Standardizing Clinical NGS Applications—Based on ACMG Standards

Seung-hyun Yoon[1, a], Chae-hee Lim[1, b, *],

[1]KOTITI Testing,48 Gwacheon-daero 7na-gil, Gwacheon-si, Gyeonggi-do,13840, Republic of Korea
a.seunghyun_11_4258@163.com, b.lim07chaehee_2023@sina.com
***Corresponding Author**

Abstract: Next-generation sequencing (NGS)technology has been applied in clinical laboratories. The development of this technology, along with the expansion of testing scale and scope, has significantly reduced the cost of genetic testing. The clinical applications of NGS are continuously expanding, including targeted gene panel testing, individual exome, or whole-genome analysis. Moreover, NGS technology has further improved treatment decision-making and prediction for high-risk populations of certain diseases. However, the development of NGS also poses challenges, including the confusion of choices brought by sequencing platforms and technological updates, and the inconsistency of clinical validation of sequencing results. To help clinical laboratories better understand NGS technology, resolve issues related to clinical validation, and correctly select NGS sequencing platforms and methods, as well as to continuously monitor NGS and ensure high-quality sequencing results and their clinical interpretation, this article summarizes and discusses the professional standards and guidelines of the American College of Medical Genetics and Genomics (ACMG)regarding the clinical application of NGS technology. This is intended for reference by molecular diagnostic professionals in China when formulating domestic industry standards.

Keywords: American College of Medical Genetics and Genomics, Next-generation Sequencing, Disease-targeted Gene Panel Testing, Exome Sequencing, Whole-genome Sequencing

1. Introduction

Next-generation sequencing (NGS)technology is a method of large-scale parallel sequencing that utilizes clonal amplification or single-molecule template techniques. The development of NGS technology has been rapid, and it is now widely used in clinical settings. NGS is characterized by its high throughput and low cost per base, but there are still many issues when using this technology for clinical testing. For example, the standardization of sequencing processes, the confusion caused by different sequencing platforms and technical principles in user selection, the variability in data analysis of sequencing results, and the inconsistency between results and clinical interpretation. To address these issues, this article summarizes and discusses the professional standards and guidelines of the American College of Medical Genetics and Genomics (ACMG)regarding NGS technology, for reference by molecular diagnostic professionals in China.

2. Overview of NGS analysis

2.1. Three levels of NGS analysis

NGS technology is categorized into three levels of complexity: disease-targeted gene panels, exome sequencing (ES), and genome sequencing (GS).

Disease-targeted gene panelsare used to detect known disease-related genes. They sequence only the genes associated with a specific disease, resulting in higher sensitivity and specificity. In terms of equipment, small benchtop sequencers are sufficient for this type of testing. The data volume generated is relatively low, and the storage requirements are minimal.

Exome sequencing (ES)is a method that sequences all the coding regions of the genome. It is mainly used to detect mutations in known disease-related genes or to identify novel pathogenic genes.ES has intermediate sequencing capacity and cost compared to disease-targeted gene panels and GS. Its analytical

https://doi.org/10.62852/csa/2025/154

sensitivity is lower than most disease-targeted gene panels, and its specificity may also decrease with reduced sequencing efficiency. Therefore, Sanger sequencing is often required to confirm and exclude false-positive mutation signals.

Genome sequencing (GS)covers both coding and non-coding regions of the entire genome. Its advantage is that the sample preparation is simple and does not require PCR or hybridization enrichment steps for target regions. However, GS is not ideal for interpreting non-coding variants, so analysis usually starts with the coding regions. If no pathogenic mutations are found, the data can be re-analyzed to search for regulatory regions in the non-coding areas. Additionally, GS generates a large volume of data, with time-consuming analysis and significant result variability. Its widespread clinical application is still not fully mature.

2.2. The workflow of NGS

The workflow of NGS includes three main steps: sample preparation, sequencing, and data analysis. Sample preparation involves the collection and processing of patient samples, DNA extraction, library construction, application of biological molecular tags(barcodes),capture and enrichment of target regions/genes, ligation, and amplification; the sequencing step involves running the prepared samples on sequencing machines; data analysis involves processing the sequence data generated after sequencing through specialized bioinformatics pipelines for detecting DNA variants, with the main processes including base calling, read alignment, variant identification, and variant annotation.

3. Factors to consider when conducting sequencing work

The main factors affecting the development of NGS testing and analysis in the laboratory include the content of sequencing, the selection of sequencing platforms and methods, the choice of data analysis tools, variant filtering, homologous region sequencing, and variant file formats, all of which are important factors that must be considered before conducting sequencing work.

3.1. Sequencing content

When conducting NGS, it is essential to select targeted gene panels that are highly associated with the disease in question, meaning that there is sufficient scientific evidence to support their relevance to the disease. If a targeted panel includes genes associated with multiple phenotypes, laboratories should offer sub-panel testing to clinical physicians to achieve greater specificity and reduce the workload of experimental analysis. Generally speaking, the larger the area of analysis, the greater the number of variants related to the clinical phenotype that will be obtained, and the greater the workload of data analysis.

3.2. Selection of sequencing platforms and methods

When choosing a sequencing platform, laboratories must carefully consider factors such as the size of the sequencing region, the required coverage, the expected sample volume, the required run time, and the cost to determine the criteria for purchasing instruments. The choice of detection platform should also be based on the number and scale of variants to be detected. There are two sequencing modes: single-end sequencing and paired-end sequencing. Paired-end sequencing can make the read length clearer, which is particularly necessary for reading base data in repetitive regions.

3.3. Data analysis tools

Each NGS detection platform is equipped with specific base-calling algorithm software, and the correctness of base calling is generally assessed using Phred-like scores. For targeted detection and exome sequencing (ES), it is recommended to perform full-length calibration with a reference genome to reduce the generation of erroneous sequencing information outside the detection panel and to increase sequencing coverage to improve the accuracy of variant identification. For genome sequencing(GS),local alignment after sequence alignment helps to identify insertion or deletion mutation information.

https://doi.org/10.62852/csa/2025/154

3.4. Variant filtration

Laboratories need to establish variant screening methods to exclude common benign variants and identify molecules associated with rare diseases. For targeted NGS panels, laboratories should develop automated classification tools to filter out common benign variants from rare, highly penetrant variants. Reference databases such as d b SNP, the NHLBI Exome Sequencing Project, and the 1000 Genomes Project can be used for automated classification. In exome sequencing (ES)or genome sequencing (GS)analyses, appropriate family members can be used as references to identify molecules associated with rare diseases. Laboratories must strike a balance between over-filtering and under-filtering: over-filtering may inadvertently exclude pathogenic mutations, while under-filtering can lead to an excessive number of variants to analyze, increasing the workload. Laboratories can proceed stepwise, first identifying clear and obvious causes, and if necessary, adjust filtering criteria to expand the search for mutated genes.

3.5. Homologous region sequencing

Homologous sequences, such as pseudogenes, can interfere with short-read sequencing methods. When reads are incorrectly aligned to homologous regions due to the limited length of NGS sequencing reads, false-positive variant calls may occur; conversely, when reads containing mutations are aligned to homologous sites, false-negative results may arise. Therefore, methods to detect pathogenic variants within known homologous regions need to be identified, such as using global alignment followed by local alignment or paired-end sequencing. In some cases, Sanger sequencing with specific complementary primer pairs can be used to confirm the variant region. Thus, laboratories must have extensive experience with NGS technologies, be fully aware of the limitations of the techniques used,and, if necessary, use orthogonal methods to confirm all disease-targeted or diagnostic test results. Sanger sequencing is the most commonly used confirmatory method.

3.6. File formats

NGS test outputs should adhere to commonly used formats (e.g.,. bam files for aligned signals and. Fast q files for recording sequence information and base quality), and the data formats used should be easily convertible to standard formats. The current standard format—Variant Call Format(.v c f)—has been adopted by the 1000 Genomes Project.

4. NGS operation validation

NGS testing equipment and reagents should have a legitimate product registration certificate. Commercial products related to NGS must be validated by clinical laboratories before they can be used as diagnostic tools. Other unapproved products should be labeled as “For Research Use Only,” “Investigational Use Only,” or “Analyte Specific Reagents.” Projects considered as laboratory-developed tests need to go through a complete approval process.

4.1. Test validation and optimization

Validation should cover sample preparation, sequencing platforms, and data analysis methods. Once the test workflow is selected, it should be repeatedly performed until all testing conditions and analysis settings are optimized. Laboratories should establish variant identification protocols by analyzing data of known sequence variants (e.g., SNPs, small insertions/deletions, large copy number variations, and structural variations). It is worth noting that using synthetic mutants can help create large-scale trial data for comparison and optimization across different sequencing platforms and for setting thresholds for each platform. Therefore, laboratories must establish standard operating procedures for testing and strictly enforce them. Additionally, the optimization process should include all clinical sample types (e.g., whole blood, saliva, formalin-fixed paraffin-embedded tissues).

Laboratories should use reference samples with characteristic features from the Sanger database to determine the validity of NGS. These samples do not need to contain specific pathogenic mutations but should be renewable resources for establishing baseline data as a reference for subsequent test calibration. For validation of disease-targeted gene panels, the focus should be on the specific association between genes and diseases, as well as a broad spectrum of pathogenic mutation genes. Additionally, the accuracy

https://doi.org/10.62852/csa/2025/154

of sequencing in highly homologous regions should be noted. For ES and GS, validation should focus on high-quality exome/genome metrics, such as achieving average coverage or the percentage of bases reaching a minimum coverage threshold. Any changes in the laboratory should be validated using previous test samples or characteristic reference samples through a complete re-testing process.

4.2. Validation of performance characteristics

The performance characteristics that need validation include analytical sensitivity, specificity, reproducibility, and repeatability. These parameters should be established at the technical level and must reflect the entire testing process.

5. Quality Control in NGS Analysis

Laboratories must establish quality control (QC)measures and apply them to monitor each test, including pre-analytical, analytical, and post-analytical quality monitoring. Monitoring content includes sample quality, test parameters, personnel qualifications, data storage, reference materials, and proficiency testing.

5.1. Sample quality requirements

DNA for NGS can be extracted from various types of samples. Laboratories should specify sample types and minimum sample volumes. The quality of DNA and mutation detection requirements vary among different sample types, and laboratories need to determine acceptable criteria for each sample type (e.g., sample volume, tissue amount). They should also develop laboratory operating manuals and quality management plans for DNA extraction and quantitative analysis, and note these parameters in patient reports. For samples that do not meet requirements, re-sampling should be conducted. If re-sampling is not possible, whole-genome amplification may be performed if the laboratory is proficient in this technique, with a note in the report about potential inherent biases to inform clinicians and patients of the limitations of the technique. In any case, laboratories should document operational times and methodological choices in their laboratory operating manuals.

5.2. Test parameter requirements

Laboratories should establish QC checkpoints for quality assurance monitoring. The content of QC monitoring includes: DNA quality, monitoring of error rates during sequencing, and assessment of data quality after run/before analysis; QC operation records should also include the instruments used and reagent batch numbers. During validation, laboratories should document deviations from standard procedures and corrective actions; they must also record the bioinformatics workflow in NGS data analysis, and document any special data in the analysis process; the analysis system should be perfected to allow laboratories to track software versions and the specific changes of each version.

5.3. Personnel qualification requirements

Technicians engaged in NGS should undergo specialized pre-job training and obtain qualifications, with rich experience in variant sequence assessment; they should be well-versed in the pathogenesis of diseases and the performance of NGS technology. If the laboratory provides ES and GS services, experts proficient in the relationship between genes, mutations, and disease phenotypes should be available.

5.4. Data storage

NGS data storage methods should comply with regulatory requirements to ensure the traceability of patient data. Laboratories should clearly specify file types and determine the retention period for each type in their procedures. It is recommended that laboratories retain each file type for at least two years to allow tracing of initial test results and re-analysis of data after improvements in the analysis process. In addition, laboratories should consider retaining.vcf files and final clinical reports and interpretations of related variant subtypes. When conditions permit, the possible significance of the relationship between mutated genes and diseases can be re-interpreted.

https://doi.org/10.62852/csa/2025/154

5.5. Reference materials

Laboratories should utilize reference materials (RMs)with well-characterized features as references for sequencing and copy number variation detection. Human cell lines provided by Coriell have highly specific genomic regions and non-shared areas, which have been used in GS experiments. However, it is important to note the potential decrease in genomic stability over time. Additionally, other organizations such as the College of American Pathologists (CAP)and the U.S. Food and Drug Administration provide RMs for evaluating instrument and test performance. In practice, genomic DNA extracted from whole blood is stable and can be amplified through whole-genome amplification techniques to generate a large quantity of DNA from a small sample for long-term use. Simulated electronic sequences obtained through computational methods can also serve as RMs and can be integrated into the QC or proficiency testing process. These simulated sequences are particularly useful for addressing specific regions, such as repetitive sequences, known insertions and deletions, and SNPs.

5.6. Proficiency testing

In accordance with the requirements of the Clinical Laboratory Improvement Amendments (CLIA), proficiency testing (PT)procedures must be established and conducted regularly. Laboratories must employ currently accepted proficiency testing methods. If no other laboratories are conducting parallel NGS testing, internal proficiency testing methods should be developed and utilized by the laboratory to validate their results.

6. Sequencing request forms, report formats, and result interpretation

6.1. Sequencing request forms

For genetic testing applications, the requesting physician is required to provide detailed clinical information about the patient, such as clinical features, family history, and physical examination findings, when requesting genetic testing. This information enables the laboratory to determine whether the detected genetic variants are consistent with the patient's clinical symptoms and signs and to assess the patient's prognosis. The laboratory will then conduct the genetic tests and report any potentially pathogenic variants detected in the genes, along with an explanation of the testing methods and operational parameters for each test.

6.2. Report format

The laboratory director determines the details of the report. The report must clearly and concisely explain any relevant results and their clinical significance to facilitate the physician's understanding of the test results and describe the advantages and limitations of the methods used.

All NGS reports must include a list of identified mutations, annotated according to the nomenclature guidelines of the Human Genome Variation Society and clinically classified in accordance with the ACMG guidelines (which are regularly updated, and laboratories should monitor changes to these guidelines). Additionally, gene nomenclature should follow the nomenclature of the Human Genome Organization. The report should detail the exon number of the transcript corresponding to each variant site, any mutations, methodology, and referenced websites. It is recommended that laboratories report variant data in a structured format. The report should also provide technical parameters and analytical limitations at the level of the targeted region, exome, or genome. For reporting incidental findings, in accordance with recognized medical practice and ethical principles, laboratories should develop relevant policies to determine whether certain types of incidental findings should be reported and when.

6.3. Result interpretation

According to the ACMG guidelines, each identified mutation in disease-targeted gene panel reports must be evaluated and classified, including an evidence-based assessment of the likelihood of the disease-related mutation gene function and its gene product. For experimental results obtained from ES and GS testing, in addition to determining whether the identified variants alter gene or protein function, the known gene breakpoint phenotypes should also be assessed for correlation with the patient's clinical

https://doi.org/10.62852/csa/2025/154

phenotype. If multiple potentially clinically significant variants are detected, the interpretation should discuss the possible correlation of each variant with the patient's phenotype and prioritize the mutations. In ES and GS test reports, for gene variants with no disease association, that is, variants of uncertain significance (VUSs), laboratories should establish procedures for reporting VUSs and report the first findings in the test objectively and truthfully until evidence of a gene-disease association is found. Laboratories are also encouraged to store clinical sequencing data in public databases.

Laboratories should establish clear policies regarding the re-analysis of genetic testing results and whether additional fees will be charged for re-analysis. It is recommended that laboratories regularly communicate with clinicians and inform them whether the knowledge related to VUSs has been updated. The interpretation of sequencing results must be cautious, and laboratories must maintain sufficient communication with clinicians to avoid unreasonable interpretations.

7. Conclusion

Investigating the causes of genetically heterogeneous diseases is a long-term and arduous task. NGS has overcome the limitations of traditional Sanger sequencing, which could not perform large-scale sequencing. However, from disease-related gene panels to the entire genome, while NGS generates high-quality sequence data, it also faces many challenges, including obstacles in processing large amounts of information, biological interpretation, and data reference. Future clinical applications of this technology still need to be further perfected.

8. References

[1] Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. Working Group of the American College of Medical Genetics and Genomics Laboratory Qualiy Assurance Commitee. ACMG clinical laboratory standards for next-generation sequencing[J]. Genet Med, 2013,15(9):733747.

https://doi.org/10.62852/csa/2025/154