Original Article
Diagnostic Genetics

Ann Lab Med 2026;46:94-103
https://doi.org/10.3343/alm.2025.0112
ISSN 2234-3806 • eISSN 2234-3814

ANNALS OF LABORATORY MEDICINE

Applying National Whole-genome Sequencing Findings for Rare Diseases in Clinical Practice: The Imperative of a Multidisciplinary Approach

Kyung Sun Park, M.D., Ph.D.[1], Sunghwan Shin, M.D.[2], Jong-Ho Park, Ph.D.[3], Young-Eun Kim, M.D., Ph.D.[4], Won Kyung Kwon, M.D.[5], Min-Kyung So, M.D., Ph.D.[6], Changhee Ha, M.D., Ph.D.[7], Ja-Hyun Jang, M.D., Ph.D.[8], Taeheon Lee, Ph.D.[9], Chang-Seok Ki, M.D., Ph.D.[9], Yoonjung Kim, M.D., Ph.D.[10], Kyung-A Lee, M.D., Ph.D.[10], Inho Park, Ph.D.[11,12], Sejoon Lee, Ph.D.[3], Hong-Hee Won, Ph.D.[13,14], the Genetic Testing Specialists Consortium, and Jong-Won Kim, M.D., Ph.D.[8]

[1]Department of Laboratory Medicine, Kyung Hee University Medical Center, Kyung Hee University College of Medicine, Seoul, Republic of Korea; [2]Department of Laboratory Medicine, Inje University Ilsan Paik Hospital, Goyang, Republic of Korea; [3]Precision Medicine Center, Seoul National University Bundang Hospital, Seongnam, Republic of Korea; [4]Department of Laboratory Medicine, College of Medicine, Hanyang University, Seoul, Republic of Korea; [5]U2 Medical Foundation, Seoul, Republic of Korea; [6]Department of Laboratory Medicine, Ewha Womans University College of Medicine, Seoul, Republic of Korea; [7]Department of Laboratory Medicine, Konkuk University Medical Center, Konkuk University School of Medicine, Seoul, Republic of Korea; [8]Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea; [9]GC Genome, Yongin, Republic of Korea; [10]Department of Laboratory Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea; [11]Department of Pathology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea; [12]Center for Precision Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea; [13]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea; [14]Samsung Genome Institute, Samsung Medical Center, Seoul, Republic of Korea

Background: As nationwide government-led whole-genome sequencing (WGS) projects progress, optimizing the clinical integration of large-scale WGS results is crucial. We explored how the initial analysis from Korea's First WGS Pilot Study for Rare Diseases was applied in clinical practice, and then we reanalyzed the data comprehensively at Samsung Medical Center (SMC) Seoul, Korea.

Methods: A prospective cohort study designed to collect WGS data under a Korean national initiative was conducted from August 2020 to December 2021. We focused on patients with rare diseases recruited from 16 university hospitals. The participants included 5,000 individuals (2,200 probands and 2,800 family members). The initial WGS data and diagnostic reference reports (from 682 probands and 484 family members), generated based on the First Korean WGS Pilot Study for Rare Diseases, were subsequently reanalyzed by SMC.

Results: The initial analysis of the First Korean WGS Pilot Study data revealed a diagnostic rate of 17%. Upon receiving these results, the SMC conducted two rounds of reanalysis, increasing the diagnostic rate from 15% in the first analysis, to 18% in the second, and finally to 24% in the third ($P = 1.6 \times 10^{-5}$). Key factors in improving the genetic diagnosis included increased detection of novel (likely) pathogenic variants ($P = 1.0 \times 10^{-4}$), improved diagnostic rates with larger family recruitment ($P = 0.004$), and refined clinical information for more precise genotype–phenotype correlation analysis (40%).

Conclusions: Although national WGS projects lay a foundation for rare disease diagnosis, hospital-level reanalysis and multidisciplinary collaborations are crucial for optimizing diagnostic outcomes.

Key Words: Cohort studies, Diagnosis, Family, Genotype, Hospitals, Korea, Patients, Phenotype, Rare Diseases, Whole Genome Sequencing

Received: February 26, 2025
Revision received: April 22, 2025
Accepted: August 13, 2025
Published online: September 17, 2025

Corresponding author:
Jong-Won Kim, M.D., Ph.D.
Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea
E-mail: culture.jkim@gmail.com

INTRODUCTION

Recent applications of whole-genome sequencing (WGS) in diagnosing rare diseases have demonstrated superior clinical utility over traditional methods, such as next-generation sequencing panels and whole-exome sequencing [1–4]. However, setting up and actively utilizing genome sequencing at the individual hospital level remains cost-prohibitive and poses significant challenges in handling vast amounts of genomic data [5, 6]. Concurrently, global initiatives designed to accumulate WGS data have been increasing through national projects [7–9]. Most of these projects prioritize rare diseases as their initial target [9, 10] to address previously undiagnosed cases. In light of these developments, hospitals now face the critical question of how to optimize the utilization of results derived from national-level analyses in real-world clinical practice to achieve conclusive outcomes.

In Korea, as part of the National Project of Bio Big Data (https://www.cirn.re.kr), the First Korean WGS Pilot Study for Rare Diseases (1st KWGS-RD) was conducted from August 2020 to December 2021. The study involved 5,000 individuals, including patients with rare diseases and their family members, and was designed to produce WGS data and issue diagnostic reference reports. Since its completion, this project has been referred to as the 1st KWGS-RD. We investigated how the initial analysis findings from the 1st KWGS-RD were translated into clinical practice and subsequently underwent a thorough reanalysis at Samsung Medical Center (SMC), Seoul, Korea, ultimately resulting in final genetic diagnoses. We identified key considerations that should be carefully addressed in this process.

MATERIALS AND METHODS

Participants

Sixteen university hospitals participated in the 1st KWGS-RD. All university hospitals obtained written informed consent from the participants, and the SMC Institutional Review Board (IRB) approved this study protocol (IRB approval number 2022-05-057).

Probands (who are affected) and their family members (who may or may not be affected) with rare diseases who met the following inclusion criteria were given priority registration at each participating hospital: (1) a clear family history of a rare disease but without accurate diagnoses in probands or other affected family members; (2) previous genetic testing with negative results; and (3) disease confirmation through previous tests but with atypical clinical manifestations or courses.

WGS data analysis and interpretation in the first analysis

The overall analysis workflow used in this study is shown in Fig. 1. Blood samples were collected from the probands and their family members participating in this project, and clinical information on rare diseases was collected using the Rare Disease Clinical Genomic Information Collection and Management System (https://bighug.kdca.go.kr/). At each hospital, registering the disease categories of the probands was mandatory. Additionally, the following medical observations were registered: current illness (with International Classification of Diseases, 10th Edition codes, Human Phenotype Ontology [HPO] terminology, and the disease name), family history, genetic testing history, physical examination results, laboratory test results, pathological findings, and imaging findings. Genomic DNA was extracted from the participants' blood, and WGS was performed using the Illumina NovaSeq 6000 platform (San Diego, California, USA) at three testing institutions (Fig. 1).

WGS data and clinical information for the registered probands were transmitted to the Genetic Testing Specialists Consortium (GTSC) for clinical interpretation. The GTSC consisted of 29 specialists in laboratory medicine dedicated to genetic testing. Details of the analytical pipelines for the WGS data are described in the Supplemental Data Methods. During the first analysis, these experts conducted clinical interpretations based on virtual panels for each disease category, following the 2015 American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines [11] and ClinGen specifications (https://clinicalgenome.org/). To construct the virtual panels, we reviewed existing classification systems for genetic diseases (https://panelapp.genomicsengland.co.uk/; https://www.england.nhs.uk/publication/national-genomic-test-directories/; and https://research.nhgri.nih.gov/CGD/, accessed on November 13, 2020) and ultimately selected a curated list of genes assigned to 19 disease categories (Supplemental Data Table S1). In the 1st KWGS-RD, the goal was to generate diagnostic reference reports within 25 days of WGS data generation during the study period. These reports were produced by GTSC specialists and delivered to each hospital.

Reanalysis (second and third analyses)

We applied the results from the national-scale analysis to the clinical setting of a specific hospital and analyzed the process leading to the final genetic diagnosis through a more in-depth reanalysis. For this purpose, we conducted two rounds of reanalysis (second and third analyses) using WGS data from the SMC, which had the highest number of registered probands (Fig. 1).

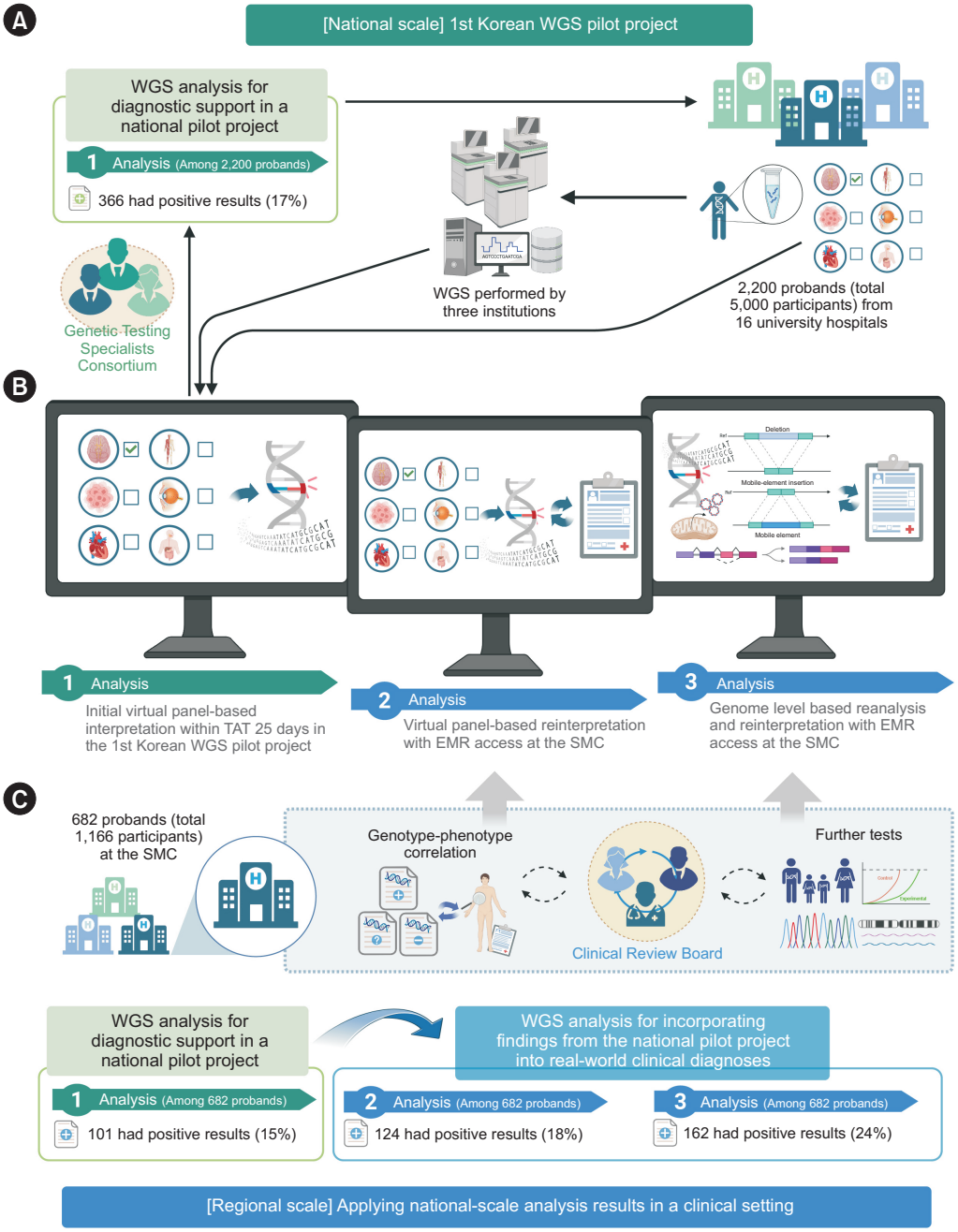

Fig. 1. Study overview. (A) Samples and clinical information for 2,200 probands with rare diseases (out of 5,000 participants) were collected at 16 university hospitals and underwent whole-genome sequencing (WGS) at three testing facilities. The WGS data and clinical information of the patients were then sent to the Genetic Testing Specialists Consortium for analysis and interpretation (first analysis, Methods section), after which diagnostic reference reports were generated and subsequently returned to the hospitals. In the First Korean Whole-Genome Sequencing Pilot Study for Rare Diseases (1st KWGS-RD), a diagnostic rate of 17% was achieved (i.e., during the initial analysis). (B) The analytical methods used for the first through third analyses are illustrated. (C) Among the 16 hospitals, the Samsung Medical Center (SMC) participated with the most patients (682 probands, 1,166 participants overall) and conducted subsequent reanalysis (second and third analyses, Supplemental Data Methods). During both rounds of reanalysis, an ongoing review of genotype–phenotype correlations was assessed by a Clinical Review Board composed of clinical experts, genetic specialists, genetic testing specialists, and clinical bioinformaticians. Based on their review, additional tests (such as Sanger sequencing, multiplex ligation-dependent probe amplification, RNA studies, and familial testing) were performed to achieve definitive interpretations. The second analysis improved the diagnostic rate to 18%, and further enhancement to 24% was achieved in the third analysis.
Abbreviations: TAT, turnaround time; EMR, electronic medical record.

https://doi.org/10.3343/alm.2025.0112

The second analysis followed the same approach used in the first analysis, i.e., interpreting coding single-nucleotide variants/small insertions and deletions (SNVs/INDELs) based on a virtual panel for each disease category. Specialists in clinical genetic variant interpretation accessed the electronic medical records (EMRs) to confirm the relevant phenotypes associated with the identified candidate genetic variants, ultimately arriving at a final diagnosis (Supplemental Data Methods). Notably, a Clinical Review Board (CRB) consisting of clinical experts, genetic specialists, genetic testing specialists, and clinical bioinformaticians reviewed the genotype–phenotype correlations. To confirm the genetic findings, additional procedures, such as Sanger sequencing, multiplex ligation-dependent probe amplification, RNA studies, familial testing, or functional studies, were performed as needed.

The third analysis involved a genome-level-based reanalysis, including copy-number variants and structural variants (CNVs/SVs), transposon elements, aberrant splicing variants at non-coding sequences, mitochondrial DNA (mtDNA) variants, and repeat expansions. The review was extended to coding SNVs/INDELs beyond the virtual panel (Supplemental Data Methods and Supplemental Data Fig. S1). This analysis was conducted for all participants enrolled in the SMC. Candidate variants identified during this stage underwent EMR review, comprehensive evaluation by the CRB, and confirmatory genetic testing where necessary.

The methods used for statistical analysis and figure generation are described in the Supplemental Data Methods section.

RESULTS

Participants

An overview of the study is presented in Fig. 1. In the 1st KWGS-RD, 5,000 participants (2,200 probands and 2,800 family members) were enrolled across 16 university hospitals (Table 1, Supplemental Data Fig. S2). Probands in three disease category groups constituted over half of all participants, with neurological and neurodevelopmental disorders (NNDDs) being most frequently represented (42%). The next most common groups were cardiovascular disorders (CVDs, 12%) and tumor syndromes (TUSs, 12%). Proband ages ranged from 0 to 88 yrs, with a mean age of 21 yrs; pediatric probands (<18 yrs) were more frequently enrolled than adult probands (≥18 yrs) (57% vs. 43%). Age and sex distributions differed significantly across disease categories ($P<1\times10^{-6}$). For example, men predominated in the CVD group (55%), male pediatric probands were prevalent in the NNDD group (43%), and women were more frequently enrolled in the TUS group (73%). Family recruitment also varied by disease category, with a greater number of family members recruited as the proband age decreased (Spearman's correlation coefficient [rho] of −0.646, $P=5.7\times10^{-260}$) (Table 1).

At the SMC, 682 probands and 484 family members participated in the 1st KWGS-RD (Supplemental Data Fig. S3, Supplemental Data Table S2). The TUS group accounted for the highest proportion of registered probands (30%), followed by the NNDD (28%) and CVD (26%) groups. Adult probands (80%) were more frequently enrolled than pediatric probands (20%), and women (53%) outnumbered men (47%). The sex distribution differed significantly by the age group ($P=0.038$), with a higher proportion of males among pediatric probands (55%) and females among adult probands (55%). Statistically significant differences were found in the distributions of age ($P=7.4\times10^{-33}$), sex ($P<1\times10^{-6}$), and family recruitment ($P<1\times10^{-6}$) with different disease categories. A negative correlation was also observed between proband age and number of recruited family members (rho = −0.715, $P=1.2\times10^{-107}$).

Initial analysis of the 1st KWGS-RD results

Sixteen batches of WGS tests were conducted with an average depth of 35.18×. On average, 91.5% of the target bases had >20× coverage. The average transition: transversion ratio was 1.94. For each individual, we obtained an average of 3,900,848 SNV and 910,478 INDEL reads.

The average turnaround time from WGS data generation to the issuance of the diagnostic reference reports was 21 days. In the first analysis of the 1st KWGS-RD, a genetic diagnosis was achieved for 17% of the 2,200 probands with rare diseases (95% confidence interval [CI]: 15–18%; Fig. 1). Of these diagnoses, 95% were based on coding SNVs/INDELs. With the initial analysis, the diagnostic rates (DRs) varied widely across the disease categories ($P=3.0\times10^{-4}$), ranging from 0–29% (Fig. 2A, Supplemental Data Table S3). Specifically, in the CVD, NNDD, and TUS groups, representing a substantial portion of the recruited probands, the DRs were 23% (95% CI: 18–30%), 18% (95% CI: 15–21%), and 10% (95% CI: 6–14%), respectively (Fig. 2).

In total, 238 causative genes were identified; 57 were reported more than once, and 181 only once, indicating a long-tailed gene distribution (Fig. 2B). The list of genes reported at least twice as causative factors for each disease category is presented in Supplemental Data Table S4. The overall DRs tended to increase with larger family structures: 15% for singletons

Table 1. Participants enrolled in the First Korean Whole-Genome Sequencing Pilot Project for Rare Diseases

Disease category	No. of probands	Family structure					Sex		Age (yrs)		
		Singleton	Duo	Trio	Quad	Quin	Male	Female	Mean	Median	Range
CVD	261	169	27	62	3	0	173	88	39.8	42	0–81
CLP	12	2	2	8	0	0	9	3	19.0	15	2–40
DED	29	11	3	13	2	0	13	16	17.9	10	0–71
DCAS	105	9	12	84	0	0	46	59	6.4	4	0–54
END	51	9	6	33	3	0	21	30	13.4	11	0–56
GAH	48	8	7	31	2	0	26	22	10.8	7.5	0–42
GRD	60	13	7	39	1	0	27	33	8.9	4	0–72
HID	105	38	13	50	4	0	56	49	17.7	14	0–73
HED	14	2	2	10	0	0	7	7	14.6	5	1–42
IND	2	1	0	1	0	0	1	1	3.5	3.5	2–5
MED	40	5	4	30	1	0	21	19	13.0	9.5	0–41
NNDD	930	158	115	639	17	1	565	365	14.7	8	0–83
OPD	92	32	17	42	1	0	50	42	34.1	36	0–78
PSD	11	0	0	10	1	0	7	4	7.0	6	4–11
RUTD	74	16	12	43	3	0	41	33	20.8	14	0–88
RED	23	11	1	11	0	0	10	13	33.4	39	0–77
RHD	37	17	3	16	1	0	18	19	21.1	18	0–55
SKD	48	8	8	31	1	0	22	26	10.8	8	0–40
TUS	257	187	20	46	4	0	55	202	41.2	43	0–82
Other	1	0	1	0	0	0	1	0	4.0	4	4
Total	2,200	696	260	1,199	44	1	1,169	1,031	21.5	13	0–88

Abbreviations: CVD, cardiovascular disorder; CLP, cytopathies; DED, dermatological disorder; DCAS, dysmorphic and congenital abnormality syndrome; END, endocrine disorder; GAH, gastroenterology and hepatology disorder; GRD, growth disorder; HID, hematological and immunological disorder; HED, hearing and ear disorder; IND, infectious disease; MED, metabolic disorder; NNDD, neurology and neurodevelopmental disorder; OPD, ophthalmological disorder; PSD, psychiatric disorder; RUTD, renal and urinary tract disorder; RED, respiratory disorder; RHD, rheumatological disorder; SKD, skeletal disorder; TUS, tumor syndrome.

(95% CI: 12–18%), 16% for duos (95% CI: 11–22%), and 18% for trios (95% CI: 16–21%) (Fig. 2C, $P=0.04$). However, the DRs differed based on family structure across different disease categories (Supplemental Data Table S3).

Strong correlations were observed between DRs based on disease category and those based on family structure (singletons vs. trios and larger family groups; rho=0.71 or 0.77, respectively), sex (rho=0.78 for both males and females), and age <18 yrs or age ≥18 yrs (rho=0.74 or 0.78, respectively) (Fig. 2D). Notably, DRs in pediatric probands showed a substantially strong correlation with trios or larger groups (rho=0.92).

Reanalysis results

In the initial analysis, the DR for the 682 probands at the SMC was 15% (95% CI: 12–18%). However, the DR increased to 18% (95% CI: 15–22%) in the second analysis and ultimately to 24% (95% CI: 20–28%) in the third analysis ($P=1.6\times10^{-5}$, Fig. 1) [12–14]. The overall average time for probands with rare diseases to receive a genetic diagnosis after experiencing symptoms and seeking medical attention was 61 months (95% CI: 53–70 months) with a median of 46 months (95% CI: 37–56 months) (Supplemental Data Fig. S4 and Supplemental Data Table S5). The changes in the results from the first to third analyses conducted at the SMC are illustrated in Supplemental Data Fig. S5. Novel pathogenic or likely pathogenic variants (PVs) identified through genome-level analyses are described in Supplemental Data Table S6.

Among the disease categories, statistically significant improvements were made in the DRs for the CVD ($P=0.007$) and NNDD groups ($P=0.015$; Fig. 3A). From the first to the third

https://doi.org/10.3343/alm.2025.0112

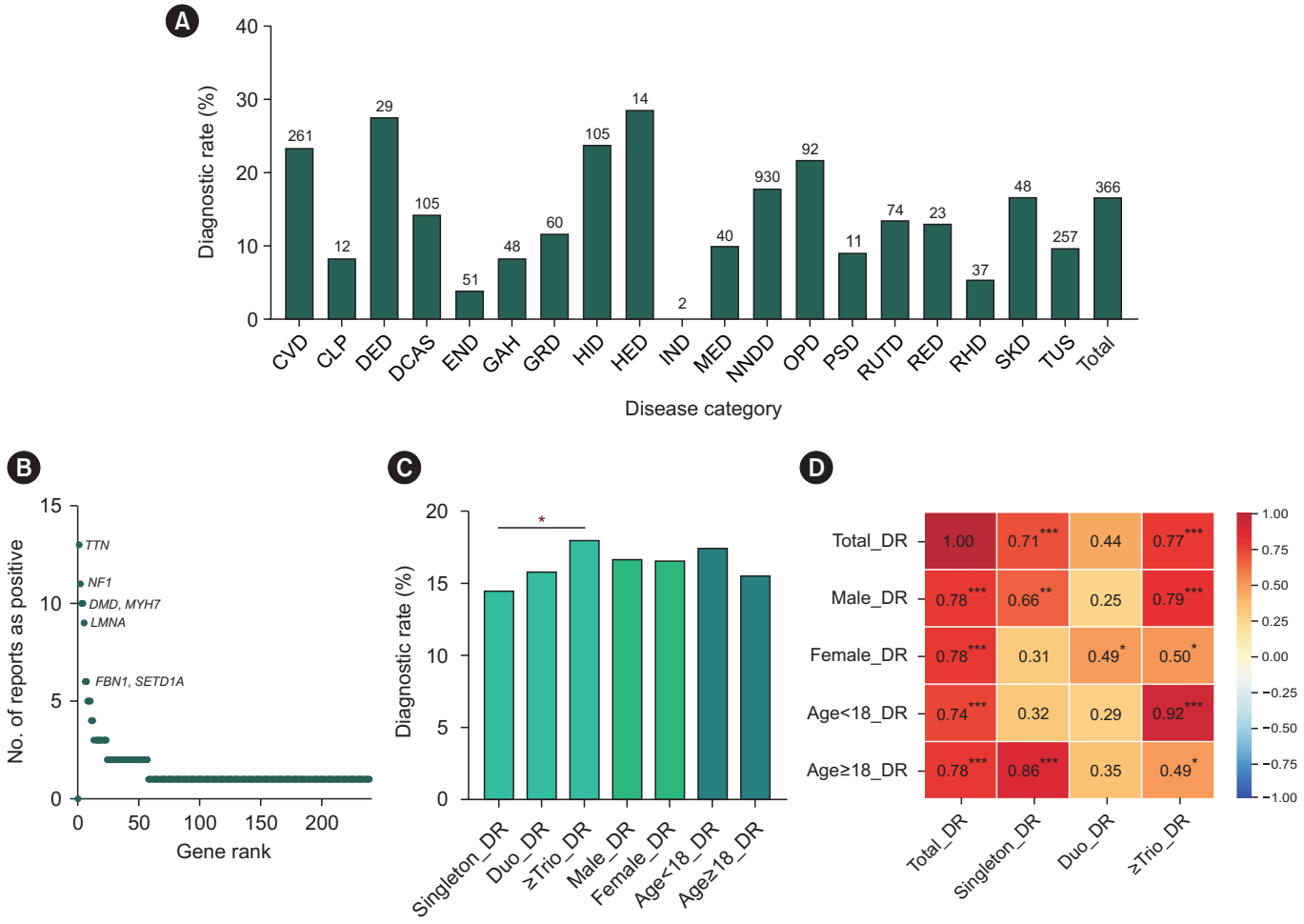

Fig. 2. Initial analysis results of the First Korean Whole-Genome Sequencing Pilot Project for Rare Diseases. (A) The histogram illustrates the DRs by disease category, ranging from 0–29%. Significant differences were observed among disease categories ($P = 3.0 \times 10^{-4}$). The values above each bar represent the number of probands enrolled. (B) The scatter plot illustrates the genes identified as causative and their frequency of occurrence during the initial analysis. (C) The histogram depicts DRs based on family structure (singletons, duos, or trios, and larger groups), sex (male or female), and age (age < 18 or ≥ 18 yrs). As the number of recruited family members increased, the DR significantly increased ($P = 0.04$). (D) The heat map illustrates correlations between DRs based on family structure, sex, and age (stratified by disease category) during the initial analysis. The numbers in the heat maps represent Spearman's correlation coefficients.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Abbreviations: DR, diagnostic rate; CVD, cardiovascular disorder; CLP, cytopathies; DED, dermatological disorder; DCAS, dysmorphic and congenital abnormality syndrome; END, endocrine disorder; GAH, gastroenterology and hepatology disorder; GRD, growth disorder; HID, hematological and immunological disorder; HED, hearing and ear disorder; IND, infectious disease; MED, metabolic disorder; NNDD, neurology and neurodevelopmental disorder; OPD, ophthalmological disorder; PSD, psychiatric disorder; RUTD, renal and urinary tract disorder; RED, respiratory disorder; RHD, rheumatological disorder; SKD, skeletal disorder; TUS, tumor syndrome.

analyses, genetic diagnoses due to novel (likely) PVs significantly increased ($P = 1.0 \times 10^{-4}$, Fig. 3B), as did those from de novo novel (likely) PVs ($P = 0.011$, Fig. 3C) (Fig. 3).

The DRs for singletons ($P = 0.007$) and family trios or larger family groups ($P = 0.001$) increased significantly after two rounds of reanalysis (Fig. 3D). The DRs for male ($P = 0.002$) and female probands ($P = 0.003$), as well as the DRs for pediatric ($P = 0.004$) and adult probands ($P = 0.001$), increased significantly after both rounds of reanalysis. In the third analysis at the SMC, the DRs significantly increased with a larger family structure; the DRs for singletons, family duos, and family trios or larger groups were 20% (95% CI: 16–25%), 24% (95% CI: 13–38%), and 31% (95% CI: 24–40%), respectively ($P = 0.004$). Unlike the initial results of the 1st KWGS-RD, the DR for pediatric probands (34%, 95% CI: 25–45%) was significantly higher than that for adult probands (21%, 95% CI: 17–25%) ($P = 0.001$). Notably, many

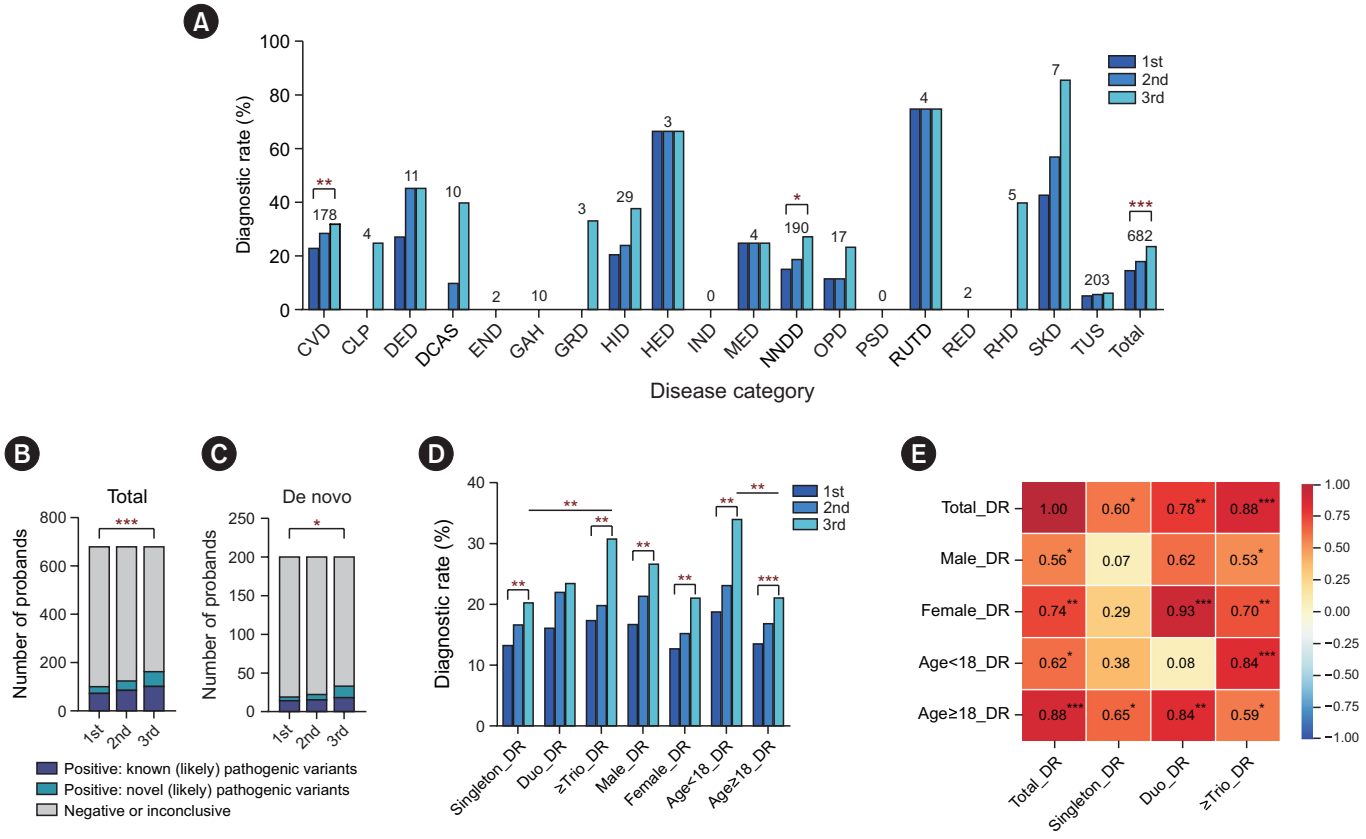

Fig. 3. Reanalyses results conducted at the Samsung Medical Center. (A) The histogram depicts the diagnostic rates (DRs) found from the first to third analyses across different disease categories. Significant differences in DRs between the first and third analyses that showed a trend are indicated with asterisks. *P<0.05; **P<0.01; ***P<0.001. (B) The histogram represents the number of probands diagnosed with a genetic disorder due to known (likely) pathogenic variants (PVs) or novel (likely) PVs from the first to third analyses. In cases where a diagnosis required two (likely) PVs in a gene, the presence of only one novel (likely) PV was considered a positive result due to a novel variant. From the first to third analyses, a statistically significant increasing trend was observed in the proportion of novel (likely) PVs (P=0.0001). (C) The histogram represents the distribution of probands genetically diagnosed due to known (likely) PVs or novel (likely) PVs resulting from de novo variants from the first to third analyses. The frequency of de novo (likely) PVs significantly increased as the reanalysis progressed (P=0.011). (D) The histogram depicts DRs based on family structure (singletons, duos, or trios, and larger groups), sex (male or female), and age (age <18 or ≥18 yrs) from the first to third analyses. Significant differences that showed a trend in DRs from the first to third analyses are indicated with asterisks. **P<0.01; ***P<0.001. (E) The heat map illustrates the correlation between DRs based on family structure, sex, and age (stratified by disease category) during the third analysis.

Abbreviations: DR, diagnostic rate; PV, pathogenic variant; CVD, cardiovascular disorder; CLP, cytopathies; DED, dermatological disorder; DCAS, dysmorphic and congenital abnormality syndrome; END, endocrine disorder; GAH, gastroenterology and hepatology disorder; GRD, growth disorder; HID, hematological and immunological disorder; HED, hearing and ear disorder; IND, infectious disease; MED, metabolic disorder; NNDD, neurology and neurodevelopmental disorder; OPD, ophthalmological disorder; PSD, psychiatric disorder; RUTD, renal and urinary tract disorder; RED, respiratory disorder; RHD, rheumatological disorder; SKD, skeletal disorder; TUS, tumor syndrome.

similarities were observed in the correlation of DRs between the third analysis conducted at the SMC and the first analysis of the national pilot project (Fig. 3E), particularly the correlation between the DRs for trios or larger groups and pediatric probands (rho=0.84).

We analyzed factors contributing to additional genetic diagnoses during our reanalyses (Table 2). The primary reason for the higher DR in the second analysis than in the first was that the clinical information registered during the first analysis was often nonspecific or unclear (71%). In contrast, the main factor driving the increased DR in the third analysis was the comprehensive analysis of large variants (CNVs/SVs, 34%). Additionally, through multiple rounds of review by the CRB, we resolved issues related to clinical information (21%) and ambiguous interpretations (16%). A review of genes excluded from the previous virtual panel-based analysis led to additional genetic diagnoses, accounting for 13% of panel-related issues. By comparing the first analysis with the third, we confirmed that meticulously analyzing

https://doi.org/10.3343/alm.2025.0112

Table 2. Factors contributing to additional genetic diagnoses in reanalyses

Variant category	Factors	Reanalysis vs. previous analysis, N (%)		
		2nd vs. 1st (N=24)	3rd vs. 2nd (N=38)	3rd vs. 1st (N=62)
Small variants in the coding sequence and its adjacent region	Panel issues	0 (0)	5 (13)	5 (8)
	Ambiguous interpretations	6 (25)	6 (16)	12 (19)
	Issues of clinical information	17 (71)	8 (21)	25 (40)
Large variants	CNVs/SVs	0 (0)	13 (34)	13 (21)
Genetic variants in non-coding sequence	Alternative splicing variants	1 (4)	3 (8)	4 (6)
	mtDNA variants	0 (0)	2 (5)	2 (3)
	Transposable elements	0 (0)	1 (3)	1 (2)

Abbreviations: CNVs/SVs, copy-number variants and structural variants; mtDNA, mitochondrial DNA.

small variants within or near the coding sequence (68%) was crucial (Table 2).

DISCUSSION

Worldwide initiatives are implementing WGS as part of national projects [7–9], helping establish its clinical utility. When conducting national WGS projects, efficient and accurate analysis of large volumes of genomic data is crucial. At the forefront of national initiatives, such as one conducted in the United Kingdom, an initial analysis based on a virtual panel was employed for diagnosing rare genetic diseases [9, 10]. Subsequently, the Exomiser application, which utilizes HPO terms [15–17], was employed for further analysis [9]. In a national pilot project targeting rare diseases conducted in the UK, a DR of 16% was reported using virtual panel-based analysis [10]. Subsequent studies utilized both virtual panels and Exomiser for WGS analysis, resulting in a DR of 25% [9]. We achieved a DR of 17% during the first analysis of the 1st KWGS-RD, and the DR obtained after two rounds of reanalysis at the SMC ranged from 15–24% ($P=1.6\times10^{-5}$), similar to the results of the UK WGS pilot project. Similar to previous results [9, 18], we demonstrated variability in the DR depending on the recruited disease cohort.

Virtual panel-based analysis requires only basic disease category information, enabling clinicians to register patients without standardized phenotypes. However, this situation can lead to inconsistencies in clinical data quality and relies on known genetic variants, which limits the scope of discoveries. Breaking virtual panel boundaries requires systematic input of diverse phenotype information [19], favoring HPO-based gene selection over disease names. The outcome depends on the accuracy of the HPO input, emphasizing the need for clinician efforts and standardized education on HPOs. Without proper training, interpretations of the results of national WGS projects may be limited. In the 1st KWGS-RD, variations occurred in the HPOs registered for patients across the participating university hospitals, which posed limitations when applying them for a comprehensive analysis; therefore, the first analysis utilized a virtual panel-based approach for efficient and rapid analysis (averaging 21 days). The main goal of the 1st KWGS-RD was to quickly provide diagnostic reference reports to support patient care, with genome-level analyses performed only upon request.

We focused on the methodological aspects of how the WGS data and analysis results produced by this national project could be applied in a clinical setting to treat patients. We consider this a crucial issue that must be addressed, as national-level WGS data accumulation is underway worldwide. The two rounds of WGS reanalysis conducted at the SMC can be considered good models for addressing these concerns. The second analysis serves as a model for deriving optimal genetic diagnoses in hospital environments, where, despite receiving annotated genetic data results from the national project, practical constraints prevented independent handling of the WGS data and establishing various analysis pipelines internally. The third analysis represents a model for identifying genetic causes through reanalysis in hospital environments capable of establishing various WGS analysis pipelines independently.

Through our reanalyses, we discovered that multidisciplinary collaboration [20] is not an option but a necessity when reviewing and applying the analyzed WGS results obtained from the national project to individual patients. Obtaining additional genetic diagnoses is important; however, verifying their correlation with the patient's phenotype is even more critical. Indeed, during reanalysis at the SMC, we encountered a false-positive result in the initial analysis, which relied solely upon registered clinical information (Supplemental Data Fig. S5). Rather than diversify-

ing pipelines to analyze various aspects of the genome, meticulously analyzing small variants located in the coding sequence was crucial. For example, 88% of all detected (likely) PVs were identified, although they were in fact (likely) pathogenic but overlooked in previous analyses based on correlations with phenotypes. Of the 42 small variants detected in the coding sequence and adjacent regions, the majority (88%) were attributed to ambiguous interpretations (N=12) and issues regarding clinical information (N=25) (Table 2).

In the second analysis, meticulous EMR review improved DRs without genome-level analysis, largely owing to in-depth CRB discussions. As current interpretations often focus on European variants, non-European populations tend to have higher rates of variants of uncertain significance (VUSs) [21, 22]. When novel variants specific to Koreans are found, additional studies (e.g., segregation, family testing, or functional studies) are often required to resolve VUSs according to the 2015 ACMG/AMP guidelines [11] or ClinGen specifications. Prioritizing VUSs and conducting further investigations within the CRB are key steps for resolving these uncertainties.

With rare diseases, unexpected clinical diagnoses are often made through genotype-driven analysis [23]. Comprehensive analysis of the entire genome is a primary objective of conducting WGS testing, and such analyses ultimately help improve DRs [24]. However, such genome-level analysis remains challenging. Currently, various pipelines exist for CNVs/SVs, aberrant splicing variants, mtDNA, and transposon elements; however, integrated or standardized specifications for utilizing such diverse pipelines are lacking [25]. Identifying causative variants among numerous candidate variants derived from each analysis is not straightforward. Genetic analysts strive to retain only the most likely candidate variants through optimal analysis, whereas clinical experts in genotype–phenotype matching are needed to review these candidate variants. Additional tests were conducted to strengthen the evidence for the pathogenicity of the selected candidate variants. Furthermore, exploring whether overlooked answers exist is necessary when using different analysis pipelines for various applications. Confirmatory tests (such as multiplex ligation-dependent probe amplification or RNA sequencing) might be prioritized when an institutional diagnostic laboratory has the necessary infrastructure, especially for candidate variants related to CNVs or alternative splicing. However, when functional validation through *in vitro* or *in vivo* studies is required, the timeline for confirming a diagnosis becomes unpredictable, and the process demands considerable resources.

Finally, we did not examine the regulatory effects of non-coding sequences or perform analyses of repeat expansions, highlighting the need for such studies [25, 26]. Additionally, more research-oriented analyses [9, 10] of disease manifestations caused by novel genes in patients with characteristic phenotypes and family histories should be conducted.

CONCLUSIONS

In the WGS era, continuous reanalysis is necessary with each update in clinical information and analysis methods. A multidisciplinary approach in real-world clinical settings is crucial for accurate interpretations. Without employing such an approach, identifying genetic answers from WGS data alone can be challenging.

SUPPLEMENTARY MATERIALS

Supplementary materials can be found via https://doi.org/10.3343/alm.2025.0112.

ACKNOWLEDGEMENTS

The authors thank all experts who participated in the GTSC.

AUTHOR CONTRIBUTIONS

Park KS and Kim JW conceived and designed the study. The GTSC conducted the primary interpretation of the WGS data. Park KS, Shin S, Kim YE, Kwon WK, So MK, Ha C, and Jang JH reanalyzed the WGS data. Shin S, Kwon WK, So MK, Ha C, Jang JH, and Kim JW performed the clinical work. Park KS, Kim Y, Lee KA, Ki CS, Park I, and Kim JW designed the workflow for the data analyses. Park JH, Lee T, Park I, Lee S, and Won HH performed computational analyses. Park KS, Shin S, Park JH, Kim YE, and Kim JW wrote the manuscript with input from all authors.

CONFLICTS OF INTEREST

None declared.

RESEARCH FUNDING

This study was supported by grants from the National Research Foundation of Korea (NRF), funded by the Korean Government (MSIT) (grant numbers 1711180477 and 2022R1A2C2091571).

REFERENCES

1. Chung CCY, Hue SPY, Ng NYT, Doong PHL, Hong Kong Genome Project, Chu ATW, et al. Meta-analysis of the diagnostic and clinical utility of exome and genome sequencing in pediatric and adult patients with rare diseases across diverse populations. Genet Med 2023;25:100896.
2. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nat Rev Genet 2018;19:253–68.
3. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genet Med 2018;20:435–43.
4. Schobers G, Derks R, den Ouden A, Swinkels H, van Reeuwijk J, Bosgoed E, et al. Genome sequencing as a generic diagnostic strategy for rare disease. Genome Med 2024;16:32.
5. Hayeems RZ, Dimmock D, Bick D, Belmont JW, Green RC, Lanpher B, et al. Clinical utility of genomic sequencing: a measurement toolkit. NPJ Genom Med 2020;5:56.
6. Marshall CR, Bick D, Belmont JW, Taylor SL, Ashley E, Dimmock D, et al. The Medical Genome Initiative: moving whole-genome sequencing for rare disease diagnosis to the clinic. Genome Med 2020;12:48.
7. Stark Z, Dolman L, Manolio TA, Ozenberger B, Hill SL, Caulfied MJ, et al. Integrating genomics into healthcare: a global responsibility. Am J Hum Genet 2019;104:13–20.
8. All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program. Nature 2024;627:340–6.
9. 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 Genomes pilot on rare-disease diagnosis in health care — preliminary report. N Engl J Med 2021;385:1868–80.
10. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, et al. Whole-genome sequencing of patients with rare diseases in a national health system. Nature 2020;583:96–102.
11. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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. Genet Med 2015;17:405–24.
12. Shin S, Lee J, Kim YG, Ha C, Park JH, Kim JW, et al. Genetic diagnosis of children with neurodevelopmental disorders using whole genome sequencing. Pediatr Neurol 2023;149:44–52.
13. Kim YG, Kwon H, Park JH, Nam SH, Ha C, Shin S, et al. Whole-genome sequencing in clinically diagnosed Charcot–Marie–Tooth disease undiagnosed by whole-exome sequencing. Brain Commun 2023;5:fcad139.
14. Ha C, Kim D, Bak M, Park JH, Kim YG, Jang JH, et al. *CRYAB* stop-loss variant causes rare syndromic dilated cardiomyopathy with congenital cataract: expanding the phenotypic and mutational spectrum of alpha-B crystallinopathy. J Hum Genet 2024;69:159–62.
15. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine JP, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res 2019;47:D1018–27.
16. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The Human Phenotype Ontology in 2021. Nucleic Acids Res 2021;49:D1207–17.
17. Smedley D, Jacobsen JO, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc 2015;10:2004–15.
18. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet 2015;47:717–26.
19. Thompson R, Papakonstantinou Ntalis A, Beltran S, Töpf A, de Paula Estephan E, Polavarapu K, et al. Increasing phenotypic annotation improves the diagnostic rate of exome sequencing in a rare neuromuscular disorder. Hum Mutat 2019;40:1797–812.
20. Macken WL, Falabella M, McKittrick C, Pizzamiglio C, Ellmers R, Eggleton K, et al. Specialist multidisciplinary input maximises rare disease diagnoses from whole genome sequencing. Nat Commun 2022;13:6324.
21. Chen E, Facio FM, Aradhya KW, Rojahn S, Hatchell KE, Aguilar S, et al. Rates and classification of variants of uncertain significance in hereditary disease genetic testing. JAMA Netw Open 2023;6:e2339571.
22. Venner E, Patterson K, Kalra D, Wheeler MM, Chen YJ, Kalla SE, et al. The frequency of pathogenic variation in the All of Us cohort reveals ancestry-driven disparities. Commun Biol 2024;7:174.
23. Austin-Tse CA, Jobanputra V, Perry DL, Bick D, Taft RJ, Venner E, et al. Best practices for the interpretation and reporting of clinical whole genome sequencing. NPJ Genom Med 2022;7:27.
24. Pagnamenta AT, Camps C, Giacopuzzi E, Taylor JM, Hashim M, Calpena E, et al. Structural and non-coding variants increase the diagnostic yield of clinical whole genome sequencing for rare diseases. Genome Med 2023;15:94.
25. Ellingford JM, Ahn JW, Bagnall RD, Baralle D, Barton S, Campbell C, et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. Genome Med 2022;14:73.
26. Shim Y, Seo J, Lee ST, Choi JR, Choi YC, Shin S, et al. Clinical application of optical genome mapping for molecular diagnosis of facioscapulohumeral muscular dystrophy. Ann Lab Med 2024;44:437–45.