Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Real-World Clinical Utility of Genome Sequencing Following the National Expansion of the Rare Disease Diagnostic Support Infrastructure in Korea

Seokhui Jang
3billion
Seung Hwan Oh
Pusan National University School of Medicine, Pusan National University Yangsan Hospital
Go Hun Seo
3billion
Jinhong Shin
Pusan National University School of Medicine, Pusan National University Yangsan Hospital
Jee Young Kim
Korea Disease Control and Prevention Agency
Seonghee Yoon
Korea Disease Control and Prevention Agency
Ye Eun Lee
Pusan National University
Hoon Kook
Chonnam National University Hwasun Hospital
Dae Hyun Jang
The Catholic University of Korea Incheon St. Mary's Hospital
Juwon Kim
Yonsei University Wonju College of Medicine
Jiye Ahn
Chosun University College of Medicine
Anna Cho
Seoul National University Bundang Hospital
Jinkyu Kim
Jeonbuk National University School of Medicine
Jeesuk Yu
Dankook University Hospital
Soonhak Kwon
Kyungpook National University School of Medicine & KNU Children's Hospital
Yena Lee

Hallym University Sacred Heart Hospital

Bo Lyun Lee

Busan Paik Hospital, Inje University College of Medicine

Do Hoon Kim

Keimyung University, Dongsan Medical Center

Jun Hwa Lee

Samsung Changwon Hospital, Sungkyunkwan University School of Medicine

Yoo Mi Kim

Chungnam National University Sejong Hospital, College of Medicine, Chungnam National University

Eungu Kang

Korea University Ansan Hospital, Korea University College of Medicine

Young Lim Shin

Soonchunhyang University Bucheon Hospital, Soonchunhyang University College of Medicine

Kyung Yeon Lee

Ulsan University Hospital, University of Ulsan College of Medicine

Mijin Hong

Konyang University Hospital, Konyang University College of Medicine

Jung Sook Yeom

Gyeongsang National University, College of MedicineGyeongsang National University Hospital

Hyo Jeong Kim

Gachon University Gil Medical Center, Gachon University College of Medicine

Jon Soo Kim

Chungbuk National University Hospital, Chungbuk National University College of Medicine

Han Hyuk Lim

Chungnam National University Hospital

Jin Kyung Kim

Daegu Catholic University School of Medicine

Young Mi Kim

Pusan National University Hospital and Biomedical Research Institute, Pusan National University School of Medicine

Geun Young Chae

3billion

Kisang Kwon

3billion

Seung Woo Ryu

3billion

Dongseok Moon

3billion

Ki Woong Kwon

3billion

Sohyun Lee

3billion

Kieop Park

3billion

Esther Youn

3billion

Yoo hyun Kim

3billion

Hane Lee

3billion

Chong Kun Cheon

chongkun@pusan.ac.kr

Pusan National University School of Medicine, Pusan National University Yangsan Hospital

---

Article

Keywords:

Posted Date: October 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11228887/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Establishment of an advanced, in-region one-stop management system for rare diseases fundamentally begins with accurate local diagnosis. However, equipping individual medical institutions with advanced diagnostic equipment presents significant practical challenges in terms of cost and efficiency. Government-led policy and administrative-economic support that enables rare disease patients to be diagnosed in their local communities, eliminating the need for long-distance travel to the capital area, forms a crucial foundation for national rare disease management. In line with this stance, the Korea Disease Control and Prevention Agency (KDCA) launched the Korean Regional Rare Disease Diagnostic Support Program (KR-RDSP) in 2023, providing genome sequencing (GS)-based diagnostics to patients residing outside the Capital area. Building upon the foundational success and positive outcomes of this regional initiative, the program underwent a significant expansion starting in 2025. This nationwide scaling integrated a total of 34 advanced general hospitals and specialized rare disease centers—now encompassing select regions within the Capital area as well—thereby strengthening the diagnostic infrastructure to ensure more equitable and comprehensive diagnostic services for all rare disease patients nationwide.

Methods

We analyzed the three-year outcomes and clinical utility of this program from 2023 to 2025. A total of 1,620 participants were enrolled across 34 regional rare disease centers (annual enrollment: 400 in 2023, 410 in 2024, and 810 in 2025) in Korea.

Results

The median age at enrollment was 9 years (mean: 13.5 years, SD: 14.8 years). Among the 1,620 participants, 1,234 (76.17%) had documented information on age of symptom onset. Symptom onset was predominantly in the early childhood period: 39.44% (639/1,620) presented with symptom onset before 2 years of age (neonatal and infancy), and 23.52% (381/1,620) between 2 and 11 years. Among these 1,234 participants, the mean interval from symptom onset to testing was 7.00 years. The overall diagnostic yield reached 35.0% (567/1,620). Notably, 20.3% of diagnosed cases provided clinically actionable findings, influencing therapeutic strategies and genetic counseling. Approximately 9.03% of all reported variants (6.35% of positively diagnosed cases) were detectable exclusively through GS, confirming its superior sensitivity over conventional sequencing methods. Family testing and systematic reanalysis utilizing AI tools identified 43 additional diagnoses, resulting in a 2.0 percentage point absolute increase in the overall diagnostic yield (from 32.96% to 35.0%). The centralized workflow achieved an efficient mean turnaround time (TAT) of 27.5 days, a significant reduction compared to the typical referral systems (>6 months) in Korea. This efficiency correlated with high stakeholder satisfaction (91.4-97.7%). However, significant variation in inter-center diagnostic yield (11.1-51.9%, p=0.007) suggests a need for further standardization of clinical workflows.

Conclusions

The national RDSP successfully demonstrates the clinical utility, scalability, and efficiency of a centralized GS-based diagnostic model in dismantling systemic barriers to precision medicine.

By providing high diagnostic accuracy and significantly reducing the time to diagnosis, this national program establishes GS as a primary diagnostic tool for rare disease patients. These findings provide further supporting evidence that a government-led GS program should continue to be expanded for the national healthcare system.

Background

Under the Rare Disease Management Act in South Korea, rare diseases are defined as conditions affecting fewer than 20,000 individuals or those with an unknown prevalence due to diagnostic challenges [1]. Although each individual condition affects only a small number of patients, more than 7,000 distinct conditions have been defined as rare diseases worldwide [2]. To diagnose this broad spectrum of conditions, diagnostic testing strategies have evolved considerably, progressing from karyotyping to next-generation sequencing (NGS). Despite these advancements, a vast number of patients remain without a definitive molecular explanation for their symptoms, enduring a prolonged 'diagnostic journey' often driven by a lack of testing and treatment opportunities and compounded by years of emotional strain.

Exome sequencing (ES) and chromosomal microarray (CMA) have been widely adopted, yet each carries intrinsic limitations. ES is constrained to protein-coding regions, which represent only approximately 1-2% of the human genome, and therefore fails to detect deep intronic variants, non-exonic repeat expansions, small intragenic copy number variants (CNVs), and complex structural rearrangements in regions of low mappability. CMA, though effective for large genomic imbalances, cannot detect balanced structural rearrangements, single-nucleotide variants (SNVs), or small insertions/deletions (INDELs). Genome sequencing (GS) is a comprehensive genomic testing strategy, capable of identifying all classes of clinically relevant variants in a single assay [3,4]. Importantly, GS has been shown to diagnose approximately 7-9% of patients through variants that are exclusively detectable by this modality [5–7].

In South Korea, because more than 50% of the population resides in the Capital area [8], advanced genomic infrastructure has remained highly centralized. Attempting to resolve this by equipping every regional hospital with independent ES/GS facilities, however, would be highly inefficient. Indeed, the healthcare gap between the capital and non-capital regions is not merely a result of these technological deficits; rather, it is largely driven by patients' strong preference for major metropolitan centers, often stemming from informational asymmetry. Because patients tend to continue their treatment at the site of diagnosis, this trend creates a vicious cycle of centralization that disrupts the national healthcare delivery system. Furthermore, this travel heavily burdens nearly 47% of rare disease patients outside the capital [9], prolonging the diagnostic journey and causing financial hardship.

To address this inequity and correct the structural imbalance in the medical delivery system, outreach diagnostic programs are designed to keep patients within their local healthcare networks, rather than simply compensating for a perceived deficit in regional clinical quality. In this context, the Korea Disease Control and Prevention Agency (KDCA) initiated the Korean Regional Rare Disease Diagnostic Support Program (KR-RDSP) in 2023. The program was designed as a pilot to evaluate whether high-quality, GS-based genomic diagnosis could be delivered efficiently to patients residing outside the Capital area through a centralized testing workflow. A pilot report covering the first year of the program (400 participants enrolled from June to December 2023) demonstrated an overall diagnostic yield of 36.3%, with satisfactory turnaround times and strong regional center satisfaction [10].

Building on this foundation, the program was substantially expanded in subsequent years, increasing from 11 participating centers and 400 samples in 2023 to 34 centers and 810 samples in 2025, thereby strengthening the infrastructure of the RDSP in Korea to ensure more equitable and comprehensive diagnostic services for all rare disease patients nationwide (Table 1). The primary objective of this study was to evaluate the clinical utility, scalability, and equity impact of the national RDSP. Here we present the first comprehensive three-year multicenter report of the KR-RDSP, encompassing 1,620 participants analyzed between 2023 and 2025. We characterize the overall diagnostic yield (including reanalysis), the spectrum of identified variants, the unique added value of GS for detecting variants undetectable by other modalities, the impact of family-based segregation analysis and periodic reanalysis, the prevalence of East Asian-specific disease-causing variants within this cohort, clinical actionability of molecular diagnoses, and illustrative case examples that highlight the distinctive contribution of GS to molecular diagnostics in this population.

Methods

Referrals and Sample Collection

The national RDSP was supported by the Division of Rare Disease Management of the KDCA from April 2023 to December 2025. Led by Pusan National University Yangsan Hospital, the number of participating regional rare disease centers increased from 11 in 2023 to 34 in 2025, spanning 12 metropolitan areas/provinces by the end of the study period. The KDCA established a centralized, secure online portal to order and receive test results. Peripheral blood samples were collected in EDTA tubes. In cases where blood sampling was unfeasible, alternative specimens were obtained, including dried blood spots (DBS; n=5), and DNA extracted from tissue (n=1). Samples were transported to the reference laboratory for GS by courier service within 2 days of collection. The TAT was defined as the period from the date of sample reception at the reference laboratory to the date the GS report was released to the ordering physician.

Genome Sequencing

Genomic DNA was extracted from the specimens using the QIAamp DNA Blood Mini Kit (Qiagen, version 4.2) or MagMAX™ DNA Multi-Sample Ultra 2.0 Kit (Thermofisher) and the sequencing library was

prepared using the TruSeq DNA PCR-Free kit (Illumina, San Diego, CA). The sequencing was performed using NovaSeq 6000 (Illumina) in 2023 and NovaSeq X system (Illumina) in 2024-2025 with 150 bp paired-end sequencing.

Sequencing Data Processing and Variant Calling

Sequencing data analysis was performed using a bioinformatics pipeline built with various publicly available tools.

For samples sequenced on the NovaSeq 6000 system (2023), binary base call (BCL) files were converted to FASTQ files using bcl2fastq (v2.20.0.422; Illumina). Sequence reads were aligned to the Genome Reference Consortium Human Build 38 (GRCh38) [11] and Revised Cambridge Reference Sequence (rCRS) [12] of the mitochondrial genome using BWA-mem2 (v2.2.1) [13], generating BAM (Binary Alignment Map) files. BAM files were subsequently sorted and recalibrated using SAMtools (v1.18) [14] and the Genome Analysis Toolkit (GATK; v4.4.0.0) [15], respectively.

For samples sequenced on the NovaSeq X system (from 2024 onward), read alignment was performed using the DRAGEN Bio-IT Platform (Illumina). Furthermore, starting in July 2025, alignment was updated to utilize a masked version of the GRCh38 reference, in which falsely duplicated regions of chromosome 21 were masked to N's based on the masking file developed in collaboration with the Genome Reference Consortium (GRC) [16].

Variant calling was performed using the following tools: GATK (v4.4.0.0) [15] for single nucleotide variants (SNVs) and small insertions and deletions (INDELs); internally developed software (3bCNV) for copy number variants (CNVs); Manta (v1.6.0) [17] for structural variants (SVs); ExpansionHunter (v5.0.0) [18] for repeat expansions; Mutect2 (v4.4.0.0) [15] for mitochondrial variants; and the Mobile Element Locator Tool (MELT, v2.2.2) [19] for mobile element insertions (MEIs). Aneuploidy was predicted based on depth-of-coverage (DOC) information, and regions of homozygosity (ROH) were detected using AutoMap (v1.3) [20].

Sequencing quality was assessed using multiple metrics, including the mapping rate, percentage of the genome covered at a minimum of 20× depth-of-coverage, total variant counts, heterozygous/homozygous variant ratio, transition/transversion (ti/tv) ratio, and sample contamination estimated by freemix. All samples were confirmed to meet quality thresholds prior to downstream analysis.

Variant Prioritization and Classification

Variant filtering, annotation, and prioritization were performed using EVIDENCE (v3.2) [21], an automated variant interpretation system developed by 3billion, Inc. (Seoul, Korea). Variants were annotated using the Variant Effect Predictor (VEP) [22], with v104.2 used from April 2023 to August 2025, and v112 used

from September 2025 onward. The selection of representative transcripts was based on the Matched Annotation from NCBI and EMBL-EBI (MANE) or canonical transcript information from the RefSeq database (and the Ensembl database when not available). In cases without canonical transcript annotations, the transcript with the most severe predicted consequence was selected.

SNVs and INDELs with an allele frequency (AF) >5% in gnomAD v3.1 (https://gnomad.broadinstitute.org/)[23], CNVs with an AF >1% in gnomAD SVs v2.1 (https://gnomad.broadinstitute.org/) and the Database of Genomic Variants (DGV) gold standard (https://dgv.tcag.ca/), and SVs with an AF >1% in the 3billion internal database were removed. Variants previously reported in the literature or classified as pathogenic (P) or likely pathogenic (LP) in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) were excluded from frequency-based filtering. The remaining variants in both coding and non-coding regions were classified as P, LP, variant of uncertain significance (VUS), likely benign (LB), or benign (B) in accordance with the ACMG and the Association for Molecular Pathology (AMP) variant interpretation guidelines [24–26], with internal customizations. Three *in silico* prediction tools were applied to assess the pathogenicity of missense variants and potential splicing effects: the Rare Exome Variant Ensemble Learner (REVEL) [27], 3Cnet [28], and SpliceAI [29]. A symptom similarity score was calculated for each gene harboring a candidate variant using the Human Phenotype Ontology (HPO) framework [30], adapting the depth-based similarity method described by Köhler et al [31]. Variants classified as LB or B were not reported. For repeat expansion variants, a variant was considered P if the repeat number met or exceeded the pathogenic threshold defined in the literature and the Online Mendelian Inheritance in Man (OMIM) database [32]. MEIs identified by MELT were considered potentially disease-causing if located within a coding exon or a disease-associated intronic region.

Variant Reporting

Report outcomes were classified as positive, inconclusive, or negative. A positive report required one or more disease-causing variants: for autosomal dominant (AD) or X-linked (XL) diseases, one heterozygous or hemizygous P/LP variant in a known disease gene consistent with the patient's phenotype; for autosomal recessive (AR) diseases, one homozygous P/LP variant or two compound heterozygous P/LP variants in a relevant gene. An inconclusive report was issued when: (1) a heterozygous or hemizygous VUS was identified in a known AD or XL disease gene; (2) potential compound heterozygous variants were found consisting of one P/LP variant and one VUS, a homozygous VUS, or only one P/LP variant in a known AR disease gene; or (3) P/LP variants were identified in a gene not yet associated with a defined disease in the OMIM and Orphanet but reported with sufficient evidence in the literature. A negative report was issued when no clinically significant variants were identified.

All candidate variants were manually reviewed in the Integrative Genomics Viewer (IGV v2.9.4) [33] to verify technical validity. Final variant selection and clinical interpretation were performed by clinical molecular geneticists and clinical geneticists, who evaluated candidate variants in the context of each patient's clinical information and selected the most clinically significant variants for reporting. Variants

were named according to the standard nomenclature of the Human Genome Variation Society (HGVS; www.hgvs.org).

Variant Confirmation and Family Testing

All diagnostic P/LP variants, specifically SNVs, small INDELs, and SVs with identifiable breakpoints, were confirmed by Sanger sequencing prior to reporting. PCR primers were designed using Primer3 (v0.4.0) [34,35] and NCBI GenBank reference sequences [36]. PCR amplification and Sanger sequencing were performed using the PCR Master Mix Kit (Thermo Fisher Scientific, Waltham, MA, USA) and the SeqStudio Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Sequencing results were analyzed using Sequence Scanner Version 1.0 (Applied Biosystems).

Validation of SVs lacking identifiable breakpoints was delegated to the respective regional centers, with recommendations for CMA, MLPA, karyotyping, or FISH. Family segregation analysis was exclusively performed using appropriate orthogonal methods.

Satisfaction Survey

A satisfaction survey was conducted to evaluate the program. While the survey initially targeted clinicians, it was expanded to include patients from 2024. In cases where the proband was unable to participate in the survey, a family member completed it on their behalf. Among clinicians, 10 of 31 (32.3%) responded in 2023, 24 of 34 (70.6%) in 2024, and 43 of 66 (65.2%) in 2025. Patient surveys were introduced in 2024, with 80 of 410 participants (19.5%) responding across 9 centers in that year, and 260 of 810 participants (32.1%) responding across 20 centers in 2025.

The questionnaire items were tailored to each participant group. For clinicians, the survey assessed five key domains: (1) satisfaction with diagnostic test results, (2) satisfaction with genetic and Sanger test results, (3) helpfulness of results for clinical management, (4) convenience of the program process, and (5) the necessity of program continuation.

For patients and their families, the survey comprehensively evaluated ten domains: (1) convenience of the program process, (2) convenience of location, (3) program promotion and information guidance, (4) reliability of the testing process, (5) satisfaction with waiting time for test results, (6) satisfaction with diagnostic test results, (7) satisfaction with counseling on additional tests (e.g., family testing), (8) alleviation of financial burden, (9) overall satisfaction with the program, and (10) the necessity of program continuation.

Statistical Analysis

Statistical analyses were performed using R (version 4.3.0) and Microsoft Excel (Microsoft, Redmond, WA, USA). The diagnostic yield was expressed as a proportion with a 95% confidence interval (CI), calculated using the Wilson score interval method.

Comparisons of diagnostic yields between subgroups-including sex, centers, and testing history (first-tier versus previously tested cohort)-were performed using the Chi-square test of independence, or Fisher's exact test where appropriate.

To evaluate whether the differences in diagnostic yields across disease categories were driven by the variation in the number of clinical features per patient, a multivariable logistic regression analysis was performed at the individual patient level. The model included the patient's diagnostic outcome (positive vs. inconclusive/negative) as the dependent variable, with the disease category and the individual number of clinical features as independent variables. To avoid complete separation in the logistic model, categories with zero diagnostic yields ('Hearing and ear disorders', 'Respiratory disorders' and 'Autoimmune or Rheumatological disorders') were excluded from this analysis. A *p*-value of less than 0.05 was considered statistically significant.

The interval between age at symptom onset and age at testing (enrollment) was calculated for all participants. Participants with unknown age of symptom onset (n=386) were excluded from this calculation, yielding an analytic sample of 1,234 participants. Among participants who received a positive diagnosis, the diagnostic journey was defined as the interval between age at symptom onset and age at diagnosis. For cases diagnosed at initial analysis, age at testing was defined as age at enrollment; for cases whose diagnosis was established through subsequent family testing or systematic reanalysis, age at testing was defined as age at the time of reanalysis. This calculation was restricted to the 432 positively diagnosed participants with documented age of symptom onset.

To determine the East Asian specificity of each variant, two independent Fisher's exact tests were performed comparing East Asian versus non-East Asian allele counts: one within gnomAD v4.1, and a second within an in-house database (consisting of 34,014 East Asian individuals out of 96,198 total samples), the latter included to compensate for the comparatively small East Asian sample size in gnomAD. For each test, a Bonferroni-corrected significance threshold was applied ($p < 6.85 \times 10^{-5}$, 0.05/729 variants tested); variants below this threshold were classified as "Significant East Asian-specific," and those with $6.85 \times 10^{-5} \leq p < 0.05$ as "Suggestive East Asian-enriched." For each variant, the more significant classification between the two independent tests was retained as the final call.

Results

Cohort Demographics and Phenotypic Characteristics

The demographic and phenotypic characteristics of the 1,620-participant cohort are summarized in Supplementary Table S1. The cohort was predominantly male, with a male-to-female ratio of 1.36:1 (935/1,620; 57.72%). Among the 1,620 participants, 1,234 (76.17%) had documented information on age of symptom onset. Symptom onset was predominantly in the early childhood period: 39.44% (639/1,620) presented with symptom onset before 2 years of age (neonatal and infancy), and 23.52% (381/1,620)

between 2 and 11 years. The median age at enrollment was 9 years (mean: 13.5 years, SD: 14.8 years). Among these 1,234 participants, the mean interval from symptom onset to testing was 7.00 years.

Neurodevelopmental disorders constituted the largest disease category (785/1,620; 48.46%), followed by neurology disorders (254/1,620; 15.68%) and dysmorphic and congenital abnormality syndromes (154/1,620; 9.51%). The most frequently reported phenotypic features were global developmental delay (n = 199), seizure (n = 162), and intellectual disability (n = 69).

Of the total participants, 52.0% (842/1,620) had undergone at least one prior genetic test. The most frequent prior tests included gene panel testing (n = 454, 53.9%), CMA (n = 317, 37.6%), and karyotyping (n = 229, 27.2%), with some participants having received multiple modalities. Additionally, 151 participants (17.9%) had a history of prior ES or GS. Notably, 50 of these participants (including those with prior ES/GS) were referred despite having previous positive results with pathogenic/likely pathogenic (P/LP) variants, typically to investigate partially explained phenotypes or suspected dual diagnoses. Re-evaluation of these 50 cases yielded the following: the previously identified variants were confirmed and reported as diagnostic in 25 cases, whereas in 18 cases, the prior findings were not reported due to low clinical correlation upon reassessment (e.g., CNVs with incomplete penetrance or variants reclassified as likely benign (LB) based on internal allele frequencies). Importantly, this study identified new or additional diagnostic variants in 4 cases and additional VUS in 3 cases. Among the 151 participants with prior ES or GS, 23 were successfully diagnosed as positive in this study through the reclassification of previous findings or the identification of new or additional diagnostic variants.

Meanwhile, 48.0% of participants (778/1,620) had never undergone any prior genetic testing, meaning that GS through this program represented their first access point to molecular diagnostics. The diagnostic yield was higher in the 'first-tier cohort' (37.3%) compared to the 'previously tested cohort' (32.9%), though this difference was not statistically significant (p-value = 0.056) (Supplementary Figure S1). Among this first-tier group, the mean interval between symptom onset and testing was 7.68 years (median 4 years), compared to 6.43 years (median 4 years) in the previously tested cohort (Welch’s t-test, p = 0.016; Mann-Whitney U test, p = 0.122).

Overall Diagnostic Yield

Across 1,620 participants, the overall diagnostic yield (positive result rate) was 35.0% (567/1,620; 95% CI: 32.7-37.3%). Among the 567 positively diagnosed patients, 13 cases (2.3%) received dual or multiple independent molecular diagnoses. This highlights the comprehensive capability of GS to capture complex genetic etiologies that sequential targeted panels often miss. An additional 13.52% (219/1,620) received inconclusive results and 51.48% (834/1,620) received negative results (Figure 1A). Among 219 participants with inconclusive findings, the most common reasons were: identification of a VUS requiring further functional studies in 40.1% (88/219), requiring RNA-seq validation in 16.4% (36/219), VUS requiring *de novo* confirmation in 16.4% (36/219) (Supplementary Table S2). Initially, 60 cases required de novo confirmation; following parental testing (e.g., Sanger sequencing), 24 of these were confirmed as de novo and reclassified as LP, reducing the number of cases still awaiting de novo confirmation to 36

in the final inconclusive group. In contrast, the 88 cases requiring functional studies remained largely unresolved over the same period, as functional validation assays generally require substantially more time and technical complexity than parental Sanger testing. The diagnostic yield was higher in female participants (253/685; 36.93%) compared with male participants (314/935; 33.58%), though this difference did not reach statistical significance (*p*-value = 0.1624) (Figure 1B).

When stratified by age of onset, diagnostic yield was highest among participants with neonatal and infancy onset (247/639; 38.65%), followed by childhood onset (132/381; 34.65%) (Figure 1C). By disease category, tumor syndromes showed the highest diagnostic yield (50.00%; 34/68), followed by ophthalmological disorders (43.33%; 13/30), although these were among the smaller groups. Notably, neurodevelopmental disorders, which accounted for approximately half of the total patient population, showed a diagnostic yield of 39.24% (308/785), which is consistent with the typical diagnostic yields reported for ES/GS in this patient group [37–39] (Figure 1D). Multivariable logistic regression analysis at the individual patient level demonstrated that the number of clinical features per patient was not significantly associated with diagnostic success (*p*-value = 0.293; CI: 0.97-1.09; OR=1.03). This finding indicates that after adjusting for disease categories, the quantity of clinical manifestations does not primarily account for the variation in diagnostic efficiency across different categories, robustly supporting our initial hypothesis (Supplementary Table S3). Diagnostic yield also varied with the length of the interval between symptom onset and testing, generally increasing from approximately 30% among participants tested within 5 years of symptom onset to over 50% among those tested 16-25 years after onset (logistic regression: OR 1.03 per additional year of delay, 95% CI 1.01-1.04, p < 0.001) (Figure 1E). Notably, among the 432 positively diagnosed participants with documented onset information, the mean interval from symptom onset to molecular diagnosis, the diagnostic journey, was 8.36 years, reflecting the prolonged diagnostic delay (typically 5 to 10 years) commonly experienced by rare disease patients [40,41].

Variant Characteristics

Among 567 participants with positive findings, nuclear genome SNV/INDELs were the most common, comprising 417 of 567 positive cases (73.54%). Large structural variants including CNVs constituted the second most prevalent class (127/567; 22.39%), followed by repeat expansions (11/567; 1.94%), mitochondrial SNV/INDELs (7/567; 1.23%), aneuploidy (2/567; 0.35%), MEIs (2/567; 0.35%), and UPiD (1/567; 0.18%). (Table 2).

Of all reported variants across positive and inconclusive cases (n = 786 total), 71 (9.03%) were classified as exclusively identifiable by GS, variant types that would be undetectable by ES. These comprised: 35 deep intronic SNV/INDELs (of which 9 were diagnostic in positive cases and 26 were candidate findings in inconclusive cases), 23 small intragenic CNVs (<50 kb or spanning <3 exons), 6 non-exonic repeat

expansions, 5 non-coding region SNV/INDELs, and 2 genomic inversions. In positive cases alone, 36 of 567 (6.35%) were attributable to GS-exclusive variants.

Importantly, among the patients with a history of prior ES who received additional positive diagnoses in this study, structural variants (SVs) emerged as the most prevalent finding, accounting for 8 cases. These included small CNVs spanning fewer than three consecutive exons (3 cases), large CNVs (3 cases), genomic inversions (1 case), and complex SVs (1 case) (Supplementary Figure S1F).

An illustrative case involved a 12-year-old male with suspected Gitelman syndrome who presented with hypokalemia at age 8. Previous ES had yielded an inconclusive result, identifying only a single heterozygous likely pathogenic variant in *SLC12A3*. However, GS successfully identified an additional pathogenic heterozygous deletion spanning exons 14-15 of *SLC12A3* (Supplementary Figure S2). Notably, as this deletion (approximately 1.8 kb) spans only two exons, it was uniquely detectable through GS, which enabled a definitive molecular diagnosis.

Additionally, we evaluated the clinical actionability of the 567 positively diagnosed cases using curated gene lists from ClinGen, Treatable-ID, and IEMbase. A total of 20.3% (n=115) of the positive cases were diagnosed with conditions involving genes with established clinical management protocols (Figure 2).

East Asian-specific and novel variants

To characterize the population-level specificity of identified variants, 729 SNVs from 627 patients were analyzed. For variants reported in gnomAD v4.1, their allele counts in the East Asian population were compared against those of other non-East Asian ancestry groups using Fisher's exact test with a Bonferroni-corrected significance threshold of $p < 6.85\times10^{-5}$ (0.05/729); variants with $p < 6.85\times10^{-5}$ were classified as significant East Asian specific (EA_significant), whereas those with $6.85\times10^{-5} \le p < 0.05$ were considered suggestive East Asian specific (EA_suggestive).

Among the 729 variants, 372 (51.0%) were unreported in both gnomAD and the in-house dataset. Of

the 287 variants absent from both ClinVar and gnomAD, 37 were subsequently identified in the in-house dataset, leaving 250 variants entirely absent from ClinVar, gnomAD, and the in-house dataset,

confirming them as entirely novel discoveries. Filtering against gnomAD alone classified 81 and 30 variants as EA_significant and EA_suggestive, respectively, while 141 were Non_specific. However, incorporating the in-house dataset refined this classification: EA_significant, EA_suggestive, and Non_specific variants increased to 99 (13.6%), 41 (5.6%), and 217 (29.8%). Detailed classification is provided in Figure 3.

Reanalysis by family testing and pipeline update

Following the initial GS reporting, we maintained rigorous diagnostic efforts to maximize the clinical utility of the genomic data through two key mechanisms: family-based segregation testing and

systematic reanalysis driven by database and pipeline updates. These proactive approaches significantly enhanced the overall diagnostic yield and refined the classification of test results. The cumulative impact of these processes on final result classification is illustrated in Figure 4. Initial analysis yielded 534 positive, 268 inconclusive, and 818 negative results; following family testing and reanalysis, results were 567 positive, 219 inconclusive, and 834 negative. Taken together, these processes resulted in a net increase of 33 positively diagnosed cases over initial analysis.

Family-based segregation analysis was performed in 353 families (711 additional subjects), encompassing a total of 366 variants whose inheritance pattern was confirmed through this analysis. This analysis directly resulted in a change of clinical interpretation for 71 cases: 32 cases were changed from inconclusive to positive, 29 from inconclusive to negative, 6 from positive to inconclusive, and 4 from positive to negative. At the variant level, 34 VUS variants were reclassified as likely pathogenic (LP) and 28 LP variants were reclassified as pathogenic (P) following segregation analysis. The inheritance patterns of the 366 confirmed variants were as follows: 96 were assumed *de novo*, 33 were confirmed *in trans* (compound heterozygous in AR disorders), and 38 were inherited from an unaffected parent. Detailed family testing results are presented in Supplementary Table S4.

Beyond segregation testing, systematic reanalysis was conducted on inconclusive or negative samples to incorporate updates in bioinformatics pipelines and gene-disease association databases (e.g., OMIM, Orphanet, and variant-specific databases). A total of 20 cases underwent reanalysis due to database/pipeline updates; among these, 9 cases were reclassified from inconclusive/negative to positive, and 6 cases from negative to inconclusive (mean time to reanalysis: 11.8 months). Drivers of reclassification included new disease-gene associations reported as gene of uncertain significance (GUS) in 4 cases, newly curated disease-gene associations on OMIM/Orphanet in 5 cases (mean time to reanalysis: 9.6 months), and variant reclassification by database/pipeline update in 11 cases (mean time to reanalysis: 12.27 months) (Supplementary Table S5). Additionally, two cases were subjected to reanalysis specifically driven by explicit clinical requests (Supplementary Table S6).

The practical impact of systematic reanalysis is illustrated by the following case: a 12-year-old male referred for short stature and skeletal dysplasia, whose initial report was issued as negative in June 2025. However, a reanalysis conducted six months later identified a VUS in the *PTBP1* gene (NM_002819.5:c.2T>C [p.Met1?]), resulting in an inconclusive report. Although a study linking *PTBP1* to skeletal dysplasia, consistent with the proband's phenotype, had been published in September 2025 (three months after the initial report) [42], the variant remained classified as 'uncertain' at the time because *PTBP1* had not yet been registered in the OMIM/Orphanet database. Following the subsequent entry of STAD syndrome (OMIM: 621495), caused by heterozygous *PTBP1* mutations, the variant was reclassified as 'pathogenic.' Consequently, the case was updated to a 'positive' diagnosis without the need for additional testing.

Although 3 cases successfully transitioned to positive diagnoses following the formal entry of their corresponding genes into OMIM or Orphanet, 17 cases currently remain classified as inconclusive due to

being harbored in established GUS genes (Supplementary Table S2). Notably, many of these remaining GUS cases possess compelling clinical, genetic, and segregation evidence strongly supporting pathogenicity, yet remain locked in inconclusive status solely due to delayed formal curation in public disease databases such as OMIM.

An illustrative example involves a 5-year-old female presenting with neurodevelopmental disorder, delayed motor and language development, and obesity. GS identified a heterozygous missense variant in *RAB11A* (NM_004663.5:c.461C>T [p.Pro154Leu]). Although *RAB11A* is not currently registered as a disease-causing gene in OMIM or Orphanet, multiple recent case reports have linked *RAB11A* variants to neurodevelopmental phenotypes similar to those observed in our patient [43,44]. Furthermore, this exact variant has already been submitted twice as Pathogenic/Likely Pathogenic in ClinVar [ClinVar: VCV002582937.27]. Subsequent family Sanger sequencing confirmed that the variant was inherited from her mother, who exhibits an identical clinical phenotype, thereby establishing co-segregation of the variant with the disease. Despite these strong lines of evidence (including candidate gene literature, existing ClinVar assertions, and confirmed familial segregation) this case remains classified as inconclusive simply because *RAB11A* has not yet been formally curated into OMIM. This highlights an urgent need for accelerated curation and active submission of novel disease-gene associations to public databases to bridge the gap between emerging clinical evidence and formal molecular diagnosis.

Additional research findings

Of the 1,602 participants who consented to receive secondary findings (1602/1620; 98.89%), screening was performed for 84 medically actionable genes as recommended by the American College of Medical Genetics and Genomics (ACMG) [45]. Pathogenic or likely pathogenic secondary findings were identified in 63 participants (63/1,602; 3.93%) across the cohort. Among these, the presence of the variant was further confirmed in family members of 19 participants through family-based Sanger sequencing.

Satisfaction Survey

The satisfaction survey results for clinicians and patients are summarized in Tables S7 and S8, respectively based on 77 clinician responses (10 in 2023, 24 in 2024, and 43 in 2025) and 340 patient responses (80 in 2024 and 260 in 2025). For clinicians, the overall satisfaction remained high across all three years, with the highest average score observed in 2024 (96.52), followed by 2025 (94.28) and 2023 (91.4). Notably, the items assessing the "necessity of program continuation" achieved perfect scores (100) across all three years.

For patients, high levels of satisfaction were also reported, with average scores of 97.74 in 2024 and 94.82 in 2025. In 2024, the "reliability of the testing process," "alleviation of financial burden," and "overall satisfaction with the program" all achieved perfect scores of 100. Across both groups, the "necessity of program continuation" received consistently high scores (97.7 to 100), demonstrating strong stakeholder support for the program's sustainability.

Discussion

Our three-year evaluation of the national RDSP demonstrated an overall diagnostic yield of 35.0%, aligning with global GS benchmarks (30-45%) [2,46]. Importantly, this diagnostic efficiency remained stable as the program scaled up from 11 to 34 centers, confirming the robustness and reproducibility of this centralized diagnostic model.

According to 2023 national patient statistics, approximately 47% of rare disease patients in South Korea reside outside the Capital area. Prior to the implementation of this program, due to the highly centralized genomic infrastructure, nearly half of these patients had to travel long distances to the capital to access advanced diagnostics, such as GS [10]. This geographical barrier inherently exacerbates the risk of diagnostic delay; indeed, previous literature indicates that patients required to consult specialists outside their residential area (OR 2.1; 95% CI 1.6-2.9) or travel to different regions at the onset of their diagnostic journey (OR 1.7; 95% CI 1.1-2.5) face a significantly higher risk of delayed diagnosis [47]. By enabling local sample collection across 34 regional centers, the national RDSP spared the 1,620 participants an average travel distance of approximately 200 km to the capital, significantly reducing logistical, financial, and emotional burdens. Notably, 48.0% of participants (778/1,620) had never undergone any prior genetic testing, indicating that GS through this program represented their first access point to molecular diagnostics. Among this first-tier group, the mean interval between symptom onset and testing was 7.68 years (median 4 years) - longer than that of the previously tested cohort (6.43 years) - suggesting that the program reached patients who had experienced prolonged, unaddressed diagnostic delay prior to program access, consistent with the pre-existing regional testing gap described above. Diagnostic yield did not decline among patients with the longest onset-to-testing intervals; rather, yield tended to increase with the duration of delay (OR 1.03 per additional year, $p<0.001$) (Figure 1E). Although this pattern may partly reflect the evolving clinical recognizability of certain syndromic conditions over time, it also suggests that the nationwide, decentralized reach of this program extended diagnostic opportunity to patients who had remained undiagnosed for extended periods, complementing the regional access gains described above.

This decentralized access model is strongly supported by high stakeholder satisfaction rates (91.4%-97.7%) and an efficient mean turnaround time of 27.5 days, a major improvement over the traditional referral pathway. However, the significant inter-center variance in diagnostic yield across the 34 participating centers (11.1% to 51.9%, $p = 0.007$) indicates that centralized laboratory testing alone cannot fully standardize outcomes. This discrepancy likely stems from differences in local clinical workflows, phenotyping depth, and clinician experience, underscoring the need for targeted capacity-building and standardized protocols across regional centers.

The 19.2% of variants reclassified as East Asian-enriched (EA_significant or EA_suggestive) only after incorporating the in-house dataset illustrates the risk of allele frequency misclassification when relying solely on globally aggregated databases such as gnomAD, whose East Asian sample size remains comparatively limited. Conversely, the 34.3% (n = 250) of variants that remain novel even after

incorporating the in-house dataset supports their genuine rarity/novelty in the Korean population, underscoring the value of ethnicity-matched resources for confidently distinguishing true novel variants from database under-representation artifacts. Although global databases like gnomAD include East Asian data, the representation remains limited, comprising only 7.05% (9,977/141,456) in v2 and 3.42% (2,604/76,156) in v3, which is insufficient to capture the full spectrum of East Asian haplotypes [23]. Therefore, to prevent the misclassification of rare pathogenic alleles as common benign variants, it is imperative to establish and utilize a dedicated Korean/East Asian-specific population variant dataset. Notably, national-level projects, such as the one described in this study, will significantly contribute to the accumulation and curation of these critical genomic datasets.

Our findings also strongly support the clinical utility of GS as a first-line genomic investigation rather than a last-resort escalation from ES or panel testing. In support of this, GS successfully identified 23 additional diagnostic cases among 151 participants with a history of prior ES or GS. Furthermore, 9.03% of all reported variants were classified as exclusively identifiable by GS, inherently beyond the detection scope of conventional testing. These data collectively support the deployment of GS as a first-line diagnostic tool, thereby sparing patients the cumulative financial costs, emotional burdens, and diagnostic delays associated with sequential testing strategies. Additionally, the identification of dual diagnoses in our cohort highlights GS's capability to untangle complex, overlapping phenotypes that might be missed or only partially resolved by sequential targeted panels.

To maximize diagnostic yield beyond the initial report, we employed two complementary post-initial strategies: targeted family-based segregation and systematic database/pipeline reanalysis. Because the national RDSP adopted a proband-only GS strategy to optimize limited resources, family-based Sanger sequencing served as a crucial, cost-effective orthogonal method. This approach systematically resolved initial uncertainties by downgrading multiple VUS findings to (likely) benign, which reclassified 29 cases from inconclusive to negative and four from positive to negative. This downgrading process holds substantial clinical value, as unresolved VUS findings often trigger severe patient anxiety and prompt unnecessary clinical surveillance. Concurrently, segregation testing upgraded 32 inconclusive cases to positive via de novo or compound heterozygosity confirmation. Notably, this included six cases where de novo status was confirmed through STR analysis introduced in 2025, successfully meeting the ACMG PS2 criterion. While family testing accounted for most of this diagnostic improvement, systematic reanalysis-though smaller in absolute impact-effectively yielded new diagnoses years after initial testing by capturing newly established gene-disease associations, such as the PTBP1 case. Together, these efforts increased the final diagnostic yield from 32.96% to 35.0% and reduced inconclusive results from 16.54% to 13.52%. These outcomes collectively underscore that a genomic diagnosis should not be treated as a single, one-time event, but rather as an evolving process that continues to develop through sustained re-evaluation.

Importantly, a molecular diagnosis through the national RDSP translates directly into actionable clinical guidance. In our cohort, 20.3% of the diagnosed patients presented actionable findings; while the most prevalent category of actionability was specialized evaluation and surveillance, several cases also

involved genes with established pharmacological treatments or nutritional therapies. Furthermore, in South Korea, patients receiving a diagnosis linked to a qualifying rare disease code may access the National Health Insurance Service special exemption program, which covers up to 90% of medical expenditure for qualifying conditions [48]. Among the positively diagnosed participants in our cohort, 443 patients were diagnosed with conditions linked to such qualifying codes, substantially reducing their financial burden and expanding their access to targeted care. These tangible clinical and financial benefits emphasize the critical need for expanded national support and funding for such genomic screening initiatives.

Beyond these direct insurance benefits, the program demonstrates substantial health-economic value. While direct economic modeling was beyond the scope of this study, terminating an average 8.36-year diagnostic journey within a 27.5-day turnaround time inherently prevents the cumulative costs associated with redundant diagnostic procedures, prolonged empirical treatments, and travel expenses. Thus, implementing GS as a primary diagnostic tool not only provides clinical clarity but also effectively curtails the economic burden on both the healthcare system and affected families.

Notably, Sanger sequencing was integrated into our workflow primarily because Korean regulatory and insurance frameworks still mandate orthogonal confirmation for GS-identified clinical variants. However, our cohort data strongly challenges the necessity and universality of this mandatory requirement. Among the 802 total cases reported as positive or inconclusive in the initial GS analysis, low-confidence metrics requiring Sanger validation-defined as variant allele fraction (VAF) < 30% or quality score < 250-were observed in only 22 cases (2.7%). This indicates that 97.3% (780/802) of diagnostic or candidate variants detected by GS already possessed high-confidence quality metrics. Consequently, routine Sanger confirmation represents a technically redundant step, a conclusion aligned with a growing body of recent literature, and serves as a source of unnecessary diagnostic delay and healthcare expenditure [49,50].

In terms of structural variation, Sanger sequencing exhibited intrinsic technical limitations as a confirmatory tool for complex genomic alterations. Among 534 positively diagnosed cases, 67 cases (12.5%) involved large structural variants or copy number variants (CNVs) with unmapped breakpoints that could not be validated or delineated by Sanger sequencing, necessitating reliance on alternative regional modalities such as CMA, MLPA, or FISH. These findings demonstrate that while high-depth GS reliably captures a comprehensive spectrum of genomic variants in a single assay, traditional Sanger sequencing exhibits inherent technical limitations as a universal confirmatory tool, particularly for complex structural alterations. Therefore, our findings highlight an urgent need for regulatory policy shifts to recognize high-confidence GS as a definitive stand-alone diagnostic tool, optimizing the requirement for Sanger confirmation specifically for high-confidence GS calls and redirecting health system resources toward resolving remaining undiagnosed cases through multi-omic approaches.

Despite the high diagnostic utility and efficiency of our GS pipeline, a subset of cases remained negative or inconclusive even after reanalysis. Long-read sequencing may be worth considering as a next step for

these cases; for example, in participants with an identified repeat expansion or a single heterozygous variant in a recessive disease gene, long-read sequencing could determine the precise repeat number or reveal a second variant undetected by short read GS. Similarly, targeted RNA analysis via RT-PCR, introduced in 2025 to validate predicted splicing alterations could not resolve 7 cases (4 with failed cDNA synthesis, 3 detected only normal transcripts); whole transcriptome RNA-seq, which was precluded during initial evaluation due to resource and sample-collection constraints, may resolve these cases and is being implemented in 2026.

Conclusion

In this comprehensive three-year study of 1,620 participants, the national RDSP demonstrated that a centralized GS workflow provides high diagnostic accuracy (35.0%) and operational efficiency. By effectively shortening the prolonged diagnostic journey and reducing travel burdens for patients with rare disease, this model firmly establishes GS as a powerful first-line diagnostic tool for rare diseases. Furthermore, the tangible clinical and health-economic benefits observed, amplified by targeted family testing and systematic reanalysis, strongly support the program's continuation. Ultimately, these findings underscore the critical necessity of expanding national genomic initiatives to dismantle regional healthcare disparities and ensure equitable access to precision medicine.

List of abbreviations

ACMG, American College of Medical Genetics and Genomics; AD, autosomal dominant; AF, allele frequency; AMP, Association for Molecular Pathology; AR, autosomal recessive; BAM, Binary Alignment Map; BCL, binary base call; CI, confidence interval; CMA, chromosomal microarray; CNV, copy number variant; DBS, dried blood spot; DGV, Database of Genomic Variants; DOC, depth-of-coverage; EA, East Asian; ES, exome sequencing; FISH, fluorescence in situ hybridization; GATK, Genome Analysis Toolkit; GRC, Genome Reference Consortium; GRCh38, Genome Reference Consortium Human Build 38; GS, genome sequencing; GUS, gene of uncertain significance; HGVS, Human Genome Variation Society; HPO, Human Phenotype Ontology; IGV, Integrative Genomics Viewer; INDEL, insertion/deletion; IRB, Institutional Review Board; KDCA, Korea Disease Control and Prevention Agency; KR-RDSP, Korean Regional Rare Disease Diagnostic Support Program; LB, likely benign; LP, likely pathogenic; MANE, Matched Annotation from NCBI and EMBL-EBI; MEI, mobile element insertion; MELT, Mobile Element Locator Tool; MLPA, multiplex ligation-dependent probe amplification; NGS, next-generation sequencing; OMIM, Online Mendelian Inheritance in Man; OR, odds ratio; P, pathogenic; PCR, polymerase chain reaction; rCRS, Revised Cambridge Reference Sequence; REVEL, Rare Exome Variant Ensemble Learner; ROH, regions of homozygosity; RT-PCR, reverse transcription polymerase chain reaction; SD, standard deviation; SNV, single-nucleotide variant; STR, short tandem repeat; SV, structural variant; TAT, turnaround time; ti/tv, transition/transversion; UPiD, uniparental isodisomy; VAF, variant allele fraction; VEP, Variant Effect Predictor; VUS, variant of uncertain significance; XL, X-linked.

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board (IRB) of Pusan National University Yangsan Hospital (55-2026-086), and written consent was obtained from all participants/families.

Consent for publication

Not applicable.

Availability of data and materials

The genomic and clinical data that support the findings of this study are available from the corresponding authors upon reasonable request. The data is not publicly available due to privacy or ethical restrictions.

Competing interests

The authors declare no conflicts of interest.

Funding

This study was supported by the Korea Disease Control and Prevention Agency (Announcement Nos. 2023-128, 2024-176, and 2025-34).

Authors' contributions

Conceptualization: S.J., J.Y.K., S.Y., J.S., S.H.O., H.L., C.K.C. Data Curation: S.J., G.H.S., S.W.R., D.M., K.W.K., S.L., K.P., E.Y., Y.H.K. Software: G.Y.C., K.K. Supervision: J.S., S.H.O., H.L., C.K.C., G.H.S, H.K., D.H.J., J.K.[8], J.A., A.C., J.K.[11], J.Y., S.K., Y.L., B.L.L., D.H.K., J.H.L., Y.M.K.[18], E.K., Y.L.S., K.Y.L., M.H., J.S.Y., H.J.K., J.S.K., H.H.L., J.K.K., Y.M.K.[28] Funding Acquisition: C.K.C. Project Administration: J.Y.K., S.Y., Y.E.L., J.S., S.H.O., H.L., C.K.C. Writing—original draft: S.J., H.L., J.S., S.H.O., and C.K.C. Writing – Review & Editing: S.J., G.H.S., G.Y.C., K.K., S.W.R., D.M., K.W.K., S.L., K.P., E.Y., Y.H.K., J.Y.K., S.Y., Y.E.L., H.K., D.H.J., J.K.[8], J.A., A.C., J.K.[11], J.Y., S.K., Y.L., B.L.L., D.H.K., J.H.L., Y.M.K.[18], E.K., Y.L.S., K.Y.L., M.H., J.S.Y., H.J.K., J.S.K., H.H.L., J.K.K., Y.M.K.[28], J.S., S.H.O., H.L., C.K.C.

Acknowledgements

We thank all the participants and their families for agreeing to share their clinical and genomic data for this program. We also thank the physicians and staff at each of the participating regional rare disease centers for their contributions in recruiting participants, providing clinical information, and collecting samples.

References

1. RARE DISEASE MANAGEMENT ACT.
2. Dawkins, H. J. S. *et al.* Progress in Rare Diseases Research 2010-2016: An IRDiRC Perspective. *Clin Transl Sci* **11**, 11–20 (2018).
3. Lionel, A. C. *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* **20**, 435–443 (2018).
4. Stavropoulos, D. J. *et al.* Whole Genome Sequencing Expands Diagnostic Utility and Improves Clinical Management in Pediatric Medicine. *NPJ Genom Med* **1**, 15012- (2016).
5. Wojcik, M. H. *et al.* Genome Sequencing for Diagnosing Rare Diseases. *N Engl J Med* **390**, 1985–1997 (2024).
6. Ewans, L. J. *et al.* Whole exome and genome sequencing in mendelian disorders: a diagnostic and health economic analysis. *Eur J Hum Genet* **30**, 1121–1131 (2022).
7. Alfares, A. *et al.* Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genet Med* **20**, 1328–1333 (2018).
8. Kim, H., Lee, N. & Kim, S.-N. Suburbia in evolution: Exploring polycentricity and suburban typologies in the Seoul metropolitan area, South Korea. *Land Use Policy* **75**, 92–101 (2018).
9. *2023 Annual Report on the Rare Disease Patients in Korea*. https://helpline.kdca.go.kr/cdchelp/ph/infNoti/selectAnnFrmDetail.do?menu=G0100&ntbdSno=962&schNtbdCcd=01.
10. Khang, R. *et al.* Genome Sequencing of Rare Disease Patients Through the Korean Regional Rare Disease Diagnostic Support Program. *Hum Mutat* **2025**, 6096758 (2025).
11. Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res* **27**, 849–864 (2017).
12. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147 (1999).
13. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/ARXIV.1303.3997 (2013).
14. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
15. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491–498 (2011).
16. Wagner, J. *et al.* Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol* **40**, 672–680 (2022).
17. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).

18. Dolzhenko, E. *et al.* ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* **35**, 4754–4756 (2019).
19. Gardner, E. J. *et al.* The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res* **27**, 1916–1929 (2017).
20. Quinodoz, M. *et al.* AutoMap is a high performance homozygosity mapping tool using next-generation sequencing data. *Nat Commun* **12**, 518 (2021).
21. Seo, G. H. *et al.* Diagnostic yield and clinical utility of whole exome sequencing using an automated variant prioritization system, EVIDENCE. *Clin Genet* **98**, 562–570 (2020).
22. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
23. Gudmundsson, S. *et al.* Variant interpretation using population databases: Lessons from gnomAD. *Hum Mutat* **43**, 1012–1030 (2022).
24. Richards, S. *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* **17**, 405–424 (2015).
25. Riggs, E. R. *et al.* Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med* **22**, 245–257 (2020).
26. McCormick, E. M. *et al.* Specifications of the ACMG/AMP standards and guidelines for mitochondrial DNA variant interpretation. *Hum Mutat* **41**, 2028–2057 (2020).
27. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877–885 (2016).
28. Won, D.-G., Kim, D.-W., Woo, J. & Lee, K. 3Cnet: pathogenicity prediction of human variants using multitask learning with evolutionary constraints. *Bioinformatics* **37**, 4626–4634 (2021).
29. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e24 (2019).
30. Gargano, M. A. *et al.* The Human Phenotype Ontology in 2024: phenotypes around the world. *Nucleic Acids Res* **52**, D1333–D1346 (2024).
31. Köhler, S. *et al.* Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet* **85**, 457–464 (2009).
32. McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet* **80**, 588–604 (2007).
33. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192 (2013).
34. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291 (2007).
35. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**, e115 (2012).

36. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res* **45**, D37–D42 (2017).
37. Srivastava, S. *et al.* Correction: Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med* **22**, 1731–1732 (2020).
38. Neuens, S. *et al.* Diagnostic yield of clinical exome sequencing in 868 children with neurodevelopmental disorders. *Eur J Med Genet* **76**, 105030 (2025).
39. Malmgren, H. *et al.* Diagnostic yield of 1000 trio analyses with exome and genome sequencing in a clinical setting. *Front Genet* **16**, 1580879 (2025).
40. Siegel, I. J. *et al.* Diagnostic delays in rare genetic disorders with neuropsychiatric manifestations: A systematic review. *Eur J Med Genet* **75**, 105016 (2025).
41. Tinker, R. J. *et al.* Diagnostic delay in monogenic disease: A scoping review. *Genet Med* **26**, 101074 (2024).
42. Masson, A. *et al.* PTBP1 variants displaying altered nucleocytoplasmic distribution are responsible for a neurodevelopmental disorder with skeletal dysplasia. *J Clin Invest* **135**, e182100 (2025).
43. Zhang, H. *et al.* A de novo mutation in RAB11A is associated with neurodevelopmental disorder accompanied by variable multisystem abnormalities. *Front Genet* **16**, 1636206 (2025).
44. Hamdan, F. F. *et al.* High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies. *Am J Hum Genet* **101**, 664–685 (2017).
45. Lee, K. *et al.* ACMG SF v3.3 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* **27**, 101454 (2025).
46. Costain, G. *et al.* Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *Eur J Hum Genet* **26**, 740–744 (2018).
47. Benito-Lozano, J. *et al.* Diagnostic Process in Rare Diseases: Determinants Associated with Diagnostic Delay. *Int J Environ Res Public Health* **19**, 6456 (2022).
48. Lee, H. *et al.* Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* **312**, 1880–1887 (2014).
49. De Cario, R. *et al.* Sanger Validation of High-Throughput Sequencing in Genetic Diagnosis: Still the Best Practice? *Front Genet* **11**, 592588 (2020).
50. Arteche-López, A. *et al.* Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes. *Sci Rep* **11**, 5697 (2021).

Tables

Table 1. Annual Summary of GS Testing Results

	2023	2024	2025
No. of Samples	400	410	810
No. of Participating Centers	11	18	34
No. of Family testing samples	121	157	433
Diagnostic Yield (%)	37.25[a]	31.71	35.56
Mean TAT (Days)	29.61	28.35	26.10

 Yield for the 2023 cohort differs from the previously reported 36.3% [10] due to additional diagnoses established through family testing and reanalysis conducted after the pilot publication.

Table 2. Characteristics of Positive and Inconclusive cases and variants identifiable only by genome sequencing

Variant Type	**Positive (n)**	**Inconclusive (n)**	**Total (n)**
SNV/INDELs (nuclear genome)	417	190	607
SNV/INDELs (mitochondrial genome)	7	2	9
Repeat expansion	11	1	12
Mobile element insertion	2	0	2
Aneuploidy	2	2	4
Large structural variants including copy number variants	127	23	150
UPiD	1	1	2
Total	**567**	**219**	**786**
Variants only identifiable by genome sequencing	**Positive (n)**	**Inconclusive (n)**	**Total (n)**
Deep intronic SNV/INDELs	9	26	35
Non-coding region SNV/INDELs	4	1	5
< 3 consecutive and < 50 kb exonic deletion/duplication	16	7	23
inversion	2	0	2
non-exon repeat	5	1	6
Total (%)	**36 (6.35%)**	**35 (15.98%)**	**71 (9.03%)**

Figures

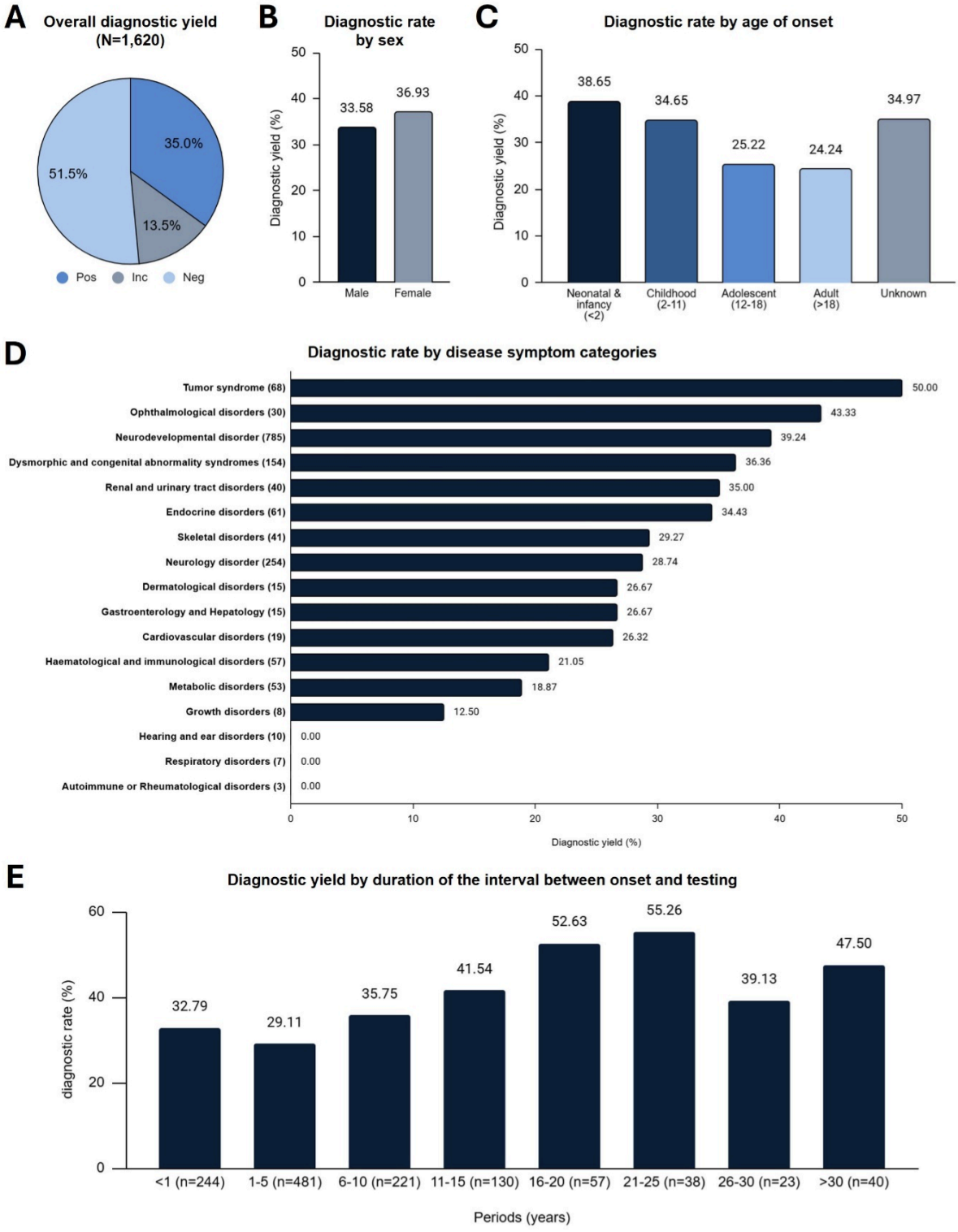


Figure 1

Diagnostic yield and test statistics.

A. Overall diagnostic yield. B. Diagnostic rate by sex. C. Diagnostic yield by age group. D. Diagnostic rate by disease symptom categories. E. Diagnostic yield by duration of the interval between onset and testing.

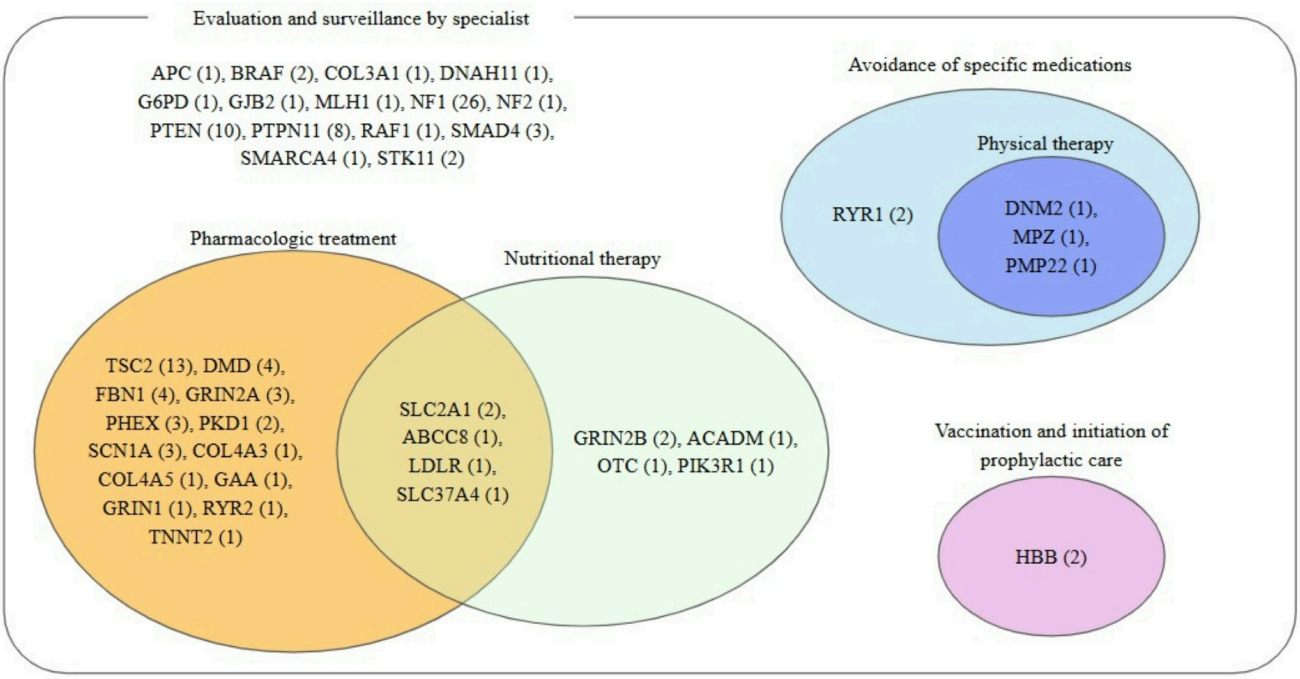


Figure 2

Clinical actionability of identified variants and classification of management categories.

The diagram illustrates the distribution and clinical actionability of molecular diagnoses among the positively diagnosed rare disease patients. Out of the 567 positively diagnosed cases, 115 (20.3%) harbored variants in genes with well-established clinical actionability and management protocols. Individual gene symbols are plotted within their respective clinical management categories, with the number of corresponding cases indicated in parentheses. The actionable interventions are classified into distinct sub-categories, including: Evaluation and surveillance by specialist (outer boundary encompassing broad screening for genes such as *NF1*,*PTEN*, and *PTPN11*); Pharmacologic treatment (orange circle); Nutritional therapy (green circle); Physical therapy (blue nested circles); Avoidance of specific medications; and Vaccination and initiation of prophylactic care (purple circle). Intersecting regions represent genes that qualify for multiple therapeutic modalities simultaneously (e.g., *SLC2A1*and *ABCC8* bridging both pharmacologic and nutritional therapies).

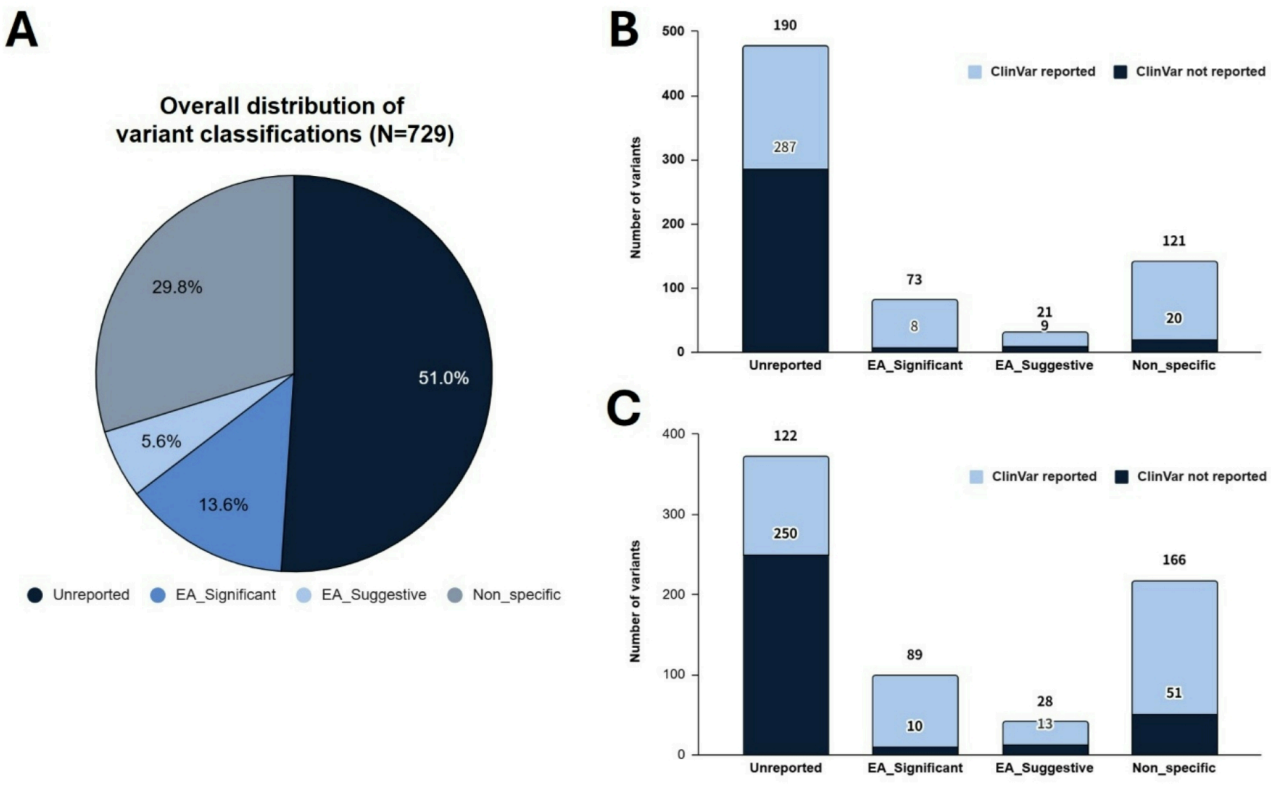


Figure 3

Population-level specificity and ClinVar annotation of the 729 identified variants.

A. Overall percentage distribution of the 729 variants classified into four categories: Unreported, EA_significant, EA_suggestive, and Non_specific. B. Distribution and ClinVar reporting status of variants when filtered against the gnomAD dataset alone. C. Refined distribution and ClinVar reporting status of variants after incorporating the Korean in-house dataset. The light blue and dark navy bars represent variants reported and not reported in the ClinVar database, respectively.

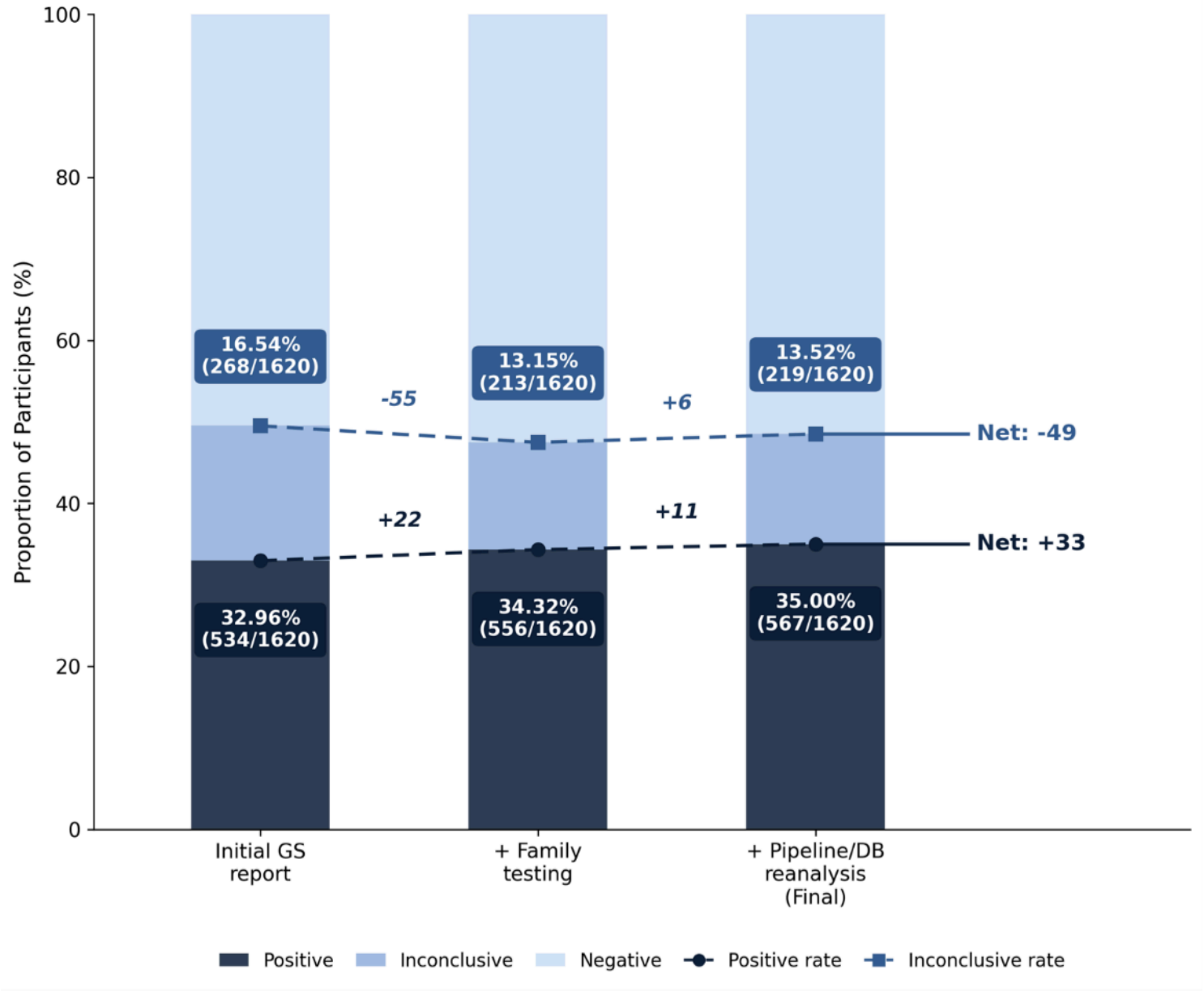


Figure 4

Changes in test results after family testing and reanalysis Stacked bars show the proportion of participants classified as positive, inconclusive, or negative at each stage: the initial GS report, after incorporating family-based segregation testing, and after systematic reanalysis driven by database/pipeline updates (final). Overlaid lines track the positive rate (black) and inconclusive rate (blue) across stages, with the net change in case counts shown between each stage and from initial to final report.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 2325RDSPGenomicMedicinesupple.docx