**), Mendeliome regions (**Table 13**), and Priority regions (**Table 14**). Results of this test using clinical samples each with 72 replicates processed on the Helix Laboratory Platform were found to pass all acceptance criteria with PPA 99.98% and TPPV 99.96% for SNVs, and PPA 99.83% and TPPV 99.12% for insertions and PPA 99.93% and TPPV 99.73% for deletions when considered in aggregate. The SNV was further stratified by zygosity (heterozygous, homozygous, hemizygous) and indels were stratified by size (1-2bp, 3-5 bp and 6-20 bp). Acceptance criteria were not applied to the stratified data, however data that was less than the acceptance criteria are shaded and shown in bold in the tables.

Table 12. Precision Using Clinical Samples - Coding Regions

Sample Name (number of replicates)	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
SF-1070-81401 (72)	SNV (het)	All	(b)(4)		973343	466	149	99.98%	99.95%
	SNV (hom)	All			564279	48	87	99.98%	99.99%
	Insertion	1-2			5184	22	4	99.92%	99.58%
		3-5			3435	18	7	99.80%	99.48%
		6-20			1289	68	1	99.92%	**94.99%**
		All			9908	108	12	99.88%	**98.92%**
	Deletion	1-2			7226	18	10	99.86%	99.75%
		3-5			4198	17	1	99.98%	99.60%
		6-20			1371	0	0	100.00%	100.00%
		All			12795	35	11	99.91%	99.73%
SF-1901-15675 (69)	SNV (het)	All			875072	322	136	99.98%	99.96%
	SNV (hom)	All			529558	81	121	99.98%	99.98%
	SNV (hemi)	All			16370	0	1	99.99%	100.00%
	Insertion	1-2			4790	18	3	99.94%	99.63%
		3-5			3149	10	1	99.97%	99.68%
		6-20			1133	35	27	**97.67%**	**97.00%**
		All			9072	63	31	99.66%	99.31%
	Deletion	1-2			5780	14	12	99.79%	99.76%
		3-5			4169	29	2	99.95%	99.31%
		6-20			1747	1	6	99.66%	99.94%
		All			11696	44	20	99.83%	99.63%

Sample Name (number of replicates)	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
SF-2711-11665 (72)	SNV (het)	All	(b)(4)		931251	387	102	99.99%	99.96%
	SNV (hom)	All			552477	36	43	99.99%	99.99%
	SNV (hemi)	All			20288	0	0	100.00%	100.00%
	Insertion	1-2			6040	54	11	99.82%	99.11%
		3-5			3212	16	0	100.00%	99.50%
		6-20			1531	21	19	**98.77%**	**98.65%**
		All			10783	91	30	99.72%	99.16%
	Deletion	1-2			6890	3	1	99.99%	99.96%
		3-5			4898	20	0	100.00%	99.59%
		6-20			1636	14	33	**98.02%**	99.15%
		All			13424	37	34	99.75%	99.73%
SF-2757-45105 (72)	SNV (het)	All			953805	486	216	99.98%	99.95%
	SNV (hom)	All			553134	107	131	99.98%	99.98%
	Insertion	1-2			5707	24	9	99.84%	99.58%
		3-5			3421	33	0	100.00%	99.04%
		6-20			1106	48	13	**98.84%**	**95.84%**
		All			10234	105	22	99.79%	**98.98%**
	Deletion	1-2			6525	4	0	100.00%	99.94%
		3-5			3917	32	2	99.95%	99.19%
		6-20			1746	2	5	99.71%	99.89%
		All			12188	38	7	99.94%	99.69%
SF-2856-29005 (60)	SNV (het)	All			786113	338	120	99.98%	99.96%
	SNV (hom)	All			460283	50	84	99.98%	99.99%
	Insertion	1-2			4445	6	0	100.00%	99.87%
		3-5			2455	6	0	100.00%	99.76%
		6-20			962	28	13	**98.67%**	**97.17%**
		All			7862	40	13	99.83%	99.49%
	Deletion Deletion Deletion Deletion	1-2			4885	7	4	99.92%	99.86%
		3-5			3694	18	1	99.97%	99.52%
		6-20			1144	1	1	99.91%	99.91%
		All			9723	26	6	99.94%	99.73%
SF-3002-89172 (72)	SNV (het)	All			948079	522	182	99.98%	99.94%
	SNV (hom)	All			549038	12	72	99.99%	100.00%
	Insertion	1-2			6103	24	2	99.97%	99.61%
		3-5			3184	30	2	99.94%	99.07%
		6-20			1258	86	3	99.76%	**93.60%**
		All			10545	140	7	99.93%	**98.69%**

Sample Name (number of replicates)	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	Deletion	1-2	(b)(4)		5557	7	0	100.00%	99.87%
		3-5			4224	30	2	99.95%	99.29%
		6-20			1332	1	3	99.78%	99.92%
		All			11113	38	5	99.96%	99.66%
SF-3312-24549 (69)	SNV (het)	All			898443	602	327	99.96%	99.93%
	SNV (hom)	All			523131	73	65	99.99%	99.99%
	Insertion	1-2			5553	14	2	99.96%	99.75%
		3-5			3238	23	7	99.78%	99.29%
		6-20			1101	36	0	100.00%	96.83%
		All			9892	73	9	99.91%	99.27%
	Deletion	1-2			6217	6	6	99.90%	99.90%
		3-5			4237	13	0	100.00%	99.69%
		6-20			1450	3	1	99.93%	99.79%
		All			11904	22	7	99.94%	99.82%
SF-3607-46036 (69)	SNV (het)	All			893750	449	184	99.98%	99.95%
	SNV (hom)	All			535016	45	51	99.99%	99.99%
	Insertion	1-2			4959	21	3	99.94%	99.58%
	Insertion	3-5			2884	36	17	99.41%	**98.77%**
		6-20			1122	69	0	100.00%	**94.21%**
		All			8965	126	20	99.78%	**98.61%**
	Deletion	1-2			5784	3	0	100.00%	99.95%
		3-5			4310	17	0	100.00%	99.61%
		6-20			1354	1	1	99.93%	99.93%
		All			11448	21	1	99.99%	99.82%
SF-5044-81848 (72)	SNV (het)	All			921112	386	144	99.98%	99.96%
	SNV (hom)	All			556208	39	99	99.98%	99.99%
	SNV (hemi)	All			17552	0	0	100.00%	100.00%
	Insertion	1-2			5933	50	3	99.95%	99.16%
		3-5			3383	27	5	99.85%	99.21%
		6-20			1542	32	11	99.29%	**97.97%**
		All			10858	109	19	99.83%	99.01%
	Deletion	1-2			6833	5	1	99.99%	99.93%
		3-5			3794	33	0	100.00%	99.14%
		6-20			1500	3	0	100.00%	99.80%
		All			12127	41	1	99.99%	99.66%
SF-5221-11824 (72)	SNV (het)	All			1052179	418	137	99.99%	99.96%
	SNV (hom)	All			556770	21	58	99.99%	100.00%

Sample Name (number of replicates)	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	SNV (hemi)	All	(b)(4)		23396	0	0	100.00%	100.00%
	Insertion	1-2			6019	35	15	99.75%	99.42%
		3-5			3788	17	5	99.87%	99.55%
		6-20			1047	33	2	99.81%	**96.94%**
		All			10854	85	22	99.80%	99.22%
	Deletion	1-2			7330	6	1	99.99%	99.92%
		3-5			3749	22	7	99.81%	99.42%
		6-20			1928	3	0	100.00%	99.84%
		All			13007	31	8	99.94%	99.76%
SF-5730-56578 (64)	SNV (het)	All			803771	440	152	99.98%	99.95%
	SNV (hom)	All			499280	154	171	99.97%	99.97%
	Insertion	1-2			4541	22	0	100.00%	99.52%
		3-5			2892	7	0	100.00%	99.76%
		6-20			1100	10	3	99.73%	99.10%
		All			8533	39	3	99.96%	99.55%
	Deletion	1-2			5741	6	3	99.95%	99.90%
		3-5			3521	25	3	99.91%	99.29%
		6-20			1246	1	0	100.00%	99.92%
		All			10508	32	6	99.94%	99.70%
SF-5744-44790 (63)	SNV (het)	All			810534	361	115	99.99%	99.96%
	SNV (hom)	All			473706	64	129	99.97%	99.99%
	SNV (hemi)	All			18275	0	0	100.00%	100.00%
	Insertion	1-2			4557	9	2	99.96%	99.80%
		3-5			2231	6	5	99.78%	99.73%
		6-20			1051	23	3	99.72%	**97.86%**
		All			7839	38	10	99.87%	99.52%
	Deletion	1-2			5373	7	8	99.85%	99.87%
		3-5			3678	16	0	100.00%	99.57%
		6-20			1319	2	3	99.77%	99.85%
		All			10370	25	11	99.89%	99.76%
SF-7430-13589 (70)	SNV (het)	All			889411	437	135	99.98%	99.95%
	SNV (hom)	All			535661	162	186	99.97%	99.97%
	SNV (hemi)	All			17194	0	0	100.00%	100.00%
	Insertion	1-2			5406	40	0	100.00%	99.27%
		3-5			3247	20	5	99.85%	99.39%
		6-20			1234	57	0	100.00%	**95.58%**
		All			9887	117	5	99.95%	**98.83%**

Sample Name (number of replicates)	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	Deletion	1-2	(b)(4)		6589	6	0	100.00%	99.91%
		3-5			4996	12	1	99.98%	99.76%
		6-20			1090	4	5	99.54%	99.63%
		All			12675	22	6	99.95%	99.83%
SF-7804-33749 (72)	SNV (het)	All			928456	449	190	99.98%	99.95%
	SNV (hom)	All			549399	154	176	99.97%	99.97%
	SNV (hemi)	All			18875	0	0	100.00%	100.00%
	Insertion	1-2			5322	21	7	99.87%	99.61%
		3-5			3771	14	0	100.00%	99.63%
		6-20			1484	14	2	99.87%	99.07%
		All			10577	49	9	99.91%	99.54%
	Deletion	1-2			6315	5	0	100.00%	99.92%
		3-5			5118	26	3	99.94%	99.49%
		6-20			2039	5	3	99.85%	99.76%
		All			13472	36	6	99.96%	99.73%
SF-8038-20902 (72)	SNV (het)	All			939766	498	314	99.97%	99.95%
	SNV (hom)	All			573278	148	192	99.97%	99.97%
	Insertion	1-2			6180	33	1	99.98%	99.47%
		3-5			3168	25	0	100.00%	99.22%
		6-20			1469	52	4	99.73%	**96.58%**
		All			10817	110	5	99.95%	**98.99%**
	Deletion	1-2			6745	11	1	99.99%	99.84%
		3-5			4367	35	0	100.00%	99.20%
		6-20			1268	1	0	100.00%	99.92%
		All			12380	47	1	99.99%	99.62%
SF-8085-16718 (72)	SNV (het)	All			908943	437	164	99.98%	99.95%
	SNV (hom)	All			541517	98	154	99.97%	99.98%
	SNV (hemi)	All			17369	0	0	100.00%	100.00%
	Insertion	1-2			4630	31	0	100.00%	99.33%
		3-5			3046	17	1	99.97%	99.44%
		6-20			1394	31	17	**98.80%**	**97.82%**
		All			9070	79	18	99.80%	99.14%
	Deletion	1-2			6305	5	0	100.00%	99.92%
		3-5			4438	13	0	100.00%	99.71%
		6-20			1684	1	0	100.00%	99.94%
		All			12427	19	0	100.00%	99.85%
	SNV (het)	All			758854	375	154	99.98%	99.95%

Sample Name (number of replicates)	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
SF-8763-29117 (64)	SNV (hom)	All	(b)(4)		556416	156	222	99.96%	99.97%
	SNV (hemi)	All			17805	0	0	100.00%	100.00%
	Insertion	1-2			4311	28	16	99.63%	99.35%
		3-5			2593	22	9	99.65%	99.16%
		6-20			1200	34	17	**98.60%**	**97.24%**
		All			8104	84	42	99.48%	**98.97%**
	Deletion	1-2			5267	5	1	99.98%	99.91%
		3-5			3229	13	2	99.94%	99.60%
		6-20			1262	3	1	99.92%	99.76%
		All			9758	21	4	99.96%	99.79%

Table 13. Precision Using Clinical Samples - Mendeliome

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
SF-1070-81401	SNV (het)	All	(b)(4)		256740	72	17	99.99%	99.97%
	SNV (hom)	All			164983	4	10	99.99%	100.00%
	Insertion	1-2			1416	5	0	100.00%	99.65%
		3-5			1133	4	0	100.00%	99.65%
		6-20			328	19	0	100.00%	**94.52%**
		All			2877	28	0	100.00%	99.04%
	Deletion	1-2			2348	10	9	99.62%	99.58%
		3-5			780	8	0	100.00%	**98.98%**
		6-20			284	0	0	100.00%	100.00%
		All			3412	18	9	99.74%	99.48%
SF-1901-15675	SNV (het)	All			245884	65	23	99.99%	99.97%
	SNV (hom)	All			145341	69	84	99.94%	99.95%
	SNV (hemi)	All			4689	0	0	100.00%	100.00%
	Insertion	1-2			1372	1	0	100.00%	99.93%
		3-5			975	4	0	100.00%	99.59%
		6-20			301	17	14	**95.56%**	**94.65%**
		All			2648	22	14	99.47%	99.18%
	Deletion	1-2			1541	1	0	100.00%	99.94%
		3-5			962	5	0	100.00%	99.48%
		6-20			412	1	0	100.00%	99.76%
		All			2915	7	0	100.00%	99.76%
SF-2711-11665	SNV (het)	All			256127	106	15	99.99%	99.96%

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	SNV (hom)	All	(b)(4)		150275	23	3	100.00%	99.98%
	SNV (hemi)	All			4934	0	0	100.00%	100.00%
	Insertion	1-2			1927	24	0	100.00%	**98.77%**
		3-5			977	1	0	100.00%	99.90%
		6-20			533	2	3	99.44%	99.63%
		All			3437	27	3	99.91%	99.22%
	Deletion	1-2			1920	1	0	100.00%	99.95%
		3-5			790	1	0	100.00%	99.87%
		6-20			471	9	13	**97.31%**	**98.13%**
		All			3181	11	13	99.59%	99.66%
SF-2757-45105	SNV (het)	All			267470	93	32	99.99%	99.97%
	SNV (hom)	All			152844	4	33	99.98%	100.00%
	Insertion	1-2			1601	7	4	99.75%	99.56%
		3-5			922	2	0	100.00%	99.78%
		6-20			445	7	0	100.00%	**98.45%**
		All			2968	16	4	99.87%	99.46%
	Deletion	1-2			1961	1	0	100.00%	99.95%
		3-5			628	5	0	100.00%	99.21%
		6-20			415	0	0	100.00%	100.00%
		All			3004	6	0	100.00%	99.80%
SF-2856-29005	SNV (het)	All			220858	57	25	99.99%	99.97%
	SNV (hom)	All			128303	3	12	99.99%	100.00%
	Insertion	1-2			1097	3	0	100.00%	99.73%
		3-5			751	5	0	100.00%	99.34%
		6-20			281	16	0	100.00%	**94.61%**
		All			2129	24	0	100.00%	98.89%
	Deletion	1-2			1376	1	0	100.00%	99.93%
		3-5			804	5	0	100.00%	99.38%
		6-20			272	0	0	100.00%	100.00%
		All			2452	6	0	100.00%	99.76%
SF-3002-89172	SNV (het)	All			264950	94	65	99.98%	99.96%
	SNV (hom)	All			156857	2	13	99.99%	100.00%
	Insertion	1-2			1559	7	0	100.00%	99.55%
		3-5			958	1	0	100.00%	99.90%
		6-20			494	24	2	99.60%	**95.37%**
		All			3011	32	2	99.93%	**98.95%**
	Deletion	1-2			1636	1	0	100.00%	99.94%

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		3-5	(b)(4)		627	10	0	100.00%	**98.43%**
		6-20			398	0	0	100.00%	100.00%
		All			2661	11	0	100.00%	99.59%
SF-3312-24549	SNV (het)	All			263378	130	92	99.97%	99.95%
	SNV (hom)	All			140912	14	1	100.00%	99.99%
	Insertion	1-2			1533	4	0	100.00%	99.74%
		3-5			858	2	0	100.00%	99.77%
		6-20			256	13	0	100.00%	**95.17%**
		All			2647	19	0	100.00%	99.29%
	Deletion	1-2			1699	1	0	100.00%	99.94%
		3-5			821	0	0	100.00%	100.00%
		6-20			312	0	0	100.00%	100.00%
		All			2832	1	0	100.00%	99.96%
SF-3607-46036	SNV (het)	All			254473	119	42	99.98%	99.95%
	SNV (hom)	All			145473	3	7	100.00%	100.00%
	Insertion	1-2			1191	10	1	99.92%	99.17%
		3-5			856	6	0	100.00%	99.30%
		6-20			447	14	0	100.00%	**96.96%**
		All			2494	30	1	99.96%	**98.81%**
	Deletion	1-2			1474	0	0	100.00%	100.00%
		3-5			1030	3	0	100.00%	99.71%
		6-20			334	1	1	99.70%	99.70%
		All			2838	4	1	99.96%	99.86%
SF-5044-81848	SNV (het)	All			264606	51	19	99.99%	99.98%
	SNV (hom)	All			152174	16	1	100.00%	99.99%
	SNV (hemi)	All			5198	0	0	100.00%	100.00%
	Insertion	1-2			1899	18	1	99.95%	99.06%
		3-5			706	4	0	100.00%	99.44%
		6-20			458	12	0	100.00%	**97.45%**
		All			3063	34	1	99.97%	**98.90%**
	Deletion	1-2			1649	3	1	99.94%	99.82%
		3-5			648	1	0	100.00%	99.85%
		6-20			461	0	0	100.00%	100.00%
		All			2758	4	1	99.96%	99.86%
SF-5221-11824	SNV (het)	All			291590	126	10	100.00%	99.96%
	SNV (hom)	All			150298	1	31	99.98%	100.00%
	SNV (hemi)	All			6286	0	0	100.00%	100.00%

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	Insertion	1-2	(b)(4)		1665	19	13	99.23%	**98.87%**
		3-5			1050	3	1	99.90%	99.72%
		6-20			428	16	2	99.53%	**96.40%**
		All			3143	38	16	99.49%	**98.81%**
	Deletion	1-2			2099	1	0	100.00%	99.95%
		3-5			861	1	0	100.00%	99.88%
		6-20			470	0	0	100.00%	100.00%
		All			3430	2	0	100.00%	99.94%
SF-5730-56578	SNV (het)	All			227706	51	11	100.00%	99.98%
	SNV (hom)	All			140042	64	78	99.94%	99.95%
	Insertion	1-2			1366	4	0	100.00%	99.71%
		3-5			932	1	0	100.00%	99.89%
		6-20			293	3	1	99.66%	**98.99%**
		All			2591	8	1	99.96%	99.69%
	Deletion	1-2			1752	0	1	99.94%	100.00%
		3-5			636	1	0	100.00%	99.84%
		6-20			303	0	0	100.00%	100.00%
		All			2691	1	1	99.96%	99.96%
SF-5744-44790	SNV (het)	All			226779	68	14	99.99%	99.97%
	SNV (hom)	All			132460	60	86	99.94%	99.95%
	SNV (hemi)	All			4615	0	0	100.00%	100.00%
	Insertion	1-2			1332	4	0	100.00%	99.70%
		3-5			784	0	0	100.00%	100.00%
		6-20			223	3	3	**98.67%**	**98.67%**
		All			2339	7	3	99.87%	99.70%
	Deletion	1-2			1459	2	0	100.00%	99.86%
		3-5			693	2	0	100.00%	99.71%
		6-20			251	2	0	100.00%	99.21%
		All			2403	6	0	100.00%	99.75%
SF-7430-13589	SNV (het)	All			238928	99	34	99.99%	99.96%
	SNV (hom)	All			155135	95	78	99.95%	99.94%
	SNV (hemi)	All			4008	0	0	100.00%	100.00%
	Insertion	1-2			1462	1	0	100.00%	99.93%
		3-5			965	5	0	100.00%	99.48%
		6-20			543	27	0	100.00%	**95.26%**
		All			2970	33	0	100.00%	**98.90%**
	Deletion	1-2			1704	2	0	100.00%	99.88%

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		3-5	(b)(4)		902	3	0	100.00%	99.67%
		6-20			323	0	1	99.69%	100.00%
		All			2929	5	1	99.97%	99.83%
SF-7804-33749	SNV (het)	All			258339	87	56	99.98%	99.97%
	SNV (hom)	All			157997	0	8	99.99%	100.00%
	SNV (hemi)	All			5001	0	0	100.00%	100.00%
	Insertion	1-2			1289	5	0	100.00%	99.61%
		3-5			972	3	0	100.00%	99.69%
		6-20			386	0	1	99.74%	100.00%
		All			2647	8	1	99.96%	99.70%
	Deletion	1-2			1734	1	0	100.00%	99.94%
		3-5			1216	4	0	100.00%	99.67%
		6-20			392	1	2	99.49%	99.75%
		All			3342	6	2	99.94%	99.82%
SF-8038-20902	SNV (het)	All			259007	111	81	99.97%	99.96%
	SNV (hom)	All			163178	33	9	99.99%	99.98%
	Insertion	1-2			1714	1	0	100.00%	99.94%
		3-5			895	6	0	100.00%	99.33%
		6-20			492	21	3	99.39%	**95.91%**
		All			3101	28	3	99.90%	99.11%
	Deletion	1-2			1373	6	1	99.93%	99.56%
		3-5			571	7	0	100.00%	**98.79%**
		6-20			486	0	0	100.00%	100.00%
		All			2430	13	1	99.96%	99.47%
SF-8085-16718	SNV (het)	All			252881	101	23	99.99%	99.96%
	SNV (hom)	All			153650	83	102	99.93%	99.95%
	SNV (hemi)	All			4239	0	0	100.00%	100.00%
	Insertion	1-2			1320	12	0	100.00%	99.10%
		3-5			889	1	0	100.00%	99.89%
		6-20			402	14	0	100.00%	**96.63%**
		All			2611	27	0	100.00%	**98.98%**
	Deletion	1-2			1514	0	0	100.00%	100.00%
		3-5			900	3	0	100.00%	99.67%
		6-20			333	0	0	100.00%	100.00%
		All			2747	3	0	100.00%	99.89%
SF-8763-29117	SNV (het)	All			208231	49	31	99.99%	99.98%
	SNV (hom)	All			158564	80	81	99.95%	99.95%

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	SNV (hemi)	All	(b)(4)		4687	0	0	100.00%	100.00%
	Insertion	1-2			1304	20	5	99.62%	**98.49%**
		3-5			801	3	0	100.00%	99.63%
		6-20			304	4	0	100.00%	**98.70%**
		All			2409	27	5	99.79%	**98.89%**
	Deletion	1-2			1623	1	1	99.94%	99.94%
		3-5			602	2	0	100.00%	99.67%
		6-20			482	0	1	99.79%	100.00%
		All			2707	3	2	99.93%	99.89%

Table 14. Precision using Clinical Samples - Priority

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
SF-1070-81401	SNV (het)	All	(b)(4)		43666	5	6	99.99%	99.99%
	SNV (hom)	All			27588	0	0	100.00%	100.00%
	Insertion	1-2			432	0	0	100.00%	100.00%
		3-5			144	0	0	100.00%	100.00%
		All			576	0	0	100.00%	100.00%
	Deletion	1-2			830	9	9	**98.93%**	**98.93%**
		3-5			72	0	0	100.00%	100.00%
		All			902	9	9	99.01%	99.01%
SF-1901-15675	SNV (het)	All			39686	3	0	100.00%	99.99%
	SNV (hom)	All			23218	0	0	100.00%	100.00%
	SNV (hemi)	All			823	0	0	100.00%	100.00%
	Insertion	1-2			344	1	0	100.00%	99.71%
		3-5			69	0	0	100.00%	100.00%
		6-20			43	0	14	**75.44%**	100.00%
		All			456	1	14	**97.02%**	99.78%
	Deletion	1-2			594	0	0	100.00%	100.00%
		3-5			138	0	0	100.00%	100.00%
		All			732	0	0	100.00%	100.00%
SF-2711-11665	SNV (het)	All			43641	1	4	99.99%	100.00%
	SNV (hom)	All			23798	17	0	100.00%	99.93%
	SNV (hemi)	All			932	0	0	100.00%	100.00%
	Insertion	1-2			425	17	0	100.00%	**96.15%**

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		All	(b)(4)		425	17	0	100.00%	**96.15%**
	Deletion	1-2			557	0	0	100.00%	100.00%
		3-5			144	0	0	100.00%	100.00%
		6-20			72	0	0	100.00%	100.00%
		All			773	0	0	100.00%	100.00%
SF-2757-45105	SNV (het)	All			46318	3	3	99.99%	99.99%
	SNV (hom)	All			24729	1	0	100.00%	100.00%
	Insertion	1-2			572	0	0	100.00%	100.00%
		3-5			144	0	0	100.00%	100.00%
		All			716	0	0	100.00%	100.00%
	Deletion	1-2			847	0	0	100.00%	100.00%
		3-5			144	0	0	100.00%	100.00%
		All			991	0	0	100.00%	100.00%
SF-2856-29005	SNV (het)	All			36193	4	0	100.00%	99.99%
	SNV (hom)	All			21576	0	1	100.00%	100.00%
	Insertion	1-2			179	1	0	100.00%	99.44%
		3-5			60	0	0	100.00%	100.00%
		All			239	1	0	100.00%	99.58%
	Deletion	1-2			579	0	0	100.00%	100.00%
		3-5			120	0	0	100.00%	100.00%
		All			699	0	0	100.00%	100.00%
SF-3002-89172	SNV (het)	All			42057	3	2	100.00%	99.99%
	SNV (hom)	All			26845	1	1	100.00%	100.00%
	Insertion	1-2			288	0	0	100.00%	100.00%
		3-5			72	0	0	100.00%	100.00%
		6-20			72	0	0	100.00%	100.00%
		All			432	0	0	100.00%	100.00%
	Deletion	1-2			616	0	0	100.00%	100.00%
		3-5			144	0	0	100.00%	100.00%
		All			760	0	0	100.00%	100.00%
SF-3312-24549	SNV (het)	All			44982	13	7	99.98%	99.97%
	SNV (hom)	All			23177	0	0	100.00%	100.00%
	Insertion	1-2			207	0	0	100.00%	100.00%
		All			207	0	0	100.00%	100.00%
	Deletion	1-2			598	0	0	100.00%	100.00%
		3-5			138	0	0	100.00%	100.00%
		All			736	0	0	100.00%	100.00%

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
SF-3607-46036	SNV (het)	All	(b)(4)		40605	3	0	100.00%	99.99%
	SNV (hom)	All			25428	0	0	100.00%	100.00%
	Insertion	1-2			344	1	0	100.00%	99.71%
		3-5			69	0	0	100.00%	100.00%
		6-20			138	0	0	100.00%	100.00%
		All			551	1	0	100.00%	99.82%
	Deletion	1-2			580	0	0	100.00%	100.00%
		3-5			138	0	0	100.00%	100.00%
		All			718	0	0	100.00%	100.00%
SF-5044-81848	SNV (het)	All			44535	3	0	100.00%	99.99%
	SNV (hom)	All			25362	16	0	100.00%	99.94%
	SNV (hemi)	All			1072	0	0	100.00%	100.00%
	Insertion	1-2			431	16	0	100.00%	**96.42%**
		3-5			72	1	0	100.00%	**98.63%**
		All			503	17	0	100.00%	**96.73%**
	Deletion	1-2			764	0	0	100.00%	100.00%
		3-5			72	0	0	100.00%	100.00%
		All			836	0	0	100.00%	100.00%
SF-5221-11824	SNV (het)	All			45197	6	0	100.00%	99.99%
	SNV (hom)	All			26751	0	0	100.00%	100.00%
	SNV (hemi)	All			1283	0	0	100.00%	100.00%
	Insertion	1-2			285	0	0	100.00%	100.00%
		3-5			0	1	0	NA	0.00%
		All			285	1	0	100.00%	99.65%
	Deletion	1-2			623	0	0	100.00%	100.00%
		3-5			216	0	0	100.00%	100.00%
		6-20			72	0	0	100.00%	100.00%
		All			911	0	0	100.00%	100.00%
SF-5730-56578	SNV (het)	All			40040	4	0	100.00%	99.99%
	SNV (hom)	All			24352	0	0	100.00%	100.00%
	Insertion	1-2			510	0	0	100.00%	100.00%
		All			510	0	0	100.00%	100.00%
	Deletion	1-2			679	0	0	100.00%	100.00%
		3-5			128	0	0	100.00%	100.00%
		All			807	0	0	100.00%	100.00%
SF-5744-44790	SNV (het)	All			38154	23	1	100.00%	99.94%
	SNV (hom)	All			22376	0	20	99.91%	100.00%

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	SNV (hemi)	All	(b)(4)		814	0	0	100.00%	100.00%
	Insertion	1-2			377	0	0	100.00%	100.00%
		3-5			126	0	0	100.00%	100.00%
		All			503	0	0	100.00%	100.00%
	Deletion	1-2			480	0	0	100.00%	100.00%
		3-5			189	0	0	100.00%	100.00%
		All			669	0	0	100.00%	100.00%
SF-7430-13589	SNV (het)	All			42258	11	25	99.94%	99.97%
	SNV (hom)	All			23196	24	0	100.00%	99.90%
	SNV (hemi)	All			899	0	0	100.00%	100.00%
	Insertion	1-2			210	0	0	100.00%	100.00%
		3-5			140	0	0	100.00%	100.00%
		6-20			70	7	0	100.00%	**90.91%**
		All			420	7	0	100.00%	**98.36%**
	Deletion	1-2			476	0	0	100.00%	100.00%
		3-5			140	0	0	100.00%	100.00%
		All			616	0	0	100.00%	100.00%
SF-7804-33749	SNV (het)	All			48093	2	1	100.00%	100.00%
	SNV (hom)	All			25810	0	0	100.00%	100.00%
	SNV (hemi)	All			1143	0	0	100.00%	100.00%
	Insertion	1-2			425	0	0	100.00%	100.00%
		3-5			144	0	0	100.00%	100.00%
		All			569	0	0	100.00%	100.00%
	Deletion	1-2			491	0	0	100.00%	100.00%
		3-5			359	0	0	100.00%	100.00%
		All			850	0	0	100.00%	100.00%
SF-8038-20902	SNV (het)	All			38901	2	2	99.99%	99.99%
	SNV (hom)	All			28811	0	1	100.00%	100.00%
	Insertion	1-2			431	1	0	100.00%	99.77%
		3-5			72	0	0	100.00%	100.00%
		All			503	1	0	100.00%	99.80%
	Deletion	1-2			624	0	0	100.00%	100.00%
		3-5			72	0	0	100.00%	100.00%
		All			696	0	0	100.00%	100.00%
SF-8085-16718	SNV (het)	All			40463	3	0	100.00%	99.99%
	SNV (hom)	All			26918	10	0	100.00%	99.96%

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	SNV (hemi)	All	(b)(4)		563	0	0	100.00%	100.00%
	Insertion	1-2			286	10	0	100.00%	**96.62%**
		3-5			72	0	0	100.00%	100.00%
		All			358	10	0	100.00%	**97.28%**
	Deletion	1-2			474	0	0	100.00%	100.00%
		3-5			216	0	0	100.00%	100.00%
		All			690	0	0	100.00%	100.00%
SF-8763-29117	SNV (het)	All			33518	0	2	99.99%	100.00%
	SNV (hom)	All			26385	76	64	99.76%	99.71%
	SNV (hemi)	All			703	0	0	100.00%	100.00%
	Insertion	1-2			318	12	0	100.00%	**96.36%**
		3-5			64	0	0	100.00%	100.00%
		All			382	12	0	100.00%	**96.95%**
	Deletion	1-2			492	0	0	100.00%	100.00%
		3-5			128	0	0	100.00%	100.00%
		All			620	0	0	100.00%	100.00%

Precision Reference– Metrics: HLP has variable coverage across the coding regions with highest coverage for select genes with the highest clinical relevance (i.e., Priority). The coverage is expected to vary depending on the genomic region. (b)(4) (b)(4) (b)(4) Callability, percentage of based with coverage ≥ 20x, and mean coverage are summarized for all samples and is shown in **Table 15a (**Cell Lines) **and Table 15b** (Saliva)**.**

Table 15a. Precision 6 Cell Line Samples

(b)(4)

Table 15b. Precision of 17 Clinical Samples

(b)(4)

iii. Between-Lot Reproducibility

The reproducibility across multiple lots of the Helix *Exome+* Reagents library preparation and enrichment reagents, and the Helix Reagent Kit, which includes clustering and sequencing reagents was evaluated. A total of 24 samples (6 cell lines and 18 saliva-derived DNAs) were selected for this study. These 24 samples, replicated 3 times on each of 3 plates, were processed through the library preparation workflow (3 replicates/sample/plate x 3 plates = 9 replicates/sample), to generate a total of 216 libraries at the end of this step (24 samples x 3 replicates/sample/plate x 3 plates = 216 libraries). Each library was used twice during enrichment, increasing the number of replicates per sample to 18, in a combination that allowed all libraries generated with different lots of reagents to be tested against all 3 lots of enrichment reagents.

Three lots of enrichment reagents were used to process (b)(4) libraries per enrichment, generating a total of (b)(4) enriched samples. Each enriched sample was clustered and sequenced in 3 independent events to produce a total of 1296 sequenced samples. Each of the 24 starting samples produced a total of 54 replicate sequences that were processed with combinatorial sets of critical reagents through the Helix workflow, so that lot variation

could be assessed (9 library prep replicates x 2 enrichments/replicate x 3 sequencing events/enriched library = 54 total replicates).

Out of 1296 samples, 1287 sample replicates passed all QC and were used in the final data analyses presented in this report. For sequenced samples to be included in data analyses, the sample(s) must pass all the evaluability thresholds. **Table 16** shows the number of samples that were evaluable. All samples used in this study that passed the evaluability thresholds were used in the analysis.

Table 16. Post-Sequencing Sample Evaluability

Group Name	Number of Biosamples	Number of Biosamples Used for Analysis
NA12877	54	54
NA12878	54	54
NA24143	54	54
NA24149	54	54
NA24385	54	54
NA24631	54	48
SF-1926-57155	54	54
SF-2911-23958	54	54
SF-3016-50811	54	54
SF-3171-12777	54	54
SF-3724-39960	54	54
SF-3762-53211	54	54
SF-4908-78821	54	53
SF-5309-54434	54	54
SF-5727-76022	54	54
SF-6482-77281	54	54
SF-7909-95368	54	54
SF-8249-10361	54	54
SF-8270-35552	54	54
SF-8472-70262	54	52
SF-8478-76128	54	54
SF-8568-55453	54	54
SF-9559-80394	54	54
SF-9984-59520	54	54

A reference sequence for each clinical sample was generated by majority call comparison over all replicates of a sample in this study. The number of SNVs, Insertions, and Deletions calculated for each sample is shown in **Table 17**.

Performance metrics in the Between-lot study for both Cell lines and Clinical specimens were evaluated. **Table 18** shows the mean callability, the mean percentage of the region with ≥ 20x voerage, and the average coverage for each region (coding, mendeliome and priority).

Table 18. Between-Lot Reproducibility Study Summary of Performance Metrics -Cell Lines and Clinical Specimens

Sample Name	Mean Callability – Coding	Mean Callability - Mendeli-ome	Mean Callability – Priority	Mean Percent >=20x Coding	Mean Percent >=20x Mendeli-ome	Mean Percent >=20x Priority	Average Coverage – Coding	Average Coverage – Mendeli-ome	Average Coverage - Priority
NA12877	(b)(4)								
NA12878									
NA24143									
NA24149									
NA24385									
NA24631									
SF-1926-57155									
SF-2911-23958									
SF-3016-50811									
SF-3171-12777									

Sample Name	Mean Callability – Coding	Mean Callability - Mendeli-ome	Mean Callability – Priority	Mean Percent >=20x Coding	Mean Percent >=20x Mendeli-ome	Mean Percent >=20x Priority	Average Coverage – Coding	Average Coverage – Mendeli-ome	Average Coverage - Priority
SF-3724-39960	(b)(4)								
SF-3762-53211									
SF-4908-78821									
SF-5309-54434									
SF-5727-76022									
SF-6482-77281									
SF-7909-95368									
SF-8249-10361									
SF-8270-35552									
SF-8472-70262									
SF-8478-76128									
SF-8568-55453									
SF-9559-80394									
SF-9984-59520									

The mean PPA and mean PPA lower bound, mean TPPV, mean TPPV lower bound, mean NPA and mean NPA lower bound for each library prep lot, enrichment lot, clustering lots, and sequencing lots and for select ClinVar variants, which also included mean NRC analytical range and mean NRC clinical variants were calculated. All pre-defined study acceptance criterial were met for SNVs, insertions and deletions combined. Acceptance criteria were not established for data when sorted into size ranges, however, as highlighted in the table and consistent with the precision studies, insertion calling (primarily (b)(4) in some samples was less accurate (**Table 19-21**).

Table 19. Between-Lot Reproducibility -Coding Region; Cell Lines and Clinical Specimens

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
NA12877	SNV (het)	All	(b)(4)						
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		6-20	(b)(4)						
		All							
NA12878	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
NA24143	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
NA24149	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
NA24385	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		All	(b)(4)						
	Deletion	1-2							
		3-5							
		6-20							
		All							
NA24631	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-1926-57155	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-2911-23958	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-3016-50811	SNV (het)	All							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	SNV (hom)	All	(b)(4)						
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-3171-12777	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-3724-39960	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-4908-78821	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		All				(b)(4)			
	Deletion	1-2							
		3-5							
		6-20							
		All							
	SNV (het)	All							
	SNV (hom)	All							
SF-4908-78821	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-5309-54434	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-5727-76022	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
SF-6482-77281	SNV (het)	All	(b)(4)						
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-7909-95368	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-8249-10361	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-8270-35552	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		All	(b)(4)						
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-8472-70262	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-8478-76128	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-8568-55453	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-9559-80394	SNV (het)	All							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	SNV (hom)	All	(b)(4)						
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-9984-59520	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							

Table 20. Between-Lot Reproducibility – Mendeliome; Cell Lines and Clinical Specimens

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
NA12877	SNV (het)	All	(b)(4)						
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
NA12878	SNV (het)	All							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
	SNV (hom)	All	(b)(4)						
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
NA24143	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
NA24149	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
NA24385	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		6-20	(b)(4)						
		All							
NA24631	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-1926-57155	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-2911-23958	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-3016-50811	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		3-5	(b)(4)						
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-3171-12777	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-3724-39960	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-3762-53211	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		6-20	(b)(4)						
		All							
SF-4908-78821	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-5309-54434	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-5727-76022	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-6482-77281	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		3-5	(b)(4)						
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-7909-95368	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-8249-10361	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-8270-35552	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		6-20	(b)(4)						
		All							
SF-8472-70262	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-8478-76128	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-8568-55453	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-9559-80394	SNV (het)	All							
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
		3-5	(b)(4)						
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							
SF-9984-59520	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		6-20							
		All							

Table 21: Between-Lot Reproducibility – Priority; Cell Lines and Clinical Specimens

Sample Name	Variant Type	Variant Length	Number of Expected Variants	Number of No-calls	TP	FP	FN	PPA	TPPV
NA12877	SNV (het)	All	(b)(4)						
	SNV (hom)	All							
	SNV (hemi)	All							
	Insertion	1-2							
		3-5							
		6-20							
		All							
	Deletion	1-2							
		3-5							
		All							
NA12878	SNV (het)	All							
	SNV (hom)	All							
	Insertion	1-2							
		3-5							
		All							
	Deletion	1-2							

			(b)(4)
		3-5	
		All	
NA24143	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		6-20	
		All	
	Deletion	1-2	
		3-5	
		All	
NA24149	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
NA24385	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
NA24631	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-1926-57155	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	

			(b)(4)
		3-5	
		All	
SF-2911-23958	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-3016-50811	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-3171-12777	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-3724-39960	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-3762-53211	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	

		3-5	(b)(4)
		All	
	Deletion	1-2	
		3-5	
		All	
SF-4908-78821	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	
		3-5	
		6-20	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-5309-54434	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		6-20	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-5727-76022	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-6482-77281	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	

			(b)(4)
SF-7909-95368	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	
		3-5	
		6-20	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-8249-10361	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-8270-35552	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		6-20	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-8472-70262	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-8478-76128	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		All	

			(b)(4)
	Deletion	1-2	
		3-5	
		All	
SF-8568-55453	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	
		3-5	
		6-20	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-9559-80394	SNV (het)	All	
	SNV (hom)	All	
	SNV (hemi)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	
SF-9984-59520	SNV (het)	All	
	SNV (hom)	All	
	Insertion	1-2	
		3-5	
		All	
	Deletion	1-2	
		3-5	
		All	

iv. Precision – GC content

The performance of variant calling based on the GC content of sequence context in coding regions was examined across the following ranges of GC content: 0-15%, >15-40%, >40-65%, >65-90%, and >90-100% in two data sets from the Precision and Between-Lot reproducibility studies. The mean PPA and mean TPPV for each sample is shown below (**Table 22, 23**). All data passed the specifications with the exception of indel calling in some samples, however this was due to indel size and not due to GC content. The sub-optimal performance of insertion calling for GC content at range of >15-40 was confounded with size of 6-20bp indels (**Figure 12**). Nonetheless, Indels in regions with GC

content >65% are excluded from reporting as these specimens were excluded from the data analyses. Acceptance criteria were the same as for all the studies and data less than the overall acceptance criteria is shown in bold below.

Table 22: Variant calling Performance Based on GC Content in the Precision studies

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
NA12877	0-15	INSERTION	(b)(4)			
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				
NA12878	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
		SNV (hom)	(b)(4)			
	>90-100	SNV (het)				
		SNV (hom)				
NA24143	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				
NA24149	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
NA24385	0-15	DELETION	(b)(4)			
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				
NA24631	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				
SF-1070-81401	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
	>15-40	DELETION	(b)(4)			
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-1901-15675	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
SF-2711-11665	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
		SNV (hom)	(b)(4)			
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
SF-2757-45105	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-2856-29005	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
	>65-90	SNV (het)	(b)(4)			
		SNV (hom)				
SF-3002-89172	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-3312-24549	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-3607-46036	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
		SNV (hom)	(b)(4)			
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-5044-81848	0-15	INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
SF-5221-11824	0-15	INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
		SNV (hemi)	(b)(4)			
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
SF-5730-56578	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-5744-44790	0-15	INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
	>65-90	SNV (hemi)	(b)(4)			
		SNV (het)				
		SNV (hom)				
SF-7430-13589	0-15	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
SF-7804-33749	0-15	INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
	>65-90	SNV (hemi)	(b)(4)			
		SNV (het)				
		SNV (hom)				
SF-8038-20902	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-8085-16718	0-15	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	0-15	INSERTION				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Percentage of Variants in Reference Data with No-Call	Mean PPA	Mean TPPV
SF-8763-29117		SNV (hemi)	(b)(4)			
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				

Table 23: Variant calling Performance Based on GC Content in the Between-Lot Reproducibility Study

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
NA12877	0-15	INSERTION	(b)(4)			
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
		INSERTION	(b)(4)			
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				
NA12878	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				
NA24143	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
	>65-90	SNV (het)		(b)(4)		
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				
NA24149	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				
NA24385	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
		SNV (het)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
	>90-100	SNV (hom)	(b)(4)			
NA24631	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
		SNV (hom)				
SF-1926-57155	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-2911-23958	0-15	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
		DELETION	(b)(4)			
		INSERTION				
	>15-40	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
		DELETION				
		INSERTION				
	>40-65	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
		SNV (hemi)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-3016-50811		DELETION				
		INSERTION				
	0-15	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
		DELETION				
		INSERTION				
	>15-40	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
		DELETION				
		INSERTION				
	>40-65	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
		SNV (hemi)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
SF-3171-12777	0-15	DELETION				
		INSERTION				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
		SNV (hemi)	(b)(4)			
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
SF-3724-39960	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
SF-3762-53211	0-15	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
		SNV (hom)	(b)(4)			
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
SF-4908-78821	0-15	INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
SF-5309-54434	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
		INSERTION	(b)(4)			
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-5727-76022	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>90-100	SNV (hom)				
SF-6482-77281	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
	>40-65	DELETION	(b)(4)			
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
SF-7909-95368	0-15	INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
SF-8249-10361	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
		SNV (hom)	(b)(4)			
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
SF-8270-35552	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
SF-8472-70262	0-15	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	0-15	INSERTION				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
SF-8478-76128		SNV (het)	(b)(4)			
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
SF-8568-55453	0-15	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				
SF-9559-80394	0-15	INSERTION				
		SNV (hemi)				

Sample Name	GC content range	Variant Type	Number of Variants in Reference Data	Mean Number of False Negatives	Mean PPA	Mean TPPV
		SNV (het)	(b)(4)			
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (hemi)				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (hemi)				
		SNV (het)				
		SNV (hom)				
SF-9984-59520	0-15	INSERTION				
		SNV (het)				
		SNV (hom)				
	>15-40	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>40-65	DELETION				
		INSERTION				
		SNV (het)				
		SNV (hom)				
	>65-90	SNV (het)				
		SNV (hom)				
	>90-100	SNV (het)				

Figure 12. Performance based on the GC content and indel size in the Precision and Between-

Lot Reproducibility Studies

b. Linearity/assay Reportable Range:

Not applicable.

c. Detection Limit

i. Limit of Detection

Not applicable. The HLP has a requirement for a minimum of 50ng of DNA.

ii. DNA Input

To assess the impact of different levels of DNA input on the HLP, 20 samples with known variants (single nucleotide variants, small insertions or small deletions) in medically related genes were tested across 4 input conditions spanning 10-fold below and 2 fold above the input amount of 50ng. In total, 20 samples were tested at 35ng, 50ng, 70ng and 100ng DNA input, each in triplicate, for a total of 240 samples (20 unique samples x 4 conditions x 3 replicates/condition = 240 samples). Of these 240 samples, 21 samples failed to meet secondary sequencing metrics QC for evaluability leaving 219 samples for use in the final data analyses. Replicates of the 50ng and 70ng (the optimal range of DNA input) input samples were used to determine the majority call at each position in the reportable range in order to generate a reference sequence for each sample. Sensitivity (PPA), specificity

(NPA) and precision (TPPV) were calculated for each sample and the mean of these values for all samples per condition was determined (**Table 24**).

This study met all acceptance criteria for mean PPA, NPA and TPPV. The study yielded concordant test results for all samples with a DNA input range between 35 to 70 ng. Additional specimen is required when there is less than 50ng.

Table 24: Comparison to Majority Call PPA (SNV and Indels)

Group Name	Number of Replicates	Mean PPA	Mean PPA Lower Bound	Mean NPA	Mean NPA Lower Bound	Mean TPPV	Mean TPPV Lower Bound
100 ng	55	0.999797	0.999704	1.000000	0.999999	0.999721	0.999615
35 ng	54	0.999756	0.999658	1.000000	0.999999	0.999712	0.999608

d. Traceability

The Helix Reference Genome is based on the Human Reference Genome Consortium Build 38 assembly with additional alternative contigs, including decoy sequences.

e. Analytical Specificity

i. Index Swapping -Barcoding

Sample index primers are used in the HLP to assign a unique barcode combination to each sample DNA during library preparation, allowing for the ability to pool multiple samples downstream for enrichment and sequencing processes. The HLP uses 2- unique index primers which are combined into 96 unique combinations (12x8) for sample assignment. Index swapping (also referred to as index hopping, index misassignment) can occur between samples and cause the reporting of incorrect DNA sequence data between patients. The percentage of index swapping occurring with the Helix platform and its impact on sequence accuracy was investigated.

The performance evaluated agreement between replicates of DNA isolated from 48 saliva samples with known variants (40 deletions, 9 insertions and 81 SNVs) and across the analytical range as a measure of whether index swapping affected accuracy. In total, 48 samples were run in triplicate, each down three columns of a library plate, testing overlapping indices across a row. Eight (8) of those samples were tested another two times, varying the position of sample by row to test overlapping indices down a column. A total of 160 libraries (40x3 plus 8x5) were generated. Performance was evaluated by a replicate pairwise comparison. Due to sample failure, one index comparison could not be made. A total of 163 pairwise comparisons met acceptance criteria for NRC. Only data passing the quality metrics (callability) and other contamination evaluation metrics that are applied prior to reporting. Of 160 total samples, 157 were used in the analyses; 2

samples failed to meet callability and coverage metrics and a third sample was flagged during a software-based error check. The mean NRC across the analytical range was 0.999 and for clinical variants it was 1.0; one SNV was a no call.

ii. Index Swapping - Carry-Over and Cross- Contamination

The Helix laboratory platform follows a high-throughput workflow that processes increasing numbers of multiplexed samples in each subsequent step of the HLP workflow. As enriched library pools undergo flow cell clustering and sequencing on the cBot2 and HiSeq instruments, respectively, it may be possible that carryover of barcoded DNAs may be transferred when subsequent samples are run on the same instrument. The HLP uses a software that verifies whether the reads in a particular file match previously known genotypes for an individual (or group of individuals) and checks whether the reads are contaminated as a mixture of two or more samples to estimate possible contamination between samples and exclude samples containing a mixture of DNA from multiple individuals. The quality metric for this evaluation is that they must be less than (b)(4)

(1) Carryover study 1: This study evaluated the potential for inter-run carryover DNA introduced at the clustering and sequencing steps of the HLP workflow.

A total of four cell line DNAs with a known reference sequence was used to create two plates of samples for library prep and enrichment. Plate 1 consists of 36 replicates of each NA12877 and NA12878 arranged in a checkerboard pattern, with alternating High DNA input (100 ng) into library prep or Low DNA input into library prep in adjacent wells. Plate 2 contained 48 replicates of NA24143 at standard production input of 70ng in rows C - F, and 24 replicates of NA24695 in rows G-H. Sample NA24695 is a female sample and is included on Plate 2 as filler to allow a full pool of 72 samples contributing to the sequencing pool from this plate. NA24695 will not be included in any analyses for this protocol.

Each plate of samples produced one sequencing pool. The sequencing pool of samples from Plate 1 were clustered on a single flow cell and sequenced on a single HiSeq. The sequencing pool from Plate 2 were clustered on a single flow cell using the same cBot as for Plate 1, and immediately after the Plate 1 clustering event was complete. The Plate 2 flow cell was sequenced immediately after the Plate 1 flow cell sequencing is complete, using the same HiSeq as the Plate 1 flow cell.

There were 48 index pairs used on both Plates 1 and 2. If inter-run carryover occurred, it is possible that the reads from Plate 1 could be assigned to a sample on Plate 2 that has the same index pair. The samples used in this study (Plate 1 = NA12878, NA12877; Plate 2 = NA24143) have known reference sequences. Plate 2 samples (NA24143) were evaluated for non-reference concordance (NRC) compared to the known reference data for variant sites that differ between the Plate 1 and Plate 2 samples. The 95% lower bound for all results show all samples had NRC >99.5% (data not shown). Samples on Plate 1 use 24 index pairs that are not

used on Plate 2. Barcode carryover between runs was determined by calculating the amount, if any, of the 24 index pairs from Plate 1 that may be present in Plate 2 data. Barcode carryover was not observed between runs.

(2) Carryover study 2: This study evaluated the potential for intra-run carryover DNA and its impact on the performance of the Helix Laboratory Platform.

Two cell line DNAs with known variants were used to create 36 replicates per sample, for a total of 72 samples tested (2 cell lines x 36 replicates per cell line = 72 libraries). DNAs were used to create a checkerboard pattern with DNA input amounts of either High DNA input (100ng) or Low DNA input (50ng) in adjacent wells, so that any spillover from high to low will be detected as contamination. Each sample was indexed with its own unique barcode. The analysis was conducted to detect whether DNA fragments from High DNA input are carried over to the Low DNA input, and to determine if carryover would affect performance of the Helix Laboratory Platform. Three analyses were performed.

All 36 female samples met secondary sequencing metric QC for evaluability. Performance was evaluated by calculating the amount of chromosome Y carryover into female samples and by calculating contamination using the Freemix value. Intra-run carryover protocol samples processed through the HLP workflow met the acceptance criteria for this study, indicating that intra-run carryover does not impact the performance of the Helix Laboratory Platform.

iii. *Interfering Substances*

(1) Endogenous Substances

A study was conducted to evaluate the potential impact of four endogenous substances common to saliva specimens (alpha-amylase, hemoglobin, IgA and total protein (albumin)) as potential sources of interference. To assess the impact of these four endogenous substances as an interfering substance on the Helix Laboratory Platform, the performance of the HLP was evaluated on donor saliva

samples after the addition of 4 different endogenous substances as compared to no treatment samples from the same donors.

Sixty (60) donors submitted two saliva samples each for this study which were then combined and aliquoted into 3 separate tubes. Donors were separated into 2 groups; one group received amylase and hemoglobin and the other group received IgA and albumin. For each group, a donor aliquot was processed without the addition of a substance (no treatment) and the two additional aliquots each had different endogenous substances added (amylase and hemoglobin, or IgA and albumin). Each of the no treatment samples had DNA extracted and were tested in triplicate. The treatment samples had DNA extracted and were tested as one replicate.

In total, 180 no treatment libraries (60 donors x 3 replicates/extracted DNA) and 120 treatment (amylase, hemoglobin, IgA and albumin) libraries (30 donors/group x 2 treatments/group x 2 groups) were generated for enrichment and sequencing on the Helix Laboratory Platform. Of the 300 samples in the study, 299 met secondary sequencing metrics QC for evaluability and were used in the final data analyses for this study. The one sample that failed QC metrics was not replaced and was not used in data analysis, as per protocol.

A reference for each saliva sample was generated by majority call comparison over all control replicates of a sample in this study. The number of SNVs, Insertions, Deletions, Reference Bases and Bases with No Majority was calculated for each sample. Majority call for each donor was used as a reference for comparison to the corresponding treated samples (albumin, amylase, hemoglobin, IgA).

Each sample in albumin, amylase, hemoglobin and IgA groups was compared to reference (majority call of all no treatment replicates for each sample) and concordance was calculated. Mean values of NPA, PPA and TPPV were calculated for all samples combined per condition. The NPA is calculated using only SNV; PPA and TPPV is calculated using SNV and indels combined. The acceptance criteria are: NPA $\geq$ 99.99%; PPA $\geq$ 99.5%; TPPV $\geq$ 99.5%; all with a 95% confidence interval lower bound of 99.0%.

All conditions passed the Acceptance Criteria suggesting that normally occurring levels of albumin, amylase, hemoglobin and IgA groups in saliva does not interfere with HLP performance (Table 25).

Table 25. Endogenous Interferents

Interferent	Number of Replicates	Mean PPA	Mean PPA Lower Bound	Mean NPA	Mean NPA Lower Bound	Mean TPPV	Mean TPPV Lower Bound
Albumin	30	0.9997	0.9996	1.0000	0.9999	0.9996	0.9995
Amylase	30	0.9997	0.9996	1.0000	0.9999	0.9996	0.9995

Hemoglobin	29	0.9997	0.9996	1.0000	0.9999	0.9996	0.9995
IgA	30	0.9997	0.9996	1.0000	0.9999	0.9996	0.9995

(2) Exogenous Substances

To assess the impact of 4 exogenous substances (eating, drinking, chewing gum and mouthwash) as an interfering substance on the Helix Laboratory Platform, the performance of the HLP was evaluated on saliva samples either prior to (control), immediately after or 30 minutes after consumption of an exogenous substance. The samples collected prior to consumption were used to construct a reference by majority call across all replicates from each donor before consumption samples. The performance of the HLP was evaluated by comparing the majority call reference to the data for samples collected immediately after consumption; food, drink, chewing gum, mouthwash, and 30 minutes after consumption; food, drink, chewing gum, and mouthwash. **Table 26** describes the number of biosamples at the start of the study and the number of evaluable biosamples that passed QC metrics.

Table 26. Post-Sequencing Sample Evaluability for Exogenous Substances

Condition Tested	Number of Biosamples	Number of Evaluable Biosamples
30 minutes after drink	(b)(4)	
30 minutes after food		
30 minutes after gum		
30 minutes after mouthwash		
Before Consumption - drink		
Before Consumption - food		
Before Consumption - gum		
Before Consumption - mouthwash		
Immediately after drink		
Immediately after food		
Immediately after gum		
Immediately after mouthwash		

Twenty (20) initial donors (5 per treatment group) submitted three samples each for this study, for a total of 60 saliva samples. An additional two donors were added for the food group because two samples failed to yield enough purified DNA from the saliva sample. In sum, there were 22 donors and 66 saliva samples collected for this study. In total,198 samples (22 donors x 3 treatments/donor x 3 replicates/extracted DNA), were processed on the Helix Laboratory Platform. Performance was evaluated by comparing the post consumption samples to the corresponding donor majority call reference for concordance.

Sensitivity (PPA), specificity (NPA) and precision (TPPV) were calculated for each replicate and the mean of these values for all samples per condition was compared to the acceptance criteria. Three groups, drink, chewing gum and mouthwash, passed all acceptance criteria at all timepoints. In the 'immediately after food' group (Condition 2), mean NPA and TPPV failed to meet acceptance criteria. This failure is attributed to one poorly performing sample which failed acceptance criteria for NPA and TPPV. The 30 minutes after food sample from this donor passed acceptance criteria, and the mean values for all samples 30 min after food passed all acceptance criteria. If this sample is removed from the group calculation, the mean NPA and TPPV pass acceptance criteria (**Table 27**).

These results indicate that saliva samples should be collected at least 30 minutes after consuming food, as recommended by the manufacturer (DNA Genotek), and as per the collection device instructions.

Table 27. Exogenous Interference

Condition Tested	Mean PPA	Mean PPA Lower Bound	Mean NPA	Mean NPA Lower Bound	Mean TPPV	Mean TPPV Lower Bound
30 minutes after drink	0.9997	0.9996	1.0000	1.0000	0.9997	0.9996
30 minutes after food	0.9997	0.9996	1.0000	0.9999	0.9996	0.9994
30 minutes after gum	0.9997	0.9996	1.0000	0.9999	0.9996	0.9995
30 minutes after mouthwash	0.9997	0.9996	1.0000	0.9999	0.9996	0.9995
Immediately after drink	0.9997	0.9996	1.0000	0.9999	0.9997	0.9995
Immediately after food	0.9988	0.9986	0.99989	0.9999	0.9362	0.9355
Immediately after gum	0.9997	0.9996	1.0000	0.9999	0.9997	0.9995
Immediately after mouthwash	0.9997	0.9996	1.0000	0.9999	0.9997	0.9995

(3) Microbial Interference

To assess the impact of bacteria and yeast as interfering substances on the Helix Laboratory Platform, bacterial DNA from a commercial source (American Type Culture Collection) was added in various amounts into 6 cell line DNA samples. This bacterial sample is comprised of 20 fully sequenced cultures that were selected by the vendor to encompass a variety of characteristics including different gram stain, GC content, genome size and spore formation capabilities. In total, 6 cell lines were tested across 5 conditions, 0%, 10%, 20%, 30%, 50% bacterial content, each in triplicate, for a total of 90 samples. Performance was evaluated by comparing test samples to known truth sample (GIAB or Platinum Genomes) for all evaluable samples (N=81). This study met all acceptance criteria for mean PPA, NPA and TPPV (**Table 28 and Table 29**).

Table 28. Accuracy of Cell line Under Varying Concentrations of Microbial Interference

% microbes	Number of replicates	Mean PPA	Mean PPA Lower Bound	Mean NPA	Mean NPA Lower Bound	Mean TPPV	Mean TPPV Lower Bound
0%	18	0.999691	0.999601	0.9999	0.9999	0.999475	0.999358
10%	18	0.999671	0.999571	0.9999	0.9999	0.999498	0.999374
20%	18	0.999667	0.999558	0.9999	0.9999	0.999478	0.999342
30%	16	0.999685	0.999568	0.9999	0.9999	0.999498	0.999352
50%	11	0.999622	0.999466	0.9999	0.9999	0.999531	0.999357

The impact of interference for clinically relevant variants was evaluated by ClinVar subset analysis of clinical variants found within the Helix reportable range (all likely-pathogenic and pathogenic variants reported in ClinVar that are present in the Helix reportable range after filtering). All likely-pathogenic and pathogenic variants reported in ClinVar that are present in the Helix reportable range after filtering plus additional clinical variants ((b)(4) variant positions total) were assessed for concordance with the known reference data of each cell line (**Table 29**). The results demonstrated that the HLP accuracy is not impacted by microbial interference.

Table 29. ClinVar subset analysis

% microbes	Number of Replicates	Mean NRC Analytical Range	Mean NRC Clinical Variants
0%	18	0.999166	1.000000
10%	18	0.999170	1.000000
20%	18	0.999146	1.000000
30%	16	0.999183	1.000000
50%	11	0.999153	1.000000

In addition, fresh saliva from three (3) donors was used to test three conditions each: 1) baseline with no microbe interferent added, 2) bacteria spiked-in, and 3) yeast spiked-in. Samples were tested in triplicate for a total of 27 samples. The actual amount of microbial contamination in the fresh saliva samples used in this study is estimated. Thirty percent (30%) bacterial DNA was added to the saliva DNA to mimic high levels of contamination, resulting in the estimated final proportion of bacterial DNA to be approximately 40%. Ten percent (10%) C. albicans DNA was added to the saliva DNA to mimic high levels of yeast contamination.

Replicates of the no treatment samples (no spike-in) were used to determine the majority call at each position in the reportable range in order to generate a reference genome sequence for each donor. Performance was evaluated by comparing the treatment samples (bacteria spike-in and yeast spike-in) to the corresponding donor majority call reference for concordance. The NPA was calculated using only SNV; PPA and TPPV is calculated using SNV and indels combined. The acceptance criteria are: NPA ≥ 99.99%; PPA ≥ 99.5%; TPPV ≥ 99.5%; all with a 95% confidence interval lower bound of 99.0%. All conditions passed the Acceptance Criteria suggesting that bacterial and yeast at the levels tested do not interfere with performance (**Table 30**).

Table 30. Accuracy of Clinical Samples with Microbial and Yeast Spike-ins

Spike-in	Number of replicates	Mean PPA	Mean PPA Lower Bound	Mean NPA	Mean NPA Lower Bound	Mean TPPV	Mean TPPV Lower Bound
10% yeast DNA	9	0.9997	0.9996	0.9999	0.9999	0.9996	0.9995
30% bacterial DNA	9	0.9997	0.9996	0.9999	0.9999	0.9996	0.9995

(4) Smoking

To assess the impact of smoking as an interfering substance on the Helix Laboratory Platform, saliva samples were evaluated 60 minutes prior to (Condition1, before smoking), immediately after (Condition 2) and 30 minutes after (Condition 3) smoking. Five saliva donors submitted three samples each for this study, for a total of 15 saliva samples. In total, 45 libraries (3 conditions x 5 donors x 3 replicates/extracted DNA) were generated for enrichment and sequencing on the Helix Laboratory Platform. The dataset from Condition 2 - immediately after smoking and Condition 3 - 30 minutes after smoking treatment groups were compared to the Condition 1- before smoking dataset to assess the impact of smoking on performance. Replicates of the baseline samples, before smoking, were used to determine the majority call at each position in the reportable range in order to generate a reference for each donor. Performance was evaluated by comparing the treatment samples (immediately after smoking, and 30 minutes after smoking) to the corresponding donor majority call reference for concordance. Sensitivity (PPA), specificity (NPA) and precision (TPPV) were calculated for each sample and the mean of these values for all samples per condition was determined (**Table 31**). This study met all acceptance criteria for mean PPA, NPA and TPPV suggesting that saliva collected 60 prior to smoking, immediately after smoking and 30 minutes after smoking does not interfere with performance.

Table 31. Smoking

Condition	Mean PPA	Mean PPA Lower Bound	Mean NPA	Mean NPA Lower Bound	Mean TPPV	Mean TPPV Lower Bound
30 minutes after smoking	0.9997	0.9996	1.00	0.9999	0.9997	0.9996
Immediately after smoking	0.9997	0.9996	1.00	0.9999	0.9997	0.9995

f. Stability

Studies were conducted to evaluate reagent stability and specimen stability.

i. Real Time Kit Stability

Helix conducted studies to confirm the stability of storage and shelf life conditions of 13 critical reagents provided by the vendor. Stability assessment for 7 of 13 (53%) critical reagents were tested up to less than one week before expiration date. Stability assessment for 2 of 13 (15.3%) were tested up to 3 months or less before expiration date. The remainder, 4 out of 13 (30.7%) critical reagents associated with long expiration date range (2 years) were assessed between 12 to 18 months before expiration date. Additionally, lots are qualified across the proposed timeline through assessment of a cohort of samples and controls. Study design is shown in **Figure 13**. All reagents are stored at -25°C to - 15°C.

Figure 13. Reagent Stability Study Schema. (a) A cohort of samples and controls is received at the Helix laboratory. (b) Samples and controls are processed through the HLP where critical reagents are used as part of routine testing. (c) The quality indicator performance metrics for each sample and control are grouped by the lot number of the critical reagent used. (d) Reagents lots are received, kept in inventory, and move to point of use on a first-in, first-out basis. (e) The quality indicator data collection takes place between the time a specific critical reagent lot commences use to the last date the lot is used.

ii. Specimen Stability

Two studies were conducted to support specimen stability: temperature and storage and freeze/thaw.

(1) Specimen Transport Stability

A study was conducted to evaluate the stability of specimens that are potentially subject to extreme transport conditions when shipped to the Helix laboratory. Specimens were shipped using the Oragene Dx OGD-610 collection device (K192920). A total of 5 saliva donors were consented for the study. Each donor collected and submitted 4 saliva samples (n = 20). For each donor, 2 samples were stored at room temperature, (20 +/- 5°C) and served as a control group, one sample was stored at high temperature (50 +/- 5°C) to simulate extreme summer transport conditions and one sample was stored at low temperature (-20 +/- 5°C) to simulate extreme winter transport conditions. Samples were stored under the testing conditions for a total of 7 days to simulate the worst-case scenario of a delivery delay during transport.

A reference for each saliva sample was generated by majority call comparison over all room temperature (RT) replicates of a sample in this study. The number of SNVs, Insertions, Deletions, reference bases and bases with no majority was calculated for each sample. Each replicate of a sample per testing condition (high and low temperatures), was compared to the majority call reference sequence generated from the appropriate ambient temperature sample. The NPA is calculated using only SNV; PPA and TPPV is calculated using SNV and indels combined. The acceptance criteria are: NPA $\geq$ 99.99%; PPA $\geq$ 99.5%; TPPV $\geq$ 99.5%; all with a 95% confidence interval lower bound of 99.0%. All conditions passed the Acceptance Criteria suggesting that performance is not negatively affected by transport of saliva in the OGD-610 collection kit at the extreme summer and

winter temperatures tested (**Table 32**).

Table 32. Comparison to Majority Call PPA (SNV and Indels)

Group Name	Mean PPA	Mean PPA Lower Bound	Mean NPA	Mean NPA Lower Bound	Mean TPPV	Mean TPPV Lower Bound
-20°C	0.99974	0.99963	1.00000	0.99999	0.99969	0.99956
50°C	0.99978	0.99968	1.00000	0.99999	0.99974	0.99963

(2) Freeze/Thaw Specimen Stability

A study was conducted to evaluate the impact of repeated freeze-thaws on performance for DNA extracted from saliva samples in the Helix Laboratory Platform. A total of 17 saliva samples with known variants, i.e., single nucleotide variants, small insertions or small deletions in medically related genes were tested after repeated DNA freeze-thaw cycles.

The test conditions included:

a) Baseline: DNA was extracted from saliva and subjected to NGS library preparation and sequencing prior to the first freezing event. The remaining DNA was stored at -20°C for 2 weeks.
b) Condition 1: DNA was subject to one (1) freeze-thaw cycle. The DNA was frozen for 2 weeks followed by thawing. A portion of the thawed DNA was subjected to NGS library preparation and sequencing. The remaining DNA was stored at -20°C for 2 weeks.
c) Condition 2: DNA from Condition 1 was subjected to an additional 3 freeze-thaw cycles (2 weeks/cycle). Upon the final thaw, the DNA was used for NGS library preparation and sequencing.

Performance was evaluated by pairwise comparison of known variants in baseline samples to samples exposed to freeze/thaw conditions. Results for this study show 100% concordance of the known variants, indicating that HLP performance is not negatively affected by repeat freeze/thaw under the conditions tested (data not shown).

3. Comparison Studies - Accuracy

A representative approach to accuracy was conducted to evaluate the accuracy of the HLP for the indicated variant types and challenging genomic contexts. Two studies were performed in support of the accuracy of the HLP: (1) an assessment of accuracy across the coding region and subsets using reference cell lines, and (2) accuracy of specific clinically relevant variants in DNA isolated from saliva specimens and cell lines.

a. Accuracy Study 1: Accuracy with Reference Cell Lines

The accuracy of the Helix Laboratory Platform was assessed through comparison of sequence data obtained with the HLP to the publicly available reference datasets for six (6)

well characterized cell lines. Five (NA12878, NA24385, NA24149, NA24143 and NA24631) cell lines were characterized by the Genome in a Bottle (GIAB) consortium and one cell line (NA12877) sequenced as part of the Platinum Genomes project (see **Table 7**).

Consistent with the HLP protocol, DNA input was (b)(4) and samples failing QC metrics at any step in the process were restarted or re-queued. Samples and process controls were required to pass the secondary analysis metrics for sample evaluation as listed in **Table 1**. The quality metrics and thresholds were evaluated as evidence of robustness in accordance with the requirements for each subset. The PPA, NPA, and TPPV along with the 95% lower limit of the confidence interval were calculated as described in **Table 6**. The pre-specified acceptance criteria for these studies for overall performance are shown in **Table 33**.

Table 33. Acceptance Criteria for the Accuracy Studies.

Variant Type	PPA Threshold	PPA Lower Bound Threshold*	NPA** Threshold	NPA Lower Bound Threshold*	TPPV Threshold	TPPV Lower Bound Threshold*
SNV	0.995	0.990	0.999	0.990	0.995	0.990
Deletion	0.990	0.980	n/a	n/a	0.990	0.980
Insertion	0.990	0.980	n/a	n/a	0.990	0.980

*95% Lower Bound Confidence Interval (CI) calculated using a binomial model with a 95% CI estimated with Wilson's method
** The NPA cannot be determined for insertions and deletions.

Accuracy was determined for each indicated variant type and challenging genomic contexts including the range of sizes for insertions and deletions, zygosity, GC content, pseudogenes, and accuracy of detection near short tandem repeats (i.e., STRs; homopolymer stretches). Results were summed over all 6 samples in the study and performance was calculated for the coding regions and the two subsets (mendiolome and priority). These three regions have different levels of coverage. The data in **Table 34** presents the total expected number of variant calls, the number of no calls, the percentage of variants with no calls, the number of true positives (TP), the number of false positives (FP), the number of false negatives (FN), the positive percent agreement (PPA) and technical positive percent agreement (TPPV) for all samples combined for each region. Calculation of PPA, NPA, and TPPV values were calculated as described in **Table 6**. The data demonstrates that all regions met the pre-defined acceptance criteria with the exception of insertions with size (b)(4) as high no call rate (>25%) and PPA <99% were observed. Therefore, insertions (and deletions) with sizes greater than 6bp are to be independently validated.

Data is also demonstrated per specimen stratified by specimen (**Table 35**); and further stratified by zygosity and indel size (**Table 36**); stratified specifically for indel size across all samples (**Table 37**); and stratified based on GC content (**Table 38**). Results for indels in regions where the GC content exceeds 65% were not included in the data and therefore are not reported. Data not meeting the acceptance criteria for the overall study is shown in **bold**. Stratified data were not required to meet the pre-specified acceptance criteria due to the

smaller number of variants for calculation, however data not meeting the pre-specified acceptance criteria was demonstrated to largely be limited to insertions (b)(4)

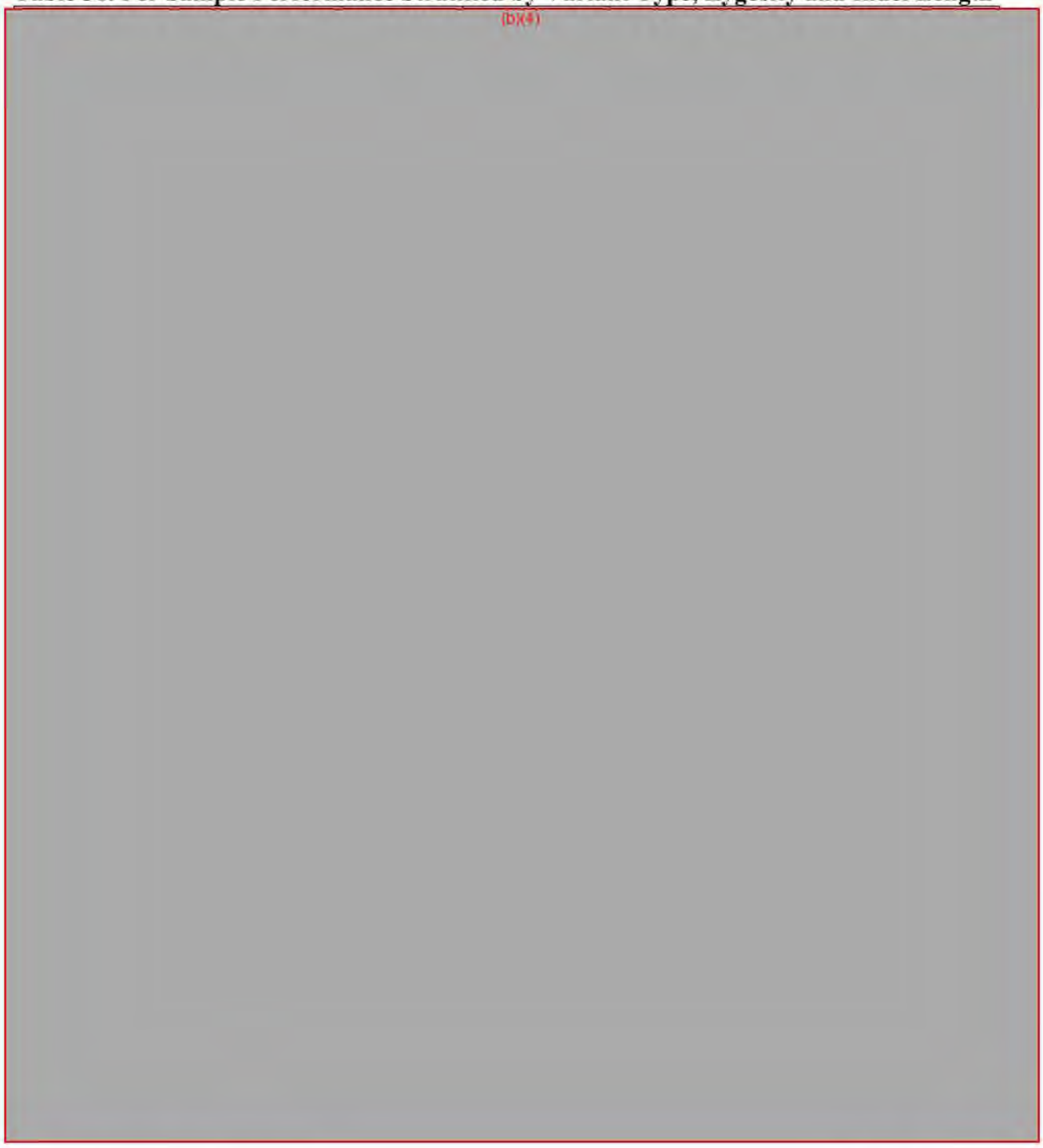
Table 34. Accuracy for the Coding Region and Distinct Subsets

Subset	Variant Type	Number of Expected Variants	Number of No-calls	% of Variants with No-calls	TP	FP	FN	PPA	PPA Lower Bound	TPPV	TPPV Lower Bound
Coding (Overall)	SNV	(b)(4)		2.00%	111516	83	118	0.999	0.9987	0.9993	0.9991
	Insertion			10.12%	815	8	2	0.998	0.9911	0.9903	0.9809
	Deletion			6.99%	929	6	3	0.999	0.9906	0.9936	0.9861
Mendeliome	SNV			1.26%	30726	26	23	0.999	0.9989	0.9992	0.9988
	Insertion			5.74%	229	6	1	0.996	0.9758	**0.9745**	0.9454
	Deletion			5.56%	204	1	0	1.000	0.9815	0.9951	0.9729
Priority	SNV			0.61%	5022	2	5	0.999	0.9977	0.9996	0.9985
	Insertion			0.00%	42	0	0	1.000	0.9162	1.0000	0.9162
	Deletion			4.84%	59	0	0	1.000	0.9389	1.0000	0.9389
ClinVar	SNV			0.00%	87	0	0	1.000	0.9577	1.0000	0.9577
	Deletion			0.00%	2	0	0	1.000	0.3424	1.0000	0.3424

Table 35. Accuracy Performance Per Sample

(b)(4)

Table 36. Per Sample Performance Stratified by Variant Type, Zygosity and Indel Length



Stratification by %GC: Results were summed over all samples and performance calculated for each %GC bin. The data below groups each variant type by %GC bin, >15-40% GC, >40-65% GC, >65-90% GC.

Table 37. Accuracy for Deletions and Insertions based on Size for All 6 Reference Samples Combined

(b)(4)

Variant Detection Near Short Tandem Repeats (STR): Stretches of homopolymer regions can lead to unreliable detection of nearby variants. To determine the HLP accuracy for detection of variants near short stretches of tandem repeats, detection of variants STRs was compared to the detection of variants not near STRs. The results are shown in **Table 38a and 38b** below. The results demonstrated no difference in performance of variant detection in the absence of STRs or near short STRs. All criteria met the prespecified acceptance criteria with the exception of insertions which was expected for the results overall (**Table 38a**). (Note: The performance of the HLP for accuracy and precision of reporting short tandem repeats was not evaluated and the HLP is not indicated for detection of STRs).

Table 38a. Variant Detection Near Short Tandem Repeats.

(b)(4)

Table 38b. Variant Detection Near Short Tandem Repeats Stratified by Indel Size.

(b)(4)

Genotype Quality (GQ) score: GQ score is one of several criteria used to control for variant calling accuracy. (b)(4) The number of false positives as a function of GQ score was evaluated for variants from Accuracy study 1. The data demonstrated an (b)(4) (b)(4) **Table 39** demonstrates that the majority of variants have high GQ scores.

Table 39. Genotype Quality Scores for SNVs, Insertions and Deletions

		GQ Score							
Variant Type	**Classification**	**[20, 30)**	**[30, 40)**	**[40, 50)**	**[50, 60)**	**[60, 70)**	**[70, 80)**	**[80, 90)**	**[90, 99]**
SNV	TP	(b)(4)							
	FP								
INSERTION	TP								
	FP								
DELETION	TP								
	FP								

b. Accuracy Study 2: Clinical Specimens

A study was conducted to evaluate the accuracy of the HLP with DNA isolated from saliva specimens. Specimens were selected for the presence of clinically relevant variants. A specimen selection protocol was used to select specimens to minimize bias as follows: Each of the samples were sequenced in the Helix lab between July 2017 and April 2018. A list of clinically relevant variants from all variants in ClinVar were identified and filtered for high confidence interpretation. The list was further refined to remove variants outside of Helix reportable range (86 were removed) and from helix blacklist (7 variants removed). The final ClinVar variant list contained 29,492 variants which were screened against the Helix database to determine availability of the variant within the Helix specimen collection. A total of 6, 427 unique variants were identified. The variant list was then randomly chosen from this list. Samples with inadequate saliva volume were removed from analysis. Two (2) samples failed to meet secondary sequencing metrics QC for evaluability based on callability and coverage metrics. Three (3) samples failed to meet secondary sequencing metrics QC based on freemix metric. One (1) sample failed to yield sufficient DNA at QC after library prep (failed at Quant QC - Quant 2). A total of 1002 clinical samples met secondary sequencing metrics QC for evaluability and were used in the final data analyses below.

A validated Sanger sequencing method was used to confirm the accuracy of 1,061 clinical variants in 1,002 clinical samples and 90 unique variants in 96 cell lines. The comparator method was assigned to sequence a region of ~125bp length, which included the variant of interest. The additional sequenced DNA flanking the variant of interest was used to assess the accuracy of the HLP for reference (wild-type sequencing). As a result of this additional sequencing, a total of 103,339 reference calls were extracted as truth for reference sequence at clinical variants sites and sites flanking the variants of interest. The variants are composed of single nucleotide variants, small insertions or small deletions (less than or equal to 20 bp) in clinically relevant genes.

Additionally, 125 clinical variants and reference calls in 96 unique cell line samples were also evaluated. Cell lines were compared to the truth genotypes taken from published variant information. Some of the samples have more than one variant and some of the variants were tested multiple times in different samples.

The study was done according to protocol and results were required to meet quality metric reporting thresholds as described previously.

Because the accuracy is contingent on the HLP results, positive predictive value (PPV) and negative predictive value (NPV) are calculated. Truth genotypes for the variants in the cell lines were taken from publicly available information. The acceptance criterion for PPA was established at 99.5% with a lower bound of 99.0% using a binomial model with a 95% CI estimated with Wilson's method. Acceptance criterion for NPA was 99%. Adjusted PPA (aPPA). and adjusted NPA (aNPA) were also calculated to adjust for potential bias in the clinical sample selection. Definitions and methods of calculation for aPPA and aNPA are shown in **Table 6**. The sampling prevalence used for the calculation of aPPA and aNPA was 0.266096.

Data were calculated for all variant and reference calls combined (All) and by variants and reference calls located within three GC bins. For a total of 104,398 possible variant calls, there were 2 incorrect calls (**Table 40**). Two heterozygous SNVs (at chr1:152313368 and at chr3:120641660) were called incorrectly as reference and classified as false negative (FN). For variant at chr1:152313368, Sanger sequencing determined this variant to be Het_Alt (A/G) and HLP called Hom_Ref (A/A). For variant at chr3:120641660, Sanger sequencing determined this variant to be Het_Alt (C/T) and HLP called Hom_Ref (C/C). The data demonstrates high degree of accuracy for the variants evaluated.

Table 40. Concordance of NGS and Sanger Results for each Variant Category

Number of Unique Clinical Samples	Variant Type	Number of Expected Calls[1]	Number of No-calls[2]	Number of Correct Calls[3]	Number of Incorrect Calls[4]	Percent Agreement including No calls	Percent Agreement[5] Excluding no calls
729	SNV (het)	(b)(4)		759	2	99.086	99.737%
81	SNV (hom)			86	0	100.000	100.000%
164	Deletion (All)			163	0	99.390	100.000%
107	Deletion (1-2bp)			106	0	100.000%	100.000%
28	Deletion (3-5bp)			28	0	100.000%	100.000%
29	Deletion (6-20bp)			29	0	100.000%	100.000%
45	Insertion (All)			45	0	100.000	100.000%
33	Insertion (1-2bp)			33	0	100.000%	100.000%
2	Insertion (3-5bp)			2	0	100.000%	100.000%
10	Insertion (6-20bp)			10	0	100.000%	100.000%
641	Reference (wild-type)			102,813	0	99.492	100.000%
1,002	All			103,866	2	99.490	99.998%

[1] Number of expected calls confirmed by Sanger sequencing (variant sites and flanking regions).
[2] Subset of expected calls which received a no-call.
[3] Number of calls agreeing with Sanger sequencing.
[4] Number of calls disagreeing with Sanger sequencing.
[5] Percent Agreement = Number of Correct Calls / (Number of Expected Calls - Number of No-calls)

Additional variants of interested present in cell lines were also evaluated. Data was also analyzed for cell lines. There were 2 homozygous calls that HLP called heterozygous (**Table 41**).

Table 41. Accuracy Study 2: Results for Cell Line Samples Reported by Variant Type.

Number of Unique Cell Line Samples	Variant Type	Number of Expected Calls[1]	Number of No-calls[2]	Number of Correct Calls[3]	Number of Incorrect Calls[4]	Percent Agreement[5]
52	SNV (het)	61	0	61	0	100.000%

18	SNV (hom)	18	0	17	1[6]	94.444%
27	Deletion	29	0	29	0	100.000%
5	Insertion	5	0	5	0	100.000%
8	Reference	8	0	8	0	100.000%
96	All	125	0	124	1[6]	99.200%

[1] Number of known calls based on published data for each cell line.
[2] Subset of expected calls which received a no-call.
[3] Number of calls agreeing with published data for each cell line.
[4] Number of calls disagreeing with published data for each cell line.
[5] Percent Agreement = Number of Correct Calls / (Number of Expected Calls - Number of No-calls)
[6] Homozygous SNV was called as heterozygous SNV.

Overall accuracy of the Saliva and Cell lines is summarized in **Table 42**. The no call rate for all data sets is shown for each region in **Table 43**.

Table 42. Accuracy Study 2: Saliva (Clinical) and Cell Line Data.

Sample	Number of Samples	TP[1]	FP[2]	TN[3]	FN[4]	NPV	TPPV	aPPA	aNPA
Clinical samples	1,002	1,053	0	102,813	2	99.998%	100.000%	99.995%	100.000%
Cell lines	96	116	0	8	1	88.889%	100.000)	99.145 (not adjusted)	100.000% (not adjusted)

[1] Number of Sanger sequencing confirmed variants that were called correctly.
[2] Number of Sanger sequencing confirmed reference calls that were incorrectly called as variants.
[3] Number of Sanger sequencing confirmed reference calls that were called correctly.
[4] Number of Sanger sequencing confirmed variants that were incorrectly called as reference or with an incorrect genotype.

No-Call Rates

The mean no-call rate of the HLP was averaged across all samples in the accuracy study and shown for the reportable region and subsets (**Table 43**). Additionally, the no call rate for variants in ClinVar were also assessed (accuracy for ClinVar subset not shown). The no call rate for the coding region considering the flanking regions in Accuracy study 2 was (b)(4) (data not shown).

Table 43. Callability in the Accuracy datasets

Reportable Window	Mean Callability	Mean No Call Rate (1- Callability)
Exome (all coding region targets)	97.0%	3.0%
Mendeliome	98.5%	1.5%
Priority	99.5%	0.5%
ClinVar Subset	99.5%	0.5%

Accuracy- Assessment of Metrics

In order to support the representative approach, data using reference samples with known sequence data, were used to demonstrate that the metrics used by the HLP establish high confidence calls (See Section L). Data demonstrating the metrics in the accuracy study were assessed to provide evidence that conclusions that the overall sequencing data accurate is supported. Mean coverage, as well as the callability and percentage of variants with (b)(4) are summarized for all samples in the accuracy study with 96 cell lines and 1002 clinical samples in **Table 44.**

Table 44. Quality Metrics for Cell Lines and Clinical Samples in the Accuracy Study

	Cell Lines	Clinical Samples
Mean Coding Callability	(b)(4)	
Range		
Mean Mendeliome Callability		
Range		
Mean Priority Callability		
Range		
Mean Coding Percent >=20x		
Range		
Mean Mendeliome Percent >=20x		
Range		
Mean Priority Percent >=20x		
Range		
Mean Coding Average Coverage		
Range		
Mean Mendiolome Average Coverage		
Range		
Mean Priority Average Coverage		
Range		

4. Clinical Performance

Not applicable. The HLP is a targeted whole exome sequencing test system intended to be used with genetic assays as part of a test system where such genetic assays are validated for specific clinical use. The HLP is being authorized in conjunction with the Helix Genetic Health Risk App test; k192063.

M. Instrument Name

Illumina HiSeq X instrument (qualified by Helix): The HiSeq sequencing instrument is a high throughput DNA analyzer that measures fluorescence signals of labeled nucleotides through the use of instrument specific reagents in conjunction with Helix proprietary reagents.

Illumina cBot system (qualified by Helix): The cBot System (Illumina Inc.) is an automated system that creates clonal clusters from single molecule DNA templates, preparing them for sequencing by synthesis (SBS) on the HiSeq instrument. The cBot isothermally amplifies cDNA fragments that have been captured by complementary adapter oligonucleotides covalently bound to the surface of Illumina

flow cells (called the Cluster Generation Process)

N. System Descriptions

1. Modes of Operation:

   Does the applicant's device contain the ability to transmit data to a computer, web server, or mobile device?
   Yes _____X___ or No _____

   Does the applicant's device transmit data to a computer, web server, or mobile device using wireless transmission?
   Yes _____X___ or No _____

2. Software:

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

   Yes ____X____ or No _____

3. Instrument software:

   - HiSeq Control Software: The HiSeq Control Software (Illumina Inc.) controls the flow cell stage, fluidics system, and flow cell temperatures. It also captures images of clusters, generating image analysis, base calling, and base call quality automatically.
   - Real Time Analysis Software: Primary analysis is performed by the Real Time Analysis (RTA) software (Illumina Inc.) and consists of base calling of each cluster at each cycle. In addition to base calling, 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 as $-10 \log_{10} P$, where P is the probability that base call is incorrect.
   - Conversion Software: The bcl2fastq conversion software (Illumina Inc.) is used to process BCL (base call log) files generated by the HiSeq instrument into de-multiplexed FASTQ files. De-multiplexing is the process of using the index sequences to assign clusters to the sample from which they originated. FASTQ is a standard text-based file format that will store the nucleotide sequences and base quality scores for each read sequenced from a sample.

   Yes _X_ or No _____

4. Specimen Identification:

   Specimens are accessioned.

5. Specimen Sampling and Handling:

   Automated

6. Calibration:

   Not applicable

7. Quality Control:

   External and internal controls

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

Analysis of Reads Coverage

Coverage varies according to genomic regions: Primary clinical regions (b)(4), Secondary clinical genes (b)(4) other coding regions (b)(4). The mean reads coverage for each study was analyzed as a function of callability to determine that callability was appropriately assessed for each reportable range. Although overall mean reads coverage varied among studies, (b)(4) (b)(4) in every study, meeting reads coverage requirements for each reportable range (**Figure 14a**).

Average Mean Callability

Due to variation of mean read coverage in different region subsets of the reportable range, callability QC thresholds for Priority, Mendeliome, and Coding regions were set at (b)(4) (b)(4) respectively. Average mean callability in each sample from all analytical studies were assessed based on region subsets. As expected, all samples in all studies met these thresholds (**Figure 14b**).

Figure 14. Average mean reads coverage and callability in each analytical study by sample types. Data in each figure was aggregated based on sample types (cell lines vs. clinical samples reportable ranges, Priority (green), Mendeliome (blue), and Coding regions (orange). (a) Average mean reads coverages in each analytical study based on genomic regions. (b) Average mean callability in each analytical study based on genomic regions.

Coverage threshold vs Callability per Reportable Ranges

As a minimum coverage of (b)(4) was used to define callability of region subsets of the reportable range, increasing the minimum coverage requirement would impact the callability of region subsets. Using different reads coverage thresholds, (b)(4) callability of each region subset, Priority, Mendeliome, and Coding, were assessed. Increasing the read coverage requirement had the most impact on coding regions as more than (b)(4) of regions are not callable (no call) in most of samples, when the threshold was set at (b)(4) (**Figure 15**).

Figure 15. Coverage threshold vs Callability per Region Subset of Reportable Range. Data were aggregated from studies, Accuracy-1, Precision, and Between-Lot Reproducibility. Y-axis, percent of regions met coverage requirements.

Non-coding region: The HLP reagents include coverage of both coding and non-coding regions. The non-coding regions are not reported. For information purposes only, the non-coding regions were covered by the HLP exhibited (b)(4) no call rate.

Change Protocols: Change protocols specify the verification and validation activities that will be performed for anticipated bioinformatic software modifications to reevaluate performance claims or performance specifications that were reviewed and determined acceptable by FDA. The protocol included a list of specific changes, the specimens (type, number and variant representation), analytical methods, statistical analysis methods, and acceptance criteria, for determination that the modifications met the performance specifications of the HLP. Assessment of the risk and the impact on results was also provided and included the processes for communicating to developers of downstream clinical genetic tests, the impact of the bioinformatics software change on the whole exome sequencing constituent system genetic data output.

P. Proposed Labeling:

The labeling supports the decision to grant the De Novo request for this device.

Q. Identified Risks to Health and Identified Mitigations

Identified Risks to Health	Mitigation Measures

Inaccurate test results and failure to provide results	Certain design verification and validation, including certain analytical studies and clinical studies. Certain labeling information, including certain performance information and device limitations.
Incorrect application or interpretation of results	Certain design verification and validation, including certain clinical studies. Certain labeling information, including certain performance information and device limitations.
User error and improper use of the device	Certain design verification and validation, including certain analytical studies and clinical studies. Certain labeling information, including certain performance information and device limitations.

R. Benefit/Risk Analysis:

Patient Perspectives

This submission did not include specific information on patient perspectives for this device.

Summary of the Assessment of Benefit
For the Proposed Indications for Use

The Helix Laboratory Platform is a device that sequences the whole exome in an individual's genomic DNA obtained from saliva. The whole exome data includes SNVs, insertions, and deletions for genetic tests validated for use on this device. The benefit of this device is large in magnitude due to the potential to enable individuals to obtain genetic information through over the counter or prescription legally marketed genetic tests that use the HLP. Such results may positively influence lifestyle decisions or optimize patient management when used in conjunction with medical consultation.

Summary of the Assessment of Risk
For the Proposed Indications for Use

Probable risks associated with the use of this device are mainly due to 1) false positive, false negatives, or failure to provide a result; 2) the incorrect application or interpretation of the data by genetic test developers that use the HLP; and 3) user error and improper use of the device

The risks of false positives are that genetic tests that use the HLP will report false results to patients and/or clinicians, which will then potentially lead to inappropriate lifestyle changes, psychological harm, and unnecessary therapies and/or medications, with the attendant risk of side effects.

The risks of false negatives are that the patient may fail to initiate appropriate medical consultation, lifestyle changes, therapeutic options, or targeted surveillance. Also, the risks of a failure to provide results are that the patient may experience a delay in getting results, and may fail to initiate appropriate medical consultation, as necessary.

There is the also risk of incorrect application or interpretation of data by genetic test developers that use the HLP, as well as user error or improper use of the device; these will result in clinically inaccurate results (false positives or false negatives) presented to the patient and/or clinicians, with the attendant risks described above.

Summary of the Assessment of Benefit-Risk
For the Proposed Indications for Use

The risks associated with providing patients incorrect genetic results may lead to delays in patient management or incorrect patient management, or failure to make lifestyle choices that may benefit an individual's well-being. For this reason, general controls are insufficient to mitigate the risks of the device. Comprehensive analytical validation studies, including accuracy studies and precision studies with reference standards and clinical specimens, were performed to support the probable benefit of this device and mitigate such risks. Additional risk mitigation factors taken into consideration were the delineation of the reportable range of the device and the analytical range of the device, which allowed for implementation of quality metric requirements for reporting, in order to mitigate the risk of incorrect results. Device design verification and validation, including appropriate descriptions of test limitations in the device labeling, support the appropriate and accurate use of the device. Finally, due to the comprehensive nature of the information, applicable protocols to support continued changes to the device were submitted and reviewed. Such change protocols enable the device to remain current with the practice of medicine and the optimization of this technology.

The probable clinical benefits outweigh the probable risks when appropriate mitigation of the risks is provided for through implementation and adherence to the special controls. The combination of the general controls and established special controls support the assertion that the probable benefits outweigh the probable risks.

S. Conclusion

The De Novo request is granted, and the device is classified under the following and subject to the special controls identified in the letter granting the De Novo request:

Product Code: QNC
Device Type: Whole exome sequencing constituent system
Class: II (special controls)
Regulation: 21 CFR 866.6000