ANNUAL REVIEWS

Annual Review of Genomics and Human Genetics

New Diagnostic Approaches for Undiagnosed Rare Genetic Diseases

Taila Hartley,[1] Gabrielle Lemire,[1,2] Kristin D. Kernohan,[1,3] Heather E. Howley,[1] David R. Adams,[4] and Kym M. Boycott[1,2]

[1]CHEO Research Institute, University of Ottawa, Ottawa, Ontario K1H 5B2, Canada; email: thartley@cheo.on.ca, glemire@cheo.on.ca, kkernohan@cheo.on.ca, hhowley@cheo.on.ca, kboycott@cheo.on.ca

[2]Department of Genetics, CHEO, Ottawa, Ontario K1H 8L1, Canada

[3]Newborn Screening Ontario, CHEO, Ottawa, Ontario K1H 9M8, Canada

[4]Office of the Clinical Director, National Human Genome Research Institute and Undiagnosed Diseases Program, National Institutes of Health, Bethesda, Maryland 20892, USA; email: david.adams@nih.gov

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

Annu. Rev. Genom. Hum. Genet. 2020. 21:9.1–9.22

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

https://doi.org/10.1146/annurev-genom-083118-015345

This is a work of the US government and not subject to copyright protection in the United States

Keywords

rare genetic diseases, diagnostic odyssey, rare disease diagnosis, data sharing, genomics, omic approaches

Abstract

Accurate diagnosis is the cornerstone of medicine; it is essential for informed care and promoting patient and family well-being. However, families with a rare genetic disease (RGD) often spend more than five years on a diagnostic odyssey of specialist visits and invasive testing that is lengthy, costly, and often futile, as 50% of patients do not receive a molecular diagnosis. The current diagnostic paradigm is not well designed for RGDs, especially for patients who remain undiagnosed after the initial set of investigations, and thus requires an expansion of approaches in the clinic. Leveraging opportunities to participate in research programs that utilize new technologies to understand RGDs is an important path forward for patients seeking a diagnosis. Given recent advancements in such technologies and international initiatives, the prospect of identifying a molecular diagnosis for all patients with RGDs has never been so attainable, but achieving this goal will require global cooperation at an unprecedented scale.

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

1. INTRODUCTION: THE REALITY OF RARE GENETIC DISEASES

Although rare by definition (average global prevalence of 1 in 2,500), collectively the more than 9,000 rare diseases (RDs) recognized to date (84) affect at least 1 in 16 people (36). RDs represent an enormous societal burden, with devastating impacts on patients, their families, and the health-care system (36). RDs are often life limiting: 57–67% of RDs lead to a reduced life expectancy (27, 56). It is widely cited that 80% of all RDs have a genetic cause, but the basis of this number is not well understood; recent review of Orphadata shows that at least 39% of RDs have an identified genetic etiology (36) and would be considered a rare genetic disease (RGD). This genetic subset of RDs often comprises serious multisystem diseases that consume a disproportionate amount of health-care resources. Even when compared with common chronic childhood diseases such as asthma and diabetes, genetic conditions incur significantly higher direct health-care costs (3.5–8.3 times higher per patient) and resource use (73); the reality of RGDs is a sobering one.

Receiving an early and accurate diagnosis is imperative for informed care in medicine, and yet patients and families living with suspected RGDs will often spend more than five years on a diagnostic odyssey. These odysseys typically consist of multiple specialist visits and invasive testing and can carry significant societal and personal costs (42). Sadly, despite a variety of available diagnostic tests, for many this odyssey will ultimately be futile: 50% of RGD patients do not receive a molecular diagnosis (105). Although fewer than 3% of RGDs have a therapy approved by the US Food and Drug Administration (36), accurate genetic diagnosis remains essential for informed care (e.g., genetic counseling, surveillance, coordinated clinical care, and family planning) and patient and family well-being (psychological closure, emotional relief, and access to resources and support) (34). For RGD patients who remain undiagnosed following all available clinical tests, we assert that there is a path forward to identify novel pathogenic mechanisms. Here, we highlight new approaches to RGDs that will hopefully enable diagnoses for all such patients in the coming decade.

2. DIAGNOSTIC CHALLENGES FOR RARE GENETIC DISEASES

Over the past 60 years, our understanding of the genetic mechanisms that can cause RGDs has increased dramatically (**Figure 1**). In 1956, it was determined that humans have 23 pairs of chromosomes, and shortly afterward the cause of Down syndrome was identified (67). Since then, our knowledge of genetic mechanisms of disease has expanded significantly to include chromosomal rearrangements, copy-number variants (CNVs, comprising microdeletions and microduplications), trinucleotide repeats, small insertions or deletions, DNA single-nucleotide variants, mitochondrial DNA mutations, and epigenetic alterations. Diagnostic tools have followed these research insights: Karyotyping and Sanger DNA sequencing have been routinely used for 40 years. The development and clinical implementation of chromosomal microarrays in the early 2000s facilitated the study of chromosomes at a higher resolution and expanded our understanding of the role of CNVs in RGDs. The emergence of next-generation DNA sequencing (NGS) approaches in the same decade allowed high-throughput approaches to DNA sequencing, enabling multiple genes to be sequenced in parallel, saving time and resources. NGS expanded to facilitate genome-wide sequencing (GWS) technologies such as exome sequencing (ES), which targets the protein-coding part of the genome, and whole-genome sequencing (WGS). The application of ES to neurodevelopmental disorders brought to light new disease mechanistic paradigms, including the high frequency of de novo mutations (30). The broader introduction of ES into the clinic has demonstrated that it can provide a diagnosis for 29–57% of RD patients, depending on the indication, technology, timing in the diagnostic pathway, and whether the DNA sequencing was done in a hospital-based or reference laboratory (26). Overall, ES has been instrumental in ushering in a new phase of RGD discovery (7).

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

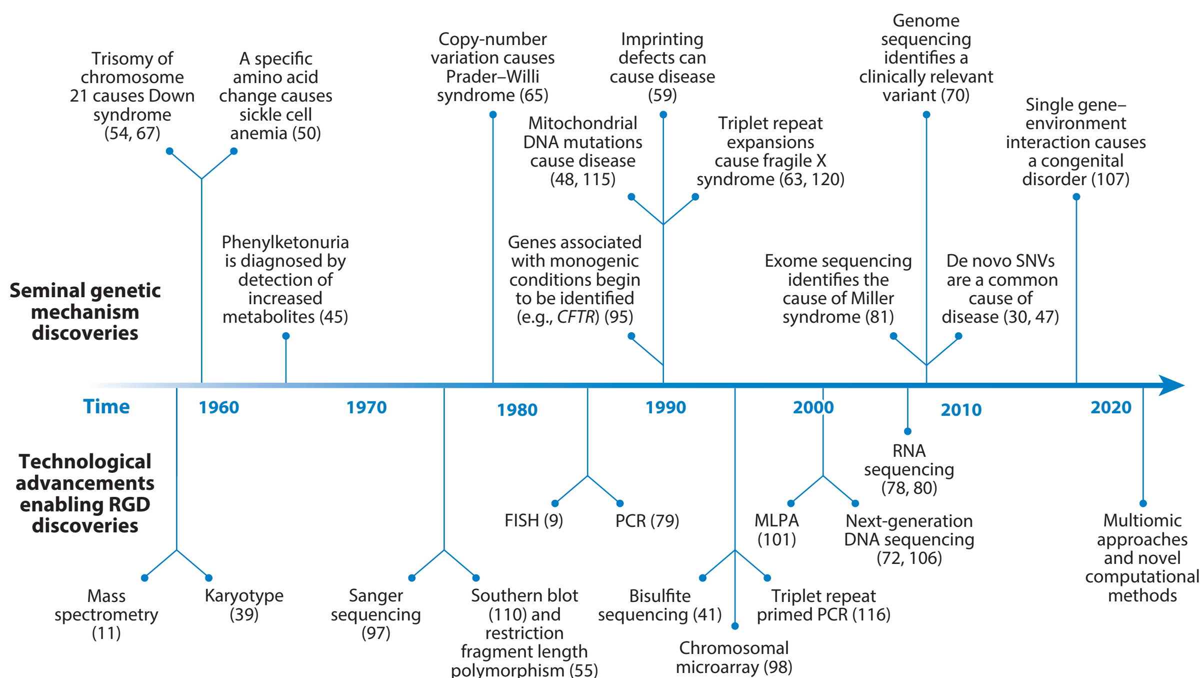

Figure 1

The past 60 years has taken us from the identification of chromosomes and the cause of Down syndrome to an understanding of the cause of more than 5,000 rare genetic diseases (RGDs). Major discovery milestones are denoted along the top of the time line, and the technological advances that enabled these discoveries are shown below it. Multiomic technologies and novel computational approaches are expected to play a significant role in RGD discoveries in the years to come. Abbreviations: FISH, fluorescence in situ hybridization; MLPA, multiplex ligation-dependent probe amplification; PCR, polymerase chain reaction; SNV, single-nucleotide variant.

Despite our increased understanding of the mechanisms of RGDs, the majority of patients remain undiagnosed because they do not have access to appropriate expertise and testing. In addition, for patients living in countries with more developed health-care systems, the current diagnostic paradigm for RGDs is not designed for those who remain undiagnosed after initial investigations because of several challenges, including interpretation of test results and limitations inherent to the paradigm.

2.1. Challenges with Access

The majority of the world's population lacks access to appropriate genetic testing, and this represents the greatest challenge for patients with an undiagnosed RGD. Many countries have recognized that people with RGDs systematically experience barriers to accessing health care (82). Whether the patient lives in a country with a seemingly advanced health-care system or in a developing country, patients with an RGD face both unique and common challenges. Many of these challenges stem from the rarity of their disease, which complicates diagnosis and appropriate care; these challenges include a lack of general awareness about RGDs, difficulty in recognizing RGDs in patients with seemingly unrelated symptoms, lack of standard criteria for diagnosis, and minimal physician training in RGDs (42). System challenges also contribute, including shortages of appropriate specialists (including clinical geneticists), uncoordinated health-care systems that make communication between physicians difficult, and limited access to diagnostic tools in certain

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

jurisdictions. The clinical translation of genomic technologies, in particular, has been widely hyped in the RGD community as an opportunity for diagnosis. Nonetheless, GWS is still accessible as a diagnostic test to only a fraction of RGD patients worldwide.

2.2. Challenges with Interpretation of Diagnostic Testing Results

For RGD patients who do receive appropriate testing, the results for a particular patient are often challenging to interpret, which delays diagnostic clarity. Contributing factors include the presence of one or multiple genomic variants of uncertain significance (VUSs), results that explain part but not all of a patient's clinical presentation, and the ongoing discovery of new disease–gene associations.

2.2.1. Variants of uncertain significance. Despite our increased knowledge of genetic mechanisms, we still do not understand the vast majority of genomic variation. The American College of Medical Genetics and Genomics defines a VUS as a genomic variant that has a probability of pathogenicity between 10% and 90% (94). The number of VUSs identified per analysis varies greatly depending on the clinical indication, the number and size of the genes analyzed, current understanding of how genomic variation influences a particular protein's function, and the availability of functional assays. For example, 10–15% of patients being evaluated with a RASopathy gene panel (which has a specific clinical indication and a small number of well-established genes) have a VUS identified (112), whereas 42% of patients being evaluated with a cardiomyopathy gene panel (which has a high rate of phenocopies and a large number of genes) have a VUS identified (85). The growing number of VUSs leads to ambiguous results for patients and represents a significant challenge in RGD diagnosis.

2.2.2. Establishing clinical validity. The clinical validity of genetic testing results is difficult to establish when some of the patient's clinical symptoms are not readily explained by the identified variants. In some cases, this may be the result of phenotypic expansion of a known disease–gene association. As more individuals undergo NGS testing, the identification of atypical presentations of known RGDs increases, broadening the phenotypic variability to better delineate the mild and severe ends of the phenotypic spectrum. For some patients, complex clinical presentations may be secondary to multiple genetic diagnoses in a single family, a phenomenon that occurs at least 5–7% of the time (6, 90).

2.2.3. New disease–gene associations. Our knowledge of disease–gene associations changes daily, with approximately 300 new discoveries reported every year (7). These new associations can be the result of genes identified that have never been previously associated with an RGD (a novel disease gene discovery) or those that have been previously associated with a distinct and different Mendelian disease (a new mechanism of disease). Several clinical laboratories will report compelling variants in genes of uncertain significance (GUSs), which then require identification of additional families with variants in the same GUS and overlapping phenotypes to support novel gene discovery and, ultimately, a diagnosis. A framework for assessing the strength of evidence for disease–gene associations has been published (111) to provide standardization and prevent the assertion of a new disease–gene association that is subsequently proven to be incorrect (32). At least one clinical laboratory has reported candidate GUSs in more than 24% of patients receiving GWS results (93). Indeed, recent estimates suggest that approximately 6,000–13,000 Mendelian conditions are yet to be defined (7). Although this suggests an enormous potential for continued

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

discovery of new disease–gene associations, these associations cannot be found using the current diagnostic paradigm.

2.3. Challenges with the Current Diagnostic Paradigm

For the past 30 years, a patient suspected of having an RGD has experienced a typical diagnostic testing approach that begins with an initial clinical consultation (medical and family history and physical examination), the output of which is a differential diagnosis and, increasingly, a genetic testing strategy. In general, this workflow is quite successful: Almost 50% of patients suspected to have an RGD receive a diagnosis within the first few tests (105). Patients who remain undiagnosed after the current diagnostic paradigm may be missed due to technological limitations of the testing used, insufficient understanding of the coding genome due to limited genomic data sharing, and pathological genetic mechanisms of RGD that are not evaluated using current methodologies.

2.3.1. Technological limitations. A wide variety of diagnostic tools are available, and all of them have technological limitations. Clinicians should therefore recognize that, despite having completed the most appropriate genetic test for their patient, one cannot truly rule out a given RGD; one can only reduce its likelihood. For example, although microarrays have vastly improved our ability to identify CNVs, their limited resolution and reliance on predefined DNA sequences and hybridization means that they can still miss some CNVs. Similarly, NGS technologies have increased our DNA-sequencing capacity, but they, like Sanger DNA sequencing, struggle with GC-rich regions, and any NGS methods that target specific regions of the genome using capture reagents (e.g., capture kits) will struggle with evenness of coverage. Furthermore, various technological limitations of NGS are inherent to the bioinformatic pipelines used to align the genomic data to a reference sequence, and approaches to annotation and filtration can impact the variants that are subsequently identified for analysis. Therefore, the diagnostic testing approach for a patient with an undiagnosed RGD needs to be carefully examined to understand the likelihood that the current methodology will miss a disease mechanism.

2.3.2. Data sharing. Millions of individuals with a suspected RGD have undergone targeted genetic testing, and more than 100,000 individuals have now had GWS (primarily ES) as part of clinical care. It is estimated that by 2025, more than 60 million patients will have had genomic sequencing in the health-care context (13). Yet most of these data sets are difficult, if not impossible, to access because health care is a highly regulated landscape due to the security and privacy issues surrounding data sharing, and possible solutions are both challenging and expensive. Even in the research setting, where patients have agreed to share their data for research, access is often a highly managed process to ensure compliance with the original consent. Strategies to de-silo existing and future data must be developed to fully realize the potential of the collective data to help interpret genomic variation for the benefit of patients.

2.3.3. Disease mechanisms associated with non-Mendelian inheritance. Current NGS and associated informatic analyses either were not designed for or are not capable of detecting genetic variation associated with non-Mendelian inheritance, such as tissue-specific somatic mosaicism, epigenetic alterations, oligogenic inheritance, or gene–environment interactions (16). For example, current informatic approaches may not readily identify mosaicism of a pathogenic variant, and detection is further challenged by tissue specificity and low-level mosaicism. Epigenetic alterations, including methylation defects, uniparental disomy, and altered parent-of-origin expression patterns at an imprinted locus, require specific technological approaches for detection.

Review in Advance first posted on
April 13, 2020. (Changes may still
occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

Furthermore, there are currently no approaches by which to reliably identify oligogenic or complex inheritance, where the cumulative effect of different common and rare variants, in conjunction with environmental triggers in the context of complex disease, causes the clinical presentation of a particular patient. The study of RGDs caused by these mechanisms is currently limited and will require focus and resources in the future.

3. CLINICAL APPROACHES TO UNDIAGNOSED RARE GENETIC DISEASES

Because the current clinical paradigm is not optimally designed for patients with an undiagnosed RGD, it is necessary to expand our approach to diagnostic care. Successfully identifying the majority of genes and variants associated with RGDs will require efficient data sharing in the clinical setting and, importantly, enabling and empowering patients to participate in data sharing an engaged stakeholders in their diagnostic journeys. We propose that patients should be offered a clinical assessment designed for undiagnosed RGDs. Integral to this assessment is recurrent reanalysis of NGS data and sharing of data to increase the likelihood of identifying a genetic diagnosis (**Figure 2**).

3.1. Deep Phenotyping Is Critical for Diagnosing Rare Genetic Diseases

Deep phenotyping of patients with an undiagnosed RGD, and family members if relevant, is critical for the diagnostic approaches that follow. Deep phenotyping in this context refers to the precise and comprehensive annotation of phenotypic abnormalities in a computer-readable format, such as the Human Phenotype Ontology (HPO) (60), a resource recognized by the International Rare Diseases Research Consortium (IRDiRC) (68). Natural history data are not captured using HPO but are available as a separate field in some phenotyping software. Environmental exposures can cause or contribute to disease, but the study of this area is still emerging, and our ability to systematically capture these data is still immature (109). In addition, deep-phenotypic assessment of family members (e.g., siblings, parents, or extended family, as appropriate) can be helpful in some instances. Based on this detailed clinical assessment of the patient and family, clinicians should examine all possible known diseases that may explain the patient's clinical presentation. It is important that the team's diagnostic approach considers atypical presentations, phenotype expansions, conflation of more than one RGD, and recently discovered disease–gene associations.

3.2. Reanalysis of Existing Test Results in Known Disease Genes

Once a patient's phenotype has been documented and analyzed, the next step is to review the testing strategy that did not yield a diagnosis. Reports from previous testing should be collected and examined to determine whether a comprehensive approach was used with regard to genes and disease mechanisms. For patients who have a compelling VUS, one would begin by looking for new evidence of pathogenicity [e.g., in the literature or databases such as ClinVar (46)] or of benign variation [e.g., a high frequency in population databases such as gnomAD (57)]. Additional evidence can come from segregation studies in the family and occasionally functional studies, though the latter are more difficult to ascertain and standardize. Reclassifying a compelling VUS as pathogenic or benign is critical for determining whether updated testing would be useful.

3.3. Reanalysis of Genome-Wide Sequencing Data

In several jurisdictions, clinical ES has become available as a funded test for the most diagnostically complex patients, but it is successful only 36% of the time despite a high index of suspicion of an

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

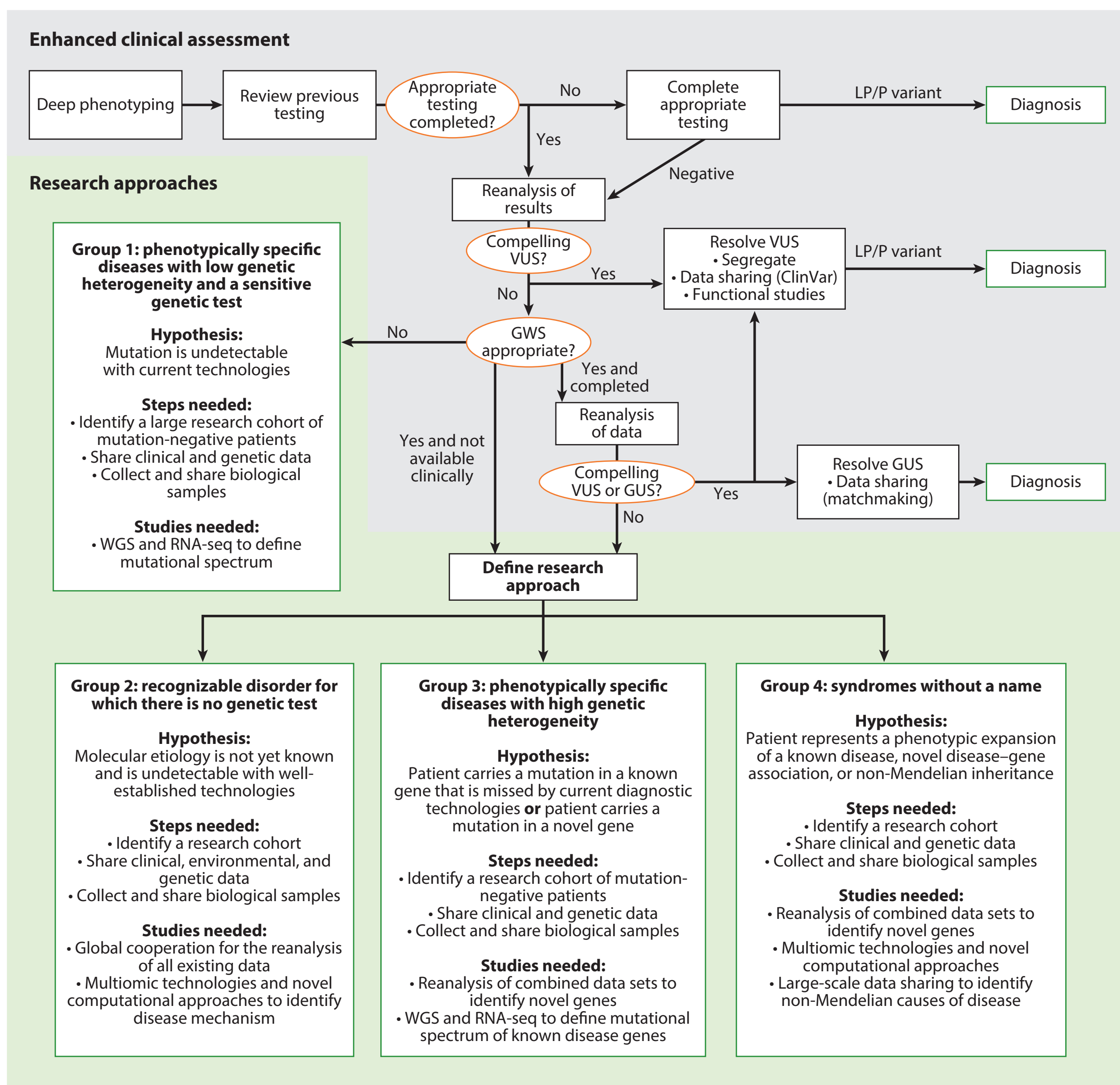

Figure 2

Opportunities for patients with a suspected but undiagnosed rare genetic disease (RGD) exist in both the clinical and research realms. The gray portion of the workflow represents an enhanced clinical assessment for undiagnosed RGD patients, and the green portion shows research approaches that could be considered for broad categories of undiagnosed RGDs. Abbreviations: GUS, gene of uncertain significance; GWS, genome-wide sequencing; LP, likely pathogenic; P, pathogenic; RNA-seq, RNA sequencing; VUS, variant of uncertain significance; WGS, whole-genome sequencing.

RGD (26). While it is not yet clear what proportion of patients harbor mutations in the protein-coding portion of their genome, it is likely far greater than this diagnostic yield. In one study that first mapped Mendelian diseases to genetic loci, more than 95% of families whose disease was correlated to a single locus harbored genic (as opposed to intergenic) variants that were potentially identifiable by ES, and reanalysis of negative clinical ES data identified likely causal variants in

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

88% of these families (104). We and others therefore argue that a negative clinical ES result is a significant starting point for a patient who remains undiagnosed, as there is a high likelihood that reanalysis will identify diagnoses in known and novel genes (7).

The systematic reanalysis of GWS data can yield additional diagnoses for several reasons. First, the immediate reanalysis of GWS data in collaboration with the referring clinician has been shown to boost diagnostic yield in known disease genes by 10%, based primarily on the clinician's understanding of the clinical features (96). Second, in a similar fashion to how new disease–variant information can aid in the reassessment of VUSs in known genes, the approximately 300 new disease–gene associations that are curated each year (7) may identify variants that were not considered at the time of clinical testing. Third, clinicians may, in fact, identify these novel disease–gene associations themselves. Indeed, it was eloquently shown that novel gene discovery via systematic reanalysis of clinical exome data can essentially double the diagnostic yield (33). Fourth, reanalysis in collaboration with a different laboratory using a distinct bioinformatics pipeline may occasionally identify variants that were missed due to technical limitations of the original laboratory. Lastly, reanalysis can allow the clinician to relax various thresholds (e.g., rarity or quality) and analyze the data using different bioinformatics tools to enable exploration of additional hypotheses (118). For example, unaffected or mildly affected parents may pass on mosaic mutations that are also present in germline cells to classically affected children, mimicking autosomal recessive inheritance, which would be easily missed in analysis. For patients who have had GWS, regardless of whether it was analyzed as a large gene panel or all 22,000 genes, the greatest opportunity for a molecular diagnosis likely comes from reanalyzing their existing data, followed by appropriate data sharing.

3.4. Data Sharing as Part of Clinical Care for Undiagnosed Rare Genetic Diseases

To aid in the interpretation of both VUSs and GUSs, significantly more data sharing is needed, and several professional organizations have highlighted the importance of this approach to improving clinical care, including the American College of Medical Genetics and Genomics (1) and the Canadian College of Medical Genetics (15). While data sharing is often considered in the context of enabling exploratory research, we advocate for data sharing to also be considered as a clinical tool to assist with RGD diagnosis. Accordingly, responsible sharing should be considered in the context of providing direct benefit to patients and viewed as an extension of clinical care in certain situations, which would therefore not require explicit research consent (31).

The sharing of genomic variants in known disease genes has been facilitated by the use of public databases such as ClinVar (46) and DECIPHER (17), and initiatives such as ClinGen (92) have worked on reconciling differences in variant interpretation. Many clinical laboratories are actively contributing to these efforts (103), and, in the best interest of patient care, laboratories that do not participate in the global understanding of RGDs should be avoided. For the 15–20% of patients with a compelling GUS following GWS, the reality is that a significant proportion will ultimately be diagnosed by finding additional unrelated cases with variants in the same gene, the same inheritance pattern, and an overlapping phenotype. Multiple projects have addressed this *n*-of-1 challenge by developing platforms to identify cases with overlapping phenotypes and candidate genes through a process referred to as two-sided matchmaking (e.g., two or more individuals are looking for a match in the same gene) (**Figure 3**). Currently, seven matchmakers worldwide are connected to one another by the Matchmaker Exchange (**https://www.matchmakerexchange.org**), a federated platform that facilitates matchmaking through a standardized application programming interface (20) and connects data from more than 70,000 patients (for a review, see Reference 88). While many of these databases can accept detailed phenotypic and genotypic information that

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

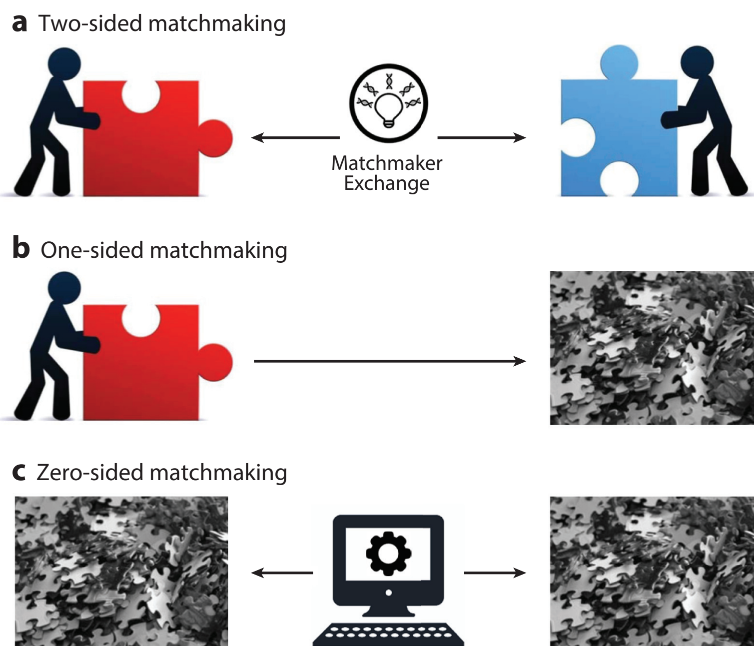

Figure 3

Different types of patient matchmaking for rare diseases. (*a*) Two-sided matchmaking, in which undiagnosed patients with the same candidate gene are identified by matchmaking platforms such as the Matchmaker Exchange. (*b*) One-sided matchmaking, in which the candidate gene identified in an undiagnosed patient is used to query a database housing genome-wide sequencing data from other undiagnosed patients. (*c*) Zero-sided matchmaking, in which computer algorithms are used to identify potentially matching undiagnosed patients with rare variants in the same gene and overlapping phenotypes.

requires explicit research consent, matchmaking via the Matchmaker Exchange is most often done using only gene names and, occasionally, high-level HPO terms and inheritance patterns. The sharing of this type of data via the Matchmaker Exchange is viewed as having minimal risk for reidentification of patients and is widely considered to be part of providing optimal care for patients (31).

3.5. Empowering Patients to Share Their Data

While clinically driven data sharing is a major goal, the reality is that such data sharing is time intensive and, unfortunately, not reimbursed. In addition, there are only approximately two geneticists per million people in the United States (71), highlighting the challenges around the availability of appropriate expertise for patients with a suspected RGD. Therefore, to reach the theoretical maximal diagnostic yield of GWS, every patient with a suspected RGD should have the opportunity to share his or her data. In fact, families with undiagnosed RGDs are often eager to share their data, as this sharing has the potential to provide direct benefit in terms of finding answers and improving their care as well as the care of other families with the same condition.

When patients and families directly share their clinical and genomic data online, they bypass associated privacy restrictions on health-related data. Current data-sharing initiatives such as the Matchmaker Exchange do not support patient-directed sharing, and families have started using social media with some remarkable successes. However, most of these patient-led efforts fail largely

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

because of the lack of suitable infrastructure and guidance on how to prioritize and share standardized information. Newer initiatives, such as MyGene2 (**https://mygene2.org**), provide a solution for many of these challenges, enabling families to share health and genetic data directly with other families, clinicians, and researchers, in the spirit of the citizen scientist (24). These types of approaches are critical for patients and families in cases where their data would otherwise not be shared by their clinician due to privacy challenges and/or time constraints.

3.6. Increasing Interdisciplinary Collaboration

There have been only a few truly integrated, interdisciplinary, translational programs for undiagnosed RGDs, but one success has been the Undiagnosed Diseases Network (91). Established in 2014, the Undiagnosed Diseases Network is a network of seven clinical sites, two DNA sequencing cores, a coordinating center, and, most recently, a central biorepository, metabolomics core, and model organisms screening center. Potential study participants apply via a web portal. Applications require a supporting summary from a health-care provider. Most or all expenses associated with study clinical work are covered by a combination of insurance reimbursement, philanthropy, and research funds. Building upon the Undiagnosed Diseases Network, the International Network for Undiagnosed Diseases brings together international and expert clinical and research centers and now includes 16 countries that agree to a set of principles, governance, and best practices (113). The Initiative on Rare and Undiagnosed Diseases in Japan is another example of a clinical RGD network (2). In England, the 100,000 Genomes Project is a dynamic national initiative established in 2013 that aims to sequence 100,000 genomes from patients with RDs and cancer (114). Such initiatives highlight the importance of interdisciplinary expertise and translational infrastructure to advance diagnostic opportunities for patients.

4. RESEARCH APPROACHES TO UNDIAGNOSED RARE GENETIC DISEASES

For patients who remain undiagnosed following a clinical assessment that includes reanalysis of existing data and appropriate data sharing within their clinical context, research brings the next opportunities. The first of these is deeper sharing of existing GWS data (primarily exome data) for discovery of new disease–gene relationships—an important approach, given the high likelihood that new pathogenic coding variants and novel disease genes are present in these data sets (7). Nonetheless, there certainly exists an undefined subset of patients whose disease mechanism remains intractable to clinically available diagnostic approaches. New approaches and technologies will be essential to provide answers for RD patients without a compelling candidate VUS or GUS from previous clinical testing. Here, we highlight emerging research approaches that are showing great promise for these patients.

4.1. Deep-Phenotypic and Genotypic Data Sharing for Discovery Purposes

The discovery potential within an undiagnosed RGD patient's ES data is significant as we continue to optimize our knowledge base and analytical processes. Disease-causing variants can go unnoticed in these data sets for several reasons, including being obscured within a large number of rare variants (due to a lack of appropriate control-population cohorts or lack of DNA-sequence data from a sufficient number of affected or unaffected family members) or within a novel gene or unannotated exon or gene. Broad sharing of deep-phenotypic (using HPO terms) and exome-sequence data to make those data discoverable by the greater RGD community provides an

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

excellent opportunity for diagnosis. Because these data are potentially more sensitive and carry a risk (albeit small) of identifiability, this level of data sharing requires consent for research (31).

In contrast to the two-sided matchmaking highlighted in Section 3.4, in which both clinicians have a compelling candidate GUS, sharing deep-phenotypic and genotypic data can enable one-sided matchmaking (in which only the querier has a GUS and is looking to find a matching patient) (**Figure 3**). One example of a database that enables such one-sided matching is PhenomeCentral (**https://www.phenomecentral.org**), a restricted-access web portal used by the Care4Rare Canada Consortium, Undiagnosed Diseases Network, and International Network for Undiagnosed Diseases (19). The Genome-Phenome Analysis Platform (**https://platform.rd-connect.eu**) works in a similar manner to make matches for the European RD-Connect project. Another example is Genotype to Mendelian Phenotype (Geno$_2$MP; **https://geno2mp.gs.washington.edu/Geno2MP**), a publicly accessible web-based query tool that searches a database of rare variants generated from ES data linked to high-level phenotypic information from a variety of Mendelian gene discovery projects, including the Centers for Mendelian Genomics in the United States. Unfortunately, all of these tools are siloed by their lack of interoperability due to both technical and governance limitations. As part of the Breaking Barriers to Health Data project, led by the World Economic Forum, four large genomics programs—Genomics England, Genomics4RD (part of the Care4Rare Canada Consortium), the Australian Genomics Health Alliance, and Intermountain Healthcare—are working toward building a governance framework that would allow for cross-border one-sided queries of each other's data sets (117).

Ultimately, the community should work toward zero-sided matchmaking, where computer algorithms are developed and optimized to identify cases with overlapping phenotypes and rare variants in the same gene (**Figure 3**). Such matching would be possible in databases such as PhenomeCentral and the Genome-Phenome Analysis Platform based on semantic similarity between clinical features and automatically prioritized genes from ES data, although no bona fide zero-sided matches have yet been confirmed. Given the daily increase in data sets related to patients with an undiagnosed RGD following clinical ES, this approach will be critically important to ensure that we can analyze such unsolved data sets at scale.

4.2. Emerging Technologies

For a subset of patients, the disease mechanism is not readily captured (if it is captured at all) by clinical testing. The literature contains multiple examples of new technologies that show great promise for RD diagnostics. Notably, all of them will require significant efforts to define their diagnostic utility and optimize their analytical validity before they are available clinically. While it is outside the scope of this review to provide significant details for each of these exciting new technologies, we present a brief overview here in the context of their promise for patients with undiagnosed RGDs.

4.2.1. Deep DNA sequencing. Mosaicism, defined as the presence of genetically different cells in an individual, has long been recognized to play a significant role in a subset of RGDs (12, 74); for example, 30% of patients with Cornelia de Lange syndrome (49, 62) and more than 85% of patients with *PIK3CA*-associated segmental overgrowth (76) carry mosaic variants as the cause of their disease. Mosaicism is technically and analytically challenging to detect. Mosaic variants will be identified by Sanger DNA sequencing only if they have a relatively high allelic fraction in the analyzed tissue (e.g., greater than approximately 10%). With NGS, sensitivity scales with depth, but in these cases it can be difficult to distinguish between technical artifacts and low-fraction mosaic variants, both of which are often filtered out by analysis pipelines due to low confidence.

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

The sequencing depth needed to reliably detect mosaic variants has not yet been established. Detection is further complicated by tissue-specific mosaicism, and thus the DNA source used for sequencing may not have a sufficient allelic fraction for identification.

Undetected mosaicism is likely a disease mechanism for a subset of patients with RGD. A retrospective analysis determined that mosaic variants were reported in approximately 2% of molecular diagnoses from approximately 12,000 clinically performed ES tests; 1.5% were mosaic in the proband, and an additional 0.3% exhibited parental mosaicism (22). Similarly, the Deciphering Developmental Disorders group looked at trio exome data from 4,293 probands with severe developmental disability and seemingly unaffected parents and identified mosaic de novo mutations for approximately 1% of patients (119). Both of these studies were retrospective, second-tier analyses to assess the contribution of mosaicism to RGD. Targeted mosaicism studies, as a first-tier approach, will need to utilize deep DNA sequencing, through either a targeted panel or ES, and consider using multiple tissue types to truly define the contribution of this genetic mechanism to undiagnosed RGDs.

4.2.2. Whole-genome sequencing. ES assesses the protein-coding portion of the genome (approximately 1% of the sequence); however, the noncoding portion contains regions vital for genome regulation and thus human health, the majority of which remain to be defined. Compared with ES, short-read WGS does not suffer from the same capture limitations and therefore more reliably covers the exome, sequences noncoding regions, and provides the ability to identify small and large CNVs and chromosomal abnormalities. With the ever-decreasing cost of NGS, short-read WGS has now become more accessible, but challenges persist with data analysis (e.g., a lack of extensive control cohorts to use as references for sequencing data; algorithms for reliably detecting structural variation, including CNVs; and in silico tools to predict the impact of noncoding variation) and data storage (e.g., cost). In addition, the lower coverage (30 reads per nucleotide, compared with more than 100 reads per nucleotide with exomes) can result in lower sensitivity for difficult-to-assess regions and mosaic variants. The reality, thus far, is that the added diagnostic utility of WGS has been underwhelming; the majority of diagnoses using WGS would have been made with better-quality ES. However, there have been some important successes, such as the identification of pathogenic variants in deep splice sites [e.g., *IGHMBP2* (23)] and regulatory regions [e.g., congenital intractable diarrhea (86)], which would have been inaccessible by ES.

To address some of the shortcomings of short-read WGS, which can sequence up to 150 base pairs per read, there are several long-read WGS technologies that can sequence more than 10,000 base pairs per read and show promise for haplotype calling; detection of genomic repeats, structural rearrangements, and CNVs; sequencing of loci with known pseudogenes; and interrogation of complex regions of the genome. Challenges for long-read WGS include a lower accuracy per DNA base compared with short-read WGS, that analytical tools and reference control cohorts are still being developed, and the prohibitive cost. The emerging approaches to combine data from short- and long-read WGS will address some of these issues. Long-read WGS has already been used successfully to identify structural variants (75, 77) and novel repeat expansions that cause neurological diseases (38, 52) and are showing significant promise for the diagnosis of RGD patients.

4.2.3. RNA sequencing. While DNA-based NGS applications are common for genetic diagnostics, RNA-based applications have yet to be leveraged in a clinical setting. RNA sequencing requires the conversion of RNA to cDNA, followed by NGS (commonly using the same Illumina short-read technology as for ES and WGS) and analysis to interrogate mRNA transcript levels, mRNA splicing, and noncoding RNA sequences. Genomic protein-coding variants can also be identified using RNA-sequencing data (89). Most genes generate multiple transcripts via complex

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

regulation of alternative mRNA splicing. Mutations that affect mRNA splicing are estimated to account for 15–60% of genetic diseases (61, 69), and regulatory mutations that affect gene expression are likely equally prevalent, yet these variants are difficult to identify from DNA sequence due to algorithm limitations. As many genes have tissue-specific regulation and expression, key challenges include the accessibility of relevant tissues for mRNA extraction, the large amount of sequencing required to detect changes among the thousands of other mRNA transcripts, and the necessity of tissue-specific large control cohorts to control for the natural variation in expression and mRNA composition observed in healthy individuals.

RNA sequencing holds promise as part of a multiomic approach with WGS to identify causative DNA variants in the vicinity of abnormal transcripts. Several studies have successfully used RNA sequencing for RGD diagnosis, predominantly in patients with well-characterized conditions in the absence of one or more pathogenic variants [e.g., neuromuscular (29, 43) and mitochondrial disorders (64)] or recessive conditions for which one mutation has been identified but the second is elusive (58), and data on its diagnostic utility are emerging (40, 66). As RNA sequencing, analysis algorithms, and control cohorts improve, the optimal positioning of this technology in the diagnostic care pathway for RGDs will become clear.

4.2.4. Methylation profiling. Aberrant methylation has been associated with several RGDs, including imprinting disorders such as Angelman syndrome, trinucleotide repeat expansion disorders such as fragile X syndrome, and a growing number of RGDs associated with germline mutations in epigenetic machinery (for a review, see Reference 14). Several methods are available for assessing genome-wide DNA methylation. These methods begin with bisulfite DNA conversion, in which unmethylated cytosine bases are deaminated to uracil. The treated DNA sequence is then assessed by either microarray (targeting specific methylation sites) or sequencing (known as bisulfite sequencing, which can be assessed genome-wide using NGS) and compared with large numbers of age- and tissue-matched control samples.

The greatest immediate potential of these methods is to diagnose patients for whom there is clinical suspicion of an imprinting disorder (4) or a germline mutation in an epigenetic regulator for which there is a defined epi-signature (3). These epi-signatures are an emerging diagnostic tool and have been defined for some classic RGDs, including CHARGE (coloboma, heart defects, atresia choanae, growth retardation, genital abnormalities, and ear abnormalities) and Kabuki syndromes (21), alpha-thalassemia mental retardation (ATRX) syndrome (99), and Sotos syndrome (25). For undiagnosed patients with a VUS in a known methylation gene, the identification (or not) of a specific methylation signature can provide functional evidence to reclassify the variant as pathogenic or benign.

For patients without a candidate DNA variant, the identification of a specific methylation profile can suggest a genetic mechanism, due to either a mutation that has been intractable to current testing approaches or a mutation affecting an epigenetic regulator that has yet to be associated with disease. In a recent study, rare epigenetic variations (also termed epivariations) were identified in 20% of patients with neurodevelopmental delay with or without congenital anomalies (8); for 24% of these patients, the authors were able to identify corresponding rare DNA-sequence variants in the differentially methylated regions. Studies of larger cohorts of patients with known as well as undiagnosed RGDs, along with appropriate controls, will demonstrate how best to use this technology as part of a multiomic approach and will ultimately identify the subset of the undiagnosed population with postzygotic epigenetic alterations as the cause of their RGD.

4.2.5. Metabolomics. Metabolomics is the study of all chemical processes within an organism, including small-molecule substrates, intermediates, and products of metabolism (18); like

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

RNA sequencing and methylation profiling, it can provide insight into the consequences of genomic variants. Metabolomics involves the separation of molecules by techniques such as gas or liquid chromatography, followed by nuclear magnetic resonance and mass spectrometry; subsequent analysis can be either targeted [to assess only particular compounds, e.g., rapid and sensitive newborn screening (83)] or untargeted. Notably, the analysis of untargeted metabolomics is incredibly complicated, requiring large cohorts to control for natural human variation, and the corresponding analytical methods are under constant development (44, 53, 102). As a result, untargeted metabolomics is only now emerging as a research tool (102), but it holds significant promise for patients with suspected inborn errors of metabolism by highlighting altered biochemical functions to facilitate prioritization of GWS variants (10, 44). Adding further utility is the linkage of this approach to biomarker identification, to ultimately facilitate diagnosis and future therapeutic studies (28). Overall, it is clear that metabolomics will hold significant diagnostic utility for patients with one of the more than 1,000 inborn errors of metabolism (37), but significant work remains to reach the full potential of this technology.

4.3. Supporting New Computational Approaches

New computational approaches to better analyze, combine, and interpret data generated from patients with an undiagnosed RGD will be essential for the resolution of all RDs. This is particularly important as we work toward understanding the noncoding portion of the genome as well as non-Mendelian genetic diseases. Consider enhancers as an example of a genomic regulatory mechanism: There are more than 100,000 enhancers in the human genome (for 20,000 genes), but their exact locations are poorly mapped, and their in vivo functions are poorly defined. We need to greatly increase our understanding of these sequences to truly begin utilizing the full breadth of data generated from WGS.

This same level of uncertainty and need for advancement exists for the other genomic regulatory mechanisms as well. Elucidation of diseases with non-Mendelian inheritance, such as oligogenic inheritance and complex (gene–environment) diseases, will require significant improvements in understanding of the noncoding genome as well as computational and functional approaches to additive risk. Development of computational approaches, including machine learning, will be limited by the availability of appropriate data. Accordingly, efforts should be made by health-care providers, clinical laboratories, and other RGD stakeholders to enable mechanisms for appropriate data sharing: All RD patients, whether diagnosed or not, should have the opportunity to contribute to future knowledge in this fashion.

4.4. Matching Patients with an Undiagnosed Rare Genetic Disease to Appropriate Approaches

Several approaches can be considered for a patient with an undiagnosed RGD. Although a diagnostic approach has not yet been established for these patients, possible strategies include developing approaches for the groups of undiagnosed patients defined by the IRDiRC Solving the Unsolved Task Force (16) (**Figure 2**). As shown in **Figure 2**, group 1 includes patients who remain undiagnosed despite having clinical features of a highly recognizable disease for which there is a very sensitive genetic test (e.g., cystic fibrosis or tuberous sclerosis). The primary hypothesis for this group is that they remain undiagnosed because their mutation is undetectable using current technology. For these patients, the best chance for a molecular diagnosis would come from sharing of biological samples (different tissues) and data with international efforts focused on using additional omic approaches (e.g., WGS or RNA sequencing) to define the full mutational spectrum of

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

the RGD. The collection of such cohorts represents a great opportunity for the RGD community to demonstrate the power and utility of these emerging technologies.

Group 2 includes patients who have a recognizable RGD where the molecular etiology has not yet been identified. These diseases have been well described in the genetics literature, and there have been tens to hundreds of cases reported, but their genetic causes remain elusive despite being the focus of many research programs [e.g., Dubowitz (51), Hallermann–Streiff (100), and PHACE (posterior fossa brain malformations, hemangioma, arterial lesions, cardiac abnormalities, and eye abnormalities) (108) syndromes]. Challenges in mechanistic delineation for these diseases may be due to extreme genetic and phenotypic heterogeneity, mosaicism, genomic dysregulation, epigenetics, gene–environment interactions, and other non-Mendelian contributions. The way forward for this group of disorders will require global cooperation for sharing of biological samples (different tissues), sharing and reanalysis of existing data, documentation of environmental exposures, and the comprehensive use of emerging and new technologies along with the development of new computational approaches to analyze multiomic data sets.

Group 3 includes patients with a clinical presentation that is associated with significant genetic heterogeneity. For example, 50–80% of nonsyndromic retinitis pigmentosa is associated with mutations in more than 50 genes (35). For diseases like this, undiagnosed patients likely have mutations in known genes that have not yet been identified by the testing methodology or in novel genes that have not yet been linked to the specific disease. Similar to group 1, these patients would be best served by the development of international cohorts of mutation-negative patients and the sharing of biological samples (different tissues) and data with international efforts focused on using additional omic approaches (e.g., WGS or RNA sequencing) to define the full mutational spectrum and identify all of the disease–gene relationships associated with the clinical presentation (e.g., retinitis pigmentosa).

Finally, group 4 includes patients who do not fit a clinically recognizable syndrome in the literature. These patients often have a constellation of features that overlap with multiple conditions and are sometimes described as having syndromes without a name (SWAN). Their RGD may be due to phenotypic expansions of known disease genes or a novel disease–gene association. Given that the number of novel genotype-driven disease–gene associations is now outpacing the number of phenotype-driven discoveries (7), this group may represent the largest number of undiagnosed RD patients. Patients in group 4 who have significant overlap with a particular disease may be diagnosed by improved assessment of the relevant gene(s). This is highlighted by a study in which three patients who remained undiagnosed following ES received diagnoses following Sanger DNA sequencing of specific genes; in one case, the variant-calling software had missed the pathogenic variant; in another, the variant was in a noncoding exon; and in the third, a bioinformatics error resulted in a frameshift variant failing to align (87). Overall, group 4, more than any other, will benefit from the broadest type of data sharing and novel computational methods or strategies, which will enable matchmaking among patients with overlapping phenotypes and ultrarare variants in the same gene. In addition, such data will facilitate the investigation of novel mechanisms of RGD (including the contribution of noncoding mutations) and oligo- or polygenic inheritance.

If clinicians or patients are not certain how to access appropriate research studies, they might consider starting with national and international research consortia that have been created to support RGD research and have studied thousands of patients using many of the approaches described above. In Canada, the Finding of Rare Disease Genes (FORGE)/Care4Rare Gene Discovery program (**http://care4rare.ca/discovery**) has connected 21 Canadian academic centers since 2011, enabling collaboration among clinicians, researchers, scientists, and bioinformaticians to work on developing approaches and tools to better evaluate patients with undiagnosed RGDs. In Europe, the Solving the Unsolved Rare Diseases (Solve-RD) project (**http://solve-rd.eu**) connects 22

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

academic centers and is fully integrated with the European Reference Networks for rare diseases. In the United States, the Centers for Mendelian Genomics (**http://mendelian.org**) are working with more than 2,000 investigators in 82 countries to identify novel disease-causing genes. These and other clinical-research consortia are driven by common goals and are united by large supportive networks such as the IRDiRC, which was established in 2011 and now brings together almost 60 national and international funding agencies, companies, and patient advocacy groups as well as hundreds of researchers and other organizations invested in RD research worldwide (5).

5. CONCLUSION

These are promising times for the RGD community; never before has the prospect of identifying a molecular diagnosis for all RGD patients been so attainable. To overcome this grand challenge, however, will require global cooperation at an unprecedented scale. We need to ensure that all undiagnosed RGD patients are identified and seen by clinicians with expertise in RGD diagnosis, have access to appropriate genetic testing, and are offered opportunities to share their phenotypic data, genotypic data, and biological samples with researchers; only then will we be able to understand the cause of each and every RGD. The IRDiRC's goal over the next 10 years—for each RD patient to receive a diagnosis within one year of coming to medical attention—will be achievable if we put the well-being of the patients and their families at the center of everything that we do.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank the IRDiRC Scientific Secretariat and members of the Solving the Unsolved Task Force, funded under the European FP7 contract SUPPORT-IRDIRC (305207), for helpful discussions. We gratefully acknowledge the Care4Rare Canada Consortium [funded by Genome Canada and the Canadian Institutes of Health Research (CIHR) under grant OGI-0147] for providing the opportunity to collectively study undiagnosed RGDs. T.H. was supported by a Frederick Banting and Charles Best Canada Graduate Scholarship Doctoral Award from CIHR. K.M.B. was supported by a CIHR Foundation Grant (FDN-154279) and a Tier 1 Canada Research Chair in Rare Disease Precision Health.

LITERATURE CITED

1. ACMG Board Dir. 2017. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. *Genet. Med.* 19:721–22
2. Adachi T, Kawamura K, Furusawa Y, Nishizaki Y, Imanishi N, et al. 2017. Japan's initiative on rare and undiagnosed diseases (IRUD): towards an end to the diagnostic odyssey. *Eur. J. Hum. Genet.* 25:1025–28
3. Aref-Eshghi E, Bend EG, Colaiacovo S, Caudle M, Chakrabarti R, et al. 2019. Diagnostic utility of genome-wide DNA methylation testing in genetically unsolved individuals with suspected hereditary conditions. *Am. J. Hum. Genet.* 104:685–700
4. Aref-Eshghi E, Schenkel LC, Lin H, Skinner C, Ainsworth P, et al. 2017. Clinical validation of a genome-wide DNA methylation assay for molecular diagnosis of imprinting disorders. *J. Mol. Diagn.* 19:848–56

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

5. Austin CP, Cutillo CM, Lau LPL, Jonker AH, Rath A, et al. 2017. Future of rare diseases research 2017–2027: an IRDiRC perspective. *Clin. Transl. Sci.* 11:21–27
6. Balci TB, Hartley T, Xi Y, Dyment DA, Beaulieu CL, et al. 2017. Debunking Occam's razor: diagnosing multiple genetic diseases in families by whole-exome sequencing. *Clin. Genet.* 92:281–89
7. Bamshad MJ, Nickerson DA, Chong JX. 2019. Mendelian gene discovery: fast and furious with no end in sight. *Am. J. Hum. Genet.* 105:448–55
8. Barbosa M, Joshi RS, Garg P, Martin-Trujillo A, Patel N, et al. 2018. Identification of rare de novo epigenetic variations in congenital disorders. *Nat. Commun.* 9:2064
9. Bauman JG, Wiegant J, Borst P, van Duijn P. 1980. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. *Exp. Cell Res.* 128:485–90
10. Beger RD, Dunn W, Schmidt MA, Gross SS, Kirwan JA, et al. 2016. Metabolomics enables precision medicine: "a white paper, community perspective." *Metabolomics* 12:149
11. Beynon JH. 1954. Qualitative analysis of organic compounds by mass spectrometry. *Nature* 174:735–37
12. Biesecker LG, Spinner NB. 2013. A genomic view of mosaicism and human disease. *Nat. Rev. Genet.* 14:307–20
13. Birney E, Vamathevan J, Goodhand P. 2017. Genomics in healthcare: GA4GH looks to 2022. bioRxiv 203554. **https://doi.org/10.1101/203554**
14. Bjornsson HT. 2015. The Mendelian disorders of the epigenetic machinery. *Genome Res*. 25:1473–81
15. Boycott KM, Hartley T, Adam S, Bernier F, Chong K, et al. 2015. The clinical application of genome-wide sequencing for monogenic diseases in Canada: position statement of the Canadian College of Medical Geneticists. *J Med. Genet.* 52:431–37
16. Boycott KM, Hartley T, Biesecker LG, Gibbs RA, Innes AM, et al. 2019. A diagnosis for all rare genetic diseases: the horizon and the next frontiers. *Cell* 177:32–37
17. Bragin E, Chatzimichali EA, Wright CF, Hurles ME, Firth HV, et al. 2014. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res*. 42:D993–1000
18. Bujak R, Struck-Lewicka W, Markuszewski MJ, Kaliszan R. 2015. Metabolomics for laboratory diagnostics. *J. Pharm. Biomed. Anal.* 113:108–20
19. Buske OJ, Girdea M, Dumitriu S, Gallinger B, Hartley T, et al. 2015. PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum. Mutat.* 36:931–40
20. Buske OJ, Schiettecatte F, Hutton B, Dumitriu S, Misyura A, et al. 2015. The Matchmaker Exchange API: automating patient matching through the exchange of structured phenotypic and genotypic profiles. *Hum. Mutat.* 36:922–27
21. Butcher DT, Cytrynbaum C, Turinsky AL, Siu MT, Inbar-Feigenberg M, et al. 2017. CHARGE and Kabuki syndromes: gene-specific DNA methylation signatures identify epigenetic mechanisms linking these clinically overlapping conditions. *Am. J. Hum. Genet.* 100:773–88
22. Cao Y, Tokita MJ, Chen ES, Ghosh R, Chen T, et al. 2019. A clinical survey of mosaic single nucleotide variants in disease-causing genes detected by exome sequencing. *Genome Med*. 11:48
23. Cassini TA, Duncan L, Rives LC, Newman JH, Phillips JA, et al. 2019. Whole genome sequencing reveals novel *IGHMBP2* variant leading to unique cryptic splice-site and Charcot-Marie-Tooth phenotype with early onset symptoms. *Mol. Genet. Genom. Med.* 7:e00676
24. Chong JX, Yu JH, Lorentzen P, Park KM, Jamal SM, et al. 2016. Gene discovery for Mendelian conditions via social networking: De novo variants in *KDM1A* cause developmental delay and distinctive facial features. *Genet. Med.* 18:788–95
25. Choufani S, Cytrynbaum C, Chung BHY, Turinsky AL, Grafodatskaya D, et al. 2015. *NSD1* mutations generate a genome-wide DNA methylation signature. *Nat. Commun.* 6:10207
26. Clark MM, Stark Z, Farnaes L, Tan TY, White SM, et al. 2018. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *npj Genom. Med.* 3:16

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

27. Costa T, Scriver CR, Childs B, Opitz JM, Reynolds JF. 1985. The effect of Mendelian disease on human health: a measurement. *Am. J. Med. Genet.* 21:231–42
28. Crowther LM, Poms M, Plecko B. 2018. Multiomics tools for the diagnosis and treatment of rare neurological disease. *J. Inherit. Metab. Dis.* 41:425–34
29. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, et al. 2017. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* 9:eaal5209
30. de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, et al. 2012. Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* 367:1921–29
31. Dyke SOM, Knoppers BM, Hamosh A, Firth HV, Hurles M, et al. 2017. "Matching" consent to purpose: the example of the Matchmaker Exchange. *Hum. Mutat.* 38:1281–85
32. Eisenberger T, Di Donato N, Baig SM, Neuhaus C, Beyer A, et al. 2014. Targeted and genomewide NGS data disqualify mutations in *MYO1A*, the "*DFNA48* gene", as a cause of deafness. *Hum. Mutat.* 35:565–70
33. Eldomery MK, Coban-Akdemir Z, Harel T, Rosenfeld JA, Gambin T, et al. 2017. Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome Med*. 9:26
34. Esquivel-Sada D, Nguyen MT. 2017. Diagnosis of rare diseases under focus: impacts for Canadian patients. *J. Community Genet.* 9:37–50
35. Fahim AT, Daiger SP, Weleber RG. 1993. Nonsyndromic retinitis pigmentosa overview. In *GeneReviews*, ed. MP Adam, HH Ardinger, RA Pagon, SE Wallace, LJH Bean, et al. Seattle: Univ. Wash. **https://www.ncbi.nlm.nih.gov/books/NBK1417**
36. Ferreira CR. 2019. The burden of rare diseases. *Am. J. Med. Genet. A* 179:885–92
37. Ferreira CR, van Karnebeek CDM, Vockley J, Blau N. 2019. A proposed nosology of inborn errors of metabolism. *Genet. Med.* 21:102–6
38. Florian RT, Kraft F, Leitão E, Kaya S, Klebe S, et al. 2019. Unstable TTTTA/TTTCA expansions in *MARCH6* are associated with familial adult myoclonic epilepsy type 3. *Nat. Commun.* 10:4919
39. Ford C, Hamerton J. 1956. The chromosomes of man. *Nature* 178:1020–23
40. Frésard L, Smail C, Ferraro NM, Teran NA, Li X, et al. 2019. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat. Med.* 25:911–19
41. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, et al. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *PNAS* 89:1827–31
42. Glob. Comm. End Diagn. Odyssey Child. Rare Dis. 2019. *Global Commission year one report*. Rep., Glob. Comm. End Diagn. Odyssey Child. Rare Dis. **https://globalrarediseasecommission.com/Report**
43. Gonorazky H, Liang M, Cummings B, Lek M, Micallef J, et al. 2016. RNAseq analysis for the diagnosis of muscular dystrophy. *Ann. Clin. Transl. Neurol.* 3:55–60
44. Graham E, Lee J, Price M, Tarailo-Graovac M, Matthews A, et al. 2018. Integration of genomics and metabolomics for prioritization of rare disease variants: a 2018 literature review. *J. Inherit. Metab. Dis.* 41:435–45
45. Guthrie R, Susi A. 1963. A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics* 32:338–43
46. Harrison SM, Riggs ER, Maglott DR, Lee JM, Azzariti DR, et al. 2016. Using ClinVar as a resource to support variant interpretation. *Curr. Protoc. Hum. Genet.* 89:8.16.1–23
47. Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, et al. 2010. De novo mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nat. Genet.* 42:483–85
48. Holt IJ, Harding AE, Morgan-Hughes JA. 1988. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* 331:717–19
49. Huisman SA, Redeker EJW, Maas SM, Mannens MM, Hennekam RCM. 2013. High rate of mosaicism in individuals with Cornelia de Lange syndrome. *J. Med. Genet.* 50:339–44
50. Ingram VM. 1959. Abnormal human haemoglobins. III. The chemical difference between normal and sickle cell haemoglobins. *Biochim. Biophys. Acta* 36:402–11
51. Innes AM, McInnes BL, Dyment DA. 2018. Clinical and genetic heterogeneity in Dubowitz syndrome: implications for diagnosis, management and further research. *Am. J. Med. Genet. C* 178:387–97

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

52. Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, et al. 2019. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat. Genet.* 51:1222–32
53. Ismail IT, Showalter MR, Fiehn O. 2019. Inborn errors of metabolism in the era of untargeted metabolomics and lipidomics. *Metabolites* 9:242
54. Jacobs PA, Court Brown WM, Baikie AG, Strong JA. 1959. The somatic chromosomes in mongolism. *Lancet* 273:710
55. Jeffreys AJ, Wilson V, Thein SL. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67–73
56. Jimenez-Sanchez G, Childs B, Valle D. 2014. The effect of Mendelian disease on human health. In *The Online Metabolic and Molecular Bases of Inherited Disease*, ed. AL Beaudet, B Vogelstein, KW Kinzler, SE Antonarakis, A Ballabio, et al. New York: McGraw-Hill. **https://doi.org/10.1036/ommbid.303**
57. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, et al. 2019. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv 531210. **https://doi.org/10.1101/531210**
58. Kernohan KD, Frésard L, Zappala Z, Hartley T, Smith KS, et al. 2017. Whole-transcriptome sequencing in blood provides a diagnosis of spinal muscular atrophy with progressive myoclonic epilepsy. *Hum. Mutat.* 38:611–14
59. Knoll JHM, Nicholls RD, Magenis RE, Graham JM, Lalande M, et al. 1989. Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. *Am. J. Med. Genet.* 32:285–90
60. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, et al. 2019. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res*. 47:D1018–27
61. Krawczak M, Reiss J, Cooper D. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* 90:41–54
62. Krawczynska N, Wierzba J, Wasag B. 2019. Genetic mosaicism in a group of patients with Cornelia de Lange syndrome. *Front. Pediatr.* 7:203
63. Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, et al. 1991. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* 252:1711–14
64. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, et al. 2017. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* 8:15824
65. Ledbetter DH, Riccardi VM, Airhart SD, Strobel RJ, Keenan BS, Crawford JD. 1981. Deletions of chromosome 15 as a cause of the Prader-Willi syndrome. *N. Engl. J. Med.* 304:325–29
66. Lee H, Huang AY, Wang L, Yoon AJ, Renteria G, et al. 2020. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet. Med.* 22:490–99
67. Lejeune J, Gautier M, Turpin R. 1959. [Study of somatic chromosomes from 9 mongoloid children]. *C. R. Hebd. Seances Acad. Sci.* 248:1721–22 (in French)
68. Lochmüller H, Le Cam Y, Jonker AH, Lau LP, Baynam G, et al. 2017. "IRDiRC Recognized Resources": a new mechanism to support scientists to conduct efficient, high-quality research for rare diseases. *Eur. J. Hum. Genet.* 25:162–65
69. López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. 2005. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett*. 579:1900–3
70. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DCY, et al. 2010. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* 362:1181–91
71. Maiese DR, Keehn A, Lyon M, Flannery D, Watson M. 2019. Current conditions in medical genetics practice. *Genet. Med.* 21:1874–77
72. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–80
73. Marshall DA, Benchimol EI, MacKenzie A, Duque DR, MacDonald KV, et al. 2018. Direct health-care costs for children diagnosed with genetic diseases are significantly higher than for children with other chronic diseases. *Genet. Med.* 21:1049–57

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

74. McNulty SN, Evenson MJ, Corliss MM, Love-Gregory LD, Schroeder MC, et al. 2019. Diagnostic utility of next-generation sequencing for disorders of somatic mosaicism: a five-year cumulative cohort. *Am. J. Hum. Genet.* 105:734–46
75. Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, et al. 2018. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet. Med.* 20:159–63
76. Mirzaa G, Conway R, Graham JM, Dobyns WB. 2013. *PIK3CA*-related segmental overgrowth. In *GeneReviews*, ed. MP Adam, HH Ardinger, RA Pagon, SE Wallace, LJH Bean, et al. Seattle: Univ. Wash. **https://www.ncbi.nlm.nih.gov/books/NBK153722**
77. Mizuguchi T, Suzuki T, Abe C, Umemura A, Tokunaga K, et al. 2019. A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing. *J. Hum. Genet.* 64:359–68
78. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5:621–28
79. Mullis KB, Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155:335–50
80. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, et al. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344–49
81. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, et al. 2010. Exome sequencing identifies the cause of a Mendelian disorder. *Nat. Genet.* 42:30–35
82. NGO Comm. Rare Dis. 2019. *Report from Rare Disease Day policy event at the United Nations. Second high level event of the NGO Committee for Rare Diseases.* Rep., NGO Comm. Rare Dis. **http://download2.eurordis.org.s3.amazonaws.com/ngocommittee/ngocommittee_report2019.pdf**
83. Ombrone D, Giocaliere E, Forni G, Malvagia S, la Marca G. 2016. Expanded newborn screening by mass spectrometry: new tests, future perspectives. *Mass Spectrom. Rev.* 35:71–84
84. Orphanet. 2019. *Prevalence and incidence of rare diseases: bibliographic data. Diseases listed by decreasing prevalence, incidence or number of published cases.* Orphanet Rep. Ser. 2, Orphanet, Paris. **http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_decreasing_prevalence_or_cases.pdf**
85. Ouellette AC, Mathew J, Manickaraj AK, Manase G, Zahavich L, et al. 2018. Clinical genetic testing in pediatric cardiomyopathy: Is bigger better? *Clin. Genet.* 93:33–40
86. Oz-Levi D, Olender T, Bar-Joseph I, Zhu Y, Marek-Yagel D, et al. 2019. Noncoding deletions reveal a gene that is critical for intestinal function. *Nature* 571:107–11
87. Pena LDM, Jiang YH, Schoch K, Spillmann RC, Walley N, et al. 2018. Looking beyond the exome: a phenotype-first approach to molecular diagnostic resolution in rare and undiagnosed diseases. *Genet. Med.* 20:464–69
88. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, et al. 2015. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum. Mutat.* 36:915–21
89. Piskol R, Ramaswami G, Li JB. 2013. Reliable identification of genomic variants from RNA-seq data. *Am. J. Hum. Genet.* 93:641–51
90. Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, et al. 2017. Resolution of disease phenotypes resulting from multilocus genomic variation. *N. Engl. J. Med.* 376:21–31
91. Ramoni RB, Mulvihill JJ, Adams DR, Allard P, Ashley EA, et al. 2017. The Undiagnosed Diseases Network: accelerating discovery about health and disease. *Am. J. Hum. Genet.* 100:185–92
92. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, et al. 2015. ClinGen—the clinical genome resource. *N. Engl. J. Med.* 372:2235–42
93. Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, et al. 2016. Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* 18:696–704
94. Richards S, Aziz N, Bale S, Bick D, Das S, et al. 2015. 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.* 17:405–23
95. Riordan JR, Rommens J, Kerem B, Alon N, Rozmahel R, et al. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245:1066–73

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

96. Salmon LB, Orenstein N, Markus-Bustani K, Ruhrman-Shahar N, Kilim Y, et al. 2018. Improved diagnostics by exome sequencing following raw data reevaluation by clinical geneticists involved in the medical care of the individuals tested. *Genet. Med.* 21:1443–51
97. Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *PNAS* 74:5463–67
98. Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–70
99. Schenkel LC, Kernohan KD, McBride A, Reina D, Hodge A, et al. 2017. Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome. *Epigenet. Chromatin* 10:10
100. Schmidt J, Wollnik B. 2018. Hallermann-Streiff syndrome: a missing molecular link for a highly recognizable syndrome. *Am. J. Med. Genet. C* 178:398–406
101. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 30:e57
102. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. 2016. Untargeted metabolomics strategies—challenges and emerging directions. *J. Am. Soc. Mass Spectrom.* 27:1897–1905
103. Shabani M, Dyke SOM, Marelli L, Borry P. 2019. Variant data sharing by clinical laboratories through public databases: consent, privacy and further contact for research policies. *Genet. Med.* 21:1031–37
104. Shamseldin HE, Maddirevula S, Faqeih E, Ibrahim N, Hashem M, et al. 2017. Increasing the sensitivity of clinical exome sequencing through improved filtration strategy. *Genet. Med.* 19:593–98
105. Shashi V, Conkie-Rosell A, Rosell B, Schoch K, Vellore K, et al. 2014. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet. Med.* 16:176–82
106. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, et al. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728–32
107. Shi H, Enriquez A, Rapadas M, Martin EMMA, Wang R, et al. 2017. NAD deficiency, congenital malformations, and niacin supplementation. *N. Engl. J. Med.* 377:544–52
108. Siegel DH. 2018. PHACE syndrome: infantile hemangiomas associated with multiple congenital anomalies: clues to the cause. *Am. J. Med. Genet. C* 178:407–13
109. Silverman EK, Allard P, Loscalzo J, Mulvihill JJ, Korrick SA, Undiagn. Dis. Netw. 2019. Reported environmental exposures are inversely associated with obtaining a genetic diagnosis in the Undiagnosed Diseases Network. *Am. J. Med. Genet. A* 179:958–65
110. Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–17
111. Strande NT, Riggs ER, Buchanan AH, Ceyhan-Birsoy O, DiStefano M, et al. 2017. Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* 100:895–906
112. Stuurman KE, Joosten M, van der Burgt I, Elting M, Yntema HG, et al. 2019. Prenatal ultrasound findings of rasopathies in a cohort of 424 fetuses: update on genetic testing in the NGS era. *J. Med. Genet.* 56:654–61
113. Taruscio D, Groft SC, Cederroth H, Melegh B, Lasko P, et al. 2015. Undiagnosed Diseases Network International (UDNI): white paper for global actions to meet patient needs. *Mol. Genet. Metab.* 116:223–25
114. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, et al. 2018. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* 361:k1687
115. Wallace D, Singh G, Lott M, Hodge J, Schurr T, et al. 1988. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* 242:1427–30
116. Warner JP, Barron LH, Goudie D, Kelly K, Dow D, et al. 1996. A general method for the detection of large CAG repeat expansions by fluorescent PCR. *J. Med. Genet.* 33:1022–26
117. World Econ. Forum. 2019. Breaking Barriers to Health Data project. *World Economic Forum*. **https://www.weforum.org/projects/breaking-barriers-to-health-data-project**

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)

118. Wright CF, McRae JF, Clayton S, Gallone G, Aitken S, et al. 2018. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genet. Med.* 20:1216–23
119. Wright CF, Prigmore E, Rajan D, Handsaker J, McRae J, et al. 2019. Clinically-relevant postzygotic mosaicism in parents and children with developmental disorders in trio exome sequencing data. *Nat. Commun.* 10:2985
120. Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, et al. 1991. Fragile X genotype characterized by an unstable region of DNA. *Science* 252:1179–81

Annu. Rev. Genom. Hum. Genet. 2020.21. Downloaded from www.annualreviews.org
Access provided by 2603:6000:8a00:7603:c3f:f54c:ea58:cb35 on 05/17/20. For personal use only.

Review in Advance first posted on April 13, 2020. (Changes may still occur before final publication.)