Title

Clinical variants paired with phenotype: A rich resource for brain gene curation.

Permalink

Journal

Authors

Chopra, Maya

Savatt, Juliann

Bingaman, Taylor

et al.

Publication Date

DOI

Peer reviewed

Clinical variants paired with phenotype: A rich resource for brain gene curation

Maya Chopra[1,2,*], **Juliann M. Savatt**[3], **Taylor I. Bingaman**[3], **Molly E. Good**[3], **Alexis Morgan**[3], **Caitlin Cooney**[3], **Allison M. Rossel**[3], **Bryanna VanHoute**[3], **Ineke Cordova**[3], **Sonal Mahida**[2], **Virginia Lanzotti**[4], **Dustin Baldridge**[4], **Christina A. Gurnett**[4], **Joseph Piven**[5], **Heather Hazlett**[5], **Scott L. Pomeroy**[2], **Mustafa Sahin**[1,2], **Philip R.O. Payne**[6], **Erin Rooney Riggs**[3], **John N. Constantino**[7],

behalf of the Brain Gene Registry Consortium

[1]Rosamund Stone Zander Translational Neuroscience Center, Boston Children's Hospital; Boston, MA

[2]Department of Neurology, Boston Children's Hospital Intellectual Disability and Research Center; Harvard Medical School; Boston, MA

[3]Autism and Developmental Medicine Institute, Geisinger, Danville, PA

[4]Washington University School of Medicine Intellectual and Developmental Disability Research Center, St. Louis, MO

[5]Department of Psychiatry, University of North Carolina Intellectual and Developmental Disability Research Center, Chapel Hill, NC

[6]Institute for Informatics Washington University in St. Louis, St. Louis, MO

[7]Division of Behavioural and Mental Health, Children's Healthcare of Atlanta, Departments of Psychiatry and Paediatrics, Emory University, Atlanta, GA

*Correspondence and requests for materials should be addressed to Maya Chopra, RSZ Translational Neuroscience Center BCH3468, Boston Children's Hospital, 300 Longwood Ave., Boston, MA 02115. maya.chopra@children.harvard.edu.

Abstract

Purpose: Clinically ascertained variants are under-utilized in neurodevelopmental disorder research. We established the Brain Gene Registry (BGR) to coregister clinically identified variants in putative brain genes with participant phenotypes. Here, we report 179 genetic variants in the first 179 BGR registrants and analyze the proportion that were novel to ClinVar at the time of entry and those that were absent in other disease databases.

Methods: From 10 academically affiliated institutions, 179 individuals with 179 variants were enrolled into the BGR. Variants were cross-referenced for previous presence in ClinVar and for presence in 6 other genetic databases.

Results: Of 179 variants in 76 genes, 76 (42.5%) were novel to ClinVar, and 62 (34.6%) were absent from all databases analyzed. Of the 103 variants present in ClinVar, 37 (35.9%) were uncertain (ClinVar aggregate classification of variant of uncertain significance or conflicting classifications). For 5 variants, the aggregate ClinVar classification was inconsistent with the interpretation from the BGR site-provided classification.

Conclusion: A significant proportion of clinical variants that are novel or uncertain are not shared, limiting the evidence base for new gene-disease relationships. Registration of paired clinical genetic test results with phenotype has the potential to advance knowledge of the relationships between genes and neurodevelopmental disorders.

Keywords

Introduction

Despite being readily accessible, clinically ascertained genetic variants are under-utilized in neurodevelopmental disorders (NDDs) research. Existing approaches to assess a gene's role in disease (gene-disease validity curation) rely heavily on published reports of rare genetic variants.[1] This context is prone to bias, either toward a particular phenotypic category or toward severity. Variants identified by sequencing a large cohort of individuals from a study focused on autism, for example, may not assess for other phenotypic features, may exclude mildly affected individuals, and often do not include standardized neurobehavioral evaluations. Furthermore, case reports beyond the initial publication asserting a gene's relationship to disease are often not considered novel and thus unlikely to be published, limiting the evidence available to confirm gene-disease relationships and/or biasing the description of the phenotypic spectrum of a condition.

Databases such as ClinVar[2,3] that make variants and their classifications publicly available are important resources for the genetics community seeking information on the role of rare variants in disease.[4] Although laboratories are encouraged to share variants with ClinVar, workflows and protocols are laboratory-dependent, and not all of them share data routinely. Given the public nature of the database and aggregate nature of laboratory submissions, laboratories that do submit often restrict the amount of the accompanying

individual-level details (eg, phenotypic information). If phenotypic information is submitted, it is often limited to that which was listed on the test requisition form, which may be inaccurate or incomplete.[2] Furthermore, there is no systematic method by which clinicians and/or consented individuals can be recontacted by those interested in a particular variant for phenotype verification, family segregation, or ascertainment of co-occurring variants. Together, these factors contribute to the under-utilization of genomic and health data generated via clinical testing in NDD research.

The scientific implications of the siloed nature of clinically ascertained data are significant. The limited pace of the accrual of variant-level evidence supporting a gene's role in disease is a major contributor to the increasing burden of reports with variants of uncertain significance (VUS) classifications, which account for up to 86% of genetic test results[5] and now outnumber pathogenic (P) findings.[6] Harnessing variants acquired through clinical care across medical specialties and pairing them with phenotype has the potential to contribute to unbiased understanding of the association between rare genetic variants and neurodevelopmental profiles of patients. A streamlined process to pair publicly available variants with clinical phenotype and familial segregation information would accelerate variant interpretation by laboratories, clinicians, and researchers. Increased understanding of a gene's role in disease and better delineation of phenotypic spectrum has the potential to accelerate genetic diagnoses, resolve uncertainty of genomic variants (thus reducing VUS reports and the associated burden to families and the health care system), improve counseling and establish a resource for the development of gene-based therapeutics.

The Eunice Kennedy Shriver Intellectual and Developmental Disability Research Centers (IDDRC) are an established network of researchers funded by the National Institutes of Child Health and Human Development that collaborate to expand basic and translational research to better understand the causes of intellectual and developmental disabilities and to develop effective therapies for these disorders.[7] The mission of the IDDRC-CTSA (Clinical Translational Science Awards) Brain Gene Registry (BGR) is to accelerate and enrich the systematic evaluation of putative "brain genes" (ie, genes implicated in neurodevelopment with varying degrees of evidence) by establishing a data repository with paired genomic and phenotypic data. In this collaborative initiative, individuals with clinically reported variants in any gene implicated in neurodevelopment are enrolled through 1 of the 10 recruiting IDDRC sites. The recruitment and study flow are shown in Figure 1. Eligibility requires that the variant be reported by a clinical laboratory with CLIA (Clinical Laboratory Improvement Amendments) certification and be classified as VUS, likely pathogenic (LP) or pathogenic (P) according to professional[8] guidelines (Figure 1). Eligibility is determined by variant only, regardless of indication for testing and regardless of phenotype. Variant data are recorded along with electronic health record (EHR)-derived phenotypic data.

Given the importance of a standardized assessment in obtaining a precise neurobehavioral phenotype, participants also undergo a rapid virtual neurobehavioral assessment (RNAP) with established cognitive and behavioral tools (Figure 1) and newly designed dysmorphology and neurology assessment tools. The RNAP consists of questionnaires and behavioral surveys completed by the parent/ caregiver or consented adult subject along with a telehealth assessment that can be completed by a registered nurse (RN) or masters'

level clinician. The questionnaires are selected based on age, are completed electronically, typically take 60 to 90 minutes, and can be completed in stages. The telehealth component typically takes 60 minutes. The results are scored electronically, and participants can be given brief feedback upon request. The assessments were selected on the basis of (1) brevity and feasibility for busy clinical settings, (2) the availability of established methods for acquiring valid, standardized phenotypic information from patients and families, and (3) broad clinical applicability, requiring no higher degree of specialty than a trained RN or master's level clinician. The RNAP systematically characterizes key domains of function relevant to comprehensive phenotypic characterization of NDDs, as well as subclinical NDD traits in the domains of cognition, adaptive functioning, motor/sensory, autism symptoms, psychiatric symptoms, physical characteristics, and neurologic concerns. This constellation of characteristics is rarely documented in the EHR.

In parallel to BGR enrollment, participants are required to register in GenomeConnect,[9,10] the Clinical Genome Resource (ClinGen)[11] patient registry; coenrollment in GenomeConnect ensures BGR participants' genomic and health survey data are submitted to ClinVar and made publicly accessible, and that participants may elect to receive updates about their results. All data (variant, EHR, RNAP, and GenomeConnect) collected via the BGR are registered in CIELO (Collaborative Informatics Environment for Learning on Health Outcomes),[12] a custom data commons platform developed and hosted by the Institute for Informatics at Washington University in St. Louis. Data may be used as case-level evidence for intellectual disability/autism (ID/autism) gene-disease validity curation under the ClinGen[11] framework, as well as for translational research efforts.

In the present study, we evaluated 179 variants from the first 179 individuals coenrolled in the BGR and GenomeConnect to determine the proportion that were novel to ClinVar at the time of submission, and the proportion that were absent from an additional 6 disease databases.

Materials and Methods

Gene selection

Although individuals with a clinically identified variant (VUS and above) in any brain gene are eligible to participate, priority is given to those with a variant in a BGR gene of interest. The BGR genes of interest list is a regularly updated set of putative brain genes selected from the published literature and from a list of approximately 900 genes on commercial ID/autism genetic testing panels queried via the Genetic Testing Registry in collaboration with ClinGen[11,13] and IDDRC investigators.

At the time of writing, there were 377 genes of interest (https://braingeneregistry.wustl.edu/genes-of-interest/). Genes that have previously been assigned a "Moderate" or "Limited" level of evidence for gene-disease validity by the ClinGen ID/ Autism gene curation expert panel (GCEP) are of the highest priority. These evidence levels encompass genes with some genetic and experimental evidence to support a relationship to disease, but additional data may increase confidence in this relationship. Genes with "Definitive" status are included on a case-by-case basis; additional cases may inform the neurodevelopmental phenotype and/or

improve understanding of genotype-phenotype relationships. Genes with "Disputed" or "Refuted" status are not included. Brief definitions of gene-disease validity evidence levels prioritized for BGR gene selection are shown in Table 1. The purpose of this prioritization approach is to select those genes which are most likely to shift classifications with additional case-level data. Results are continuously reviewed such that new genes with "Moderate" or "Limited" classifications are added periodically.

Genes that have never been curated but that are commercially tested on ID/autism panels are also included. Inclusion of these genes is based on neurodevelopmental phenotypic assertion in the literature, greater number of panels in the Genetic Testing Registry, and higher proportion of VUS classifications in ClinVar. Quantitative metrics or numeric cutoffs for these categories are not used. Rather, these characteristics are evaluated in aggregate by the BGR team to determine those genes most likely to benefit from additional case-level data from the BGR.

Although individuals with multi-gene deletions and duplications are not precluded from participating, recruitment is currently intentionally focused on individuals with variants affecting a single gene, in order to enrich the data set with case-level evidence that can be used for gene-disease validity curation.

Participant identification

Ten participating IDDRC sites - Baylor College of Medicine (BCM), University of Washington (UW), University of North Carolina (UNC), Albert Einstein College of Medicine, Children's National Hospital, Washington University (WUSTL), Waisman Center of the University of Wisconsin-Madison (UW-Madison), the MIND Institute of the University of California Davis (UC Davis), Children's Hospital of Philadelphia, and Kennedy Krieger Institute (KKI) carried out searches for individuals with clinically identified variants with American College of Medical Genetics and Genomics/ Association for Molecular Pathology (ACMG/AMP) classification[8] of VUS, LP, or P in the BGR genes of interest. Search methods varied between sites and are described in detail below. Two sites (UW and KKI) used 2 search methodologies.

Laboratory database search—BCM: BCM searched for BGR-eligible variants in a series of queries for selected genes through the database of Baylor Genetics Laboratory, by interrogating CLIA reports for VUS, LP, and P variants reported since 2012. From this search, only patients local to Texas Children's Hospital were eligible to participate.

EHR queries—UW, University of North Carolina (entire Hospital system), Albert Einstein College of Medicine, WUSTL, UW-Madison, Children's Hospital of Philadelphia, KKI, and Children's National Hospital: custom EHR queries based on gene name for the preceding 5 years were performed, followed by manual review of patient EHRs to confirm presence of an eligible variant in a CLIA report.

UW, UC Davis, and KKI: Internal databases were used to search all clinically ordered genetic tests for the preceding 5, 10, and 12 years, respectively. The databases were

manually screened and corresponding EHR records were manually reviewed to confirm presence of an eligible variant in a CLIA report.

Recruitment

After informed consent, participants coenrolled into the BGR (in which genetic test results were coregistered with phenotypic data from the RNAP and the EHR), and GenomeConnect (in which genetic results and health survey data were collected). Within GenomeConnect, participants are asked to complete a general health survey that collects health information across 17 body systems. Additional, optional surveys are assigned based on participant initial responses and can be completed as participants are able. These surveys are intended to supplement the data collected via the EHR and RNAP. Variants are collected from participant reports by GenomeConnect team members using a structured survey tool (https://clinicalgenome.org/docs/genomeconnect-genetic-report-review-standard-operatingprocedure) and submitted to ClinVar with case-level phenotypic and segregation data. The reporting laboratory name, that laboratory's reported classification, and the classification date are submitted.[9] Variants are not independently classified by the GenomeConnect team.

Assessment of variant presence in databases

Variants were manually assessed from March to April 2023 to determine whether they were present or absent at the time of GenomeConnect's submission to ClinVar. In addition to ClinVar, 6 other rare disease databases, listed and described in Table 2, were interrogated for presence or absence of the variants.

In instances when the variant was previously present in ClinVar, the "aggregate ClinVar classification" at the time of submission of the BGR variant was noted. Although ClinVar does not curate submissions, the aggregate classification from submitters is provided for each variant (https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/clinsig_agg). If multiple classifications are present and they differ by major classification category (VUS, P/LP, and benign/likely benign [B/LB]), the aggregate classification is listed as "conflicting interpretations of pathogenicity." Importantly, because the BGR variants are not independently classified by the registry team, they are not included in the aggregate ClinVar classification.[9]

When the laboratory classification on the BGR participants' genetic report was the same as the aggregate ClinVar classification, this was noted as "consistent with aggregate ClinVar classification." Conversely, if the laboratory classification differed from the aggregate ClinVar classification by major category, this was noted as "inconsistent with aggregate ClinVar classification." Finally, if the laboratory report classification differed from other previous ClinVar submission(s) from the same laboratory for the same variant, this was noted as "intralaboratory discrepancy." These individuals meet the threshold for recontact according to the GenomeConnect protocol and are alerted to contact their clinician for further discussion of any updated genetic information and, if indicated, for reevaluation.[9,14] Interlaboratory discrepancies (eg, the classification from a participant's reporting lab is in conflict with the classification from a different lab) are not currently shared with participants

because their variant's classification may still represent their reporting laboratory's current classification. However, these discrepancies may be resolved through the ClinGen Sequence Variant Inter-Laboratory Discrepancy Resolution group and relevant variant curation expert panels, and prompt future participant updates.[15]

Results

At the time of this analysis, there were 179 individuals with 179 variants in 76 genes coenrolled in the BGR and GenomeConnect. The unique variant count is listed in Supplemental Table 1; variants are also available on the BGR ClinVar submission page (https://www.ncbi.nlm.nih.gov/clinvar/submitters/508359/). There were 7 variants identified in multiple participants, 3 related parent-child pairs with VUS, and 4 unrelated pairs of participants with P variants. Additionally, there were 7 participants with a variant in more than 1 brain gene. The BGR participantprovided classification was VUS for 59 variants, LP for 35 variants, and P for 82 variants.

The presence or absence of variants in ClinVar at time of submission and other reference databases is shown in Table 2. Overall, 76 (42.5%) variants (listed in Supplemental Table 1 with their classifications) were absent from ClinVar, and 62 (34.6%) variants were absent from all analyzed disease databases, highlighting the value of the BGR in contributing novel genomic data. Such data can be vital to assessing and establishing gene-disease validity and can also aid clinical laboratories in their classification of variants.

Of the 103 (57.5%) variants already present in ClinVar, as shown in Figure 2, 28 (26.9%) had an aggregate ClinVar classification of VUS and 11 (9.6%) had an aggregate ClinVar classification of "conflicting interpretation of pathogenicity." Expert review of paired phenotypic and genotypic data has the potential to resolve uncertainty and shift these types of variants to more definitive classifications, demonstrating the utility of sharing additional observations of variants already in the public sphere in addition to those that are novel. (Figure 2).

Recognizing that some variants may have been classified and reported to participants before incorporation of relevant guidelines, the availability of population databases, or availability of other relevant variant data,[8,16] the BGR participant-provided report classifications were compared with the ClinVar aggregate classifications (Figure 3). Of the 92 variants present in ClinVar with an aggregate classification that did not fall under "conflicting interpretation of pathogenicity," 5 (5.4%) had an aggregate ClinVar classification that was inconsistent with the BGR participant-provided classification (Figure 3A). Further, an analysis was undertaken to assess for "intralaboratory discrepancies" (Figure 3B). There were 9 variants for which the BGR participant-provided classification was discrepant with the current ClinVar classification from the same reporting laboratory. For 6 variants, this was confidence discrepancy (for example, a difference between LP and P), whereas for 3, the discrepancy was between the three major categories (VUS, P/ LP, and B/LB). These included 3 variants reported as VUS on the BGR participant report; 2 of them had ClinVar classifications of LB from the reporting laboratory, and the third had a ClinVar classification of P from the reporting laboratory. In all 3 instances of a major category discrepancy and in 3 confidence

discrepancies, the participant's report date preceded the laboratory's last evaluation date of the variant, suggesting that the variant had been reclassified after the participant's testing. These participants may or may not be aware of this updated information and meet the threshold for recontact according to the previously published GenomeConnect protocol for providing variant classification updates.[9,14] In the other 3 confidence discrepancies, the participants report date was more recent than the laboratory's ClinVar submission highlighting the registry's ability to contribute more recent variant classifications to ClinVar.

Discussion

In the IDDRC-CTSA BGR initiative, rare, clinically identified genetic variants in putative brain genes are coregistered with standardized neurodevelopmental phenotype and EHR data into an extensible, cloud-based data common (CIELO).[12] Of note, the CIELO environment provides the ability to submit, link, index, and then track access and analysis of data sets, thus enabling multi-site data intensive research. Parallel enrollment into ClinGen's GenomeConnect program provides a pathway for entry of coregistered variant data in ClinVar and offers participants a mechanism to receive updates about their genetic results. Although the focus of the IDDRCs is on research for NDDs, the platform we describe— pairing genetic data with standardized phenotype ascertained through centers of expertise in a centralized data commons—could be translated across other disciplines, such as cardiovascular genetics or inborn errors of metabolism.

On analysis of the first 179 variants registered through this study, 76 (42.5%) were novel to ClinVar. There were 62 variants (34.6%) that were absent from all databases analyzed, most of which retrieve variants from the published literature (Table 2). Our study highlights the novelty of clinical genetic data siloed within health systems and laboratory databases. Such variant-level data—with an established pathway for pairing with standardized neurobehavioral phenotyping and EHR data—is a unique resource for elucidating unbiased associations between genotype and phenotype for NDDs. Although there are other resources in which paired genotype and phenotype exist, such as DECIPHER,[17] the BGR is unique in its requirement for clinically ascertained variants, the standardized nature of phenotypic data derived from the RNAP and the EHR, and the ability to recontact participants through GenomeConnect.

The standard ClinGen framework for evaluation of gene-disease validity involves ranking evidence from "Definitive" to "Refuted" and making this publicly available via its website. BGR data may be used for the curation of genes that have not yet been evaluated by ClinGen for their contribution to NDD, as well as for the reevaluation of those that previously only reached "Moderate" or "Limited" levels of evidence. The BGR ID/Autism GCEP team was launched in parallel to the registry's establishment to integrate clinically derived genomic and health data into brain gene curation. For example, the gene *ASH1L* (HGNC:19316) was recently curated for complex NDD by the BGR ID/Autism GCEP Team, with a "Definitive" result, confirming the role of this gene in neurodevelopment (https://search.clinicalgenome.org/kb/genes/HGNC:19088). Two BGR probands, each with de novo truncating variants dated 2019 and 2020, were included in the curation. Had these variants been shared earlier, "Definitive" status could have been assigned 3 years ago.

Furthermore, the BGR participants manifested previously unreported phenotypic features, including ataxic gait in one and depression, anxiety, and language disorder in the other. This illustrates that inclusion of variants identified in a clinical context into gene curation, rather than only those that have been published, significantly expands the scope and spectrum of scorable variant-level evidence and has the potential to accelerate resolution of gene-disease validity.

Brain gene curation directly informs outcomes of clinical testing for children and adults with NDD presentations. Although recent evidence-based ACMG guidelines recommend exome/genome sequencing as a first- (or second-) line test for ID,[18] gene sequencing panels for ID/autism remain commonly used in clinical practice, and the approach to gene inclusion varies between laboratories. It has been recommended that genes with a level of evidence below "Moderate" for gene-disease validity be excluded from gene panels.[19] By accelerating and enriching the evaluation of ID /autism genes, BGR data have potential impact on decisions around gene inclusion and diagnostic outcomes of such testing in the clinical context. As the diagnostic paradigm for NDDs continues to shift from panel to exome/genome sequencing, brain gene curation will also be critical in informing variant classification and clinical interpretation.

Expert review of paired genotype and phenotype data derived from the BGR can also be used to inform variant classification. It is recommended that variants in genes with "Limited" evidence for a gene-disease relationship be classified no higher than VUS, and variants in genes with a "Moderate" level of evidence be classified no higher than LP under ACMG criteria.[8,20] The burden of VUS following exome sequencing is reported to be in the range of 25.3% to 86%,[5,21,22] and this classification has a potential negative impact on parental understanding, physician comfort, and clinical decision-making around genomic results.[6,14] Further, VUS classifications pose a significant drain on health care resources, requiring time and expertise for analysis, interpretation, and reinterpretation.[6] Of the 103 BGR variants that were already present in ClinVar, 28 (26.9%) had an aggregate ClinVar classification of VUS and 11 (10.6%) had "conflicting interpretations of pathogenicity." The VUS category includes, firstly, variants in genes for which the relationship between the gene and disease has not been established (candidate gene or gene of uncertain significance) and, secondly, variants in established disease genes for which there is conflicting or insufficient evidence for a LP or LB classification. The BGR will aid in the resolution of VUS in genes for which the relationship with disease has not been established. BGR enrollment also has the potential to add further information about specific variants in known disease genes. The absence of detailed phenotypic and segregation information in most laboratory-submitted ClinVar entries can limit ability to resolve uncertainty around these variants. The BGR initiative will enrich this data set and introduce a clear means for obtaining further information and performing multi-scale analyses of such data. Thus, the BGR will inform the resolution of uncertainty around variant pathogenicity with ensuing potential for clinical impact.

For P/LP variants identified in genes with an established role in neurodevelopment, the BGR can contribute to the phenotypic and variant spectrum. The inclusion of variants identified in clinical settings, regardless of indication for testing, reduces the likelihood of ascertainment

or publication bias seen in the literature to date. This approach will facilitate a broader, more accurate understanding of the disease spectrum, by including, for example, individuals with mild or atypical phenotypes, or those whose indication for genetic testing was for reasons other than a neurodevelopmental presentation. In instances where a variant in a gene implicated in an autosomal dominant (or X-linked) NDD is found to be inherited from an apparently unaffected, or mildly affected parent, enrollment is also be offered to the parent, broadening the understanding of variable expressivity and reduced penetrance of such disorders. Three parent-child pairs have been enrolled to date with VUS in *SYN1* (HGNC:11494), *MYT1L* (HGNC:7623), and *CNTN4* (HGNC:2174) (Supplemental Table 1).

An example of the potential to broaden the understanding of the phenotypic spectrum is the *CTCF* (HGNC: 13723) gene, which, when initially evaluated by the ID/Autism GCEP in 2017, had an evidence level of "Moderate" for a relationship to autosomal dominant syndromic ID. Since its inclusion in the BGR, CTCF has been recurated (2021) with more recent publications[23,24] and reached a "Definitive" level of evidence. However, the spectrum of severity has broadened,[24] extending from mild developmental delay or normal IQ to severe ID, with variable presence of features suggestive of autism spectrum disorder. For this gene, the BGR focus has shifted from collecting case-level evidence to support gene-disease validity to accumulating a spectrum of cases to better define the neurobehavioral phenotype and the degree of variability in severity. At the time of writing, there are 6 participants with clinically identified variants in this gene, coregistered with the BGR and GenomeConnnect. Interested researchers may request access to the neurobehavioral assessment data, for example, to better define the phenotypic spectrum of this disorder. The registry thus has utility from the discovery phase (demonstrating gene-disease validity) to the delineation phase (informing phenotypic and mutational spectrum) of putative brain genes.

As demonstrated in our study, the siloed nature of clinical variants is a limitation to translational research for rare genetic disorders, particularly for the rapid identification of research eligible individuals. In addition to resolving questions around the relationship of rare brain gene variants and disease, the registry will evolve into a resource for enrollment into natural history studies and clinical trials for individuals with rare monogenic disorders. Further, there may be direct benefits to participants because those who have consented to be recontacted through GenomeConnect may receive updates to their genetic test results arising from shifts in gene-disease validity evaluation or variant reclassifications utilizing BGR data.[14] Laboratory and clinical protocols for reassessing and recontacting patients with variant reclassifications differ; as such, participants may not be aware of this updated information. The participants harboring variants with an intralaboratory discrepancy and BGR report date that precedes the reporting laboratory's ClinVar classification are currently being recontacted by GenomeConnect. Participants that opt to receive updates from GenomeConnect are informed that there may be an update to their genetic results and are encouraged to contact their provider for a review and to request an updated genetic testing report. This recontact can trigger diagnoses and changes in management, in the case of an upgrade from VUS to Pand, in the case of VUS to LB, can prompt additional evaluation.[9,14] Via coenrollment in BGR and GenomeConnect, patients can contribute to the genomics knowledgebase and can be informed of changes in their results.

Finally, ClinVar provides a critical conduit for BGR data requests for stakeholders outside of the program. Researchers who require more information (on, for example, phenotype or inheritance) can request access via the study website (https://braingeneregistry.wustl.edu/). Following evaluation and approval, deidentified data can be accessed via CIELO (including genetic variant, RNAP data, EHR data, and GenomeConnect health and genetic survey results), in accordance with IRB-approved guidelines.

There are several limitations to this study. First, the BGR aims for unbiased recruitment of individuals with rare variants in brain genes regardless of phenotype and regardless of clinical interpretation of causality of a given variant, but search methodologies may have skewed ascertainment. For IDDRC sites that utilized EHR-based searches, there may have been enrichment for variants clinically interpreted as causative for neurodevelopmental presentations. For sites that used internal databases housed in neurology, there may have been a bias toward identifying and recruiting individuals with neurodevelopmental presentations. A further limitation is that, although we were able to determine the timeline of previous submission to ClinVar, this was not possible for other disease databases.

Future goals of the BGR include broadening use of the data set to other GCEPs and improving interregistry operability. Efforts are underway to better automate the process of data extraction so that BGR data can be harnessed for other curation groups. A pilot project to map Human Phenotype Ontology terms from the RNAP data is nearing completion. Given the overlap between the BGR and other rare neurodevelopmental disorder registries, future goals also include feasibility evaluations of interregistry operability. Efforts to harmonize data governance structures, avoid duplication, limit participation burden, and share unique data assets will be explored.

In summary, we present the first 179 variants from participants coenrolled in the BGR and GenomeConnect. Deidentified participant phenotypic data, derived from a standardized neurobehavioral assessment and the EHR, are available upon request and approval. Of the 179 variants identified in enrolled participants, 76 (42.5%) were novel to ClinVar. This confirms our hypothesis that clinically identified variants represent an under-utilized resource for enhancing scientific knowledge about the association between rare genetic variants and neurodevelopmental outcomes. Coregistration of such variants with standardized neurobehavioral phenotyping and EHR data, as has been accomplished through the BGR, is poised to advance understanding of gene-disease relationships and variant pathogenicity and will serve as a unique resource for translational advances relating to rare NDDs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grypp, Jessica Kinard, Abigail Moradel Higareda, Paul Deppen, Tiffany Hartman, Anna Bican, Maddie Rockouski, Jordan Goodman, Emily Schneider, Elizabeth Forsen, Christina Sargent, Isaac Horn, and Madeline Thompson.

Data Availability

The genetic variants included in this analysis are available in the Supplemental File and are also available at https://www.ncbi.nlm.nih.gov/clinvar/submitters/508359/

References

1. 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. 10.1016/j.ajhg.2017.04.015 [PubMed: 28552198]

2. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062–D1067. 10.1093/nar/gkx1153 [PubMed: 29165669]

3. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44(D1):D862–D868. 10.1093/nar/gkv1222 [PubMed: 26582918]

4. Landrum MJ, Kattman BL. ClinVar at five years: delivering on the promise. Hum Mutat. 2018;39(11):1623–1630. 10.1002/humu.23641 [PubMed: 30311387]

5. Shashi V, McConkie-Rosell A, Schoch K, et al. Practical considerations in the clinical application of whole-exome sequencing. Clin Genet. 2016;89(2):173–181. 10.1111/cge.12569 [PubMed: 25678066]

6. Burke W, Parens E, Chung WK, Berger SM, Appelbaum PS. The challenge of genetic variants of uncertain clinical significance: a narrative review. Ann Intern Med. 2022;175(7):994–1000. 10.7326/M21-4109 [PubMed: 35436152]

7. Walkley SU, Abbeduto L, Batshaw ML, et al. Intellectual and developmental disabilities research centers: fifty years of scientific accomplishments. Ann Neurol. 2019; 86(3): 332–343. 10.1002/ana.25531 [PubMed: 31206741]

8. 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. 10.1038/gim.2015.30 [PubMed: 25741868]

9. Savatt JM, Azzariti DR, Faucett WA, et al. ClinGen's GenomeConnect registry enables patient-centered data sharing. Hum Mutat. 2018;39(11):1668–1676. 10.1002/humu.23633 [PubMed: 30311371]

10. 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. 10.1002/humu.22838 [PubMed: 26178529]

11. Rehm HL, Berg JS, Brooks LD, et al. ClinGen—the clinical genome resource. N Engl J Med. 2015;372(23):2235–2242. 10.1056/NEJMsr1406261 [PubMed: 26014595]

12. Payne P, Lele O, Johnson B, Holve E. Enabling open science for healthresearch: collaborative informatics environment for Learning on health outcomes (CIELO). J Med Internet Res. 2017;19(7):e276. 10.2196/jmir.6937 [PubMed: 28760728]

13. Riggs ER, Bingaman TI, Barry CA, et al. Clinical validity assessment of genes frequently tested on intellectual disability/autism sequencing panels. Genet Med. 2022;24(9):1899–1908. 10.1016/j.gim.2022.05.001 [PubMed: 35616647]

14. Savatt JM, Azzariti DR, Ledbetter DH, et al. Recontacting registry participants with genetic updates through GenomeConnect, the ClinGen patient registry. Genet Med. 2021;23(9):1738–1745. 10.1038/s41436-021-01197-8 [PubMed: 34007001]

15. Harrison SM, Dolinsky JS, Knight Johnson AE, et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. Genet Med. 2017;19(10):1096–1104. 10.1038/gim.2017.14 [PubMed: 28301460]

16. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2019. 10.1101/531210. Published online August 13.

17. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. Am J Hum Genet. 2009;84(4):524–533. 10.1016/ j.ajhg.2009.03.010 [PubMed: 19344873]

18. Manickam K, McClain MR, Demmer LA, et al. 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. 2021;23(11):2029–2037. 10.1038/s41436-021-01242-6 [PubMed: 34211152]

19. 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. 10.1038/s41436-019-0666-z [PubMed: 31732716]

20. Thaxton C, Good ME, DiStefano MT, et al. Utilizing ClinGen gene-disease validity and dosage sensitivity curations to inform variant classification. Hum Mutat. 2022;43(8):1031–1040. 10.1002/ humu.24291 [PubMed: 34694049]

21. Srivastava S, Cohen JS, Vernon H, et al. Clinical whole exome sequencing in child neurology practice. Ann Neurol. 2014;76(4):473483. 10.1002/ana.24251

22. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. Eur J Hum Genet. 2017;25(2):176–182. 10.1038/ ejhg.2016.146 [PubMed: 27848944]

23. Chen F, Yuan H, Wu W, et al. Three additional de novo CTCF mutations in Chinese patients help to define an emerging neurodevelopmental disorder. Am J Med Genet C Semin Med Genet. 2019;181(2):218–225. 10.1002/ajmg.c.31698 [PubMed: 30893510]

24. Konrad EDH, Nardini N, Caliebe A, et al. CTCF variants in 39 individuals with a variable neurodevelopmental disorder broaden the mutational and clinical spectrum. Genet Med. 2019;21(12):2723–2733. 10.1038/s41436-019-0585-z [PubMed: 31239556]
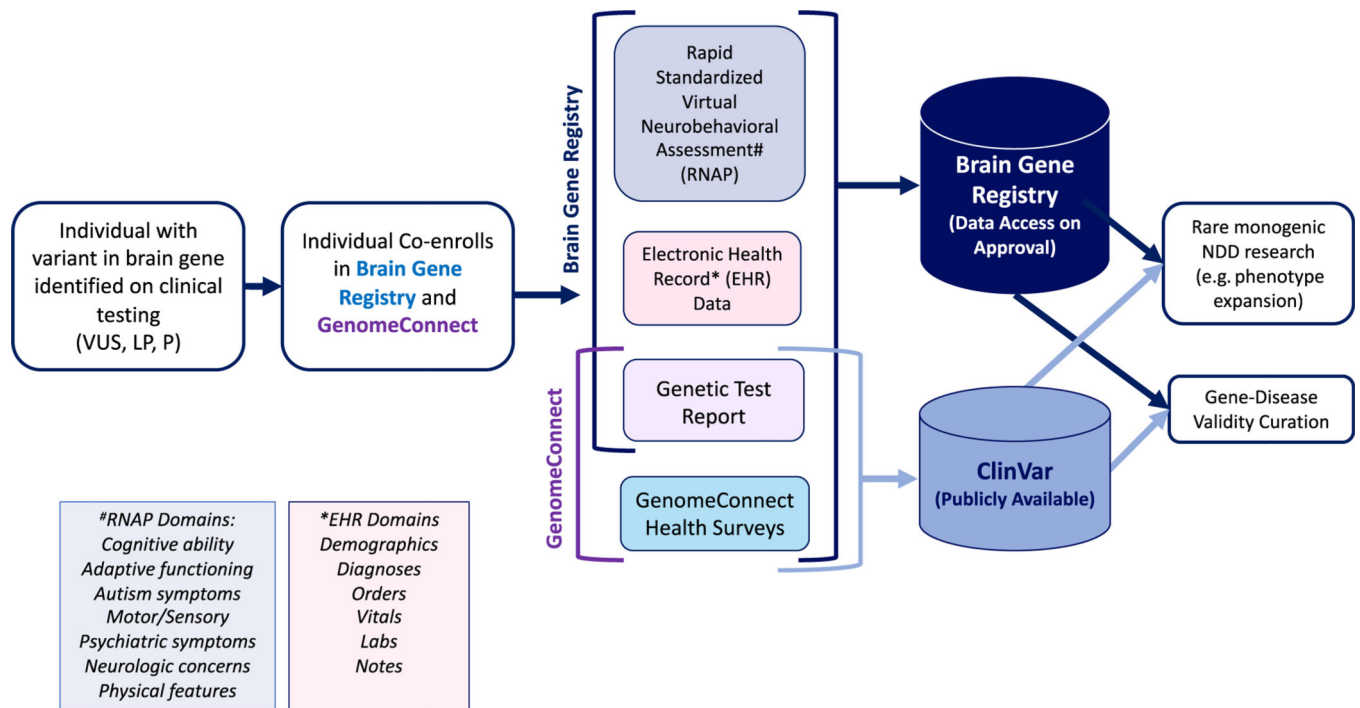
Figure 1. The brain gene registry.
Brain Gene Registry Graphic demonstrating coregistration of RNAP, EHR data, genetic test report, and parallel enrollment into GenomeConnect. These data will enrich rare monogenic NDD research and gene-disease validity. The RNAP and EHR domains are shown in the purple and pink boxes, respectively. EHR, electronic health record; LP/P, likely pathogenic/pathogenic; NDD, neurodevelopmental disorder; RNAP, rapid neurobehavioral assessment protocol; VUS, variant of uncertain significance.
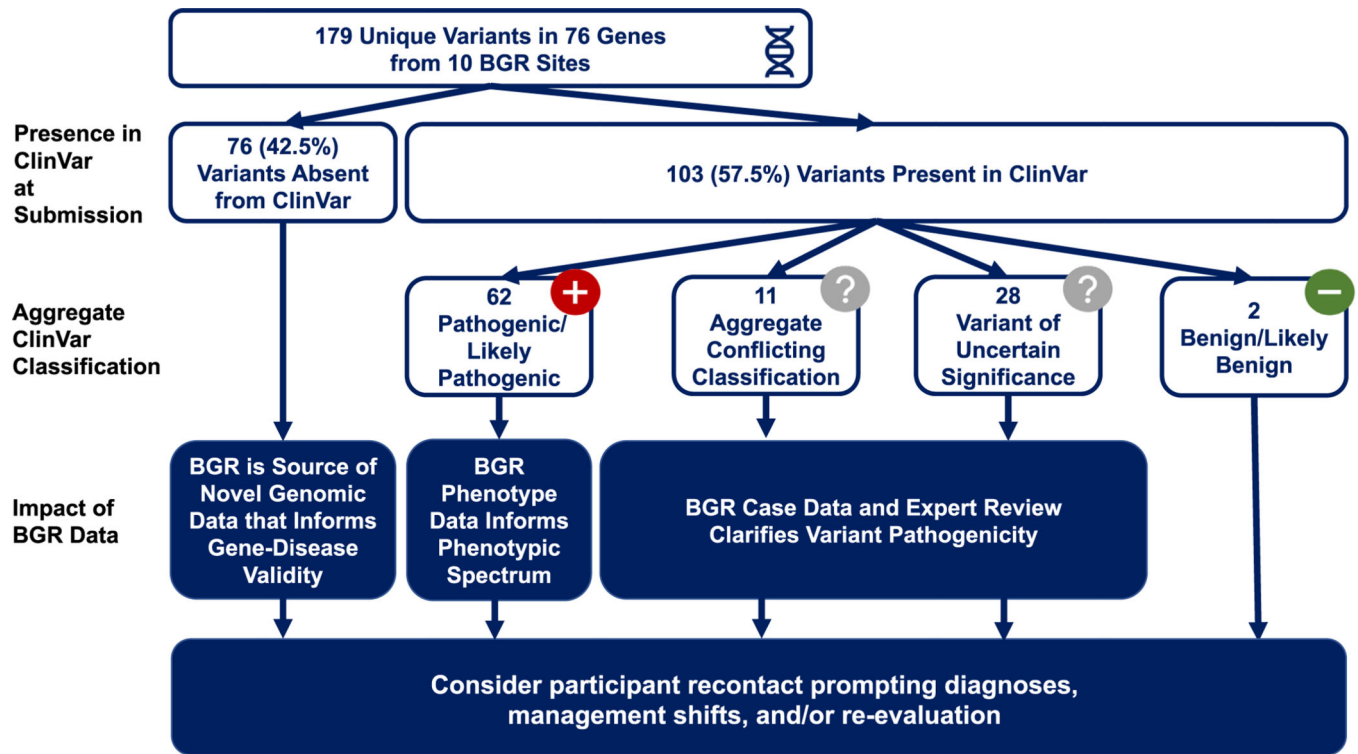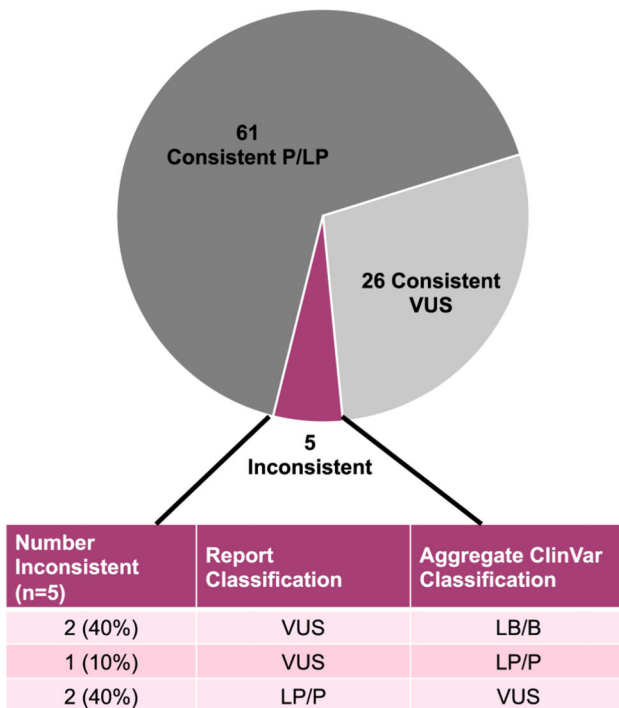
Figure 2. ClinVar Aggregate Classification at Time of Submission.

Of the 179 identified unique variants, 41.9% were absent in ClinVar at time of the BGR submission. Of the 103 variants present in ClinVar, the aggregate ClinVar classification was assessed. Enrollment of subjects into the BGR and coregistration of rare genetic variants, neurobehavioral, and EHR data have impact regardless of presence in ClinVar or ClinVar classification. BGR, Brain Gene Registry.

A Conflict Between Aggregate ClinVar Classification and BGR Report

Number Inconsistent (n=5)	Report Classification	Aggregate ClinVar Classification
2 (40%)	VUS	LB/B
1 (10%)	VUS	LP/P
2 (40%)	LP/P	VUS

B Intra-laboratory Discrepancy: Discrepancy Between BGR Report and Reporting Laboratory ClinVar Classification

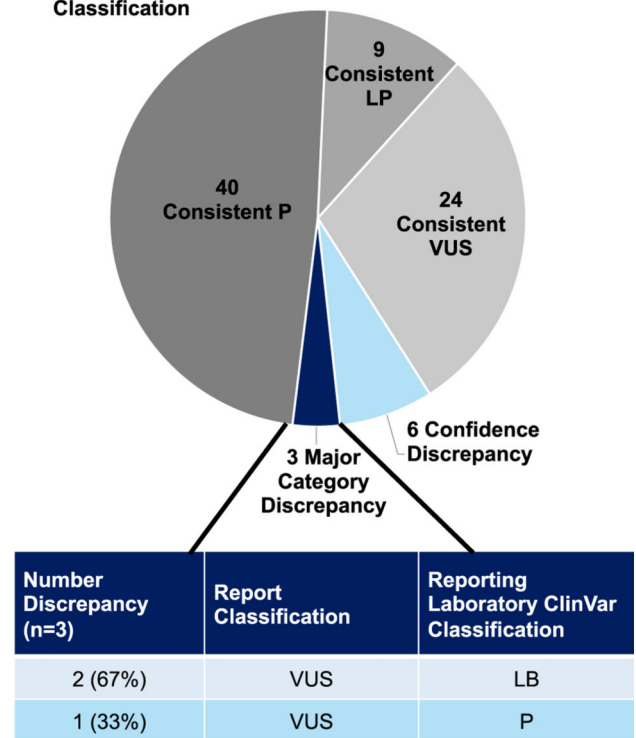| Number Discrepancy (n=3)	Report Classification	Reporting Laboratory ClinVar Classification
2 (67%)	VUS	LB
1 (33%)	VUS	P

Figure 3. Inconsistencies in Classification.
A. inconsistency between aggregate ClinVar classification and BGR report at time of BGR submission. Variants included are those with a record in ClinVar, at least 1 provided classification, and an overall non-conflicting classification ($n = 93$). B. Intralaboratory discrepancy: discrepancy between BGR report and reporting laboratory ClinVar classification. Variants included are those with a record in ClinVar from the BGR participant reporting laboratory ($n = 84$). BGR, Brain Gene Registry; LB/B, likely benign/benign; LP/P, likely pathogenic/pathogenic; VUS, variant of uncertain significance.

Table 1

Gene disease validity classification

Definitive

    • Role has been repeatedly demonstrated in research and clinical diagnostic settings

    • Upheld over time (in general, at least 3 years)

    • No convincing contradictory evidence

Strong

At least 2 separate studies with the following:

    • Numerous unrelated probands harboring variants with evidence for causality

    • Typically, experimental evidence is also present (but not required)

    • No convincing contradictory evidence

Moderate

  1 independent study with the following:

    • Several unrelated probands with pathogenic variant

    • Some supporting experimental data

    • No convincing contradictory evidence

Limited

  1 independent study with the following:

    • <3 unrelated probands with pathogenic variants OR

    • Multiple variants reported in unrelated probands but without sufficient evidence for pathogenicity

    • No convincing contradictory evidence

Gene-Disease Validity Classification levels and their descriptions, as per the Standard Operating Procedures Document for Gene-Disease Validity (May 2022 https://www.clinicalgenome.org/site/assets/files/5391/version_9_gene_curation_sop_final2.pdf). Genes that have previously been curated for neurodevelopmental disorder assertions with "Moderate" or "Limited" levels of evidence are given the highest priority for BGR gene selection.

Table 2

Presence in disease databases

Database	Present	Absent	Percent Absent
ClinVar submission driven database of genomic variants and their relationship to health (PMID: 29165669)	103	76	42.5%
Denovodb collection of germline de novo variants identified in the human genome from the published literature (including preprints) (https://denovo-db.gs.washington.edu/) SNV Only	12	151	92.6%
Varicarta database of variants identified in individuals diagnosed with Autism Spectrum Disorder (ASD) and reported in peer-reviewed, published literature. (PMID: 3170562) SNV Only	13	150	92.0%
ADMI DB resource for researchers and clinicians that provides genotype and phenotype data from neurodevelopmental disorders obtained from published literature (PMID: 26817790)	18	161	89.9%
LitVar resource that searches and retrieves variant information from the biomedical literature SNV Only	30	133	81.6%
DECIPHER database contains data from 42,678 patients with rare disease who have given consent for broad data sharing (PMID: 19344873)	22	157	87.7%
HGMD Professional collection of published germline variants in nuclear genes that are thought to be associated with human disease (PMID: 32596782)	68	111	62.0%
All Databases	117	62	34.6%

Disease databases analyzed for presence or absence of variants, together with descriptions of each of the databases. $N = 179$ variants.

HGMD, Human Gene Mutation Database; *SNV*, single nucleotide variants.