Title

A dyadic approach to the delineation of diagnostic entities in clinical genomics

Journal

American Journal of Human Genetics, 108(1)

ISSN

0002-9297

Authors

Biesecker, Leslie G

Adam, Margaret P

Alkuraya, Fowzan S

et al.

Publication Date

2021

DOI

10.1016/j.ajhg.2020.11.013

Peer reviewed

A dyadic approach to the delineation of diagnostic entities in clinical genomics

Leslie G. Biesecker,[1,*] Margaret P. Adam,[2] Fowzan S. Alkuraya,[3] Anne R. Amemiya,[4] Michael J. Bamshad,[5,6] Anita E. Beck,[7,8] James T. Bennett,[9] Lynne M. Bird,[10,11] John C. Carey,[12] Brian Chung,[13] Robin D. Clark,[14] Timothy C. Cox,[15] Cynthia Curry,[16] Mary Beth Palko Dinulos,[17] William B. Dobyns,[18] Philip F. Giampietro,[19] Katta M. Girisha,[20] Ian A. Glass,[21] John M. Graham, Jr.,[22] Karen W. Gripp,[23] Chad R. Haldeman-Englert,[24] Bryan D. Hall,[25] A. Micheil Innes,[26] Jennifer M. Kalish,[27,28] Kim M. Keppler-Noreuil,[29] Kenjiro Kosaki,[30] Beth A. Kozel,[31]

(Author list continued on next page)

Summary

The delineation of disease entities is complex, yet recent advances in the molecular characterization of diseases provide opportunities to designate diseases in a biologically valid manner. Here, we have formalized an approach to the delineation of Mendelian genetic disorders that encompasses two distinct but inter-related concepts: (1) the gene that is mutated and (2) the phenotypic descriptor, preferably a recognizably distinct phenotype. We assert that only by a combinatorial or dyadic approach taking both of these attributes into account can a unitary, distinct genetic disorder be designated. We propose that all Mendelian disorders should be designated as "*GENE*-related phenotype descriptor" (e.g., "*CFTR*-related cystic fibrosis"). This approach to delineating and naming disorders reconciles the complexity of gene-to-phenotype relationships in a simple and clear manner yet communicates the complexity and nuance of these relationships.

Introduction

Rapid advances in the understanding and delineation of Mendelian conditions have outstripped our prior conceptions of these disorders. While the genetic basis of only a few dozen Mendelian disorders was known by 1990, genes associated with thousands of such conditions are now known and the numbers continue to climb.[1] Concurrently, there is an increasing recognition of the complexity and nuance of the relationship of genes to phenotypes. Whereas three decades ago a commonly taught aphorism in genetics was "one gene, one disease," we now know that this is incorrect for many genes. For example, variants in the fibrillin 1 gene (*FBN1* [MIM: 134797]) are associated with several apparently phenotypically distinct disorders that involve multiple organ systems (see OMIM in Web resources).[2] Some phenotypes comprise recognizably distinct pleiotropic syndromes, such as Marfan syndrome (MIM: 154700), but other variants in that same gene can be associated with non-syndromic ectopia lentis (MIM: 129600). In some cases, the various phenotypic associations are related to distinct mechanisms of pathogenesis (e.g., loss of function versus gain of function), while in others, they are instead due to a range of severity of a single mode of pathogenesis (e.g., from mild to complete loss of

[1]Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; [2]Division of Genetic Medicine, Department of Pediatrics, University of Washington School of Medicine, Seattle, WA 98105, USA; [3]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia; [4]Computercraft Corporation, McLean, VA 20005, USA; [5]Department of Pediatrics and Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [6]Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA; [7]Pediatric Genetic Medicine Service, University of Washington, Seattle, WA 98195, USA; [8]Seattle Children's Hospital, Seattle, WA 98015, USA; [9]Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute and Division Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, WA 98101, USA; [10]Department of Pediatrics, University of California San Diego, San Diego 92123, USA; [11]Rady Children's Hospital, San Diego, CA 92123, USA; [12]Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT 84108, USA; [13]Department of Paediatrics and Adolescent Medicine, Hong Kong Children's Hospital, Queen Mary Hospital, LKS Faculty of Medicine, University of Hong Kong, Hong Kong, China; [14]Loma Linda University School of Medicine, Department of Pediatrics, Division of Medical Genetics, Loma Linda, CA 92354, USA; [15]Department of Oral and Craniofacial Sciences, School of Dentistry and Department of Pediatrics, School of Medicine, University of Missouri-Kansas City, Kansas City, MO 64108, USA; [16]Genetic Medicine, Department of Pediatrics, University of California, Fresno, Fresno, CA 93701, USA; [17]The Geisel School of Medicine at Dartmouth, Department of Pediatrics, Section of Genetics and Child Development, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, USA; [18]Division of Genetics, Department of Pediatrics, University of Minnesota, Minneapolis, MN 55455, USA; [19]Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ 08901, USA; [20]Department of Medical Genetics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal 576104, India; [21]Department of Pediatrics and Medicine, University of Washington School of Medicine, Seattle, WA 98195, USA; [22]Cedars-Sinai Medical Center and Harbor-UCLA Medical Center, David Geffen School of Medicine at UCLA, Los Angeles, CA 90048, USA; [23]Division of Medical Genetics, Department of Pediatrics, AI DuPont Hospital for Children/Nemours, Wilmington, DE 19803, USA; [24]Mission Fullerton Genetics Center, Asheville, NC 28803, USA; [25]Greenwood Genetic Center, Greenwood, SC 29646, USA; [26]Department of Medical Genetics and Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB T3B 6A8, Canada; [27]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [28]Departments of Pediatrics and Genetics, Perelman School of Medicine, University of Pennsylvania,

(Affiliations continued on next page)

Ghayda M. Mirzaa,[6,32] John J. Mulvihill,[33,34] Malgorzata J.M. Nowaczyk,[35] Roberta A. Pagon,[2] Kyle Retterer,[36] Alan F. Rope,[37] Pedro A. Sanchez-Lara,[38] Laurie H. Seaver,[39] Joseph T. Shieh,[40] Anne M. Slavotinek,[40] Andrew K. Sobering,[41] Cathy A. Stevens,[42] David A. Stevenson,[43] Tiong Yang Tan,[44] Wen-Hann Tan,[45,46] Anne C. Tsai,[47] David D. Weaver,[48] Marc S. Williams,[49] Elaine Zackai,[50] and Yuri A. Zarate[51]

function). In yet other genes, disruption of distinct functional domains leads to loss of function of specific subsets of the activity of the protein. Another facet of the complexity is highlighted in conditions such as Bardet-Biedl syndrome (MIM: 209900)—what initially seemed to be a single phenotype has been associated with pathogenic variants in more than two dozen genes (see Forsyth and Gunay-Aygun GeneReviews in Web resources). The rasopathies are a further dimension of this—a family of seven phenotypes associated with variants in at least 13 genes.[3,4] We are very far from "one gene, one disease" pedagogy, and it is clear that the relationship of genes to phenotypes is complex and heterogeneous and continues to evolve. An additional important observation is that the identification of molecularly targeted therapies supports the need for developing a molecular taxonomy of disease, that is, a classification and organization of disease entities by the molecular pathophysiologic attributes rather than only considering their overt phenotype.

It is essential that disease taxonomy and naming conventions accurately reflect biologic reality while supporting clinicians in the diagnosis and treatment of genetic disorders. The current haphazard and inconsistent approach of naming some Mendelian disorders on the basis of their manifestations, phenotypic attributes (e.g., acronyms), and eponyms and others on the basis of the mutated gene serves none of these goals. We conclude that a more systematic categorization of Mendelian disorders that supports the intended uses of the information associated with the gene is needed. Here, we propose a standardized approach to this challenge for Mendelian disorders.

Proposal

We propose that a unitary diagnostic entity be a dyad of (1) a molecular etiology in the form of a HGNC/HUGO-approved gene descriptor and (2) a phenotype, preferably a recognizably distinct phenotype whenever that is feasible and justified. We propose that these dyadic descriptors should be used for all Mendelian disorders with a known genetic basis. This nomenclature should be applied in the clinical care setting, in textbooks and journals, in genetic databases and data repositories, and in the research environment.

General principles

We begin to address this complex question by focusing on Mendelian (sometimes called "single-gene") disorders. We define these as disorders for which a large proportion of the phenotypic variance is explained by the presence of a pathogenic variant(s) in a single gene. These disorders can also be recognized as those that are inherited in one of the classic Mendelian patterns of autosomal dominant, autosomal recessive, or X-linked, and in the case of mitochondrial disease, maternal inheritance.

Next, we introduce the concept of a "unitary, distinct, Mendelian disease entity." This entity comprises a label or descriptor that is used to define a disease (diagnosis, syndrome, or condition) that is considered to be unitary. It is a single entity that is discrete and meaningfully different from other diagnostic entities. It is critical to recognize that all such descriptors are heuristics (practical methods that are not guaranteed to be optimal but that are sufficient)—these descriptors may not accurately and comprehensively reflect biologic reality, but they serve as useful tools for clinicians to identify, characterize, and manage affected individuals. That a Mendelian disease entity is unitary and distinct ultimately reflects the wisdom and judgment of practitioners and researchers in the field to determine that one such entity is sufficiently distinct from others to make that

Philadelphia, PA 19104, USA; [29]Division of Genetics, Children's National Medical Center, Washington, DC 20010, USA; [30]Keio University School of Medicine, Tokyo 160-8582, Japan; [31]Translational Vascular Medicine Branch, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA; [32]Center for Integrative Brain Research, Seattle Children's Research Institute, Department of Pediatrics, University of Washington, Seattle, WA 98101, USA; [33]University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA; [34]National Human Genome Research Institute, Bethesda, MD 20892, USA; [35]Molecular Medicine & Pathology and Pediatrics, McMaster University, Hamilton, ON L8S 3K9, Canada; [36]GeneDx, Gaithersburg, MD 20877, USA; [37]Genome Medical, South San Francisco, CA 94080, USA; [38]Department of Pediatrics, Cedars-Sinai Medical Center and David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90048, USA; [39]Spectrum Health Medical Genetics and Genomics/Helen Devos Children's Hospital, Department of Pediatrics and Human Development, Michigan State University College of Human Medicine, Grand Rapids, MI 49503, USA; [40]Institute for Human Genetics and Division of Medical Genetics, Department of Pediatrics Benioff Children's Hospital, University of California, San Francisco, San Francisco, CA 94143, USA; [41]Department of Biochemistry, St. George's University School of Medicine, St. George's, Grenada; [42]Department of Pediatrics, University of Tennessee College of Medicine, Chattanooga, TN 37403, USA; [43]Division of Medical Genetics, Department of Pediatrics, Stanford University, Palo Alto, CA 94305, USA; [44]Victorian Clinical Genetics Services, Murdoch Children's Research Institute and Department of Paediatrics, University of Melbourne, Melbourne, VIC 3052, Australia; [45]Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA 02115, USA; [46]Harvard Medical School, Boston, MA 02115, USA; [47]Section of Genetics, Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA; [48]Department of Medical and Molecular Genetics, Indiana University School of Medicine, 975 W. Walnut Street, Indianapolis, IN 46202, USA; [49]Genomic Medicine Institute, Geisinger, Danville, PA 17822, USA; [50]Division of Human Genetics, Department of Pediatrics, Children's Hospital of Philadelphia, PA 19104, USA; [51]Section of Genetics and Metabolism, Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR 72202, USA
*Correspondence: lesb@mail.nih.gov
https://doi.org/10.1016/j.ajhg.2020.11.013.

distinction clinically useful. The descriptor should also take affected individuals' perspectives into account and avoid pejorative terminology.

The concept underlying this proposal is not novel as it leverages approaches developed in two essential genetics resources: Online Mendelian Inheritance in Man (OMIM) and GeneReviews. OMIM has wisely chosen to catalog genes and phenotypes separately. We propose to adopt and standardize a version of the approach used in several entries in the online NCBI Bookshelf GeneReviews (see Web resources). There, Pagon and colleagues have adopted what we describe as a dyadic approach to the definition of a disorder. Here, we use the term dyad in its literal form, which is a set of two elements (here genotype and phenotype) treated as one. Either element alone is considered an incomplete descriptor of the entity. GeneReviews has begun to adopt this approach by designating a number of Mendelian disease entities with the general dyadic structure of "(gene)-(phenotype)" or "(gene)-related (phenotype)" or "(gene)-associated (phenotype)," which takes advantage of the separate elements of genotype and phenotype, each of which are OMIM entities. The "gene" part of the descriptor is the HUGO gene nomenclature committee (HGNC/HUGO)-approved gene name, which should be a single OMIM entity. The "phenotype" part of the descriptor is a phrase that describes the clinical features of the disorder. Whenever possible, this should be a recognizably distinct phenotype; however, we recognize that this is not always possible.

In the following, we provide examples of the structure and utility of the proposed approach.

(1) *CHST3*-related skeletal dysplasia
(2) *NGLY1*-related congenital disorder of deglycosylation
(3) *PAFAH1B1*-associated lissencephaly/subcortical band heterotopia

These first three entities are current listings in GeneReviews, and we would consider each of them to be unitary and distinct Mendelian disease entities (the OMIM entries for these entities are *CHST3* [MIM: 603799]; *NGLY1* [MIM: 610661]; "congenital disorder of glycosylation" [MIM: 615273]; and *PAFAH1B1* [MIM: 601545]). Note that GeneReviews sometimes uses the word "associated," whereas most of their entries use "related," and they are now evolving using "related" in future updates. We do not ascribe a significant distinction of these minor formatting differences but would encourage uniformity and a focus on the concept of specifying both gene and phenotype and that the gene name strictly adhere to the preferred Human Gene Nomenclature Committee designation. Given that there is only a single phenotype or spectrum associated with each of these genes described in entities 1–3 (based on current knowledge) and only a single gene associated with each of these phenotypes, we recognize that specifying both the gene and the phenotype is redundant. Despite that redundancy, we advocate for describing them in this way because knowledge will change (these genes may be associated with other phenotypes in the future) as well as for uniformity with the nomenclature of disorders involving more complex gene-phenotype relationships (examples below). Indeed, these three obscure disorders were selected as examples here because they comprise some of the very few entities in GeneReviews where there is unique mapping of phenotype to gene, and vice versa. It is interesting to note that these three entities represent a range of phenotypic descriptors from the relatively simple "skeletal dysplasia" to an apparent hybrid or spectrum descriptor of "lissencephaly/subcortical band heterotopia." In the latter, the authors and editors are making the case that the two descriptors of neuronal migration defects lie on a spectrum and should be considered a single phenotypic spectrum. We also note that the phrase "congenital disorder of deglycosylation" focuses on a biochemical or laboratory test finding rather than an overt, physical, or clinical manifestation of a disorder. This reflects the wide range of concepts that we are not discussing that underlie phenotypic descriptors (see Discussion).

Two additional examples illustrate the use of the system when the gene-to-phenotype relationship is more complex.

(4) *GLI3*-related Pallister-Hall syndrome
(5) *GLI3*-related Greig cephalopolysyndactyly syndrome

These descriptors denote that these entities are considered to be two unitary, distinct Mendelian disorders (*GLI3* [MIM: 165240]; Pallister-Hall syndrome [MIM: 146510]; and Greig cephalopolysyndactyly syndrome [MIM: 175700]). This reflects the current judgment and state of knowledge in the field that despite their being allelic, these two phenotypes comprise a sufficiently distinct range of features that they are better designated as distinct entities rather than as a continuum. In this case, the phenotypic distinction also reflects an underlying difference in the pathogenetic mechanism of disease for these two phenotypes (dominant negative versus haploinsufficiency). However, not all such phenotypic distinctions will be correlated with a mechanism of disease or the mechanistic distinction may not be known. Inconsistency in this approach is evident in the example of *LMNA* in GeneReviews. The cardiomyopathy attributed to variants in *LMNA* (MIM: 150330) is designated as *LMNA*-related dilated cardiomyopathy ("cardiomyopathy, dilated, 1A" [MIM: 115200]), while other distinct disorders associated with *LMNA*, such as "lipodystrophy, familial partial, type 2" (MIM: 151660), "Hutchinson-Gilford progeria" (MIM: 176670), and "muscular dystrophy, congenital" (MIM: 613205), are not so designated.

In the following three examples, another aspect of the gene-to-phenotype complexity is addressed.

(6) *BBS1*-related Bardet-Biedl syndrome
(7) *MKKS*-related Bardet-Biedl syndrome
(8) *MKKS*-related McKusick-Kaufman syndrome

As noted above, Bardet-Biedl syndrome has been associated with pathogenic variants in many genes, two of which (*BBS1* and *MKKS*) are included here as examples (*BBS1* [MIM: 209901], Bardet-Biedl syndrome 1 [MIM: 209900]; *MKKS* [MIM: 604896], Bardet-Biedl syndrome 6 [MIM: 605231]; McKusick-Kaufman syndrome [MIM: 236700]). To our knowledge, there are no currently recognized, clinically distinct phenotypic subtypes of Bardet-Biedl syndrome. Thus, there is no distinction between the phenotypic descriptors in entities 6 and 7, even though the genes are distinct. Entities 6 and 8 do not share a gene or a phenotype label, although some would most likely consider them to be in the same gene-disorder family. Entities 7 and 8 share a gene association, but as in entities 4 and 5, the two phenotypes are considered sufficiently distinct to warrant classification as two unitary, distinct Mendelian disorders, again, given current knowledge because the number of clinical reports of individuals with McKusick-Kaufman syndrome is small.

In some cases, the determination that a diagnostic entity is distinct from another may be based on a combination of distinct phenotype and distinct mechanism, as is the case for Pallister-Hall and Greig cephalopolysyndactyly syndromes. In others, it may be based only on phenotypic distinction, as is the case for McKusick-Kaufman syndrome and Bardet-Biedl syndrome—mechanistic knowledge is not necessary for such a distinction, although it can be considered. Another example would be the various pathophysiologic mechanisms of cystic fibrosis variants. There are six classes of variants in this gene associated with the phenotype of cystic fibrosis (MIM: 219700) (absent protein, defective processing, defective regulation, defective conduction, reduced transcript level, and instability at the cell surface).[5] The field does not consider these various mechanisms sufficiently distinct to warrant distinct diagnostic descriptors, so all would be labeled similarly.

(9) *CFTR*-related cystic fibrosis

However, cystic fibrosis is pleiotropic and individuals with various *CFTR* (MIM: 602421) variants and genotypes can be affected with clinically distinct phenotypes, such as chronic pancreatitis and congenital bilateral absence of the vas deferens (CBAVD [MIM: 277180]), and have no other signs of cystic fibrosis.[6,7] These could warrant distinct dyadic descriptors.

(10) *CFTR*-related chronic pancreatitis
(11) *CFTR*-related CBAVD

Finally, we cite an example from GeneReviews that uses the dyadic concept of this proposal but has some challenges requiring more in-depth discussion.

(12) *ARID1B*-related disorder

This entity, posted in May 2019 in GeneReviews, suggests that there are descriptors in that volume that may benefit from modifications. Although it initially appears to follow the general structure that we advocate, on closer inspection, it does not. "Disorder" is not a valid phenotypic descriptor. In effect, this descriptor is not dyadic because it appears to define the entity solely by the gene that is mutated. We also note that there is a separate entity in GeneReviews with the descriptor of "Coffin-Siris syndrome" (MIM: 135900). In this separate entry on Coffin-Siris syndrome (CSS), it is stated that the CSS phenotype is associated with variants in six genes, of which *ARID1B* (MIM: 614556) is the most common (see Vergano et al. GeneReviews in Web resources). And in the entry for *ARID1B*-related disorder, they state "*ARID1B*-related disorder constitutes a clinical continuum, from classic Coffin-Siris syndrome to intellectual disability (ID) with or without nonspecific dysmorphic features." Therefore, "Coffin-Siris syndrome" associated with *ARID1B* variants must be considered a subset of "*ARID1B*-related disorder" and the former is not distinct from the latter. By themselves, the descriptors "Coffin-Siris syndrome" and "*ARID1B*-related disorder" do not meet our definition of unitary, distinct, Mendelian entities. This conundrum could be addressed via one of two approaches. We are not advocating for one or the other approach for this particular disorder and provide them only as an illustration of approaches that could be used. First, one could define a single spectrum disorder.

(13) *ARID1B*-related Coffin-Siris spectrum disorder

This spectrum could encompass the complete phenotypic range from non-syndromic ID to classic Coffin-Siris syndrome. The second approach could be to designate two entities.

(14) *ARID1B*-related intellectual disability with congenital anomalies spectrum
(15) *ARID1B*-related Coffin-Siris syndrome

The "Coffin-Siris syndrome" phenotype could be precisely defined to include a minimum set of diagnostic criteria (anomalies and ID) (see Vergano et al. GeneReviews in Web resources).[8] Individuals not meeting those criteria but having some of the features and a pathogenic variant in *ARID1B* would be described as having "*ARID1B*-related intellectual disabilities with anomalies spectrum," while those meeting stringent phenotypic syndrome criteria would be diagnosed with *ARID1B*-related Coffin-Siris syndrome. Others might argue that there should be the following two entities in addition to entity 15:

(16) *ARID1B*-related intellectual disability
(17) *ARIB1B*-related intellectual disability and non-specific anomalies syndrome.

Others might argue that either entity 15 or 17 are sufficient alone because the distinction between the two is clinically unimportant. These are difficult judgements that can only be made by consensus of expert clinicians. Importantly, the dyadic approach clarifies that this is an issue of phenotypic lumping and splitting, and it separates the entity from other forms of Coffin-Siris syndrome that could be designated, such as *SOX11*-related Coffin-Siris syndrome (*SOX11* [MIM: 600896]). Because it can take some time to work out some of these challenges, we would not object to interim or placeholder entities such as "*ARID1B*-related disorder," created for heuristic and editorial purposes until consensus is developed on these questions. But by specifying that these are distinct entities, the dyadic approach can clarify and catalyze that debate.

The dyadic approach is amenable to disorders caused by mosaic variants in single genes. Recently, one of us has piloted this approach with Proteus syndrome[9] (MIM: 176920). We found this to be useful because there is a continuous range of variability of the Proteus syndrome phenotype, which is attributable to the timing and tissue of origin of the somatic variant. We recognized that there needed to be an arbitrary but clear lower bound of the phenotype designated as Proteus syndrome, as well as distinguishing *AKT1*-related overgrowth from *AKT1*-related cancers (*AKT1* [MIM: 164730]). We developed a points-based approach to this entity and designated individuals with ten or more points as having

(18) *AKT1*-related Proteus syndrome.

Those with two to nine points would be designated as having

(19) *AKT1*-related overgrowth spectrum.

We asserted that there was a valid clinical distinction of "Proteus syndrome" from the less-specific designation of "overgrowth spectrum" despite the fact that this distinction is arbitrary. This is the same challenge that was addressed above regarding entities 14–17. It has always been challenging to define clinically meaningful boundaries between phenotypic entities. The dyadic approach does not solve that problem but clarifies the debate.

Entity 19 was intentionally similar to

(20) *PIK3CA*-related overgrowth spectrum.

This recognizes that it can be difficult to distinguish mildly affected individuals with mosaic overgrowth associated with *AKT1* versus *PIK3CA* variants[10] (*PIK3CA* [MIM: 171834]), although when properly designated as a dyad, entities 19 and 20 are unitary and distinct disorders. We propose that this could be extended to other disorders, including those that can manifest either as mosaic or constitutional disorders.

Discussion

The taxonomy of human disease represents an extraordinarily complex, nuanced, and confounding admixture of concepts, labels, and history that continues to evolve over time without explicit guidance or structure. A comprehensive and completely rational taxonomy with internal consistency and external validity is practically impossible as it would require complete knowledge of physiology and pathophysiology, which is far into the future if it is ever to be achieved. Nevertheless, it is essential and tractable to have a taxonomy that represents current knowledge and can sustain amendments as knowledge improves. The consequence is that the taxonomic description of human disease is heuristic—a descriptive system that is close to current understanding of biological reality and useful to practicing clinicians, even if it is neither comprehensive nor precise. Adding to the complexity is the fact that essentially all biologic variation is continuous, not categorical, while in medicine, it is necessary to have categories because categorical descriptors are used for making prognoses and selecting treatments. Fitting a system with continuous variation into a categorical taxonomic structure means that the latter cannot perfectly represent the former. Diagnoses are categories of pathophysiology, and while these categories are useful and essential, there are inherent challenges with variation within them and overlap among them. Within a given diagnostic entity, there will be variation. Among diagnoses, there must be defined boundaries so that one diagnosis is distinct from another.

We have adopted an approach to the taxonomy that is applicable to Mendelian diseases, defined as those in which the majority of the phenotypic variance can be explained by a genotype comprising pathogenic variants (one or two, depending on inheritance pattern) in a single gene. We have also taken the step of defining what we think should be the leaf on this taxonomic tree. As noted above, this is a categorical descriptor of pathology that is unitary and distinct. We suggest that these two attributes are critical to what we are doing as diagnosticians, which is to lump and split disease entities in a rational manner. That they are unitary is critical—we do not want a single disease descriptor to encompass more than a single disease or pathophysiologic entity. That they are distinct is equally important—we do not want to have multiple labels for what is understood to be a single pathophysiologic entity.

This dyadic approach to Mendelian disease taxonomy does not resolve the question of what are meaningfully distinct phenotypes, how they should be named, and how one designates distinct, recognizable phenotypes from clinical spectra, which is a challenge that existed well before this nomenclature system was introduced and is not resolved by it. While we prefer designation of recognizably distinct phenotypes, we recognize that is not always possible or optimal. For example, there are many genes associated with each of the following non-specific

phenotypes: autism, ID, and pigmentary retinopathy. There is little reason at this time to think that these categories of phenotypes will be resolved into subgroups of recognizably distinct phenotypes. This is part of the lumping and splitting debate that is ongoing for many disorders–trading off specificity of finer designations versus their generalizability. These issues will be challenging to resolve, and various clinicians and geneticists may find it difficult to reconcile their views. Some value the insight of broad, overarching umbrella descriptors that unify entities, whereas others value the specificity of the narrow categories that distinguish subtly distinct phenotypes. These competing views can only be resolved by disease experts gathering to debate and reconcile their competing views and priorities and developing a consensus. Neither does this system have anything to say about the diverse nosology of syndrome naming that includes researcher eponyms (Bardet-Biedl), place names (Floating-Harbor [MIM: 136140]), family name designations (Cowden [MIM: 158350]), acronyms (CHARGE [MIM: 214800]), allegorical descriptors (Proteus), histopathological descriptors (sickle cell [MIM: 603903]), or other phenotypic descriptors (incontinentia pigmenti [MIM: 308300]). Nor does it address the concerns that some phenotype or gene names may be pejorative (e.g., hyperphophatasia-mental retardation syndrome [MIM: 239300]). This debate will most likely continue but does not affect the fundamental utility of the proposed dyadic system. An attribute of the dyadic system is that by anchoring phenotypes to genes, one can more readily version and track dyadic entities as our genetic understanding of them and our phenotypic naming conventions evolve over time. Large-scale efforts to acquire deep phenotyping data, supported by the human phenotype ontology, will facilitate this evolution of our understanding of phenotypic complexity. Although less of an issue than phenotypes, gene names can also be a challenging issue for a dyadic system. They probably change less often than do phenotype descriptors, but stability of HGNC/HUGO-approved gene names is important for a dyadic system to be useful. Like phenotypes, gene names can be pejorative (e.g., sonic hedgehog, *SHH* [MIM: 600725]) or reflect obsolete understanding of the function of the gene (e.g., *PAFAH1B1*), both of which were problematic before the dyadic proposal and are not resolved by it. It will be important that gene names are as stable as possible, taking into account up-to-date biology and considerations of the implications of a gene name on those who will carry it as a diagnosis.

The phenotype component of the dyadic approach described here is consistent with the recent commentary by Rasmussen and Hamosh.[11] We would agree that phenotypes should not be described with gene names—doing so would fundamentally confuse what we believe are two important but distinct attributes. We would also agree that naming of phenotypes and naming of genes should continue independently with their distinct attributes. However, it is important to recognize that the examples provided here convincingly show that a phenotype is not a unitary and distinct diagnostic entity. It is also worth emphasizing that the dyadic approach is built from the gene and phenotype foundations provided by OMIM—that those are robust and properly documented is essential to the dyadic approach. That they are separately cataloged is extremely useful in that they can be addressed and modified independently. Rasmussen and Hamosh did not address the crucial question of determining the attributes of a unitary, distinct, Mendelian disease entity. The dyadic proposal properly frames this crucial question, and we believe it is a valid and practically useful response to that crucial question but goes further than Rasmussen and Hamosh by integrating these two attributes.

It is also worth noting that debates over naming phenotypes can be distressing for the individuals who are affected by these disorders—they do not like for the name of "their" disease to be changed. The dyadic proposal has the advantage that the current phenotypic descriptor (e.g., McKusick-Kaufman syndrome) does not have to be changed; only the associated gene must be appended to the now dyadic descriptor. More substantial revisions of phenotypic descriptors should take the views of affected individuals into account.

This proposal has similarities to several prior formulations. It is a variation of the multiaxis diagnostic system that was proposed by one of us in 2001.[12] The multiaxis system had positive attributes, but it garnered little enthusiasm and was not adopted. The approach we propose here, as noted above, emulates and extends the GeneReviews approach to disease description. An implicit form of this system is also intrinsic to the abbreviations describing many disorders in OMIM. In that catalog, one finds disease entries such as BBS1 and BBS8 (MIM: 615985). Here, they are encoding a phenotype of Bardet Biedl syndrome (BBS) but adding a numerical suffix to distinguish them on the basis of a different genetic cause. Unfortunately, this can be confusing because in some cases the genotype-phenotype descriptor is the same as the HGNC-approved gene name (as in BBS1, which is associated with pathogenic variants in *BBS1*) but in others it is distinct (as in BBS8, which is associated with pathogenic variants in the gene *TTC8* [MIM: 608132]). We suggest that the explicit use of the gene name in our dyadic approach is simpler and easier to understand and remember and facilitates the gathering of additional information from sources that contain the gene name as structured data rather than the numerical code suffixes of OMIM. We also note that the ClinGen consortium has organized its variant curation expert panels into dyadic groupings—a gene and an associated phenotype, recognizing as well that it is this dyad that defines a unitary and distinct entity of disease. A similar approach has been proposed for Mendelian genetic renal diseases.[13] These authors have also proposed a dyadic descriptor, "ADPKD-PKD1" for autosomal dominant polycystic kidney disease—associated with *PKD1* (ADPKD1 [MIM: 173900];

PKD1 [MIM: 601303]). These authors recognize that it is the dyad that comprises a distinct and unitary disease taxonomic entry. As they state, "We therefore propose categorization of patients with a *phenotypic and genotypic descriptor* that will clarify etiology, provide prognostic information, and better describe atypical cases" (emphasis added).[13] That they code the phenotype first and always abbreviate it is a minor difference from our proposed convention, which would designate the entity as "*PKD1*-related polycystic kidney disease" (formally, diseases do not have the attribute of being autosomal dominant—inheritance patterns have that attribute, and since PKD1 is only associated with autosomal dominant inheritance, we would endorse the OMIM phenotypic descriptor of "polycystic kidney disease" and not include the inheritance pattern). Interestingly, these authors extend the taxonomic categorization into categories of mutations/modes of pathogenesis, for example dividing the PKD1 entity into subcategories of those caused by truncating versus non-truncating *PKD1* variants: ADPKD-PKD1^{T} and ADPKD-PKD1^{NT}. Similar arguments could be made for incorporating concepts such as gain of function or loss of function into phenotype descriptors. A potential issue for the dyadic approach is that not all variants that are associated with Mendelian disorders are in genes. For example, the ZRS enhancer of *SHH* lies in an intron of the *LMBR1* (MIM: 605522) gene and, when mutated, is associated with a Mendelian disorder of preaxial polydactyly II (MIM: 174500). A recent study cataloged 453 non-coding variants that have been associated with a Mendelian disorder.[14] While this is a small number compared to the more than 150,000 pathogenic single-nucleotide variants in ClinVar (accession date September 21, 2020), this is an important class of pathogenic variants that is likely to increase in the future. While we do not propose to do so here, a consideration could be to modify the genetic etiology half of the dyad along the lines of *SHH*^{ZRS}-related preaxial polydactyly, similar to the renal disease proposal discussed above. This categorization scheme could be extended to particular alleles or haplotypes of genes or downstream molecular etiologies to further clarify and specify the molecular etiology as well. While we find this to be consistent with our general approach, we would suggest that this might be too fine of a categorization for many practitioners and could quickly become cumbersome as a result of lengthy descriptors. These considerations warrant additional study of their potential utility. We believe that there is a balance to be struck between the completeness and generalizability of a descriptor and its usability. While we readily concede that for single gene disorders, collapsing molecular etiology to an HGNC/HUGO gene name does sacrifice some information, we believe that the brevity of the gene-phenotype dyad is critical for its usability.

In conclusion, the approach we advocate for is derived from, and consistent with, multiple other conceptual proposals for categorizing mendelian genetic disorders as distinct, dyadic entities and we suggest it is fundamentally correct. What is needed now is a consistent, standardized approach for so designating such diseases, as we have proposed here.

Acknowledgments

The authors thank Robert L. Nussbaum for critical review of a draft of this manuscript. L.G.B. was supported by the National Institutes of Health (NIH) (ZIA HG200388 06). M.P.A., A.R.A., and R.A.P. are supported by a subcontract from Computercraft Corporation to the University of Washington for editing *GeneReviews*. T.C.C. was supported by NIH R01 DE027879. J.M.K. was supported by Alex's Lemonade Stand Foundation, Damon Runyon Cancer Research Foundation, and NIH K08CA193915. G.M.M. was supported by NIH K08NS092898 and Jordan's Guardian Angels and the Brotman Baty Institute.

Declaration of interests

L.G.B. receives in-kind research support from ArQule (now a wholly owned subsidiary of Merck) and Pfizer and receives honoraria from Cold Spring Harbor Press. K.W.G. is Chief Medical Officer for FDN (Face to DNA). K.R. is an employee of GeneDx, a subsidiary of Opko Health, and a shareholder of Opko Health. The other authors declare no competing interests.

Web resources

GeneReviews, https://www.ncbi.nlm.nih.gov/books/NBK1116/?report=reader

GeneReviews, Forsyth, R.L., and Gunay-Aygun, M. (2003) Bardet-Biedl Syndrome Overview, https://www.ncbi.nlm.nih.gov/books/NBK1363/

GeneReviews, Vergano, S.A., van der Sluijs, P.J., and Santen, G. (2019). ARID1B-Related Disorder, https://www.ncbi.nlm.nih.gov/books/NBK541502/

Human Gene Nomenclature Committee of HUGO, https://www.genenames.org

OMIM, https://omim.org

References

1. Bamshad, M.J., Nickerson, D.A., and Chong, J.X. (2019). Mendelian Gene Discovery: Fast and Furious with No End in Sight. Am. J. Hum. Genet. *105*, 448–455.
2. Sakai, L.Y., Keene, D.R., Renard, M., and De Backer, J. (2016). FBN1: The disease-causing gene for Marfan syndrome and other genetic disorders. Gene *591*, 279–291.
3. Gross, A.M., Frone, M., Gripp, K.W., Gelb, B.D., Schoyer, L., Schill, L., Stronach, B., Biesecker, L.G., Esposito, D., Hernandez, E.R., et al. (2020). Advancing RAS/RASopathy therapies: An NCI-sponsored intramural and extramural collaboration for the study of RASopathies. Am. J. Med. Genet. A. *182*, 866–876.
4. Amberger, J.S., Bocchini, C.A., Scott, A.F., and Hamosh, A. (2019). OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res. *47* (D1), D1038–D1043.
5. Zielenski, J. (2000). Genotype and phenotype in cystic fibrosis. Respiration *67*, 117–133.
6. Steiner, B., Rosendahl, J., Witt, H., Teich, N., Keim, V., Schulz, H.U., Pfützer, R., Löhr, M., Gress, T.M., Nickel, R., et al. (2011).

Common CFTR haplotypes and susceptibility to chronic pancreatitis and congenital bilateral absence of the vas deferens. Hum. Mutat. *32*, 912–920.

7. LaRusch, J., Jung, J., General, I.J., Lewis, M.D., Park, H.W., Brand, R.E., Gelrud, A., Anderson, M.A., Banks, P.A., Conwell, D., et al.; North American Pancreatitis Study Group (2014). Mechanisms of CFTR functional variants that impair regulated bicarbonate permeation and increase risk for pancreatitis but not for cystic fibrosis. PLoS Genet. *10*, e1004376.
8. Schrier, S.A., Bodurtha, J.N., Burton, B., Chudley, A.E., Chiong, M.A., D'avanzo, M.G., Lynch, S.A., Musio, A., Nyazov, D.M., Sanchez-Lara, P.A., et al. (2012). The Coffin-Siris syndrome: a proposed diagnostic approach and assessment of 15 overlapping cases. Am. J. Med. Genet. A. *158A*, 1865–1876.
9. Sapp, J.C., Buser, A., Burton-Akright, J., Keppler-Noreuil, K.M., and Biesecker, L.G. (2019). A dyadic genotype-phenotype approach to diagnostic criteria for Proteus syndrome. Am. J. Med. Genet. C. Semin. Med. Genet. *181*, 565–570.
10. Keppler-Noreuil, K.M., Rios, J.J., Parker, V.E., Semple, R.K., Lindhurst, M.J., Sapp, J.C., Alomari, A., Ezaki, M., Dobyns, W., and Biesecker, L.G. (2015). *PIK3CA*-related overgrowth spectrum (PROS): diagnostic and testing eligibility criteria, differential diagnosis, and evaluation. Am. J. Med. Genet. A. *167A*, 287–295.
11. Rasmussen, S.A., Hamosh, A.; and OMIM curators (2020). What's in a name? Issues to consider when naming Mendelian disorders. Genet. Med. *22*, 1573–1575.
12. Robin, N.H., and Biesecker, L.G. (2001). Considerations for a multiaxis nomenclature system for medical genetics. Genet. Med. *3*, 290–293.
13. Cornec-Le Gall, E., Torres, V.E., and Harris, P.C. (2018). Genetic Complexity of Autosomal Dominant Polycystic Kidney and Liver Diseases. J. Am. Soc. Nephrol. *29*, 13–23.
14. Smedley, D., Schubach, M., Jacobsen, J.O.B., Köhler, S., Zemojtel, T., Spielmann, M., Jäger, M., Hochheiser, H., Washington, N.L., McMurry, J.A., et al. (2016). A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease. Am. J. Hum. Genet. *99*, 595–606.