Clinical validity assessment of genes frequently tested on intellectual disability/autism sequencing panels

Erin Rooney Riggs, MS[1], Taylor I. Bingaman, BS[1], Carrie-Ann Barry, BS[2], Andrea Behlmann, PhD[3], Krista Bluske, PhD[4], Bret Bostwick, MD[5], Alison Bright, PhD[6], Chun-An Chen, PhD[5], Amanda R. Clause, PhD[4], Avinash V. Dharmadhikari, MS, PhD[7, 8], Mythily Ganapathi, PhD[9], Claudia Gonzaga-Jauregui, PhD[10], Andrew R. Grant, BA[11, 12], Madeline Y. Hughes, BA[13], Se Rin Kim, BA[14], Amanda Krause, MD, PhD[15], Jun Liao, PhD[9], Aimé Lumaka, MD, PhD[16], Michelle Mah, MSc[17], Caitlin M. Maloney, BS[18], Shruthi Mohan, PhD[19], Ikeoluwa A. Osei-Owusu, PhD[11], Emma Reble, MSc[20], Olivia Rennie, BS[21], Juliann M. Savatt, MS[1], Hermela Shimelis, PhD[1], Rebecca K. Siegert, BSN[11], Tam P. Sneddon, PhD[22], Courtney Thaxton, PhD[22], Kelly A. Toner, MA[2], Kien Trung Tran, PhD[23], Ryan Webb, BS[11], Emma H. Wilcox, BA[24], Jiani Yin, PhD[25], Xinming Zhuo, PhD[26], Masa Znidarsic, MD[27], Christa Lese Martin, PhD[1], Catalina Betancur, MD, PhD[28], Jacob A.S. Vorstman, MD, PhD[21], David T. Miller, MD, PhD*[29], Christian P. Schaaf, MD, PhD*[5, 30]

Affiliations

1. Autism & Developmental Medicine Institute, Geisinger, Danville, PA, USA
2. Drexel University College of Medicine, Philadelphia, PA, USA
3. Invitae, San Francisco, CA, USA
4. Illumina, Inc., San Diego, CA, USA
5. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
6. Natera, San Carlos, CA, USA
7. Children's Hospital of Los Angeles, Los Angeles, CA, USA
8. Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
9. Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY, USA

10. Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Juriquilla, Querétaro, Mexico
11. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
12. New York Medical College, Valhalla, NY, USA
13. University of Illinois, Chicago, Chicago, IL, USA
14. National Human Genome Research Institute, Bethesda, MD, USA
15. Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, The University of the Witwatersrand, Johannesburg, South Africa
16. Laboratoire de Génétique Humaine, University of Liège, Liège, Belgium
17. Trillium Health Partners, Mississauga, Ontario, Canada
18. University of Washington, Seattle, WA, USA
19. University of North Carolina, Chapel Hill, North Carolina, USA
20. St. Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada
21. Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada
22. Department of Pathology and Laboratory Medicine, The University of North Carolina, Chapel Hill, NC, USA
23. Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia
24. The Warren Alpert Medical School of Brown University, Providence, RI, USA
25. Department of Neurology, University of California Los Angeles, Los Angeles, CA, USA
26. Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
27. University Medical Center, Ljubljana, Slovenia
28. Neuroscience Paris Seine, Institut de Biologie Paris Seine, Sorbonne Université, INSERM, CNRS, Paris, France

29. Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA

30. Institute of Human Genetics, Heidelberg University, Heidelberg, Germany

* denotes co-senior author

Corresponding Author

Erin Rooney Riggs, MS, CGC

Autism & Developmental Medicine Institute

Geisinger

100 N Academy Avenue

MC 60-36

Danville, PA 17822

570-522-9427

eriggs@geisinger.edu

Conflict of Interest Notification

1. Dr. Alison Bright is a shareholder of and employed by Natera. Dr. Bright has also been an employee of Invitae and Quest Diagnostics commercial laboratories.
2. Dr. Amanda Clause and Dr. Krista Bluske are shareholders of and employed by Illumina, Inc.
3. Dr. Andrea Behlmann is a shareholder of and is employed by Invitae.
4. Dr. Bret Bostwick has received research support from Biomarin Pharmaceuticals. He is currently employed by and is a shareholder of Alnylam Pharmaceuticals.
5. All other authors declare no conflict of interest.

Abstract

Purpose: Neurodevelopmental disorders (NDDs), such as intellectual disability (ID) and autism spectrum disorder (ASD), exhibit genetic and phenotypic heterogeneity, making them difficult to differentiate without a molecular diagnosis. The ClinGen ID/Autism Gene Curation Expert Panel (GCEP) utilizes systematic curation to distinguish ID/ASD genes that are appropriate for clinical testing (i.e., with substantial evidence supporting their relationship to disease) from those that are not.

Methods: Using the ClinGen gene-disease validity curation framework, the ID/Autism GCEP classified genes frequently included on clinical ID/ASD testing panels as Definitive, Strong, Moderate, Limited, Disputed, Refuted, or No Known Disease Relationship.

Results: As of September 2021, 156 gene-disease pairs have been evaluated. Though most (75%) were determined to have definitive roles in NDDs, 22 (14%) genes evaluated had either Limited or Disputed evidence. Such genes are currently not recommended for use in clinical testing due to the limited ability to assess the impact of identified variants.

Conclusion: Our understanding of gene-disease relationships evolves over time; new relationships are discovered, and previously-held conclusions may be questioned. Without periodic re-examination, inaccurate gene-disease claims may be perpetuated. The ID/Autism GCEP will continue to evaluate these claims to improve diagnosis and clinical care for NDDs.

Introduction

Neurodevelopmental disorders (NDDs) represent a spectrum of disease manifestations affecting normal brain development and daily functioning, and are often attributable to a genetic etiology. NDDs may include, but are not limited to, global developmental delay (GDD), intellectual disability (ID), autism spectrum disorder (ASD), and epilepsy, with or without additional features such as dysmorphic features or other congenital anomalies. While genomic variation is hypothesized to play a role in most NDDs, currently up to 50% of cases have a genetic etiology identifiable by current molecular testing methodologies.[1,2] In clinical practice, genetic testing to identify an etiology in individual patients is important for providing a genetic diagnosis for the family to inform prognostic information, potential treatment strategies, and family planning.[3-5] Multiple neurodevelopmental phenotypes can be associated with the same genomic variant and can even present in the same individual.[6] There can also be variable expressivity and incomplete penetrance, even within families.[7] In addition, some NDDs are part of broader, syndromic presentations; the additional features present in these syndromes may be recognized by experienced clinicians, which may result in a more straightforward diagnostic course. However, many NDDs often have widely variable, overlapping features and exhibit a great deal of genetic heterogeneity, making it difficult to distinguish among them clinically without the aid of molecular diagnostics.[1] Diagnostic yields for genetic testing related to NDDs, such as ID and ASD, have increased as Next Generation Sequencing (NGS) technologies have enabled multiple genes to be sequenced on a single platform, paving the way for large multi-gene testing panels and exome or genome sequencing. Recent studies have shown that exome sequencing has increased diagnostic yield for NDDs to approximately 36%.[1] In the case of gene panels for ID and ASD, the list of genes varies considerably between laboratories, partly due to lack of consensus about what constitutes an established gene-disease relationship.[8,9] Similarly, for exome and genome sequencing, there is a need to identify genes known to be related to particular disorders to inform variant classification and clinical

interpretation. Phenotype-based filtering strategies help clinical laboratories narrow the focus to genes/variants most likely to be of clinical relevance. As such, differentiating genes with substantial evidence supporting their role in NDDs from those with limited or disputed evidence becomes critical. Including genes of uncertain significance (GUS) in clinical testing pipelines can lead to difficulty with variant interpretation and ambiguous test results, delaying patient diagnosis. GUS cannot be assessed using the 2015 ACMG/AMP sequence variant interpretation guidelines[10], and therefore variants identified in these genes should be classified as variants of uncertain significance (VUS). Therefore, it can be challenging for both patients and clinicians to receive or report a VUS from clinical genetic testing because of its ambiguous nature.[11]

Many genes have *bona fide* evidence supporting their relationship with disease, while others can be described as candidate genes at best. In an effort to help laboratories and clinicians distinguish and stratify the evidence level of genomic data, the Clinical Genome Resource (ClinGen), an NIH-funded initiative dedicated to determining clinically relevant genes and variants for application in medical and research fields[12], has developed a framework to systematically assess the evidence supporting or refuting gene-disease relationships.[13] ClinGen's Gene Curation Expert Panels (GCEPs) apply this evidence-based framework in order to evaluate the clinical validity of gene-disease relationships within various clinical domains, including ID and ASD.

The mission of the ClinGen Intellectual Disability/Autism Gene Curation Expert Panel (ID/Autism GCEP) (https://clinicalgenome.org/affiliation/40006/) is to provide the community with systematic, evidence-based generated data for NDD gene-disease relationships, specifically ID and ASD. Our initial goal was to evaluate the evidence supporting genes currently included on clinical genetic testing panels marketed for ID and/or ASD. We have since expanded our scope to include those genes newly associated with ID and/or ASD that may be appropriate for inclusion on testing panels, as well as genes that have previously been implicated in disease for which current evidence suggests such claims may be disputed

or refuted. The ID/Autism GCEP anticipates that these publicly available, evidence-based curations will provide laboratories with valuable guidance when determining which genes to include or remove from diagnostic panels and guide filters for exome/genome analysis, resulting in increased consistency for clinical care across laboratories.

Materials and Methods

Identifying Relevant Genes

We queried the Genetic Testing Registry (GTR)[14] in 2017 and again in 2019 for any multi-gene, NGS panels designed with ID and/or ASD as one of the conditions for which the test was offered. The 2019 query returned 65 different panels and 4,962 unique genes. The working group coordinator (ER) and co-chairs (DM, CS) manually reviewed these results in order to identify those panels for which ID and/or ASD was the primary indication for testing. We excluded panels that were too broad in scope (e.g., tests designed to identify disorders affecting infants in the neonatal intensive care unit, or tests designed to target all neurological disorders, including neuromuscular and neurodegenerative disorders). Following this review process, we identified 30 NGS panels primarily focused on ID and/or ASD including 972 unique genes. A total of 498 of these genes (51%) appeared only on a single panel, and are not included in the analysis presented in this paper. We initially opted to curate genes in descending order of frequency, starting with those most frequently included on ID/ASD testing panels. Over time, we also incorporated genes submitted by ID/Autism GCEP committee members and ClinGen users, including those with newly described relationships with ID and/or ASD, as well as those with relationships that were thought to be disputed. In total, 156 gene-disease pairs were curated as of September 30, 2021. Of note, some genes included on this initial list have presentations that overlap with the scope of other ClinGen GCEPs (Epilepsy GCEP, Inborn Errors of Metabolism GCEP, etc.); in some of these cases, the

primary curation was completed by the other GCEP, and the ID/Autism GCEP was listed as a secondary contributor.

Determining Disease Entities (Precuration)

Before beginning curation, curators reviewed the Online Mendelian Inheritance in Man (OMIM)[15] and Orphanet[16] databases to determine the disease entity(ies) purported to be related to the gene under evaluation. If a gene was associated with more than one disease entity in OMIM and/or Orphanet, the ClinGen precuration process (https://tinyurl.com/lumpandsplit) was followed to determine whether the disease entities should be lumped into a single, all-encompassing term or split into separate curations (Supplemental Figure 1). Briefly, factors such as mode of inheritance, disease mechanism, and inter-/intra-familial variability were considered when opting to lump or split previously asserted disease entities. In general, if the conditions in question could not be distinguished at the molecular level (e.g., the same variant(s) results in different presentations within and between families), they were lumped; if there were some distinguishing characteristics (e.g., loss-of-function variants result in one presentation, gain-of-function variants result in a different presentation; both autosomal dominant and recessive inheritance patterns were observed, etc.), they were split. Following review of the precuration information, the GCEP selected an appropriate disease term from the Monarch Disease Ontology (Mondo)[17,18], an ontology that harmonizes multiple disease resources, to represent the disease entity being curated. Of note, Mondo is the disease ontology used for all ClinGen gene-disease validity curations; OMIM terms (when available) are provided within this manuscript as a point of reference along with the Mondo term used for curation.

In accordance with the ClinGen precuration guidelines, the ID/Autism GCEP opted not to perform separate curations for different neurodevelopmental presentations observed within a given gene (e.g., ID, ASD, seizures). There were no cases in which these presentations could be reliably distinguished from

one another at the variant level (e.g., where, for a given gene, truncating variants would always and only result in ID, while missense variants would always and only result in ASD)[19], therefore they were lumped. If a well-accepted disease term was available in these circumstances (particularly if other consistent, non-neurodevelopmental features were also present), this term was used (e.g., Sotos syndrome (MONDO:0019349, MIM:117550), Phelan-McDermid syndrome (MONDO:0011652, MIM:606232), Cornelia de Lange syndrome (MONDO:0016033, MIM:122470)). Otherwise, a more general term was used (e.g., complex neurodevelopmental disorder (MONDO:0100038), syndromic ID (MONDO:0000508), non-syndromic X-linked ID (MONDO:0019181)); these general terms do not have corresponding MIM numbers.

Curation and Expert Review

Once the genes and disease terms were identified, evidence supporting or refuting each gene-disease pair was gathered and evaluated in accordance with the ClinGen gene-disease validity curation process[13], and classifications were assigned. GCEP curators performed the preliminary evidence evaluation and classification according to the Standard Operating Procedures (SOP) document versions 5-8 (depending on the date) (https://tinyurl.com/genediseasesop). If the curator of a well-established gene-disease pair was able to reach Definitive with no major questions or concerns, they submitted the results of the curation to the experts to review and approve via email. For all other genes, the curator presented the results of their preliminary curation to the entire GCEP during twice monthly teleconferences for review. Following expert approval, all curations are made publicly available through the ClinGen website (https://search.clinicalgenome.org/kb/gene-validity/). A listing of all current curations completed by the ID/Autism GCEP (updated in real time) can be accessed at the following URL: https://search.clinicalgenome.org/kb/affiliate/10006.

Re-curation

ClinGen's procedure for re-curation was developed to periodically reassess gene-disease relationships since new supporting or conflicting evidence may emerge over time. The intervals recommended for recuration differ based upon the initial classification (https://tinyurl.com/yc3sfr7k). For example, Definitive gene-disease relationships are reassessed on an as-needed basis (when/if additional information becomes available), while Limited gene-disease relationships should be reassessed every three years. If a classification changes as the result of a re-evaluation, an updated report is published to the ClinGen website (www.clinicalgenome.org). Each gene-disease record receives an updated "date last evaluated" on the ClinGen website, regardless of whether or not the final classification has changed.

Results

The ClinGen clinical validity framework utilizes both genetic and experimental evidence to quantitatively analyze the strength of evidence for a gene-disease relationship.[13] The classifications Definitive, Strong, Moderate, and Limited are used if there is evidence found to support the gene-disease relationship. The classifications No Known Disease Relationship, and Disputed or Refuted indicate no relationship or conflicting evidence, respectively. As of September 30, 2021, 156 gene-disease pairs were evaluated for gene-disease validity by the ID/Autism GCEP and are publicly available on the ClinGen website. Individual genes evaluated are noted in Table 1; a full listing of each gene along with its condition, mode of inheritance, and evaluation scores is provided in Supplemental Table 1.

Gene-Disease Relationships with Sufficient Supporting Evidence

Of the 156 gene-disease pairs evaluated by the ID/Autism GCEP, 117 (75%) were classified as Definitive using the ClinGen gene-disease validity framework (Figure 1). Examples of Definitive gene-disease

relationships include *CREBBP*/Rubinstein-Taybi syndrome (MONDO:0019188, MIM:180849), *RAI1*/Smith-Magenis syndrome (MONDO:0008434, MIM:182290), *NIPBL*/Cornelia de Lange syndrome (MONDO:0016033, MIM:122470), and *ANK2*/complex neurodevelopmental disorder (MONDO:0100038) (see Supplemental Table 1 for complete list). Within the ClinGen gene-disease validity framework[13], the difference between Definitive and Strong is replication over time; there were no gene-disease relationships classified as Strong, as each of these 117 gene-disease pairs had at least three years pass since the initial report and multiple observations from independent sources.

Seventeen gene-disease pairs (11%) were classified as having "Moderate" evidence to support their causative role in disease (see Supplemental Material for complete list). In general, genes with a classification of Moderate are considered appropriate for inclusion on multi-gene testing panels or for exome/genome analysis pipelines.[20] However, these gene-disease pairs need additional genetic and/or gene-level experimental evidence to reach Strong or Definitive. The ClinGen ID/Autism GCEP plans to re-evaluate Moderate gene-disease pairs every two years from the date of last review to investigate whether enough new evidence has emerged to update their classifications (see section on re-curation below).

Gene-Disease Relationships with Little Supporting Evidence

Among the 156 gene-disease pairs, three (2%) had Limited evidence to support the gene's role in disease (*NTNG1*/complex neurodevelopmental disorder (MONDO:0100038), *CACNG2*/complex neurodevelopmental disorder (MONDO:0100038), and *LAS1L*/X-linked syndromic ID (MONDO:0020119)) (Table 2). The three Limited genes (*NTNG1*, *CACNG2,* and *LAS1L*) each have extremely limited genetic evidence supporting their relationships with disease (three or fewer probands meeting our thresholds for scoring); however, no significant contradictory evidence was identified to dispute or refute these claims. While the current evidence is sparse, we have not yet ruled out the possibility that variation in

these genes could cause these conditions, and it is possible that additional evidence could bolster these claims. Given the limited knowledge available, interpreting variation identified in these genes during the course of clinical testing would be difficult; such genes should be considered GUS and should generally not be included on diagnostic gene sequencing analyses per guidance from the American College of Medical Genetics and Genomics.[20]

Conflicting Evidence: Disputed and Refuted Gene-Disease Relationships

The gene-disease validity classifications of Disputed and Refuted are reserved for those gene-disease pairs with conflicting evidence reported since the time of initial association between a gene and disease. Gene-disease pairs with a classification of Disputed have conflicting evidence, but this information is not necessarily convincing enough to negate the possibility of the gene's role in the disease. Gene-disease pairs with a classification of Refuted have conflicting evidence that significantly outweighs any supporting evidence, or evidence against a gene's role in disease. As of September 2021, no genes have been classified as Refuted by the ID/Autism GCEP; however, 19 (12%) gene-disease pairs were considered Disputed (Table 2).

Many of these genes were initially implicated in disease years prior to the widespread availability of population variation resources, such as gnomAD.[21] Often, variation in the gene would be identified based on limited evaluation of the initial proband(s) (e.g., screening a small number of candidate genes within a linkage region, or screening a cohort of individuals for variants in a gene that was identified as possibly being disrupted in a translocation case). Once proposed in the literature, others would search for variation within the genes amongst their cohorts, often perpetuating claims of a gene-disease relationship. When evaluated utilizing current data and the ClinGen framework, most of these variants were discounted due to their high frequency in the general population, inheritance from a reportedly unaffected parent, non-segregation within the family, and/or the presence of other disease-causing

variant(s) identified in the proband. For those genes in which variation was observed frequently enough to warrant case-control studies (e.g., *CNTNAP2* [heterozygous variants], *RELN* [heterozygous variants], *SLC6A4*, *EN2, MET*), the studies either did not demonstrate difference in variation rates between cases and controls, or initial positive findings in association studies of common variants with small sample sizes could not be replicated in larger datasets. In scenarios like these where previously published genetic evidence had been ruled out, experimental evidence was typically scored at 0 points; the group felt that, without a clear link to human disease, it was difficult to assess how well any experimental evidence correlated with said disease. As with the Limited genes described above, these genes should also be considered GUS[20] and not included in diagnostic testing for ID and/or ASD. Note, however, that some genes that were determined not to be involved in autosomal dominant disorders, such as *CNTNAP2, RELN,* and *LAMC3*, are involved in recessive disorders associated with ID/ASD and, as such, should be included in diagnostic testing for those disease relationships.

Re-curation

There were 15 genes re-curated by the ID/Autism GCEP during the course of our initial analyses, including *ZNF292*. *ZNF292* was first reported in relation to autosomal dominant complex neurodevelopmental disorder (MONDO:0100038) in 2012.[22,23] It was originally curated by the ID/Autism GCEP in 2018 and was found at the time to have Limited evidence to support this gene-disease relationship. The disease mechanism at the time was unclear, and other sources cataloging gene-disease relationships (OMIM, Orphanet) had no documented disease relationships for this gene. In 2020, the ID/Autism GCEP received a request for re-curation from GenomeConnect[24,25], ClinGen's online patient registry; a participant enrolled with a variant in *ZNF292* reported as a "candidate gene," and the registry could identify that newer information had become available since the last evaluation. As a result, the ID/Autism GCEP agreed to re-curate this gene outside of the typical timeline for Limited gene-disease

classifications. With the addition of the information from the Mirzaa et al. publication[26], this gene-disease classification reached Definitive, and the GenomeConnect participant was ultimately issued an updated report from the laboratory.

Discussion

The ClinGen ID/Autism GCEP was established in order to provide standardized assessments of the level of evidence available to support purported relationships between specific genes and diseases involving ID and/or ASD. Here we present the results of our first 156 gene-disease evaluations; the results of our assessments are made publicly available immediately after review through the ClinGen website (https://clinicalgenome.org/). It is important to note that, due to the syndromic nature of many NDDs, some NDD genes may be curated by other ClinGen GCEPs (e.g., Epilepsy, Brain Malformations, Inborn Errors of Metabolism, etc.). Our ultimate goal across ClinGen is to provide such assessments for all genes suspected of being involved in NDDs, in order to clearly distinguish between genes with sufficient evidence to warrant evaluation during clinical genetic testing from those without. In this initial set of evaluations, we applied a rigorous quantitative approach and identified several gene-disease pairs currently being included on clinical testing panels that lack the evidence necessary to solidify their role in NDDs. We hope this information will be taken into consideration as laboratories update their test offerings; current guidance suggests that only those gene-disease pairs with classifications of Moderate or above should be included in clinical testing.[20]

There are multiple groups now engaged in the process of curating gene-disease relationships. The ClinGen gene-disease curation process differs from other general and ID/ASD-specific initiatives that have previously been used to identify genes involved in disease, which could explain why this process has identified genes on established panels lacking solid evidence (n=22) (Figure 2). Both OMIM and Orphanet are commonly used general resources that catalog gene-disease relationships. While some

evidence supporting these claims is often included in the descriptions of genes/diseases available through these sites, there is no formal effort to quantify the strength of said evidence (or lack thereof); as a result, there is no simple way to distinguish between those gene-disease pairs with little supporting evidence from those with substantial supporting evidence. This poses difficulties for laboratories trying to make decisions regarding which genes to include on diagnostic panels, or which gene/phenotype relationships to examine as part of exome or genome sequencing. Including genes with gene-disease validity classifications of Limited, Disputed, or Refuted can result in an increased number of variants being classified as "uncertain," or variants being classified as "likely pathogenic" or "pathogenic" inappropriately per ACMG technical standards.[20] Other general curation efforts, such as Genomics England PanelApp (https://panelapp.genomicsengland.co.uk/)[27], PanelApp Australia (https://panelapp.agha.umccr.org/)[28], the Transforming Genomic Medicine Initiative's (TGMI) Gene2Phenotype (G2P) (https://www.ebi.ac.uk/gene2phenotype)[29], etc. are also working to evaluate gene-disease validity. Users of these resources should note that each resource has different evaluation metrics, different approaches to evidence aggregation, and even different terminology to describe the results. In an effort to harmonize this information for the genomics community, ClinGen has partnered with these and other organizations to form the Gene Curation Coalition (GenCC) (https://thegencc.org/), a resource that aims to facilitate the consistent assessment of gene-disease relationships. ClinGen submits all of its gene-disease validity assessments to the GenCC database so that the community may evaluate them in the context of assessments from other submitters.

The ClinGen ID/Autism GCEP is serving an unmet need for the NDD gene curation community. Within the ID/ASD field specifically, there are several ongoing efforts focused on identifying genes associated with specific neurodevelopmental presentations.[30] Resources such as VariCarta (https://varicarta.msl.ubc.ca/index)[31] and denovo-db (https://denovo-db.gs.washington.edu/denovo-db/)[32] catalog variants reported in the literature; the former focuses on variants reported in individuals

diagnosed with ASD, while the latter documents *de novo* variants in the broader NDDs and other disorders. Neither of these resources provide assessments of the validity of the genes' relationships with ASD or other NDDs. Resources such as the Geisinger Developmental Brain Disorders (DBD) Database (https://dbd.geisingeradmi.org/)[33] and the Simons Foundation Autism Research Initiative (SFARI) Gene (https://gene.sfari.org/)[34] do provide such assessments. Geisinger DBD characterizes genes as either "high confidence" or "emerging" candidate genes using a tiered system based on the number of truncating variants identified across various NDDs; SFARI Gene uses the number of *de novo*, likely-gene-disrupting variants observed in individuals with ASD to assign a numeric score signifying the group's confidence in the gene's role in ASD. SFARI includes a fourth category, "syndromic," to denote those genes associated with presentations beyond the characteristics required for an ASD diagnosis.

The information from these resources provides a useful snapshot of evidence that may be available in the literature, and often serve as a starting point for identifying relevant literature for ClinGen curations. Simply cataloging variants reported in the literature is highly valuable in and of itself, as this information is often buried in supplemental material and not easily discoverable utilizing conventional methods of searching. However, classification of genes based solely on counts of variants reported in the literature does not always provide a complete or accurate view of the role of that gene in disease, and may not account for other essential parameters, such as the gene's constraint for truncating and missense variation, the frequency of the variant in control populations, the segregation of the variant with the phenotype, the mode of inheritance (dominant, recessive, or X-linked), and the biological sex of the probands in X-linked disorders. ClinGen's comprehensive approach takes these variables into consideration to provide a more robust assessment of gene-disease validity for use in clinical applications.

One clear limitation to such a comprehensive approach is the amount of time it takes to review each gene. On average, the ID/Autism GCEP completes approximately 4 gene-disease validity assessments

per month; at this pace, it would take approximately 6.6 years to evaluate the 318 remaining genes from our initial list included on more than one clinical testing panel. However, by embracing collaborative community approaches, such as GenCC (described above), it is possible that the gene-disease validity community as a whole (not limited to a single effort) can provide evaluations of these genes in a shorter timeframe. Distributed effort can also allow all groups to also monitor the needs for recuration as new information arises instead of focusing solely on providing new evaluations.

Our understanding of gene-disease relationships is evolving over time; as new data become available, it can either illuminate previously undiscovered gene-disease relationships, or cause us to question previously-held conclusions. If we do not periodically re-examine these relationships, we run the risk of perpetuating inaccurate gene-disease relationships in the literature, in clinical testing, and in patient care. For example, a gene may have been reported as a putative cause of ID/ASD as the result of single-gene sequencing studies years ago, and "verified" against a control set of a few hundred individuals. Because of this, such a gene could have been included on a clinical testing panel, or referenced as a disease gene in subsequent publications. A new laboratory trying to develop an ID/ASD testing panel may do so by incorporating the genes tested on other panels and/or searching the literature for reported ID/ASD gene-disease associations, leading to the example gene being included on additional panels. Meanwhile, new information becomes available (e.g., population databases like gnomAD) that changes our perspective on the initial information. This new information might reveal that variants in this example gene that were initially deemed "disease-causing" are common in the general population, calling into question the gene-disease relationship. This information is publicly available, but if not reported in some way to the community, either through literature or curation efforts such as ClinGen, it is possible that inaccurate information is still getting disseminated and potentially impacting patient results. The ClinGen ID/Autism GCEP serves as a resource to continuously evaluate these claims and

make them publicly available in the hopes of ultimately improving clinical testing and care for individuals with NDDs.

Data Availability

The data set supporting the current study is included as a supplemental table. The ClinGen ID/Autism GCEP makes all curations publicly available on the ClinGen website (https://search.clinicalgenome.org/kb/gene-validity/).

Acknowledgements

This work was supported by the National Human Genome Research Institute of the National Institutes of Health (NIH) under award number: U24HG006834. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author Information

Funding Acquisition, Conceptualization, Methodology: E.R.R., C.L.M.; Formal Analysis, Supervision, Validation, Writing – original draft: E.R.R., T.I.B., C.B., J.A.S.V, D.T.M, C.P.S.; Project Administration, Visualization: E.R.R., T.I.B.; Data Curation, Investigation, Writing – review & editing: E.R.R., T.I.B., C.A.B., A.B., K.B., B.B., A.R.B., C.A.C., A.R.C., A.V.D., M.G., C.G.J., A.R.G., M.Y.H., S.R.K., A.K., J.L., A.L., M.M., C.M.M., S.M., I.A.O.O., E.R., O.R., J.M.S., H.S., R.K.S., T.P.S., C.T., K.A.T., K.T.T., R.W., E.H.W., J.Y., X.Z., M.Z., C.L.M., C.B., J.A.S.V., D.T.M., C.P.S.

Ethics Declaration

Data used to make the curation decisions described in this manuscript comes from review of previously published literature. As this work did not involve any patient interaction, no consent was needed.

Figure Legends

Figure 1. Clinical validity classifications of the 156 gene-disease pairs evaluated as of September 2021.

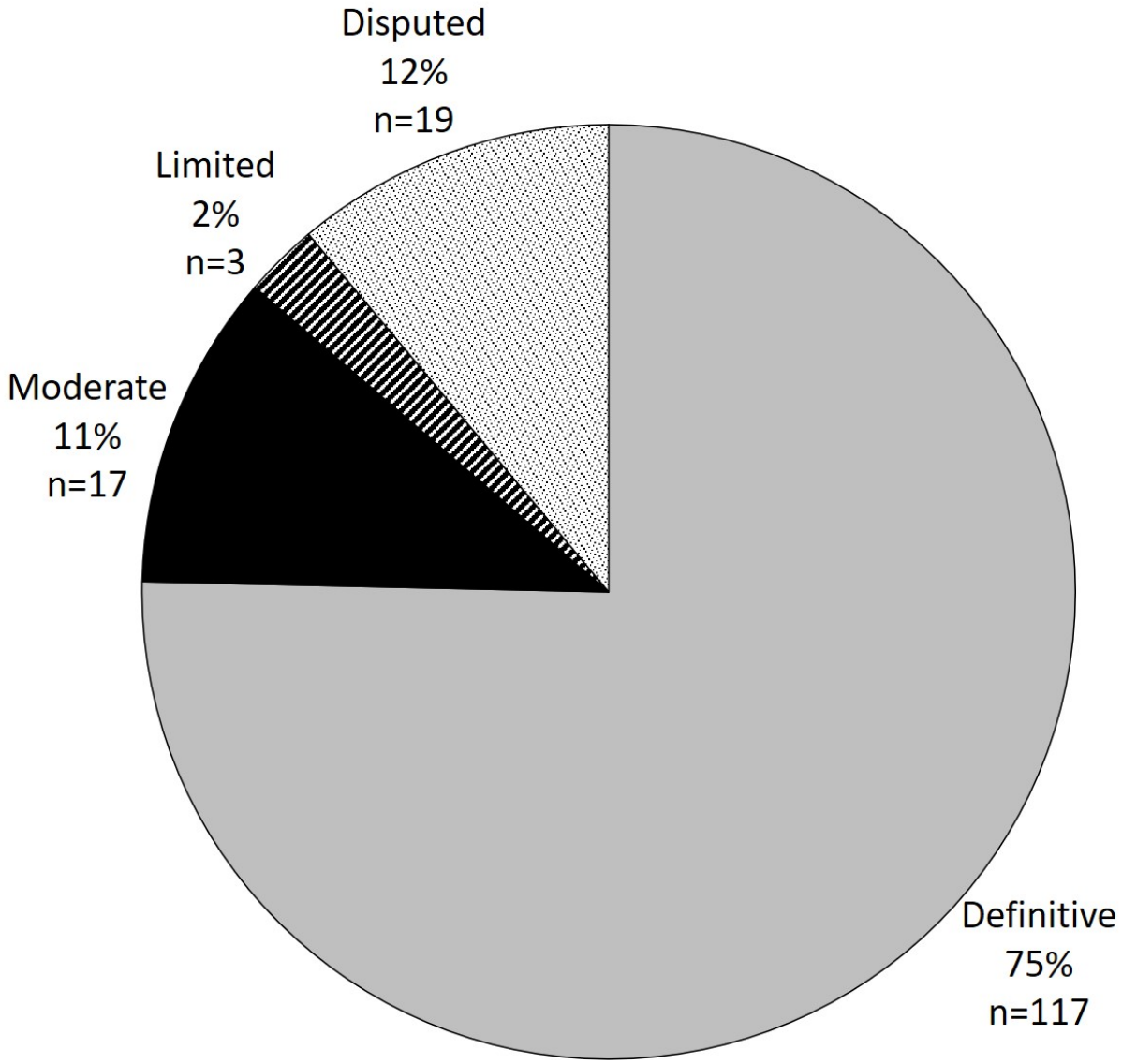
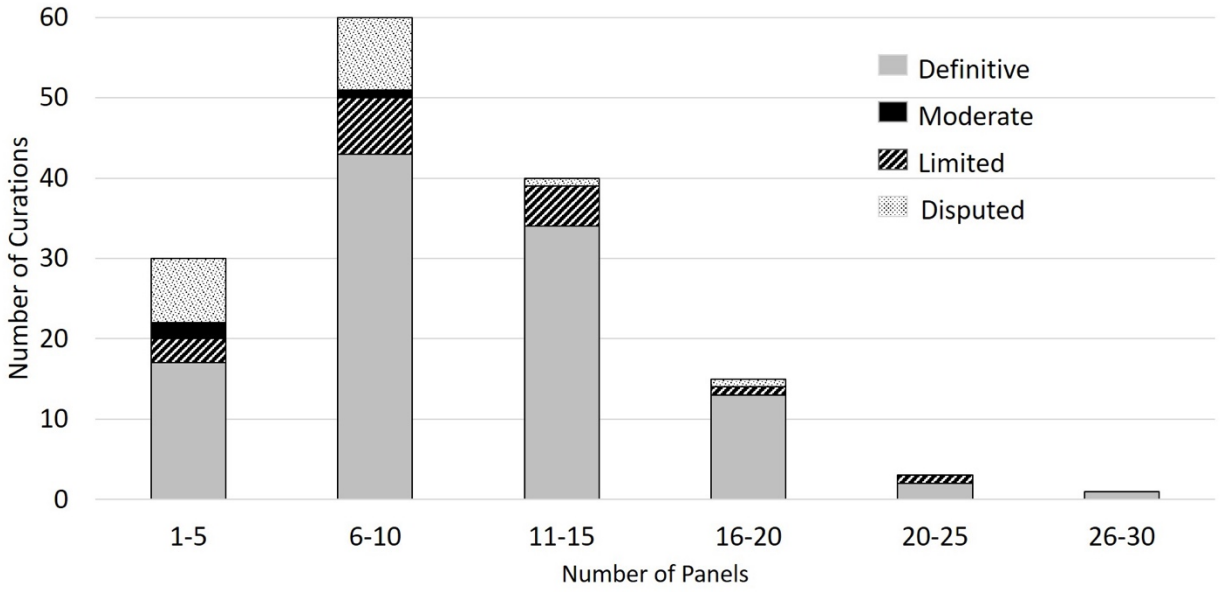


Figure 2. Curated gene-disease pairs plotted according to the number of clinical genetic testing panels on which they appear. Number of panels was obtained by querying the Genetic Testing Registry (GTR) in September 2019 for any multi-gene next-generation sequencing panel marketed for ID and/or ASD.



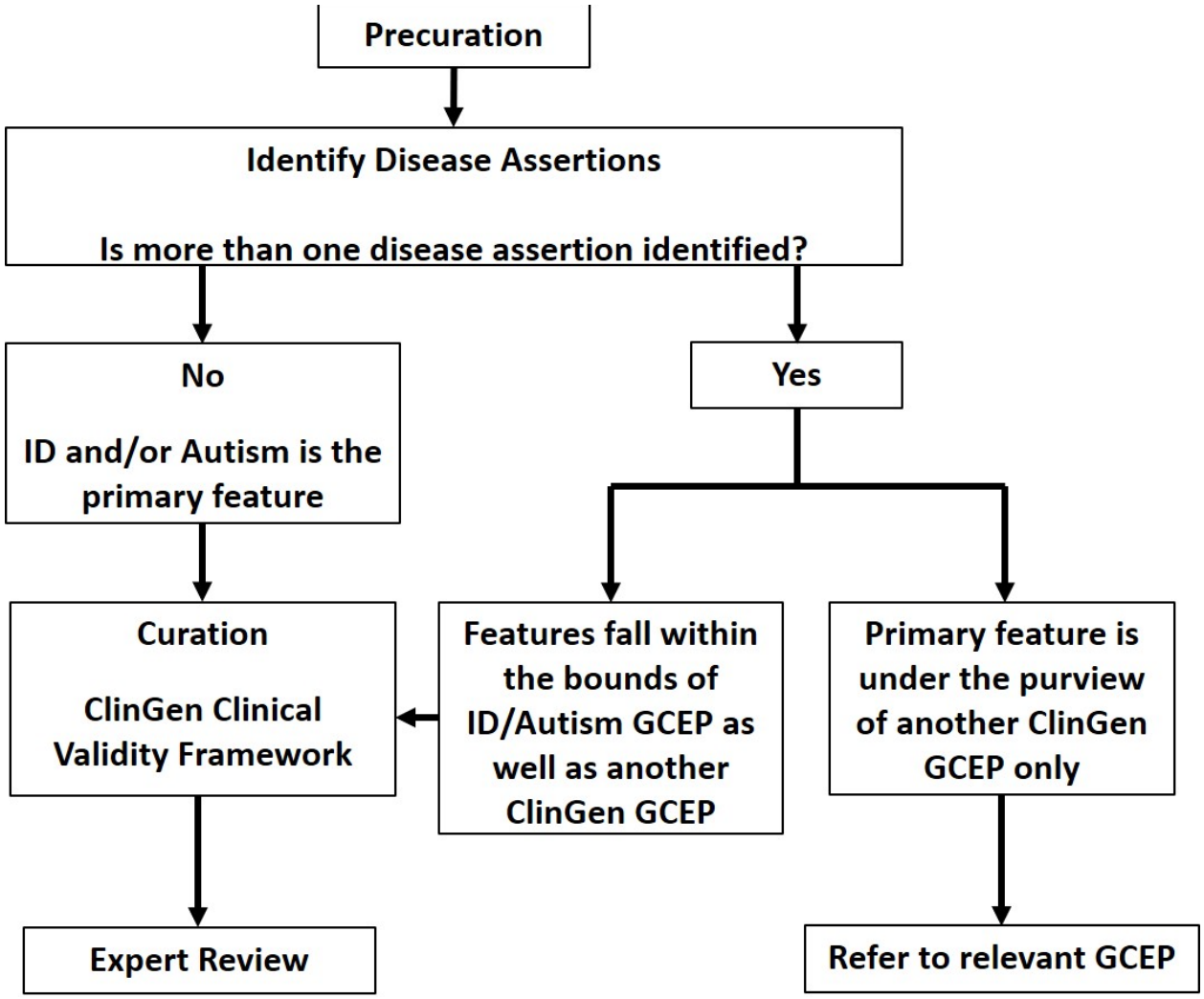

Precuration
Identify Disease Assertions
Is more than one disease assertion identified?
No
ID and/or Autism is the primary feature
Yes
Curation
ClinGen Clinical Validity Framework
Features fall within the bounds of ID/Autism GCEP as well as another ClinGen GCEP
Primary feature is under the purview of another ClinGen GCEP only
Expert Review
Refer to relevant GCEP

Table 1. Genes evaluated by the ClinGen ID/Autism GCEP as of September 2021 organized by classification. Please see Supplemental Table 1 for full detail on the conditions/modes of inheritance.

Classifications	Genes (n=156)
Definitive (n=117)	*ADNP, ADSL, AFF2, ALDH5A1, ANK2, ANKRD11, AP1S2, AP4B1, AP4E1, AP4M1, ARID1A, ARID1B, ARX, ASXL1, ASXL2, ATP6AP2, ATP7A, ATRX, AUTS2, BCL11A, BRSD2, BRWD3, CASK, CC2D1A, CHD8, CLCN4, CNKSR2, CRADD, CREBBP, CTCF, CTNNB1, CUL3, CUL4B, DDX3X, DHCR7, DKC1, DLG3, DYRK1A, EHMT1, FGD1, FLNA, FMR1, FOLR1, FOXP1, FOXP2, GNAI1, GPC3, GRIA3, HCFC1, HDAC8, HOXA1, HPRT1, HUWE1, IDS, IL1RAPL1, IQSEC2, KDM5C, KIF1A, L1CAM, MAN1B1, MAOA, MBTPS2, MED12, MED13L, MEF2C, MID1, MYT1L, NAA10, NBEA, NDP, NEXMIF, NHS, NIPBL, NLGN4X, NR4A2, NRXN1, NSD1, OCRL, OFD1, PACS1, PAK3, PHF6, PHF8, PLP1, POGZ, PORCN, PQBP1, PTCHD1, RAB39B, RAD21, RAI1, RPS6KA3, SATB2, SETBP1, SETBP1, SHANK2, SHANK3, SLC16A2, SLC2A1, SMARCA2, SMARCA4, SMC1A, SMC3, SMS, TAOK1, TBL1XR1, TBR1, TNRC6B, TRAPPC9, TUSC3, UBE2A, UPF3B, VPS13B, ZC4H2, ZDHHC9, ZEB2, ZNF292*
Moderate (n=17)	*ACSL4, ANK3, ARHGEF9, CRBN, FTSJ1, GDI1, MBD5, MED23, NLGN3, NSDHL, NSUN2, RPL10, ST3GAL3, SYN1, SYP, TSPAN7, ZNF711*
Limited (n=3)	*CACNG2, LAS1L, NTNG1*
Disputed (n=19)	*AGTR2, ARHGEF6, CDH15, CLIC2, CNTNAP2, DPP6, EN2, IGBP1, KATNAL2, LAMC3, MET, RELN, SLC6A4, SLC9A9, SHROOM4, ZDHHC15, ZNF41, ZNF674, ZNF81*

Table 2. Detailed listing of all Limited and Disputed curations as of September 2021. Note that some genes that were determined not to be involved in autosomal dominant disorders, such as *CNTNAP2, RELN,* and *LAMC3,* are involved in recessive disorders associated with ID/ASD. Abbreviations: SOP (Standard Operating Procedures), GTR (Genetic Testing Registry)

Gene	Disease Name	MONDO ID	Mode of Inheritance	Classification	SOP Version	Date Last Evaluated	Number of Panels in GTR (2019)	Genetic Evidence Points	Experimental Evidence Points	Total Points
CACNG2	Complex Neurodevelopmental Disorder	0100038	Autosomal Dominant	Limited	8	7/29/2021	4	1.5	0	1.5
NTNG1	Complex Neurodevelopmental Disorder	0100038	Autosomal Dominant	Limited	8	2/2/2021	6	0.1	2.5	2.6
LAS1L	X-Linked Syndromic Intellectual Disability	0020119	X-Linked	Limited	8	9/21/2021	4	1.7	0	1.7
CDH15	Intellectual Disability	0001071	Autosomal Dominant	Disputed	8	2/17/2021	5	0	0	0
CNTNAP2	Complex Neurodevelopmental Disorder	0100038	Autosomal Dominant	Disputed	8	3/16/2021	19	0	0	0
DPP6	Complex Neurodevelopmental Disorder	0100038	Autosomal Dominant	Disputed	8	5/5/2021	5	0	0	0
EN2	Complex Neurodevelopmental Disorder	0100038	Autosomal Dominant	Disputed	8	2/16/2021	5	0	0	0
KATNAL2	Complex Neurodevelopmental Disorder	0100038	Autosomal Dominant	Disputed	7	5/20/2020	7	0	0	0
LAMC3	Complex Neurodevelopmental Disorder	0100038	Autosomal Dominant	Disputed	7	9/1/2020	6	0	0	0
MET	Complex Neurodevelopmental Disorder	0100038	Autosomal Dominant	Disputed	8	1/19/2021	4	0	0	0

RELN	Complex Neurodevelopmental Disorder	0100038	Autosomal Dominant	Disputed	8	3/17/2021	11	0	0	0
SLC6A4	Autism Spectrum Disorder	0005258	Autosomal Dominant	Disputed	8	1/6/2021	7	0	0	0
SLC9A9	Autism Spectrum Disorder	0005258	Autosomal Dominant	Disputed	8	10/8/2020	8	0	0	0
AGTR2	X-Linked Complex Neurodevelopmental Disorder	0100148	X-Linked	Disputed	7	6/2/2020	7	0	0	0
ARHGEF6	Non-Syndromic X-Linked Intellectual Disability	0019181	X-Linked	Disputed	8	10/20/2020	8	0	0	0
CLIC2	X-Linked Complex Neurodevelopmental Disorder	0100148	X-Linked	Disputed	8	2/16/2021	5	0	0	0
IGBP1	Corpus Callosum Agenesis – Intellectual Disability – Coloboma – Micrognathia Syndrome	0010333	X-Linked	Disputed	8	2/2/2021	6	0	0	0
SHROOM4	X-Linked Complex Neurodevelopmental Disorder	0100148	X-Linked	Disputed	8	1/13/2021	9	0	0	0
ZDHHC15	Complex Neurodevelopmental Disorder	0100038	X-Linked	Disputed	8	7/30/2020	3	0	0	0
ZNF41	Non-Syndromic X-Linked Intellectual Disability	0019181	X-Linked	Disputed	7	3/16/2021	6	0	0	0
ZNF674	X-Linked Intellectual Disability	0100284	X-Linked	Disputed	8	5/4/2021	5	0	0	0
ZNF81	X-Linked Intellectual Disability	0019181	X-Linked	Disputed	8	1/26/2021	5	0	0	0

References

1. Clark MM, Stark Z, Farnaes L, et al. 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.* 2018;3:16-8. eCollection 2018. doi:10.1038/s41525-018-0053-8.

2. Srivastava S, Love-Nichols JA, Dies KA, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med.* 2019;21(11):2413-2421. doi:10.1038/s41436-019-0554-6.

3. Bilkey GA, Burns BL, Coles EP, et al. Genomic Testing for Human Health and Disease Across the Life Cycle: Applications and Ethical, Legal, and Social Challenges. *Front Public Health.* 2019;7:40. doi:10.3389/fpubh.2019.00040.

4. Hayeems RZ, Babul-Hirji R, Hoang N, Weksberg R, Shuman C. Parents' Experience with Pediatric Microarray: Transferrable Lessons in the Era of Genomic Counseling. *J Genet Couns.* 2016;25(2):298-304. doi:10.1007/s10897-015-9871-3.

5. Iglesias A, Anyane-Yeboa K, Wynn J, et al. The usefulness of whole-exome sequencing in routine clinical practice. *Genet Med.* 2014;16(12):922-931. doi:10.1038/gim.2014.58.

6. Srivastava S, Cohen JS, Vernon H, et al. Clinical whole exome sequencing in child neurology practice. *Ann Neurol.* 2014;76(4):473-483. doi:10.1002/ana.24251.

7. Finucane B, Challman TD, Martin CL, Ledbetter DH. Shift happens: family background influences clinical variability in genetic neurodevelopmental disorders. *Genet Med.* 2016;18(4):302-304. doi:10.1038/gim.2015.92.

8. Erger F, Schaaf CP, Netzer C. Which genes to assess in the NGS diagnostics of intellectual disability? The case for a consensus database-driven and expert-curated approach. *Mol Cell Probes.* 2019;45:84-88. doi:S0890-8508(18)30328-1.

9. Hoang N, Buchanan JA, Scherer SW. Heterogeneity in clinical sequencing tests marketed for autism spectrum disorders. *NPJ Genom Med.* 2018;3:27-3. eCollection 2018. doi:10.1038/s41525-018-0066-3.

10. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-424. doi:10.1038/gim.2015.30.

11. Mahon SM. Management of Patients with a Genetic Variant of Unknown Significance. *Oncol Nurs Forum.* 2015;42(3):316-318. doi:10.1188/15.ONF.316-318.

12. Rehm HL, Berg JS, Brooks LD, et al. ClinGen--the Clinical Genome Resource. *N Engl J Med.* 2015;372(23):2235-2242. doi:10.1056/NEJMsr1406261.

13. Strande NT, Riggs ER, Buchanan AH, et al. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J Hum Genet.* 2017;100(6):895-906. doi:S0002-9297(17)30160-X.

14. Rubinstein WS, Maglott DR, Lee JM, et al. The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. *Nucleic Acids Res.* 2013;41(Database issue):925. doi:10.1093/nar/gks1173.

15. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). World Wide Web URL: https://omim.org/.

16. Orphanet: an online database of rare diseases and orphan drugs. Copyright, INSERM 1997. Available at http://www.orpha.net. Accessed 30 December 2021.

17. Shefchek KA, Harris NL, Gargano M, et al. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.* 2020;48(D1):D704-D715. doi:10.1093/nar/gkz997.

18. Mungall CJ, McMurry JA, Kohler S, et al. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.* 2017;45(D1):D712-D722. doi:10.1093/nar/gkw1128.

19. Myers SM, Challman TD, Bernier R, et al. Insufficient Evidence for "Autism-Specific" Genes. *Am J Hum Genet.* 2020;106(5):587-595. doi:S0002-9297(20)30113-0.

20. Bean LJH, Funke B, Carlston CM, et al. Diagnostic gene sequencing panels: from design to report-a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med.* 2020;22(3):453-461. doi:10.1038/s41436-019-0666-z.

21. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7.

22. de Ligt J, Willemsen MH, van Bon BW, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med.* 2012;367(20):1921-1929. doi:10.1056/NEJMoa1206524.

23. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature.* 2012;485(7397):242-245. doi:10.1038/nature11011.

24. Savatt JM, Azzariti DR, Faucett WA, et al. ClinGen's GenomeConnect registry enables patient-centered data sharing. *Hum Mutat.* 2018;39(11):1668-1676. doi:10.1002/humu.23633.

25. Kirkpatrick BE, Riggs ER, Azzariti DR, et al. GenomeConnect: matchmaking between patients, clinical laboratories, and researchers to improve genomic knowledge. *Hum Mutat.* 2015;36(10):974-978. doi:10.1002/humu.22838.

26. Mirzaa GM, Chong JX, Piton A, et al. De novo and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder. *Genet Med.* 2020;22(3):538-546. doi:10.1038/s41436-019-0693-9.

27. Martin AR, Williams E, Foulger RE A Genomics England, London, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet.* 2019;51(11):1560-1565. doi:10.1038/s41588-019-0528-2.

28. Stark Z, Foulger RE, Williams E, et al. Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. *Am J Hum Genet.* 2021;108(9):1551-1557. doi:S0002-9297(21)00268-8.

29. Thormann A, Halachev M, McLaren W, et al. Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat Commun.* 2019;10(1):2373-3. doi:10.1038/s41467-019-10016-3.

30. Schaaf CP, Betancur C, Yuen RKC, et al. A framework for an evidence-based gene list relevant to autism spectrum disorder. *Nat Rev Genet.* 2020;21(6):367-376. doi:10.1038/s41576-020-0231-2.

31. Belmadani M, Jacobson M, Holmes N, et al. VariCarta: A Comprehensive Database of Harmonized Genomic Variants Found in Autism Spectrum Disorder Sequencing Studies. *Autism Res.* 2019;12(12):1728-1736. doi:10.1002/aur.2236.

32. Turner TN, Yi Q, Krumm N, et al. denovo-db: a Compendium of Human De Novo Variants. *Nucleic Acids Res.* 2017;45(D1):D804-D811. doi:10.1093/nar/gkw865.

33. Gonzalez-Mantilla AJ, Moreno-De-Luca A, Ledbetter DH, Martin CL. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *JAMA Psychiatry.* 2016;73(3):275-283. doi:10.1001/jamapsychiatry.2015.2692.

34. Abrahams BS, Arking DE, Campbell DB, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism.* 2013;4(1):36-36. doi:10.1186/2040-2392-4-36.