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.

RareBoost: A Research-Integrated Program for Rare and Undiagnosed Diseases in Türkiye

Mehmet Mert Topaloğlu
Izmir Biomedicine and Genome Center
Ayca Yigit
Izmir Biomedicine and Genome Center
Aliye Kubra Unal
Izmir Biomedicine and Genome Center
Emre Ozzeybek
Izmir Biomedicine and Genome Center
Mert Pekerbas
Izmir Biomedicine and Genome Center
Barıs Salman
Izmir Biomedicine and Genome Center
Kutay Bulut
Izmir International Biomedicine and Genome Institute
Inci Yaprak
Izmir International Biomedicine and Genome Institute
Begum Utz
Izmir Biomedicine and Genome Center
Emine Eylul Taskin
Izmir Biomedicine and Genome Center
Noor Fatima
Izmir Biomedicine and Genome Center
Sinem Aktug
Izmir Biomedicine and Genome Center
Tugce Batur
Izmir Biomedicine and Genome Center
Burcu Akman
Izmir Biomedicine and Genome Center
Evin Iscan
Izmir Biomedicine and Genome Center
Emrah Yucesan
Istanbul University

Ugur Ozbek

ugur.ozbek@ibg.edu.tr

Izmir Biomedicine and Genome Center

Research Article

Keywords: rare diseases, undiagnosed diseases, genomic diagnosis, whole-genome sequencing

Posted Date: June 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9790284/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

Advanced genomic sequencing has improved rare disease (RD) diagnostics, yet many patients remain undiagnosed despite extensive molecular testing. Increasingly, the major challenge lies not in data generation, but in the integration and reinterpretation of clinical and genomic information over time. Here, we describe RareBoost (RB), a research-integrated framework for unresolved RDs based on coordinated multidisciplinary evaluation and systematic genomic reinterpretation in Türkiye.

Methods

RB established a nationwide multidisciplinary diagnostic program for patients with unresolved RDs, integrating expertise in clinical genetics, molecular biology, and bioinformatics through a monthly advisory board. Most enrolled cases had remained unsolved despite prior extensive evaluations, frequently including whole-exome sequencing (WES). The RB workflow included deep phenotyping, reanalysis of prior genomic data, trio-based whole-genome sequencing (WGS), and complementary transcriptomic or functional investigations when indicated.

Results

The RB registry includes 1,057 individuals from 385 families, of whom 201 completed the diagnostic workflow. A definitive molecular diagnosis was established in 27.4% of cases, while clinically relevant variants of uncertain significance were identified in an additional 32.3%. Reanalysis of prior WES data alone resolved a subset of cases without additional sequencing. In unresolved WES-negative cases, WGS identified additional pathogenic findings, including deep intronic and structural variants. Diagnostic resolution was achieved through iterative reinterpretation within a longitudinal multidisciplinary framework.

Conclusions

Many unresolved RD cases benefit from systematic reinterpretation of existing genomic data within structured multidisciplinary systems. Our findings suggest that real-world diagnostic limitations increasingly arise not only from sequencing access, but from fragmentation of clinical and genomic interpretation processes over time. The RB framework supports a shift from one-time testing to integrated, iterative diagnostic models for unresolved RDs.

Introduction

Although rare diseases (RDs) are individually rare, they collectively affect millions of individuals worldwide and represent a major global health challenge [1]. Despite major advances in genomic sequencing, many patients remain undiagnosed due to extensive genetic and phenotypic heterogeneity, difficulties in variant interpretation, and limited integration between clinical practice and research environments [2].

Improvements in sequencing capacity and bioinformatics are progressively shifting the diagnostic bottleneck from data generation toward variant interpretation, deep phenotyping, functional validation, and longitudinal integration of clinical and research data [3]. Although the dominant bottlenecks differ between countries and institutions, these challenges remain globally relevant and have led to the emergence of multidisciplinary Undiagnosed Disease Programs (UDPs) for patients who remain undiagnosed despite extensive clinical and molecular investigation [4–7]. Over the past two decades, these efforts evolved from isolated national initiatives into internationally coordinated diagnostic and research networks (Fig. 1).

Beyond diagnostic yield alone, the clinical success of these programs is increasingly measured by their ability to shorten the prolonged interval between first presentation and molecular diagnosis, often referred to as the “diagnostic odyssey”. RD patients still experience average diagnostic delays of approximately 5–8 years globally, with some remaining undiagnosed for decades [8–10]. Many undergo repeated evaluations across multiple institutions, reflecting fragmented care and contributing to frequent misdiagnoses, delayed interventions, missed therapeutic opportunities, and substantial long-term psychological and socioeconomic burden [11]. Structured UDPs aim to address these challenges through coordinated, multidisciplinary evaluation and systematic genomic reanalysis, with initiatives such as Solve-RD highlighting the value of reinterpretation strategies [12].

Alongside these global challenges, each country faces its own structural and demographic factors affecting RD diagnostics. In Türkiye, a relatively high burden of rare recessive disorders associated with consanguinity increases both the number and diversity of unresolved RD cases, further emphasizing the need for coordinated multidisciplinary evaluation and research-integrated diagnostic approaches [13]. Furthermore, as an upper–middle-income country, Türkiye exhibits considerable regional variability in access to genomic infrastructure, specialized expertise, and emerging technologies, creating a resource-variable diagnostic landscape that continues to complicate the evaluation of unsolved cases, even in well-established centers [11, 13].

RareBoost (RB) is a project funded by the European Union’s Horizon 2020 ERA Chairs Program (https://rareboost.ibg.edu.tr/) and implemented at the Izmir Biomedicine and Genome Center (IBG). Through RB, IBG aims to strengthen its research and innovation capacity in RDs and establish itself as a center of excellence in RD research and diagnostics. To support these goals, RB established the Rare and Undiagnosed Diseases Platform (RUDiP), a multidisciplinary platform focused on RD diagnostics and research through integrated clinical and research-based approaches. Beyond case evaluation, RB also aims to strengthen national and international collaborations, expand translational research capacity,

and support education and training activities in RD genomics. In parallel, the project seeks to increase awareness of RDs among healthcare professionals, researchers, and the public through various communication and dissemination activities [14].

A major component of RUDiP is its multidisciplinary advisory model for unresolved RD cases. Inspired by NIH-UDN and UDNI frameworks [5, 7], the RB model was adapted for a resource-variable diagnostic landscape through a research-integrated reevaluation model designed to facilitate collaborative interpretation and multidisciplinary expert input.

Accordingly, rather than focusing solely on individual genetic discoveries or technical methodologies, this study examines the establishment, structure, and first two years of the RB model for rare and undiagnosed diseases. We argue that sustainable progress in RD diagnostics depends not only on technological advances but also on coordinated organizational structure, interdisciplinary collaboration, and reinterpretation.

Methods

Ethical approval

The RareBoost Project was approved by the IBG Non-interventional Research Ethics Board, with protocol number 2023-039, in compliance with the principles of the Declaration of Helsinki [15]. Informed consent was obtained from all participants or their legal guardians prior to participation in the study.

Study design and conceptual framework

RB was established as a research-integrated framework to address structural limitations in diagnostics for rare and undiagnosed diseases and to improve diagnostic pathways and research capacity for unresolved cases in Türkiye. The RB team integrates expertise in clinical genetics, molecular biology, and bioinformatics, and is led by a senior researcher with over 20 years of experience in RDs in Türkiye. A dedicated analysis unit coordinates clinician communication, phenotyping, genomic interpretation, and genotype–phenotype correlation, supported by bioinformatics specialists who develop and maintain in-house pipelines and data resources. The laboratory unit oversees sample tracking, biobanking, and experimental studies, while molecular genetics expertise supports variant interpretation and the design of functional assays, particularly for non-coding and splicing variants. Administrative staff ensure operational coordination across the project, whereas the training unit organizes educational activities and transferable-skills programs related to RDs. Communication and dissemination activities are coordinated by the outreach unit to enhance project visibility, stakeholder engagement, and public awareness of RDs. The multidisciplinary organization and iterative diagnostic workflow established within RB are summarized in Fig. 2a.

Clinical network and case enrollment

RB monthly advisory board was initiated through targeted outreach to clinicians and healthcare institutions, initially within the Izmir region and later expanded nationwide. This network facilitates referral of patients with unresolved conditions.

Referrals are initiated through direct clinician–researcher interaction, including physical or virtual consultation, or via the RB multidisciplinary advisory board. Convening monthly, the advisory board evaluated presubmitted cases through detailed discussions. To date, 22 advisory board meetings have been conducted, during which 100 patients have been reviewed (Fig. 2b).

Referring physicians submitted clinical information using a standardized phenotyping form, which is reviewed by RB clinical geneticists to ensure consistent and comprehensive phenotypic assessment. All cases are evaluated by clinical geneticists prior to acceptance.

Patients were included if they remained undiagnosed despite previous clinical and molecular investigations prior to referral to RB, frequently including WES. Case acceptance was based on (i) persistence of diagnostic uncertainty despite standard diagnostic approaches available within the healthcare system, (ii) availability of sufficient phenotypic data for meaningful reevaluation, and (iii) expert suspicion of an underlying genetic etiology potentially amenable to further genomic or functional investigation.

Following inclusion, cases typically undergo trio-based short-read whole-genome sequencing (WGS). Genomic data are analyzed using the RB bioinformatics pipeline and interpreted within a multidisciplinary setting. In unresolved individuals following trio-WGS, or in those with clinically relevant VUS, complementary approaches such as RNA sequencing are used to generate additional diagnostic evidence. Long-read WGS and additional functional studies are planned for future implementation in selected unresolved individuals.

Samples and biobanking

Whole blood, tissue biopsy, and/or buccal swab samples, along with clinical and demographic data, were collected from patients and their available affected/unaffected parents following informed consent. Consent includes the use of stored materials in future research, ensuring the long-term utility of the cohort. Samples were processed and stored in accordance with institutional standard operating procedures to ensure traceability, integrity, and compliance with ethical and legal regulations. Depending on sample availability and clinical indication, DNA and RNA were isolated from biospecimens, and primary peripheral blood mononuclear cells were isolated from whole blood and archived under controlled conditions. All samples were stored at IBG-Biobank, the national biobank node of Türkiye within the BBMRI-ERIC network, operating in accordance with ISO 20387:2018 (Biotechnology — Biobanking requirements).

Analytical pipeline and data governance

Fragmented workflows and limited integration between clinical and research data remain major obstacles in rare disease diagnostics, consistent with findings from our previous nationwide needs assessment study in Türkiye [13]. These challenges include disconnected data records, inconsistent expert review, limited longitudinal follow-up, non-uniform genomic data processing, and difficulties integrating information across multiple clinical domains. To address this, two internal tools were developed: RarePipe, which standardizes genomic data processing, and RareIndex, which tracks patients and decisions over time.

RarePipe is an in-house, nextflow-based tertiary analysis pipeline. It integrates internal genomic data, filters variants by depth and allele frequency, and prioritizes them using Exomiser with HPO terms. Variant annotation is performed with Ensembl VEP, and a parallel module enables CNV detection [16].

RareIndex is a web-based, on-premises system with restricted access to authorized personnel. It is designed for RD research and requires structured data across multiple clinical domains and sample types [17]. RareIndex tracks patients and biospecimens, records pipeline and tool versions for each analysis, and logs clinical data, analyses, pending actions, and council decisions.

Data is stored according to FAIR principles [18] and complies with Turkish data privacy regulations (KVKK). Patient identifiers are replaced with internal codes at intake, and all sensitive clinical data are stored in an encrypted format within RareIndex.

Results

Cohort characteristics

Prior to enrollment in the RB registry, all cases had undergone extensive genetic testing; nearly all probands had previously undergone WES without a molecular diagnosis. To date, the RB registry cohort included 1,057 individuals from 385 families. Biological samples from 914 individuals were available through the IBG Biobank.

A total of 201 families from across Türkiye completed the RB diagnostic workflow and received diagnostic evaluation reports (Fig. 3a–c). All subsequent analyses refer to this subset unless otherwise specified. Although some families included more than one affected individual, all statistical analyses were based on a single representative proband per family. Accordingly, family-level counts throughout the manuscript correspond to analyses based on representative probands.

Among these 201 probands, 113 (56.2%) were male, and 88 (43.8%) were female. Consanguinity was reported in 83 families (41.3%), absent in 104 (51.7%), and unknown in 14 (7.0%). A family history of similar disease manifestations was present in 57 cases (28.4%), while 123 (61.2%) were isolated cases; data for previous tests were unavailable for 21 families (10.4%).

Approximately 170 probands (84.5%) were pediatric (<18 years) at the time of enrollment. The mean age at evaluation was 11.3 years. The mean time between symptom onset and study enrollment was approximately 11 years, indicating a substantial diagnostic delay in a predominantly pediatric-onset cohort. Among the 120 families with clinically relevant findings (pathogenic/likely pathogenic or VUS), 17 received reports containing more than one variant relevant to the phenotype.

Diagnostic outcomes

The RB diagnostic workflow was completed for 201 families in the study cohort (201/385). A definitive molecular diagnosis (P/LP variants) was established in 55 cases (27.4%), while clinically relevant variants of uncertain significance (VUS) were identified in 65 (32.3%). The remaining 81 cases (40.3%) remained unsolved (Fig. 3b). Diagnostic outcomes varied across clinical subgroups, with differences in the rates of definitive molecular diagnosis and in the distribution of clinically relevant VUS findings across disease categories (Fig. 3c). A subset of the RB cohort and selected molecular findings were previously reported in a genomics-focused study describing diagnostic variants, analytical approaches, and illustrative cases in greater detail [19].

Reanalyses of previous WES

Reanalysis of previously generated WES data was performed in 106 cases prior to WGS. Unlike the prospectively generated trio-WGS data within the RB workflow, prior WES datasets were heterogeneous in sequencing age, testing strategy, and originating centers. Nevertheless, all available WES data were systematically reanalyzed before proceeding to trio-WGS.

Reanalysis yielded P/LP variants in 15 cases (14.2%), 25 clinically relevant VUSs in 25 cases (23.6%), and negative results in 66 cases (62.3%). The identified P/LP findings were established without additional sequencing and resulted from a combination of updated variant classification, renewed variant calling, and deeper phenotyping.

Among the 66 unresolved cases following WES reanalysis, 38 subsequently underwent trio-based WGS. Within this subgroup, WGS identified 6 pathogenic or likely pathogenic findings that were not possible to be detected by prior WES analyses, including deep intronic and structural variants. In addition, 12 clinically relevant VUS findings were identified, while 20 cases remained unresolved after WGS.

Whole genome sequencing

WGS was performed for 133 families, including 38 families previously evaluated by WES reanalysis. This yielded P/LP variants in 40 cases (30.1%), clinically relevant VUSs in 40 cases (30.1%), and 53 unsolved cases (39.8%). Unresolved cases were either selected for additional investigations, including long-read genome sequencing, or remained under follow-up for future reanalysis.

Among the 80 families with clinically relevant variants identified through WGS (P/LP or VUS), 20 harbored variants considered not reliably detectable by adequately performed trio-based WES. These included deep intronic variants, copy-number variants, intergenic variants, and variants in RNU genes that were not routinely captured by earlier WES designs. The remaining 60 cases (75.0%) involved variants detectable by trio-based WES data.

Illustrative case: Resolving a 25-year diagnostic odyssey

The impact of the research-integrated UDP model is best illustrated by a 26-year-old female patient referred via virtual consultation from a remote province in Türkiye. Born to consanguineous parents, the patient developed symptoms early in infancy, presenting with congenital hypoglycemia, hepatomegaly, and neutropenia. Despite a strong clinical suspicion of glycogen storage disease (GSD), she remained undiagnosed for 25 years, having previously undergone multiple evaluations, including WES and MLPA, all of which were inconclusive.

Upon enrollment in RB, since prior WES raw data were unavailable for reanalysis, the patient was prioritized for trio-based WGS. Although the initial automated bioinformatic pipeline did not flag any candidate variants, the high clinical index of suspicion for GSD type Ib led to manual visualization of the *SLC37A4* (MIM *602671) locus using the Integrative Genomics Viewer (IGV). This manual curation identified a homozygous pathogenic variant that had been systematically missed. Retrospective analysis revealed that the pipeline misannotated the variant to an overlapping gene due to a more severe predicted consequence in that alternative transcript. This technical issue might explain the false-negative results in previous WES reports.

The identification of this variant confirmed the diagnosis of GSD type Ib (MIM #232220) at age 26. Following the molecular diagnosis, the patient's clinical management was immediately adjusted to a disease-specific treatment protocol, effectively ending a 25-year diagnostic odyssey.

This case reflects a broader trend within the RB cohort regarding clinical actionability. Among the 55 families in which a P/LP variant was identified, 12 (21.8%) involved actionable genes, in which the molecular diagnosis directly informed or altered clinical management (Supplementary Table 1). This corresponds to 12 of the 120 families (10.0%) with clinically relevant findings in their reports (P/LP and VUS cases combined), for whom the RB framework provided findings that could support immediate and meaningful medical action.

Capacity Building, Outreach, and Networking Activities

Regular education and training activities were conducted to support capacity building and improve awareness and understanding of RDs among clinicians, researchers, trainees, patients, and patient organizations. Over the past two years, RB organized RD-focused seminars, workshops, and symposia

aimed at strengthening research capacity and facilitating knowledge exchange. These activities covered a broad range of topics, including clinical practice, diagnostics, research methodologies, ethics, health policies, and patient management. In addition, up to 20 transferable skills training sessions were conducted, engaging approximately 2,000 early-career researchers.

Communication and dissemination activities were also conducted to increase the visibility of RUDiP's research and strengthen stakeholder engagement. Over the last two years, RB has organized more than 50 RD-focused events with almost 4,000 attendees. These included conferences, symposia, seminars, and hackathons targeting researchers, clinicians, patients, patient representatives, students, and early-career professionals, not only nationally, but also participants across the MENA (Middle East and North Africa) region and underrepresented countries.

RB is also actively involved in ERDERA (European Rare Diseases Research Alliance, https://erdera.org/) activities within work streams focused on the clinical research network, education and training, and international capacity alignment, in which RB coordinates the national mirror group Türkiye, serving as a hub that unites around 30 key national stakeholders in rare diseases. These activities have helped connect RUDiP's diagnostic and research efforts with the broader RD community, thereby supporting RB's overarching objectives.

Discussion

RB was established following nationwide surveys and stakeholder workshops that identified major barriers in RD diagnostics in Türkiye and translated these findings into a research-integrated diagnostic framework [13]. Starting the project with a clearer understanding of real-world diagnostic limitations shaped the model in several important ways. Although access to genomic technologies was, and still remains, a significant challenge in Türkiye, our findings support the view that technology itself is no longer the only limiting factor. In many real-world settings, the greater challenge now is integration, reinterpretation, and reanalysis of existing clinical and genomic data. Diagnostic success, therefore, increasingly relies on sustained multidisciplinary interpretation rather than single-point testing approaches. [20].

Consistent with findings from previous RD cohorts, most patients experienced prolonged diagnostic odysseys before inclusion, despite extensive prior evaluations that frequently included WES, repeated specialist consultations, and assessments across multiple centers [21]. The cohort was predominantly pediatric, reflecting the early onset of many RDs, yet the mean diagnostic delay still approached 11 years. This means that many patients remained undiagnosed throughout critical developmental periods and often progressed into later disease stages before reaching a structured multidisciplinary evaluation framework.

Within this context, iterative reanalysis proved particularly valuable. Re-evaluation of existing genomic data identified pathogenic or likely pathogenic variants in 15 cases without additional sequencing, highlighting how unsolved cases can benefit from continued reinterpretation as analytical tools and

clinical knowledge evolve. Beyond its near-zero marginal diagnostic cost, this approach also increases the long-term value of the initial genomic test. These findings underscore the importance of maintaining accessible, well-structured genomic datasets that can be systematically revisited over time rather than treated as static outputs of single testing events [22].

Supporting infrastructures such as RarePipe [16] and RareIndex [17] reduced fragmentation between laboratory analysis and clinical interpretation. Standardized analytical workflows, patient tracking, and the integrated organization of clinical and genomic data improved continuity across repeated evaluations and facilitated coordinated reinterpretation across teams and disciplines. In many unsolved cases, this kind of structure proved just as important as the sequencing itself, particularly in a resource-variable setting where fragmentation of analyses and metadata can easily disrupt continuity of evaluation [23].

Within this framework, a definitive molecular diagnosis was achieved in 27.4% of families (55/201). When clinically relevant VUS findings were included, the proportion of potentially meaningful findings increased to 59.7%. While VUS findings do not constitute definitive diagnoses, they often represent intermediate steps toward future resolution, particularly in unresolved cohorts where supporting evidence remains limited. Collectively, these observations support the view that diagnostic resolution in RDs is often progressive rather than binary. Diagnostic outcomes also varied across clinical subgroups, with the highest definitive diagnostic yields observed in inborn errors of metabolism and neuromuscular disorders, whereas neurodevelopmental and heterogeneous conditions were more likely to remain unsolved (Fig.3b). This pattern likely reflects differences in genetic architecture, phenotypic specificity, and the current limits of variant interpretation. The relatively high rate of consanguinity within the cohort also influenced the diagnostic strategy and contributed to the utility of trio-based WGS in unresolved cases [24]. In many patients, the integration of trio-WGS with deep phenotyping and multidisciplinary reinterpretation improved candidate variant prioritization and facilitated the identification of homozygous and compound-heterozygous pathogenic variants, particularly in complex or previously unsolved cases [25,26]. These findings further emphasize that in unresolved RDs, genetic testing alone is often insufficient without integrating genomic data, phenotypic context, and longitudinal expert reevaluation.

Beyond molecular resolution itself, another outcome of diagnosis was potential clinical actionability. In our cohort, 12 of 201 families (6.0%) had findings with direct or potential clinical implications, corresponding to 12 of the 55 definitively diagnosed families (21.8%). For these patients, molecular diagnosis had the potential to extend beyond genetic counseling to inform treatment decisions, surveillance strategies, dietary interventions, and preventive management. The case of a 26-year-old patient diagnosed with GSD type Ib after 25 years illustrates the clinical importance of molecular diagnosis. In this patient, diagnostic resolution directly enabled changes in disease management.

Because the RB cohort consisted predominantly of previously unresolved and highly selected referral cases, diagnostic yields reported here should not be interpreted as representative of unselected RD

populations.

Beyond individual patient impact, structured multidisciplinary diagnostic programs may also reduce the broader healthcare burden [7,12,20]. Prolonged diagnostic odysseys are frequently associated with repeated testing, fragmented referrals, unnecessary procedures, and persistent uncertainty for both patients and clinicians. Prolonged diagnostic odysseys create a substantial long-term economic burden for both patients and healthcare systems. Earlier molecular diagnosis and coordinated diagnostic evaluation may therefore improve clinical outcomes while reducing unnecessary healthcare utilization [1,11].

A major strength of the RB framework was its ability to sustain interaction among clinicians, researchers, and bioinformatics teams throughout the diagnostic process. This collaborative structure was particularly important in a resource-variable setting, where diagnostic fragmentation across institutions and disciplines might otherwise limit the continuity of evaluation.

Despite these advances, a substantial proportion of cases remained unsolved. This reflects both the biological complexity of rare diseases and the practical challenges of implementing comprehensive genomic investigations within a large collaborative framework [27]. Sustained coordination among clinicians, laboratories, bioinformatics teams, and researchers remains demanding over extended periods, particularly in systems with geographically distributed centers and follow-up. In some cases, incomplete clinical data, limited access to additional family members, or patient loss restricted opportunities for segregation studies, reinterpretation, or additional sampling. Functional validation of candidate variants remains resource-intensive and a major bottleneck in interpreting unresolved RDs [28]. As unsolved cases become more individualized, the expertise required for interpretation also becomes more specialized. Tissue-specific disorders add another layer of difficulty, since obtaining appropriate biopsy material is often not feasible, thereby limiting transcriptomic, proteomic, or other functional analyses in the affected tissue [29]. Funding mechanisms remain largely dependent on relatively slow project-based structures, which require careful prioritization of which candidate variants or unresolved cases to advance to downstream functional investigation. Many of these challenges are inherent to unsolved RD diagnostics and remain difficult to overcome even in highly specialized settings.

Conclusion and Future Perspectives

Our findings show that many unresolved RD cases should no longer be considered permanently unsolved after a single negative genomic test, but instead be revisited through structured multidisciplinary reinterpretation. The RareBoost framework demonstrates that combining clinical evaluation, genomic reanalysis, and collaborative expert interpretation can recover diagnoses even in patients who remained unsolved despite years of prior investigations.

Future development of the RB framework will increasingly focus on unresolved cases requiring interpretation beyond genomic testing alone. While periodic reanalysis, WGS, and improved bioinformatic pipelines can improve diagnostic yield, many cases involving novel missense variants, non-

coding variants, genes not yet associated with human disease, and complex splicing alterations require additional functional evidence. As clinically relevant VUS findings continue to accumulate within the analyzed subgroup of our cohort, the scope of the project is planned to expand beyond variant identification toward functional validation workflows. RNA-based analyses, transcriptomic approaches, long-read sequencing, genome-wide methylation profiling, minigene assays for splice-disrupting variants, and selected patient-derived cellular models are among the techniques currently being implemented within our laboratory and through collaborations with other centers. Long-read WGS is also planned for unresolved cases lacking clinically relevant candidate variants. Periodic genomic reinterpretation and continued clinical reevaluation will remain central components of the RB strategy, as evolving phenotypes, functional evidence, and analytical advances may enable resolution of previously unsolved cases.

As analytical tools and biological knowledge continue to evolve, maintaining long-term access to genomic and phenotypic data will become increasingly important for future diagnostic resolution. In this context, research-integrated, iterative diagnostic programs such as RB may represent an important next step in rare disease diagnostics.

Declarations

Corresponding Author

Correspondence to Ugur Ozbek

Ethics Declarations

Ethical Approval and consent to participate

The RareBoost Project was approved by the IBG Non-interventional Research Ethics Board, with protocol number 2023-039, in compliance with the principles of the Declaration of Helsinki. All families have been informed and have agreed to this publication of history and data. No images and videos of the patient have been included in this work.

Funding

We would like to acknowledge the support of the RareBoost project, funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 952346. The authors also acknowledge the support of TÜSEB for enabling genome sequencing analyses for patients included in this study. These initiatives have substantially strengthened research and diagnostic capacity in rare and undiagnosed diseases at the Izmir Biomedicine and Genome Center and across Türkiye, fostering progress in diagnostics and further studies of rare and undiagnosed conditions.

Author Contribution

MMT: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, and review & editing. AY: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft and review & editing. AKU: Investigation, Visualization, Writing - original draft and review & editing.EO: Data curation, Investigation, Methodology, Visualization, Writing - original draft and review & editing. MP: Data curation, Formal analysis, Investigation, Writing - original draft, and review & editing. BS: Data curation, Investigation, Methodology, Visualization, Writing - original draft, review & editing. KB: Data curation, Visualization, review & editing. IY: Investigation, Writing - original draft, review & editing. BU: Investigation, data curation, review & editing.EET: Investigation, data curation, review & editing.NF: Investigation, data curation, review & editing.SA: Data curation, Methodology, Writing - original draft, review & editing.TB: Investigation, review & editing. BA: Investigation, review & editing. EI: Review & editing. EY: Investigation, Supervision, review & editing. UO: Conceptualization, Investigation, Supervision, review & editing.

Acknowledgement

The authors thank the Izmir Biomedicine and Genome Center Biobank and Biomolecular Resources Platform (IBG-Biobank) for use of its facilities and services, as well as for collecting, processing, storing, and making available the biological samples and data used in this study. This work was funded by the RB Project, funded by the European Union's Horizon 2020 research and innovation programme (Grant No. 952346; https://rareboost.ibg.edu.tr/), and by the ERDERA Project (Grant No. 101156595).We also gratefully acknowledge TÜSEB for enabling genome sequencing analyses for our patients. We sincerely thank all clinicians, researchers, and members of the Rare and Undiagnosed Disease Platform (RUDiP) who contributed to the Rare and Undiagnosed Patients Advisory Board meetings. Furthermore, we acknowledge Gennext and Genomics&More platforms (PhiTech Inc.) for their genomic and transcriptomic services.We are especially grateful to the patients and their families for their trust, participation, and contribution to this work.

Data Availability

The data that support the findings of this study are available from the corresponding author on request. The data are not publicly available due to privacy or ethical restrictions.

References

1. Chung CCY, Project HKG, Chu ATW, Chung BHY. Rare disease emerging as a global public health priority. Front Public Health. 2022;10:1028545. https://doi.org/10.3389/fpubh.2022.1028545.

2. Boycott KM, Hartley T, Biesecker LG, Gibbs RA, Innes AM, Riess O, et al. A Diagnosis for All Rare Genetic Diseases: The Horizon and the Next Frontiers. Cell. 2019;177:32–7. https://doi.org/10.1016/j.cell.2019.02.040.
3. Kernohan KD, Boycott KM. The expanding diagnostic toolbox for rare genetic diseases. Nat Rev Genet. 2024;25:401–15. https://doi.org/10.1038/s41576-023-00683-w.
4. Wallis M, Bodek SD, Munro J, Rafehi H, Bennett MF, Ye Z, et al. Experience of the first adult-focussed undiagnosed disease program in Australia (AHA-UDP): solving rare and puzzling genetic disorders is ageless. Orphanet J Rare Dis. 2024;19:288. https://doi.org/10.1186/s13023-024-03297-5.
5. Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, et al. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. Genet Sci. 2012;14:51–9. https://doi.org/10.1038/gim.0b013e318232a005.
6. Cortial L, Nguyen C, Julkowska D, Cocqueel-Tiran F, Moliner AM, Blin O, et al. Managing rare diseases: examples of national approaches in Europe, North America and East Asia. Rare Dis Orphan Drugs J. 2022;1:10. https://doi.org/10.20517/rdodj.2022.02.
7. Taruscio D, Baynam G, Cederroth H, Groft SC, Klee EW, Kosaki K et al. The Undiagnosed Diseases Network International: Five years and more! Mol Genet Metab. 2020. pp. 243–54. https://doi.org/10.1016/j.ymgme.2020.01.004
8. EURORDIS Rare Barometer. Diagnostic journey of people living with a rare disease. Rare Dis Europe (EURORDIS). 2024. https://www.eurordis.org
9. Splinter K, Adams DR, Bacino CA, Bellen HJ, Bernstein JA, Cheatle-Jarvela AM, et al. Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. N Engl J Med. 2018;379:2131–9. https://doi.org/10.1056/NEJMoa1714458.
10. The 100,000 Genomes Project Pilot Investigators. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. N Engl J Med. 2021;385:1868–80. https://doi.org/10.1056/NEJMoa2035790.
11. Hatirnaz Ng O, Sahin I, Erbilgin Y, Ozdemir O, Yucesan E, Erturk N, et al. Obstacles and expectations of rare disease patients and their families in Turkiye: ISTisNA project survey results. Front Public Health. 2022;1049349. https://doi.org/10.3389/fpubh.2022.1049349.
12. Zurek B, Ellwanger K, Vissers LELM, Schüle R, Synofzik M, Töpf A, et al. Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases. Eur J Hum Genet. 2021;29:1325–31. https://doi.org/10.1038/s41431-021-00859-0.
13. Durmus S, Yucesan E, Aktug S, Utz B, Caglayan AO, Gencpinar P, et al. Management of rare and undiagnosed diseases: insights from researchers and healthcare professionals in Türkiye. Front Public Health. 2025. https://doi.org/10.3389/fpubh.2024.1501942.
14. Utz B, Taskin EE, Yigit A, Ozzeybek E, Bulut K, Sonmezler E, et al. The Rare Hackathon: An extracurricular educational event to promote rare disease awareness among students in medicine and life sciences. Rare. 2025;3:100106. https://doi.org/10.1016/j.rare.2025.100106.

15. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;2191–4. https://doi.org/10.1001/jama.2013.281053.
16. Salman B, Ozzeybek E. Ozbek-Lab/rarepipe: Initial Release (v0.0.1). Zenodo;2026. https://doi.org/10.5281/zenodo.18169888
17. Ozzeybek E, Salman B. Ozbek-Lab/rareindex: v0.9.0 (v0.9.0). Zenodo;2026. https://doi.org/10.5281/zenodo.20326160
18. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.
19. Yigit A, Pekerbas M, Salman B, Bulut K, Ozzeybek E, Topaloglu MM et al. Integrative and systematic genomic approaches to improve diagnosis in rare and undiagnosed diseases: results from the RareBoost project. Research Square; 2026. Preprint. Available from: https://doi.org/10.21203/rs.3.rs-8703835/v1
20. Laurie S, Steyaert W, De Boer E, Polavarapu K, Schuermans N, Sommer AK, et al. Genomic reanalysis of a pan-European rare-disease resource yields new diagnoses. Nat Med. 2025;31:478–89. https://doi.org/10.1038/s41591-024-03420-w.
21. Faye F, Crocione C, Anido De Peña R, Bellagambi S, Escati Peñaloza L, Hunter A, et al. Time to diagnosis and determinants of diagnostic delays of people living with a rare disease: results of a Rare Barometer retrospective patient survey. Eur J Hum Genet. 2024;32:1116–26. https://doi.org/10.1038/s41431-024-01604-z.
22. Costain G, Jobling R, Walker S, Reuter MS, Snell M, Bowdin S, et al. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. Eur J Hum Genet. 2018;26:740–4. https://doi.org/10.1038/s41431-018-0114-6.
23. Rockowitz S, LeCompte N, Carmack M, Quitadamo A, Wang L, Park M, et al. Children's rare disease cohorts: an integrative research and clinical genomics initiative. npj Genom Med. 2020;5:29. https://doi.org/10.1038/s41525-020-0137-0.
24. Khayat AM, Alshareef BG, Alharbi SF, AlZahrani MM, Alshangity BA, Tashkandi NF. Consanguineous Marriage and Its Association With Genetic Disorders in Saudi Arabia: A Review. Cureus [Internet]. 2024 [cited 2026 May 22]; https://doi.org/10.7759/cureus.53888
25. Akgun-Dogan O, Tuc Bengur E, Ay B, Ozkose GS, Kar E, Bengur FB, et al. Impact of deep phenotyping: high diagnostic yield in a diverse pediatric population of 172 patients through clinical whole-genome sequencing at a single center. Front Genet. 2024;1347474. https://doi.org/10.3389/fgene.2024.1347474.
26. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Abel HJ, Regier AA, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort, including 602 trios. Cell. 2022;185:3426–e344019. https://doi.org/10.1016/j.cell.2022.08.004.

27. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am J Hum Genet. 2017;100:695–705. https://doi.org/10.1016/j.ajhg.2017.04.003.
28. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet. 2017;101:315–25. https://doi.org/10.1016/j.ajhg.2017.07.014.
29. Hou H, Yuki KE, Costain G, Szuto A, Barnes S, Ramani AK et al. Assessing the diagnostic impact of blood transcriptome profiling in a pediatric cohort previously assessed by genome sequencing [Internet]. Genetic and Genomic Medicine; 2024 [cited 2026 Jan 6]. https://doi.org/10.1101/2024.12.03.24317221

Figures

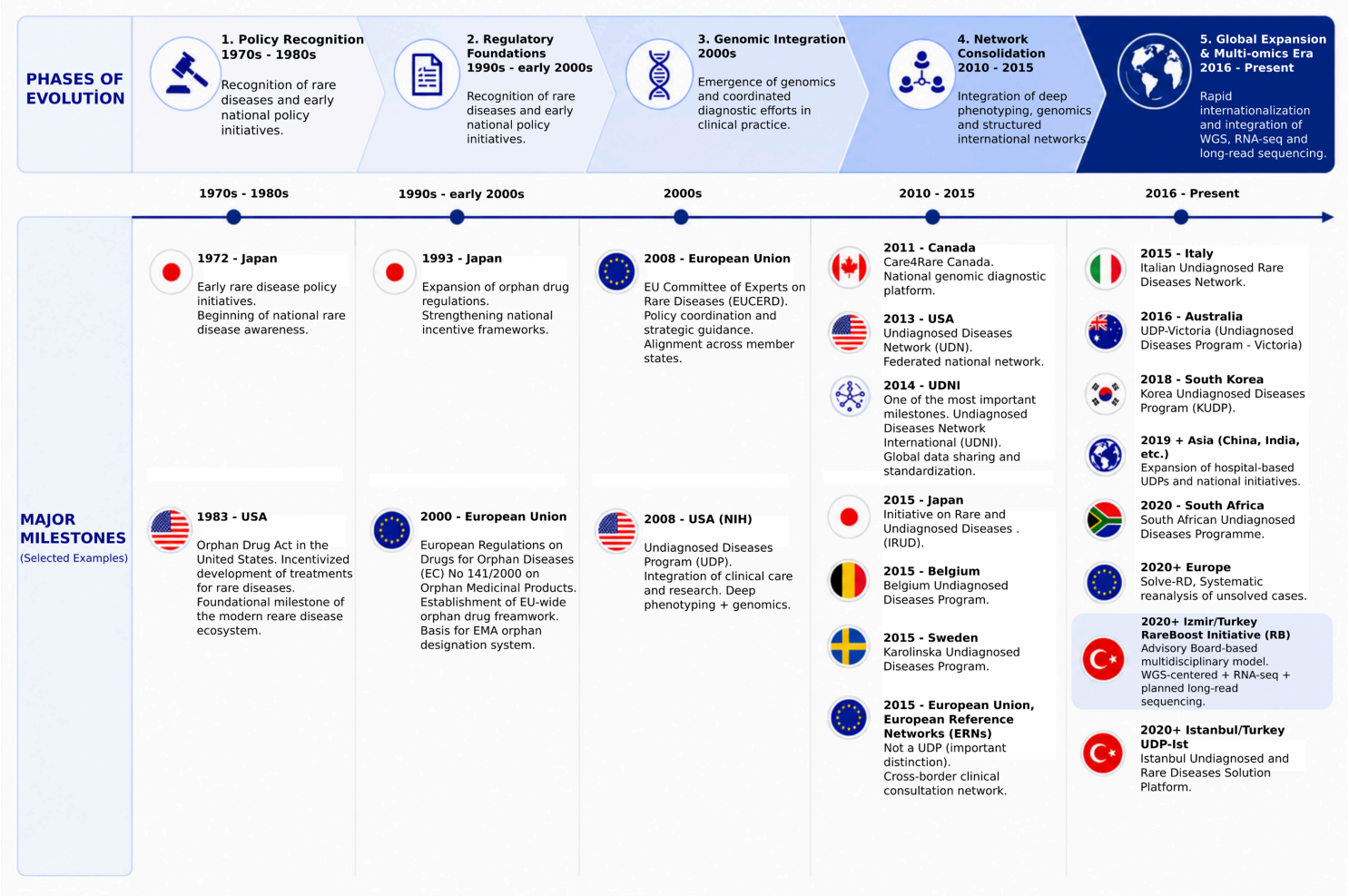


Figure 1

Global evolution of undiagnosed diseases programs (UDPs) and the emergence of RareBoost within the international RD diagnostic landscape from the 1970s to the present. The timeline illustrates major milestones in RD policy, orphan drug regulation, genomic integration, and international network

development across five phases: Policy Recognition, Regulatory Foundations, Genomic Integration, Network Consolidation, and the Global Expansion/Multi-omics Era. Selected international examples demonstrate the progressive expansion of coordinated UDP initiatives and the integration of advanced genomic technologies, including WGS, RNA-seq, and long-read sequencing.

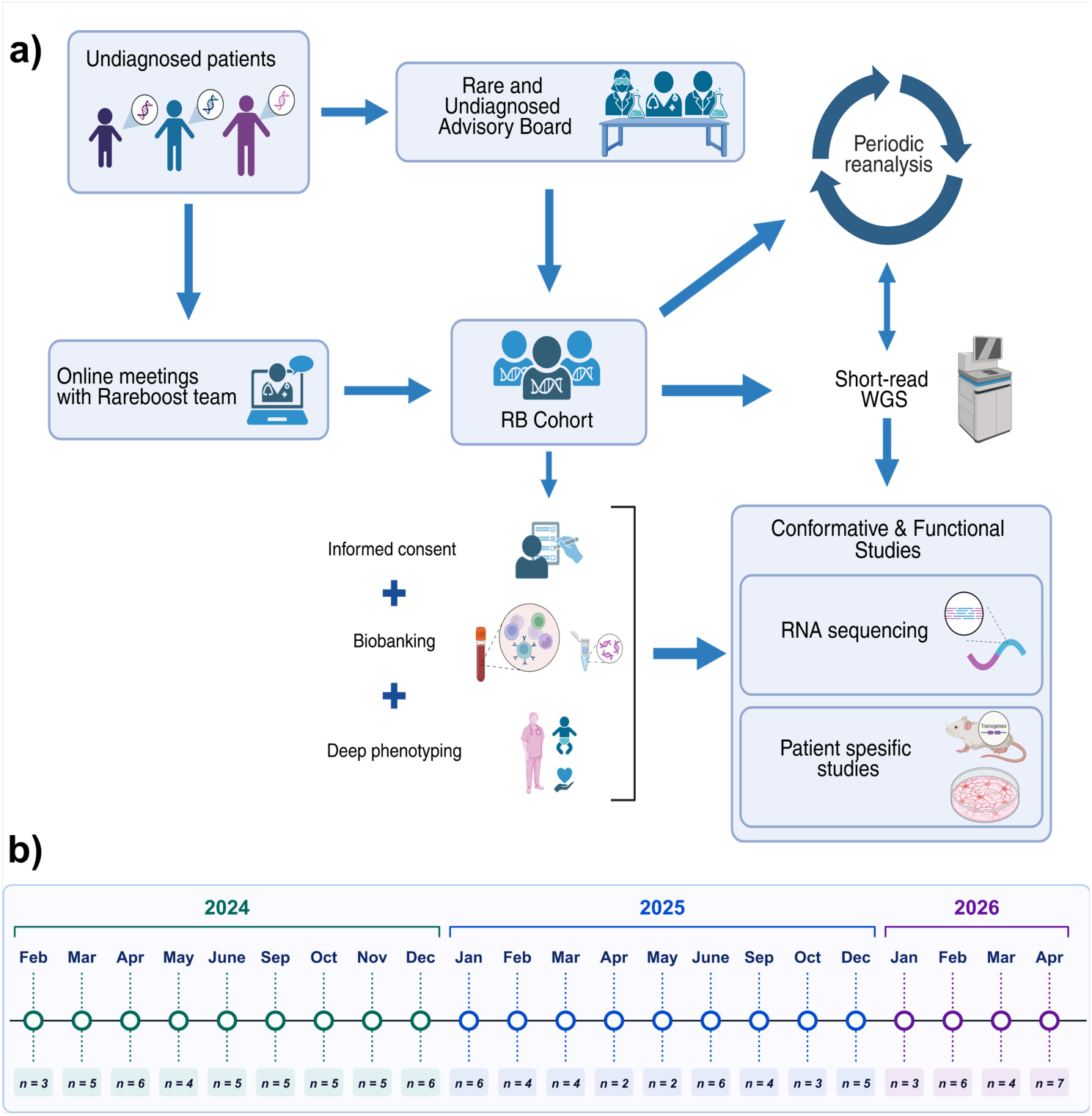


Figure 2

a) Schematic overview of the RB workflow for unresolved RD cases, including multidisciplinary advisory board evaluation, online meetings, biobanking, deep phenotyping, genomic testing, confirmatory and functional studies, and periodic reanalysis. **b)**Timeline of RB advisory board meetings and the number of cases presented over time.

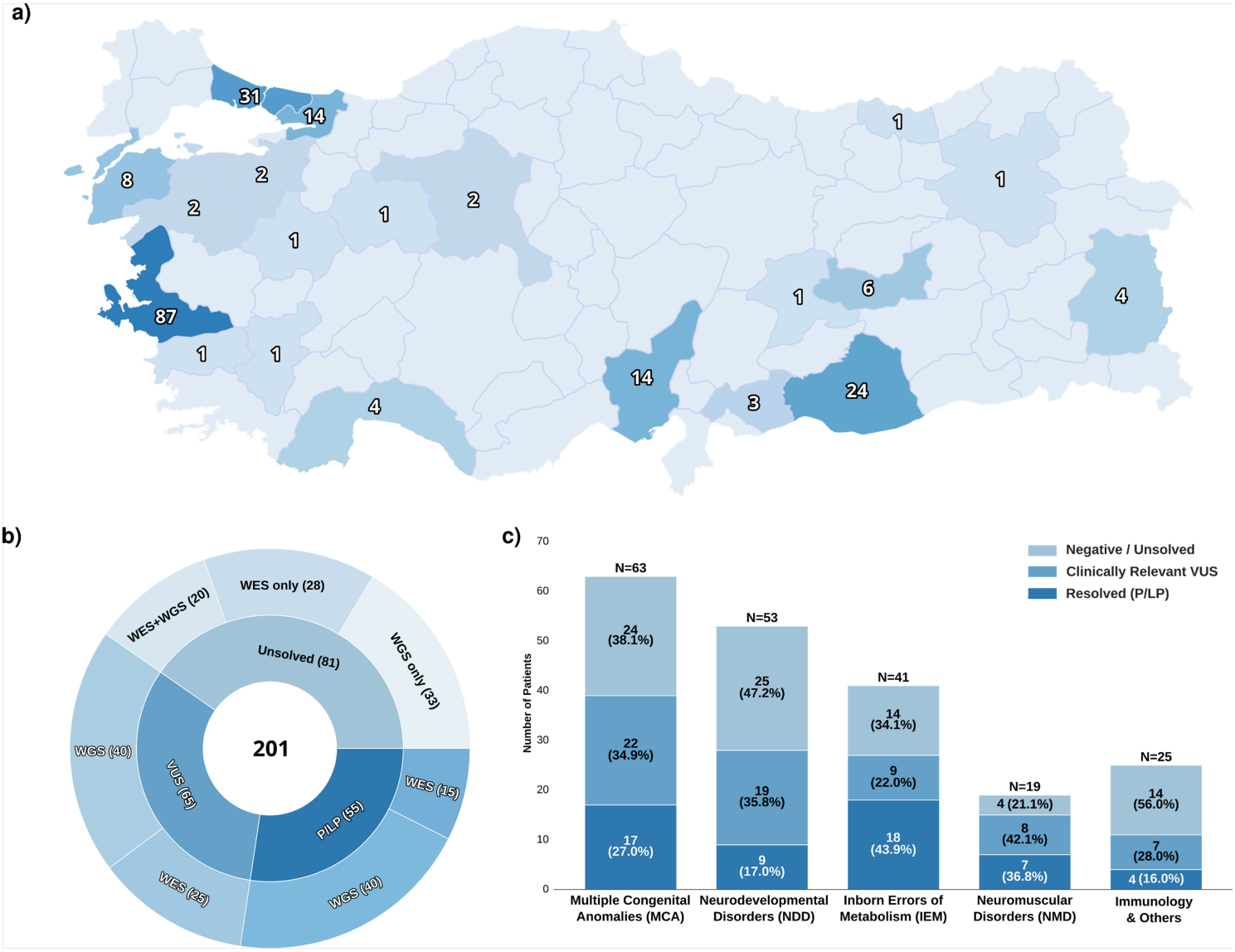


Figure 3

a) Geographic distribution of 201 families across Türkiye. **b)**Diagnostic outcomes of the 201 families stratified by method (WES reanalysis vs. WGS). **c)** Diagnostic outcomes of the 201 families stratified by disease group

Supplementary Files

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

- Supplementarytable1.docx