Six years’ accomplishment of the Initiative on Rare and Undiagnosed Diseases: nationwide project in Japan to discover causes, mechanisms, and cures

Yuji Takahashi, Hidetoshi Date, Hideki Oi, Takeya Adachi, Noriaki Imanishi, En Kimura, Hotake Takizawa, Shinji Kosugi, Naomichi Matsumoto, Kenjiro Kosaki, Yoichi Matsubara, IRUD Consortium, and Hidehiro Mizusawa.

The identification of causative genetic variants for hereditary diseases has revolutionized clinical medicine and an extensive collaborative framework with international cooperation has become a global trend to understand rare disorders. The Initiative on Rare and Undiagnosed Diseases (IRUD) was established in Japan to provide accurate diagnosis, discover causes, and ultimately provide cures for rare and undiagnosed diseases. The fundamental IRUD system consists of three pillars: IRUD diagnostic coordination, analysis centers (IRUD-ACs), and a data center (IRUD-DC). IRUD diagnostic coordination consists of clinical centers (IRUD-CLS) and clinical specialty subgroups (IRUD-CSSs). In addition, the IRUD coordinating center (IRUD-CC) manages the entire IRUD system and temporarily operates the IRUD resource center (IRUD-RC). By the end of March 2021, 6301 pedigrees consisting of 18,136 individuals were registered in the IRUD. The whole-exome sequencing method was completed in 5136 pedigrees, and a final diagnosis was established in 2247 pedigrees (43.8%). The total number of aberrated genes and pathogenic variants was 657 and 1718, among which 1113 (64.8%) were novel. In addition, 39 novel disease entities or phenotypes with 41 aberrated genes were identified. The 6-year endeavor of IRUD has been an overwhelming success, establishing an all-Japan comprehensive diagnostic and research system covering all geographic areas and clinical specialties/subspecialties. IRUD has accurately diagnosed diseases, identified novel aberrated genes or disease entities, discovered many candidate genes, and enriched phenotypic and pathogenic variant databases. Further promotion of the IRUD is essential for determining causes and developing cures for rare and undiagnosed diseases.

INTRODUCTION

Identification of the causative genetic variants of hereditary diseases has revolutionized clinical medicine to enhance diagnostic accuracy, understand disease pathogenesis, and develop therapies. The major technical breakthrough behind the revolution was the development of whole-genome sequencing (WGS) and whole-exome sequencing (WES) methods employing next-generation sequencing (NGS), enabling identification of causative genetic variants by simultaneously capturing all candidate variants potentially causing diseases in affected individuals. Additionally, tremendous amounts of variant data have been generated using sequencers and remarkable advances have been made in bioinformatics exploiting such big data with powerful computational data analysis methods. Accumulation of variant data via registration in public databases has also accelerated the discovery of pathogenic variants by filtering known variants that are not associated with diseases.

However, there still remain many diseases for which causative genetic variants have not been identified. According to Online Mendelian Inheritance in Man (OMIM) (URL: https://www.omim.org/), 9514 hereditary diseases were registered as of November 2021, of which 3288 diseases with (suspected) Mendelian basis are classified as having an unknown molecular basis. A complete understanding of the molecular basis of these diseases is one of the ultimate goals of human molecular genetics, which remains highly challenging even in the NGS era. The difficulty lies in identifying multiple pedigrees with pathogenic variants in the same genes to verify causality, particularly when researchers attempt to discover causative genetic variants for ultra-rare diseases, which are thought to comprise most unsolved diseases [1]. Even when
a novel candidate gene is identified in a given pedigree with an ultra-rare disorder, it is almost impossible to discover a second pedigree within a single institute or a single consortium, known as the N-of-1 problem. Extensive data sharing across multiple institutions and international collaboration is key to overcoming this difficulty.

To solve the N-of-1 problem, the formation of extensive collaborative frameworks with international cooperation has become a global trend [2]. The Undiagnosed Diseases Network in the USA [1], Genomics England in England [3] and Finding of Rare Disease Genes in Canada [4] are the three leading projects that have achieved unprecedented success in identifying causative genetic variants of many rare disorders. In 2011, the International Rare Disease Research Consortium (IRDiRC) (https://irdirc.org/about-us/history) established a worldwide network to connect individual projects to conduct international collaborative studies, further promoting the discovery of causative genetic variants [5]. In 2015, the Agency for Medical and Research Development (AMED) participated in IRDiRC as the first organization from Japan [6] and set specific action goals to promote international collaboration through further data sharing to contribute to the AMED-affiliated IRDiRC 10-year plan [7]. Furthermore, Undiagnosed Disease Network International was established in 2015 to build a consensus framework of principles, best practice, and governance involving these projects [8]. One of the key factors behind the success of these projects has been the development of the MatchMaker Exchange (MME), an extensive data-sharing system connecting genomic and phenotypic databases based on a unified computational architecture and common application programming interface [9]. MME adopts human phenotype ontology (HPO) as the standardized language to describe phenotypes [10] and facilitates the computation of phenotype and genotype matching to identify multiple pedigrees with the same aberrated genes.

Even when causative genetic variants have been established, genetic diagnosis of rare diseases remains difficult in clinical settings. One reason for this is that a limited number of patients undergo WGS- or WES-based genetic diagnosis. Accessibility is among the key factors that can drive the utilization of this innovative technology in clinical genetics, necessitating a nationwide infrastructure to regionally equalize this testing. Additionally, determining pathogenic variants among a large number of variants yielded by WGS/WES and establishing the final diagnosis in which the pathogenic variants fully account for clinical manifestations is labor-intensive. Particularly, a substantial number of rare and undiagnosed diseases present with complex phenotypes with multiple affected organs, making the determination difficult by researchers in a single specialty. Therefore, it is recommended that diagnostic boards composed of physicians with a wide range of specialties and geneticists should discuss the final diagnosis based on the phenotypes and WGS/WES data.

Enhancing the diagnostic accuracy of rare diseases has been vigorously pursued in Japan. Remarkable achievements have been made in the research and countermeasures for rare and intractable diseases, designated as “Nan-byo,” by the Ministry of Health, Labor and Welfare in Japan, which was established in 1972 after the subacute myelo-optic neuropathy endemic. In 2015, a new intractable disease law was enforced to expand “Nan-byo” from 56 to 333 diseases, further promoting the diagnosis of rare and intractable diseases. Nevertheless, two surveys conducted by AMED showed that more than 37,000 cases remained undiagnosed [11].

To address these issues, the Initiative on Rare and Undiagnosed Diseases (IRUD) was launched in 2015 as a nationwide project in Japan supported by AMED [6]. IRUD aims to establish accurate diagnoses, discover causes, and ultimately provide cures for diseases through nationwide coverage of comprehensive diagnostic systems, utilization of innovative tests including NGS, and construction of an internationally sharable clinical database [11]. Initially, IRUD for pediatric patients (IRUD-P) was launched in July 2015, followed by IRUD for adult patients (IRUD-A) in October. In 2017, the two were integrated into one project as the IRUD to make the project more extensive and comprehensive. This study describes the accomplishments of the 6-year effort of the IRUD project, illustrating the whole diagnostic system, diagnostic yield and pathogenic variant landscape of rare and undiagnosed diseases, novel genes/disease entities, and human resource development.

MATERIALS AND METHODS

IRUD entry criteria

The IRUD entry criteria are as follows [6].

1. The patient remains undiagnosed for ≥6 months (not necessary for infants) and suffers from disabilities in daily life, AND
2. 1–2. Objective signs exist that cannot be attributed to a single organ; OR
2–2. Direct or indirect evidence exists of a genetic etiology (e.g., similar symptoms found in the patient’s relatives) [6].

Here, an undiagnosed disease is clearly distinguished from an undetermined disease in which a clinical diagnosis has been made but its causative genetic variants have not been confirmed. For example, if spinocerebellar degeneration is clinically diagnosed, although its causative genetic variants have not been analyzed and disease type has not been determined, it is classified as an undetermined disease and excluded from the IRUD. This clearly distinguishes the IRUD from genetic diagnostic services.

Genomic analysis

Blood samples were obtained from the participants fulfilling the criteria mentioned above with informed consent, DNA samples were extracted, and B-cell lymphoblast cell lines were established. Genomic DNA was subjected to enrichment of exonic sequences using Agilent SureSelect (Agilent Technologies, Santa Clara, CA, USA). Massively parallel sequencing (100-base pair paired-end reads) was performed using NGS (Hiseq2500; Illumina, San Diego, CA, USA). The sequences were aligned to the reference genome (GRCh37/hg19) using the Burrows-Wheeler Aligner or NovoAlign. Removal of potential polymerase chain reaction duplicates, recalibration of base quality values, local realignment, and variation calls were performed using ANNOVAR [11] and GenRead (https://annovar.openbioinformatics.org/) together with RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/). 1000 Genomes Project Database (http://www.1000genomes.org/), dbSNP135 (http://www.ncbi.nlm.nih.gov/projects/SNP/), gnomAD (http://www.gnomad.broadinstitute.org), and in-house databases.

Data sharing

A data-sharing platform, IRUD Exchange, was designed by incorporating the Patient Archive system [12] that complies with HPO and can be linked with the MME. The architecture of the IRUD Exchange allows IRUD researchers to conduct similarity searches using pattern-matching algorithms as a powerful tool to address ‘N-of-1’ problems of rare diseases. The IRUD Exchange also facilitates the registration of HPO-based phenotypes by adopting a user-friendly interface that automatically translates clinical summaries written in Japanese into English and highlights relevant HPO terms.

Central ethics committee

Initially, the IRUD-P started with approval from the individual institutional ethics committee. Subsequently, a central ethics committee (CEC) was established in IRUD-A as one of the leading model projects in AMED to facilitate the ethical review process in multi-institutional large-scale collaborative research. All but one of the individual institutional ethical committees in IRUD-A delegated the reviewing process to the CEC. The delegation process had been further promoted upon the integration of IRUD-A and IRUD-P.
IRUD Diagnostic Coordination

IRUD Clinical Center (IRUD-CL), IRUD Diagnostic Board (IRUD-DB), IRUD Clinical Specialists Subgroup (IRUD-CSS), IRUD Data Center (IRUD-DC), IRUD Analysis Center (IRUD-AC), IRUD Resource Center (IRUD-RC), and IRUD Coordination Center (IRUD-CC).

IRUD Coordination Center (IRUD-CC).

The principal role of IRUD-CC is administration of the whole system through monthly IRUD-PB meetings as the highest decision-making organization. The constituents of IRUD-PB include representatives of the IRUD-CC (principal investigator), AMED (program officers) as the funding agency, IRUD-CLS, IRUD-CSSs, IRUD-ACs, and IRUD-DCs. IRUD-CC drafts a unified research protocol that is ratified by IRUD-PB and subjected to CEC, with the one-step approval process contributing to timely modification of the IRUD research. In addition, IRUD-CC monitors the progress of the entire research by conducting a regular survey. IRUD-CC also operates the sample and information logistics system described in detail in a subsequent section.

IRUD Clinical Center (IRUD-CL) and Semi-Clinical Center (IRUD-SCL). IRUD-CL/SCL is the only contact site for patients with IRUD. Upon patient entry, parents are recruited so that trio analysis can be performed. IRUD-CL/SCL operates the IRUD Diagnostic Board (IRUD-DB), which manages the process from patient entry decisions to final diagnosis establishment. IRUD-SCLs are designed to fill the geographic gaps of IRUD-CLS, although no funding was provided. IRUD Cooperative Hospitals refer candidates for IRUD entry to IRUD-CL/IRUD-SCL. The IRUD-DB plays a central role in the regional diagnostic network with IRUD Cooperative Hospitals. The IRUD-DB is composed of pediatricians and physicians of various specialties/subspecialties for adults, clinical geneticists, genetic counselors, and data scientists. The IRUD protocol stipulates the participation of representative physicians from local medical associations in each IRUD-DB to promote regional cooperation.

IRUD-DB holds regular meetings during which thorough pre-entry evaluation is conducted based on the clinical information described in a regular format on a ‘patient sheet’ to determine whether the candidate is suitable for IRUD entry and if sufficient investigation has already been completed, including whether clinical workups and available genetic tests such as chromosome analysis or gene-panel analysis have been performed. Similarly, post-analysis evaluation is conducted at the regular meetings to determine whether the pathogenic variant reported by IRUD-AC fully accounts for the clinical phenotypes leading to the final clinical diagnosis. In addition, the activities of the IRUD-DB include genetic counseling, further follow-up and reevaluation of the pedigree, public relations, and human resource development. Thus, IRUD-CL/SCL plays an essential role in the clinical aspects of the IRUD system.

IRUD Clinical Specialty/Subspecialty Subgroup (IRUD-CSS). IRUD-CSSs are organized by assembling members of IRUD-DBs across entire IRUD-CLS according to their specialties/subspecialties. IRUD-CSSs support the activities of individual IRUD-DBs and provide professional advice based on their specialties/subspecialties for cases that cannot be resolved by IRUD-CLS alone.

Thus, IRUD-CLS and IRUD-CSSs form the IRUD Diagnostic Coordination covering entire geographic areas and specialty/subspecialty fields in Japan.

IRUD Analysis Center (IRUD-AC). Each IRUD-CL/IRUD-SCL sends DNA samples via an outsourcing provider to a corresponding single IRUD-AC, which conducts comprehensive genomic analysis, identifies pathogenic variants, and reports to the IRUD-CL/IRUD-SCL via the IRUD-CC. When pathogenic variants are undetermined, further intensive research is conducted to identify novel aberrated genes/pathogenic variants using WGS, multi-omics analysis, and functional studies.

IRUD Data Center (IRUD-DC). IRUD-DC operates the IRUD Exchange, the data-sharing platform described above, promotes data sharing among IRUD researchers, and serves as a gateway to domestic and international collaboration. Phenotypic and genomic information has been accumulated to promote the establishment of new causative genetic variants and disease concepts. The IRUD Exchange is also used as a database to understand the overall epidemiologist landscape of rare and undiagnosed diseases registered in the IRUD. All IRUD-CLS and IRUD-SCLs have a computer terminal for the IRUD-Exchange, and phenotypic and genotype data are transferred to IRUD-DC through a specific virtual private network to ensure security.

IRUD Resource Center (IRUD-RC). IRUD-RC establishes a resource repository for clinical information and DNA samples/B cell lines and manages a utilization committee for examining the utilization of repositories. The IRUD-CC temporarily serves as the IRUD-RC,
which is planned to be established as an independent facility assigned by AMED.

**IRUD workflow of samples/information**

IRUD-CC organizes a unified workflow to facilitate sample and information sharing among the IRUD-CLs/IRUD-SCLs, IRUD-ACs, and IRUD-CC and establishes a centralized repository in IRUD-RC (Fig. 2).

Upon entry into the IRUD, an individual identification (ID) number, composed of a three-alphabetical institutional code and five-digit numerical number, is allotted to each registrant, whose DNA/cell line samples, clinical information, and analysis results are handled with a specific ID. The workflow utilizes an outsourcing provider to extract genomic DNA samples and establish B lymphoblast cell lines. The genomic DNA samples are sent to the IRUD-ACs and IRUD-CC, and B lymphoblast cell lines are sent to the IRUD-CC. Each IRUD-CL/IRUD-SCL sends the samples to a specific IRUD-AC designated by IRUD-CC. Clinical information in the form of a 'patient sheet' is also sent to IRUD-ACs via an outsourcing provider. The analysis results and clinical information are sent from the IRUD-ACs to the IRUD-CC and reported to the corresponding IRUD-CLs. Clinical information is accumulated in the IRUD-DC through IRUD-Exchange from the IRUD-CLs. Text-based clinical data on the patient sheet delivered from the IRUD-CL/SCL via the IRUD-AC and analysis reports delivered from the IRUD-AC are also stored in the IRUD-CC. Genomic DNA samples and B lymphoblast cell lines are deposited in the IRUD-RC.

**Present status and activities of the IRUD system**

In March 2021, the IRUD diagnostic system comprised 450 institutions consisting of 37 IRUD-CLs, 15 IRUD-SCLs, and 398 cooperative hospitals (Fig. 3). Five IRUD-CLs also serve as IRUD-ACs, one of which also serves as the IRUD-DC. The National Center of Neurology and Psychiatry serves as the IRUD-CC, IRUD-CL, and IRUD-RC. Twenty-one IRUD-CCs included 497 clinical specialists to support IRUD-DBs in the IRUD-CLs (Table 1). The IRUD-RC established resource repositories, including 4489 genomic DNA samples and 3017 lymphoblastic cell lines. Phenotypes and genetic data of 5378 pedigrees have been registered on the IRUD Exchange, among which 62 are shared internationally through the MME. Thirty-two IRUD-CLs have delegated the ethics review.
process to CEC, the remaining 5 utilize their own institutional review boards.

**Diagnostic yield and pathogenic variant landscape**

In March 2021, 6301 pedigrees consisting of 18136 individuals were registered in the IRUD. WES was completed in 5136 pedigrees, with a final diagnosis established in 2247 pedigrees (43.8%) (Fig. 4). The total numbers of aberrated genes and pathogenic variants in these pedigrees were 657 and 1718, respectively; 1113 (64.8%) of the variants were novel (Fig. 5a). Among the 2247 pedigrees, the most frequently identified causative gene was CHD7, which was identified in 31 pedigrees, followed by MEFV in 27 pedigrees and ARID1B in 25 pedigrees. In contrast, 298 aberrated genes were causative in single pedigrees, exhibiting a long-tail distribution of gene frequencies (Fig. 5b) (Supplementary Table 1). Most pathogenic variants are unique, whereas some have been identified in multiple pedigrees. Notably, a known pathogenic variant p.E148Q in MEFV and known p.N308S mutation in PTPN11 were identified in eight and five pedigrees, respectively, making these pathogenic variants relatively common among undiagnosed diseases in the Japanese population.

**Novel aberrated genes or disease entities**

Thirty-nine novel disease entities or phenotypes along with 41 aberrated genes were identified in the IRUD (Fig. 4, Supplementary Table 2). These genes were classified into three categories. Category 1: Novel disease entities with novel aberrated genes. Category 2: Novel disease entities to which unique OMIM IDs are assigned to novel pathogenic variants in known aberrated genes. Category 3: Novel phenotypes in known disease entities with novel pathogenic variants of known aberrated genes. Category 1 comprised 23 disease entities with 24 genes [13–32], category 2 comprised 4 disease entities with 4 genes [33–36], and category 3 comprised 12 disease entities with 13 genes [37–47]. In addition, among the remaining 2889 families in whom causative genetic variants had not been established, 313 pedigrees belonged to the N-of-1 category, in which only one candidate gene per pedigree...
was determined by WES but remained as a single family, awaiting further discovery of additional pedigrees.

**Human resource development**

The IRUD offers an outstanding opportunity for human resource development through on-the-job training. In fiscal years 2018, 2019, and 2020, there were 78, 87, and 24 IRUD members and related people who received promotions in their institutes; 26, 34, and 0 were promoted in other institutes; 12, 47, and 8 became staff inside their institutes; 10, 5, and 0 were given jobs in companies associated with clinical genetics; 56, 62, and 38 achieved certification by the Japanese Board of Medical Genetics and Genomics, Clinical Genetics; and 17, 50, and 12 obtained certification as Genetic Counselors, respectively. IRUD has also contributed to the training of data scientists: 64, 65, and 5 medical doctors and 32, 38, and 4 non-medical doctors participated in genome informatics analysis as data scientists in 2018, 2019, and 2020, respectively.

**DISCUSSION**

The 6-year endeavor of the IRUD has been an overwhelming success, establishing an all-Japan comprehensive diagnostic and research system for rare and undiagnosed diseases covering entire geographic areas and clinical specialties/subspecialties. The IRUD has led to the accurate diagnoses, identification of novel aberrated genes or disease entities, discovery of many candidate genes, enrichment of phenotypic and pathogenic variant databases, and development of treatments and cures. It also has established a fundamental infrastructure for both centralized
governance with unified protocols, logistics, biorepositories, data sharing and ethics and individual autonomous research activities among the IRUD-AC, IRUD-DC, and IRUD-CL/SCL. Thus, the IRUD is a unified, sustainable medical and research system that can be expanded to all fields of genomic medicine.

The IRUD has established an accurate diagnosis for a large number of patients and, in some cases, led to the use of specific therapies with positive effects [40]. Pathogenic variants have been identified in 657 genes, encompassing more than one-tenth of all aberrated genes registered in the OMIM. The diagnostic yield is 43.8% (2247 in 5136 pedigrees), which is comparable to that of the Undiagnosed Diseases Network (30%; 427 in 1413 pedigrees) (https://undiagnosed.hms.harvard.edu/about-us/facts-and-figures/). Particularly, approximately one-half of pedigrees with an established diagnosis possess unique aberrated genes. Importantly, such diseases are individually ‘ultra-rare’ but not as a whole, necessitating further rigorous endeavors to provide accurate diagnosis and develop therapeutic measures for individual ‘ultra-rare’ diseases.

The IRUD has had important impacts on basic research by identifying many novel aberrated genes, establishing novel disease entities, and detecting novel pathogenic variants in known aberrated genes. Disease-causing genes are the most reliable pathogenic molecules with greatest impact on the development, course and prognosis of the disease. Identification of these genes has led to an increased understanding of disease pathogenesis and identification of druggable seeds, promoting research of rare and undiagnosed diseases, common diseases, and human physiology. Furthermore, the IRUD has greatly contributed to genomic medicine by identifying a large number of novel pathogenic variants in known aberrated genes. Approximately two-thirds of the identified pathogenic variants were novel, supporting the indispensable role of the IRUD in genomic medicine in addition to its diagnostic services. Aggregation of phenotypic and pathogenic variant data can enrich variant databases such as ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), LOVD (https://www.lovd.nl/), and MGeND (https://mgend.med.kyoto-u.ac.jp/) as well as disease databases such as OMIM and Orphanet (https://www.orpha.net/).

A remarkable feature of the IRUD is the extensive data sharing through the IRUD Exchange, which has accumulated HPO-based phenotypic and genetic data to enable searches for similar cases automatically by pattern-matching algorithms and solving N-of-1 problems. The IRUD Exchange is compatible with MME and functions as a gateway to domestic and international collaborative networks. The IRUD is one of the principal eight nodes of international genomic research projects connected via the MME. The collaboration network provides a collective dataset spanning more than 150,000 cases from more than 11,000 contributors in 88 countries [48]. Thus, the IRUD greatly contributes to international data sharing.

The IRUD-CL/SCL plays an essential role in covering broad aspects of genomic medicine as the only contact site for patients, where the IRUD-D8 conducts pre-entry and post-analysis evaluation; establishes a final clinical diagnosis; and offers opportunities for genetic counseling, patient follow-up, public relations, regional cooperation, transitional medicine, and human resource development. Thus, the IRUD-CL/SCL is a clinical core facility not only in genomic medicine but also in medical systems for rare and intractable diseases, Nan-byo. Moreover, IRUD-D8 has substantially improved both the clinical and research levels of IRUD in a cost-effective manner. Cost-effectiveness has been achieved by utilizing the diagnostic process of the national healthcare insurance system in Japan so that pre-screening for IRUD entry is covered by the system, enabling funding for the IRUD to be concentrated on research. Therefore, IRUD-D8s play a major role on both the quality and cost-effectiveness by conducting thorough pre-entry investigation within the health care system to maintain the optimum standard for IRUD entry best-suited for comprehensive genomic analysis and discovery of novel disease genes/entities.

Finally, IRUD has promoted human resource development for genome medicine or research and for rare diseases (Nan-byo). Activities in the IRUD encompass entire fields of genome medicine or research, including phenotyping, genetic analysis, informatics, diagnosis, and counseling. Experiences in the IRUD would be of great help in individual career development. The IRUD also provides excellent educational opportunities for medical geneticists, genetic counselors, and data scientists. Therefore, the IRUD is not only a nationwide diagnostic and research system, but also a sustainable human resource development system in Japan.

Notwithstanding the exhaustive WES, nearly 3000 pedigrees remained undiagnosed. Particularly, 313 pedigrees belonged to the N-of-1 category, for which discovery of another pedigree with the same candidate genes should be definitely necessary. Further promotion of international collaboration is the key to address the N-of-1 issue. On top of that, reasons for the undetermined causes are thought to include mosaicism, genomic alterations, gene regulation, and complex inheritance, all of which are difficult to capture by WES [49]. To address these issues, the IRUD has begun to adopt the cutting-edge strategies of WGS, long-read sequencing, RNA sequencing, epigenetics, proteomics, and metabolomics analysis. IRUD is one of the leading projects of the National Execution Plan for WGS, a national project promoting WGS-based genomic medicine. Furthermore, IRUD Beyond has been launched to provide a prioritized opportunity for IRUD to conduct functional studies using animal models and therapeutic studies using induced pluripotent stem cells or genome editing [50]. Taken together, IRUD continues to move forward until carrying out its mission to determine causes and provide cures for all the rare and undiagnosed diseases.

DATA AVAILABILITY
All the data and materials are available for academia, researchers or private enterprises, either domestic or abroad, upon examination and permission of usage proposals by IRUD Promotion Board (IRUD-PB) and CEC. Usage for profitable researches is restricted to those which contribute to the progress of medical fields. The sequence data of known pathogenic variants are available in Genetic Medicine Japan Variant Database (MGeND) (https://mgend.med.kyoto-u.ac.jp/) supported by AMED.

REFERENCES
1. Gahl WA, Wise AL, Ashley EA. The Undiagnosed Diseases Network of the National Institutes of Health: a national extension. JAMA. 2015;314:1797–8.
2. 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.
3. Turnbull C. Introducing whole-genome sequencing into routine cancer care: the Genomics England 100 000 Genomes Project. Ann Oncol. 2018;29:784–7.
4. Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, et al. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. Am J Hum Genet. 2014;94:809–17.
5. Ayme S. IRDRIC-recommended. Eur J Hum Genet. 2016;24:955.
6. Adachi T, Kawamura K, Furusawa Y, Nishizaki Y, Imanishi N, Umehara S, et al. Japan’s initiative on rare and undiagnosed diseases (IRUD): towards an end to the diagnostic odyssey. Eur J Hum Genet. 2017;25:1025–8.
7. Austin CP, Dawkins HS. Medical research: Next decade’s goals for rare diseases. Nature. 2017;548:158.
8. Taruscio D, Groff SC, Cederroth H, Melegh B, Lasko P, Kosaki K, et al. Undiagnosed Diseases Network International (UDNI): White paper for global actions to meet the N-of-1 problem, logistics, biorepositories, data sharing and ethics and individual autonomous research activities among the IRUD-AC, IRUD-DC, and IRUD-CL/SCL. Thus, the IRUD is a unified, sustainable medical and research system that can be expanded to all fields of genomic medicine.
9. Adachi T, Kawamura K, Furusawa Y, Nishizaki Y, Imanishi N, Umehara S, et al. Japan’s initiative on rare and undiagnosed diseases (IRUD): towards an end to the diagnostic odyssey. Eur J Hum Genet. 2017;25:1025–8.
10. Austin CP, Dawkins HS. Medical research: Next decade’s goals for rare diseases. Nature. 2017;548:158.
11. Taruscio D, Groff SC, Cederroth H, Melegh B, Lasko P, Kosaki K, et al. Undiagnosed Diseases Network International (UDNI): White paper for global actions to meet the N-of-1 problem, logistics, biorepositories, data sharing and ethics and individual autonomous research activities among the IRUD-AC, IRUD-DC, and IRUD-CL/SCL. Thus, the IRUD is a unified, sustainable medical and research system that can be expanded to all fields of genomic medicine.
12. Adachi T, Kawamura K, Furusawa Y, Nishizaki Y, Imanishi N, Umehara S, et al. Japan’s initiative on rare and undiagnosed diseases (IRUD): towards an end to the diagnostic odyssey. Eur J Hum Genet. 2017;25:1025–8.
13. Austin CP, Dawkins HS. Medical research: Next decade’s goals for rare diseases. Nature. 2017;548:158.
11. Adachi T, Imamishi N, Ogawa Y, Furusawa Y, Izumida Y, Izumi Y, et al. Survey on patients with undiagnosed diseases in Japan: potential patient numbers benefiting from Japan’s initiative on rare and undiagnosed diseases (IRUD). Orphanet J Rare Dis. 2018;13:208.

12. Groza T, Kohler S, Doeßen K, Collier N, Oellrich A, Smedley D, et al. Automatic concept recognition using the human phenotype ontology reference and test suite corpora. Database J Biol Databases Curation. 2015;2015:bav0085.

13. Niñori T, Ouchi-Uchiyama M, Sasahara Y, Kaneko T, Hashii Y, Irie M, et al. Mutations in MECOM, encoding oncoprotein EV1, cause radialular synostosis with amegakaryocytic thrombocytopenia. Am J Hum Genet. 2015;97:848–54.

14. Takenouchi T, Kosaki R, Nizuma T, Hata K, Kosaki K. Macrothrombocytopenia and developmental delay with a de novo CDCA42 mutation: Yet another locus for thrombocytopenia and developmental delay. Am J Med Genet A. 2015;167A:2822–5.

15. Takenouchi T, Miura K, Uehara T, Mizuno S, Kosaki K. Establishing SON in 21q22.11 as a cause of a new syndromic form of intellectual disability: Possible contribution to Braddock-Carey syndrome phenotype. Am J Med Genet A. 2016;170:2587–90.

16. Makrythanasis P, Kato M, Zaki MS, Saitsu H, Nakamura K, Santoni FA, et al. RRAS2 mutations cause Noonan syndrome. Am J Hum Genet. 2019;104:1233–40.

17. Minatogawa M, Takenouchi T, Tsuyusaki Y, Iwasaki F, Uehara T, Kurosawa K, et al. CDK19-related mutations in three male patients with West syndrome. J Hum Genet. 2016;61:653–61.

18. Mizukami M, Takano K, Tsusuki Y, Yoshitomi S, Shimono M, Aoki Y, et al. WDR45 mutations cause early infantile epileptic encephalopathy. Ann Clin Transl Neurol. 2019;6:2334.

19. Hiraide T, Nakashima M, Yamoto K, Fukuda T, Kato M, Ikeda H, et al. De novo SF3B1 variants in SETD1B are associated with intellectual disability, epilepsy and autism. Am J Hum Genet. 2018;102:1960.

20. Lardelli RM, Schaffer AE, Eggens VR, Zaksis M, Grainger S, Sauh T, et al. Biallelic variants in the 3′ exonucleosome TOE1 cause pontocerebellar hypoplasia and uncover a role in snRNA processing. Nat Genet. 2017;49:457–64.

21. Shaw ND, Brand H, Kuchpinska Z, Bengani H, Plummer L, Jones TI, et al. SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arthria and Bosma arthria microphthalmia syndrome. Nat Genet. 2017;49:328–48.

22. Groza T, Kohler S, Doelken S, Collier N, Oellrich A, Smedley D, et al. Automatic concept recognition using the human phenotype ontology reference and test suite corpora. Database J Biol Databases Curation. 2015;2015:bav0085.

23. Wada Y, Kikuchi A, Arai-Ichinoi N, Sakamoto O, Takezawa Y, Iwasawa S, et al. Pathogenic variants in PIGG cause intellectual disability with seizures and hypotonia. Am J Hum Genet. 2016;98:615–26.

24. Levy-Carvalho A, Casas MT, Sequeira-Navarro I, Climent-Pujol MV, González-Morcillo L, et al. Mutations in PIGG cause intellectual disability with seizures and hypotonia. Am J Hum Genet. 2016;98:615–26.

25. Hiraide T, Nakashima M, Yamato K, Fukuda T, Kato M, Ikeda H, et al. De novo variants in SETD1B are associated with intellectual disability, epilepsy and autism. Am J Hum Genet. 2018;102:1960.
