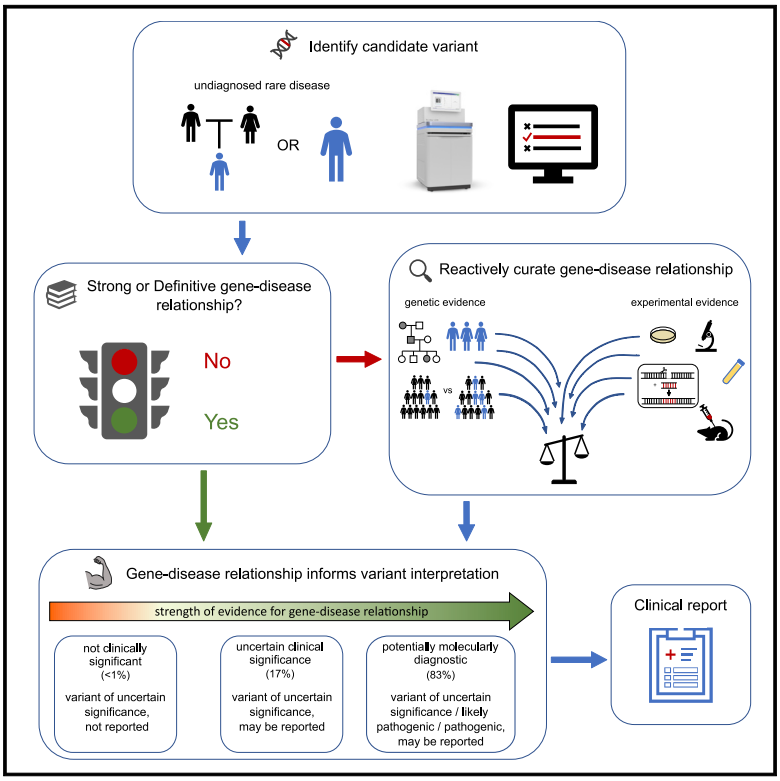
Cell Genomics

Short Article

Reactive gene curation to support interpretation and reporting of a clinical genome test for rare disease: Experience from over 1,000 cases

Graphical abstract



Highlights

- Evaluation of gene-disease relationships is essential for variant interpretation
- Experience using reactive gene curation to support a clinical genome test
- Rigorous reactive gene curation enables robust clinical decision making

Authors

Amanda R. Clause, Julie P. Taylor, Revathi Rajkumar, ..., Ryan J. Taft, Denise L. Perry, Alison J. Coffey

Correspondence

acoffey@illumina.com

In brief

Clause et al. describe a clinical laboratory's experience implementing reactive evaluation of gene-disease relationships within the interpretation and reporting workflow of a clinical genome sequencing test for rare genetic diseases. This work demonstrates that rigorous, reactive gene curation aids in variant interpretation and meets a critical need for clinical reporting.

Clause et al., 2023, Cell Genomics *3*, 100258
February 8, 2023
https://doi.org/10.1016/j.xgen.2023.100258

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Short Article

Reactive gene curation to support interpretation and reporting of a clinical genome test for rare disease: Experience from over 1,000 cases

Amanda R. Clause,[1] Julie P. Taylor,[1] Revathi Rajkumar,[1] Krista Bluske,[1] Maren Bennett,[1] Laura M. Amendola,[1] ICSL Interpretation and Reporting Team[1] David R. Bentley,[1] Ryan J. Taft,[1] Denise L. Perry,[1] and Alison J. Coffey[1,2,*]

[1]Medical Genomics Research, Illumina, Inc., San Diego, CA 92122, USA

[2]Lead contact

*Correspondence: acoffey@illumina.com

https://doi.org/10.1016/j.xgen.2023.100258

SUMMARY

Current standards in clinical genetics recognize the need to establish the validity of gene-disease relationships as a first step in the interpretation of sequence variants. We describe our experience incorporating the ClinGen Gene-Disease Clinical Validity framework in our interpretation and reporting workflow for a clinical genome sequencing (cGS) test for individuals with rare and undiagnosed genetic diseases. This "reactive" gene curation is completed upon identification of candidate variants during active case analysis and within the test turn-around time by focusing on the most impactful evidence and taking advantage of the broad applicability of the framework to cover a wide range of disease areas. We demonstrate that reactive gene curation can be successfully implemented in support of cGS in a clinical laboratory environment, enabling robust clinical decision making and allowing all variants to be fully and appropriately considered and their clinical significance confidently interpreted.

INTRODUCTION

Clinical genome sequencing (cGS) is a comprehensive genetic test that is emerging as a first-tier diagnostic tool for patients with rare and undiagnosed genetic diseases (RUGDs).[1–6] Patients for whom cGS is indicated present with a wide range of clinical manifestations, requiring knowledge across an increasing number of genes. In addition to confirming the genetic etiology of disease, evaluating the clinical validity of evidence for a gene-disease relationship (GDR) is considered an essential component of variant classification, where variant pathogenicity should not be determined without first establishing the strength of the relationship between the gene and the disease.[7,8] Gene curation is thus a key part of robust clinical reporting practices.

In 2017, Strande et al.[9] published the Clinical Genome Resource[10] (ClinGen) Gene-Disease Clinical Validity framework, a semiquantitative method to assess the strength of publicly available genetic and experimental evidence supporting or contradicting a GDR. Gene curation efforts by ClinGen and others[11–18] have largely taken the approach of curating a predefined list of genes proposed to be associated with a particular disorder ("proactive" gene curation). As of September 1, 2022, 1,480 unique genes curated by ClinGen are publicly available.[19] In addition to the efforts of ClinGen, PanelApp is a curation platform that employs crowdsourcing[15,16] and has the advantage of distributing the workload to reach consensus on which genes have sufficient evidence for disease association to be included on genetic testing panels. These centralized repositories of large numbers of GDRs (genes and genomic entities as of September 1, 2022: Genomics England PanelApp, 6,122; PanelApp Australia, 5,548) are currently in use across many diagnostic labs. However, even with these combined efforts, it is unlikely that the over 10,000 genes projected to be causal for monogenic disease[20] will be able to be proactively curated in a timely fashion. The absence of a fully curated genome creates an ongoing need for gene curation by clinical laboratories, particularly those who offer comprehensive and unbiased testing approaches like GS, which interrogate the entire genome and for which proactive curation of all genes of potential interest is not currently feasible.

Here, we describe the Illumina Clinical Services Laboratory (ICSL)'s implementation of the ClinGen framework into the interpretation and reporting workflow for a cGS test for individuals with a suspected RUGD. In contrast to proactive approaches, "reactive" gene curation is completed within the test turn-around time for genes in which specific candidate variants are identified during active case analysis. We demonstrate that rigorous, reactive gene curation aids in the triage, classification, and reporting of variants for a cGS test, enabling robust clinical decision making, including reporting of potentially relevant findings in genes with emerging evidence.

RESULTS

Reactive gene curation supports a genome sequencing test for RUGDs

The implementation of reactive gene curation into the clinical workflow is illustrated in Figure 1A. When analysis of the genomic

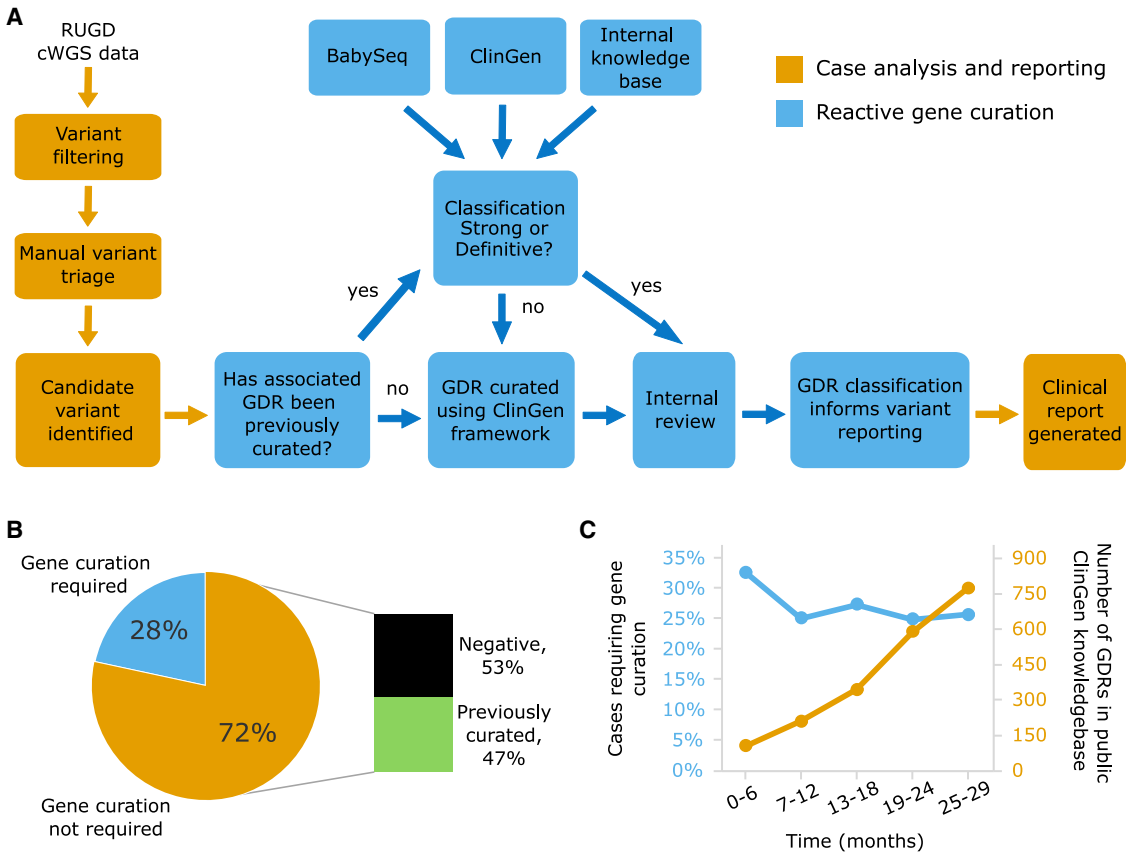


Figure 1. Reactive gene curation supports a cGS test for RUGDs

(A) Flow diagram outlining the incorporation of reactive gene curation into ICSL's clinical interpretation and reporting workflow.

(B) Percentage of cases for which gene curation was required.

(C) Percentage of cases requiring gene curation as a function of time relative to the first analyzed case (blue) compared with the growth in GDR classifications publicly available through ClinGen (orange).

sequence of the proband (case analysis) identifies a candidate variant that may explain all or some of the proband's phenotype, the associated GDR is reactively curated within the interpretation and reporting window, with the exception of GDRs previously classified as strong or definitive by the BabySeq Project,[11] ClinGen, or internally; these classifications are used for reporting decisions without additional curation. GDRs that have not been previously curated are curated using the ClinGen framework and subjected to internal review, and the final GDR classification is used to inform variant classification and reporting decisions.

Over 29 months, ICSL completed cGS analysis for the probands of 1,037 unrelated families. Reactive gene curation was required for 28% of cases (Figure 1B), for a total of 286 GDRs. Of the 72% that did not require gene curation, approximately half were negative cases with no candidate variants identified (53%), and half had a variant identified in a gene with a GDR that was previously classified as strong or definitive (47%). Despite considerable growth in ClinGen's public knowledgebase of gene-disease clinical validity over the same period, the percentage of cases requiring gene curation held largely stable over time (Figure 1C).

Our test population includes individuals with a suspected RUGD, which are often multi-system in nature. To assess consistency in outcomes between reactive and proactive gene curation approaches across the considerable diversity in clinical presentation and biological mechanism seen in our laboratory and curated reactively (Figure S1), we evaluated concordance between our GDR classifications and assertions now publicly available from ClinGen, which had not been available at the time of our initial curation. We found a high degree of concordance. Of 58 GDRs for which ClinGen had published a classification for the same disorder, 48 were fully concordant (Table S1). For six of the remaining nine GDRs, the discrepancy could be attributed to differences in evidence available at the time of curation. For three GDRs, the difference was between definitive and strong, which is not clinically significant. Only one GDR, involving the *LAMA4* gene, was truly discordant (no known disease relationship [NKDR]-animal model only [ICSL] versus limited [ClinGen]) due to our more conservative approach to scoring case data in this instance.

Genetic and experimental evidence supporting GDRs for RUGDs

The ClinGen framework takes both genetic and experimental evidence into account. The distribution of evidence across the

274 GDRs with a classification of limited or above is shown in Figure 2A. The maximum of 12 points (pts) genetic evidence was awarded for 148 of the 192 GDRs classified as strong or definitive (77%). Fewer than 12 pts genetic evidence was awarded in 23%, such that experimental evidence was the difference between a moderate and strong/definitive classification. Fourteen GDRs classified as limited or above lacked experimental evidence. These GDRs encompassed the full range of classifications, from limited to definitive. For an additional 23 GDRs, $\leq$0.5 pts experimental evidence was awarded; this evidence typically reflected a biochemical function or expression pattern consistent with the disease. Experimental evidence resulted in an upgrade from limited in 40% of the 47 GDRs classified as moderate.

Curation of animal models is prioritized during reactive curation because they can provide strong evidence linking a gene to disease as well as insight into the underlying biological mechanisms.[21–23] A range of organisms was encountered, including mouse, *Xenopus*, zebrafish, *Drosophila*, *C. elegans*, and rarely others such as rat, ferret, dog, pig, calf, chick, and yeast. An animal model was curated for 79% of GDRs (Figure 2B). However, among these, 12% were not scored. Reasons for not scoring included a mismatch in molecular mechanism with the curated disease, embryonic lethality without further investigation, and lack of homology between the human and model organism gene. Mouse models were by far the most common organism scored and more often strongly recapitulated the features observed in human patients (Figure 2C).

Co-segregation of variants with a disease phenotype within a multi-generational family provides additional support for a GDR. Within the framework, segregation can be scored as genetic evidence ($\leq$3 pts). Segregation evidence was curated in 32% of GDRs (Figure 2D), but points were awarded for only half. In the remaining cases, the estimated or author-reported LOD score did not meet the framework's scoring threshold, genotyping of the family was incomplete, or the testing methodology did not account for genetic heterogeneity.

Large cohort studies identified through matchmaking networks and tools, such as GeneMatcher[24] and Matchmaker Exchange,[25] are playing an increasing role in identifying new GDRs. This type of genotype-first approach may be particularly impactful for RUGDs and other suspected genetic disorders that are not yet clinically recognizable based on presentation. 15% (n = 42) of curated GDRs included a publication describing a case series coordinated through GeneMatcher (Figure 2E). For 18 of these GDRs, the publication arising from GeneMatcher-facilitated collaborations was the source of all case data included in the curation, and for 14, the GeneMatcher publication alone provided sufficient data to score the maximum allotted to genetic evidence, enabling a GDR classification of strong or definitive. In cases of a definitive classification, the GeneMatcher publication involved collaboration across multiple institutions, often internationally, and was taken to meet the framework's requirement for replication.

Impact of gene curation on variant interpretation and reporting

The use of GDR classifications in variant classification and reporting decisions is illustrated in Figure 3A. As suggested by the American College of Medical Genetics and Genomics (ACMG),[7] the classification of variants in genes with a moderate GDR is capped at likely pathogenic (LP). The classification of variants in genes with a GDR classification of limited, NKDR, NKDR-animal model only, disputed, and refuted is capped at variant of uncertain significance (VUS). Variants in genes with a GDR classification of NKDR, disputed, or refuted are typically not reported, although in rare circumstances, variants in genes with an NKDR GDR can be reported as a candidate for research. Genes with a limited or NKDR (with or without an animal model) GDR classification may be submitted to GeneMatcher if the proband, variant, and gene meet internal criteria.[26]

Of the 286 GDRs curated, 83% (n = 238) were classified as moderate, strong, or definitive and therefore had sufficient evidence to support reporting a variant as a potential molecularly diagnostic finding. 17% (n = 47) were classified as limited, NKDR, or NKDR-animal model only and considered to be of uncertain clinical significance (Figures 3B and 3C). Only one GDR was classified as disputed; this curation was conducted to evaluate an asserted disease association for heterozygous variants in a gene (*SAMHD1*) with an established relationship with an autosomal recessive disorder.

In total, 379 variants were reported for reactively curated GDRs (Figure 3D). For definitive GDRs, the number of reported variants classified as LP or pathogenic (P) exceeded that classified as VUS, suggesting that accumulation of evidence in support of a GDR may help reduce numbers of reported VUSs. Clinical judgment was applied to classify three variants in genes with a moderate GDR as P despite the recommended cap at LP: for all three GDRs, the variant identified in the proband had been reported in a significant number or even all published cases of the disorder.

80% of curated GDRs were reported only once. Of the 56 recurrent GDRs for which variants were reported in at least two unrelated probands, only 10 were reported three or more times (Figure 3E; Table S2). Most recurrent GDRs (84%) were classified as strong or definitive. GDRs associated with nonspecific phenotypes common in our cohort or with a clinical presentation with variable expressivity or reduced penetrance may be hard to rule out as being contributory to the proband's phenotype, particularly in the absence of many previously described patients or detailed phenotyping.

Curating GDRs with limited evidence for RUGDs

While gene panels aim to interrogate primarily genes with an established link to a given phenotype, cGS includes analysis of novel genetic etiologies in addition to well-characterized relationships. As a result, variants in genes with little or emerging evidence to suggest a potential association and/or relevance for the proband can be identified. Variants in genes that lack an established relationship to disease are valuable to report as research candidates provided there is valid evidence that the GDR is relevant to the proband's phenotype and supportive of variant pathogenicity. We consider GDRs classified as limited, NKDR, and NKDR-animal model only to be "genes of uncertain clinical significance"[7] but eligible for reporting. VUSs were reported for 29 of the 47 GDRs with these classifications (62%). Most often, the decision to report was due to compelling overlap in phenotype

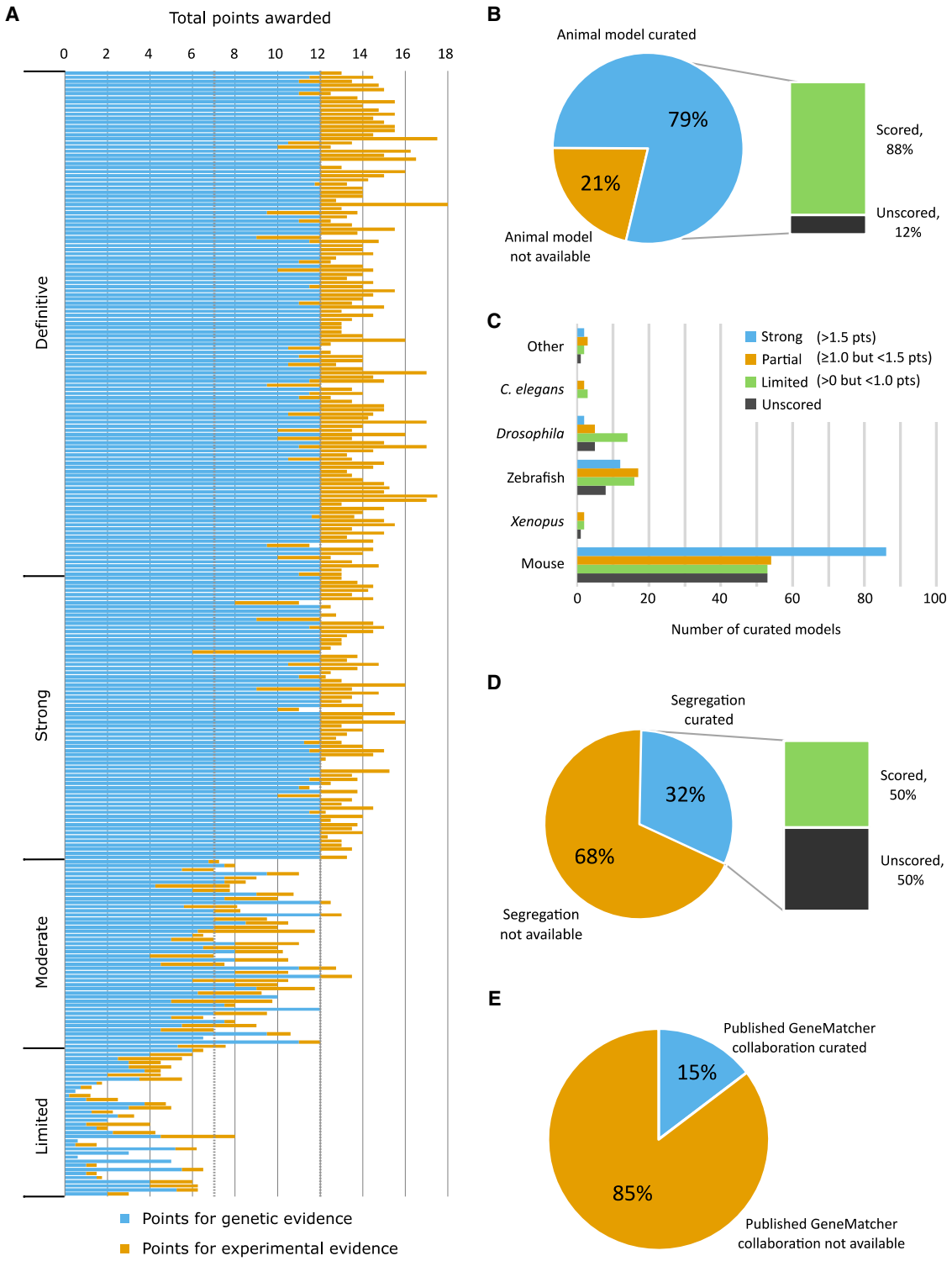


Figure 2. Evidence supporting GDRs for rare disease

(A) Points awarded to genetic and experimental evidence across the 274 curated GDRs with a classification of limited and above. Each row represents a single GDR. Dashed vertical lines show the boundary between limited and moderate (7 pts) and moderate and strong/definitive (12 pts).

(B) Percentage of GDRs for which animal models were curated and scored.

(C) Number of curated models plotted according to the degree of phenotype recapitulation and species.

(D) Percentage of GDRs for which segregation evidence was curated and scored.

(E) Percentage of GDRs for which a cohort study organized through GeneMatcher was curated.

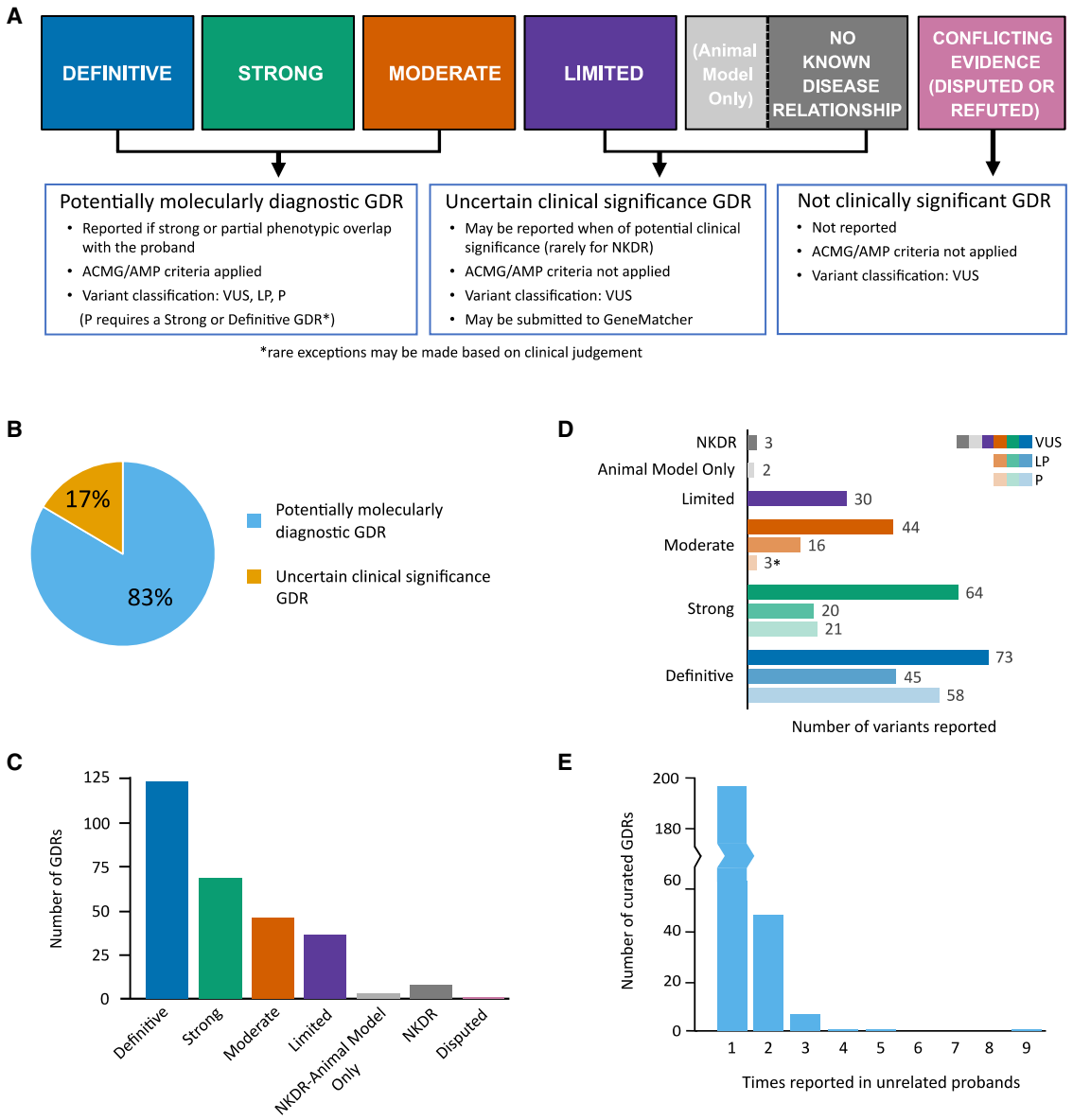


Figure 3. Reactive gene curation supports clinical reporting for a cGS test for RUGDs

(A) Use of GDR classification in variant classification and reporting.

(B) Categorization of GDRs for clinical reporting as a percentage of total GDRs curated. The GDR classified as disputed is not depicted.

(C) Number of curated GDRs within each gene-disease validity classification.

(D) Number of variants reported for curated GDRs, stratified according to variant and GDR classification. *Variant reported as P for moderate GDR based on clinical judgment.

(E) Recurrent GDRs for which variants were reported in multiple unrelated probands.

between the proband and previously reported cases (Tables 1 and S3). Consistency between the variant consequence and proposed disease mechanism and between variant inheritance and family history were also key factors. In the three GDRs classified as NKDR-animal model only, variants in the gene had not yet been reported in humans, but a strong overlap between the proband's phenotype and an animal model led to a decision to report.

For the remaining 18 GDRs, close examination of the evidence led to a variant not being reported. In many of these cases, curation uncovered evidence not appreciated at the time of variant triage that enabled subsequent confident rule out of the variant. Reasons not to report included limited or incomplete phenotype overlap with reported patients and/or animal model; insufficient evidence for the GDR due to unconvincing literature reports or contradictory evidence; insufficient evidence for the variant identified in the case proband, such as inconsistency with the proposed disease mechanism; and identification of another variant with stronger evidence.

Table 1. Clinical reporting decisions for a subset of GDRs of uncertain clinical significance

Gene	Disease	Reporting rationale
Reported		
AQP11[a]	polycystic kidney disease (MONDO: 0020642)	compound heterozygous rare missense variants in a proband with a highly specific phenotype that overlaps that in mouse models
DDR2	Warburg-Cinotti syndrome (MIM: 618175)	same variant as in previously reported cases in a proband with strong phenotype overlap
ENTPD1	hereditary spastic paraplegia 64 (MIM: 615683)	compound heterozygous variants (frameshift and splice region variant) in a proband with significant phenotypic overlap with previously reported cases
SYP	X-linked intellectual disability (MIM: 300802)	hemizygous rare missense variant in affected brothers with phenotype overlap with previously reported cases
TSHZ1	congenital aural atresia (MIM: 607842)	heterozygous rare start-lost variant in a proband with highly specific phenotype overlap with previously reported cases
UNC13A	congenital nervous system disorder (MONDO: 0002320)	*de novo* rare missense variant in a proband with phenotype overlap with previously reported cases
Not reported		
COL12A1	Bethlem myopathy (MIM: 616471)	*de novo* rare missense variant predicted damaging not reported due to unclear/insufficient phenotype overlap
KCND2	complex neurodevelopmental disorder (MONDO: 0100038)	highly conserved variant not present in single parent available in a proband not reported due to partial/incomplete phenotype overlap
LAMA4[a]	dilated cardiomyopathy (MONDO: 0014095)	rare essential splice variant inherited from an unaffected parent in a proband with phenotype overlap but not reported due to weak evidence for GDR
TGM6	spinocerebellar ataxia (MIM: 613908)	rare frameshift variant not present in available parent in proband with limited phenotype overlap with previously reported cases in the context of weak and contradictory evidence for GDR
TNIK	intellectual disability (MIM: 617028)	compound heterozygous missense variants not reported due to limited phenotype overlap and weak evidence for GDR
TUBB1	macrothrombocytopenia (MIM: 613112)	inherited frameshift variant not reported due to inheritance and mismatch with proposed disease mechanism

See also Table S3. Unless otherwise noted, GDRs were classified as limited.
[a]NKDR-animal model only.

For most GDRs classified as NKDR, multiple published case reports asserted an association, but the cases were not scored due to insufficient evidence of variant pathogenicity. Further expert review may enable reclassification of some of these as disputed. Clinical judgment led to the reporting of VUSs in three genes with a GDR classified as NKDR (Table S3). In each case, the decision to report was based on evidence that could not be formally incorporated within the ClinGen framework, resulting in a lower classification than expected. This included unpublished case-level evidence in GeneMatcher for two cases and an inability to score reported variants due to the involvement of genetic material outside the gene of interest for the third.

Impact of recurating GDRs

We recurated 26 GDRs, including four originally classified as limited, 10 as moderate, and 12 as strong. The median time between the original curation and reevaluation was 24.5 months (range 6–35 months). Recuration led to reclassification of 77%, with an upgrade of 19 GDRs (limited to moderate, n = 2; moderate to strong/definitive, n = 6; strong to definitive, n = 11) (Figure S2A; Table S4). The time elapsed between the original curation and the availability of new evidence ranged from 1 to 31 months (median 10 months), excluding the seven GDRs for which new evidence had not been published at the time of recuration (Figure S2B).

DISCUSSION

This study presents our clinical laboratory's experience incorporating reactive gene curation into an interpretation and reporting workflow for a rare disease cGS test. Incorporating the ClinGen framework into the clinical interpretation and reporting process meets the need for principled evaluation of GDRs, which is essential to deliver rigorous and consistent evaluation of variants. Our cohort spans a range of ages, ethnicities, geographic regions, and indications for testing, which may differ from that seen in other clinical laboratories but speaks to the general applicability of gene curation across diverse populations. We have demonstrated the feasibility of incorporating reactive gene curation into case analysis for a clinical test and have outlined strategies that may enable others to adopt a similar approach.

Many ClinGen gene curation expert panels focus on a single condition of interest, leaving a gap in the availability of robustly evaluated, publicly available GDR classifications for many of the

multi-system disorders seen in this study. While multi-system disorders are often individually rare and show variable expressivity, exome and genome sequencing are more likely to be recommended and are increasingly utilized early in the diagnostic evaluation for patients with complex presentations.[27] This diversity necessitates a unique expert-generalist approach where curators are experienced in the application of the framework and critical evaluation of supportive evidence rather than experts in a specific disease area. The broad applicability of the framework across disease areas, as illustrated by the diversity of gene curation expert panels as well as our own curations, demonstrates the framework's strength and enables consistent and accurate curations with the expert-generalist approach. Within ClinGen, the Syndromic Disorders Gene Curation Expert Panel[28] was established to address multi-system GDRs and also adopted the expert-generalist approach. However, further community efforts are required to meet the need for timely curation of all multi-system GDRs.

Focusing on the most impactful evidence is an effective strategy to implement reactive gene curation while accommodating the time constraints of a clinical interpretation and reporting workflow. A confident clinical validity classification can be reached more quickly by prioritizing case data and high-impact experimental evidence, such as animal models. Comprehensive experimental evidence curation is most impactful when it can resolve a borderline classification. The high concordance between our classifications and those from ClinGen demonstrates that these strategies do not compromise GDR classifications and validates our implementation of reactive gene curation within the test turn-around time.

Public sharing of gene curation data is critical to reduce the numbers of reactive gene curations required in a clinical workflow. For example, the 1,500 GDRs curated by the BabySeq project,[11,29] a critical step in implementing GS in newborns, provided a foundational knowledgebase to reduce the curation burden for RUGDs, particularly for pediatric cases. Nevertheless, we saw only a modest decline in gene curation burden over an ~2.5-year period despite growth in public knowledgebases of gene-disease validity. This largely stable requirement for gene curation likely reflects the relatively low rate of recurrence of GDRs for RUGDs, the fast pace of discovery of new disease genes,[20] and a lack of focus of community gene curation efforts on the types of disorders observed in our cohort. However, the recent launch of the GenCC,[30] of which we are a contributing member, has led to an order of magnitude increase in the public GDR knowledgebase by bringing together and harmonizing assessments of GDRs from groups across the globe regardless of methodology. Sharing can also enable the identification and resolution of conflicts in classification, as demonstrated by efforts by PanelApp in the UK and Australia.[15] We piloted use of GenCC data and found they eliminated the need for 13 gene curations in 4 months, highlighting the direct benefit of a centralized and transparent "ClinVar for genes" on clinical interpretation as well as the importance of data sharing to avoid duplication of effort.

Gene curation is valuable for all GDRs, but arguably the biggest impact of gene curation is in giving confidence in reporting decisions for variants identified in genes of uncertain significance (GUSs). Although such variants are necessarily VUSs, they can be considered research candidates, highlight a possibly novel diagnosis, or act as a flag for future reanalysis. Within the global clinical genetics community, there is currently a lack of consensus regarding reporting of VUSs in GUSs, with US-based recommendations from the Medical Genome Initiative[31] advocating for reporting of VUSs in GUSs when they are considered strong candidates and guidelines from the European Society of Human Genetics[32] suggesting that such findings should not be reported clinically and instead be restricted to a separate research report. While GDRs with less evidence may not always be suitable for inclusion on genetic testing panels[7,8] and can thus be avoided for panel tests, the potential to uncover research candidates can be considered a benefit of unbiased testing approaches like cGS, particularly for the patient populations for whom this type of test is increasingly indicated as a first-tier approach.[1,2,5] Reporting of VUSs in GUSs, where there is a limited but valid suggestion of significance, is particularly impactful for individuals with RUGDs, whose chance of receiving a potentially informative finding should not be constrained by the absence of large numbers of previously described cases. For ultra-rare diseases, there is an ethical imperative to return VUSs in GUSs when of potential clinical relevance. Principled consideration of these GDRs prevents candidate variants from being unnecessarily ruled out but, critically, prevents the harm of returning VUSs in genes without some evidence of a causal relationship with disease. Furthermore, the strength of the GDR also guides variant classification, appropriately limiting assertions of pathogenicity based on the level of evidence for the GDR.[7]

It is important to monitor for the publication of new supporting or contradictory evidence, additional disease assertions, or data impacting the designation of multiple unique disease entities associated with a single gene, which may result in recuration of GDRs. However, resource constraints within a clinical laboratory can limit recuration efforts. Reactive recuration based on automated literature searches or recurrence of a GDR in an active case, rather than predefined timescales, may be an efficient approach. In the clinical setting, before a recuration is initiated, preliminary evaluation of whether newly published evidence will result in a change in GDR classification, and variant reclassification and an amended report, can further target resource allocation and be more impactful for patients.

The rapid growth in novel gene-disease discoveries, including those arising from large collaborative efforts like GeneMatcher, is an exciting development in human genetics.[33] These publications often contain substantial case data, enabling a GDR to be quickly classified as strong or definitive without extensive experimental data. Prioritizing and enabling publication of aggregated cases without the requirement for experimental support should be encouraged to allow more rapid sharing of nascent GDRs, which is beneficial for the triage and curation of candidate variants and helps to increase the chance that patients and their families will receive a potentially informative finding.

While the gene curation still needed for a fully curated genome could be considered a finite undertaking, this would necessitate sizable resources. ClinGen, PanelApp, and the GenCC harness the power of combined community efforts. At the level of individual genes, however, accelerating gene curation efforts requires

development of software and automation tools to aggregate publicly available data and identify relevant literature, a time-consuming part of the process, so curators can focus on evidence interpretation rather than gathering.

Rigorous gene curation is necessary for accurate and consistent clinical interpretation across variant triage, classification, and reporting decisions. A fully curated set of GDRs for the human genome would be a tremendous asset to clinical interpretation of GS data, improve standardization and consistency in gene panels,[7,34,35] and potentially increase diagnostic rates.[36] We have demonstrated that until we have a fully curated genome, focused reactive gene curation, regardless of methodology, can be implemented to meet a critical need for clinical reporting.

Limitations of the study

Reactive gene curation aims to provide accurate and up-to-date evaluation of the clinical validity of a GDR to support variant interpretation and reporting. However, the ClinGen framework limits consideration to publicly available evidence, which could lead to underestimation of the strength of the GDR and possibly variant pathogenicity in some cases. The ClinGen framework is also designed for use with monogenic, Mendelian disorders. As such, it will not address all situations, such as risk factors, that may be encountered during the interpretation of genetic findings by a clinical laboratory. We have implemented the framework as generalists rather than experts in a particular disease area, and while our concordance analysis supports the legitimacy of this approach, expert input could influence the interpretation and scoring of evidence in some cases. Implementing reactive gene curation, particularly using a rigorous method like the ClinGen framework, requires investment in personnel training and procedures to ensure consistency. While this could pose a barrier to adoption, we find that the skills required for variant triage, curation, and interpretation are transferable to gene-level analysis and argue that reactive gene curation is valuable regardless of the methodology used. Finally, this study reflects the experience of a single clinical laboratory; the experiences of other laboratories adopting a similar approach may differ due to differences in patient populations, personnel experience and expertise, and resources.

CONSORTIA

ICSL interpretation and reporting team: Carolyn Brown, Matthew P. Brown, Amanda Buchanan, Brendan Burns, Nicole J. Burns, Anjana Chandrasekhar, Aditi Chawla, Katie Golden-Grant, Akanchha Kesari, Alka Malhotra, Becky Milewski, Samin A. Sajan, Zinayida Schlachetzki, Sarah Schmidt, Brittany Thomas, Erin Thorpe.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Genome sequencing, analysis, and interpretation
  - Gene curation and clinical reporting
  - Re-evaluation of curated GDRs
  - Data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2023.100258.

ACKNOWLEDGMENTS

We thank J. Avecilla, V. Rajan, B. Juan, and S. Ajay for computational support and J. Belmont for thoughtful review.

AUTHOR CONTRIBUTIONS

Conceptualization, A.R.C., J.P.T., R.R., K.B., M.B., L.M.A., D.R.B., and A.J.C.; investigation, A.R.C., J.P.T., R.R., K.B., ICSL Interpretation and Reporting Team, and A.J.C.; project administration, M.B. and R.R.; supervision, D.L.P., D.R.B., R.J.T., and A.J.C.; writing & editing, A.R.C., J.P.T., L.M.A., D.L.P., R.J.T., and A.J.C.

DECLARATION OF INTERESTS

All authors are employees and shareholders of Illumina, Inc.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: June 28, 2022
Revised: September 13, 2022
Accepted: January 6, 2023
Published: January 31, 2023

REFERENCES

1. Lionel, A.C., Costain, G., Monfared, N., Walker, S., Reuter, M.S., Hosseini, S.M., Thiruvahindrapuram, B., Merico, D., Jobling, R., Nalpathamkalam, T., et al. (2018). Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genet. Med. *20*, 435–443. https://doi.org/10.1038/gim.2017.119.
2. Scocchia, A., Wigby, K.M., Masser-Frye, D., Del Campo, M., Galarreta, C.I., Thorpe, E., McEachern, J., Robinson, K., Gross, A., et al.; ICSL Interpretation and Reporting Team (2019). Clinical whole genome sequencing as a first-tier test at a resource-limited dysmorphology clinic in Mexico. NPJ Genom. Med. *4*, 5. https://doi.org/10.1038/s41525-018-0076-1.
3. Turro, E., Astle, W.J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Allen, H.L., Sanchis-Juan, A., Frontini, M., Thys, C., et al. (2020). Whole-genome sequencing of patients with rare diseases in a national health system. Nature *583*, 96–102. https://doi.org/10.1038/s41586-020-2434-2.
4. Bertoli-Avella, A.M., Beetz, C., Ameziane, N., Rocha, M.E., Guatibonza, P., Pereira, C., Calvo, M., Herrera-Ordonez, N., Segura-Castel, M., Diego-Alvarez, D., et al. (2021). Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort. Eur. J. Hum. Genet. *29*, 141–153. https://doi.org/10.1038/s41431-020-00713-9.

5. NICUSeq Study Group; Krantz, I.D., Medne, L., Weatherly, J.M., Wild, K.T., Biswas, S., Devkota, B., Hartman, T., Brunelli, L., Fishler, K.P., et al. (2021). Effect of whole-genome sequencing on the clinical management of acutely ill infants with suspected genetic disease: a randomized clinical trial. JAMA Pediatr. *175*, 1218–1226. https://doi.org/10.1001/jamapediatrics.2021.3496.

6. Incerti, D., Xu, X.-M., Chou, J.W., Gonzaludo, N., Belmont, J.W., and Schroeder, B.E. (2022). Cost-effectiveness of genome sequencing for diagnosing patients with undiagnosed rare genetic diseases. Genet. Med. *24*, 109–118. https://doi.org/10.1016/j.gim.2021.08.015.

7. Bean, L.J.H., Funke, B., Carlston, C.M., Gannon, J.L., Kantarci, S., Krock, B.L., Zhang, S., and Bayrak-Toydemir, P.; ACMG Laboratory Quality AssuranceCommittee (2020). Diagnostic gene sequencing panels: from design to report-a technical standard of the American College of Medical Genetics and Genomics (ACMG). Genet. Med. *22*, 453–461. https://doi.org/10.1038/s41436-019-0666-z.

8. Thaxton, C., Good, M.E., DiStefano, M.T., Luo, X., Andersen, E.F., Thorland, E., Berg, J., Martin, C.L., Rehm, H.L., Riggs, E.R., et al. (2022). Utilizing ClinGen gene-disease validity and dosage sensitivity curations to inform variant classification. Hum. Mutat. *43*, 1031–1040. https://doi.org/10.1002/humu.24291.

9. Strande, N.T., Riggs, E.R., Buchanan, A.H., Ceyhan-Birsoy, O., DiStefano, M., Dwight, S.S., Goldstein, J., Ghosh, R., Seifert, B.A., Sneddon, T.P., 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. https://doi.org/10.1016/j.ajhg.2017.04.015.

10. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen–the clinical genome resource. N. Engl. J. Med. *372*, 2235–2242. https://doi.org/10.1056/NEJMsr1406261.

11. Holm, I.A., Agrawal, P.B., Ceyhan-Birsoy, O., Christensen, K.D., Fayer, S., Frankel, L.A., Genetti, C.A., Krier, J.B., LaMay, R.C., Levy, H.L., et al. (2018). The BabySeq project: implementing genomic sequencing in newborns. BMC Pediatr. *18*, 225. https://doi.org/10.1186/s12887-018-1200-1.

12. Farwell Hagman, K.D., Shinde, D.N., Mroske, C., Smith, E., Radtke, K., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E.C., Alcaraz, W.A., et al. (2017). Candidate-gene criteria for clinical reporting: diagnostic exome sequencing identifies altered candidate genes among 8% of patients with undiagnosed diseases. Genet. Med. *19*, 224–235. https://doi.org/10.1038/gim.2016.95.

13. Adler, A., Novelli, V., Amin, A.S., Abiusi, E., Care, M., Nannenberg, E.A., Feilotter, H., Amenta, S., Mazza, D., Bikker, H., et al. (2020). An international, multicentered, evidence-based reappraisal of genes reported to cause congenital long QT syndrome. Circulation *141*, 418–428. https://doi.org/10.1161/CIRCULATIONAHA.119.043132.

14. Hosseini, S.M., Kim, R., Udupa, S., Costain, G., Jobling, R., Liston, E., Jamal, S.M., Szybowska, M., Morel, C.F., Bowdin, S., et al. (2018). Reappraisal of reported genes for sudden arrhythmic death: evidence-based evaluation of gene validity for brugada syndrome. Circulation *138*, 1195–1205. https://doi.org/10.1161/CIRCULATIONAHA.118.035070.

15. Stark, Z., Foulger, R.E., Williams, E., Thompson, B.A., Patel, C., Lunke, S., Snow, C., Leong, I.U.S., Puzriakova, A., Daugherty, L.C., et al. (2021). Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. Am. J. Hum. Genet. *108*, 1551–1557. https://doi.org/10.1016/j.ajhg.2021.06.020.

16. Martin, A.R., Williams, E., Foulger, R.E., Leigh, S., Daugherty, L.C., Niblock, O., Leong, I.U.S., Smith, K.R., Gerasimenko, O., Haraldsdottir, E., et al. (2019). PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat. Genet. *51*, 1560–1565. https://doi.org/10.1038/s41588-019-0528-2.

17. Thormann, A., Halachev, M., McLaren, W., Moore, D.J., Svinti, V., Campbell, A., Kerr, S.M., Tischkowitz, M., Hunt, S.E., Dunlop, M.G., et al. (2019). Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. Nat. Commun. *10*, 2373. https://doi.org/10.1038/s41467-019-10016-3.

18. Transforming Genetic Medicine Initiative (TGMI) Gene2Phenotype. https://www.ebi.ac.uk/gene2phenotype.

19. ClinGen Gene-Disease Validity Curations (2022). https://search.clinicalgenome.org/kb/gene-validity?page=1&size=25&search=.

20. Bamshad, M.J., Nickerson, D.A., and Chong, J.X. (2019). Mendelian gene discovery: fast and furious with No end in sight. Am. J. Hum. Genet. *105*, 448–455. https://doi.org/10.1016/j.ajhg.2019.07.011.

21. Adamson, K.I., Sheridan, E., and Grierson, A.J. (2018). Use of zebrafish models to investigate rare human disease. J. Med. Genet. *55*, 641–649. https://doi.org/10.1136/jmedgenet-2018-105358.

22. Hmeljak, J., and Justice, M.J. (2019). From gene to treatment: supporting rare disease translational research through model systems. Dis. Model. Mech. *12*, dmm039271. https://doi.org/10.1242/dmm.039271.

23. Brown, S.D.M. (2021). Advances in mouse genetics for the study of human disease. Hum. Mol. Genet. *30*, R274–R284. https://doi.org/10.1093/hmg/ddab153.

24. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. *36*, 928–930. https://doi.org/10.1002/humu.22844.

25. Philippakis, A.A., Azzariti, D.R., Beltran, S., Brookes, A.J., Brownstein, C.A., Brudno, M., Brunner, H.G., Buske, O.J., Carey, K., Doll, C., et al. (2015). The Matchmaker Exchange: a platform for rare disease gene discovery. Hum. Mutat. *36*, 915–921. https://doi.org/10.1002/humu.22858.

26. Taylor, J.P., Malhotra, A., Burns, N.J., Clause, A.R., Brown, C.M., Burns, B.T., Chandrasekhar, A., Schlachetzki, Z., Bennett, M., Thorpe, E., et al. (2022). A clinical laboratory's experience using GeneMatcher-Building stronger gene-disease relationships. Hum. Mutat. *43*, 765–771. https://doi.org/10.1002/humu.24356.

27. Manickam, K., McClain, M.R., Demmer, L.A., Biswas, S., Kearney, H.M., Malinowski, J., Massingham, L.J., Miller, D., Yu, T.W., and Hisama, F.M.; ACMG Board of Directors (2021). Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). Genet. Med. *23*, 2029–2037. https://doi.org/10.1038/s41436-021-01242-6.

28. Syndromic Disorders Gene Curation Expert Panel - ClinGen | Clinical Genome Resource https://clinicalgenome.org/affiliation/40060/.

29. Ceyhan-Birsoy, O., Machini, K., Lebo, M.S., Yu, T.W., Agrawal, P.B., Parad, R.B., Holm, I.A., McGuire, A., Green, R.C., Beggs, A.H., and Rehm, H.L. (2017). A curated gene list for reporting results of newborn genomic sequencing. Genet. Med. *19*, 809–818. https://doi.org/10.1038/gim.2016.193.

30. DiStefano, M.T., Goehringer, S., Babb, L., Alkuraya, F.S., Amberger, J., Amin, M., Austin-Tse, C., Balzotti, M., Berg, J.S., Birney, E., et al. (2022). The Gene Curation Coalition: a global effort to harmonize gene-disease evidence resources. Genet. Med. *24*, 1732–1742. https://doi.org/10.1016/j.gim.2022.04.017.

31. Austin-Tse, C.A., Jobanputra, V., Perry, D.L., Bick, D., Taft, R.J., Venner, E., Gibbs, R.A., Young, T., Barnett, S., Belmont, J.W., et al. (2022). Best practices for the interpretation and reporting of clinical whole genome sequencing. npj Genom. Med. *7*, 27–13. https://doi.org/10.1038/s41525-022-00295-z.

32. Souche, E., Beltran, S., Brosens, E., Belmont, J.W., Fossum, M., Riess, O., Gilissen, C., Ardeshirdavani, A., Houge, G., van Gijn, M., et al. (2022). Recommendations for whole genome sequencing in diagnostics for rare diseases. Eur. J. Hum. Genet. *30*, 1017–1021. https://doi.org/10.1038/s41431-022-01113-x.

33. Boycott, K.M., Azzariti, D.R., Hamosh, A., and Rehm, H.L. (2022). Seven years since the launch of the Matchmaker Exchange: the evolution of

genomic matchmaking. Hum. Mutat. *43*, 659–667. https://doi.org/10.1002/humu.24373.

34. Chambers, C., Jansen, L.A., and Dhamija, R. (2016). Review of commercially available epilepsy genetic panels. J. Genet. Counsel. *25*, 213–217. https://doi.org/10.1007/s10897-015-9906-9.
35. Ben-Shachar, R., Svenson, A., Goldberg, J.D., and Muzzey, D. (2019). A data-driven evaluation of the size and content of expanded carrier screening panels. Genet. Med. *21*, 1931–1939. https://doi.org/10.1038/s41436-019-0466-5.
36. Wright, C.F., FitzPatrick, D.R., and Firth, H.V. (2018). Paediatric genomics: diagnosing rare disease in children. Nat. Rev. Genet. *19*, 253–268. https://doi.org/10.1038/nrg.2017.116.
37. Vanderver, A., Bernard, G., Helman, G., Sherbini, O., Boeck, R., Cohn, J., Collins, A., Demarest, S., Dobbins, K., Emrick, L., et al. (2020). Randomized clinical trial of first-line genome sequencing in pediatric white matter disorders. Ann. Neurol. *88*, 264–273. https://doi.org/10.1002/ana.25757.
38. Terry, S.F., and Taft, R. (2021). iHope genetic health: enabling genomic medicine across the globe. Genet. Test. Mol. Biomarkers *25*, 733–734. https://doi.org/10.1089/gtmb.2021.0287.
39. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., 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*, 405–424. https://doi.org/10.1038/gim.2015.30.
40. Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F., and South, S.T.; Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee (2011). American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet. Med. *13*, 680–685. https://doi.org/10.1097/GIM.0b013e3182217a3a.
41. Riggs, E.R., Andersen, E.F., Cherry, A.M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D.I., South, S.T., Thorland, E.C., et al. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet. Med. *22*, 245–257. https://doi.org/10.1038/s41436-019-0686-8.
42. Brandt, T., Sack, L.M., Arjona, D., Tan, D., Mei, H., Cui, H., Gao, H., Bean, L.J.H., Ankala, A., Del Gaudio, D., et al. (2020). Adapting ACMG/AMP sequence variant classification guidelines for single-gene copy number variants. Genet. Med. *22*, 336–344. https://doi.org/10.1038/s41436-019-0655-2.
43. Thaxton, C., Goldstein, J., DiStefano, M., Wallace, K., Witmer, P.D., Haendel, M.A., Hamosh, A., Rehm, H.L., and Berg, J.S. (2022). Lumping versus splitting: how to approach defining a disease to enable accurate genomic curation. Cell Genom. *2*, 100131. https://doi.org/10.1016/j.xgen.2022.100131.
44. ClinGen Standard Gene-Disease Relationship Recuration Procedure(s) https://clinicalgenome.org/site/assets/files/2164/clingen_standard_gene-disease_validity_recuration_procedures_v1.pdf.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
ICSL curated gene-disease validity classifications	This paper	https://search.thegencc.org/submitters/GENCC_000105
ClinGen published gene-disease validity classifications	ClinGen	https://search.clinicalgenome.org/kb/gene-validity
BabySeq Project gene-disease validity classifications	Ceyhan-Birsoy et al.[29]	N/A
GeneMatcher	Sobreira et al.[24]	https://genematcher.org/
Software and algorithms		
Excel	Microsoft	N/A
Other		
clinical WGS data	Illumina Clinical Services laboratory	The gene curations conducted for this study were initiated from clinical WGS data that are available from Illumina Clinical Services Laboratory upon request, but restrictions apply to the availability of these sequence data because they contain sensitive information that could compromise patient privacy/consent and so are not publicly available.

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Alison Coffey (acoffey@illumina.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Gene-disease validity classifications are publicly available as of the date of publication through the GenCC (https://thegencc.org/) and can be accessed by Submitter (Illumina): https://search.thegencc.org/submitters/GENCC_000105. A subset of GDRs was also shared directly with ClinGen (https://search.clinicalgenome.org/); these curations are reviewed in accordance with the standard procedures of the relevant GCEP (primarily the Syndromic Disorders GCEP) before being published by the GCEP. GDRs published through ClinGen GCEPs do not always acknowledge individual contributors. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Data analysis was completed under an institutional review board research exemption (ID Number: ICSL-001). Informed consent was not required since no human subjects were enrolled; all study samples were de-identified residual specimens leftover from routine clinical care.

METHOD DETAILS

Genome sequencing, analysis, and interpretation

Samples from individuals with a suspected RUGD are received from clinical and research partners across the globe,[2,5,37] including those served by the iHope program, a philanthropic initiative to enable access to genome sequencing to underserved families with children facing rare and undiagnosed genetic diseases.[2,38] Among the 1,033 cases within our cohort, 67% were received as trios, 13% as duos, 13% as proband only, and 7% as quads or higher order family structures. Interpretation and reporting are conducted within the context of the proband's phenotype, and a clinical report is returned to the ordering physician. This test is designed to detect and report on single nucleotide variants (SNVs), small insertion and deletion events, copy number variants (CNVs),

homozygous loss of *SMN1*, mitochondrial SNVs, and a set of short tandem repeat expansions with known associations with genetic disorders. cGS is performed to a minimum coverage of 40X.

Variants are filtered and prioritized for manual review (or triage) using a combination of phenotype- and genotype-driven approaches based on multiple factors, including population allele frequency, variant consequence, evolutionary conservation, *in silico* predictions, occurrence in a gene whose associated disease or function overlaps the proband's reported phenotype, and inheritance, as appropriate. Variants that pass filtering are triaged by the case analyst (range ~50–200 depending on family structure) to identify candidate variants for further investigation based on their potential clinical significance for the proband. For most cases, only a small number of variants are flagged as candidates (typically 1–2 per case). Candidate SNVs and small insertion and deletion events are curated and classified according to the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) guidelines[39] with additional guidance, including that from the ClinGen Sequence Variant Interpretation Working Group and Variant Curation Expert Panels. Candidate CNVs are curated and classified according to the ACMG/ClinGen standards for the interpretation and reporting of CNVs current at the time of reporting.[40,41] Additional guidelines by Brandt et al. (2020) are used when classifying CNVs that affect a single gene.[42]

Gene curation and clinical reporting

Gene curations are conducted by the case analyst upon identification of a candidate SNV, small insertion or deletion, or CNV affecting a single gene if the GDR of interest has not been previously classified as Strong or Definitive by the BabySeq Project, ClinGen, or internally. GDRs are evaluated using the ClinGen framework for Gene-Disease Clinical Validity.[9] Genetic and experimental evidence is gathered from the public domain and scored according to the current ClinGen Standard Operating Procedures (SOP). Genetic evidence (maximum of 12 pts) includes variant and case reports, segregation data, and case-control studies, though the latter are rarely available for RUGD. Experimental evidence (maximum of 6 pts) includes biochemical function, expression, protein interaction, functional alteration studies in patient and non-patient cells, non-human organism and cell culture models, and genetic rescue. For genes that have been linked to more than one disease phenotype, ClinGen's lumping and splitting guidelines[43] are used to define the disease entity for curation.

Curators are clinical genomics scientists and genetic counselors who are already well-versed in the principles of evaluating evidence linking genomic variants to disease. They are further trained in the evaluation of gene-disease validity by a small working group of scientists who specialize in gene curation. In our experience, the skills underlying variant curation and interpretation (review of clinical literature, evaluation of functional studies) are transferable to gene-level analysis. During training, curators typically complete practice curations under close supervision by the gene curation working group, who provide feedback. In addition, before a GDR classification is used in variant classification or reporting decisions, the curation is presented by the curator at one of a series of daily meetings where members of the working group confirm completeness, review scoring of the evidence, and approve the final classification. Fully trained curators often volunteer as ClinGen biocurators and participate in ClinGen gene curation expert panels.

Curators aim to document the breadth of evidence available. Reactive gene curation is performed as part of active case analysis within the deadlines imposed by clinical reporting. As such, all available evidence may not be assessed, particularly for GDRs that reach a Strong or Definitive classification. When available, curators score the maximum data attributable to genetic evidence and include non-human animal model data and evidence across at least two of three experimental evidence categories: function, functional alteration, and models and rescue. If the available genetic evidence does not reach the maximum, experimental evidence is typically curated until a Strong/Definitive classification or the maximum possible points is reached.

Curated GDRs are stored in an internal knowledgebase and shared with the community through ClinGen and the Gene Curation Coalition (GenCC).

Re-evaluation of curated GDRs

A subset of GDRs were recurated (Table S4) in accordance with ClinGen's recommended time frame,[44] using the SOP current at the time of recuration. Recuration includes a literature search for, and evaluation of, new evidence as well as review of scoring of previously curated evidence.

Data analysis

The disease entities curated for RUGD are commonly complex disorders affecting multiple organs or body systems and include conditions with multiple congenital anomalies or structural birth defects which span many clinical domains, necessitating curation across a diversity in presentation. The primary clinical features of each curated disease entity were determined through review of published literature and Online Mendelian Inheritance in Man (OMIM). Disorders characterized primarily by dysfunction of a single affected body system were assigned one of 23 top-level ancestor terms within the phenotypic abnormality subontology of the Human Phenotype Ontology (Table S5 and Figure S1). Disorders characterized by significant dysfunction of two or more body systems were described as Multi-system.

ICSL's dataset of genes curated for gene-disease validity was compared with ClinGen's publicly available curations. Genes curated by both were then reviewed to determine whether the curations were completed for the same disease entity. Disease entities lacking an exact name match were confirmed to be functionally equivalent through review of PMIDs used in the curation.

Cases for sequencing are received from probands with multiple family structures. When tallying the numbers of variants reported across the sample cohort for inclusion in this study, a variant identified and reported in multiple affected family members was counted only once. Variant tallies include those related to the indication for testing as well as incidental findings. Variants identified as secondary findings were excluded from tallies.