HHS Public Access

Author manuscript

Cell Genom. Author manuscript; available in PMC 2022 June 23.

Published in final edited form as:

Cell Genom. 2022 May 11; 2(5): . doi:10.1016/j.xgen.2022.100131.

Lumping versus splitting: How to approach defining a disease to enable accurate genomic curation

Courtney Thaxton[1,8,*],

Jennifer Goldstein[1],

Marina DiStefano[2],

Kathleen Wallace[1],

P. Dane Witmer[3],

Melissa A. Haendel[4],

Ada Hamosh[5],

Heidi L. Rehm[6,7],

Jonathan S. Berg[1]

[1]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA

[2]Precision Health Program, Geisinger, Danville, PA 17822, USA

[3]Johns Hopkins Genomics, Johns Hopkins University, Baltimore, MD 21218, USA

[4]Center for Health AI, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

[5]McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21205, USA

[6]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA

[7]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[8]Lead contact

SUMMARY

The dilemma of how to categorize and classify diseases has been debated for centuries. The field of medical genetics has historically approached nosology based on clinical phenotypes observed in patients and families. Advances in genomic sequencing and understanding of genetic contributions to disease often provoke a need to reassess these classifications. The Clinical

*Correspondence: courtney_thaxton@med.unc.edu.

AUTHOR CONTRIBUTIONS

Conceptualization, C.T., J.S.B., H.L.R., A.H., and M.A.H.; formal analysis, C.T., J.G., M.D., K.W., and P.D.W.; funding acquisition, J.S.B., H.L.R., A.H., and M.A.H.; investigation, C.T., J.G., M.D., K.W., and P.D.W; methodology, C.T., J.G., M.D., K.W., P.D.W., J.S.B., H.L.R., A.H., and M.A.H.; writing – original draft, C.T. and J.S.B.; writing – review & editing, C.T., J.G., M.D., K.W., P.D.W., J.S.B., H.L.R., A.H., and M.A.H.

SUPPLEMENTAL INFORMATION

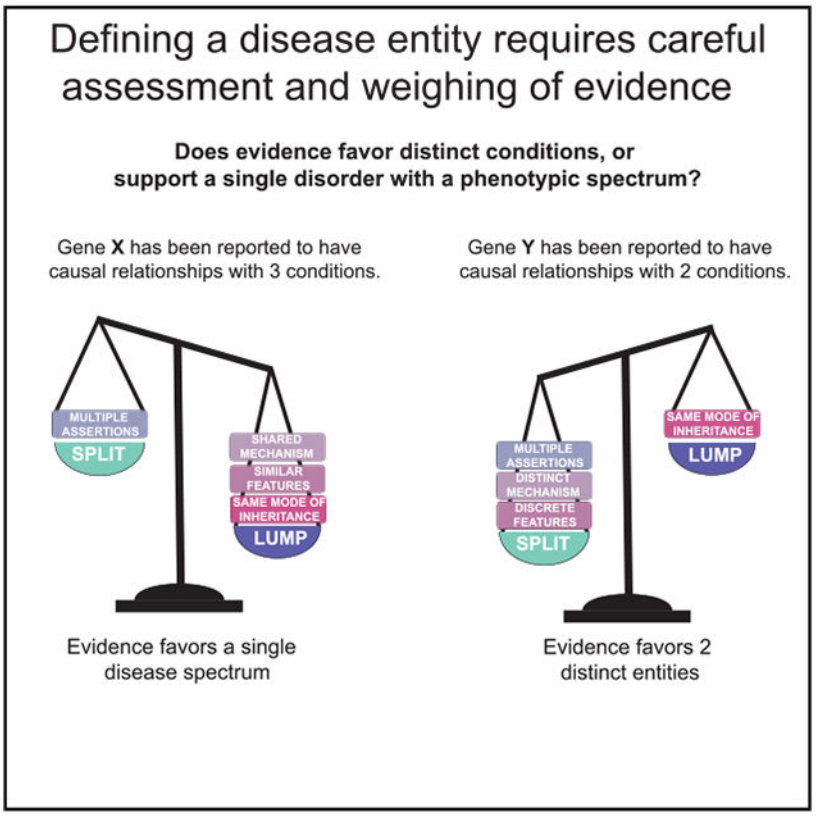
Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2022.100131.

DECLARATION OF INTERESTS

H.L.R. is a member of the *Cell Genomics* advisory board. The other authors declare no competing interests.

Genome Resource (ClinGen) has developed frameworks to classify the strength of evidence underlying monogenic gene-disease relationships, variant pathogenicity, and clinical actionability. It is therefore necessary to define the disease entity being evaluated, which can be challenging for genes associated with multiple conditions and/or a broad phenotypic spectrum. We therefore developed criteria to guide "lumping and splitting" decisions and improve consistency in defining monogenic gene-disease relationships. Here, we outline the precuration process, the lumping and splitting guidelines with examples, and describe the implications for clinical diagnosis, informatics, and care management.

Graphical abstract



In brief

Thaxton et al., outline four criteria (assertion, molecular mechanism, phenotypic expressivity, and inheritance pattern) as key pieces of evidence to determine the appropriate disease entity for use in gene-disease clinical validity classifications.

INTRODUCTION

Taxonomy, the process of classifying entities or naming them, requires determining whether to lump together entities in one category or split them apart. Charles Darwin articulated this challenge in a letter to his colleague J.D. Hooker in 1857,[1] in which he wrote of being "extremely interested in tabulating the size of genera and species" and how "it is good to

have hair splitters and lumpers." Lumpers are individuals who classify broadly and generally allow for ranges of characteristics to be classified into fewer entities, whereas splitters classify based on specific defining characteristics and/or nuances and thus create multiple classifications to reflect these distinctions.

In medicine, the nosology of disease entities has historically relied on clinical/phenotypic features but adapted to technological advances by incorporating other information that could be gained from physiological measurements, imaging, histology, biomarkers, and other data. While individual phenotypic features may be relevant to the clinical diagnosis of disease, they are not necessarily specific to the underlying cause. This is certainly true in the subset of conditions that are considered to have a Mendelian pattern of inheritance; many of these entities were defined phenotypically long before a clear genetic etiology was identified. A seminal paper in 1969 by Victor McKusick[2] addressed the nosology of genetic conditions and recognized pleiotropism and genetic heterogeneity as two opposing factors that could influence disease classification. As defined by McKusick, pleiotropism refers to multiple phenotypic features in different organ systems arising from variation in a single gene, while genetic (locus) heterogeneity refers to the situation in which variation in different genes can result in nearly identical phenotypic features.

We have seen in past decades dramatic increases in the discovery of monogenic gene-disease relationships and greater appreciation of the phenotypic spectrum of these diseases,[3,4] while detailed elucidation of tissue expression and protein function have provided a mechanistic understanding of pleiotropism in many cases. However, the nature of many monogenic conditions introduces challenges for the curation and evaluation of evidence. Variable expressivity results in a spectrum of phenotypic features and/or disease severity observed in different individuals with a given disorder. The clinical features present in any given patient at a particular point in time represent only a static description, and cases reported in the scientific literature may therefore present an incomplete view of the overall phenotypic spectrum. These allelic and phenotypic intricacies can confound the distinction of disease entities for the gene in question, in some cases leading to multiple apparently different conditions attributed to the same gene, thus fundamentally complicating the naming of conditions and evaluating the underlying evidence.[5]

The Clinical Genome Resource (ClinGen) consortium was established by the National Institutes of Health (NIH) to define the clinical relevance of genes and variants for use in precision medicine.[6] To accomplish this goal, ClinGen has established frameworks by which evidence can be systematically evaluated to define the clinical validity of gene-disease relationships,[7] variant pathogenicity (https://clinicalgenome.org/working-groups/sequence-variant-interpretation/), and clinical actionability.[8] Defining the disease entity to be curated is critical to the evidence collection, assessment, and scoring rubric for final classification. A significant challenge inherent in this process is how to approach genes with multiple disease relationship assertions or broad phenotypic spectra. A curator may ask whether two or more of the gene-disease relationships for a single gene reflect variable expressivity within the same condition, or whether one previously described entity actually represents two or more conditions caused by different types of variation in the same gene. Thus, before any assessment can be done, it is necessary to determine when to lump or to split. Excessive

splitting to evaluate all of the asserted phenotypic features observed among individuals with variants in a given gene could result in *ad infinitum* curations and classifications that fail to reflect the overall nature of the disease entity. In contrast, lumping all of the genotype-phenotype observations for a given gene into one broad disease entity may fail to recognize the intricacies of the molecular mechanism and yield classifications that are neither specific enough for clinical decision making nor representative of true differences in disease entities. Defining gene-disease relationships in a standardized, reproducible, and fully provenanced manner is critical for accurate clinical diagnosis, prognosis, and management. Of note, rare genetic diseases are often defined heterogeneously across sources,[9] which can lead to inconsistent interpretation of the clinical relevance of genes and variants.

To assist in defining the most appropriate disease entity to evaluate for a given gene, ClinGen collaborated with Online Mendelian Inheritance in Man (OMIM) and the Monarch Initiative to develop guidance regarding the definition of monogenic disease entities. Agreement on a consistent system for naming (or at least defining) disease entities, although challenging, will ultimately facilitate the grouping of those entities ontologically, thus enhancing the scientific discovery and understanding of disorders with an underlying genetic component.[10,11] This focus on lumping and splitting is not intended to detract from the role of the physician in clinical diagnosis, but to improve variant classification and case-level interpretation of genetic findings. Certainly, disease nomenclatures must continue to evolve with the accumulation of more evidence, and we expect that the process of lumping and splitting will be ongoing. Here, we outline our general principles, criteria, and processes that were developed to assist clinical domain experts and curators in defining the most appropriate disease entity for use in various evaluation frameworks.

RESULTS

ClinGen has organized gene curation expert panels (GCEPs) to review genes that fall within a specific clinical subdomain. These groups often focus on clinical phenotypes that could represent an isolated disease entity or a group of related disorders; could be observed with both syndromic and non-syndromic presentations; or could result from monogenic or multifactorial causes. Overall, our analysis indicated a tendency for some expert groups to split gene curations based on a phenotypic feature of interest driven by current genetic testing practices (gene panels offered for certain clinical indications) rather than to lump and curate for the broader cardinal disease entity (see Table 1 for examples). While it may be informative to understand the level of evidence supporting the association with a particular phenotypic feature of interest, the natural bias within the literature for novelty, the frequent lack of complete medical histories in articles from narrowly focused specialty journals, and the difficulty in obtaining longitudinal data from large numbers of affected individuals will likely result in inaccurate classifications attributed to such gene-phenotypic feature relationships. This is especially concerning for genes associated with multiple phenotypic features that make up a syndrome, in which a classification based on a single phenotypic feature may vary significantly from that of the syndrome that includes the feature of interest. This is exemplified in the findings for the *MEN1* and *CRYAB* genes, in which splitting of associated phenotypic features resulted in a lower level clinical validity classification that was not a true reflection of the available evidence supporting a broader disease entity (see

the supplemental materials for detailed explanation). Such disparity could result in confusion among various stakeholders, including the clinical community.

Together, these findings guided the development of our general principle: Genes should be curated for a single condition (i.e., lumped) unless there are clear indications to split diseases into separate entities for curation of gene-disease validity and variant pathogenicity. Expert groups wishing to better define the spectrum and frequency of individual features of a syndrome should conduct a second stage of evaluation using carefully phenotyped individuals with pathogenic variants, rather than applying the gene-disease validity framework to accomplish this goal.

Lumping and splitting criteria

We identified four distinct criteria that should be considered when curators first begin to define the gene-disease entity to be evaluated (Figure 1).

Assertion—The first criterion asks whether the gene has already been asserted to be involved in more than one disease entity and/or cataloged by nosological (e.g., OMIM, Orphanet) and ontological (e.g., Monarch Initiative) authorities. Curators should review these resources as a starting point for the evaluation of the current assertions for any given gene, followed by authoritative reviews and primary literature as a secondary measure to evaluate any new assertions from experts in the field. If multiple assertions or named conditions have been associated with the gene, then the curator must then proceed with subsequent steps to decide which disease entity to curate, or whether to lump two or more asserted disease phenotypes associated with the gene of interest into one broad disease entity. If only one primary gene-disease relationship has been asserted, then curation should focus on that entity unless there are compelling reasons to define a split (see other criteria for further details). The Lumping and Splitting Working Group (LSWG) determined that the existence of a gene on a diagnostic panel for a particular phenotype was insufficient reason to create a split curation, given that it is standard practice for diagnostic laboratories to include conditions with broadly overlapping phenotypic spectra on testing panels to maximize clinical sensitivity, especially in the context of variable expressivity or testing paradigms in which syndromic features may be underrecognized.

Molecular mechanism—Given that ClinGen classifications are primarily focused on diseases with a monogenic underpinning, the molecular mechanism of disease is the second criterion for disease entity determination. Differences in the molecular mechanism between asserted disease entities for a gene in question may include loss-of-function (LOF) versus gain-of-function (GOF) variants, effects of variants in distinct transcripts, and variants occurring in distinct protein functional domains or gene regions. The molecular mechanism should be evaluated initially at the genetic level (e.g., variant type), followed by any biochemical evidence, since at the protein level, the determination of LOF versus GOF ultimately depends on the availability of functional assays to test the mechanism of pathogenicity. The absence of a clear consensus on the molecular mechanism underlying the role of a gene in disease, or lack of clarity about the differences in molecular mechanism between multiple asserted gene-disease entities, is considered a strong reason to lump those

entities together, unless other compelling evidence exists in favor of a split. This initial lumping does not preclude future splitting of disease entities once more evidence becomes available.

Phenotypic variability—As noted above, pleiotropy and variable expressivity are common features of many genetic conditions. This fundamental property can lead to the impression that variation within a given gene gives rise to disparate collections of phenotypic features and therefore distinct conditions. To evaluate this criterion, curators analyze the phenotypic features in individuals within the same family harboring the same genetic variant (intrafamilial variability), compared to the features seen in unrelated individuals with distinct variants (interfamilial variability). The specificity of phenotypes observed within families, or between unrelated individuals harboring the same variant, may be an indication of an entity that is distinct from other disease phenotypes observed among individuals with variants in the same gene, even in the absence of data to suggest distinct molecular mechanisms. However, inconsistent or less specific phenotypic variation may be indicative of variable expressivity observed in many diseases, presumably due to additional genetic and non-genetic modifiers. These other factors should be taken into consideration when determining the relevant disease entity.

Inheritance pattern—Inheritance pattern (essentially the monoallelic or biallelic requirement for disease expression) is often the most recognizable characteristic of a monogenic disorder, and many genes have been asserted to have distinct disease relationships depending on the combination of variants or inheritance patterns (see Table 1). Curators should evaluate whether the evidence suggests that the inheritance patterns observed in families with a given disease entity or entities have readily discernible phenotypic features and/or distinct clinical management. Alternatively, the evidence may suggest that the apparently distinct inheritance patterns represent a continuum of disease with differences in severity and/or age of onset, where the observed phenotypic features and the risk of developing phenotypic manifestations are correlated with the zygosity, and therefore dosage, of the disease-causing variants (e.g., low density lipoprotein receptor [LDLR] variants associated with familial hypercholesterolemia).

General guidance and rationale that may help to determine when to lump or split are outlined in Table 2.

It is also useful to characterize the disease entity being curated as (1) a simple disease entity, limited to one phenotypic feature in one organ system; (2) a syndrome, in which a gene is associated with multiple phenotypic features arising in multiple organ systems with or without variable expressivity; or (3) an intermediate category in which multiple related phenotypic features arise but are limited to one organ system (e.g., different cardiac manifestations associated with *ACTN2*). This structure is highlighted in Figure 2 and led the LSWG to coin a new nosological classification of “variable phenotype, single organ system.”

We developed a “precuration” approach for evaluating evidence in each of the four categories to determine the most appropriate disease entity for curation

(https://clinicalgenome.org/docs/lumping-and-splitting-precuration-template-blank/). This step occurs before formal evidence curation for gene-disease validity, as it guides the evidence that will be included and scored or excluded from the gene-disease assessment. In some cases, the expert panel may wish to curate a split disease entity for the purpose of disputing or refuting one of the assertions. Ultimately, the decision to lump or split is a balance of criteria, in which all of the evidence should be weighed to determine the appropriate disease entity/entities for classification (Figure 1). It is important to note that this strategy does not restrict the number of disease entities that can be related to a gene, but rather provides a way to assess the relevant condition(s) to curate for a gene given the available data and evidence at the time. In addition, the Mondo ontology aims to archive the evidence and provenance of such decisions, to maintain links between related concepts defined by various organizations.

DISCUSSION

Taxonomy has great importance for defining the nature of monogenic disorders, the pathogenic variants that cause them, and the evidence that supports such assertions. Accurate designation of disease entities is required for patients to obtain specific diagnoses that can inform an understanding of their disease, facilitate optimal management, and eventually enable therapeutic development. Fundamentally, the goal of genomic medicine is to establish specific molecular diagnoses that inform prognosis and management of patients. This requires clear delineation of gene-disease entities and the evidence supporting their clinical validity, variant pathogenicity, and actionability.[6]

The guidance presented here informs the curation of gene-disease relationships and subsequently the curation of variant pathogenicity. Importantly, the American College of Medical Genetics and Genomics now recommends that only genes with a gene-disease clinical validity classification of moderate or above, based on the current framework of ClinGen, be included on disease-focused genetic testing panels.[60] If curations were performed based on excessive splitting for each individual feature, then many genes associated with syndromes could be erroneously left off testing panels due to a perceived lack of evidence for a particular phenotypic feature (given the variable expressivity of many conditions and the limited phenotyping published in many case reports) (see Data S1). Ultimately, this could lead to misdiagnosis and inadequate patient care.

Harmonizing assertions among different expert groups requires concepts to be well defined so that comparisons can be made between like entities. In the process of formulating the above criteria, three distinct groups (ClinGen, OMIM, and Monarch) were able to better refine the nosological and ontological representations of several monogenic diseases, and we expect to further elaborate on this harmonization process to facilitate the analysis of genetic variants and the diseases they cause. Indeed, several ClinGen expert panels have actively pursued involvement in the refinement and restructuring of ontology relationships within a given clinical domain, specifically within the Monarch Initiative, allowing for collaboration with a broader ontological community that participates in the Monarch Initiative activities, including Human Phenotype Ontology and Orphanet.[61] In addition, the Gene Curation Coalition (GenCC: http://thegencc.org) aims to compare and harmonize

gene-disease relationships across multiple curatorial groups. The lumping and splitting process that defines each disease entity for curation will be critical to allow for direct mapping of the clinical validity assertions provided by each participating group.

Further augmentation of monogenic disease taxonomy and nomenclature will be enabled by computational resources that track the provenance and history of gene-disease assertions. The lumping and splitting decisions of ClinGen are available to the public via the website, including references to OMIM identifiers that were included in lumped or split entries. In addition, improved phenotype representation methods such as the Global Alliance for Genomics and Health Phenopackets exchange standard (http://phenopackets.org/) will allow a more systematic approach to classifying diseases based upon individual-level data and support computational approaches in genomic medicine. This type of work may also benefit future development and harmonization of disease codes (e.g., ICD-10) with biologically relevant ontologies, which could have significant effects on precision medicine in the future.[10]

The lumping and splitting process is expected to be dynamic and collaborative between ClinGen and other curatorial entities, as scientific advancements necessitate the re-evaluation of current gene-disease relationships. In some cases, these decisions will require considering re-naming disease entities, in accordance with a more complete understanding of their molecular underpinnings. Overall, the process of lumping and splitting has significant implications for monogenic gene-disease relationships and beyond, and careful consideration of these guidelines at the level of curation or the level of clinical data collection will be critical for defining each unitary and distinct monogenic disease entity.

Limitations of the study

The gene-disease validity work performed by ClinGen GCEPs, including the precuration step to inform the disease entities upon which curation will be performed, relies on published data, which can be inherently biased toward novelty for a new gene-disease relationship while lacking the necessary thorough phenotyping of patients longitudinally, as well as appropriate functional assays to determine the mechanism of the specific variant or gene as a whole in disease. Existing cases may have been ascertained based on similarity of phenotype leading to a bias toward a single disease entity. Therefore, the determination of the disease entity by ClinGen GCEPs is anticipated to evolve over the years as more information is revealed through enhanced phenotype-genotype studies, genotype first approaches to ascertaining individuals with disease, and functional assays to determine the mechanisms of disease. Finally, the lumping and splitting process is not meant to be broadly applied to change the clinical diagnoses of patients, but to bring a greater understanding to the genomic underpinnings of monogenic (i.e., Mendelian) disorders, and to increase awareness and understanding as genomic analysis enters predictive diagnostics in the future.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Courtney Thaxton, Ph.D (courtney_thaxton@med.unc.edu).

Materials availability—This study did not generate new unique reagents.

Data and code availability—All gene-disease relationships can be accessed from the Clinical Genome Resource website at www.clinicalgenome.org. The gene-disease clinical validity classifications are updated periodically, and the most up-to-date evaluations will be published to the website. Note, some curations may not appear readily on the website as they may be in the evaluation process by one of the gene curation expert panels for final gene-disease validity classification prior to publication.

METHOD DETAILS

To develop recommendations for defining the appropriate disease entity/entities for any given gene, nosological experts, ontological experts, clinicians, clinical molecular geneticists, and biocurators were assembled into a working group termed the Lumping and Splitting Working Group (LSWG; https://clinicalgenome.org/working-groups/lumping-and-splitting/). Initial discussions focused on the lumping and splitting concept and exemplar genes associated with multiple disease entities, including several that had been previously evaluated and classified by a small subset of ClinGen groups prior to the implementation of these criteria. These genes included *FHL1, LDLR, SLC26A4, FH, MSH2, ATM, RET, MEN1, SCN8A, CAV3, CRYAB, ACTN2, LAMP2,* and *PLN*. Each gene was reviewed for evidence supporting and/or contradicting involvement with each of the disease entities asserted in OMIM (www.OMIM.org), Orphanet (www.orpha.net), and the Monarch Initiative (www.monarchinitiative.org), through literature review and application of the 2017 version (SOP 5; https://www.clinicalgenome.org/docs/gene-disease-validity-sop-version-5/) of the ClinGen Gene-Disease Clinical Validity Standard Operating Procedure.[7] In accordance with the framework, a clinical validity classification of limited (0.1-6 points), moderate (7-11 points), strong (12-18 points), definitive (12-18 points and replication over time), disputed, or refuted was assigned to each gene-disease relationship, with the diseases defined based upon the working lumping and splitting guidance. For genes that had a prior ClinGen-approved classification, the LSWG reviewed the evidence and compared it with the evidence and classification(s) obtained using the disease(s) defined by the developing lumping and splitting guidance. Discrepancies in classifications and the potential ramifications of lumping and splitting decisions for clinical diagnostics were discussed and taken into account. Consistent findings began to emerge from the working examples which guided the LSWG decisions for determining the relevant disease entity (see Table 1 for the results of the evaluations for lumping and splitting). Four criteria emerged from this effort: assertion, molecular mechanism, phenotypic variability, and inheritance pattern (see recommendations below). In some cases, the final classifications for these genes may have changed and the most up-to-date classifications can be found at www.clinicalgenome.org.

The initial lumping and splitting guidance was formally incorporated into the gene-disease clinical validity process in 2018 (SOP version 6) and the latest guidance can be found on the Lumping and Splitting Working Group webpage at https://clinicalgenome.org/working-groups/lumping-and-splitting/, including the current guidelines, a precuration example and template (blank and example), and a video tutorial. The data presented here represents the precuration stage and suggestions by this working group for the gene-disease validity relationship for gene curation expert panels (GCEP) to consider for the final evaluation, classification approval, and publishing to the website; therefore all genes presented in this article may not be currently reflected on www.clinicalgenome.org based on their priority for curation within their respective GCEP. Furthermore, the ClinGen website represents a living document of the current classifications and will include the latest version and date for the represented gene-disease clinical validity classification, however new information may have been published since the last clinical validity assessment for all genes listed. ClinGen has procedures detailing our process for re-evaluation (recuration) for gene-disease clinical validity relationship (https://clinicalgenome.org/docs/gene-disease-validity-recuration-process/).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This publication was supported by the National Human Genome Research Institute (NHGRI) of the NIH under award numbers U01HG007437 (to J.S.B.), U24HG009650 (to J.S.B), U41HG006627 (to A.H.), and U24HG006834 (to H.L.R.) and the Office of Director R24OD011883 (to M.A.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

1. Darwin (1857). Darwin Project (University of Cambridge). https://www.darwinproject.ac.uk/letter/DCP-LETT-2130.xml#back-mark-Lfoot.f6.
2. McKusick VA (1969). On lumpers and splitters, or the nosology of genetic disease. Perspect. Biol. Med 12, 298–312. 10.1353/pbm.1969.0039. [PubMed: 4304823]
3. Amberger JS, Bocchini CA, Scott AF, and Hamosh A (2019). OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res. 47, D1038–D1043. 10.1093/nar/gky1151. [PubMed: 30445645]
4. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, Brookes AJ, Brudno M, Carracedo A, den Dunnen JT, et al. (2017). International cooperationto enable the diagnosis of all rare genetic diseases. Am. J. Hum. Genet 100, 695–705. 10.1016/j.ajhg.2017.04.003. [PubMed: 28475856]
5. Biesecker LG (1998). Lumping and splitting: molecular biology in the genetics clinic. Clin. Genet 53, 3–7. 10.1034/j.1399-0004.1998.531530102.x. [PubMed: 9550355]
6. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, et al. (2015). ClinGen–the clinical Genome resource. N. Engl. J. Med 372, 2235–2242. 10.1056/nejmsr1406261. [PubMed: 26014595]
7. Strande NT, Riggs ER, Buchanan AH, Ceyhan-Birsoy O, DiStefano M, Dwight SS, Goldstein J, Ghosh R, Seifert BA, Sneddon TP, et al. (2017). Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the clinical Genome resource. Am. J. Hum. Genet 100, 895–906. 10.1016/j.ajhg.2017.04.015. [PubMed: 28552198]

8. Hunter JE, Irving SA, Biesecker LG, Buchanan A, Jensen B, Lee K, Martin CL, Milko L, Muessig K, Niehaus AD, et al. (2016). A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. Genet. Med 18, 1258–1268. 10.1038/gim.2016.40. [PubMed: 27124788]

9. Haendel M, Vasilevsky N, Unni D, Bologa C, Harris N, Rehm H, Hamosh A, Baynam G, Groza T, McMurry J, et al. (2020). How many rare diseases are there? Nat. Rev. Drug Discov 19, 77–78. 10.1038/d41573-019-00180-y. [PubMed: 32020066]

10. Haendel MA, Chute CG, and Robinson PN (2018). Classification, ontology, and precision medicine. N. Engl. J. Med 379, 1452–1462. 10.1056/nejmra1615014. [PubMed: 30304648]

11. Haendel MA, McMurry JA, Relevo R, Mungall CJ, Robinson PN, and Chute CG (2018). A census of disease ontologies. Annu. Rev. Biomed. Data Sci 1, 305–331. 10.1146/annurev-biodatasci-080917-013459.

12. Simonin C, Devos D, Vuillaume I, de Martinville B, Sablonnière B, Destée A, Stoppa-Lyonnet D, and Defebvre L (2008). Attenuated presentation of ataxia-telangiectasia with familial cancer history. J. Neurol 255, 1261–1263. 10.1007/s00415-008-0857-z. [PubMed: 18575927]

13. Stankovic T, Kidd AM, Sutcliffe A, McGuire GM, Robinson P, Weber P, Bedenham T, Bradwell AR, Easton DF, Lennox GG, et al. (1998). ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of mutant ATM and the risk of leukemia, lymphoma, and breast cancer. Am. J. Hum. Genet 62, 334–345. 10.1086/301706. [PubMed: 9463314]

14. Zhang BQ, Si N, and Liu DF (2015). Identification of anovel four and a half LIM domain 1 mutation in a Chinese male presented with hypertrophic cardiomyopathy and mild skeletal muscle hypertrophy. Chin. Med. J. (Engl) 128, 2269–2270. 10.4103/0366-6999.162493. [PubMed: 26265627]

15. Gueneau L, Bertrand AT, Jais JP, Salih MA, Stojkovic T, Wehnert M, Hoeltzenbein M, Spuler S, Saitoh S, Verschueren A, et al. (2009). Mutations of the FHL1 gene cause Emery-Dreifuss muscular dystrophy. Am. J. Hum. Genet 85, 338–353. 10.1016/j.ajhg.2009.07.015. [PubMed: 19716112]

16. Knoblauch H, Geier C, Adams S, Budde B, Rudolph A, Zacharias U, Schulz-Menger J, Spuler A, Yaou RB, Nürnberg P, et al. (2010). Contractures and hypertrophic cardiomyopathy in a novel FHL1 mutation. Ann. Neurol 67, 136–140. 10.1002/ana.21839. [PubMed: 20186852]

17. Schoser B, Goebel HH, Janisch I, Quasthoff S, Rother J, Bergmann M, Müller-Felber W, and Windpassinger C (2009). Consequences of mutations within the C terminus of the FHL1 gene. Neurology 73, 543–551. 10.1212/wnl.0b013e3181b2a4b3. [PubMed: 19687455]

18. Windpassinger C, Schoser B, Straub V, Hochmeister S, Noor A, Lohberger B, Farra N, Petek E, Schwarzbraun T, Ofner L, et al. (2008). An X-linked myopathy with postural muscle atrophy and generalized hypertrophy, termed XMPMA, is caused by mutations in FHL1. Am. J. Hum. Genet 82, 88–99. 10.1016/j.ajhg.2007.09.004. [PubMed: 18179888]

19. Wagnon JL, Barker BS, Ottolini M, Park Y, Volkheimer A, Valdez P, Swinkels MEM, Pate IMK, and Meisler MH (2017). Loss-of-function variants of *SCN8A* in intellectual disability without seizures. Neurol. Genet 3, e170. 10.1212/nxg.0000000000000170. [PubMed: 28702509]

20. Gertler TS, and Carvill GL (2019). *SCN8A*: when neurons are so excited, they just can't hide it. Epilepsy Curr. 19, 269–271. 10.1177/1535759719858338. [PubMed: 31288536]

21. Meisler MH, Helman G, Hammer MF, Fureman BE, Gaillard WD, Goldin AL, Hirose S, Ishii A, Kroner BL, Lossin C, et al. (2016). SCN8A encephalopathy: research progress and prospects. Epilepsia 57, 1027–1035. 10.1111/epi.13422. [PubMed: 27270488]

22. Drilon A, Hu ZI, Lai GGY, and Tan DSW (2018). Targeting RET-driven cancers: lessons from evolving preclinical and clinical landscapes. Nat. Rev. Clin. Oncol 15, 151–167. 10.1038/nrclinonc.2017.175. [PubMed: 29134959]

23. Plaza-Menacho I, Burzynski GM, de Groot JW, Eggen BJ, and Hofstra RM (2006). Current concepts in RET-related genetics, signaling and therapeutics. Trends Genet. 22, 627–636. 10.1016/j.tig.2006.09.005. [PubMed: 16979782]

24. Di Zanni E, Adamo A, Belligni E, Lerone M, Martucciello G, Mattioli G, Pini Prato A, Ravazzolo R, Silengo M, Bachetti T, and Ceccherini I (2017). Common PHOX2B poly-alanine contractions

impair RET gene transcription, predisposing to Hirschsprung disease. Biochim. Biophys. Acta Mol. Basis Dis 1863, 1770–1777. 10.1016/j.bbadis.2017.04.017. [PubMed: 28433712]

25. Fitze G, König IR, Paditz E, Serra A, Schläfke M, Roesner D, Ziegler A, and Schackert HK (2008). Compound effect of PHOX2B and RET gene variants in congenital central hypoventilation syndrome combined with Hirschsprung disease. Am. J. Med. Genet. A 146A, 1486–1489. 10.1002/ajmg.a.32300. [PubMed: 18438890]
26. Pelet A, Geneste O, Edery P, Pasini A, Chappuis S, Atti T, Munnich A, Lenoir G, Lyonnet S, and Billaud M (1998). Various mechanisms cause RET-mediated signaling defects in Hirschsprung's disease. J. Clin. Invest 101, 1415–1423. 10.1172/jci375. [PubMed: 9502784]
27. Bordeaux MC, Forcet C, Granger L, Corset V, Bidaud C, Billaud M, Bredesen DE, Edery P, and Mehlen P (2000). The RET proto-oncogene induces apoptosis: a novel mechanism for Hirschsprung disease. EMBO J. 19, 4056–4063. 10.1093/emboj/19.15.4056. [PubMed: 10921886]
28. Brautbar A, Leary E, Rasmussen K, Wilson DP, Steiner RD, and Virani S (2015). Genetics of familial hypercholesterolemia. Curr. Atheroscler. Rep 17, 20. 10.1007/s11883-015-0491-z.
29. The molecular, biochemical, and cellular basis of genetic disease. In Thompson and Thompson Genetics in medicine eighth edition, Nussbaum RL, McInnes RR, and Willard HF, eds. (Elsevier), p. 227.
30. Catteruccia M, Sanna T, Santorelli FM, Tessa A, Di Giacopo R, Sauchelli D, Verbo A, Monaco ML, and Servidei S (2009). Rippling muscle disease and cardiomyopathy associated with a mutation in the CAV3 gene. Neuromuscul. Disord 19, 779–783. 10.1016/j.nmd.2009.08.015. [PubMed: 19773168]
31. Fee DB, So YT, Barraza C, Figueroa KP, and Pulst SM (2004). Phenotypic variability associated with Arg26Gln mutation in caveolin3. Muscle Nerve. 30, 375–378. 10.1002/mus.20092. [PubMed: 15318349]
32. Fischer D, Schroers A, Blümcke I, Urbach H, Zerres K, Mortier W, Vorgerd M, and Schröder R (2003). Consequences of a novel caveolin-3 mutation in a large German family. Ann. Neurol 53, 233–241. 10.1002/ana.10442. [PubMed: 12557291]
33. Woodman SE, Sotgia F, Galbiati F, Minetti C, and Lisanti MP (2004). Caveolinopathies: mutations in caveolin-3 cause four distinct autosomal dominant muscle diseases. Neurology 62,538–543. 10.1212/wnl.62.4.538. [PubMed: 14981167]
34. Gazzerro E, Sotgia F, Bruno C, Lisanti MP, and Minetti C (2010). Caveolinopathies: from the biology of caveolin-3 to human diseases. Eur. J. Hum. Genet 18, 137–145. 10.1038/ejhg.2009.103. [PubMed: 19584897]
35. Vatta M, Ackerman MJ, Ye B, Makielski JC, Ughanze EE, Taylor EW, Tester DJ, Balijepalli RC, Foell JD, Li Z, et al. (2006). Mutant caveolin-3 induces persistent late sodium current and is associated with long-QT syndrome. Circulation 114, 2104–2112. 10.1161/circulationaha.106.635268. [PubMed: 17060380]
36. Chiu C, Bagnall RD, Ingles J, Yeates L, Kennerson M, Donald JA, Jormakka M, Lind JM, and Semsarian C (2010). Mutations in alpha-actinin-2 cause hypertrophic cardiomyopathy. J. Am. Coll. Cardiol 55, 1127–1135. 10.1016/j.jacc.2009.11.016. [PubMed: 20022194]
37. Bagnall RD, Molloy LK, Kalman JM, and Semsarian C (2014). Exome sequencing identifies a mutation in the ACTN2 gene in a family with idiopathic ventricular fibrillation, left ventricular noncompaction, and sudden death. BMC Med. Genet 15, 99. 10.1186/s12881-014-0099-0. [PubMed: 25224718]
38. Girolami F, Iascone M, Tomberli B, Bardi S, Benelli M, Marseglia G, Pescucci C, Pezzoli L, Sana ME, Basso C, et al. (2014). Novel α-actinin 2 variant associated with familial hypertrophic cardiomyopathy and juvenile atrial arrhythmias: a massively parallel sequencing study. Circ. Cardiovasc. Genet 7, 741–750. 10.1161/circgenetics.113.000486. [PubMed: 25173926]
39. Wémeau JL, and Kopp P (2017). Pendred syndrome. Best Pract. Res. Clin. Endocrinol. Metab 31, 213–224. 10.1016/j.beem.2017.04.011. [PubMed: 28648509]
40. Jensen MK, Havndrup O, Christiansen M, Andersen PS, Diness B, Axelsson A, Skovby F, Køber L, and Bundgaard H (2013). Penetrance of hypertrophic cardiomyopathy in children and adolescents: a 12-year follow-up study of clinical screening and predictive genetic testing. Circulation 127, 48–54. 10.1161/circulationaha.111.090514. [PubMed: 23197161]

41. Vicart P, Caron A, Guicheney P, Li Z, Prévost MC, Faure A, Chateau D, Chapon F, Tomé F, Dupret JM, et al. (1998). A missense mutation in the αB-crystallin chaperone gene causes a desmin-related myopathy. Nat. Genet 20, 92–95. 10.1038/1765. [PubMed: 9731540]
42. Chen Q, Liu JB, Horak KM, Zheng H, Kumarapeli AR, Li J, Li F, Gerdes AM, Wawrousek EF, and Wang X (2005). Intrasarcoplasmic amyloidosis impairs proteolytic function of proteasomes in cardiomyocytes by compromising substrate uptake. Circ. Res 97, 1018–1026. 10.1161/01.res.0000189262.92896.0b. [PubMed: 16210548]
43. Diokmetzidou A, Soumaka E, Kloukina I, Tsikitis M, Makridakis M, Varela A, Davos CH, Georgopoulos S, Anesti V, Vlahou A, and Capetanaki Y (2016). Desmin and αB-crystallin interplay in the maintenance of mitochondrial homeostasis and cardiomyocyte survival. J. Cell Sci 129, 3705–3720. 10.1242/jcs.192203. [PubMed: 27566162]
44. Dubin RA, Wawrousek EF, and Piatigorsky J (1989). Expression of the murine alpha B-crystallin gene is not restricted tothe lens. Mol. Cell Biol 9, 1083–1091. 10.1128/mcb.9.3.1083. [PubMed: 2725488]
45. Rajasekaran NS, Connell P, Christians ES, Yan LJ, Taylor RP, Orosz A, Zhang XQ, Stevenson TJ, Peshock RM, Leopold JA, et al. (2007). Human alpha B-crystallin mutation causes oxido-reductive stress and protein aggregation cardiomyopathy in mice. Cell 130, 427–439. 10.1016/j.cell.2007.06.044. [PubMed: 17693254]
46. Sanbe A, Osinska H, Saffitz JE, Glabe CG, Kayed R, Maloyan A, and Robbins J (2004). Desmin-related cardiomyopathy in transgenic mice: a cardiac amyloidosis. Proc. Natl. Acad. Sci. U S A 101, 10132–10136. 10.1073/pnas.0401900101. [PubMed: 15220483]
47. Wang X, Osinska H, Klevitsky R, Gerdes AM, Nieman M, Lorenz J, Hewett T, and Robbins J (2001). Expression of R120G–αB-crystallin causes aberrant desmin and αB-crystallin aggregation and cardiomyopathy in mice. Circ. Res 89, 84–91. 10.1161/hh1301.092688. [PubMed: 11440982]
48. Wójtowicz I, Jabłońska J, Zmojdzian M, Taghli-Lamallem O, Renaud Y, Junion G, Daczewska M, Huelsmann S, Jagla K, and Jagla T (2015). Drosophila small heat shock protein CryAB ensures structural integrity of developing muscles, and proper muscle and heart performance. Development 142, 994–1005. 10.1242/dev.115352. [PubMed: 25715399]
49. Schröder R, and Schoser B (2009). Myofibrillar myopathies: a clinical and myopathological guide. Brain Pathol. 19 (3), 483–492. [PubMed: 19563540]
50. Haghighi K, Kolokathis F, Pater L, Lynch RA, Asahi M, Gramolini AO, Fan GC, Tsiapras D, Hahn HS, Adamopoulos S, et al. (2003). Human phospholamban null results in lethal dilated cardiomyopathy revealing a critical difference between mouse and human. J. Clin. Invest 111, 869–876. 10.1172/jci17892. [PubMed: 12639993]
51. Coman D, Kranc KR, and Christodoulou J (2006). Fumarate hydratase deficiency. [updated 2020 Apr 23]. In GeneReviews® [Internet], Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, and Amemiya A, eds. (University of Washington, Seattle), pp. 1993–2021.
52. Su D, Singer EA, and Srinivasan R (2015). Molecular pathways in renal cell carcinoma: recent advances in genetics and molecular biology. Curr. Opin. Oncol 27, 217–223. 10.1097/cco.0000000000000186. [PubMed: 25811348]
53. Wimmer K, Kratz CP, Vasen HFA, Caron O, Colas C, Entz-Werle N, Gerdes AM, Goldberg Y, Ilencikova D, Muleris M, et al. ; on behalf of the EU-Consortium Care for CMMRD C4CMMRD (2014). Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'care for CMMRD' (C4CMMRD). J. Med. Genet. 51, 355–365. 10.1136/jmedgenet-2014-102284. [PubMed: 24737826]
54. Lynch HT, Lynch PM, Pester J, and Fusaro RM (1981). The cancer family syndrome. Rare cutaneous phenotypic linkage of Torre's syndrome. Arch. Intern. Med 141, 607–611. 10.1001/archinte.1981.00340050059016. [PubMed: 7224741]
55. Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, and Boland CR (2009). Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin. Genet 76, 1–18. 10.1111/j.1399-0004.2009.01230.x.
56. Sanbe A (2011). Molecular mechanisms of .ALPHA.-Crystallinopathy and its therapeutic strategy. Biol. Pharm. Bull 34, 1653–1658. 10.1248/bpb.34.1653. [PubMed: 22040875]

57. van der Zwaag PA, van Rijsingen IA, Asimaki A, Jongbloed JD, van Veldhuisen DJ, Wiesfeld AC, Cox MG, van Lochem LT, de Boer RA, Hofstra RM, et al. (2012). Phospholamban R14del mutation in patients diagnosed with dilated cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy: evidence supporting the concept of arrhythmogenic cardiomyopathy. Eur. J. Heart Fail 14, 1199–1207. 10.1093/eurjhf/hfs119. [PubMed: 22820313]

58. Haghighi K, Bidwell P, and Kranias EG (2014). Phospholamban interactome in cardiac contractility and survival: a new vision of an old friend. J. Mol. Cell Cardiol 77, 160–167. 10.1016/j.yjmcc.2014.10.005. [PubMed: 25451386]

59. Selcen D (2011). Myofibrillar myopathies. Neuromuscul Disord. 21 (3), 161–492. 10.1016/j.nmd.2010.12.007. [PubMed: 21256014]

60. Bean LJH, Funke B, Carlston CM, Gannon JL, Kantarci S, Krock BL, Zhang S, and Bayrak-Toydemir P; ACMG laboratory quality assurance committee (2020). Diagnostic gene sequencing panels: from design to report-a technical standard of the American College of medical genetics and genomics (ACMG). Genet. Med 22, 453–461. [PubMed: 31732716]

61. Helbig I (2014). The ILAE Commission Genetic Blog. http://epilepsygenetics.net/2020/10/14/omim-to-retire-eiee-classification-an-important-step-to-overhaul-terminology-for-genetic-epilepsies/.

Highlights

- Guidance for defining a disease entity for gene and variant curations
- Highlights harmonization across nosological and ontological authorities
- ClinGen framework enables consistency of curations across multiple clinical domains

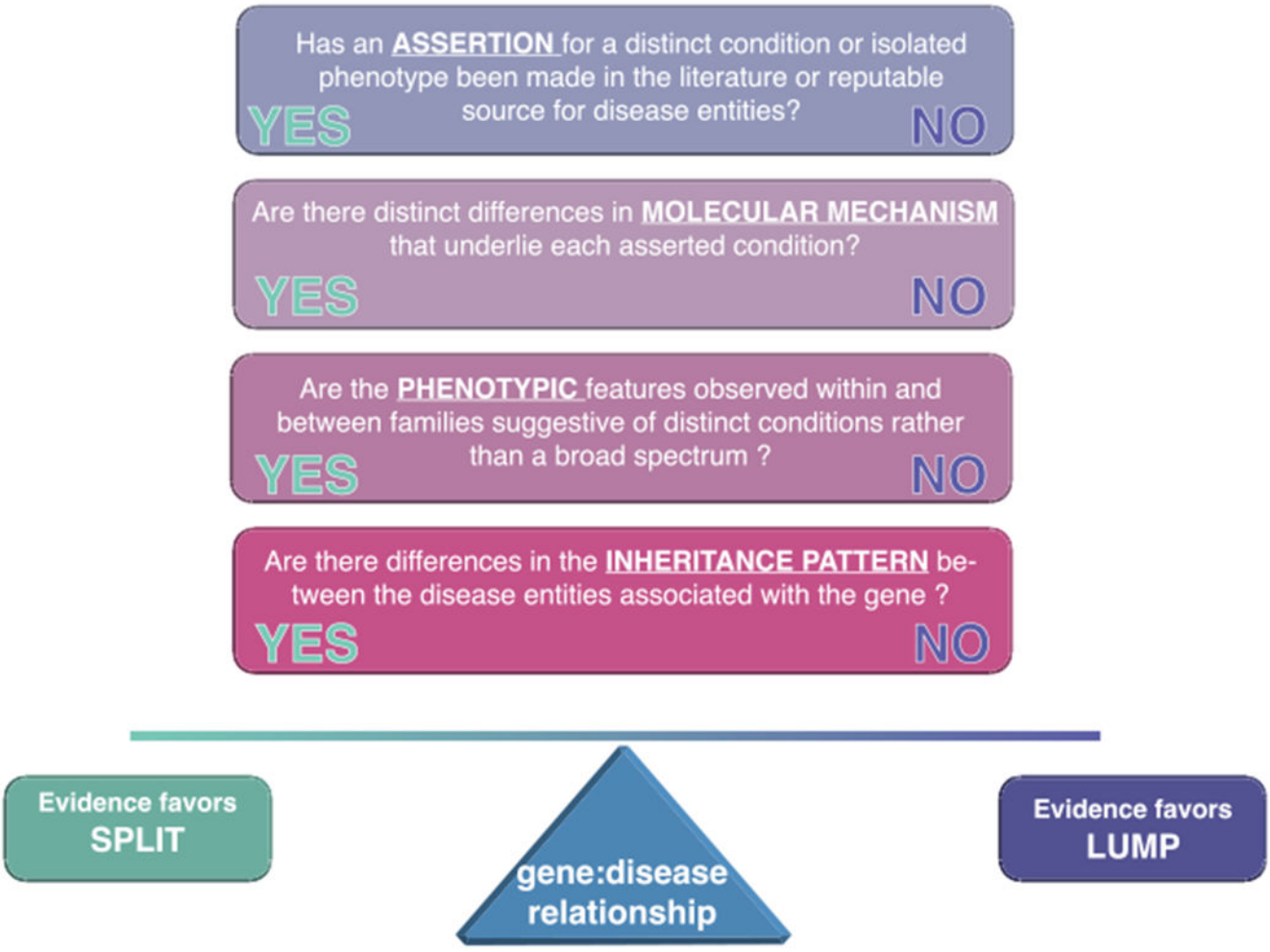


Figure 1. Weighing the evidence

The four criteria for lumping and splitting should be assessed and weighed as a balance. If only a single assertion has been made in the literature and/or databases of monogenic diseases (e.g., OMIM, Monarch [Mondo ontology], Orphanet), it is possible that no further steps are needed. However, some groups find it useful to precurate genes with a single disease entity to discuss disease nomenclature and review any new evidence or assertions that have not yet been formally captured in nosological and ontological resources. If multiple distinct disease entities have been asserted, then the curator will evaluate the evidence for the molecular mechanism, phenotypic expressivity, and inheritance pattern to determine whether to lump certain entities for curation as a syndrome or an organ-specific complex phenotype or to keep them separate as split disease entities. If the evidence is equally balanced between lumping or splitting, then experts should be consulted to compare the relevant weight of each piece of evidence.

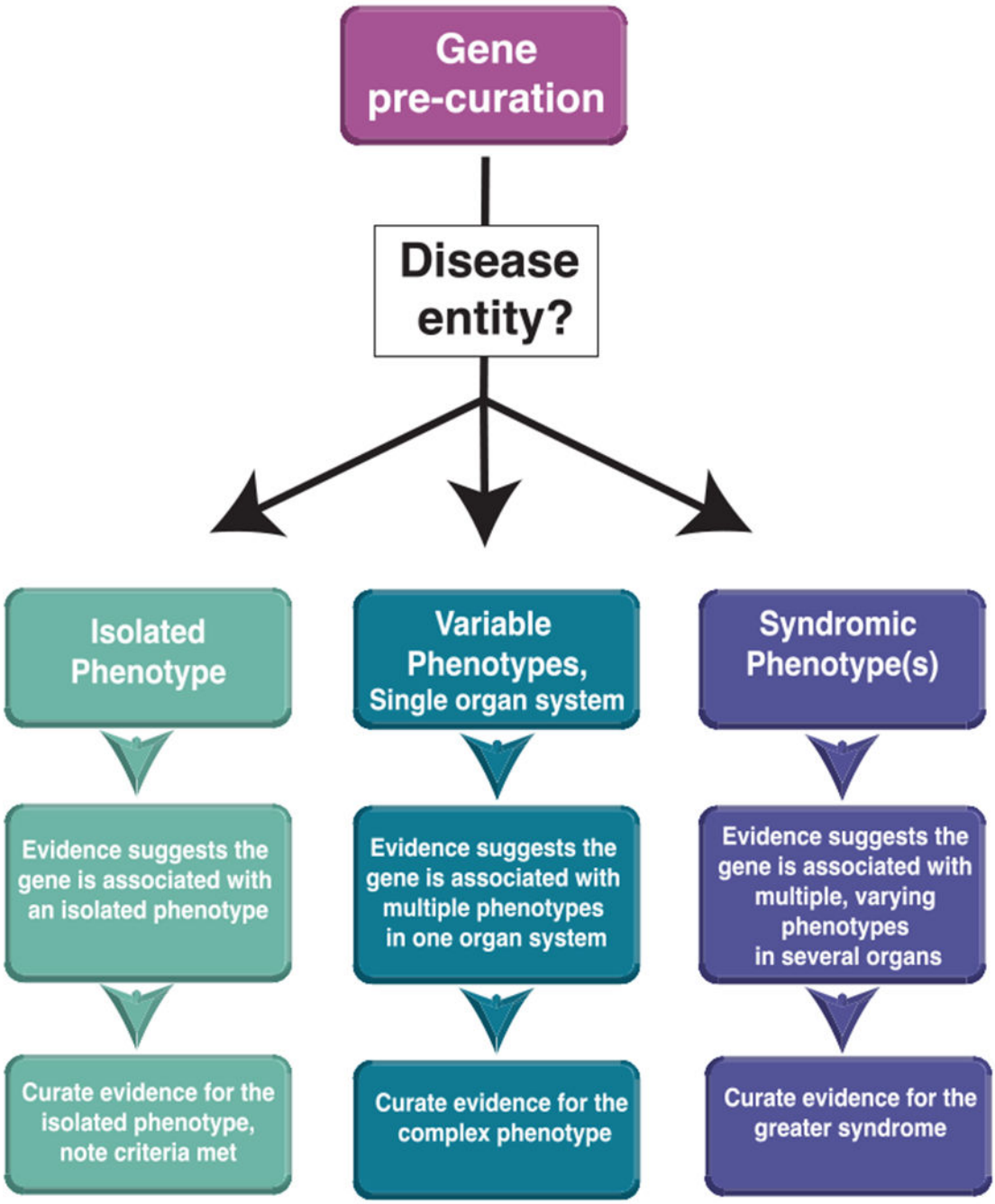


Figure 2. Lumping and splitting conundrum: defining a disease entity

When assessing the involvement of any given gene in disease, several possibilities for a disease entity may exist, including the following: (1) an isolated phenotype, in which 1 phenotype (or phenotypic feature) arises in a single organ system with no risk of other phenotypes arising in that organ system or elsewhere; (2) variable phenotypes in a single organ, in which multiple related phenotypes (or phenotypic features) arise in a single organ system; or (3) a syndromic phenotype, in which multiple, varying phenotypes occur

in multiple organs. Assessing the appropriate disease entity or entities to curate can be challenging, thus requiring the use of defined criteria.

Table 1.

Representative genes evaluated to determine lumping and splitting criteria

Gene *Date of OMIM record	Disease entity(s) per OMIM	Inheritance pattern	MIM#	Lump or split	Criteria met	Reference
ATM *as of 2018	{breast cancer, susceptibility to}	AD, Smu	114480	split	Assertion Phenotypic variability Inheritance pattern	Simonin et al.[12]; Stankovic et al.[13]; Zhang et al.[14]
	ataxia-telangiectasia	AR	208900	split	Assertion Phenotypic variability Inheritance pattern	
	lymphoma, B-cell non-Hodgkin, somatic	Som	n/a	not included as they are not germline	n/a	n/a
	lymphoma, mantle cell, somatic	Som	n/a			
	T cell prolymphocytic leukemia, somatic	Som	n/a			
FHL1 *as of 2018	Emery-Dreifuss muscular dystrophy 6, X-linked	XL	**300696**	lump	Assertion Molecular mechanism	Gueneau et al.[15]; Knoblauch et al.[16]; Schoser et al.[17]; Windpassinger et al.[18]; Wagnon et al.[19]
	myopathy, X-linked, with postural muscle atrophy	XL	**300696**		Phenotypic variability	
	reducing body myopathy, X-linked 1a, severe, infantile, or early childhood onset	XL	300717			
	reducing body myopathy, X-linked 1b, with late childhood or adult onset	XL	300718			
	scapuloperoneal myopathy, X-linked dominant	XL	300695			
SCN8A *as of 2018	cognitive impairment with or without cerebellar ataxia	AD	614306	split	Molecular mechanism, Phenotypic variability	Gertler et al.[20]; Meisler et al.[21]; Drilon et al.[22]
	epileptic encephalopathy, early infantile, 13	AD	614558	lump	Molecular mechanism Phenotypic variability	
	seizures, benign familial infantile, 5	AD	617080		Inheritance pattern	
RET *as of 2018	medullary thyroid carcinoma	AD	155240	lump into MEN2A and MEN2B	Molecular mechanism Phenotypic variability Inheritance pattern	Plaza-Menacho et al.[23]; Di Zanni et al.[24]; Fitze et al.[25]; Pelet et al.[26]; Bordeaux et al.[27]; Brautbar et al.[28]
	pheochromocytoma	AD	171300			
	multiple endocrine neoplasia IIA (MEN2A)	AD	171400	split	Molecular mechanism	
	multiple endocrine neoplasia IIB (MEN2B)	AD	162300			

Gene *Date of OMIM record	Disease entity(s) per OMIM	Inheritance pattern	MIM#	Lump or split	Criteria met	Reference
	central hypoventilation syndrome, congenital	AD	209880	split	to dispute	
	{Hirschsprung disease, susceptibility to, 1}	AD	142623	n/a	not evaluated, likely not Mendelian	
LDLR *as of 2017	hypercholesterolemia, familial, 1	AD, AR	143890	lump	Molecular mechanism Phenotypic variability Inheritance pattern	Nussbaum et al.[29]; Catteruccia et al.[30]
	LDL cholesterol level QTL2	AR, AR	143890			
CAV3 *as of 2017	cardiomyopathy, familial hypertrophic	AD	192600	lump	Assertion Molecular mechanism Phenotypic variability Inheritance pattern	Fee et al.[31]; Fischer et al.[32]; Woodman et al.[33]; Gazzerro et al.[34]; Vatta et al.[35]
	creatine phosphokinase, elevated serum	AD	123320			
	muscular dystrophy, limb-girdle, type 1C	AD	607801			
	myopathy, distal, Tateyama type	AD	614321			
	rippling muscle disease 2	AD	606072			
	long QT syndrome 9	AD	611818	split	Assertion Molecular mechanism	Chiu et al.[36]
ACTN2 *as of 2017	cardiomyopathy, dilated, 1AA, with or without LVNC	AD	**612158**	lump	Assertion Molecular mechanism Phenotypic variability Inheritance pattern	Bagnall et al[37]; Girolami et al[38]; Wémeau et al[39]
	cardiomyopathy, hypertrophic, 23, with or without LVNC	AD	**612158**			
CRYAB	cardiomyopathy, dilated, 1II myopathy, myofibrillar, 2 myopathy, myofibrillar, fatal infantile hypertonic, alpha-B crystallin-related	AD	615184	lump	Assertion Molecular mechanism Phenotypic variability Inheritance pattern	Jensen et al[40]; Vicart et al[41]; Chen et al[42]; Diokmetzidou et al[43]; Dubin et al[44]; Rajasekaran et al[45]; Sanbe et al[46]; Wang et al[47]; Wójtowicz et al[48]; Schröder et al.[49] Haghighi et al[50]
	AD	608810				
	AR	613869				
	cataract 16, multiple types	AD, AR	613763	needs evaluation	needs assessment by ocular experts as cataracts are observed in the myofibrillar myopathy	
SLC26A4	deafness, autosomal recessive 4, with enlarged vestibular aqueduct	AR	600791	lump	Assertion Molecular mechanism Phenotypic variability Inheritance pattern	Coman et al[51]
	Pendred syndrome	AR	274600			
FH	fumarase deficiency	AR	606812	split	Assertion Phenotypic variability Inheritance pattern	Su et al[52]
	leiomyomatosis and renal cell cancer	AD	150800	split	Assertion Phenotypic variability Inheritance pattern	Wimmer et al[53]
MSH2 *As of 2017	mismatch repair cancer syndrome 2	AR	619096	split	Phenotypic variability Inheritance pattern	Lynch et al[54]

Gene *Date of OMIM record	Disease entity(s) per OMIM	Inheritance pattern	MIM#	Lump or split	Criteria met	Reference
	colorectal cancer, hereditary nonpolyposis, type 1 (aka Lynch syndrome)	AD	120435	lump	Assertion Molecular mechanism Phenotypic variability Inheritance pattern	Lynch et al[55]; Sanbe et al[56]
	Muir-Torre syndrome	AD	158320			
LAMP2	Danon disease	XL	300257	n/a	follows general principle	n/a
PLN	cardiomyopathy, dilated, 1P	not noted	609909	lump	Assertion Molecular mechanism Phenotypic variability Inheritance pattern	van der Zwaag et al[57]; Haghighi et al[58]; Selcen et al[59]
	cardiomyopathy, hypertrophic, 18	AD	613874			
MEN1	adrenal adenoma, somatic	Som	n/a	not included as they are not germline	n/a	n/a
	angiofibroma, somatic	Som	n/a			
	carcinoid tumor of lung	Som	n/a			
	lipoma, somatic	Som	n/a			
	parathyroid adenoma, somatic	Som	n/a			
	multiple endocrine neoplasia 1	AD	131100	n/a	follows general principle	

Boldface MIM phenotype numbers indicate where OMIM has asserted the phenotypes are seemingly part of a single disease spectrum.

AD, autosomal dominant; AR, autosomal recessive; n/a, not applicable; Som, somatic variation; these are not included in the current gene-disease clinical validity assessments, as the framework is restricted to germline gene-disease relationships; XL, X-linked inheritance.

Cell Genom. Author manuscript; available in PMC 2022 June 23.

Table 2.

Reasons to lump or split gene-disease assertions

Reasons to lump	Reasons to split
An assertion for only 1 disease entity has been made in the literature	An assertion for >1 distinct disease entity has been made in the literature
No difference in molecular mechanism is observed among the disease entities	A well-established difference in molecular mechanism(s) between ≥2 disease entities is observed
Interfamilial phenotypic variability is no more pronounced than intrafamilial variability	Interfamilial phenotypic variability is more pronounced than intrafamilial variability
The difference in the inheritance pattern for the disease entities is representative of a continuum of disease (i.e., mild phenotypic features are observed in heterozygotes for recessive disease or dosage impacts are observed for dominant disease [more severe phenotype in homozygotes])	The representative disease entities between differing inheritance patterns are distinguishable, with notable varying phenotypes and/or clinical management distinctions
The disease entities in question are seemingly part of a variable phenotype observed within a single organ system and there is insufficient evidence for any single phenotype • If variants for each entity are VUSs, and no distinguishing phenotype is observed, then lump for a broader phenotype	To dispute or refute a disease entity assertion for the gene in question • Must have convincing evidence to dispute or refute • This would be a very rare occurrence, and the isolated disease entity being disputed or refuted cannot be included as part of the phenotypic spectrum observed in a syndrome associated with the gene of interest