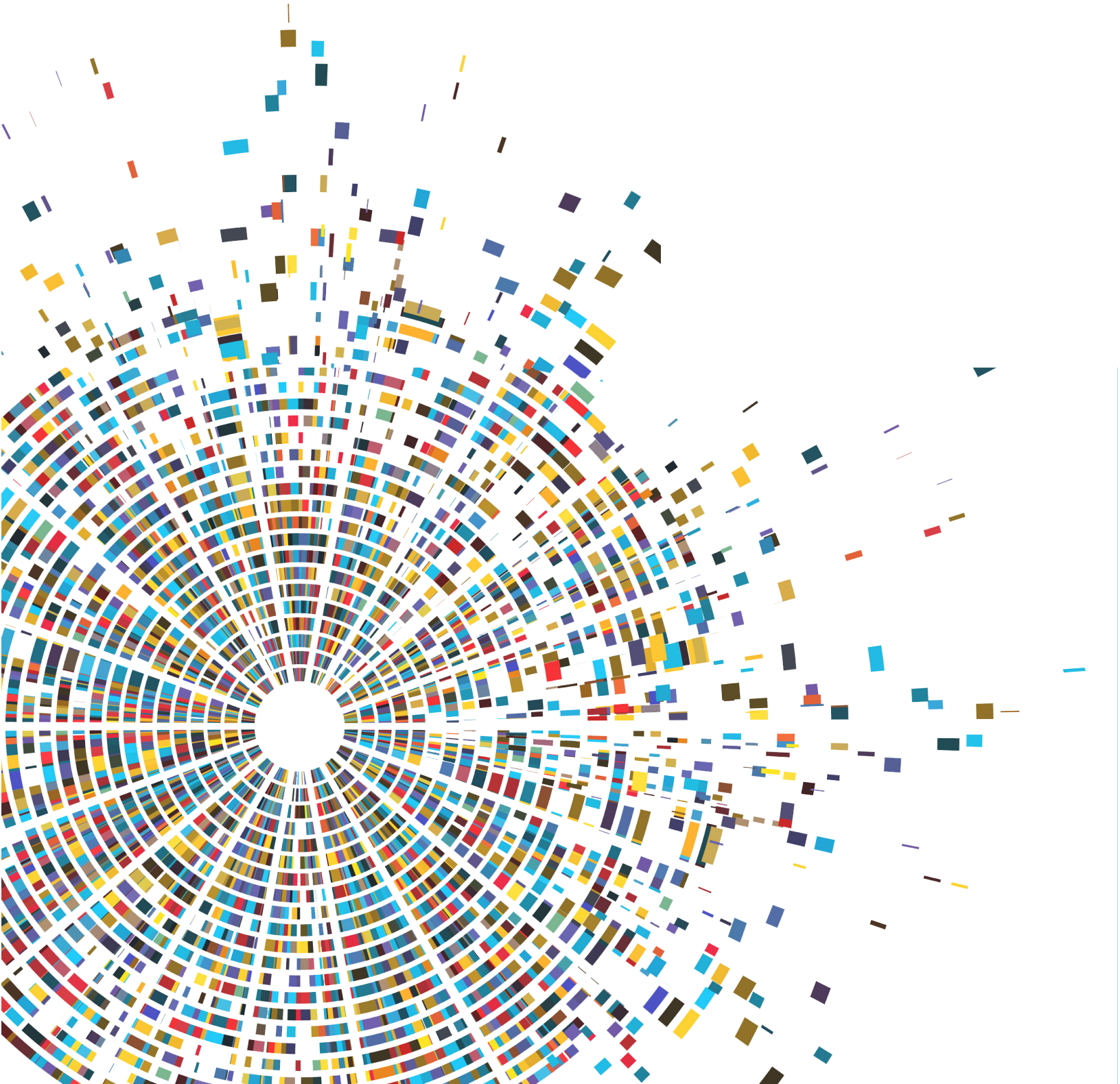
variantyx

# Genomic Unity® Test

## A Whole Genome Clinical Validation Study



## Overview

Variantyx applies whole genome sequencing (WGS) in tandem with its proprietary Genomic Intelligence® analysis pipeline and software for the identification of clinically relevant variants. Nuclear WGS identifies single nucleotide variants (SNVs), small insertions and deletions (indels), structural variants (SVs) and short tandem repeat expansions (STRs), while mitochondrial genome sequencing identifies SNVs and heteroplasmy. This white paper describes the validation performed, providing the sensitivity and specificity for each variant type.

## Background

Historically the detection of SNVs and indels have been performed using multiple sequencing platforms such as Sanger sequencing and Next Generation Sequencing (NGS), including whole exome and whole genome sequencing (WES/WGS) methodologies. However, in the broad scope of all genomic testing, this represents only a subset of methodologies used to identify pathogenic variants in the human genome. Other methodologies include: Southern blots to detect large STRs and large deletions, PCR and capillary electrophoresis to detect shorter STRs, qPCR and MLPA to detect deletions and duplications, and arrayCGH/microarray, FISH and karyotypes to detect gross chromosomal deletions.

A major concern in using multiple methods to detect different types of genetic variants is that one might miss a connection in a patient with a combination of changes that were identified by different method platforms at different times and often at different laboratories. For example, detection of a heterozygous deletion in a recessive gene in one lab and a heterozygous pathogenic SNV in another lab might be reported as "carrier" by each lab. To assimilate these independent lines of evidence into a positive result requires a comprehensive understanding of molecular biology and the complexities of genetic testing. This connection might be missed, even by a technically savvy ordering clinician.

Moreover, since each test has its own turn-around-time and is often ordered sequentially as a reflex off a prior result for financial reasons, disjointed tests offered at different labs place a burden on the family for multiple sample collections as well as extending the time to diagnosis, which can delay clinical management implementation. Therefore, consolidating previous multi-platform testing into one unified report is highly desirable in the clinical genomics diagnostic space to improve diagnostic yields, shorten patient turnaround times and provide the earliest possible window for clinical management and intervention.

With its PCR free DNA preparation and consistent read depth, WGS provides the opportunity for such a consolidated test. At Variantyx, we have developed the Genomic Unity® test to provide comprehensive analysis of the variant types for which validation is described in this document.

## Methodology

30X whole genome sequencing was performed on the Illumina platform using the Illumina TruSeq PCR Free DNA Preparation Kit. Data analysis was performed with version 2.7.0.0 of Variantyx's Genomic Intelligence® analysis pipeline and software.

Whenever possible, validation was performed using NIST Genome in a Bottle (GIAB) reference datasets. On average, about 1.4% of the genome does not have coverage sufficient for reliable variant calling (<8X), however only about 0.4% of clinically relevant variants (based on HGMD Professional and ClinVar annotations) are not adequately covered. See the Appendix for coverage and sequence depth metrics for GIAB samples. When acceptable "gold standard" true positive data sets did not exist, as in the case of structural variants and short tandem repeats, sourced controls were used as described.

## Single nucleotide variants (SNVs)

Analysis of SNVs is performed using GATK best practice guidelines. Validation was performed on three GIAB samples: NA24149, NA24143 and NA12878. The reported validation statistics are based on the mean of true/false positive/negative values calculated for the three samples.

| | |
|---|---|
| True positive (TP) | 2,723,422 |
| False positive (FP) | 4,709 |
| True negative (TN) | 2,188,155,955 |
| False negative (FN) | 3,146 |
| Sensitivity | 0.99885 |
| Specificity | 1.00000 |
| Positive predictive value | 0.99827 |
| Accuracy | 1.00000 |

The following tables provide validation statistics for SNVs divided into heterozygous and homozygous (reference and alternate alleles combined) calls.

### Heterozygous SNVs

| | |
|---|---|
| True positive (TP) | 1,656,415 |
| False positive (FP) | 4,648 |
| True negative (TN) | 2,189,182,961 |
| False negative (FN) | 2,573 |
| Sensitivity | 0.99846 |
| Specificity | 1.00000 |
| Positive predictive value | 0.99720 |
| Accuracy | 1.00000 |

**Homozygous SNVs**

| | |
|---|---|
| True positive (TP) | 1,067,006 |
| False positive (FP) | 61 |
| True negative (TN) | 2,189,772,370 |
| False negative (FN) | 573 |
| Sensitivity | 0.99946 |
| Specificity | 1.00000 |
| Positive predictive value | 0.99994 |
| Accuracy | 1.00000 |

Sensitivity (>99.8%), specificity (100%), positive predictive value (>99.7%) and accuracy (100%) of SNVs called by Genomic Unity® test are at or above those typically reported for whole exome sequencing by multiple clinical laboratories.

## Small insertions and deletions (indels)

Analysis of indels is performed using GATK best practice guidelines. Validation was performed on three GIAB samples: NA24149, NA24143 and NA12878. The reported validation statistics are grouped by indel size and based on the mean of true/false positive values calculated for the three samples. Indel sizes of 1-4 bp, 5-15 bp and ⩾16 bp are considered. However, there is not a hard size limit as the largest indel included was 172 bp.

1-4 bp

| | |
|---|---|
| True positive (TP) | 389,023 |
| False positive (FP) | 4,917 |
| Sensitivity | 0.98552 |

5-15 bp

| | |
|---|---|
| True positive (TP) | 31,393 |
| False positive (FP) | 546 |
| Sensitivity | 0.98384 |

⩾16 bp

| | |
|---|---|
| True positive (TP) | 6,187 |
| False positive (FP) | 128 |
| Sensitivity | 0.97176 |

The following tables provide validation statistics for indels divided into heterozygous and homozygous (reference and alternate alleles combined) calls.

**Heterozygous indels**

1-4 bp

| | |
|---|---|
| True positive (TP) | 242,396 |
| False positive (FP) | 2,425 |
| Sensitivity | 0.98685 |

5-15 bp

| | |
|---|---|
| True positive (TP) | 20,677 |
| False positive (FP) | 245 |
| Sensitivity | 0.98074 |

⩾16 bp

| | |
|---|---|
| True positive (TP) | 4,200 |
| False positive (FP) | 78 |
| Sensitivity | 0.96772 |

**Homozygous indels**

1-4 bp

| | |
|---|---|
| True positive (TP) | 146,627 |
| False positive (FP) | 2,492 |
| Sensitivity | 0.98332 |

5-15 bp

| | |
|---|---|
| True positive (TP) | 10, 716 |
| False positive (FP) | 301 |
| Sensitivity | 0.98985 |

⩾16 bp

| | |
|---|---|
| True positive (TP) | 1,987 |
| False positive (FP) | 50 |
| Sensitivity | 0.98042 |

Sensitivity of the Genomic Unity® test in calling of indels of any size is at or above 96%.

## Mitochondrial heteroplasmy

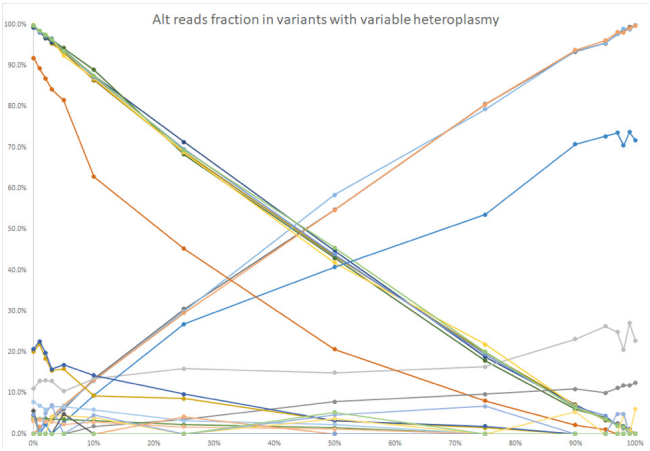
30X whole genome sequencing produces approximately 2,000X coverage of the mitochondrial genome. Mitochondrial variants were detected in 15 samples obtained by mixing DNA of two known samples in different ratios and sequencing the mixes. Greater than 1,000X coverage was observed at every nucleotide in all tested samples.

The following table shows how the percentage of alternate reads detected changes as the composition of the sequencing mixes progress from containing 0% to 100% of the second known sample. For each case, the listed value is the percentage of alternate reads detected out of the total number of reads, which reflects the heteroplasmy of the called variants. In some cases, both samples are homozygous reference, consistently producing a score of 0% (for example, chrM:10925T>G). In other cases, both samples are homozygous alternate, consistently producing a score of 100% (for example, chrM:750A>G). It is the cases where one or both samples are heterozygous, and in some cases at varying degrees of heteroplasmy, that provide the greatest insight into the detectable level of heteroplasmy (for example, chrM:6776T>C). Note that in cases where the variant is not called due to the fraction of alternate reads falling below the calling threshold, the number of alternate reads is defaulted to 0 in turn producing a value of 0% regardless of the actual number of supporting reads.

| Variant name | 0% | 1% | 2% | 3% | 5% | 10% | 25% | 50% | 75% | 90% | 95% | 97% | 98% | 99% | 100% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chrM:750A>G | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| chrM:15326A>G | 99.9% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 99.9% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| chrM:16519T>C | 100.0% | 99.9% | 100.0% | 100.0% | 99.9% | 99.9% | 100.0% | 99.9% | 100.0% | 100.0% | 99.9% | 99.9% | 100.0% | 100.0% | 100.0% |
| chrM:310T>TC | 100.0% | 99.9% | 100.0% | 99.9% | 99.5% | 99.5% | 99.5% | 100.0% | 99.6% | 99.7% | 100.0% | 99.8% | 99.7% | 100.0% | 99.9% |
| chrM:1438A>G | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| chrM:8860A>G | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 99.9% | 100.0% | 100.0% | 100.0% | 99.9% | 100.0% |
| chrM:263A>G | 99.8% | 100.0% | 100.0% | 99.8% | 100.0% | 100.0% | 100.0% | 100.0% | 99.9% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| chrM:4769A>G | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 99.9% | 100.0% | 100.0% | 100.0% |
| chrM:10535T>C | 0.0% | 1.5% | 2.8% | 3.4% | 6.1% | 13.5% | 30.5% | 54.8% | 80.6% | 93.4% | 95.5% | 97.7% | 98.4% | 99.5% | 99.7% |
| chrM:6293T>A | 99.5% | 98.4% | 97.3% | 95.4% | 93.3% | 86.4% | 68.9% | 43.6% | 19.8% | 7.1% | 3.3% | 2.3% | 1.4% | 0.8% | 0.0% |
| chrM:200A>G | 99.3% | 98.3% | 96.7% | 95.8% | 93.7% | 87.5% | 71.3% | 44.8% | 18.8% | 7.0% | 3.4% | 2.4% | 1.9% | 1.0% | 0.0% |
| chrM:8602T>C | 99.9% | 98.3% | 97.0% | 96.4% | 94.3% | 89.0% | 68.3% | 43.0% | 17.9% | 6.0% | 4.0% | 2.3% | 1.6% | 0.9% | 0.0% |
| chrM:15314G>A | 0.0% | 1.0% | 2.9% | 4.1% | 6.4% | 12.9% | 30.0% | 58.4% | 79.3% | 93.6% | 95.5% | 97.8% | 99.0% | 98.8% | 99.9% |
| chrM:6776T>C | 0.0% | 1.7% | 2.7% | 4.1% | 6.9% | 13.1% | 29.5% | 54.7% | 80.5% | 93.8% | 96.1% | 98.2% | 98.0% | 99.2% | 99.8% |
| chrM:302A>AC | 11.1% | 12.9% | 13.1% | 12.9% | 10.4% | 13.5% | 15.9% | 14.9% | 16.4% | 23.2% | 26.3% | 24.9% | 20.5% | 27.0% | 22.7% |
| chrM:14212T>C | 99.9% | 98.2% | 97.4% | 96.6% | 92.4% | 86.9% | 68.8% | 41.8% | 21.8% | 6.8% | 3.6% | 2.3% | 1.3% | 0.8% | 0.0% |
| chrM:8937T>C | 99.8% | 98.1% | 97.3% | 96.6% | 93.6% | 87.2% | 69.6% | 43.9% | 19.4% | 6.6% | 4.4% | 2.1% | 1.4% | 0.0% | 0.0% |
| chrM:3010G>A | 99.8% | 98.4% | 97.6% | 96.4% | 93.6% | 87.5% | 69.3% | 45.5% | 20.1% | 6.4% | 3.6% | 2.0% | 1.7% | 0.0% | 0.0% |

| Variant name (continued) | 0% | 1% | 2% | 3% | 5% | 10% | 25% | 50% | 75% | 90% | 95% | 97% | 98% | 99% | 100% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chrM:302A>ACC | 0.0% | 0.0% | 2.3% | 0.0% | 2.8% | 9.4% | 26.8% | 40.7% | 53.6% | 70.8% | 72.8% | 73.6% | 70.5% | 73.8% | 71.7% |
| chrM:16189T>C | 91.8% | 89.3% | 86.8% | 84.2% | 81.5% | 62.8% | 45.3% | 20.7% | 8.1% | 2.2% | 1.1% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:9053G>A | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 1.8% | 3.5% | 7.8% | 9.7% | 10.9% | 10.0% | 11.1% | 11.8% | 11.9% | 12.4% |
| chrM:16183A>ACC | 20.1% | 21.9% | 18.3% | 15.5% | 15.8% | 9.3% | 8.6% | 3.1% | 1.5% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:16183A>AC | 20.7% | 22.6% | 19.8% | 15.8% | 16.8% | 14.3% | 9.7% | 3.2% | 1.9% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:10785T>C | 3.6% | 3.6% | 3.8% | 3.5% | 3.5% | 3.2% | 2.3% | 1.4% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:16188CT>C | 7.8% | 6.9% | 6.0% | 6.7% | 6.6% | 5.9% | 3.2% | 2.3% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:7162G>A | 3.1% | 3.4% | 3.0% | 3.0% | 2.3% | 2.9% | 1.7% | 1.2% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:1552G>A | 1.7% | 1.8% | 1.8% | 1.7% | 1.8% | 1.6% | 1.6% | 0.8% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:13762T>G | 4.0% | 0.0% | 0.0% | 4.2% | 4.5% | 3.9% | 0.0% | 3.7% | 0.0% | 5.3% | 0.0% | 0.0% | 4.6% | 0.0% | 6.1% |
| chrM:539T>A | 0.0% | 0.0% | 4.9% | 7.1% | 0.0% | 4.6% | 0.0% | 4.6% | 6.8% | 0.0% | 0.0% | 4.9% | 4.9% | 0.0% | 0.0% |
| chrM:814A>G | 0.8% | 0.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:10933C>CGGGGGGGGG | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 2.4% | 0.0% | 0.0% | 0.0% | 0.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:10569G>A | 0.0% | 0.7% | 0.0% | 0.0% | 0.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:10935A>G | 5.7% | 0.0% | 0.0% | 0.0% | 4.8% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:13768T>G | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 3.1% | 0.0% | 0.0% |
| chrM:301A>C | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 3.9% | 0.0% | 0.0% |
| chrM:10936C>G | 4.5% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:3776G>A | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.8% | 0.0% |
| chrM:3577A>C | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 4.2% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:10926T>G | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 2.1% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:302A>ACCC | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 20.5% | 0.0% | 0.0% |
| chrM:10941T>G | 4.6% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:484A>C | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 5.2% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| chrM:10925T>G | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |

The data from the table on the previous page is displayed graphically in the figure to the right.




The following table shows the same data, expressed as the number of reference and alternate reads (displayed as reference;alternate) measured in the sample in order to provide full sight into the coverage at each variant site.

Note that in cases where the variant is not called due to the fraction of alternate reads falling below the calling threshold, the value is defaulted to 0;0 regardless of the actual number of supporting reads.

| Variant name | 0% | 1% | 2% | 3% | 5% | 10% | 25% | 50% | 75% | 90% | 95% | 97% | 98% | 99% | 100% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chrM:750A>G | 0;1706 | 0;1718 | 0;1845 | 1;2539 | 0;2569 | 0;2018 | 0;1995 | 0;2217 | 0;1872 | 0;1689 | 0;2285 | 0;1620 | 0;1860 | 0;1714 | 0;2753 |
| chrM:15326A>G | 1;1525 | 1;2369 | 0;2608 | 0;1549 | 0;2245 | 0;1557 | 0;1845 | 1;1461 | 0;1480 | 0;2064 | 0;1587 | 0;1467 | 0;2654 | 0;1497 | 0;1469 |
| chrM:16519T>C | 0;1759 | 1;1582 | 0;1902 | 0;1537 | 1;1494 | 1;1801 | 0;1451 | 1;1717 | 0;1748 | 0;2262 | 1;1817 | 1;1805 | 0;3758 | 0;1783 | 0;1419 |
| chrM:310T>TC | 0;763 | 1;787 | 0;861 | 1;777 | 4;779 | 5;997 | 4;826 | 0;861 | 3;724 | 3;871 | 0;848 | 2;895 | 3;1069 | 0;912 | 1;777 |
| chrM:1438A>G | 1;2045 | 0;2879 | 0;2649 | 0;2709 | 0;2934 | 0;2064 | 0;2591 | 0;2530 | 0;3172 | 1;3359 | 0;2175 | 0;3176 | 0;3094 | 0;3091 | 0;2486 |
| chrM:8860A>G | 0;1837 | 0;1743 | 0;1648 | 0;1701 | 0;1753 | 0;1498 | 0;1751 | 0;1612 | 0;1707 | 1;1765 | 0;1713 | 0;1539 | 0;1638 | 1;1966 | 0;2204 |
| chrM:263A>G | 2;1297 | 0;1321 | 0;1435 | 3;1748 | 0;942 | 0;1143 | 0;990 | 0;952 | 1;1248 | 0;1096 | 0;1027 | 0;1056 | 0;1415 | 0;1484 | 0;895 |
| chrM:4769A>G | 0;2340 | 0;1777 | 0;2250 | 0;2134 | 0;2753 | 0;2252 | 0;2098 | 0;3009 | 0;2204 | 0;1587 | 0;1887 | 1;1966 | 0;2015 | 0;2922 | 0;1961 |
| chrM:10535T>C | 0;0 | 2482;37 | 2406;70 | 2700;94 | 2340;151 | 2688;419 | 1440;631 | 1461; 1768 | 384;1596 | 179;2543 | 84;1765 | 75;3207 | 46;2794 | 10;1904 | 6;2376 |
| chrM:6293T>A | 10;2025 | 43;2663 | 70;2484 | 93;1948 | 151;2094 | 369;2345 | 908;2013 | 1949; 1504 | 1733;428 | 3424;262 | 3458;118 | 3871;90 | 3838;56 | 3604;30 | 0;0 |
| chrM:200A>G | 22;3226 | 57;3359 | 111;3255 | 131;2959 | 222;3322 | 311;2176 | 1082; 2691 | 2054; 1665 | 2719;630 | 3456;261 | 3779;135 | 3772;94 | 3851;76 | 3335;35 | 0;0 |
| chrM:8602T>C | 1;1391 | 32;1832 | 72;2340 | 90;2404 | 116;1903 | 308;2483 | 601;1294 | 1540; 1163 | 1601;348 | 1589;101 | 1474;61 | 2621;62 | 2948;49 | 2489;23 | 0;0 |
| chrM:15314G>A | 0;0 | 2504;26 | 2409;71 | 1513;64 | 2239;154 | 1345;200 | 1429;612 | 628;882 | 312;1196 | 131;1924 | 82;1736 | 34;1482 | 24;2397 | 22;1750 | 2;1510 |
| chrM:6776T>C | 0;0 | 2358;40 | 2382;66 | 2479;105 | 2356;174 | 2001;302 | 1868;782 | 1158; 1398 | 512;2118 | 132;1993 | 110;2723 | 53;2875 | 55;2765 | 22;2740 | 4;1821 |
| chrM:302A>AC | 649;81 | 642;95 | 679;102 | 649;96 | 628;73 | 736;115 | 491;93 | 399;70 | 281;55 | 212;64 | 188;67 | 193;64 | 0;0 | 181;67 | 177;52 |
| chrM:14212T>C | 2;1401 | 28;1531 | 38;1418 | 70;1972 | 145;1768 | 182;1208 | 552;1215 | 1274;915 | 1082;302 | 1583;116 | 1349;50 | 1387;32 | 1625;22 | 1501;12 | 0;0 |
| chrM:8937T>C | 5;2045 | 37;1898 | 42;1520 | 64;1823 | 126;1838 | 215;1467 | 564;1292 | 1080;844 | 1358;327 | 1967;139 | 2054;95 | 1950;41 | 1788;26 | 0;0 | 0;0 |
| chrM:3010G>A | 6;2910 | 49;3085 | 38;1532 | 114;3013 | 183;2665 | 412;2874 | 629;1417 | 1168;975 | 1906;479 | 2239;152 | 2604;98 | 2312;46 | 1489;26 | 0;0 | 0;0 |

| Variant name (continued) | 0% | 1% | 2% | 3% | 5% | 10% | 25% | 50% | 75% | 90% | 95% | 97% | 98% | 99% | 100% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chrM:302A>ACC | 0;0 | 0;0 | 679;16 | 0;0 | 628;18 | 736;76 | 491;180 | 399;274 | 281;324 | 212;514 | 188;502 | 193;538 | 267;637 | 181;509 | 177;448 |
| chrM:16189T>C | 54;606 | 71;590 | 92;605 | 107;572 | 127;559 | 301;509 | 560;463 | 862;225 | 1285;113 | 1354;30 | 1402;15 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:9053G>A | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 2108;39 | 1553;57 | 1916;163 | 2029;217 | 1350;166 | 1549;173 | 1539;193 | 1893;254 | 1530;206 | 1488;211 |
| chrM:16183A>ACC | 428;108 | 417;117 | 463;104 | 480;88 | 494;93 | 634;65 | 817;77 | 991;32 | 1266;19 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:16183A>AC | 428;112 | 417;122 | 463;114 | 480;90 | 494;100 | 634;106 | 817;88 | 991;33 | 1266;24 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:10785T>C | 2078;78 | 2459;92 | 1479;58 | 2024;74 | 2439;89 | 2501;84 | 2605;60 | 3170;46 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:16188CT>C | 616;52 | 624;46 | 662;42 | 641;46 | 661;47 | 772;48 | 998;33 | 1071;25 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:7162G>A | 2705;87 | 2335;83 | 2061;63 | 2828;87 | 2370;56 | 2506;74 | 2184;38 | 2312;27 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:1552G>A | 2132;36 | 3029;55 | 3084;56 | 2933;50 | 2943;53 | 3358;56 | 3195;53 | 3461;29 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:13762T>G | 988;41 | 0;0 | 0;0 | 1083;48 | 1464;69 | 1023;42 | 0;0 | 927;36 | 0;0 | 1014;57 | 0;0 | 0;0 | 966;47 | 0;0 | 1134;74 |
| chrM:539T>A | 0;0 | 0;0 | 1083;56 | 1091;83 | 0;0 | 1082;52 | 0;0 | 1087;53 | 1750; 127 | 0;0 | 0;0 | 1034;53 | 1125;58 | 0;0 | 0;0 |
| chrM:814A>G | 1460;12 | 1756;13 | 0;0 | 0;0 | 0;0 | 0;0 | 1836;13 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:10933C> CGGGGGGGGG | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 1671;41 | 0;0 | 0;0 | 0;0 | 821;6 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:10569G>A | 0;0 | 2633;18 | 0;0 | 0;0 | 2484;18 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:10935A>G | 1110;67 | 0;0 | 0;0 | 0;0 | 1197;61 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:13768T>G | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 983;31 | 0;0 | 0;0 |
| chrM:301A>C | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 1170;48 | 0;0 | 0;0 |
| chrM:10936C>G | 1105;52 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:3776G>A | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 1447;11 | 0;0 |
| chrM:3577A>C | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 1724;75 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:10926T>G | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 1737;37 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:302A>ACCC | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 267;69 | 0;0 | 0;0 |
| chrM:10941T>G | 1014;49 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:484A>C | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 957;53 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |
| chrM:10925T>G | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 | 848;6 | 0;0 | 0;0 | 0;0 | 0;0 | 0;0 |

This assay demonstrates the ability of the Genomic Unity® test to correctly identify mitochondrial variants down to 5% heteroplasmy.

This study does not present validation for reproducible calling of large deletions of the mitochondrial genome, due to too small of a number of available controls. However, it should be noted that both CAP proficiency large mitochondrial deletion samples were accurately identified and reported (see structural variant table below).

**Structural variants (SVs)**

Analysis of SVs is performed using a proprietary combination of break point analysis of split and discordant reads and read depth analysis, which is described in detail in the publication by Neerman et al[1].

Unlike for SNVs and indels, no true positive "gold standard" variant set of acceptable quality is currently available for the secondary validation of SVs. The closest is the dataset published by Parikh et al[2] which uses the svclassify method to establish bench mark structural variant calls. Different sequencing and variant calling methods applied by different research groups produced sets of "true positive" variants which vary between each other by nearly an order of magnitude in terms of quantity and overlap of detected variants. Close examination of a representative group of "true positive" SVs identified by different approaches revealed large numbers of false positives and false negatives, therefore making use of this data less than ideal for analytical validation of this clinical test. The statistics below are based on the available true positive data set NA24385 and will be used until better quality true positive SV data becomes available.

| | | | | |
|---|---|---|---|---|
| CNV <1,000 | True positive (TP) | 1,586 | False negative (FN) | 472 |
| | False positive (FP) | 574 | Sensitivity | 0.77100 |
| CNV 1,000-10,000 | True positive (TP) | 486 | False negative (FN) | 82 |
| | False positive (FP) | 311 | Sensitivity | 0.85600 |
| CNV 10,000-100,000 | True positive (TP) | 27 | False negative (FN) | 11 |
| | False positive (FP) | 111 | Sensitivity | 0.71000 |
| CNV >100,000 | True positive (TP) | 1 | False negative (FN) | 0 |
| | False positive (FP) | 0 | Sensitivity | 1.00000 |

To more accurately reflect the true sensitivity and specificity of the algorithms, we analyzed known structural variants which are summarized in the table below. The following data set comprises true positive (causative pathogenic SVs confirmed by orthogonal detection techniques) and true negative clinical samples (those of healthy individuals or affected individuals with another type of causative genetic variant). Most were obtained from public collections, while some originated from other sources. This data set is discussed in more detail in the publication by Neerman et al[1].

| Type of variant (Size, bp) | Variantyx detected (Size, bp) |
|---|---|
| 24 variants including 2 SVs (126bp, 343bp) | Yes (126), Yes (break point) |
| Trisomy Chr 21 | Yes (whole chromosome) |
| Trisomy Chr 18 | Yes (whole chromosome) |
| Trisomy Chr 13 | Yes (whole chromosome) |
| SNV + gross deletion (10kbp) | Yes (9,789 bp) |
| Compound het point mutation with large deletion, Krabbe disease, GALC (31kbp) | Yes (31,666 bp) |
| Deletion chrX male (8kbp) | Yes (8,749 bp) |
| Het deletion GGA compound with splicing point mutation (8kbp) | Yes (8,267 bp) |
| arr 22q11.22q11.23(21,390,449-21,978,719)x1 | Yes (679,500 bp) |
| arr 2q37.2q37.3(236,218,793-242,654,701)x3,1 1q25(131,542,057-134,434,130)x1 | Yes (6,484,000 bp), Yes (2,925,000 bp) |

| | |
|---|---|
| arr[hg19] 5q33.1q35.3(152,281,639-180,686,444)x3,7q36.2q36.3(155,040,999-159,123,167)x1 | Yes (4,117,500 bp), Yes (337,000 bp, and few additional smaller duplication within the same locus) |
| Deletion DMD exon 44 | Yes (195,012 bp) |
| Gross deletion (6 mbp) | Yes (5,979,236 bp) |
| Gross deletion (340 kbp) | Yes (340,153 bp) |
| 45 bp deletion | Yes (45bp) |
| Gross duplication (700 kbp) | Yes (712,000 bp) |
| Chromosome deletion | Yes (16,095,000 bp) |
| Duplicated chromosome copy number variation (CNV) reference panel | Yes (17,983,000 bp) |
| Translocated chromosome | Yes (3,393,500 bp deletion; 8,112,000 bp duplication) |
| Recombinant chromosome copy number variantion (CNV) reference panel 02 | Yes (1,108,000 bp deletion; 13,109,000 bp duplication) |
| Mitochondrial deletion | Yes (8,959 bp) |
| Mitochondrial deletion | Yes (7,647 bp) |

The above samples include 18 SVs >100kb, of which all were detected, supporting a sensitivity and specificity of detection of large structural variants of 100%.

The Genomic Unity® test detects pathogenic SVs of multiple types with over 96% clinical sensitivity, with examples of reported pathogenic variants ranging from a 45 bp deletion to a complex rearrangement involving millions of base pairs on three different chromosomes.

**Short tandem repeat expansions (STRs)**

Analysis of STRs is performed using the application of three paired-end read strategies: de novo assembly of spanning reads, alignment and counting of repeats in anchored reads and counting and statistical normalization of reads containing only repeat sequence.

No true positive pathogenic STRs are available within GIAB samples, thus these types of variants were validated using samples from patients with confirmation of disease and a confirmed genetic diagnosis made by an orthogonal technology at CLIA-certified laboratories. Further details are found in the table below.

| Sample | Gene (region) | Repeat unit | Reference count | Variantyx count | Reporting threshold |
|---|---|---|---|---|---|
| 1 | AR | CAG | 51 | 52 | 35 |
| 2 | ATXN1 | CAG | 15/65 | 19/68 | 36 |
| 3 | ATXN1 | CAG | 29/52 | 30/61 | 36 |
| 4 | ATXN1 | CAG | 31/43 | 32/38 | 36 |
| 5 | ATXN1 | CAG | 16/68 | 20/67 | 36 |
| 6 | ATXN1 | CAG | 32/60 | 33/69 | 36 |
| 7 | ATXN2 | CAG | Intermediate (27-33) | 23/27 | 27 |
| 8 | ATXN2 | CAG | Intermediate (27-33) | 23/27 | 27 |
| 9 | ATXN2 | CAG | Intermediate (27-33) | 22/27 | 27 |
| 10 | ATXN2 | CAG | Intermediate (27-33) | 22/31 | 27 |
| 11 | ATXN2 | CAG | Intermediate (27-33) | Intermediate 23/27 | 27 |

| Sample | Gene (region) | Repeat unit | Reference count | Variantyx count | Reporting threshold |
|---|---|---|---|---|---|
| 12 | ATXN2 | CAG | Intermediate (27-33) | Intermediate 23/27 | 27 |
| 13 | ATXN2 | CAG | Intermediate (27-33) | Intermediate 23/27 | 27 |
| 14 | ATXN2 | CAG | Intermediate (27-33) | Intermediate 22/31 | 27 |
| 15 | ATXN3 | CAG | 24/74 | 57/63 | 27 |
| 16 | ATXN7 | CAG | 8/62 | 7/76 | 27 |
| 17 | C9ORF72 | GGGGCC | Positive | 2/1062 | 25 |
| 18 | C9ORF72 | GGGGCC | Positive | 2/903 | 25 |
| 19 | C9ORF72 | GGGGCC | Positive | 2/336 | 25 |
| 20 | C9ORF72 | GGGGCC | Positive C9Orf72 5/84/>145 | Positive C9Orf72 5/176 | 25 |
| 21 | C9ORF72 | GGGGCC | Positive C9Orf72 5/84/>145 | Positive C9Orf72 5/176 | 25 |
| 22 | C9ORF72 | GGGGCC | Positive C9Orf72 8/>145 | Positive C9Orf72 8/268 | 25 |
| 23 | C9ORF72 | GGGGCC | Positive C9Orf72 8/>145 | Positive C9Orf72 8/268 | 25 |
| 24 | C9ORF72 | GGGGCC | Positive C9Orf72 10/>145 | Positive C9Orf72 10/114 | 25 |
| 25 | C9ORF72 | GGGGCC | Positive C9orf72 10/>145 | Positive C9Orf72 10/114 | 25 |
| 26 | C9ORF72 | GGGGCC | Positive C9Orf72 4/>145 | Positive C9Orf72 4/141 | 25 |
| 27 | C9ORF72 | GGGGCC | Positive C9Orf72 4/>145 | Positive C9Orf72 4/141 | 25 |
| 28 | C9ORF72 | GGGGCC | Intermediate C9Orf72 11/28 | Intermediate C9Orf72 11/32 | 25 |
| 29 | C9ORF72 | GGGGCC | Intermediate C9Orf72 11/28 | Intermediate C9Orf72 11/32 | 25 |
| 30 | DMPK | CTG | Normal/over 1500 | 14/2340 | 36 |
| 31 | DMPK | CTG | Normal/over 1500 | 5/2255 | 36 |
| 32 | DMPK | CTG | 1000 | 5/1022 | 36 |
| 33 | DMPK | CTG | 500-1500 | 12/684 | 36 |
| 34 | DMPK | CTG | 500 | 532 | 36 |
| 35 | DMPK | CTG | Normal/66 | 12/71 | 36 |
| 36 | DMPK | CTG | 5/1700 | 5/1661 | 36 |
| 37 | DMPK | CTG | 500-1000 | 12/689 | 36 |
| 38 | DMPK | CTG | Normal/500 | 13/482 | 36 |
| 39 | DMPK | CTG | Affected | 12/572 | 36 |
| 40 | DMPK | CTG | 700 | 21/517 | 36 |
| 41 | DMPK | CTG | 340 | 21/288 | 36 |
| 42 | DMPK | CTG | Normal/150-160 | 5/416 | 36 |
| 43 | DMPK | CTG | Affected | 14/119 | 36 |
| 44 | DMPK | CTG | 50-80 | 14/119 | 36 |
| 45 | DMPK | CTG | 1500-2000 | 13/3620 | 36 |
| 46 | DMPK | CTG | Normal/80-90 | 22/127 | 36 |
| 47 | DMPK | CTG | Normal/130-140 | 5/405 | 36 |
| 48 | FMR1 | CGG | 30 | 30 | 45 |
| 49 | FMR1 | CGG | 477 | 87 | 45 |
| 50 | FMR1 | CGG | 80-93 | 67 | 45 |
| 51 | FMR1 | CGG | 23 | 23 | 45 |
| 52 | FMR1 | CGG | 20/29 | 20/29 | 45 |
| 53 | FMR1 | CGG | Affected | 101 | 45 |
| 54 | FMR1 | CGG | 23 | 23 | 45 |
| 55 | FMR1 | CGG | Affected | 112 | 45 |
| 56 | FMR1 | CGG | 23 | 23 | 45 |
| 57 | FMR1 | CGG | 23/30 | 23/30 | 45 |
| 58 | FMR1 | CGG | Affected | 174 | 45 |

| Sample | Gene (region) | Repeat unit | Reference count | Variantyx count | Reporting threshold |
|---|---|---|---|---|---|
| 59 | FMR1 | CGG | Affected | 174 | 45 |
| 60 | FMR1 | CGG | 32/107 | 33/73 | 45 |
| 61 | FMR1 | CGG | Affected | 69 | 45 |
| 62 | FMR1 | CGG | 23/95-140 | 23/80 | 45 |
| 63 | FMR1 | CGG | 24, 99 | 24/66 | 45 |
| 64 | FMR1 | CGG | 29, 69 | 29/56 | 45 |
| 65 | FMR1 | CGG | 31, >200 | 31/113 | 45 |
| 66 | FMR1 | CGG | 34, >200 | 34/98 | 45 |
| 67 | FMR1 | CGG | 69 | 68 | 45 |
| 68 | FMR1 | CGG | >200 | 81 | 45 |
| 69 | FMR1 | CGG | >200 | 91 | 45 |
| 70 | FMR1 | CGG | 20/183-193 | 20/74 | 45 |
| 71 | FMR1 | CGG | 30/80 | 59/70 | 45 |
| 72 | FMR1 | CGG | 31/46 | 31/54 | 45 |
| 73 | FMR1 | CGG | 29/93-110 | 29/66 | 45 |
| 74 | FMR1 | CGG | Affected | 105 | 45 |
| 75 | FMR1 | CGG | 30/75-89 | 30/71 | 45 |
| 76 | FMR1 | CGG | 76 | 56 | 45 |
| 77 | FMR1 | CGG | Affected | 97 | 45 |
| 78 | FMR1 | CGG | 931-940 | 117 | 45 |
| 79 | FMR1 | CGG | 29/41 | 29/29 | 45 |
| 80 | FMR1 | CGG | 53 | 58 | 45 |
| 81 | FMR1 | CGG | 23/30 | 23/30 | 45 |
| 82 | FMR1 | CGG | 46 | 54 | 45 |
| 83 | FMR1 | CGG | 645 | 115 | 45 |
| 84 | FMR1 | CGG | 29/29 | 29/29 | 45 |
| 85 | FMR1 | CGG | 100-104 | 71 | 45 |
| 86 | FMR1 | CGG | 23/30 | 23/30 | 45 |
| 87 | FMR1 | CGG | 23/95 | 24/60 | 45 |
| 88 | FMR1 | CGG | Affected | 126 | 45 |
| 89 | FMR1 | CGG | 29/31 | 29/31 | 45 |
| 90 | FMR1 | CGG | 80-96 | 63 | 45 |
| 91 | FMR1 | CGG | 501-550 | 100 | 45 |
| 92 | FMR1 | CGG | 29/30 | 29/30 | 45 |
| 93 | FMR1 | CGG | 100-118 | 80 | 45 |
| 94 | FMR1 | CGG | Affected | 127 | 45 |
| 95 | FMR1 | CGG | 23 | 23 | 45 |
| 96 | FMR1 | CGG | Affected | 87 | 45 |
| 97 | FMR1 | CGG | 41 | 35 | 45 |
| 98 | FMR1 | CGG | >200 | 89 | 45 |
| 99 | FMR1 | CGG | 31/53 | 31/52 | 45 |
| 100 | FMR1 | CGG | 30 | 30 | 45 |
| 101 | FMR1 | CGG | 23/29 | 24/29 | 45 |
| 102 | FMR1 | CGG | 29/45 | 29/40 | 45 |
| 103 | FMR1 | CGG | 23/29 | 23/29 | 45 |
| 104 | FMR1 | CGG | 30/73 | 30/57 | 45 |
| 105 | FMR1 | CGG | Carrier | 32/107 | 45 |

| Sample | Gene (region) | Repeat unit | Reference count | Variantyx count | Reporting threshold |
|---|---|---|---|---|---|
| 106 | FMR1 | CGG | 30/78 | 30/54 | 45 |
| 107 | FMR1 | CGG | 117 | 70 | 45 |
| 108 | FMR1 | CGG | 23/70 | 23/56 | 45 |
| 109 | FMR1 | CGG | 29/over 200 | 29/71 | 45 |
| 110 | FXN | GAA | 670/830 | 22/169 | 60 |
| 111 | FXN | GAA | 670/830 | 6/193 | 60 |
| 112 | FXN | GAA | Normal/830 | 8/90 | 60 |
| 113 | FXN | GAA | Normal/670 | 19/96 | 60 |
| 114 | FXN | GAA | 200/500 | 116/186 | 60 |
| 115 | FXN | GAA | 760/830 | 113/178 | 60 |
| 116 | FXN | GAA | Normal/700 | 8/84 | 60 |
| 117 | FXN | GAA | Normal/830 | 9/112 | 60 |
| 118 | FXN | GAA | 800/800 | 102/155 | 60 |
| 119 | FXN | GAA | 280/830 | 8/183 | 60 |
| 120 | FXN | GAA | 500 | 8/113 | 60 |
| 121 | FXN | GAA | Normal/760 | 9/111 | 60 |
| 122 | FXN | GAA | 600/700 | 100/150 | 60 |
| 123 | FXN | GAA | 330/380 | 97/146 | 60 |
| 124 | FXN | GAA | 420/541 | 99/147 | 60 |
| 125 | FXN | GAA | Normal/420 | 8/94 | 60 |
| 126 | FXN | GAA | 580/580 | 112/174 | 60 |
| 127 | FXN | GAA | Normal/830 | 79/109 | 60 |
| 128 | FXN | GAA | 650/1030 | 111/175 | 60 |
| 129 | FXN | GAA | Normal/830 | 9/101 | 60 |
| 130 | FXN | GAA | 9/1285+50 | 9/85 | 60 |
| 131 | FXN | GAA | 670/1170 | 116/183 | 60 |
| 132 | FXN | GAA | 630/830 | 109/168 | 60 |
| 133 | FXN | GAA | Normal/830 | 8/103 | 60 |
| 134 | FXN | GAA | 530/530 | 98/146 | 60 |
| 135 | HTT | CAG | 17/48 | 17/61 | 36 |
| 136 | HTT | CAG | 15/52 | 15/58 | 36 |
| 137 | HTT | CAG | 22/58 | 22/63 | 36 |
| 138 | HTT | CAG | 16/66 | 16/82 | 36 |
| 139 | HTT | CAG | 15/70 | 15/78 | 36 |
| 140 | HTT | CAG | 22/50 | 22/60 | 36 |
| 141 | HTT | CAG | 16/44 | 16/52 | 36 |
| 142 | HTT | CAG | 45/47 | 43/56 | 36 |
| 143 | HTT | CAG | 15/49 | 15/62 | 36 |
| 144 | HTT | CAG | 17/45 | 17/43 | 36 |
| 145 | HTT | CAG | 16/46 | 16/52 | 36 |
| 146 | HTT | CAG | 15/55 | 15/61 | 36 |
| 147 | HTT | CAG | 15/44 | 15/58 | 36 |

The Genomic Unity® test detects pathogenic STRs of multiple types with over 99% clinical sensitivity.

Repeat variance was calculated for each expansion type:

**CAG / CTG repeats (reverse allele, includes DMPK)**
The following genes in this data set have CAG repeat expansions: HTT, ATXN1, ATXN2, ATXN3, ATXN7, AR.

The variance between actual and detected CAG repeats from 1-40 in size is +/- 1.

The variance between actual and detected CAG repeats from >40 in size is +/- 20%.

**GAA repeats**
The following gene in this data set has a GAA repeat expansion: FXN.

FXN GAA repeats detected as 1-60 in size are reported as "normal". The detected repeat size is reported with a variance of +/- 10%.

Due to the abundance of naturally occurring GAA repeats in the genome, repeats >60 in size are reported as "heterozygous" or "homozygous" based on the correlation between detected and actual repeat size determined during validation. Repeats detected as 61-140 in size are indicative of carrier status due to heterozygous expansion of one allele and are reported as "heterozygous". Repeats detected as >140 in size are indicative of a homozygous or compound heterozygous expansion of both alleles and are reported as "homozygous".

Orthogonal confirmation is required to determine exact allele sizes of >60.

**GGGGCC repeats**
The following gene in this data set has a GGGGCC repeat expansion: C9orf72.

The variance between actual and detected GGGGCC repeats from 1-30 in size is +/- 1.

The variance between actual and detected GGGGCC repeats >30 in size is +/- 20%. Note that the variance for repeats >30 in size is challenging to determine as large expansions called with orthogonal technologies do not provide exact repeat counts.

**CGG repeats**
The following gene in this data set has a CGG repeat expansion: FMR1.

FMR1 CGG repeats detected as 1-55 in size are reported as "normal". The detected repeat size is reported with a variance of +/- 1.

FMR1 CGG repeats detected as >55 in size are reported as "pre-mutation or full mutation". These expansions are not reported as an exact number of repeats due to the challenging nature of sequencing and aligning this region.

Orthogonal confirmation is recommended to determine whether the repeat is of pre-mutation (56-200) or full mutation (>200) size.

**Conclusion**

This validation document provides data supporting the ability to reliably identify clinically relevant SNVs, indels, mitochondrial variants, SVs, and STRs using whole genome sequencing and Variantyx's proprietary Genomic Intelligence® software.

**References**

1. Neerman et al. A clinically validated whole genome pipeline for structural variant detection and analysis. BMC Genomics. 2019 Jul 16;20(Suppl 8):545.

2. Parikh H et al. svclassify: A method to establish benchmark structural variant calls. BMC Genomics. 2016 Jan 16;17:64.

## Appendix

The following table provides coverage and sequence depth metrics for three GIAB samples across all loci.

| | NA24149 | NA24143 | NA12878 |
|---|---|---|---|
| Total coverage | 118,834,360,234 | 104,452,111,795 | 103,137,319,942 |
| Total bases | 2,937,658,094 | 2,937,658,094 | 2,937,658,094 |
| <20x coverage bases | 141,669,917 | 116,038,970 | 128,704,868 |
| <8x coverage bases | 23,534,080 | 48,176,144 | 51,596,090 |
| No coverage bases | 6,174,731 | 26,258,294 | 27,964,065 |
| Annotated <8x coverage bases | 1,372 | 1,348 | 1,894 |
| HGMD <8x coverage bases | 770 | 777 | 1,092 |
| ClinVar <8x coverage bases | 739 | 763 | 1,033 |
| Average coverage | 40.45 | 35.56 | 35.11 |
| Median coverage | 40 | 35 | 35 |
| <20x coverage percent | 4.82255% | 3.95005% | 4.38121% |
| <8x coverage percent | 0.80112% | 1.63995% | 1.75637% |
| No coverage percent | 0.21019% | 0.89385% | 0.95192% |
| Annotated <8x percent | 0.37112% | 0.36463% | 0.51232% |
| HGMD <8x percent | 0.37286% | 0.37625% | 0.52878% |
| ClinVar <8x percent | 0.24882% | 0.25690% | 0.34781% |

The following table provides coverage and sequence depth metrics for three GIAB samples in the coding sequences. This is defined as the sum of all coding sequences in the genome, regardless if targeted or not (as defined by Ensembl).

| | NA24149 | NA24143 | NA12878 |
|---|---|---|---|
| Total coverage | 35,574,430 | 35,574,430 | 35,574,430 |
| Total bases | 1,515,714,707 | 1,334,122,315 | 1,196,498,356 |
| <20x coverage bases | 1,183,332 | 1,023,561 | 1,528,704 |
| <8x coverage bases | 234,903 | 344,624 | 444,046 |
| No coverage bases | 36,474 | 101,797 | 110,238 |
| Annotated <8x coverage bases | 989 | 972 | 1,382 |
| HGMD <8x coverage bases | 662 | 673 | 941 |
| ClinVar <8x coverage bases | 460 | 457 | 641 |
| Average coverage | 42.61 | 37.50 | 33.63 |
| Median coverage | 43 | 37 | 34 |
| <20x coverage percent | 3.32636% | 2.87724% | 4.29720% |
| <8x coverage percent | 0.66031% | 0.96874% | 1.24822% |
| No coverage percent | 0.10253% | 0.28615% | 0.30988% |
| Annotated <8x percent | 0.35755% | 0.35140% | 0.49963% |
| HGMD <8x percent | 0.36345% | 0.36949% | 0.51662% |
| ClinVar <8x percent | 0.21912% | 0.21769% | 0.30534% |