GeneMind

Sentieon

# Integrated High-Accuracy Genomic Analysis Workflow:

## GeneMind SURFSeq 5000 CMS Sequencing Chemistry Optimized with Sentieon DNAscope Custom Machine Learning Models

### Breakthrough Accuracy

CMS Chemistry + DNAscope AI achieve Q50 precision.

### Population Power

1KCP integration optimizes Indel detection for East Asian genomics.

### Proven Stability

Superior batch consistency for large-scale clinical and research cohorts.

# Executive Summary

Short-read whole-genome (WGS) and whole-exome sequencing (WES) form the backbone of clinical genetic testing, population genomics, and human disease research, yet persistent technical barriers limit reliable variant detection across structurally complex genomic regions[1]. Traditional sequencing workflows suffer from uneven coverage dropout in non-B DNA conformations, platform-specific sequencing error signatures, and reference bias when analyzing genetically diverse cohorts—compounding false-positive and false-negative variant calls for single-nucleotide polymorphisms (SNPs), insertions/deletions (Indels), structural variants (SVs), and copy number variants (CNVs).

To resolve these limitations, GeneMind Biosciences and Sentieon co-developed an end-to-end genomics pipeline pairing the SURFSeq 5000 sequencing platform with proprietary Cross Mountains and Seas (CMS) sequencing chemistry[2,3], alongside three fully retrained, platform-specific Sentieon DNAscope machine learning models: DNAscope Pangenome for graph-reference WGS, linear-reference DNAscope for legacy WGS workflows, and a dedicated DNAscope WES model for targeted clinical sequencing[4,5].

CMS chemistry biochemically mitigates sequencing failure in challenging loci (G-quadruplexes, hairpin secondary structures, segmental duplications) to deliver uniform genome-wide coverage and Q50 base calling accuracy. Complementing this hardware improvement, Sentieon' s custom gradient boosting decision tree (GBDT) models are trained exclusively on SURFSeq 5000 sequencing data to recognize platform-unique error profiles, eliminating systematic variant calling artifacts observed when generic models are applied to third-party sequencing data.

Comprehensive benchmarking against standardized reference datasets—including GIAB v4.2.1, GIAB v5.0 telomere-to-telomere assemblies, and the Chinese Quartet population reference panel—confirms the integrated GeneMind-Sentieon workflow outperforms industry-standard Illumina DRAGEN, MGI DeepVariant, and Illumina GATK pipelines across all variant classes. Unique integration of the 1000 Chinese Pangenome (1KCP) dataset during model training delivers population-specific accuracy gains for East Asian cohorts, with marked improvements in Indel detection compared to European-reference-derived analytical tools[6,7]. This co-optimized sequencing and computational stack delivers consistent, cost-effective, high-fidelity variant detection scalable to large population biobanks and routine clinical diagnostic sequencing.

# ◎ Introduction

## 1.1 Core Limitations of Conventional Short-Read Genomics

Whole-genome and whole-exome short-read sequencing have revolutionized hereditary disease risk assessment, clinical rare disease diagnosis, and human evolutionary population analysis over the past decade. Standard short-read platforms rely on canonical sequencing-by-synthesis chemistries optimized for ideal linear B-form DNA templates, yet a significant fraction of the human genome contains structurally complex loci that deviate from this conformation, including GC-rich folded domains, G-quadruplexes, hairpin loops, and segmental duplications. These non-canonical DNA structures create three pervasive technical bottlenecks for genomic analysis:

**Biochemical sequencing failure**

Secondary DNA structures stall polymerase progression, reduce fluorescent signal uniformity, and generate localized coverage dropout, creating blind spots for variant detection.

**Reference alignment bias**

Linear reference assemblies such as GRCh38/GRCh37 represent limited haplotype diversity and can introduce reference bias, ambiguous read mapping, and reduced variant detection sensitivity, particularly for populations underrepresented in genomic reference datasets..

**Mismatched analytical modeling**

Every sequencing platform generates unique signal noise and error signal. Variant calling machine learning models trained on data from one sequencing instrument cannot optimally discriminate true genetic variation from platform-specific technical artifacts when applied to data from alternative hardware systems.

As precision medicine and large-scale population genomics demand exhaustive, error-minimized genome characterization, incremental improvements to standalone sequencing hardware or generic bioinformatic pipelines are no longer sufficient. Concurrent advancement of sequencing biochemistry and platform-tailored computational analysis is required to resolve complex genomic loci and eliminate population-specific reference bias.

## 1.2 Integrated Solution: GeneMind SURFSeq 5000 CMS Chemistry + Sentieon Custom DNAscope Models

GeneMind Biosciences engineered the SURFSeq 5000 next-generation sequencing platform, powered by proprietary CMS sequencing chemistry, to directly address biochemical barriers in structurally challenging genomic regions. Unlike traditional sequencing reagents, CMS chemistry optimizes template denaturation efficiency, polymerase processivity, and fluorophore signal stability to resolve folded non-B DNA templates, delivering uniform genome-wide coverage and ultra-high Q50 base calling accuracy across hard-to-sequence loci. The platform matches the performance benchmarks of mainstream Illumina sequencing systems while delivering reduced operational costs, supporting both large population cohort WGS and targeted clinical WES workflows.

To translate high-quality SURFSeq 5000 raw sequencing data into clinically actionable variant calls, Sentieon developed three fully retrained DNAscope machine learning pipelines optimized exclusively for GeneMind sequencing output:

**DNAscope Pangenome (WGS)**

Graph-based alignment utilizing human pangenome reference data to mitigate linear reference bias, optimized for comprehensive SNP, Indel, SV, and CNV detection.

**Linear DNAscope (WGS)**

GRCh38/GRCh37 compatible pipeline for legacy genomic analysis workflows without pangenome infrastructure.

**DNAscope WES**

Specialized model calibrated for variable exome capture sequencing depths (50–300×), tuned for clinical targeted variant screening.

All three models leverage Gradient Boosting Decision Tree (GBDT) architectures trained on SURFSeq 5000 reference genome data, with dedicated machine learning classifiers to distinguish true germline variation from CMS platform-specific sequencing noise. This work presents rigorous multi-standard benchmarking of the combined GeneMind-Sentieon workflow, validating synergistic gains in sequencing biochemistry and AI-powered variant calling for complex and population-specific genomic analysis.

## 1.3 Pangenome and Population Genomics Context

For two decades, linear haploid human reference assemblies have formed the foundation of genomic variant calling, but their single consensus sequence introduces systemic alignment bias when analyzing genetically diverse global populations. Graph-based human pangenome resources—including the Human Pangenome Reference Consortium (HPRC) draft assembly and the 1000 Chinese Pangenome (1KCP)—integrate thousands of population-specific haplotypes to eliminate reference bias[8]. The Sentieon DNAscope Pangenome pipeline constructs personalized graph references for each sample, deploying selective optimized alignment to improve read mapping across complex breakpoints, then back-converts graph alignments to standard GRCh38 coordinates to maintain compatibility with downstream clinical analysis tools.

Notably, most existing sequencing and variant calling models are trained primarily on European genomic data, leading to reduced accuracy for East Asian cohorts. This study incorporates 1KCP population data during model training and benchmarking against the Chinese Quartet reference panel to quantify population-specific accuracy improvements for Chinese and East Asian genomic research and clinical diagnostics.

# ◉ Study Overview

## 2.1 Reference Sample Panel Selection

To generate unbiased, standardized performance metrics, the integrated GeneMind-Sentieon workflow was validated against two tiers of reference datasets:

**1.Global GIAB Standard Genomes**
HG001–HG005 cell lines with authoritative, high-confidence variant truth sets curated by the National Institute of Standards and Technology (NIST). These samples enable direct comparison to industry-standard sequencing and analytical pipelines across both GIAB v4.2.1 (high-confidence accessible genomic regions) and GIAB v5.0 (telomere-to-telomere assemblies including complex repetitive loci).

**2.Chinese Population Reference Panel**
The Chinese Quartet haplotype-resolved assembly dataset, a gold standard benchmark for East Asian population genomics, supplemented with information from 1000 Chinese Pangenome (1KCP) data to quantify ancestry-specific variant calling performance.

All sequencing experiments included both PCR-amplified and PCR-free library preparation protocols for WGS and WES to evaluate workflow robustness across common library construction methods.

## 2.2 DNAscope Model Training Workflow

Model training was executed using Sentieon CLI v1.6.3 and supporting software suite v202503.03. Training input data consisted of SURFSeq 5000 WGS reads generated from HG001, HG002, HG004, and HG005 reference genomes, processed with a stochastic data partitioning strategy:
- 20% of sequencing reads reserved for internal cross-validation during model training;
- Chromosome 20 fully isolated as an independent held-out test dataset to avoid overfitting.

Upstream read alignment was performed via two Sentieon alignment engines: Pan-align (for pangenome workflows) and BWA-Turbo (linear reference workflows), with full quality control filtering applied to all output BAM alignment files.

To simulate the full range of sequencing depths encountered in routine research and clinical labs, aligned BAM files were computationally downsampled to replicate standard coverage regimes:
- WGS: 15× to 40× uniform sequencing depth;
- WES: 50× to 300× targeted exome coverage.

Ultra-sensitive raw variant candidate outputs from the base DNAscope algorithm were used to train the GBDT classification model, which learns to separate true germline variants from platform-specific sequencing artifacts. Final variant calling accuracy metrics (precision, recall, F1-score, false positive/negative counts) were calculated via Aardvark comparison against GIAB v4.2.1 and v5.0 gold-standard VCF datasets; CNV detection performance was quantified using validated integrated alignment-variant calling methodology published in prior Sentieon research[9].

## 2.3 Benchmark Comparison Controls

Direct head-to-head comparative benchmarking was performed against three mainstream industry sequencing and analytical stacks:
- Illumina NovaSeq paired with DRAGEN v4.5 or Sentieon DNAscope;
- MGI T7 sequencer paired with DeepVariant or Sentieon DNAscope;
- Illumina sequencing paired with GATK (standard legacy clinical pipeline).

# ◎ Results

## 3.1 DNAscope Pangenome WGS Performance

### 3.1.1 Benchmarking on GIAB v4.2.1 (Accessible High-Confidence Genomic Regions)

GIAB v4.2.1 remains the universal industry standard for validating germline variant calling, restricted to well-mapped, low-complexity genomic intervals. Benchmarking at standard 30× WGS depth demonstrated the co-optimized GeneMind-Sentieon Pangenome pipeline delivered total variant error counts of approximately 6,000, comparable to Illumina NovaSeq data processed with DRAGEN v4.5 or native Sentieon DNAscope. Critically, competing workflows (MGI T7 + DeepVariant) exhibited substantially higher cumulative false-positive and false-negative variant errors. Direct error quantification confirmed the GeneMind custom pangenome model reduced total variant errors relative to both Illumina DRAGEN and MGI' s recommended analytical stack.

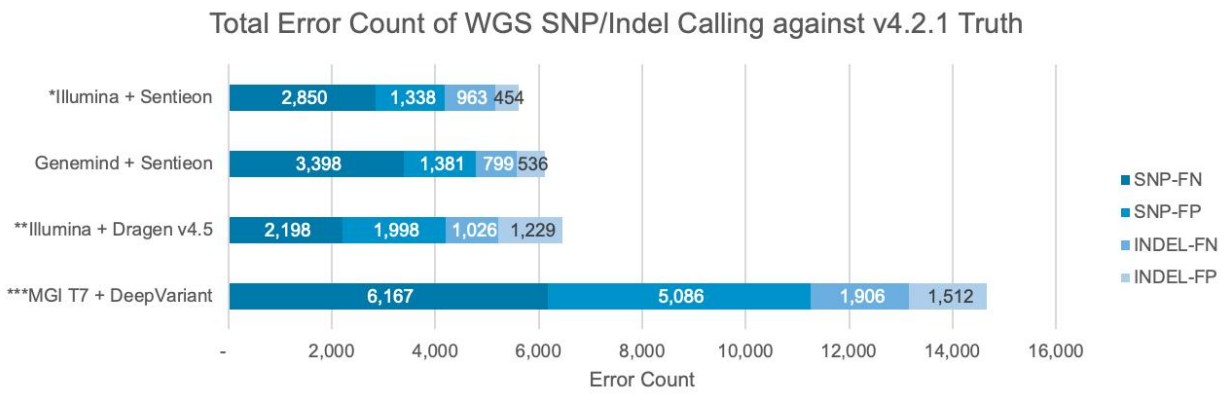


**Figure 1. Total germline variant error count comparison for HG002 30× WGS using GeneMind-trained DNAscope Pangenome models, benchmarked against GIAB v4.2.1 high-confidence accessible genomic regions.** The integrated GeneMind-Sentieon workflow exhibits lower cumulative false-positive and false-negative variant errors relative to Illumina NovaSeq + DRAGEN v4.5 and MGI T7 + DeepVariant alternative pipelines. Source:

*https://www.sentieon.com/news/sentieon-2026q2-update-datasheet;
**https://developer.illumina.com/news-updates/dragen-v4-5-illuminating-the-dark-genome-with-advanced-variant-detection;
***https://www.biorxiv.org/content/10.64898/2026.04.22.720142v5.full.pdf

The performance advantage originates from two synergistic layers: CMS sequencing chemistry eliminates coverage bias in local genomic intervals to generate more complete, uniform read data, while the platform-trained pangenome model distinguishes SURFSeq 5000-specific noise from genuine genetic variation more effectively than generic ML models trained on alternate sequencing hardware.

### 3.1.2 Rigorous Validation with GIAB v5.0 (T2T Complex Genomic Loci)

GIAB v5.0 incorporates complete telomere-to-telomere genome assemblies and expands gold-standard truth sets to segmental duplications, repetitive sequences, and medically relevant hard-to-sequence genes—representing a far more stringent test of sequencing chemistry and variant calling robustness. Under this high-stringency benchmark, the GeneMind-Sentieon workflow maintained consistent performance with a total variant error count of ~42,000, outperforming all three comparative industry pipelines tested.

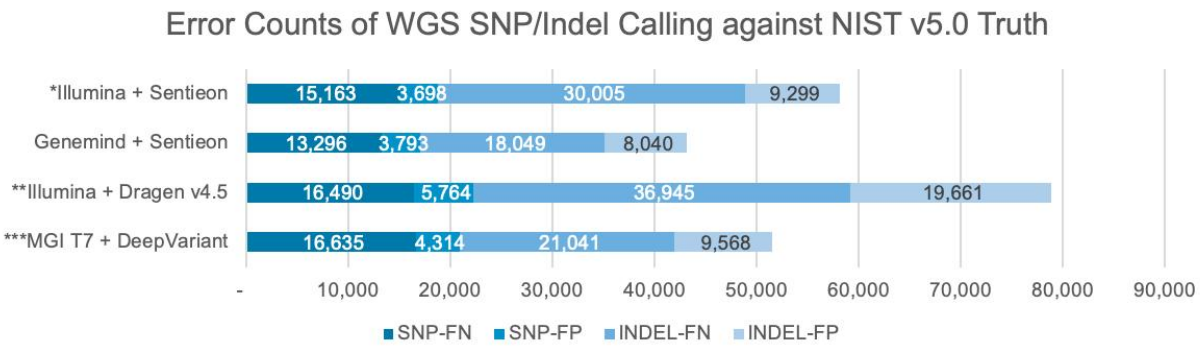



**Figure 2. Total germline variant error count comparison for HG002 30× WGS using GeneMind-trained DNAscope Pangenome models, benchmarked against GIAB v5.0 telomere-to-telomere gold-standard assemblies encompassing complex repetitive, duplicated, and non-B DNA loci.** The GeneMind-Sentieon workflow achieves the lowest cumulative variant error count across three competing industry sequencing and analytical stacks. Source:

*https://www.sentieon.com/news/sentieon-2026q2-update-datasheet;
**https://developer.illumina.com/news-updates/dragen-v4-5-illuminating-the-dark-genome-with-advanced-variant-detection;
***https://www.biorxiv.org/content/10.64898/2026.04.22.720142v5.full.pdf

CMS chemistry drives this performance gain by resolving physical barriers to sequencing non-B DNA structures: improved polymerase kinetics and stabilized fluorescent signal readouts eliminate localized coverage dropout in complex loci, removing a primary source of alignment ambiguity and variant calling failure prior to computational analysis. The uniform, high-quality raw read data generated by SURFSeq 5000 provides a superior input foundation for the pangenome graph alignment and machine learning variant classification steps.

## 3.2 Population-Specific Performance: Chinese Quartet Reference Panel

Nearly all commercial sequencing and variant calling models are trained on European-derived genomic datasets, leading to measurable accuracy loss for East Asian cohorts, particularly for small Indel variants that drive many population-specific clinical phenotypes. To address this gap, the GeneMind DNAscope Pangenome model integrated information derived from the 1000 Chinese Pangenome (1KCP) project during training.

Benchmarking against the Chinese Quartet reference panel confirmed substantial accuracy improvements relative to the Illumina + GATK gold-standard pipeline for East Asian genomics:

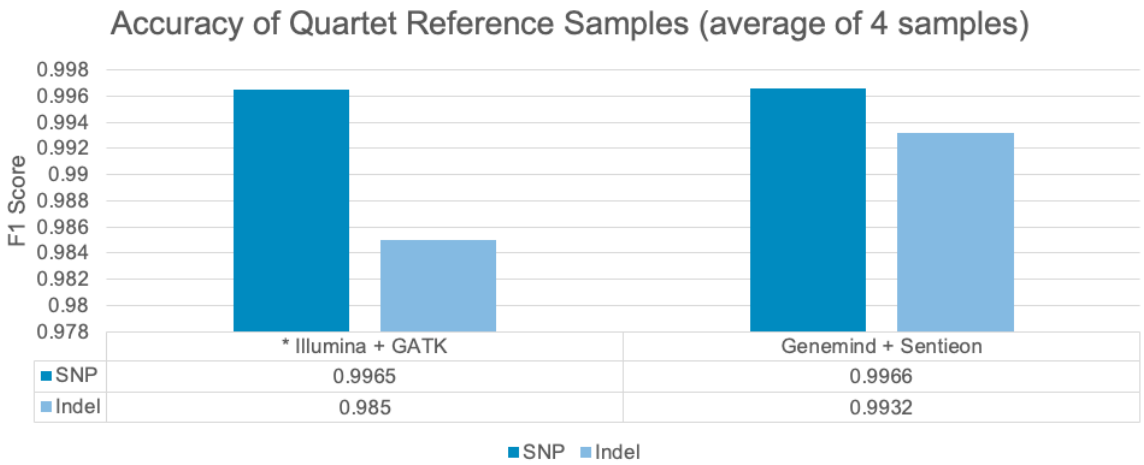



**Figure 3. Average SNP and Indel F1-score comparison across 30x sequencing dataset from four reference samples (D5, D6,F7, ) of the Chinese Quartet population reference panel, comparing GeneMind SURFSeq 5000 + Sentieon DNAscope Pangenome to the published Illumina sequencing + GATK pipeline baseline.** Marked Indel detection performance gains are observed for the GeneMind integrated workflow, driven by 1000 Chinese Pangenome model training integration. Source:

*https://link.springer.com/article/10.1186/s13059-023-03109-2#additional-information

- SNP F1-score: GeneMind + Sentieon = 0.9966 vs. Illumina + GATK = 0.9965;
- Indel F1-score: GeneMind + Sentieon = 0.9932 vs. Illumina + GATK = 0.9850.

The dominant performance gain is observed for Indel detection, the variant class most susceptible to reference bias and uneven coverage in population-specific complex loci. This result validates the utility of integrating regional population pangenome data during platform-specific model training for clinical and research applications focused on Chinese cohorts.

## 3.3 Structural Variant (SV) and Copy Number Variant (CNV) Detection

Beyond small SNP and Indel variation, comprehensive clinical genome analysis requires reliable detection of large-scale structural rearrangements and copy number alterations, which frequently occur at complex genomic breakpoints with poor mapping performance on linear reference assemblies. The DNAscope Pangenome workflow constructs personalized graph references for each sample, resolving ambiguous read alignment at SV breakpoints and copy-variable loci.

Benchmark SV detection metrics comparing GeneMind-Sentieon to Illumina DRAGEN v4.5.

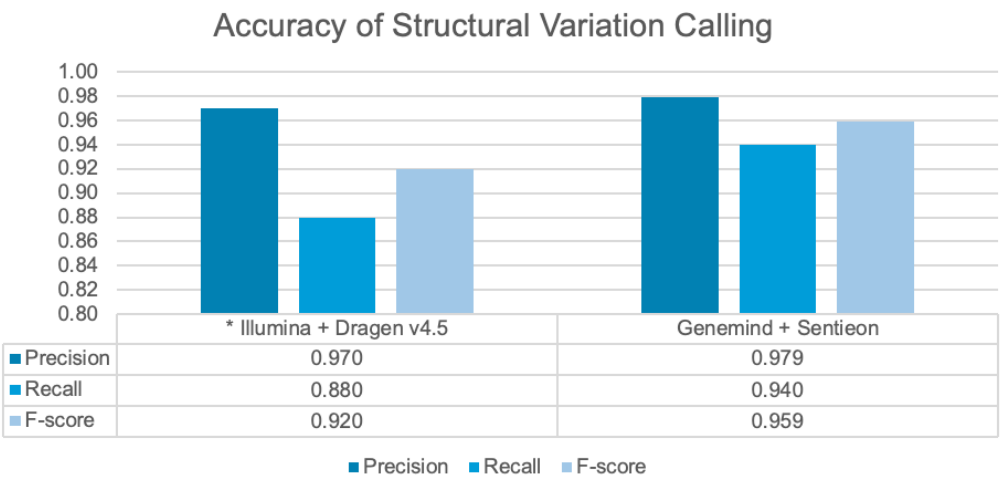



**Figure 4. Structural variant (SV) detection precision, recall, and overall F1-score comparison between 30x WGS dataset generated and analyzed by Illumina NovaSeq + DRAGEN v4.5 and GeneMind SURFSeq 5000 + Sentieon DNAscope Pangenome WGS.** The GeneMind workflow delivers simultaneous improvements in precision and recall for large structural rearrangements via CMS uniform coverage and graph-based breakpoint alignment. Source:

* https://help.connected.illumina.com/dragen/dragen-v4.5/reference/release-notes-readme/dragen-crn-v4.5.4;

The integrated workflow delivers simultaneous improvements in precision and recall for structural variants, driven by uniform CMS coverage across duplicated and breakpoint regions and pangenome graph alignment resolving ambiguous split reads. CNV detection performance was similarly elevated across both copy-loss and copy-gain events relative to Illumina DRAGEN benchmarks, with consistent reduction in false copy number calls in segmentally duplicated gene families.

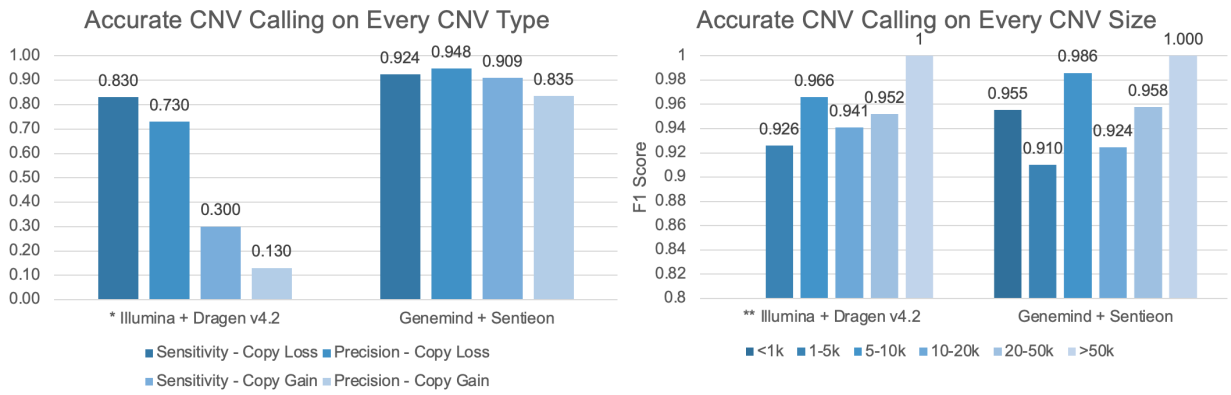



**Figure 5. Copy number variant (CNV) detection accuracy comparison for GeneMind SURFSeq 5000 WGS processed with DNAscope Pangenome relative to Illumina DRAGEN v4.2 benchmark data, evaluating both copy-loss and copy-gain events across segmentally duplicated gene loci.** Source:

* https://pmc.ncbi.nlm.nih.gov/articles/PMC12005901/pdf/vbaf071.pdf;
** https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/dragen-v4-accuracy-app-note-m-gl-01016/dragen-v4-accuracy-app-note-m-gl-01016.pdf;

## 3.4 Linear Reference DNAscope WGS (Legacy GRCh38/GRCh37 Workflows)

To support existing genomic analysis pipelines reliant on linear human reference assemblies without pangenome infrastructure, a dedicated linear-reference DNAscope model was trained exclusively on SURFSeq 5000 sequencing data. Cross-platform benchmarking against MGI T7 and Illumina NovaSeq sequencing data processed with identical Sentieon linear DNAscope parameters demonstrated the GeneMind workflow produced the lowest cumulative variant error counts:
- SNP False Negatives (FN): 18,397
- SNP False Positives (FP): 3,104
- Indel FN: 1,645
- Indel FP: 1,022

The GeneMind platform generates the lowest total variant error counts for legacy linear-reference genomic analysis.

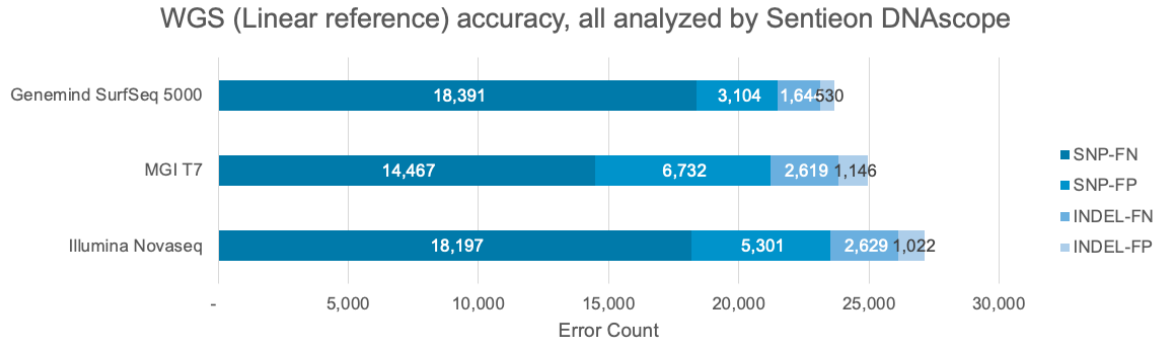
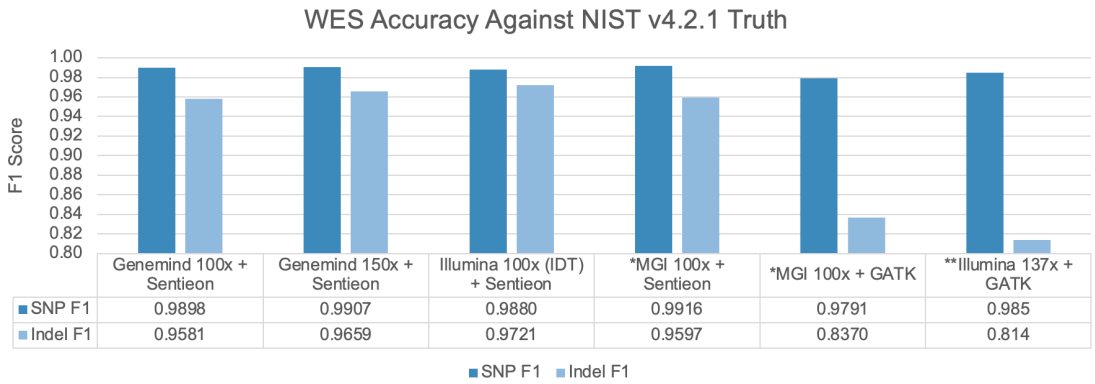



**Figure 6. Cumulative SNP and Indel false positive/negative error counts for linear-reference GRCh38 WGS workflows, comparing 30x WGS datasets from GeneMind SURFSeq 5000 to MGI T7 and Illumina NovaSeq sequencing platforms processed with Sentieon linear DNAscope corresponding models..** The GeneMind platform generates the lowest total variant error counts for legacy linear-reference genomic analysis.

Even when restricted to conventional linear reference alignment, the uniform coverage and reduced systematic sequencing bias from CMS chemistry reduce baseline technical noise, while the platform-specific machine learning model more efficiently filters sequencing artifacts to minimize false variant calls compared to generic models trained on alternate sequencing platforms. These results confirm SURFSeq 5000 delivers superior raw sequencing data quality independent of pangenome graph alignment infrastructure.

## 3.5 DNAscope WES Performance for Clinical Targeted Sequencing

Whole-exome sequencing is the primary assay for routine clinical rare disease diagnostics, requiring robust variant detection across capture depths ranging from 50× to 300×. A specialized GeneMind WES DNAscope model was trained and benchmarked across all standard clinical coverage depths against competing sequencing and analytical stacks (Figure 7).




| | Genemind 100x + Sentieon | Genemind 150x + Sentieon | Illumina 100x (IDT) + Sentieon | *MGI 100x + Sentieon | *MGI 100x + GATK | **Illumina 137x + GATK |
|---|---|---|---|---|---|---|
| SNP F1 | 0.9898 | 0.9907 | 0.9880 | 0.9916 | 0.9791 | 0.985 |
| Indel F1 | 0.9581 | 0.9659 | 0.9721 | 0.9597 | 0.8370 | 0.814 |

**Figure 7. SNV and combined Indel F1-score performance across clinical WES sequencing depths (100×, 150×) for GeneMind SURFSeq 5000 + custom Sentieon WES DNAscope, compared to Illumina (IDT), MGI, and Illumina GATK control pipelines benchmarked against GIAB v4.2.1 exonic truth sets.** The GeneMind workflow delivers consistent F1-score advantages across all tested capture depths, with substantially higher Indel detection accuracy than standard GATK clinical pipelines. Source:

*https://www.sentieon.com/news/sentieon-has-released-the-dnascope-mgi-wes-model-assist-complete-genomics-to-expand-sequencing-market/;
** https://www.salus-bio.com/uploadfiles/2025/02/20250212165046962.pdf;

**Key quantitative findings:**

- Consistent cross-depth superiority: The GeneMind custom WES model outperforms Illumina-matched Sentieon WES models at every tested coverage tier (50×, 100×, 150×, 200×, 300×) for both SNV and combined Indel detection.
- Progressive F1-score gains with increasing coverage: SNV F1-score advantage widens from +0.05% at 50× depth to +0.09% at 300× depth; combined Indel F1-score improvements range from +0.28% to +0.36%, representing a 3–4x larger performance gain than SNVs.
- Quantifiable error reduction: Per sequencing replicate, the GeneMind workflow reduces SNV errors by 49–84 and Indel errors by 20–26, driven by balanced improvements in recall (primary for SNVs) and precision (primary for Indels).
- Superiority to legacy GATK pipelines: At standard clinical depths (100×, 150×), GeneMind WES + DNAscope matches top-tier competing sequencing platforms paired with Sentieon, and substantially outperforms Illumina and MGI data processed with standard GATK variant calling workflows.

These results establish the integrated GeneMind CMS sequencing + custom DNAscope WES pipeline as a high-performance, cost-effective alternative for clinical exome diagnostic sequencing, with consistent error reduction across the full range of standard clinical capture depths.

# ◎ Discussion

This work establishes a clear causal chain linking CMS sequencing biochemistry, platform-tailored machine learning, and pangenome graph alignment to measurable improvements in germline variant detection across all variant classes and genomic contexts:

1. CMS chemistry resolves fundamental sequencing limitations: By optimizing denaturation, polymerase kinetics, and fluorescent signal stability, CMS eliminates coverage dropout and systematic bias in non-B DNA loci, generating uniform, high-fidelity raw sequencing reads that form a higher-quality input for all downstream computational analysis. This biochemical improvement reduces baseline false-positive and false-negative calls before alignment or variant classification.

2. Platform-specific ML models eliminate cross-platform modeling mismatch: All sequencing hardware generates unique error signatures; generic variant calling models trained on Illumina or MGI data cannot optimally separate true variation from SURFSeq 5000 technical noise. Retraining Sentieon DNAscope GBDT classifiers exclusively on GeneMind sequencing data maximizes precision and recall for SNPs, Indels, SVs, and CNVs.

3. Pangenome graph alignment mitigates linear reference bias: Conventional single-haplotype reference assemblies create mapping ambiguity for population-specific haplotypes. The DNAscope Pangenome pipeline integrates graph-based personalized references to improve read mapping at complex breakpoints, with back-conversion to GRCh38 coordinates preserving compatibility with standard clinical interpretation tools.

4. 1KCP integration addresses East Asian population bias: The majority of genomic analytical models are trained on European reference data, reducing diagnostic accuracy for Chinese cohorts. Integration information derived from the 1000 Chinese Pangenome during model training delivers targeted Indel detection improvements validated on the Chinese Quartet reference panel, filling a critical gap for Asian population genomics research and clinical testing.

Collectively, these layers of co-optimization differentiate the GeneMind-Sentieon workflow from standalone sequencing instruments or generic bioinformatic pipelines. Performance gains are not limited to well-mapped, low-complexity genomic regions—consistent superiority is observed on the stringent GIAB v5.0 telomere-to-telomere benchmark, which captures repetitive, duplicated, and medically critical hard-to-sequence loci.

For end-users, the integrated workflow delivers two practical advantages beyond raw accuracy metrics:
1. Cost efficiency: SURFSeq 5000 delivers performance parity with Illumina sequencing systems at reduced operational cost, scalable for large population biobank WGS and high-volume clinical WES testing.
2. Dual workflow compatibility: Parallel optimized pipelines for pangenome WGS and legacy linear-reference WGS enable seamless adoption across labs with or without graph-based reference infrastructure.

# Summary

This comprehensive benchmarking study validates three Sentieon DNAscope machine learning models purpose-built for GeneMind' s SURFSeq 5000 sequencing platform powered by CMS proprietary sequencing chemistry: DNAscope Pangenome for graph-reference whole-genome sequencing, linear-reference DNAscope for legacy WGS analysis, and a dedicated DNAscope WES model for clinical targeted exome sequencing. The end-to-end integrated sequencing and computational workflow delivers industry-leading germline variant detection accuracy across SNPs, Indels, structural variants, and copy number variants, with consistent performance advantages in structurally complex genomic regions inaccessible to conventional short-read pipelines.

Synergistic gains originate from two core innovations: CMS sequencing chemistry biochemically resolves non-B DNA sequencing barriers to generate uniform, high-quality raw read data, while platform-retrained Sentieon machine learning models precisely distinguish true genetic variation from SURFSeq 5000-specific sequencing noise. Integration of the 1000 Chinese Pangenome dataset during model training delivers unique population-specific accuracy improvements validated on the Chinese Quartet reference panel, addressing reference bias limitations prevalent in European-derived analytical stacks.

Benchmarking against GIAB v4.2.1, GIAB v5.0 telomere-to-telomere assemblies, and cross-platform comparative industry workflows confirms the GeneMind-Sentieon pipeline outperforms Illumina DRAGEN, MGI DeepVariant, and Illumina GATK solutions across all standard WGS and WES sequencing depths. This co-optimized hardware and software stack provides a scalable, reliable, cost-effective framework for high-accuracy human genomics, suitable for both large-scale population research and routine clinical genetic diagnostic applications.

# References


1. Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics, 17(5), 333–351. https://doi.org/10.1038/nrg.2016.49
2. Sun Y, Zhao X, Fan X, Wang M, Li C, Liu Y, Wu P, Yan Q and Sun L (2024) Assessing the impact of sequencing platforms and analytical pipelines on whole-exome sequencing. Frontiers in Genetics. 15:1334075. doi: 10.3389/fgene.2024.1334075
3. Li Q, Liu L, Lin Q, et al. CMS: Achieving Uniform and High-Quality Sequencing across Challenging Non-canonical Genomic Regions. bioRxiv. 2026. doi: https://doi.org/10.64898/2026.04.24.720553
4. Freed, D., Pan, R., Chen, H., et al. DNAscope: High accuracy small variant calling using machine learning. bioRxiv. https://doi.org/10.1101/2022.05.20.492556
5. Sentieon. Pangenome analysis without the pain: fast, accurate and affordable germline variant calling. Medium. https://medium.com/@frank.hu_11452/pangenome-analysis-without-the-pain-fast-accurate-and-affordable-germline-variant-calling-with-224c69005f32
6. Jia P, Dong L, Yang X, et al. Haplotype-resolved assemblies and variant benchmark of a Chinese Quartet. Genome Biology. 2023 Dec 4;24(1):277. doi: 10.1186/s13059-023-03116-3. PMID: 38049885; PMCID: PMC10694985.
7. Wang, Y., Duan, Z., Chen, D. et al. The 1000 Chinese Pangenome empowers medical and population genetics. Nature 654, 121–130 (2026). https://doi.org/10.1038/s41586-026-10315-y
8. Liao, WW., Asri, M., Ebler, J. et al. A draft human pangenome reference. Nature 617, 312–324 (2023). https://doi.org/10.1038/s41586-023-05896-x
9. Hu J, Freed D, Feng H, et al. A novel and accelerated method for integrated alignment and variant calling from short and long reads. Frontiers in Bioinformatics. 5:1691056. doi: 10.3389/fbinf.2025.1691056



GeneMind

**GeneMind Biosciences Co., Ltd.**

Technical Support: **+86-400-822-3660**

Website: en.genemind.com Email: info@genemind.com

Address: Room 502A/502B/602, Luohu Investment Holding Building, 112 Qingshuihe 1st Road, Luohu District, Shenzhen, Guangdong, China