# State of the Art NGS Variant Callers in Clinical Practice: A Comprehensive Review

**K-Dense Web**
contact@k-dense.ai
k-dense.ai

December 24, 2025

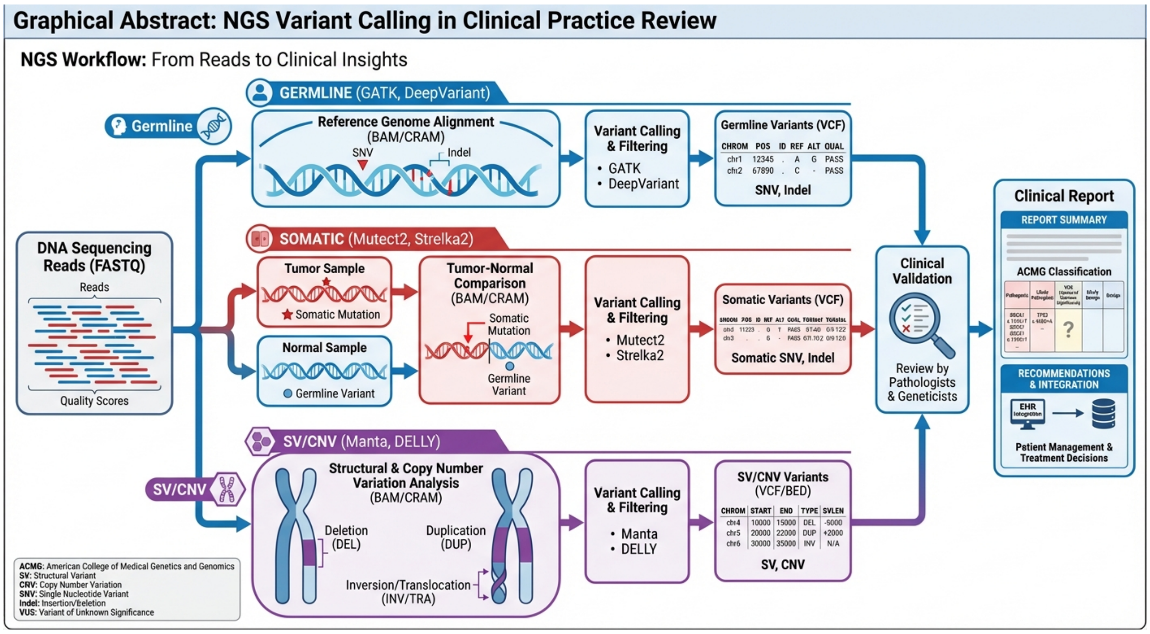


Figure 1: **Graphical Abstract.** Overview of the NGS variant calling ecosystem in clinical practice. The workflow illustrates the three primary variant calling domains—germline, somatic, and structural/copy number variants—each requiring specialized bioinformatics tools and validation strategies. All pathways converge through ACMG/AMP classification frameworks before integration into clinical reporting and electronic health record systems.

### Abstract

Next-generation sequencing (NGS) has transformed clinical diagnostics, enabling comprehensive detection of genetic variants underlying inherited disorders and cancer. Central to the clinical utility of NGS is the accuracy and reliability of variant calling algorithms, which translate raw sequencing data into actionable genetic findings. This review provides a comprehensive examination of the current state of the art in NGS variant calling for clinical applications, addressing germline variant detection for constitutional disorders, somatic variant identification in oncology, and the emerging challenge of structural variant and copy number variation analysis. We evaluate leading variant calling tools including GATK HaplotypeCaller, DeepVariant, Mutect2, Strelka2, Manta, and DELLY, analyzing their performance characteristics, clinical sensitivity and specificity, and suitability for different clinical contexts. Special attention is given to the integration of artificial intelligence and deep learning approaches, which have demonstrated significant improvements in variant calling accuracy. We synthesize current regulatory frameworks including ACMG/AMP guidelines and CAP/CLIA requirements for clinical validation, providing practical guidance for laboratory implementation. Technical challenges including low variant allele frequency detection, tumor heterogeneity management, and complex genomic region analysis are discussed alongside emerging solutions. Finally, we address practical considerations for clinical pipeline validation, reproducibility assurance, and laboratory information system integration. This review serves as a resource for clinical geneticists, molecular pathologists, and laboratory directors seeking to implement or optimize NGS variant calling workflows in clinical diagnostic settings.

**Keywords:** next-generation sequencing, variant calling, clinical genomics, DeepVariant, GATK, somatic mutations, structural variants, ACMG guidelines, precision medicine

## Contents

# 1 Introduction

The advent of next-generation sequencing (NGS) technologies has fundamentally transformed the landscape of clinical molecular diagnostics, enabling unprecedented access to comprehensive genomic information for patient care [Goodwin et al., 2016, Shendure et al., 2017]. Over the past decade, NGS has evolved from a research tool to a cornerstone of precision medicine, with applications spanning inherited disease diagnosis, pharmacogenomics, oncology, and infectious disease management [Mardis, 2019]. The democratization of genomic sequencing, driven by dramatic reductions in cost and turnaround time, has made whole-exome sequencing (WES) and targeted gene panel testing routine components of clinical care in many healthcare systems worldwide.

At the heart of any NGS-based clinical assay lies the bioinformatics pipeline responsible for transforming raw sequencing reads into clinically interpretable variant calls. Variant calling—the computational process of identifying positions in the genome where an individual differs from the reference sequence—represents a critical bottleneck in the translation of sequencing data to clinical action [Yohe and Thyagarajan, 2017]. The accuracy, sensitivity, and specificity of variant calling directly impact diagnostic yield, with false negatives potentially resulting in missed diagnoses and false positives leading to unnecessary clinical interventions or patient anxiety. Given these high stakes, the selection and validation of variant calling algorithms for clinical use demands rigorous attention to analytical performance characteristics and regulatory compliance.

The variant calling landscape has undergone substantial evolution since the introduction of early statistical methods. Traditional approaches, exemplified by the Genome Analysis Toolkit (GATK) developed at the Broad Institute, employ probabilistic models incorporating base quality scores, mapping quality, and local sequence context to distinguish true variants from sequencing artifacts [McKenna et al., 2010, DePristo et al., 2011]. These methods have established strong track records in clinical laboratories, with GATK HaplotypeCaller achieving F-scores exceeding 0.99 in benchmark evaluations for germline single nucleotide variants (SNVs) and small insertions/deletions (indels) [Krusche et al., 2019]. However, the emergence of deep learning-based variant callers, most notably Google's DeepVariant, has challenged the supremacy of classical statistical methods [Poplin et al., 2018]. By reformulating variant calling as an image classification problem, where pileup visualizations of aligned reads are processed through convolutional neural networks, DeepVariant has demonstrated superior performance in multiple benchmarking challenges, including the FDA-sponsored precisionFDA Truth Challenge [Olson et al., 2022].

The clinical variant calling ecosystem must address fundamentally different biological and technical challenges depending on the application domain. Germline variant calling, focused on identifying inherited constitutional variants, benefits from the expectation that true variants will be present at approximately 50% (heterozygous) or 100% (homozygous) variant allele frequency (VAF). In contrast, somatic variant calling in oncology must contend with tumor heterogeneity, variable tumor purity, and the detection of subclonal mutations present at VAFs as low as 1-5% or even lower in liquid biopsy applications [Cibulskis et al., 2013, Wan et al., 2017]. Structural variant (SV) detection, including large deletions, duplications, inversions, and translocations, presents yet another set of challenges, requiring algorithms capable of integrating multiple signals including discordant read pairs, split reads, and read depth changes [Chen et al., 2016, Rausch et al., 2012].

Regulatory and professional standards provide essential frameworks for clinical implementation of NGS variant calling. The American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) have established joint guidelines for variant interpretation that have become the de facto standard in clinical genetics laboratories [Richards et al., 2015]. Complementary standards for somatic variant interpretation have been developed through collaboration between AMP, the American Society of Clinical Oncology (ASCO), and the College of American Pathologists (CAP) [Li et al., 2017]. The AMP and CAP have also published joint recommendations for validating NGS bioinformatics pipelines, providing a roadmap for clinical laboratories seeking to establish robust, reproducible workflows [Roy et al., 2018].

This review aims to provide clinical geneticists, molecular pathologists, and laboratory directors with a comprehensive overview of the current state of the art in NGS variant calling for clinical practice. We examine the leading tools for germline, somatic, and structural variant detection, evaluating their performance characteristics and clinical applicability. We synthesize current regulatory requirements and validation frameworks, addressing practical considerations for pipeline implementation. Finally, we discuss emerging trends including the integration of artificial intelligence, long-read sequencing technologies, and the challenges of laboratory information system integration in an era of increasing genomic data complexity.

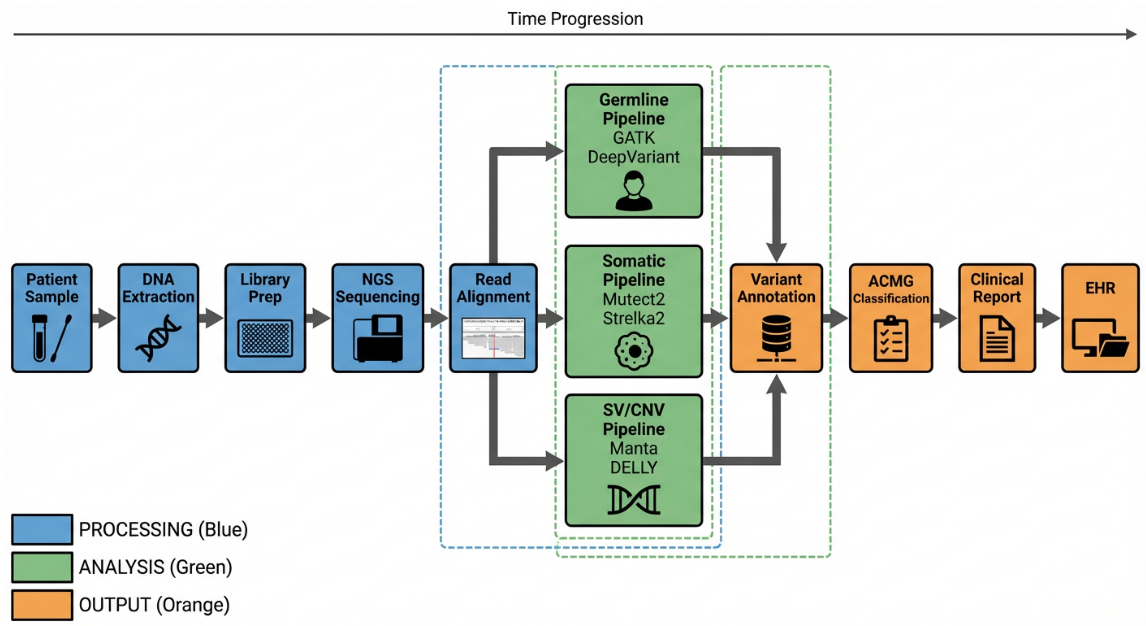


Figure 2: **Clinical NGS variant calling workflow.** Comprehensive overview of the clinical NGS pipeline from sample collection through final clinical reporting. The workflow encompasses wet lab processes (sample collection, DNA extraction, library preparation, sequencing) and bioinformatics analysis (read alignment, variant calling across germline, somatic, and structural variant domains, annotation, and classification). Integration with laboratory information systems (LIS) and electronic health records (EHR) enables seamless clinical reporting.

# 2 Regulatory Frameworks and Clinical Guidelines

## 2.1 ACMG/AMP Standards for Germline Variant Interpretation

The 2015 ACMG/AMP standards for interpretation of sequence variants represent a landmark consensus document that has fundamentally shaped clinical variant classification practices [Richards et al., 2015]. This framework establishes a five-tier classification system—pathogenic, likely pathogenic, uncertain significance, likely benign, and benign—based on the systematic evaluation of multiple lines of evidence. The guidelines provide 28 criteria organized into categories of evidence supporting pathogenicity (very strong, strong, moderate, supporting) and evidence supporting benign classification, with rules for combining evidence to reach final classifications.

The implementation of ACMG/AMP criteria requires integration of diverse data sources including population frequency databases such as gnomAD [Karczewski et al., 2020], computational predictions of variant impact, functional studies, segregation data in affected families, and de novo occurrence. The Clinical Genome Resource (ClinGen), a National Institutes of Health-funded initiative, has played a central role in refining these standards through the development of gene-specific and disease-specific variant curation expert panels (VCEPs) that provide detailed specifications for applying ACMG/AMP criteria to particular genes [Rehm et al., 2015]. These specifications address ambiguities in the original guidelines and provide quantitative thresholds tailored to the biology of individual genes.

ClinVar, the public archive of relationships among human genetic variants and phenotypes, serves as a critical repository for variant classifications [Landrum et al., 2018]. Clinical laboratories routinely query ClinVar during variant interpretation, and many contribute their own classifications, creating a collaborative ecosystem that continuously refines variant knowledge. However, inter-laboratory discordance in variant classification remains a significant challenge, with studies demonstrating that a substantial proportion of variants receive conflicting interpretations across submitting laboratories.

## 2.2 Somatic Variant Interpretation Standards

The interpretation of somatic variants in cancer presents distinct challenges from germline variant classification, as the clinical significance of a mutation depends on its role as a driver versus passenger event, its association with prognosis, and its potential as a therapeutic target or resistance mechanism. The 2017 AMP/ASCO/CAP joint guidelines established a four-tier system for categorizing somatic variants based on clinical significance: tier I (variants of strong clinical significance with FDA-approved therapies or inclusion in professional guidelines), tier II (variants of potential clinical significance based on clinical trial evidence or established biological function), tier III (variants of unknown clinical significance), and tier IV (benign or likely benign variants) [Li et al., 2017].

Knowledge bases such as CIViC (Clinical Interpretation of Variants in Cancer) and OncoKB provide curated, evidence-based annotations of somatic variants to support clinical interpretation [Griffith et al., 2017, Chakravarty et al., 2017]. These resources catalog the therapeutic, diagnostic, and prognostic implications of specific variants across cancer types, with evidence ratings based on the strength of supporting clinical and preclinical data. The integration of such knowledge bases into clinical workflows enables systematic, evidence-based variant interpretation while acknowledging that the rapidly evolving landscape of precision oncology requires continuous updating of variant annotations.

## 2.3 Laboratory Validation Requirements

The Clinical Laboratory Improvement Amendments (CLIA) regulations and CAP accreditation standards establish the regulatory framework for clinical laboratory testing in the United States. For NGS-based tests, the AMP and CAP joint recommendations for bioinformatics pipeline validation provide detailed guidance on establishing and documenting analytical performance characteristics [Roy et al., 2018]. These guidelines emphasize the distinction between analytical validation (demonstrating that the assay accurately detects variants within the defined reportable range) and clinical validation (establishing the relationship between test results and clinical phenotypes or outcomes).

Key performance metrics for variant calling validation include sensitivity (the proportion of true variants correctly identified), specificity (the proportion of non-variant positions correctly called as reference), positive predictive value (the proportion of called variants that are true positives), and reproducibility across runs, operators, and instruments. The Genome in a Bottle (GIAB) consortium has developed extensively characterized reference materials, including well-studied cell lines such as NA12878 (HG001) and a diversity panel spanning multiple populations, that serve as gold-standard truth sets for validation studies [Zook et al., 2014, 2019]. These reference materials, coupled with high-confidence variant call sets, enable rigorous benchmarking of variant calling pipelines across diverse genomic regions.

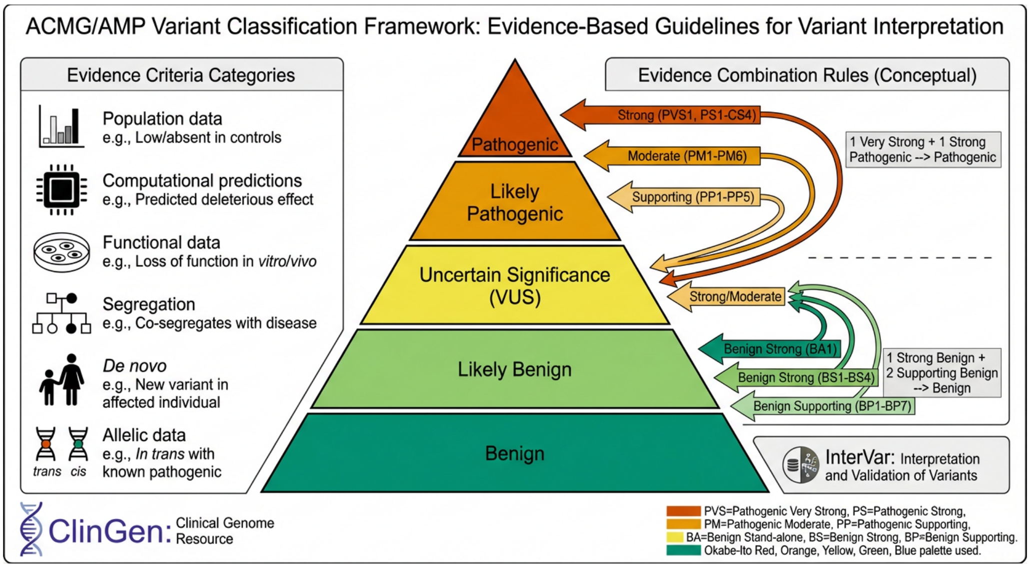


Figure 3: **ACMG/AMP variant classification framework.** The five-tier classification system established by ACMG/AMP guidelines for germline variant interpretation. Classification is determined by combining evidence criteria from multiple categories including population frequency data, computational predictions, functional studies, segregation analysis, de novo occurrence, and allelic data. The framework provides rules for combining evidence to reach final pathogenicity classifications.

# 3 Germline Variant Calling

## 3.1 Traditional Statistical Approaches

The Genome Analysis Toolkit (GATK), developed at the Broad Institute, has long represented the gold standard for germline variant calling in clinical and research applications [McKenna et al., 2010, DePristo et al., 2011]. The current GATK HaplotypeCaller employs a local re-assembly approach that constructs haplotypes from aligned reads in regions showing evidence of variation, followed by pair hidden Markov model (HMM) evaluation to calculate the likelihood of observed reads given candidate haplotypes. This approach provides robust handling of complex variants including multi-nucleotide polymorphisms and indels in close proximity.

The GATK Best Practices workflow encompasses multiple processing steps including duplicate marking, base quality score recalibration (BQSR), variant calling, and variant quality score recalibration (VQSR). VQSR applies machine learning to distinguish true variants from artifacts based on site-level annotations, though recent evidence suggests that hard filtering approaches may perform comparably or superiorly in many contexts, particularly for targeted sequencing applications. The computational requirements of the full GATK pipeline are substantial, though optimized implementations including the Intel Genomics Kernel Library (GKL) and GPU-accelerated versions have improved throughput.

Other established statistical variant callers include FreeBayes, which implements a Bayesian haplotype-based approach without local re-assembly [Garrison and Marth, 2012], and Platypus, which integrates mapping, assembly, and haplotype-based methods [Rimmer et al., 2014]. Comparative evaluations have generally demonstrated that GATK HaplotypeCaller and FreeBayes achieve similar accuracy for SNV detection, with differences in indel calling performance depending on the dataset and genomic context.

### 3.2 Deep Learning-Based Variant Callers

The introduction of DeepVariant by Google in 2017 marked a paradigm shift in variant calling methodology [Poplin et al., 2018]. DeepVariant reconceptualizes variant calling as an image classification problem: aligned reads are represented as RGB pileup images encoding base identity, base quality, and mapping quality, which are then processed through an Inception-v3 convolutional neural network to classify candidate variants as homozygous reference, heterozygous variant, or homozygous variant. This approach leverages the power of deep learning to implicitly learn complex features distinguishing true variants from artifacts without explicit feature engineering.

DeepVariant achieved recognition through its performance in the precisionFDA Truth Challenge, where it demonstrated superior accuracy compared to traditional methods [Olson et al., 2022]. Subsequent versions have extended support to long-read sequencing platforms including Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), with platform-specific models trained on appropriate training data. DeepTrio extends the DeepVariant architecture to leverage trio information (proband, mother, father), enabling improved variant calling through explicit modeling of Mendelian inheritance patterns.

Comparative studies have demonstrated that DeepVariant consistently outperforms GATK HaplotypeCaller for both SNV and indel detection, with particularly notable improvements for indel calling [Chen et al., 2022]. A study comparing the tools using trio sequencing found that DeepVariant achieved significantly lower Mendelian error rates (3.09% vs. 5.25% for GATK) and higher transition/transversion ratios, suggesting a greater proportion of true positive calls. However, GATK has demonstrated advantages in detecting rare variants, which are often of particular clinical interest in the context of rare disease diagnosis.

Additional deep learning variant callers have emerged, including Clair3, which employs a multitask learning framework supporting both short-read and long-read data [Zheng et al., 2022], and DNAscope, which combines GATK-style haplotype calling with AI-based genotyping [Li et al., 2025]. The DRAGEN platform from Illumina integrates hardware-accelerated alignment and variant calling with deep learning-enhanced accuracy, achieving performance comparable to DeepVariant with substantially reduced computational requirements [Ho et al., 2020].

### 3.3 Performance Benchmarking and Clinical Considerations

Rigorous benchmarking against gold-standard truth sets is essential for evaluating variant caller performance. The GIAB consortium's high-confidence call sets provide extensively validated reference data across multiple human genomes [Zook et al., 2019]. The precisionFDA Truth Challenge V2 extended benchmarking to challenging genomic regions including segmental duplications, tandem repeats, and regions with poor mappability, where variant calling accuracy is typically reduced [Olson et al., 2022].

Current state-of-the-art tools achieve SNP F1-scores exceeding 99.9% on benchmark datasets, while indel calling performance ranges from approximately 99% to 99.5% depending on the tool and dataset. However, it is crucial to recognize that benchmark performance on well-characterized samples may not fully reflect real-world clinical performance, where sample quality variations, sequencing artifacts, and complex variant types present additional challenges. Clinical validation using samples with confirmed variants, particularly in disease-relevant genes, remains essential.

For clinical laboratories, practical considerations beyond raw accuracy influence tool selection. GATK benefits from extensive documentation, widespread adoption, and established validation in clinical settings, while DeepVariant offers superior accuracy but requires GPU resources for optimal performance. Hybrid approaches combining multiple callers, with intersection or union strategies for variant selection, may offer advantages in sensitivity or specificity depending on clinical requirements.

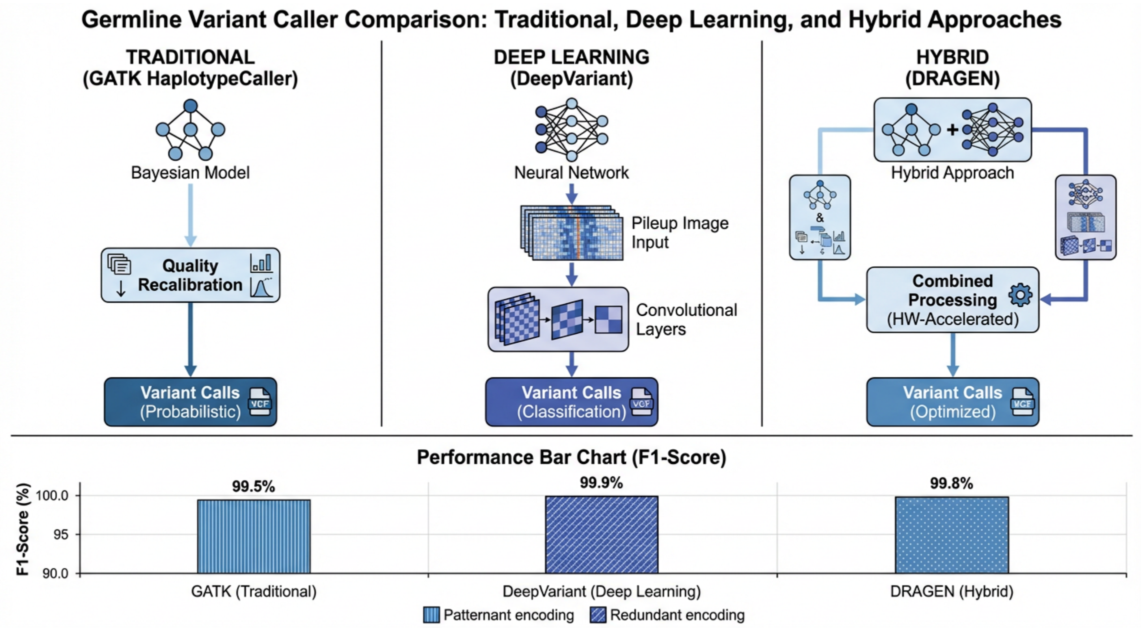


Figure 4: **Comparison of germline variant calling approaches.** Illustration of three paradigms for germline variant calling: (Left) Traditional statistical methods exemplified by GATK HaplotypeCaller, employing Bayesian probability models and hidden Markov models with quality score recalibration; (Center) Deep learning approaches represented by DeepVariant, which processes pileup images through convolutional neural networks; (Right) Hybrid methods such as DRAGEN and DNAscope that combine statistical modeling with AI-enhanced genotyping. Performance metrics from benchmark evaluations demonstrate the competitive accuracy achieved by modern deep learning approaches.

# 4 Somatic Variant Calling

## 4.1 Challenges in Somatic Variant Detection

Somatic variant calling for cancer genomics presents fundamentally different challenges compared to germline analysis. Tumor samples typically represent heterogeneous mixtures of neoplastic and normal cells, with variant allele frequencies determined by the interaction of tumor purity (the proportion of neoplastic cells in the sample) and clonal architecture (the fraction of tumor cells carrying a given mutation). Subclonal mutations, present in only a subset of tumor cells, may occur at VAFs well below the 50% expectation for germline heterozygous variants, potentially approaching the technical noise floor of sequencing.

The detection of low-VAF variants requires careful attention to error modeling, as sequencing and PCR artifacts can generate false positive signals indistinguishable from true low-frequency mutations. Approaches to managing these artifacts include the use of unique molecular identifiers (UMIs) that enable consensus calling to suppress PCR duplicates and sequencing errors, optimized library preparation protocols, and sophisticated error models that account for sequence context-dependent error rates. The emergence of cell-free DNA (cfDNA) analysis for liquid biopsy applications has further intensified these challenges, as circulating tumor DNA (ctDNA) may represent only a small fraction of total cfDNA, requiring detection of variants at VAFs of 0.1% or lower [Wan et al., 2017].

## 4.2 Tumor-Normal Paired Analysis

The availability of matched normal tissue (typically peripheral blood or adjacent normal tissue) substantially improves somatic variant calling accuracy by enabling direct discrimination of somatic mutations from germline polymorphisms and systematic artifacts. Mutect2, the current somatic variant calling module within GATK, employs a Bayesian approach that models the probability of observing the read data given somatic, germline, or artifact hypotheses [Cibulskis et al., 2013, Benjamin et al., 2019]. By jointly analyzing tumor and normal samples, Mutect2 can effectively filter germline variants while maintaining sensitivity for true somatic events.

Strelka2 provides an alternative approach for somatic variant calling from tumor-normal pairs, offering both SNV and indel detection with competitive accuracy and substantially faster runtime compared to Mutect2 [Kim et al., 2018]. Strelka2 employs a mixture model framework that explicitly accounts for tumor purity and models the probability of each candidate variant being somatic, germline heterozygous, or germline homozygous. The tool has demonstrated strong performance in benchmarking evaluations and is widely adopted in clinical and research settings.

VarScan2 represents another established somatic variant caller, employing a heuristic approach based on coverage and variant allele frequency thresholds [Koboldt et al., 2012]. While less sophisticated in its statistical modeling compared to Mutect2 or Strelka2, VarScan2 offers computational efficiency and has been successfully deployed in numerous clinical pipelines. However, default parameters require careful adjustment for low-VAF detection, as the standard VAF threshold of 20% would miss many clinically relevant subclonal mutations.

## 4.3 Tumor-Only Analysis

In many clinical scenarios, matched normal tissue is unavailable, necessitating tumor-only analysis pipelines. This approach introduces additional complexity, as somatic variants must be distinguished from germline polymorphisms without direct comparison. Strategies for tumor-only calling typically involve filtering against population databases (gnomAD, dbSNP) to remove common germline variants, though this approach may inappropriately filter somatic variants occurring at known polymorphic sites or miss rare germline variants absent from databases [Karczewski et al., 2020].

Panel of Normals (PoN) approaches, wherein a collection of normal samples processed through the same laboratory workflow serves as a reference for identifying recurrent artifacts and germline variants, can substantially improve tumor-only calling specificity. The construction of laboratory-specific PoN datasets capturing workflow-specific artifacts is particularly valuable for clinical pipelines. Additionally, analysis of variant allele frequency distributions can aid in discriminating clonal somatic mutations (typically present at VAFs reflecting tumor purity) from subclonal events and germline variants.

For cfDNA applications, tumor-only calling is the norm, as the circulating DNA itself derives from multiple tissue sources including tumor cells. Specialized tools such as eVIDENCE have been developed for cfDNA variant filtering, employing molecular barcoding information to distinguish true low-VAF variants from artifacts [Wan et al., 2017]. Machine learning approaches training on features including read depth, mapping quality, strand bias, and variant position within reads have shown promise for improving the discrimination of true ctDNA variants from false positives.

## 4.4 Ensemble Approaches and Clinical Implementation

The recognition that individual somatic variant callers have different strengths and weaknesses has motivated ensemble approaches that combine results from multiple tools. A recent benchmarking study identified optimal ensembles of four callers for SNV detection (MuSE, Mutect2, Strelka2) and indel detection (Mutect2, Strelka2, VarScan2), with voting thresholds requiring agreement between at least two tools [Van der Sanden et al., 2024]. Such ensemble strategies can improve both sensitivity (through union approaches) and specificity (through intersection requirements), with the optimal strategy depending on clinical priorities.

For clinical oncology laboratories, the selection of somatic variant calling approaches must balance accuracy against practical considerations including computational requirements, turnaround time, and validation burden. Many laboratories employ tiered approaches, with high-sensitivity methods applied to clinically actionable genes and more conservative filtering for genome-wide or exome-wide analysis. Integration with variant annotation and interpretation tools such as VEP [McLaren et al., 2016] and knowledge bases such as OncoKB [Chakravarty et al., 2017] is essential for translating variant calls into clinical recommendations.

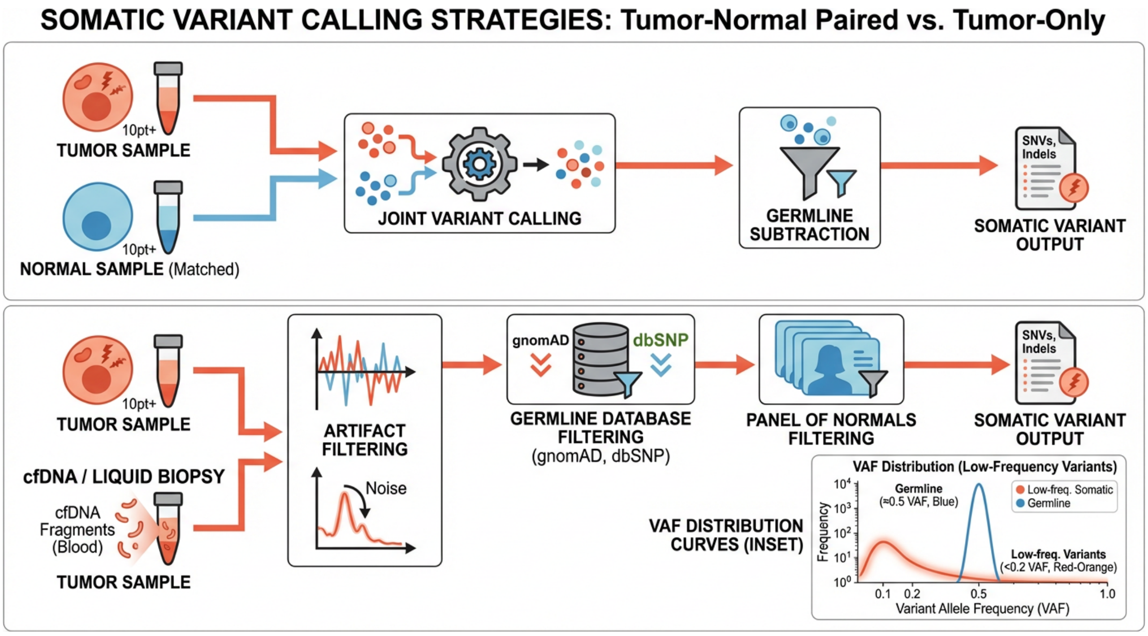


Figure 5: **Somatic variant calling strategies.** Comparison of tumor-normal paired and tumor-only analysis approaches. The tumor-normal paired workflow enables direct identification of somatic variants through comparison with matched germline, while tumor-only analysis requires filtering against population databases and panels of normals to distinguish somatic from germline variants. The schematic illustrates the additional considerations for cell-free DNA (cfDNA) liquid biopsy applications, where detection of low variant allele frequency (VAF) tumor-derived variants requires specialized artifact filtering strategies.

# 5 Structural Variants and Copy Number Variations

## 5.1 Detection Methodologies

Structural variants (SVs), defined as genomic alterations larger than 50 base pairs, encompass deletions, duplications, insertions, inversions, and translocations. These variants contribute significantly to human genetic diversity and disease, yet their detection from short-read sequencing data remains technically challenging compared to small variant calling. The difficulty arises from the fundamental limitations of short reads in spanning or uniquely mapping across large genomic rearrangements, particularly in repetitive regions where SVs are enriched.

Three primary signals are exploited for SV detection from short-read data: discordant read pairs (read pairs with unexpected insert size or orientation), split reads (individual reads that align discontinuously across a breakpoint), and read depth changes (coverage increases or decreases reflecting copy number gains or losses) [Kosugi et al., 2019]. Tools differ in their utilization and integration of these signals, with implications for sensitivity across different SV types and sizes.

Manta, developed by Illumina, employs a comprehensive approach integrating paired-end and split-read information with local assembly for breakpoint refinement [Chen et al., 2016]. The tool demonstrates particularly strong performance for deletion detection and has established clinical utility in constitutional genetics and oncology applications. Manta's local assembly capability enables precise breakpoint resolution and improved detection of novel insertions compared to purely alignment-based methods.

DELLY implements a similar multi-signal approach, combining paired-end and split-read analysis for SV discovery [Rausch et al., 2012]. Originally developed for cancer genomics, DELLY has been extended to germline applications and demonstrates competitive performance across SV types. The tool's population-mode analysis enables joint genotyping across multiple samples, facilitating rare variant identification in cohort studies.

LUMPY employs a probabilistic framework that integrates multiple alignment signals, weighting different evidence types according to their informativeness for each candidate SV [Layer et al., 2014]. This flexible approach enables incorporation of diverse signal types including paired-end, split-read, and even read-depth information from external tools.

## 5.2 Copy Number Variation Detection

Copy number variations (CNVs)—gains or losses of genomic segments—require approaches emphasizing read depth analysis. CNVnator employs a read-depth approach using fixed-size genomic bins, analyzing coverage patterns to identify regions of gain or loss [Abyzov et al., 2011]. This method is particularly effective for large CNVs (multiple kilobases and larger) but may miss smaller events or struggle in regions with variable mappability.

For targeted sequencing and gene panel applications, specialized CNV calling tools optimize performance for the reduced, non-contiguous target space. Tools such as ExomeDepth, CONTRA, and XHMM employ sample-to-reference or sample-to-cohort comparisons to identify CNVs within captured regions. The limited capture uniformity and potential for batch effects require careful attention to reference sample selection and normalization procedures.

The integration of multiple SV and CNV calling tools through ensemble approaches has demonstrated improved performance compared to individual callers. A recent comprehensive evaluation found that merging calls supported by a majority of tools improved precision compared to the best individual caller, though at some cost to recall [Zhang et al., 2024]. For clinical applications, the tradeoff between sensitivity and specificity must be calibrated to the clinical context, with higher sensitivity typically preferred for diagnostic applications and higher specificity for screening.

### 5.3 Clinical Implementation Challenges

The clinical implementation of SV and CNV calling presents several challenges beyond those encountered with small variants. First, the variable performance of callers across SV types and sizes necessitates careful validation across the spectrum of variants relevant to clinical practice. Second, the interpretation of SVs requires integration of multiple annotation sources, including gene content, regulatory element disruption, and literature evidence, without the standardized frameworks available for small variants.

The ACMG has published technical standards for chromosomal microarray analysis and recommendations for the clinical interpretation of copy number variants, providing guidance on CNV classification analogous to the sequence variant guidelines [Richards et al., 2015]. However, the application of these standards to NGS-based CNV detection requires careful attention to the technical differences between methodologies, including resolution limits and detection sensitivity.

For clinical laboratories, the validation of SV/CNV calling pipelines requires representative samples spanning the range of clinically relevant variant types. Orthogonal confirmation methods, including chromosomal microarray, multiplex ligation-dependent probe amplification (MLPA), and optical genome mapping, may be employed to establish truth sets for validation. Ongoing participation in external quality assessment programs and proficiency testing provides additional evidence of pipeline performance.

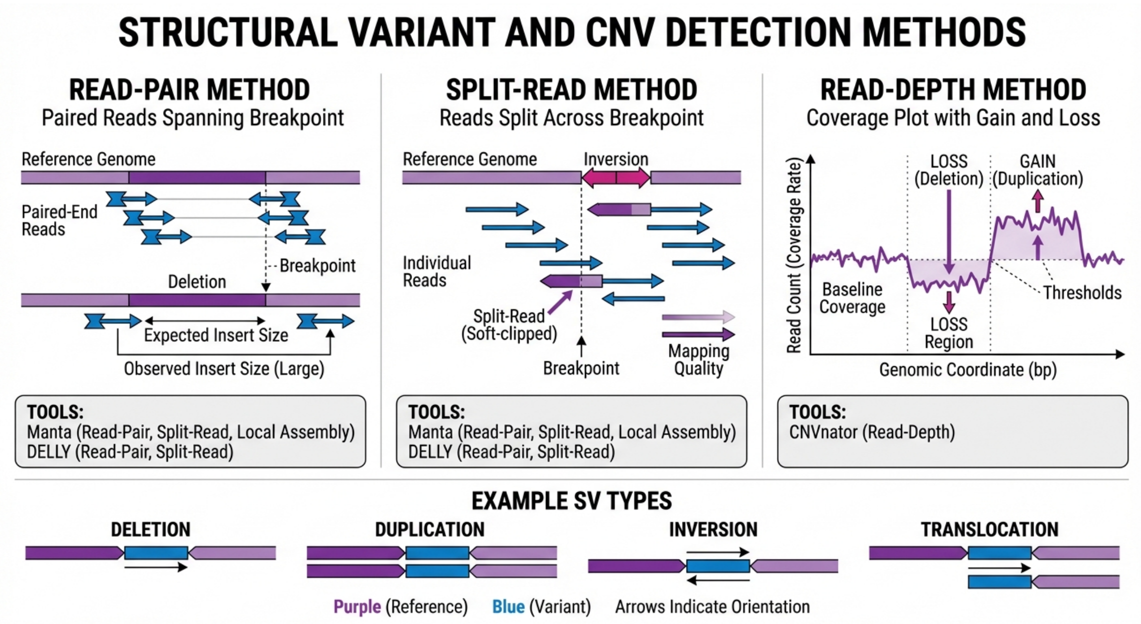


Figure 6: **Structural variant and copy number variation detection methods.** Overview of the three primary approaches for SV/CNV detection from short-read sequencing data. The read-pair method identifies variants through discordant paired-end alignments; the split-read method detects breakpoints through reads aligning discontinuously; the read-depth method identifies copy number changes through coverage analysis. Different tools emphasize different signal types: Manta and DELLY integrate read-pair and split-read signals with local assembly, while CNVnator focuses on read-depth analysis. The schematic illustrates canonical SV types including deletions, duplications, inversions, and translocations.

# 6 Clinical Implementation and Pipeline Validation

## 6.1 Analytical Validation Requirements

The analytical validation of NGS variant calling pipelines constitutes a foundational requirement for clinical implementation, ensuring that assays perform reliably within defined parameters. The AMP/CAP joint recommendations provide a comprehensive framework addressing key validation elements including accuracy, precision, analytical sensitivity, analytical specificity, and reportable range [Roy et al., 2018].

Accuracy assessment requires comparison of variant calls against established reference standards. The GIAB consortium reference materials provide well-characterized truth sets for common variants, while synthetic or spike-in controls may be necessary for rare or challenging variant types. Clinical laboratories should evaluate accuracy across the full spectrum of variant types within the reportable range, with particular attention to indels (which typically show lower accuracy than SNVs) and variants in challenging genomic contexts.

Precision studies evaluate reproducibility across multiple dimensions including within-run (repeatability), between-run, between-operator, and between-instrument variation. For variant calling, precision is typically assessed through concordance of variant calls across replicate analyses. Laboratories should establish acceptance criteria for precision metrics and monitor ongoing performance through quality control procedures.

Analytical sensitivity and specificity characterize the assay's ability to correctly identify true positives and true negatives, respectively. For variant calling, sensitivity is typically high for well-covered regions but may decrease at the edges of capture targets or in regions with reduced coverage. Defining a minimum coverage threshold for variant calling and excluding regions below this threshold from the reportable range represents a common approach to ensuring adequate sensitivity.

The limit of detection (LoD) is particularly important for somatic variant calling, where clinically relevant variants may be present at low VAFs. LoD determination requires analysis of samples with known variants at varying frequencies, typically generated through dilution of positive controls or synthetic spike-ins. Laboratories should clearly communicate the VAF threshold below which variant detection is not reliable.

## 6.2 Quality Metrics and Monitoring

Ongoing quality monitoring ensures sustained pipeline performance after initial validation. Key metrics for variant calling quality include:

- **Coverage depth and uniformity**: Adequate coverage across the target region is prerequisite for variant detection. Metrics including mean coverage, percentage of target bases covered at defined thresholds (e.g., 20x, 50x), and coefficient of variation characterize coverage distribution.
- **Mapping quality**: The proportion of reads uniquely mapping to the reference genome and the distribution of mapping quality scores provide insight into alignment quality and potential issues with sample contamination or library preparation.
- **Base quality**: Base quality score distributions and the proportion of bases exceeding quality thresholds inform variant calling confidence.
- **Transition/transversion ratio**: For whole-exome or genome data, the Ti/Tv ratio provides a global quality metric, with ratios substantially deviating from expected values (approximately 2.0-2.1 for exomes) suggesting potential quality issues.

- **Variant call rates**: Monitoring the total number of variants called per sample, stratified by type (SNV, indel) and classification (pathogenic, VUS, benign), enables detection of systematic shifts potentially indicating pipeline or sample issues.

Automated quality control systems that flag samples or runs failing defined thresholds enable proactive identification of quality issues before results release. Integration of quality metrics into laboratory information systems facilitates trend monitoring and documentation for regulatory compliance.

### 6.3 Pipeline Reproducibility and Version Control

Reproducibility—the ability to regenerate consistent results from the same input data—is essential for clinical variant calling pipelines. Software version control, documented configuration parameters, and containerized execution environments (using tools such as Docker or Singularity) support reproducibility by ensuring consistent runtime environments across analyses.

Reference genome versioning requires particular attention, as differences between genome builds or even between releases of the same build can impact variant calling. Clinical laboratories should standardize on specific reference genome versions and document all ancillary resources (gene annotations, variant databases) used in analysis. Migration between genome builds requires careful re-validation to ensure maintained performance.

Pipeline updates, whether for bug fixes, performance improvements, or incorporation of new features, must be managed through formal change control procedures. Significant changes require re-validation proportionate to the scope of modification, ranging from targeted verification for minor changes to comprehensive re-validation for major updates. Documentation of validation studies, including test data, results, and acceptance criteria, provides essential evidence for regulatory review.

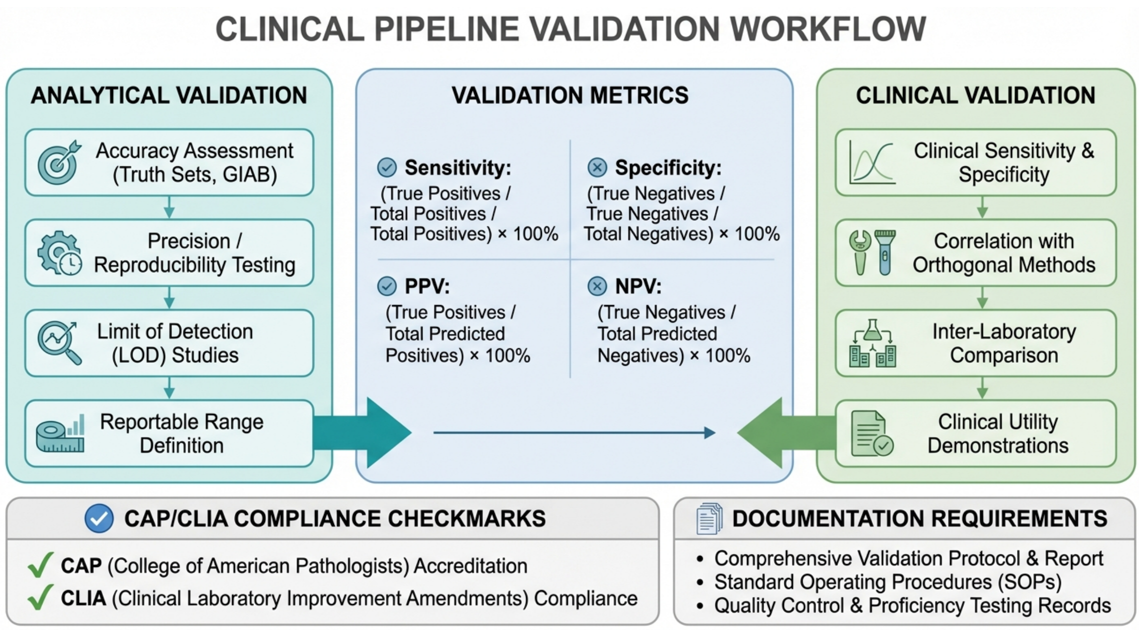


Figure 7: **Clinical pipeline validation workflow.** Overview of analytical and clinical validation requirements for NGS variant calling pipelines. Analytical validation encompasses accuracy assessment using reference materials (GIAB), precision studies evaluating reproducibility, limit of detection determination for low-frequency variants, and reportable range definition. Clinical validation establishes the relationship between variant detection and clinical outcomes. Ongoing quality monitoring ensures sustained performance, with metrics feeding into CAP/CLIA compliance documentation.

## 6.4 Laboratory Information System Integration

The integration of NGS bioinformatics pipelines with laboratory information systems (LIS) represents a critical but often underappreciated aspect of clinical implementation. Effective integration enables seamless sample tracking from accessioning through result reporting, automated data transfer between systems, and audit trail documentation for regulatory compliance.

Standards-based interfaces facilitate interoperability between NGS systems and broader healthcare information infrastructure. Health Level Seven (HL7) messaging standards, including HL7 v2 for traditional laboratory reporting and the emerging Fast Healthcare Interoperability Resources (FHIR) standard for genomic data exchange, provide frameworks for structured data transmission. The GA4GH (Global Alliance for Genomics and Health) has developed specifications for genomic data representation and exchange that complement healthcare standards.

Variant curation workflows, wherein identified variants undergo expert review and classification, require specialized interfaces supporting efficient review of variant evidence, documentation of classification rationale, and generation of clinical reports. Several commercial and open-source platforms provide variant curation functionality, with varying degrees of integration with upstream analysis and downstream reporting systems.

The complexity of NGS data and the nuanced nature of variant interpretation create challenges for standard laboratory reporting templates. Clinical reports must communicate essential information including variants identified, classification rationale, and clinical implications while remaining comprehensible to non-specialist ordering providers. Emerging approaches including interactive electronic reports and integration with clinical decision support systems offer potential for improved communication of complex genomic information.

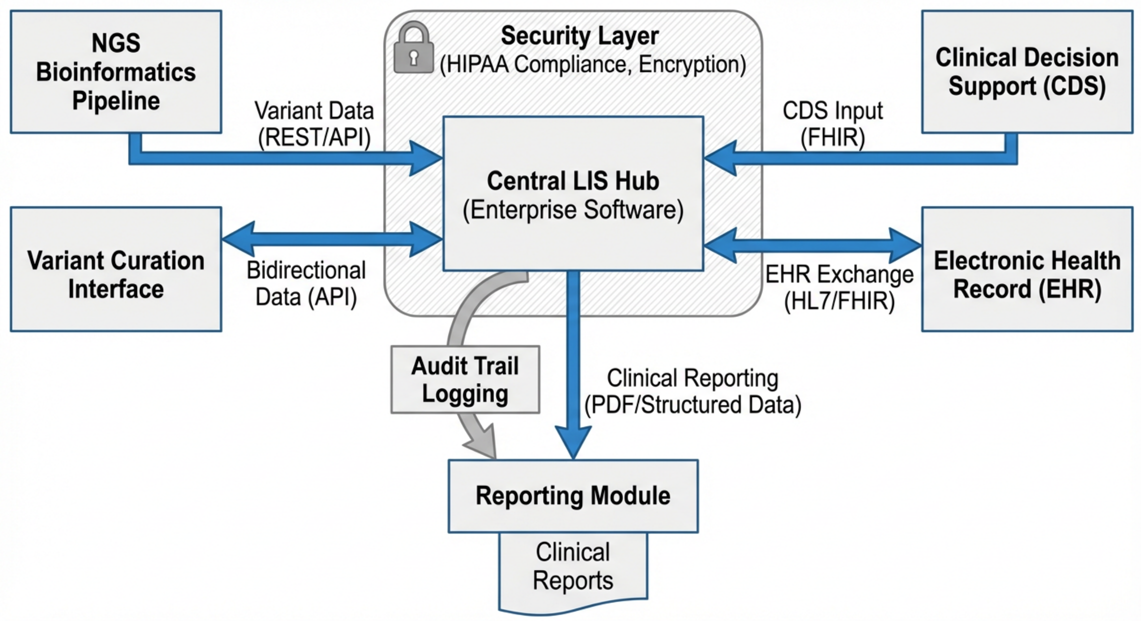


Figure 8: **Laboratory information system integration architecture.** Schematic illustrating the integration of NGS bioinformatics pipelines with laboratory information systems (LIS) and electronic health records (EHR). The architecture encompasses data flow from NGS analysis through variant curation interfaces to clinical decision support and reporting modules. Standards-based interfaces (HL7/FHIR) enable interoperability with healthcare information systems. Security layers ensure HIPAA compliance and audit trail documentation.

# 7 Emerging Trends and Future Directions

## 7.1 Long-Read Sequencing Technologies

The maturation of long-read sequencing technologies from Pacific Biosciences (PacBio HiFi) and Oxford Nanopore Technologies (ONT) is reshaping the variant calling landscape. Long reads spanning thousands of base pairs enable resolution of complex structural variants, repetitive regions, and haplotype phasing that remain challenging for short-read approaches. PacBio HiFi reads, with circular consensus accuracy exceeding 99.9%, now approach short-read accuracy while providing dramatically improved contiguity.

Variant calling tools have been developed or adapted for long-read data. DeepVariant provides models optimized for both PacBio and ONT platforms, while Clair3 implements a unified architecture supporting multiple sequencing technologies [Zheng et al., 2022, Poplin et al., 2018]. For structural variants, long reads provide direct evidence of large insertions and complex rearrangements that short reads can only indirectly infer. The precisionFDA Truth Challenge V2 demonstrated the superior performance of long-read approaches in challenging genomic regions [Olson et al., 2022].

Clinical adoption of long-read sequencing is accelerating, particularly for applications where short-read limitations have constrained diagnostic yield. Constitutional genetics applications benefiting from long-read approaches include repeat expansion disorders, structural variant-driven syndromes, and pharmacogenomic haplotyping. In oncology, long-read sequencing of tumor samples can resolve complex structural rearrangements and improve characterization of fusion genes.

### 7.2 Artificial Intelligence and Machine Learning Advances

The success of deep learning for variant calling has stimulated broader application of artificial intelligence across clinical genomics workflows. Beyond primary variant calling, AI approaches are being developed for variant prioritization and interpretation, clinical phenotype extraction from medical records, and prediction of variant pathogenicity from sequence features.

Foundation models trained on large genomic datasets show promise for diverse downstream applications. Models such as those emerging from genomic language model research can capture complex sequence patterns relevant to variant impact prediction, potentially improving upon existing computational prediction tools. The integration of multi-modal data—combining genomic variants with gene expression, protein structure predictions, and clinical features—offers opportunities for more comprehensive variant assessment.

However, the deployment of AI systems in clinical settings requires careful attention to validation, interpretability, and ongoing performance monitoring. Regulatory frameworks for AI/ML-based medical devices are evolving, with the FDA developing approaches for continuous learning systems that update based on real-world data. Clinical laboratories implementing AI-enhanced pipelines must establish appropriate validation procedures and ensure transparency in how AI contributions influence clinical decisions.

### 7.3 Population-Scale Genomics and Collaborative Networks

The accumulation of population-scale genomic datasets is transforming variant interpretation. Resources such as gnomAD, now encompassing data from over 140,000 exomes and genomes, provide increasingly refined estimates of variant frequency across diverse populations [Karczewski et al., 2020]. This information is essential for filtering common polymorphisms and identifying variants that are rare enough to potentially cause disease.

Collaborative networks for variant interpretation, exemplified by ClinGen's expert panels and variant curation groups, enable the systematic accumulation of evidence across contributing laboratories [Rehm et al., 2015]. Data sharing initiatives and harmonized classification standards are progressively reducing inter-laboratory discordance, though substantial challenges remain, particularly for rare variants and genes with limited functional characterization.

The expansion of genomic testing across diverse populations has highlighted the need for ancestry-inclusive reference databases and variant interpretation frameworks. Variants common in underrepresented populations may be misclassified as potentially pathogenic due to their absence from historically Eurocentric databases. Addressing these disparities requires deliberate efforts to expand reference databases and ensure that variant interpretation tools perform equitably across populations.

## 8 Conclusion

NGS variant calling has matured from an emerging technology to an established cornerstone of clinical molecular diagnostics. The tools and methodologies reviewed in this paper—from traditional statistical approaches such as GATK HaplotypeCaller to deep learning innovations exemplified by DeepVariant, from matched tumor-normal somatic calling with Mutect2 to structural variant detection with Manta and DELLY—collectively enable comprehensive genetic variant detection across clinical applications in constitutional genetics and oncology.

Several key themes emerge from this review. First, the continued improvement in variant calling accuracy, driven particularly by deep learning approaches, has narrowed the gap between sequencing data generation and reliable variant identification. Modern tools achieve SNP F1-scores exceeding 99.9% and indel F1-scores above 99%, with ongoing refinements addressing challenging genomic regions and variant types. Second, the availability of robust reference standards, most notably from the GIAB consortium, provides essential foundations for pipeline validation and benchmarking. Third, regulatory frameworks including ACMG/AMP guidelines for variant interpretation and AMP/CAP standards for pipeline validation establish clear expectations for clinical implementation.

Despite these advances, significant challenges remain. Low variant allele frequency detection in tumor samples and liquid biopsies continues to require careful attention to error modeling and artifact filtering. Structural variant calling, while improved, remains less mature than small variant detection, with variable performance across SV types and sizes. The interpretation of variants of uncertain significance constitutes an ongoing bottleneck, with functional characterization and clinical correlation data accumulating more slowly than variant discovery.

For clinical geneticists and laboratory directors, the practical implications of this review include several recommendations. Pipeline selection should balance accuracy against practical considerations including computational requirements, validation burden, and integration with existing systems. Ensemble approaches combining multiple callers may offer advantages in sensitivity or specificity depending on clinical priorities. Robust validation against appropriate reference materials and ongoing quality monitoring are essential for sustained performance. Finally, engagement with collaborative networks and shared resources supports the progressive refinement of variant interpretation.

Looking forward, the integration of long-read sequencing, AI-enhanced interpretation, and population-scale data resources promises continued improvement in clinical variant detection and interpretation. As genomic testing expands to broader patient populations and new clinical applications, the variant calling tools and frameworks described in this review will continue to evolve, requiring ongoing attention from clinical laboratories to maintain state-of-the-art capabilities.

## Acknowledgments

This review was generated using K-Dense Web (k-dense.ai), an AI-powered scientific writing platform.

## Conflict of Interest

The author declares no conflicts of interest.

## References

Alexej Abyzov, Alexander E Urban, Michael Snyder, and Mark Gerstein. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research*, 21(6):974–984, 2011. doi: 10.1101/gr.114876.110.

David Benjamin, Takuto Sato, Kristian Cibulskis, Gad Getz, Chip Stewart, and Lee Lichtenstein. Calling somatic SNVs and indels with Mutect2. *bioRxiv*, page 861054, 2019. doi: 10.1101/861054.

Debyani Chakravarty, Jianjiong Gao, Sarah M Phillips, Ritika Kundra, Hongxin Zhang, Jiaojiao Wang, Julia E Rudolph, Rona Yaeger, Tara Soumerai, Moriah H Nissan, et al. OncoKB: a precision oncology knowledge base. *JCO Precision Oncology*, 1:1–16, 2017. doi: 10.1200/PO.17.00011.

Jing Chen, Xue Li, Hai Zhong, Yuhui Meng, and Hongli Du. Comparison of GATK and DeepVariant by trio sequencing. *Scientific Reports*, 12(1):1929, 2022. doi: 10.1038/s41598-022-05833-4.

Xiaoyu Chen, Ole Schulz-Trieglaff, Richard Shaw, Bret Barnes, Felix Schlesinger, Morten Källberg, Anthony J Cox, Semyon Kruglyak, and Christopher T Saunders. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 32(8):1220–1222, 2016. doi: 10.1093/bioinformatics/btv710.

Kristian Cibulskis, Michael S Lawrence, Scott L Carter, Andrey Sivachenko, David Jaffe, Carrie Sougnez, Stacey Gabriel, Matthew Meyerson, Eric S Lander, and Gad Getz. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, 31(3):213–219, 2013. doi: 10.1038/nbt.2514.

Mark A DePristo, Eric Banks, Ryan Poplin, Kiran V Garimella, Jared R Maguire, Christopher Hartl, Anthony A Philippakis, Guillermo del Angel, Manuel A Rivas, Matt Hanna, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5):491–498, 2011. doi: 10.1038/ng.806.

Erik Garrison and Gabor Marth. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*, 2012.

Sara Goodwin, John D McPherson, and W Richard McCombie. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6):333–351, 2016. doi: 10.1038/nrg.2016.49.

Malachi Griffith, Nicholas C Spies, Kilannin Krysiak, Joshua F McMichael, Adam C Coffman, Arpad M Danos, Benjamin J Ainscough, Cynthia A Ramirez, Damian T Rieke, Lynzey Kujan, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nature Genetics*, 49(2):170–174, 2017. doi: 10.1038/ng.3774.

Celia Y Ho et al. Accuracy and efficiency of germline variant calling pipelines for human genome data. *Scientific Reports*, 10(1):20940, 2020. doi: 10.1038/s41598-020-77218-4.

Konrad J Karczewski, Laurent C Francioli, Grace Tiao, Beryl B Cummings, Jessica Alföldi, Qingbo Wang, Ryan L Collins, Kristen M Laricchia, Andrea Ganna, Daniel P Birnbaum, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443, 2020. doi: 10.1038/s41586-020-2308-7.

Sangtae Kim, Konrad Scheffler, Aaron L Halpern, Mitchell A Bekritsky, Eunho Noh, Morten Källberg, Xiaoyu Chen, Yeonbin Kim, Daniel Beez, Semyon Kruglyak, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods*, 15(8):591–594, 2018. doi: 10.1038/s41592-018-0051-x.

Daniel C Koboldt, Qunyuan Zhang, David E Larson, Dong Shen, Michael D McLellan, Ling Lin, Christopher A Miller, Elaine R Mardis, Li Ding, and Richard K Wilson. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, 22(3):568–576, 2012. doi: 10.1101/gr.129684.111.

Shunichi Kosugi, Yukihide Momozawa, Xiaoxi Liu, Chikashi Terao, Michiaki Kubo, and Yoichiro Kamatani. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biology*, 20(1):117, 2019. doi: 10.1186/s13059-019-1720-5.

Peter Krusche, Len Trigg, Paul C Boutros, Christopher E Mason, Francisco M De La Vega, Benjamin L Moore, Mar Gonzalez-Porta, Michael A Eberle, Zivana Tezak, Samir Lababidi, et al. Best practices for benchmarking germline small-variant calls in human genomes. *Nature Biotechnology*, 37(5):555–560, 2019. doi: 10.1038/s41587-019-0054-x.

Melissa J Landrum, Jennifer M Lee, Mark Benson, Garth R Brown, Chen Chao, Shanmuga Chitipiralla, Baoshan Gu, Jennifer Hart, Douglas Hoffman, Wonhee Jang, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46(D1):D1062–D1067, 2018. doi: 10.1093/nar/gkx1153.

Ryan M Layer, Colby Chiang, Aaron R Quinlan, and Ira M Hall. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biology*, 15(6):R84, 2014. doi: 10.1186/gb-2014-15-6-r84.

Marilyn M Li, Michael Datto, Eric J Duncavage, Shashikant Kulkarni, Neal I Lindeman, Somak Roy, Apostolia M Tsimberidou, Cindy L Vnencak-Jones, Daniel J Wolff, Anas Younes, and Marina N Nikiforova. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *The Journal of Molecular Diagnostics*, 19(1):4–23, 2017. doi: 10.1016/j.jmoldx.2016.10.002.

Shuang Li et al. Artificial intelligence in variant calling: a review. *Frontiers in Bioinformatics*, 5:1574359, 2025. doi: 10.3389/fbinf.2025.1574359.

Elaine R Mardis. The impact of next-generation sequencing on cancer genomics: from discovery to clinic. *Cold Spring Harbor Perspectives in Medicine*, 9(9):a036269, 2019. doi: 10.1101/cshperspect.a036269.

Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, and Mark A DePristo. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303, 2010. doi: 10.1101/gr.107524.110.

William McLaren, Laurent Gil, Sarah E Hunt, Harpreet Singh Riat, Graham RS Ritchie, Anja Thormann, Paul Flicek, and Fiona Cunningham. The Ensembl variant effect predictor. *Genome Biology*, 17(1):122, 2016. doi: 10.1186/s13059-016-0974-4.

Nathan D Olson, Justin Wagner, Jennifer McDaniel, Sarah H Stephens, Samuel T Westreich, Anish G Prasanna, Elaine Johanson, Emily Boja, Ezekiel J Maier, Omar Nevins, et al. precisionFDA Truth Challenge V2: Calling variants from short and long reads in difficult-to-map regions. *Cell Genomics*, 2(5):100129, 2022. doi: 10.1016/j.xgen.2022.100129.

Ryan Poplin, Pi-Chuan Chang, David Alexander, Scott Schwartz, Thomas Colthurst, Alexander Ku, Dan Newburger, Jojo Dijamco, Nam Nguyen, Pegah T Afshar, et al. A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36(10):983–987, 2018. doi: 10.1038/nbt.4235.

Tobias Rausch, Thomas Zichner, Andreas Schlattl, Adrian M Stütz, Vladimir Benes, and Jan O Korbel. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 28(18):i333–i339, 2012. doi: 10.1093/bioinformatics/bts378.

Heidi L Rehm, Jonathan S Berg, Lisa D Brooks, Carlos D Bustamante, James P Evans, Melissa J Landrum, David H Ledbetter, Donna R Maglott, Christa Lese Martin, Robert L Nussbaum, et al. ClinGen—the clinical genome resource. *New England Journal of Medicine*, 372(23): 2235–2242, 2015. doi: 10.1056/NEJMsr1406261.

Sue Richards, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W Grody, Madhuri Hegde, Elaine Lyon, Elaine Spector, Karl Voelkerding, and Heidi L Rehm. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5):405–424, 2015. doi: 10.1038/gim.2015.30.

Andy Rimmer, Hang Phan, Iain Mathieson, Zamin Iqbal, Stephen RF Twigg, Andrew OM Wilkie, Gil McVean, and Gerton Lunter. Integrating mapping-, assembly-and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics*, 46 (8):912–918, 2014. doi: 10.1038/ng.3036.

Somak Roy, Christopher Coldren, Arivarasan Karunamurthy, Nathaniel S Kip, Eric W Klee, Stephen E Lincoln, Annette Leon, Madhuri Pullambhatla, Robyn L Temple-Smolkin, Karl V Voelkerding, Chen Wang, and Alexis B Carter. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: a joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. *The Journal of Molecular Diagnostics*, 20(1):4–27, 2018. doi: 10.1016/j.jmoldx.2017.11.003.

Jay Shendure, Shankar Balasubramanian, George M Church, Walter Gilbert, Jane Rogers, Jeffery A Schloss, and Robert H Waterston. DNA sequencing at 40: past, present and future. *Nature*, 550(7676):345–353, 2017. doi: 10.1038/nature24286.

Bert Van der Sanden et al. A benchmarking study of individual somatic variant callers and voting-based ensembles for whole-exome sequencing. *Briefings in Bioinformatics*, 26(1): bbae697, 2024. doi: 10.1093/bib/bbae697.

Jonathan CM Wan, Charles Massie, Javier Garcia-Corbacho, Florent Mouliere, James D Brenton, Carlos Caldas, Simon Pacey, Richard Baird, and Nitzan Rosenfeld. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nature Reviews Cancer*, 17(4): 223–238, 2017. doi: 10.1038/nrc.2017.7.

Sophia Yohe and Bharat Thyagarajan. Review of clinical next-generation sequencing. *Archives of Pathology & Laboratory Medicine*, 141(11):1544–1557, 2017. doi: 10.5858/arpa.2016-0501-RA.

Lei Zhang, Wenling Bai, Ningning Yuan, and Zhenglin Du. Comparison of structural variant callers for massive whole-genome sequence data. *BMC Genomics*, 25(1):284, 2024. doi: 10.1186/s12864-024-10239-9.

Zhenxian Zheng, Shumin Li, Junhao Su, Amy WS Leung, Tak-Wah Lam, and Ruibang Luo. Clair3: Integrating nanopore and illumina sequencing data for comprehensive variant detection. *Nature Computational Science*, 2:797–807, 2022. doi: 10.1038/s43588-022-00387-x.

Justin M Zook, Brad Chapman, Jason Wang, David Mittelman, Oliver Hofmann, Winston Hide, and Marc Salit. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nature Biotechnology*, 32(3):246–251, 2014. doi: 10.1038/nbt.2835.

Justin M Zook, Jennifer McDaniel, Nathan D Olson, Justin Wagner, Hemang Parber, Leandros Antoniou, Charles Hock, Marc L Salit, et al. An open resource for accurately benchmarking small variant and reference calls. *Nature Biotechnology*, 37(5):561–566, 2019. doi: 10.1038/s41587-019-0074-6.