TECHNICAL NOTE

Validating TruSight™ Whole Genome

Highly reproducible data meets the requirements of an IVDR-compliant workflow for clinical whole-genome sequencing

illumina®

Introduction

Whole-genome sequencing (WGS) provides the most comprehensive view of the human genome, and often includes regions not interrogated by other methods, such as PCR, chromosomal microarray, and targeted methods such as single-gene tests, multigene next-generation sequencing (NGS) panels, and whole-exome sequencing (WES). PCR-free WGS provides superior coverage of difficult regions and enables simultaneous analysis of thousands of genes with known or suspected disease associations. Discovery of novel causative variants and continued appreciation of intronic variants influence on human disease suggest a differentiated approach with WGS. Numerous publications in the scientific literature have recognized the superior diagnostic potential of WGS, compared to other targeted approaches in patients with broad phenotypes.[1]

Clinical labs transitioning to, or incorporating, a European Union (EU) *In Vitro* Diagnostic Regulation (IVDR)–compliant genomic profiling assay face significant challenges. Specifically, they must deal with the burdens of developing controls, building bioinformatic pipelines, and performing costly and time-consuming analytical validation studies. To address these challenges, Illumina offers TruSight Whole Genome, providing a comprehensive, DNA-to-variant call format (VCF) workflow for clinical WGS (Figure 1). This wet lab-to-secondary analysis solution streamlines assay validation with internal controls, automated variant calling, and analytical validation studies.

This technical note highlights one of the several studies conducted as part of the extensive process Illumina performed to validate TruSight Whole Genome. In addition to the reproducibility study highlighted here, additional studies characterized analytical accuracy, analytical sensitivity, and robustness to potential sources of variability to ensure accurate and consistent performance.[2]

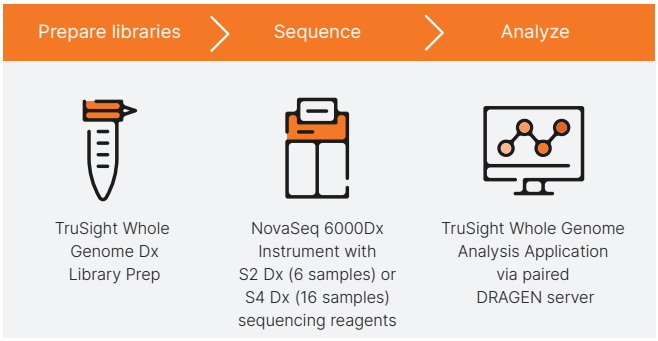

Figure 1: The comprehensive, DNA-to-VCF TruSight Whole Genome workflow.

Methods

The validation study was conducted in accordance with Clinical and Laboratory Standards Institute (CLSI) guidelines across three laboratory sites. Site-specific external reproducibility accounted for sequencing system, operator, day-to-day, and run-to-run variability. Factors tested within laboratory precision included library prep reagent lot, sequencing consumable lot, sequencing system, operator, and run-to-run performance.

Sample set and library preparation

Sixteen unique samples were used to evaluate reproducibility for detection of the variant types validated with TruSight Whole Genome. Eight noncontrived samples used to evaluate known single nulceotide variants (SNVs), copy number variants (CNVs), and runs of homozygosity (ROH) were derived from both males and females of self-identified Caucasian, African, and Asian ancestry to provide a diverse sample set. Remaining samples allowed for evaluation of the absence of *SMN1* c.840C, and low and high levels of mitochondrial SNVs allele frequency or short tandem repeats (STR) loci expansion size.

Libraries were prepared with the TruSight Whole Genome Dx Library Prep Kit. Additional details can be found within the TruSight Whole Genome Dx Library Prep User Guide.

Sequencing and data analysis

Prepared libraries were run on the NovaSeq™ 6000Dx instrument using the 16-plex workflow, with S4 Dx flow cells and reagents. The assay is validated to achieve the same performance using a 6-plex workflow with S2 Dx flow cells and reagents.

Sequencing data was automatically transferred and analyzed on the paired DRAGEN server using the settings defined in the TruSight Whole Genome Analysis Application. Built-in quality controls evaluate defined thresholds of sequencing, FASTQ, and sample library performance. Powered by DRAGEN, secondary analysis to perform highly accurate variant calling across a range of variant types delivers Genome VCF (gVCF) files for flexible downstream germline applications.

Reproducibility of variant calling

The sample pass rate across 576 total libraries (16 unique samples across 36 sequencing runs), defined as the number of samples passing sequencing quality control (QC) metrics on the first attempt, was 99.1%. All test results are based on this initial testing.

Reproducibility of SNVs, insertions/deletions (indels), CNVs, and ROH was assessed by comparing data to a reference call set selected as the sample run with mid-level autosomal coverage from three characterization runs (Table 1). Reproducibility of STR expansions, the *SMN1* variant, and mSNVs was assessed by comparing data to known status (Table 2).

Table 1: Reproducibility of TruSight Whole Genome for small variants, CNVs, and ROH

Variant type stratification	Concordant positive calls[a] Positive calls[b]			Concordant negative calls[c] Negative calls[d]			Average positive agreement (%)[e]	Average negative agreement (%)[f]
	Site 1	Site 2	Site 3	Site 1	Site 2	Site 3		
Small variants of high confidence								
SNVs	6.87e+8 6.88e+8	6.66e+8 6.67e+8	6.88e+8 6.88e+8				99.9%	
Ins 1-5 bp	3.40e+7 3.41e+7	3.30e+7 3.30e+7	3.40e+7 3.41e+7	4.86e+11 4.86e+11	4.70e+11 4.71e+11	4.86e+11 4.86e+11	99.9%	> 99.9%
Del 1-5 bp	4.40e+7 4.42e+7	4.27e+7 4.28e+7	4.41e+7 4.42e+7				99.6%	
Small variants of intermediate confidence								
SNVs	4.22e+7 4.27e+7	4.09e+7 4.13e+7	4.22e+7 4.27e+7				98.8%	
Ins 1-5 bp	1.10e+7 1.12e+7	1.07e+7 1.08e+7	1.10e+7 1.12e+7				98.9%	
Ins 6-15 bp	4.30e+6 4.33e+6	4.17e+6 4.20e+6	4.30e+6 4.34e+6				99.3%	
Ins ≥ 16 bp	6.11e+5 6.32e+5	5.93e+5 6.12e+5	6.12e+5 6.32e+5	1.72e+10 1.74e+10	1.66e+10 1.68e+10	1.72e+10 1.74e+10	96.8%	99.0%
Del 1-5 bp	2.45e+7 2.48e+7	2.38e+7 2.40e+7	2.45e+7 2.48e+7				98.9%	
Del 6-15 bp	8.73e+6 8.90e+6	8.47e+6 8.62e+6	8.74e+6 8.90e+6				98.2%	
Del ≥ 16 bp	3.59e+6 3.77e+6	3.48e+6 3.66e+6	3.59e+6 3.78e+6				95.0%	
CNVs gains ≥ 10 kb	7883 8275	7664 8012	7916 8282	5.92e+11 5.92e+11	5.73e+11 5.73e+11	5.92e+11 5.92e+11	95.5%	> 99.9%
CNVs losses ≥ 10 kb	11,517 12,089	11,248 11,777	11,516 12,113	5.92e+11 5.92e+11	5.74e+11 5.74e+11	5.92e+11 5.92e+11	95.3%	> 99.9%
ROH ≥ 500 kb	6641 6765	6519 6663	6616 6756	5.42e+11 5.47e+11	5.25e+11 5.30e+11	5.43e+11 5.47e+11	98.0%	99.2%

a. Total number of concordant positive calls = Query Concordant Positive (QCP) + Reference Concordant Positive (RCP).

b. Total number of positive calls = QCP + Query Exclusive Positive (QEP) + RCP + Reference Exclusive Positive (REP).

c. Total number of concordant negative calls = 2 × Concordant Negative (CN).

d. Total number of negative calls = 2 × CN + Reference Exclusive Negative (REN) + Query Exclusive Negative (QEN).

e. Average positive agreement = concordant positive calls/positive calls.

f. Average negative agreement = concordant negative calls/negative calls.

Abbreviations: Ins, insertions; del, deletions.

Table 2: Reproducibility of TruSight Whole Genome for STR, *SMN1*, and mSNVs

Variant type stratification	Total expected positive calls	Positive calls			Total expected negative calls	Negative calls			Percent positive calls (PPC)	Percent negative calls (PNC)
		Site 1	Site 2	Site 3		Site 1	Site 2	Site 3		
STR expansions: high LoD										
FMR1	35	12	11	12	—	—	—	—	100%	—
HTT	36	12	12	12	—	—	—	—	100%	—
FMR1 and *HTT*	71	24	23	24	—	—	—	—	100%	—
STR expansions: low LoD										
FMR1	36	11	10	11	—	—	—	—	88.9%	—
HTT	36	12	12	12	—	—	—	—	100%	—
FMR1 and *HTT*	72	23	22	23	—	—	—	—	94.4%	—
STR expansions: 28 main target STR loci combined					285	96	93	96	—	100%
Absence of SMN1 c.840C	71	24	24	23	285	96	93	96	100%	100%
mSNVs, high LoD	1080	360	360	360	457,524	152,491	152,489	152,484	100%	> 99.9%
mSNVs, low LoD	1080	360	359	360	457,524	152,481	152,489	152,483	99.9%	> 99.9%

Abbreviations: LoD, limit of detection.

Summary

TruSight Whole Genome is a clinical WGS workflow that provides reproducible detection of germline variants. This technical note presents a summary of external reproducibility data generated as part of the assay validation. The percentage of samples that met or exceeded passing QC criteria was > 99% in 576 total samples. Little variance was observed when evaluating multiple factors, including site, instrument, operators, preparations, runs, and replicates. Thus, this study demonstrates that TruSight Whole Genome generates highly reproducible results, providing a robust, consistent DNA-to-VCF workflow. This consistency facilitates adoption by streamlining the process of performing costly and time-consuming analytical validation studies.

Learn more

TruSight Whole Genome

NovaSeq 6000Dx Instrument

DRAGEN secondary analysis

References

1. 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med*. 2021;385(20):1868-1880. doi:10.1056/NEJMoa2035790.
2. Marshall CR, Chowdhury S, Taft RJ, et al. Best practices for the analytical validation of clinical whole-genome sequencing intended for the diagnosis of germline disease. *NPJ Genom Med*. 2020;5:47. Published 2020 Oct 23. doi:10.1038/s41525-020-00154-9.

Intended use statement

TruSight Whole Genome is a qualitative *in vitro* diagnostic device intended for whole-genome sequencing and detection of single nucleotide variants, insertion/deletions, copy number variants, runs of homozygosity, short tandem repeat expansions, and mitochondrial variations in human genomic DNA extracted from blood.

TruSight Whole Genome includes the TruSight Whole Genome Dx Library Prep with UD Indexes and the TruSight Whole Genome Analysis Application Software. The device is intended to be used with compatible downstream germline applications to develop *in vitro* diagnostic assays, and by qualified laboratory personnel and assay developers.

TruSight Whole Genome is intended to be used on the NovaSeq 6000Dx Instrument.

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-EMEA-01010 v1.0