Analytical and clinical validation of a targeted-enhanced whole genome sequencing-based comprehensive genomic profiling test.

Stephanie Ferguson[1], Shruthi Sriram[1], Jonathan Kyle Wallace[1], Jeonghoon Lee[1], Jung-Ah Kim[1], Yoonsuh Lee[1], Brian Baek-Lok Oh[1], Won Chul Lee[1], Sangmoon Lee[1] Erin Connolly-Strong[1]

[1] Genome Insight Inc, San Diego CA

Abstract:

Objective: To evaluate the analytical and clinical performance of the targeted enhanced whole-genome sequencing (TE-WGS) assay for comprehensive genomic profiling in oncology.

Methods: Analytical validation of the TE-WGS assay focused on sensitivity and specificity for single nucleotide variants (SNVs), insertions and deletions (indels), and structural variants (SVs). Clinical performance was benchmarked against established orthogonal methods for somatic variant detection.

Results: Analytical validation revealed a sensitivity of 99.8% for SNVs and 98.7% for indels. The positive predictive value (PPV) was 99.3% for SNVs and 98.7% for indels. Clinical validation demonstrated high concordance with reference methods in variant characterization.

Conclusion: The TE-WGS assay is a clinically valuable tool, providing precise and validated comprehensive genomic profiling that enhances personalized cancer management strategies, with the flexibility to integrate new oncological markers.

Keywords: Comprehensive genomic profiling, whole-genome sequencing, precision medicine.

1. Introduction:

Clinical application of next-generation sequencing (NGS) is transforming the management of diseases by offering precise genetic insights for personalized therapy. High-throughput NGS aims to pinpoint unique mutational landscapes in tumors to inform tailored treatment strategies or enrollment in specific clinical trials. Studies have demonstrated that precision medicine—leveraging genomic and molecular profiling to tailor treatments to specific tumor alterations—can significantly enhance patient survival, quality of life, and economic outcomes compared to single-gene testing approaches[1,2].

In clinical settings, NGS assays generally target a select set of pivotal genes, employ cancer-type-specific targeting methods, or utilize whole-exome sequencing that focuses solely on

the coding regions of the genome[3]. Whole-genome sequencing (WGS), on the other hand, offers a non-selective and exhaustive analysis of the entire genome, probing both protein-coding and non-coding regions, which may reveal a broader array of actionable therapeutic opportunities [4].

WGS stands out from targeted panel or exome sequencing by eliminating biases associated with sequence capture, thus providing a more complete picture of the genomic landscape [5]. It is particularly noteworthy that tumor-only testing methods can exhibit higher false-positive rates, an issue that is pronounced in patients of non-European ancestry. The standard practice for filtering out benign germline variants in tumor analyses involves referencing public SNP databases, which predominantly represent European-derived genetic information and are therefore less effective for other ethnicities [6,7]. Recent evidence suggests that relying solely on population databases to filter germline variations can lead to an overestimation of tumor mutational burden (TMB)[8], skewing the results of clinical trials and affecting patient outcomes in the context of FDA-approved immunotherapies.

This report details the analytical and clinical validation of the CancerVision assay, which employs targeted enhanced whole genome sequencing (TE-WGS) for comprehensive solid tumor analysis. It discusses the assay's reproducibility, sensitivity, and detection limits for single nucleotide variants (SNVs), copy number variations, and structural variations. The unique methodology of CancerVision integrates TE-WGS with concurrent tumor and germline analysis, enabling the accurate characterization of mutational landscapes and structural variations within the tumor, as well as germline contributions. By effectively differentiating between tumor-specific alterations and inherited sequences, the assay significantly increases the accuracy of somatic variant detection across diverse populations. The assay's robust performance is augmented by its integration with a database of FDA-approved treatments and clinical trials, enhancing the decision-making process for targeted therapies. By enabling the identification of both established and novel genomic alterations, CancerVision aims to refine the landscape of genomic profiling in cancer treatment, tailoring therapeutic approaches to the intricate genomic profile of each patient's tumor.

2. Materials and Methods

2.1 Test principle and intended use

The Targeted Enhanced Whole Genome Sequencing (TE-WGS) assay (CancerVision -Genome Insight, San Diego, CA), offers an integrated approach for genomic profiling by combining comprehensive whole genome analysis with a focused examination of targeted biomarker genes (refer to **Supplementary Table 1** for the gene list). This assay systematically identifies a spectrum of genomic aberrations in solid tumor specimens, encompassing single nucleotide variants (SNVs), multiple nucleotide variants (MNVs), small insertions and deletions (indels), copy number alterations (CNAs), and structural variations (SVs). Accompanying these detections, the TE-WGS provides a detailed report on the mutational landscape, encompassing mutational signatures, tumor mutational burden (TMB), microsatellite instability (MSI), and homologous recombination deficiency (HRD). The intended use of TE-WGS for the

identification of genomic alterations with established and potential clinical relevance in solid tumor malignancies.

2.2 Patient and Reference Materials

In this study, a dual approach was employed incorporating rigorously characterized reference materials alongside patient-derived samples. The reference materials comprised cell lines from Genome in a Bottle, commercially sourced FFPE blocks and DNA isolated from FFPE specimens, obtained from Horizon Discovery, United Kingdom and the American Type Culture Collection Manassas, VA. The NA12878, NA24694, NA24695, NA2463, NA24143 cell lines from Genome in a Bottle (National Institute of Standards and Technology, US), were utilized to evaluate analytical performance. Additionally, the reference samples from Horizon Discovery (OncoSpan FFPE) were selected to represent a spectrum of histopathological conditions and encompassed both single nucleotide variants (SNVs) and insertions-deletions (indels) with previously validated allele frequencies.

Clinical validation of the TE-WGS assay was performed using a selected cohort of residual patient samples (N = 56) obtained from a prospective cohort registry study (Ajou University Medical Center), based on the availability of adequate residual formalin-fixed paraffin-embedded (FFPE) material, blood, and/or DNA for comprehensive analysis. The chosen samples met the following inclusion criteria: a confirmed prior cancer diagnosis and patient age of ≥18 years at the time of sample collection. The study protocol received approval from the Ajou University Medical Center Independent Review Board, with all participants providing written informed consent in line with ethical standards set forth in the Declaration of Helsinki.

For the purpose of this study, each cancer patient's tumor sample comprised a minimum of five unstained slides at 5-micron thickness, alongside a corresponding hematoxylin and eosin (H&E) stained slide for tumor tissue delineation. A board-certified pathologist conducted a histopathological evaluation to ascertain the neoplastic content within these tissues.

To ensure assay precision, the NA12878 cell line from the Coriell Institute, Camden NJ, known for its well-documented human genome reference, was incorporated as an external quality control across the validation process.

2.2 Regulatory standards

TE-WGS is performed at the CLIA-certified Genome Insight in the United States. CLIA provides federal regulatory standards that apply to all clinical laboratory testing performed on humans in the United States, excluding clinical trials and basic research.

2.3 Library Preparation

DNA libraries were constructed using the Watchmaker DNA Library Preparation Kit (Watchmaker Genomics, Boulder CO). The process began with enzymatic fragmentation of the DNA, followed by adapter ligation. Post-ligation, a bead-based cleanup was executed, and subsequent library amplification was carried out. Quality control measures included assessing library size distribution, targeting an average fragment size of approximately 300 bp, and

quantifying yield, with a minimum threshold of 15 ng/uL. Library size was determined by the TapeStation 4200 System (Agilent Technologies, Santa Clara, CA). Concentration determinations were made using the Qubit DNA Assay Kit in conjunction with a Qubit 2.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA). Prepared libraries were preserved at ≤ –20°C when not immediately processed to the capture stage, adhering to the kit manufacturer's storage guidelines

2.3 Sequencing and Data Analysis Pipeline

The comprehensive genomic analysis was performed using the TE-WGS CancerVision system (Genome Insight Inc., San Diego, CA, USA). Sequencing of the prepared DNA libraries was carried out on the Illumina NovaSeq 6000 sequencing system (Illumina Inc.), achieving an average depth of coverage of 40x for tumor samples and 20x for matched blood samples for whole-genome sequencing (WGS). Target-enhanced sequencing for tumor DNA was facilitated by xGen Custom Hybridization Probes (IDT, Inc., Coralville, IA, USA), covering a genomic region of 2.76 Mb, and sequenced to an average depth of 500x.

The resultant raw sequences were mapped to the human reference genome build GRCh38 using the BWA-MEM algorithm. PCR duplicate reads were excised using SAMBLASTER [9]. Variant calling for germline small variants utilized HaplotypeCaller and Strelka2[10,11], while somatic small variant detection employed Strelka2 and Mutect2 [10,11]. Structural variant identification was conducted via Manta[12]. All variants, both germline and somatic, were annotated using the Variant Effect Predictor (VEP) [13] and subjected to rigorous manual review and curation within Genome Insight's proprietary genome browser.

Tumor purity, ploidy, and allele-specific copy number were determined using whole-genome sequencing (WGS) data. These parameters were inferred based on the depth ratio of tumor to normal data and B allele frequency. The primary tool used for these calculations was Sequenza, with supplementary steps implemented to correct tumor purity for copy number stable tumors and to mitigate noise in formalin-fixed paraffin-embedded (FFPE) specimens.

Final FASTQ, VCF, and CRAM files were packaged, encrypted, and transferred to a secure, long-term storage solution, ensuring data integrity and confidentiality.

3. Results

3.1. Analytical Sensitivity and Positive Predictive Value for SNV and Indel Detection of Whole Genome Sequencing

A true-positive (TP) was a mutation of interest (MOI) in the commercial cell lines identified by both the GIAB consortium and the WGS assay. Conversely, a false-negative (FN) was a mutation recognized by the GIAB consortium but missed by the WGS assay. The assay's sensitivity was quantified using the formula: TP / (TP + FN) = Assay Sensitivity. This calculation, applied to data from six reference cell lines (Genome in a Bottle), resulted in an

analytical sensitivity of 99.8% for SNVs (95% CI: 99.79% - 99.83%) and 98.7% for indels (95% CI: 99.14% - 99.72%)—refer to **Table 1**.

False positives (FP) were defined as mutations reported by the WGS assay at loci confirmed as wild-type in the commercial cell lines by GIAB consortium. The assay's positive predictive value (PPV) was determined with the formula: TP / (TP + FP) = Assay PPV. The analytical PPV, evaluated against GIAB standard, demonstrated the WGS assay's PPV of 99.3% for SNVs (95% CI: 99.28% - 99.38%) and 98.7% for indels (95% CI: 98.63% - 98.78%)—refer in Table 1.

3.2 Analytical Sensitivity for SNV and Indel Detection

The CancerVision Targeted Enhanced Whole Genome Sequencing (TE-WGS) test underwent analytical validation assessing numerous parameters, including nucleic acid extraction, sequencing, and data analysis. The assay's performance was gauged using a range of tumor-derived cell lines and standards from established commercial sources, alongside clinical FFPE samples corroborated by orthogonal testing methods. TE-WGS achieved an average mean depth of coverage of 55.3x (ranging from 42.8x to 67.1x) for WGS and 810.6x (ranging from 650.6x to 1147.4x) for TPS.

The reference FFPE DNA (Horizon) included SNVs and indels with validated allele frequencies. The reference sample comprised 25 established variants. Excluding two variants (KIT D816V and PIK3CA E545K) with potentially lower sequencing efficiency due to vector sequences, comparative analysis of 23 variants between expected and detected allelic frequencies via the TE-WGS assay revealed strong agreement, with a correlation coefficient R=0.99 for SNVs and R=0.98 for indels (**Figure 1A and B**)

3.3 Limit of Detection (LOD)

Single Nucleotide Variants (SNVs): The LOD for SNVs within the TE-WGS assay was assessed using a serial dilution approach. A KRAS G12V Reference Standard from Horizon, initially at a 50% variant allele frequency (VAF), was methodically diluted to generate VAFs of 1%, 5%, 10%, and 15%. NA12878 DNA, which harbors no KRAS variants, served as the diluent. As demonstrated in **Figure 2**, the TE-WGS assay's detection of the KRAS G12V variant was consistent and linearly proportional to the expected VAFs. The minimum LOD for SNVs, based on these evaluations, has been established at a VAF of 5%.

Insertions and Deletions (Indels): For indels, the LOD was established by analyzing variant frequencies in OncoSpan FFPE samples from Horizon. The defined minimum LOD for indel detection stands at a VAF of 10%.

Structural Variations (SVs): The LOD for SVs was determined using the Structural Multiplex Reference standard (FFPE curl) from Horizon, which encompasses a variety of known structural variations at distinct VAFs. From this analysis, the assay's LOD for SVs has been set at a minimum VAF of 15%.

The above-stated LODs reflect the assay's robustness in detecting variants of different types and sizes at low allelic frequencies, as substantiated by the comprehensive analytical validation data.

3.4 Reproducibility

Inter-precision Reproducibility: To evaluate Inter-precision reproducibility, a concordance analysis across 56 paired tumor and normal patient specimens was performed. Each specimen's genomic DNA (gDNA) was divided into three aliquots to facilitate parallel processing: two for intra-laboratory testing by different operators and one for external orthogonal validation. Upon examination, an inter-precision reproducibility rate of 95.7% (67/70) was observed.

3.5 Clinical Validation

Clinical validation of the assay encompassed an analysis of 28 paired tumor and normal clinical samples, yielding a cohort of 56 specimens (Table 2). The concordance of somatic variant identification was assessed by comparing the variant calling outcomes from our assay against those derived from a parallel orthogonal clinically validated sequencing assay, conducted by a CLIA-certified laboratory.

Inter-assay precision was determined by measuring the concordance between results from TE-WGS assay and those obtained from Orthogonal Testing. Of the 78 variants assessable by both platforms, TE-WGS assay failed to identify two variants, which were also not reported by Orthogonal Testing, resulting in an overall inter-assay precision of 70/72, or 97.2%.

In the context of Copy Number Variations (CNVs), both the TE-WGS assay and Orthogonal Testing examined 13 CNVs. The TE-WGS assay missed the identification of 1 CNV, which resulted in a CNV inter-assay precision of 12/13, or 92.3%. Notably, the TE-WGS assay identified 64 CNVs not detected by Orthogonal Testing.

For Single Nucleotide Variants (SNVs), the comparison covered 54 SNVs. The TE-WGS assay did not detect one SNV, leading to an SNV inter-assay precision of 53/54, or 98.1%. Moreover, the TE-WGS assay identified 101 additional SNVs that were not reported by Orthogonal Testing.

We estimate microsatellite instability (MSI) by genome-widely examining microsatellite regions. The score reports the number of somatic insertions and deletions per Mb in microsatellite regions across the whole genome of the tumor. A tumor is considered microsatellite stable (MSS) if the score is < 20, and MSI-High if > 20. To validate the MSI test we performed accuracy studies using 28 patient samples tested by an orthogonal approach. Patient samples were selected to demonstrate a range of tumor content from 23% up to 80% (mean 46%) by tumor estimate. MSI status was all stable by TE-WGS, of which concordance rate with the orthogonal test was 100% (28/28).

The Tumor Mutation Burden (TMB) is quantitatively determined using the targeted enhanced whole-genome sequencing (TE-WGS) method, which incorporates a germline subtraction technique to ascertain the number of somatic changes within the entire genome (~2.6Gb). A

TMB value ≥ 10 mutations/Mb is classified as 'high'. Comparison of TMB between two replicates of TE-WGS showed excellent correlation (R = 0.99), which implied the consistency of the TMB calculation by TE-WGS. Comparisons of TMB assessments between TE-WGS and an alternative tumor-only NGS methodology that uses an extensive panel revealed a notable correlation coefficient (R=0.81 in the replicate 1 and R=0.80 in the replicate 2). Despite the general agreement, there were discrepancies; specifically, 3 cases (10.7%) that the orthogonal testing classified as 'high' TMB were identified as 'low' TMB (<10 mutations/Mb) in the TE-WGS analysis.

3.6 Clinical Utility

The potential clinical utility of the TE-WGS assay was evaluated by quantifying the incidence of clinically actionable genetic alterations identified within the patient cohort. Clinically actionable alterations are characterized as those associated with FDA-approved therapeutics for on- or off-label use, as well as pathogenic or likely pathogenic germline variations. Within the study group, twelve patients (42.9%) harbored genomic aberrations for which there are commercially available therapies, excluding investigational treatments (accessible only via clinical trials). Notably, six patients (21.4%) presented with genomic modifications that were concordant with FDA-approved on-label pharmacological interventions specific to their cancer type. Moreover, one patient (3.6%) exhibited germline alterations deemed pathogenic or likely pathogenic according to established clinical criteria. Notably, two patients (7.1%) harbored complex rearrangements in the actionable genes, such as *BRIP1*, *ATR*, and *RAD51B*, which were challenging to detect with a targeted panel approach.

4. Discussion

In the evolving landscape of precision oncology, the utility of comprehensive genomic profiling (CGP) to direct targeted therapies and inform prognosis is well-established [14,15]. The CancerVision targeted enhanced whole-genome sequencing (TE-WGS) assay has been developed and analytically validated to support this paradigm by facilitating a nuanced examination of the genomic alterations that drive tumor behavior. The TE-WGS assay has undergone analytical and clinical validation, proving to be a reliable and accurate tool for comprehensive genomic profiling in oncology. The assay demonstrates high sensitivity and specificity for SNVs, indels, and SVs, with validated limits of detection that enable the identification of clinically relevant mutations at low allelic frequencies. Such sensitivity is crucial for detecting mutations that may be present in only a subset of tumor cells but could potentially drive therapeutic resistance.

The clinical validation of the TE-WGS assay confirms its high accuracy in identifying somatic variants, with results consistent with other reputable methods. Unlike fixed-panel NGS assays, which are restricted to known genomic alterations and quickly become outdated as new biomarkers are identified, the TE-WGS assay's comprehensive approach ensures it captures a broader range of clinically significant mutations. This allows patients to benefit from the most

current genomic insights without delays caused by the need to constantly update and validate new test panels [3,16].

While newer comprehensive genomic tests may utilize tumor/germline comparison, test such as whole exome sequencing do not achieve the sequencing coverage of the TE-WGS assay. This could lead to decreased accuracy and sensitivity for identifying significant clinical variants. Such potential limitations in detecting relevant mutations underscore the challenges seen in TMB analysis, where some tumor-only assays or panel test might overestimate TMB[8]. The precise quantification of TMB is crucial, as it has become an important biomarker for the eligibility of patients for immunotherapy, a rapidly advancing field in cancer treatment. The recent FDA approval of pembrolizumab for high TMB tumors exemplifies the clinical relevance of accurate TMB assessment, despite the ongoing need for more standardized measurement and reporting practices in this area[17]. While this study showed a good correlation between NGS panel test and TE-WGS approaches, instances of discordance were observed, indicating that TE-WGS methodology may have the potential to reduce exposure to ineffective therapeutic approaches.

5. Conclusion

The comprehensive results of this study demonstrate that the TE-WGS is a robust and reliable assay that accurately and reproducibly detects a patient's genomic landscape. These data support the validity of TE-WGS in the clinical decision-making process of solid tumor patients. With a focus on adaptability to new oncological markers and a 14-day turnaround time, it is equipped to support a personalized cancer management approach.

Acknowledgments

The authors would like to thank all patients and Dr. Minsuk Kwon at Ajou University who participated in the clinical study.

Ethics approval and consent to participate

This work combined data from the parent studies (IRB numbers: AJOUIRB-SMP-2022-278 and AJOUIRB-SMP-2021-576) that had previously obtained informed consent from participants and IRB approval. The authors state that they have followed the principles outlined in the Declaration of Helsinki.

Data availability

The patients participating in this study did not consent to public release of sequencing data. The TE-WGS pipeline and associated algorithms are proprietary to Genome Insight Inc.

Funding

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. 2022M3A9G101451221) along with Genome Insight Inc.

References:

[1] Leroy K, Audigier Valette C, Alexandre J, et al. Retrospective analysis of real-world data to evaluate actionability of a comprehensive molecular profiling panel in solid tumor tissue samples (REALM study). PLoS One. 2023;18(9):e0291495.

[2] Agarwala V, Khozin S, Singal G, et al. Real-World Evidence In Support Of Precision Medicine: Clinico-Genomic Cancer Data As A Case Study. Health Aff (Millwood). 2018 May;37(5):765-772.

[3] Malone ER, Oliva M, Sabatini PJB, et al. Molecular profiling for precision cancer therapies. Genome Med. 2020 Jan 14;12(1):8.

[4] Schwartzberg L, Kim ES, Liu D, Schrag D. Precision Oncology: Who, How, What, When, and When Not? Am Soc Clin Oncol Educ Book. 2017;37:160-169.

[5] Vijay P, McIntyre AB, Mason CE, et al. Clinical Genomics: Challenges and Opportunities. Crit Rev Eukaryot Gene Expr. 2016;26(2):97-113.

[6] Garofalo A, Sholl L, Reardon B, et al. The impact of tumor profiling approaches and genomic data strategies for cancer precision medicine. Genome Med. 2016 Jul 26;8(1):79.

[7] Halperin RF, Carpten JD, Manojlovic Z, et al. A method to reduce ancestry related germline false positives in tumor only somatic variant calling. BMC Med Genomics. 2017 Oct 19;10(1):61.

[8] Parikh K, Huether R, White K, et al. Tumor Mutational Burden From Tumor-Only Sequencing Compared With Germline Subtraction From Paired Tumor and Normal Specimens. JAMA Netw Open. 2020 Feb 5;3(2):e200202.

[9] Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 2014 Sep 1;30(17):2503-5.

[10] McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297-303.

[11] Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018 Aug;15(8):591-594.

[12] Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016 Apr 15;32(8):1220-2.

[13] McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Jun 6;17(1):122.

[14] Hagemann IS, Devarakonda S, Lockwood CM, et al. Clinical next-generation sequencing in patients with non-small cell lung cancer. Cancer. 2015 Feb 15;121(4):631-9.

[15] Simons M, Ramaekers B, Peeters A, et al. Observed versus modelled lifetime overall survival of targeted therapies and immunotherapies for advanced non-small cell lung cancer patients - A systematic review. Crit Rev Oncol Hematol. 2020 Sep;153:103035.

[16] Steuten L, Goulart B, Meropol NJ, et al. Cost Effectiveness of Multigene Panel Sequencing for Patients With Advanced Non-Small-Cell Lung Cancer. JCO Clin Cancer Inform. 2019 Jun;3:1-10.

[17] Chan TA, Yarchoan M, Jaffee E, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Ann Oncol. 2019 Jan 1;30(1):44-56.

Figures and Tables

Table 1: Overall performance of whole genome sequencing

	Variant	Specification
Analytical Sensitivity	SNV	99.8% (CI 99.79-99.83)
	Indel	99.3% (CI 99.14-99.72)
Analytical PPV	SNV	99.3% (CI 99.28-99.38)
	Indel	98.7% (CI 98.63-98.78)

Indel, insertion and deletion; PPV, positive predictive value; SNV, single nucleotide variant

Table 2: Tumor types of the samples for clinical validation

Tumor types	Number of samples (percent)
Breast cancer	12 (42.9%)
Stomach cancer	6 (21.4%)
Head and neck cancer	3 (10.7%)
Bladder cancer	2 (7.1%)
Gallbladder cancer	2 (7.1%)
Lung cancer	1 (3.6%)
Colorectal cancer	1 (3.6%)
Small bowel cancer	1 (3.6%)
Total	28 (100%)

Figures

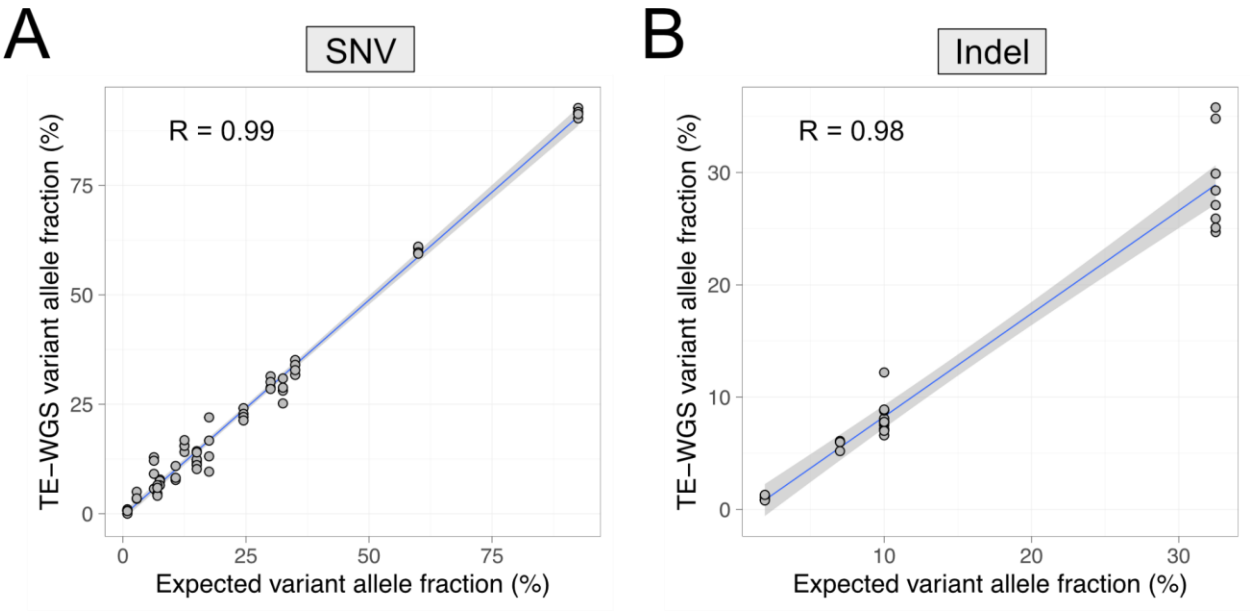

Figure 1. Correlation of expected and observed variant allele fractions in the reference material. Observed variant allele fraction single nucleotide variants (A) and insertion and deletion variants (B) by TE-WGS were concordant with the expected variant allele fractions provided by reference material manufacturer.

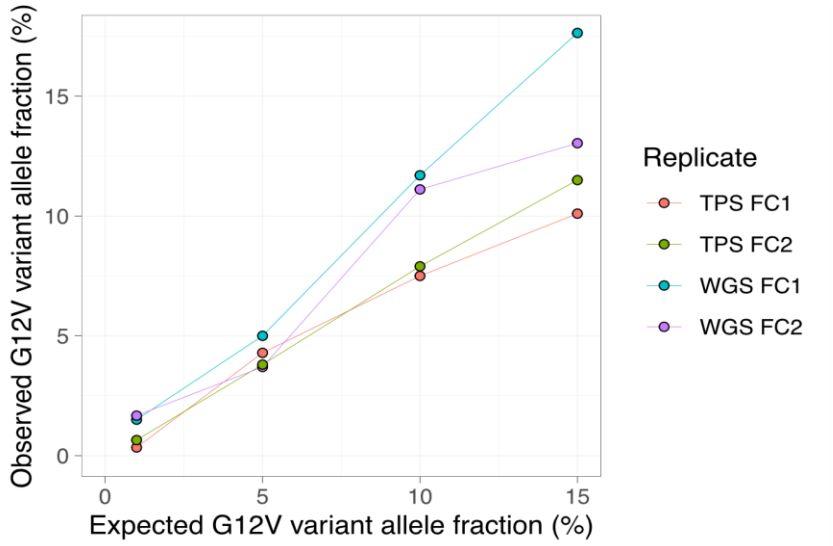

Figure 2. KRAS G12V reference standard serial dilution to assess limit of detection (LOD). The minimum LOD for SNVs has been established at a VAF of 5%.